



OPEN Using pet insurance claims to predict occurrence of vector-borne and zoonotic disease in humans in the United States

Janice O'Brien¹, Aliya McCullough², Christian Debes³ & Audrey Ruple¹✉

Taking a One Health approach to infectious diseases common to both dogs and people, pet insurance claims from 2008 to 2022 in the United States were compared to publicly available CDC-based data on human cases for Lyme disease, giardia, and Valley Fever (coccidioidomycosis). Despite having very different causative agents and etiologies, the disease trends for these three diseases were very similar between people and dogs both geographically and temporally. We furthermore demonstrated that adding dog data as a predictor variable in addition to the human data improves prediction models for those same diseases when investigating incidence over time. With machine learning prediction tools for the pet insurance to predict changes in disease incidence sooner and give public health officials more time to prepare, pet insurance data could be a helpful tool to predict and detect diseases by estimating even earlier the effects of these common exposure diseases on human health. We also show the spatiotemporal distribution of intestinal worm diagnoses in dogs, and while it could not be directly compared to human data because the corresponding disease in humans (soil-transmitted helminths) has not been well monitored recently. However, these data can help inform researchers and public health workers.

In today's highly interconnected world where human, animal, and environmental health are deeply related, novel disease detection methods are crucial. Identifying emerging and re-emerging diseases at their earliest stages enables earlier public health interventions and therefore impacts more lives. Additionally, monitoring and modeling diseases provides insights into the epidemiology of these diseases, and doing so in multiple species helps to reveal the dynamics of the diseases in these different populations. The animal in the United States that shares the most environmental exposures with humans is arguably pet dogs.

Dogs have been explored as a sentinel of human disease across many different disease types: vector-borne pathogens,¹ toxin exposures,² and cancer³ This is well covered by Sexton in the exploration of the shared exposome of people and dogs⁴ Here we examined them as a potential sentinel for four infectious diseases with different types of pathogens and different etiologies: Lyme disease, giardiasis, Valley Fever (coccidioidomycosis), and soil-transmitted helminths.

Lyme disease is caused by the *Borrelia burgdorferi* bacterium transmitted by Ixodes ticks. It was first described in 1977 in Connecticut, but was present in ticks collected in Long Island in the 1940s⁵ and in Massachusetts white-footed mice collected in 1894,⁶ and so has been plaguing the northeastern United States for centuries. Since its recognition and development of appropriate testing methods in both dogs and people, Lyme disease has been apparently spreading geographically within the United States due to the spread of the Ixodes tick vector into new areas⁷ The Ixodes spread has been shown to be secondary to factors like global climate change,⁸ spread through bird migration patterns,⁹ and deforestation and reforestation patterns¹⁰ In humans, Lyme disease is frequently mis-diagnosed. In one study by Ferrouillet, only 40% of human general practitioners could accurately diagnose an erythema migrans rash¹¹, which is considered pathognomonic for the disease. In another study by Gasmi,¹² the misdiagnosis of erythema migrans rashes was 62.8%. This frequent misdiagnosis leads to people living with the disease for prolonged periods causing severe morbidity, and can even lead to the potentially lethal condition of Lyme carditis.¹³ Understanding the misdiagnosis rates is important to improve the clinical outcomes of those with this disease, and early detection methods for the movement of this disease into new areas or for detecting annual spikes in cases is one important way to improve public health agency preparedness

¹Department of Population Health Sciences, Virginia-Maryland College of Veterinary Medicine, Virginia Tech, Blacksburg, VA, USA. ²Fetch Insurance Services, LLC, New York, NY, USA. ³Spryfox GmbH, Darmstadt, Germany. ✉email: aruple@vt.edu

and increase both clinician and public awareness. In dogs, Lyme typically causes polyarthritis, fatigue, and loss of appetite. Untreated Lyme disease in dogs can lead to Lyme nephritis and death. The disease is zoonotic, but it cannot be shared directly between people and dogs – bites from infected ticks spread the disease. However, dogs and people living in the same areas share exposure to these ticks. Previous work in modeling Lyme disease in people has focused mostly on the tick vectors and their movements^{14–16}. Modeling Lyme in dogs has focused necessarily on seropositivity not cases because of the inherent difficulty in determining the population at risk for surveillance testing data¹⁷. Comparative analyses between Lyme disease in dogs and people has used total prevalence not incidence rates across time^{18,19}.

Giardiasis is a diarrheal disease caused by a protozoan parasite transmitted by the fecal-oral route, particularly via water contaminated with infected feces. Giardiasis in people and dogs is typically caused by different assemblages of the organism. Humans are usually infected with A and B, while dogs typically contract C and D, although both species can contract each other's assemblages so zoonotic infection can occur²⁰. In both species, it is most frequently diagnosed in the young – children and puppies – and can be both severe and asymptomatic, depending on the individual infected. Both people and dogs that are otherwise compromised, such as those with weakened immune systems or the very young and very old, are more at risk for severe complications and death from the disease.

Valley Fever (coccidioidomycosis) is caused by inhaling spores of the *coccidioides* fungus, which typically grows in the soil. Dogs and people are infected by the same agent, and the disease course is very similar between the two species with primary infection occurring in the lungs, which can spread to other areas of the body. The infectious form is only found outside the body, so the disease is not contagious between individuals of either species, except in very rare cases. In people, the disease is often mis-diagnosed as a viral or bacterial infection, or even metastatic cancer,²¹ causing significant morbidity for those with delayed diagnoses. Previous studies have studied dogs as a possible sentinel of this agent when it moves into new geographic areas, as it has been proposed that dogs with digging and dirt-sniffing behaviors are more exposed to the disease than humans typically are. In Washington, cases were first identified in dogs before they were identified in people²². In Texas, a study of 6087 of dogs used serological testing to identify geographic areas of exposure²³. A study conducted at the Veterinary Medical Teaching Hospital at University of California, Davis compared the geographic distribution of exposures among both dogs and people in California which were nearly identical²⁴. Despite all these previous studies proposing dogs as a sentinel for coccidioidomycosis, none have shown a direct comparison between human and canine incidence rates over time.

Soil-transmitted helminths are considered a neglected parasitic disease, and are the most common neglected tropical disease globally. Because of the morbidity caused by these infections, they have the capability to perpetuate poverty²⁵. Soil-transmitted helminth infections are thought to be rare in the USA, however recent work has demonstrated that there are likely persisting pockets of infection^{26,27}. The worms which cause disease in humans can come from many animals: wild canids, both feral and domestic cats, racoons, pigs, but domestic dogs are hosts to many of these parasites (*Ancylostoma caninum*, *A. braziliense*, *A. ceylanicum*, *Uncinaria stenocephala*, and *Toxocara canis*). When infected dogs defecate and those feces are left in the environment, contact or ingestion of the infective eggs or larvae can cause these infections in humans. The symptoms people experience are due to the bodily location in which the worms are migrating – through the skin, eyes, abdomen, etc. – and these migrations can cause significant health problems for the infected persons.

Dogs have previously been demonstrated as a sentinel of disease for people for several infectious diseases, so using pet insurance information to monitor the rates of these diseases in dogs is a reasonable step in a One Health approach to preventing and managing these diseases. Examining pet insurance data for disease insights and outcomes is beneficial because pet owners with pet insurance are more likely to opt for veterinary recommended diagnostics and treatment^{28,29} and less likely to elect economic euthanasia^{30,31}. Therefore, the clinical picture of disease in the pet is more likely to be complete.

The purpose of this study was to examine a diverse category of diseases across both humans and dogs to determine whether pet insurance claims and information could be a potentially useful tool in the One Health approach to modeling these diseases. To that end, these diseases were compared between dogs and people using pet insurance data and publicly available data from the Centers for Disease Control, respectively. They were compared geographically, seasonally, and using incidence over time. Simple general linear models were used to model the nationwide yearly incidence rates to assess whether adding dog data as a predictor improved the linear model as compared to human-only data.

Results

Pet population and diagnosis data

1,254,522.5 dog-years were represented in this population of insured dogs between 2008 and 2022, with 1,743,709 unique dogs. Among those dogs, there were a total of 36,419 diagnoses in the infectious conditions of tickborne disease (21%), giardia (37%), fungal disease (15%), and intestinal parasites (26%).

Within tickborne diseases, the diagnosis codes were: Lyme disease (3856, 49%), Babesiosis (37, 0.5%), Ehrlichiosis (763, 10%), Anaplasmosis (1367, 18%), and tick bite (1779, 23%). Within the intestinal parasites category, the top four diagnosis codes represented were: internal parasites (2885, 30%), coccidiosis (1560, 16%), hookworm (1324, 14%), and tapeworm (1286, 13%). Within the fungal disease group, the three diagnosis codes represented were: Valley Fever (Coccidioidomycosis) (4413, 80%), aspergillosis (241, 4%), and fungal infection (885, 16%).

The geographic distribution of the pet insurance dataset is generally representative of the US population density (see Fig. 1). The age distribution of the insured population (see Table 1) was more heavily weighted toward younger dogs (the puppy and young adult age groups are the largest groups). There were more owner-reported purebred dogs compared to mixed breed dogs (63–36%).

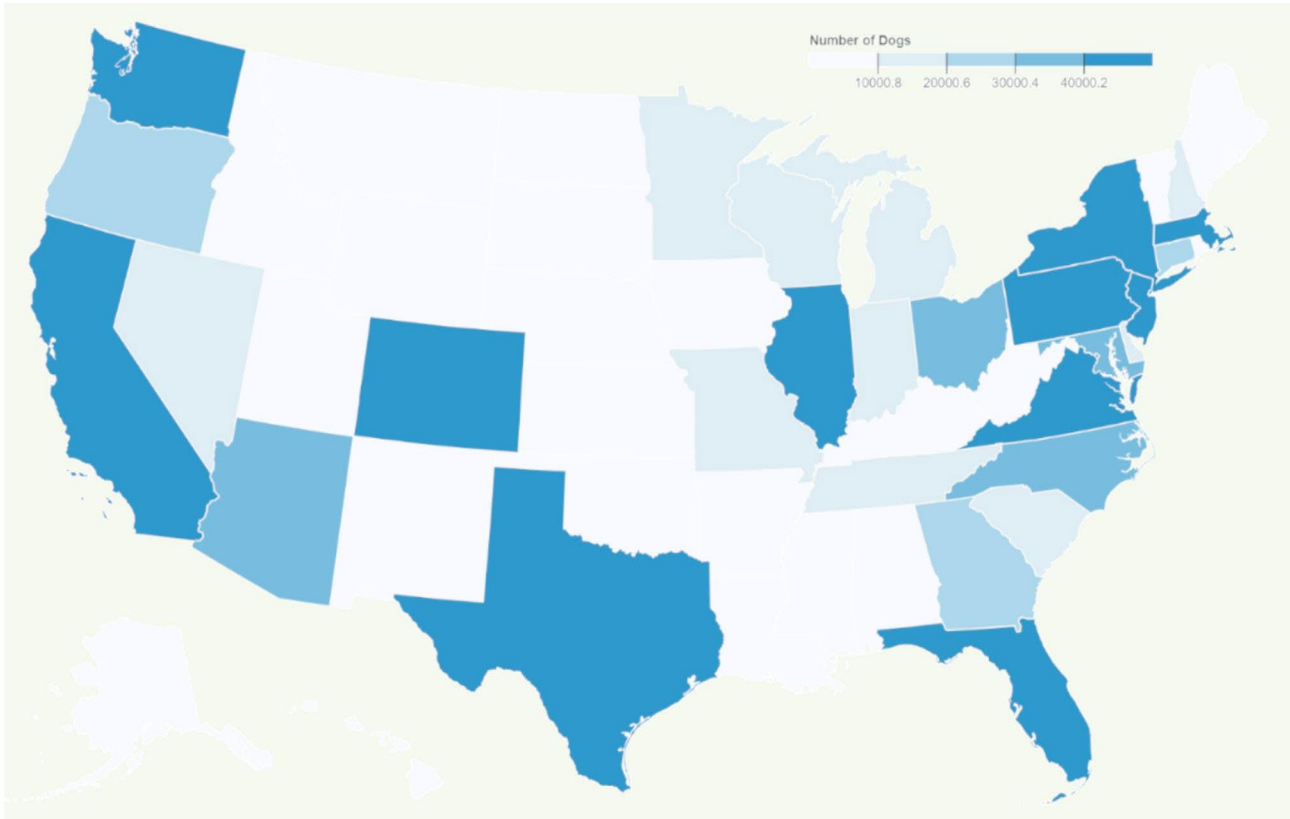


Fig. 1. Geographic distribution of the dogs in the pet insurance dataset.

	Category	Dogs diagnosed with Tickborne (7802)	Dogs diagnosed with Giardia (13489)	Dogs diagnosed with fungal (5539)	Dogs diagnosed with intestinal parasites (9589)	Total population (1743709)
Life stage	Puppy	26% (2025)	83% (11204)	12% (646)	68% (6525)	42% (382853)
	Young Adult	43% (3336)	12% (1571)	46% (2548)	21% (2045)	37% (338087)
	Mature Adult	22% (1723)	4% (485)	31% (1694)	7% (699)	15% (133982)
	Geriatric	9% (718)	2% (229)	12% (651)	3% (320)	6% (54240)
Breed	Purebred	60% (4715)	60% (8048)	65% (3594)	57% (5469)	63% (1105223)
	Mixed	40% (3087)	40% (5441)	35% (1945)	43% (4120)	36% (638481)

Table 1. Distribution of dogs in the pet insurance dataset by life stage and breed status across the disease categories.

Pet demographics across disease categories

Examining the changes in the affected population across the disease categories reveals both anticipated and unexpected findings, as shown in Table 1. Puppies were overwhelmingly the largest group diagnosed with Giardia and intestinal parasites. Young adult dogs were the largest group diagnosed with tickborne and fungal diseases. There does not appear to be a difference between pure- or mixed- breed status and diagnoses with any of the diseases. There were some small differences between diseases, but in general the percentages of purebred were approximately 60% and mixed breed were approximately 40%.

Comparison to human disease trends

Here within each overall disease category, we highlight a specific disease of interest.

Lyme

The overall incidence rate of Lyme disease in dogs is approximately 44 times higher than that of humans, with the average human incidence rate nationwide averaging 10.7 per 100,000 persons between 2008 and 2020, and the average nationwide incidence in dogs averaging 471.3 per 100,000 dogs between 2008 and 2020 for this population of insured pet dogs. Dogs have a much longer season for the disease – starting in March and ending in September – while people mostly appear to have diagnoses starting in May. Dogs appear to have a late fall

spike in cases in October, while this spike was absent in the human Lyme cases. The geographic distribution of Lyme cases is nearly identical for the northeastern states and Minnesota/Wisconsin, but notably the incidence rate among dogs is higher in adjacent states like South Dakota, Illinois, Ohio, Michigan. The human and canine Lyme disease rates appear to follow the same general trends (looking at both national rates and for the top-15 states), with one important difference: that the canine rate changes appear to precede the changes in humans by approximately 1–2 years (see Fig. 2).

Giardia

Human and Canine *Giardia* incidence trends appear to follow a similar pattern, where canine cases peak in 2009, 2011, and 2015, and human cases peak in 2010 and 2016. Both species experienced lower relative incidence rates in 2013 and 2017 (see Fig. 3).

The insurance data from dogs show that dogs have a peak incidence in the spring - specifically in March, and the incidence slowly decreases from there. Similarly to how canine Lyme cases have a more spread out seasonality compared to human Lyme cases, canine giardia is much more spread out compared to human cases throughout the year. Humans have a season starting in May, and then peaking in August. The geographic distributions of diagnoses also appear similar during these years - with the northeast and northwest being most impacted and the southeast less impacted. Overall the *Giardia* incidence rate among insured dogs has been gradually increasing, with a trough year in 2015.

Coccidioidomycosis (valley fever)

The geographic distribution of *Coccidioidomycosis* cases in dogs is nearly identical to that in humans, with California and Arizona being the hot-spots within the US. The trend lines for human and canine cases across years appear to be matching each other (see Fig. 4), with peaks in both species occurring in 2011, and troughs occurring in 2014–2015. Data after 2019 for human cases was not publicly available at the time of analysis.

Intestinal parasites

The intestinal worm incidence rates were highest in New York, New Jersey, South Carolina, Oregon, and Hawaii (see Fig. 5). Overall the incidence rates have been declining since 2008, with some increases in 2015 and 2020. Diagnoses among dogs are less common in the winter months and more common in the summer, with peak diagnoses occurring in July. The authors also attempted to compare incidence rates for canine intestinal parasites in dogs to the incidence rates for the human conditions caused by these parasites (termed soil-transmitted helminths). Unfortunately the most recently conducted studies of soil-transmitted helminths in the human population were performed in the 1980s, so recent endemic incidence rates for people are not readily available^{19,26}.

Machine learning

The models showed that the mean absolute error (MAE) for predictions across all the years of data available decreased for each of the models when dog incidence data was added to the model (see Table 2). The amount that the MAE is reduced in each case varies, but the addition of dog data improves the model predictions in every case.

Discussion

Four very different pathogens were investigated: a vector-borne bacterium, an airborne soil fungus, a water-borne protozoa, and soil-transmitted helminths. The transmission routes in people and dogs are diverse between these diseases, and yet each disease for which human data was available demonstrated a similarity in disease trends between humans and dogs. Unfortunately, human data on soil-transmitted helminths in the United States is sparse after the 1980s³² so a direct comparison between canine and human incidence rates was not possible. However, since dogs are the reservoir host for many of the soil-transmitted helminths, active monitoring of intestinal worm diagnoses in dogs would likely assist researchers in the surveillance of these neglected parasitic diseases.

Examining both intestinal pathogens in the canine insurance data, puppies were overwhelmingly the largest group diagnosed with giardia and intestinal parasites, which is consistent with previous studies regarding the prevalence of intestinal parasites in puppies³³. The overall differences in life stages diagnosed with different diseases is helpful to understand which dogs are most susceptible in the epidemiology of these disease groups and can inform both public health and veterinary practitioners. The incidence rate of intestinal worm diagnoses is overall decreasing, which may be due to increased use of monthly preventative medications. The increases in 2015 and 2020 may have been due to momentarily decreased preventative use due to economic factors (e.g. the 2015 stock market selloff, or the COVID-19 pandemic, shutdowns, and decreased access to well care for pets). The rise in diagnoses in summer months may correspond with more pets being outdoors and therefore being more exposed to infection, or also environmental temperatures being conducive for transmission.

With coccidioidomycosis, the peaks and troughs in people and dogs occur in the same years, which is logically consistent with the method of transmission: environmental conditions influence the levels of fungal spores in the air, and when the spores are more present, more dogs and humans are infected by this single agent. Here we successfully showed that incidence rates in both species are temporally similar, suggesting that at least for endemic areas spikes in canine coccidioidomycosis rates should correlate with spikes in human rates during the same years. Monitoring cases for both species could make the detection of epidemic years more sensitive. It was also demonstrated that amongst this pet insurance dataset, the geographic trends of the disease are identical to the geographical set in humans.

In spite of the difference in giardia assemblages primarily responsible for disease in both people and dogs, the overall incidence rates for both species from 2008 to 2020 follow similar patterns with peaks and troughs in the

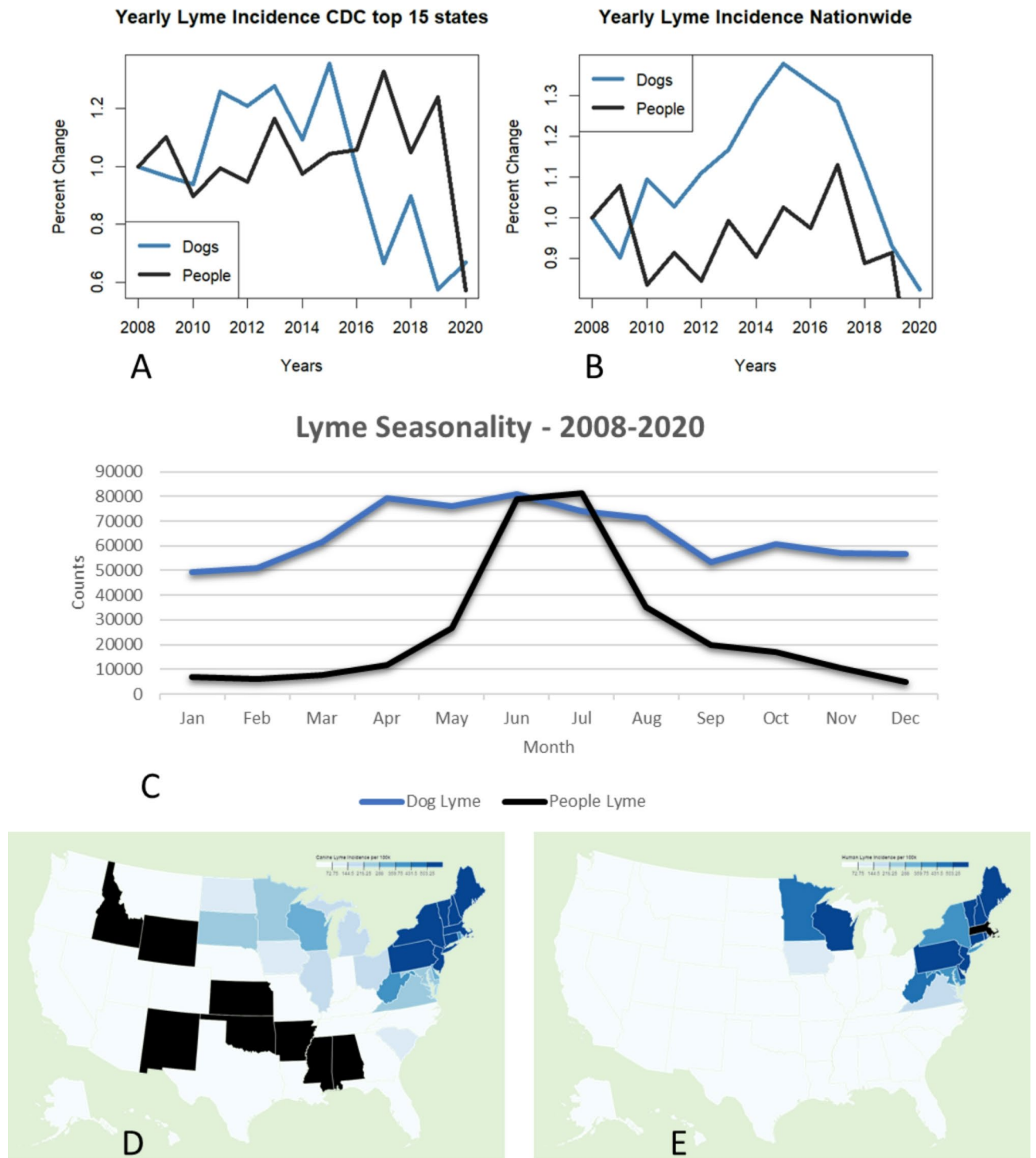


Fig. 2. Yearly Lyme incidence rates for both humans and dogs for (A) CDC top-15 Lyme states (B) Nationally graphed relative to their incidence in 2008 (C) Seasonal distribution of diagnosed Lyme case counts – 2008–2020 data overlaid. Canine cases multiplied by 200 so graphs can be compared relatively (D) dog and (E) human geographic distribution incidence per 100k in the United States (black-colored states denote absence of data).

trends occurring in similar years. Likely, the conditions that allow for canine giardia to cause a spike in infections are likely the same environmental conditions that allow for human giardia infections to spike within the same year. Conveniently, canine giardia is seasonally highest in spring before human infections begin to rise in May. If it can be demonstrated through future research that spikes in canine giardia in spring precede a spike in human giardia in the same year, active monitoring of canine giardia diagnoses could help public health professionals

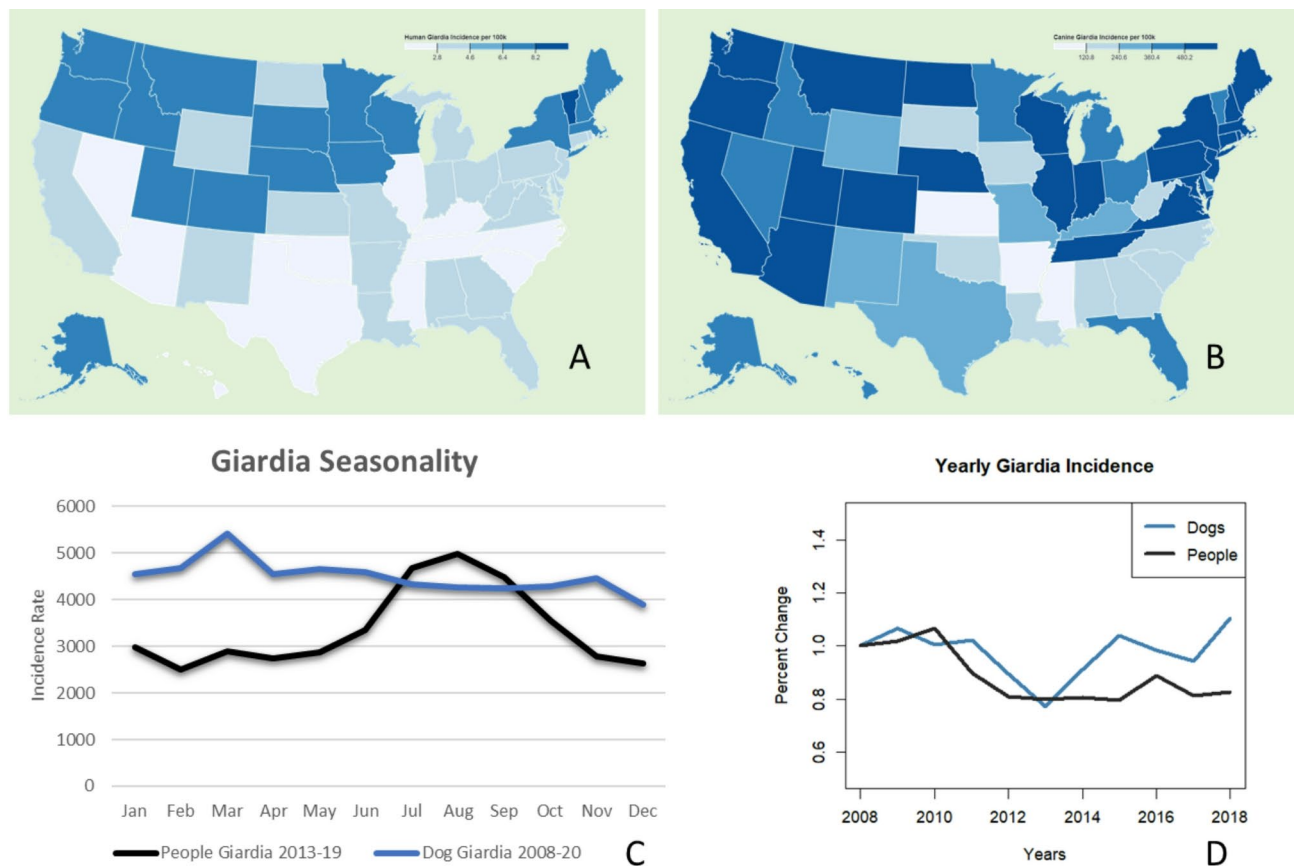


Fig. 3. Geographical distribution of Giardia for (A) dog incidence rates and (B) human incidence rates. (C) Seasonal trends in Giardia in humans and dogs – canine incidence rate per 1000, human incidence rate per 100,000. (D) Yearly trends in Giardia incidence in humans and dogs.

to implement interventions to protect people with weeks to months of warning.—It is unclear whether a high Giardia spring season for dogs precedes a high Giardia Summer season for humans, but if this is the case, it is promising again that pet insurance data could be a surveillance tool for human public health.

Of all diseases investigated, Lyme disease trends in the pet insurance data presented the largest possible benefit to public health. Unlike coccidioidomycosis and giardiasis, canine Lyme incidence trends appear to precede human Lyme incidence trends by two years at both the national and top-15 state level. Furthermore, the overall incidence rate in dogs is magnitudes higher than the rate in people, making spikes in the disease within this species easier to detect. If active monitoring of canine Lyme diagnoses can give public health agencies a 2-year lead time on emergences of Lyme, there is the potential that budget planning can earmark funds specifically for Lyme disease interventions. The reason for this 2-year gap in trends is unclear through this analysis: perhaps the 2-year lifecycle of the Ixodes vector plays a role, perhaps the 2-year lag is representative of the length of time required to obtain a diagnosis in people, or some other reason. Regardless of the cause for the 2-year difference, knowing canine Lyme trends can help public health agencies to intervene in advance to improve human health.

The use of pet insurance data in this study is beneficial as it is a large dataset of dogs used to achieve statistically significant results. Insured dogs also have varied breed heritage including purebreds and mixed breeds, live in varied geographic locations, and visit different veterinary practice types (i.e. general practice and referral care facilities). This allows comparison of disease across these variables to paint a cohesive picture for infectious disease modeling and management. For example, this insurance dataset was able to provide comprehensive temporal and geographical data for dog Valley Fever diagnoses without reliance on reference laboratory results and veterinary teaching hospital records, which present additional barriers to data acquisition.

Insured dogs in the data set tended to be puppies and young adults. This is likely due to the lower premium cost for younger dogs and the low probability of a pet having a pre-existing condition. Most pet insurers do not cover medical conditions that occurred prior to the inception of the policy.

There are drawbacks of using pet insurance data as well. Pet owners with pet insurance are able to afford both the premium and the reimbursement payment model and have access to veterinary care. Therefore, opting to use pet insurance data may impact the representation of certain demographics for example, residing in a rural area and retirement status of pet owners may impact whether a pet owner has pet insurance³⁴ This demographic information is important to account for because knowing how and when pet owners seek veterinary care has important implications for the prevention of the spread of infectious disease. For these demographic reasons, perhaps insurance is not the best tool to monitor soil-transmitted helminths in under-served areas.

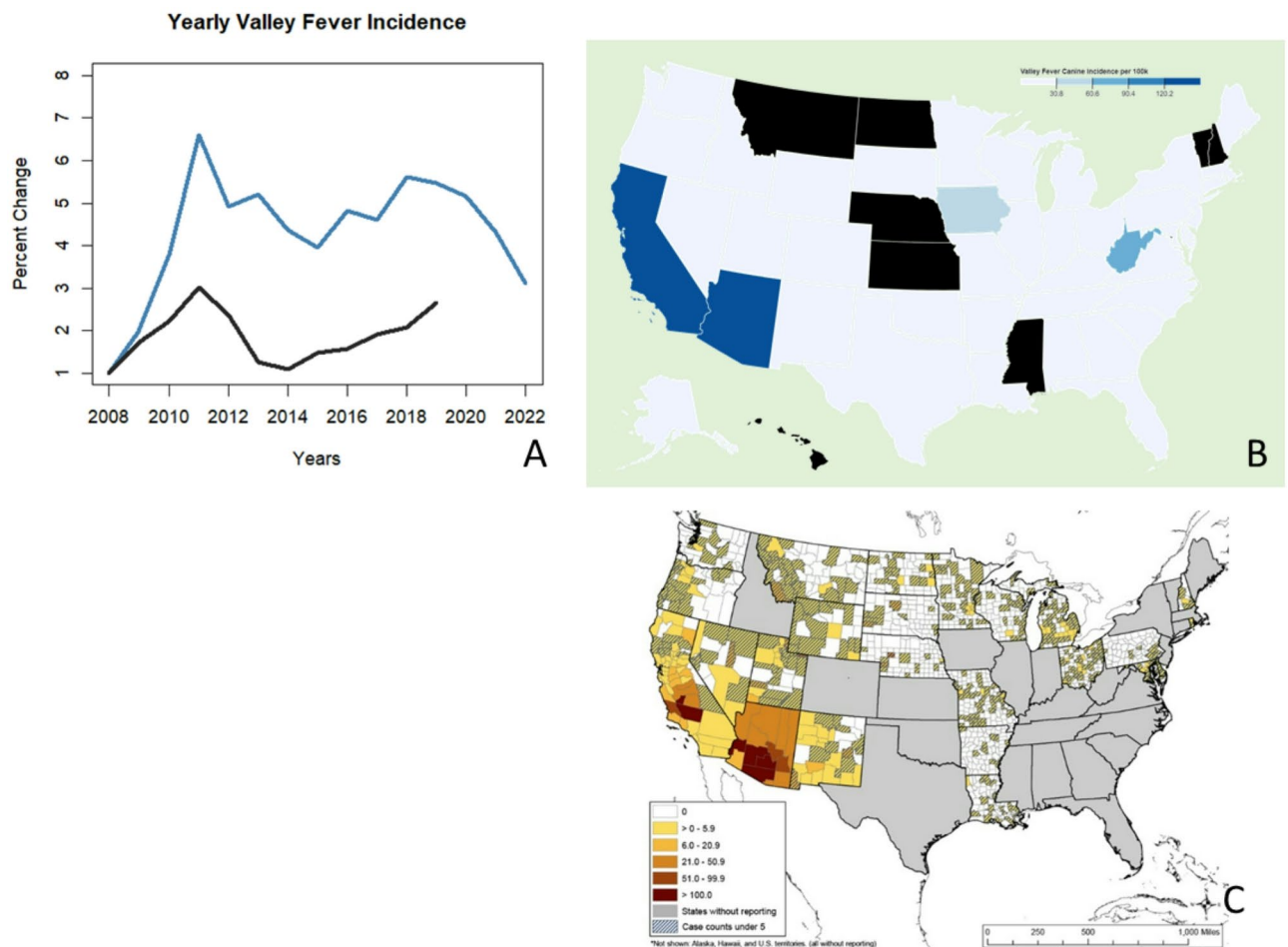


Fig. 4. (A) Yearly Coccidioidomycosis trends in dogs and humans, (B) case counts of the geographic distribution of coccidioidomycosis, (C) county-level human data from CDC (black or grey-colored states denote absence of data).

Additionally, some cases of disease may go unreported because a policyholder did not file a claim after seeking veterinary care. Cases could also be missed if the date of the veterinary invoice occurred prior to the policy inception date, during the policy waiting period, or beyond the 90 day claim submission deadline. In these cases, a condition code may not be assigned. In spite of these drawbacks to pet insurance data, it preliminarily appears to be a very useful tool for monitoring at least three diseases that both humans and dogs share.

The general linear models that were trained with both human and dog incidence information had a lower error score, indicating that across all prediction years, adding dog information to the model improves the predictions overall. These models were very simple, given that there were only a decade worth of data for both humans and dogs, and the yearly incidences were a single data point for each year. It would be valuable to further demonstrate with incidence dates across human and dog data (and not just overall yearly rates) whether the dog data improves the human prediction models for these diseases more granularly – especially given the difference in seasonality between dogs and humans for both Lyme and Giardia.

The selected linear model is based on the assumption of a linear relationship between the predictor variables and the outcome variable, as well as normally distributed residuals with constant variance. However, these assumptions may not fully capture the complexity of disease dynamics. Considering the limited dataset (annual incidence rates over a ten-year period) and the primary objective of the modeling effort (to demonstrate improvement by incorporating dog-related data), the linear model was deemed suitable for this analysis.

Fetch pet insurance currently utilizes a machine learning algorithm which predicts disease probabilities based on a pet's information. This algorithm is individual pet focused at the moment, directed at improving the customer's experience. However this algorithm could be retooled to be a syndromic surveillance database with population-level predictions for diseases within the pet dog population. If such a tool can accurately model and predict these shared diseases within the dog population, areas of emergence and times of re-emergence could be quickly detected using pet insurance data, and used to alert human health practitioners and public health workers. If active monitoring of pet insurance data can detect, or even predict, bad seasons for canine giardia, Lyme, soil transmitted helminths, or Valley Fever, health departments could be alerted to the issues before human cases begin rising excessively. In the case of Lyme disease, this alert could be several years in advance of

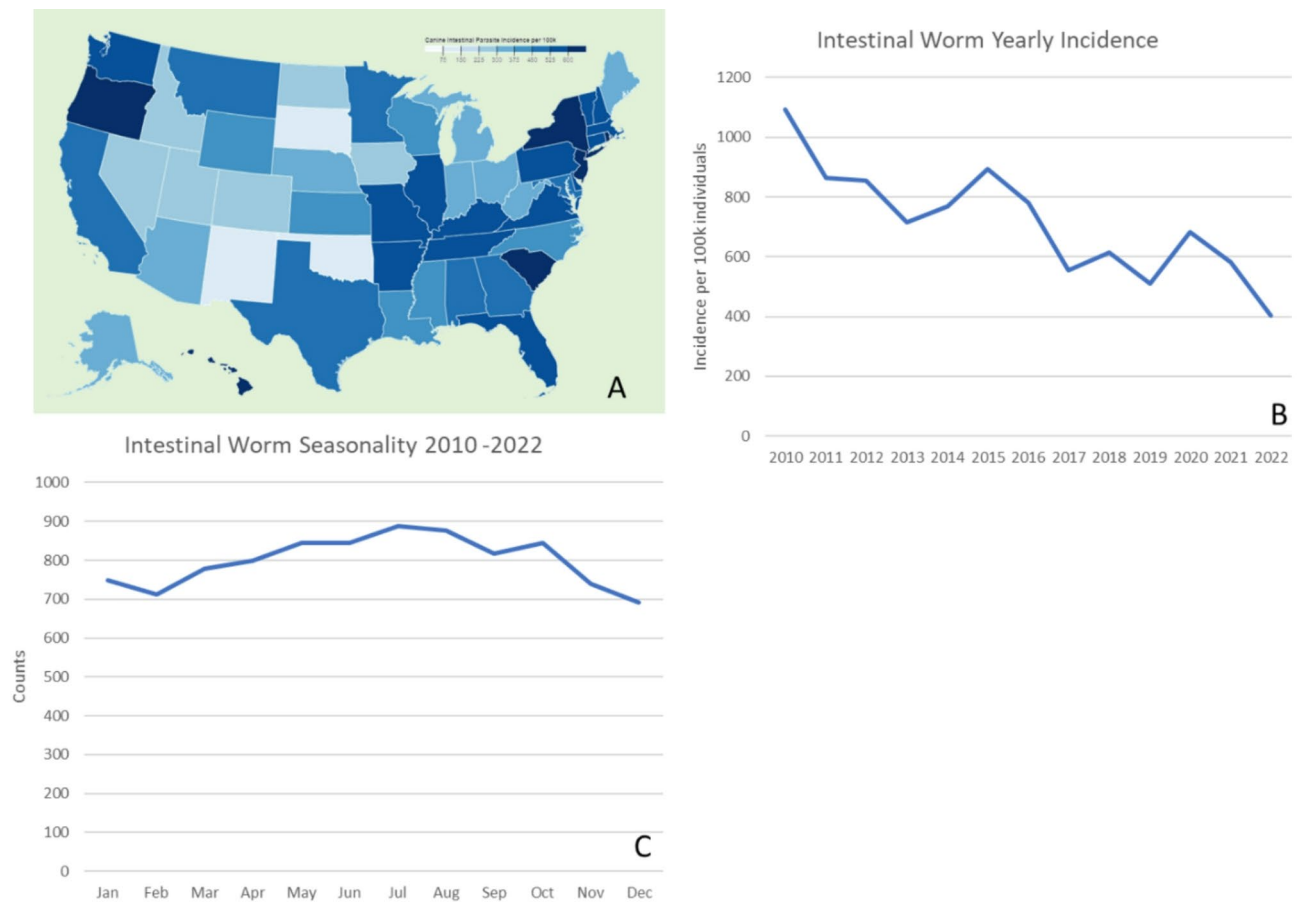


Fig. 5. Intestinal worm diagnoses within this population of insured pet dogs: (A) geographically, (B) yearly incidence per 100k, and (C) seasonally from 2010–2022.

	MAE human only	MAE human and dog	P-value whole model human data coefficient	P-value whole model dog data coefficient
Lyme top 15 states	0.69 ± 0.57	0.55 ± 0.55	0.09	0.14
Lyme nationwide	0.62 ± 0.76	0.61 ± 0.77	0.98	0.76
Valley Fever	0.64 ± 0.53	0.52 ± 0.48	0.08	0.15
Giardia	0.54 ± 0.46	0.28 ± 0.30	0.09	0.01

Table 2. Mean absolute error (MAE) ± standard deviation for the constructed models using human-only incidence information compared to using both human and dog incidence information as covariates.

when human cases start rising, so public health agencies have the opportunity to earmark funding specifically for tick prevention and Lyme disease awareness education.

Methods
Pet insurance data claims codes

The claims used in this study are from dogs insured by Fetch, Inc. from all 50 United States and 12 of 13 Canadian Provinces (Fetch, Inc does not offer pet insurance coverage in Quebec) from years 2008–2022. Fetch maintains a list of disease codes that denotes the medical condition name or clinical signs of disease. Once a policyholder files a claim, a disease code is assigned as stated on the accompanying medical records, invoice, and/or policyholder declaration. The condition codes for each disease included in this study are: Tickborne Diseases (Lyme disease, babesiosis, ehrlichiosis, anaplasmosis, tick bite), Giardiasis (one condition code: Giardia), Fungal Diseases (Coccidioidomycosis, Aspergillosis, general fungal), Intestinal Parasites (ascarids, coccidiosis, intestinal parasites, toxoplasmosis, whipworms, hookworm, roundworm/strongyle, tapeworm, internal parasites). Fetch groups dog ages by life stages, as defined by the American Animal Hospital Association (AAHA) life stage guidelines, which accounts for both age and size of dog³⁵.

Pet insurance data analysis

The counts for each disease were tabulated by state/province, year, and month within year. The overall population data was also sorted into these categories to determine the total number of active policies (dog-time). For example, a dog with an insurance policy who was insured for half of the year would have a dog-time of 0.5. A dog who was insured for the entire year had a dog-time of 1. The number of active policies in terms of dog time was used as the denominator data for normalization. The counts data were normalized and incidence rates were calculated – this was done using a combination of the Excel platform with pivot tables and R studio (see Supplemental Table 1 for a summary of this data).

Human data analysis

Data regarding the yearly trends and monthly trends of Lyme, Giardia, and Coccidioidomycosis (valley fever) were obtained from the CDC website^{36–38} – this data is publicly available for download. Giardia data for the years 2013–2019 were available at the time of publication. Lyme data was available for 2008–2021, and Coccidioidomycosis data was available for 1998–2019. The geographic trends for Lyme and Coccidioidomycosis were also obtained from the CDC site. The geographic trends for giardia for the years 2011–2016 were recreated from the choropleth map displayed by Coffey³⁹. The CDC Lyme counts per state were normalized by the 2020 total state population according to the 2020 census data⁴⁰. See supplemental Table 1 for a summary of this data.

Comparing human data to dog data

Creating direct comparisons of incident rates for every disease group was not possible as the incident rates were not available for the human population. Therefore for seasonal Lyme trends, seasonal Giardia trends, and geographic distribution of coccidioidomycosis were compared based on count data alone. When plotting yearly trends, the incident rates were plotted relative to the 2008 rate for that species. That is, the yearly incident rate was divided by the 2008 incident rate for the species and disease, and those relative measures were plotted on the same indexed chart to compare relative human disease trends and canine disease trends. This method was chosen over the dual-axis chart to emphasize relative trends of the diseases rather than changes to total counts. Specifically for the seasonality of Lyme disease, the counts of dog Lyme diagnoses required a 200x multiplier in order to display the trends lines on the same chart. The human data for Lyme seasonality was available only in counts, so therefore the canine data was also presented in counts for this comparison. In order to display the Giardia incidence rates relative to each other on the same chart, the canine incidence rates are displayed per 1000 dogs, while the human incidence rates are per 100,000 persons.

Modelling

General linear models were trained and evaluated using machine learning techniques via the “caret” package in R. The mean absolute error for the models were calculated using the package “metrics.” Two models were trained for each disease for which both human and dog data were available. The yearly incidence rates for the disease in question were input, and a new variable for the slope of the dog incidence rate were calculated. The columns were normalized using a scaler function.

Model 1 (only human data):

$$y(n) = \beta_1 y(n-1) + \epsilon$$

Outcome variable $y(n)$: Yearly incidence rate for humans in year n .
Predictor:

- $y(n-1)$: Previous year's human incidence.

Model 2 (adding dog data):

$$y(n) = \beta_1 y(n-1) + \beta_2 \Delta(n) + \epsilon$$

Outcome variable $y(n)$: Yearly incidence rate for humans in year n .
Predictors.

- $y(n-1)$: Previous year's human incidence.
- $\Delta(n)$: The difference in dog incidences
 - For Lyme: $\Delta(n) = d(n) - d(n-2)$
 - For all others: $\Delta(n) = d(n) - d(n-1)$

The first model was trained using the previous year's human incidence information to predict the year's incidence rate. The second model was trained using the previous year's human incidence information and the previous year's dog slope to predict the year's incidence rate for humans. The notable exception is that for the Lyme data, the two years ago dog slope was used instead of the previous year's, since the trends were offset by 2 years.

The models were evaluated using a five-fold cross-validation. The mean absolute error for each fold was calculated and the error for the two models for each disease were compared. General linear models were used because the data were relatively simple (yearly incidence for two host species) and limited in size, making more complex modeling unnecessary. The models were evaluated using a five-fold cross-validation procedure, and the mean absolute error for each fold was calculated. The errors for the two models for each disease were then compared.

Given that (a) the primary objective of this study was to demonstrate that incorporating dog-related data enhances the predictability of human incidents, and (b) the sample size was limited, a linear model with a restricted set of predictors was selected. It is worth noting that with the potential availability of more fine-grained data in the future, more complex models and a larger set of predictors could be utilized, likely resulting in improved model performance.

Data availability

Data is provided within the manuscript and all datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 27 June 2024; Accepted: 17 April 2025

Published online: 25 April 2025

References

1. Davila, E. et al. Domestic dogs as sentinels for West Nile virus but not Aedes-borne flaviviruses, Mexico. *Emerg. Infect. Dis.* **28**, 1071–1074 (2022).
2. Backer, L. C., Grindem, C. B., Corbett, W. T., Cullins, L. & Hunter, J. L. Pet dogs as sentinels for environmental contamination. *Sci. Total Environ.* **274**, 161–169 (2001).
3. Gavazza, A., Presciutti, S., Barale, R., Lubas, G. & Gugliucci, B. Association between canine malignant lymphoma, living in industrial areas, and use of chemicals by dog owners. *J. Vet. Intern. Med.* **15**, 190–195 (2001).
4. Sexton, C. & Ruple, A. Canine sentinels and our shared exposome. *Science* **384**, 1170–1172 (2024).
5. Persing, D. H. et al. Detection of *Borrelia burgdorferi* DNA in museum specimens of Ixodes dammini ticks. *Science* **249**, 1420–1423 (1990).
6. Marshall, W. F. Detection of *Borrelia burgdorferi* DNA in museum specimens of *Peromyscus leucopus*. *J. Infect. Dis.* **170**, 1027–1032 (1994). 3rd ed.
7. Stone, B. L., Tourand, Y. & Brissette, C. A. Brave new worlds: the expanding universe of Lyme disease. *Vector Borne Zoonotic Dis.* **17**, 619–629 (2017).
8. Ostfeld, R. S. & Brunner, J. L. Climate change and Ixodes tick-borne diseases of humans. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, (2015).
9. Ogden, N. H. et al. How Far North are migrant birds transporting the tick *Ixodes scapularis* in Canada? Insights from stable hydrogen isotope analyses of feathers. *Ticks Tick. Borne Dis.* **6**, 715–720 (2015).
10. Larsen, A. E., MacDonald, A. J. & Plantinga, A. J. Lyme disease risk influences human settlement in the wildland-urban interface: evidence from a longitudinal analysis of counties in the Northeastern United States. *Am. J. Trop. Med. Hyg.* **91**, 747–755 (2014).
11. Ferrouillet, C., Milord, F., Lambert, L., Vibien, A. & Ravel, A. Lyme disease: knowledge and practices of family practitioners in Southern Quebec. *Can. J. Infect. Dis. Med. Microbiol.* **26**, 151–156 (2015).
12. Gasmí, S. et al. Practices of Lyme disease diagnosis and treatment by general practitioners in Quebec, 2008–2015. *BMC Fam Pract.* **18**, 65 (2017).
13. Rivera, O. J. & Nookala, V. Lyme carditis. in *StatPearls* (StatPearls Publishing, Treasure Island (FL), (2023)).
14. Eisen, L. & Eisen, R. J. Changes in the geographic distribution of the blacklegged tick, *Ixodes scapularis*, in the United States. *Ticks Tick. Borne Dis.* **14**, 102233 (2023).
15. Burtis, J. C. et al. Predicting distributions of blacklegged ticks (*Ixodes scapularis*), Lyme disease spirochetes (*Borrelia burgdorferi* sensu stricto) and human Lyme disease cases in the Eastern United States. *Ticks Tick. Borne Dis.* **13**, 102000 (2022).
16. Little, E. A. H. et al. Predicting spatiotemporal patterns of Lyme disease incidence from passively collected surveillance data for *Borrelia burgdorferi* sensu lato-infected *Ixodes scapularis* ticks. *Ticks Tick. Borne Dis.* **10**, 970–980 (2019).
17. Watson, S. C. et al. A Bayesian spatio-temporal model for forecasting the prevalence of antibodies to *Borrelia burgdorferi*, causative agent of Lyme disease, in domestic dogs within the contiguous United States. *PLoS One* **12**, e0174428 (2017).
18. Cook, M. J. & Puri, B. K. Estimates for Lyme borreliosis infections based on models using Sentinel canine and human seroprevalence data. *Infect. Dis. Model.* **5**, 871–888 (2020).
19. Herrin, B. H., Beall, M. J., Feng, X., Papeş, M. & Little, S. E. Canine and human infection with *Borrelia burgdorferi* in the New York City metropolitan area. *Parasit. Vectors* **11**, 187 (2018).
20. Zajackowski, P. et al. The controversies surrounding *Giardia intestinalis* assemblages A and B. *Curr. Res. Parasitol. Vector Borne Dis.* **1**, 100055 (2021).
21. Gharakhanian, R., Ohno, A., Craig, D. & Wang, S. Valley fever, mimicker of malignancy. *Heliyon* **9**, e12900 (2023).
22. James, A. E., Pastenkos, G., Bradway, D. & Baszler, T. Autochthonous transmission of coccidioides in animals, Washington, USA. *Emerg. Infect. Dis.* **25**, 123–125 (2019).
23. Gautam, R. et al. Identifying areas of high risk of human exposure to coccidioidomycosis in Texas using serology data from dogs. *Zoonoses Public Health* **60**, 174–181 (2013).
24. Grayzel, S. E., Martínez-López, B. & Sykes, J. E. Risk factors and spatial distribution of canine coccidioidomycosis in California, 2005–2013. *Transbound. Emerg. Dis.* **64**, 1110–1119 (2017).
25. Lynn, M. K., Morrissey, J. A. & Conserve, D. F. Soil-transmitted helminths in the USA: a review of five common parasites and future directions for avenues of enhanced epidemiologic inquiry. *Curr. Trop. Med. Rep.* **8**, 32–42 (2021).
26. Tyungu, D. L. et al. Toxocara species environmental contamination of public spaces in New York City. *PLoS Negl. Trop. Dis.* **14**, e0008249 (2020).
27. Hare, A. Q. & Franco-Paredes, C. Ocular larva migrans: A severe manifestation of an unseen epidemic. *Curr. Trop. Med. Rep.* **1**, 69–73 (2014).
28. Chaumet, A. C. S. G., Rossi, T. A., Murphy, L. A. & Nakamura, R. K. Evaluation of owners' attitudes towards veterinary insurance in a specialty hospital. *J. Small Anim. Pract.* **62**, 805–809 (2021).
29. Williams, A., Williams, B., Hansen, C. R. & Coble, K. H. The impact of pet health insurance on dog owners' spending for veterinary services. *Anim. (Basel)* **10**, (2020).
30. Kipperman, B. S., Kass, P. H. & Rishniw, M. Factors that influence small animal veterinarians' opinions and actions regarding cost of care and effects of economic limitations on patient care and outcome and professional career satisfaction and burnout. *J. Am. Vet. Med. Assoc.* **250**, 785–794 (2017).
31. Boller, M. et al. The effect of pet insurance on presurgical euthanasia of dogs with gastric Dilatation-Volvulus: A novel approach to quantifying economic euthanasia in veterinary emergency medicine. *Front. Vet. Sci.* **7**, 590615 (2020).
32. Starr, M. C. & Montgomery, S. P. Soil-transmitted helminthiasis in the United States: a systematic review–1940–2010. *Am. J. Trop. Med. Hyg.* **85**, 680–684 (2011).
33. Sweet, S. et al. A 3-year retrospective analysis of canine intestinal parasites: fecal testing positivity by age, U.S. Geographical region and reason for veterinary visit. *Parasit. Vectors* **14**, 173 (2021).
34. Williams, A., Coble, K. H., Williams, B. R., Dicks, M. & Knippenberg, R. Consumer preferences for Pet Health Insurance. (2016).

35. Creevy, K. E. et al. 2019 AAHA canine life stage guidelines. *J. Am. Anim. Hosp. Assoc.* **55**, 267–290 (2019).
36. CDC. Lyme disease surveillance data. *Lyme Disease* (2024). <https://www.cdc.gov/lyme/data-research/facts-stats/surveillance-data-1.html>.
37. CDC. Nationally notifiable waterborne diseases. *Waterborne Disease and Outbreak Surveillance Reporting* (2024). https://www.cdc.gov/healthy-water-data/about/nationally-notifiable-waterborne-diseases.html?CDC_AAref_Val=https://www.cdc.gov/healthywater/surveillance/nndss.html.
38. CDC. Facts and Stats about Valley fever (Coccidioidomycosis). *Valley Fever (Coccidioidomycosis)* (2024). https://www.cdc.gov/valley-fever/php/statistics/?CDC_AAref_Val=https://www.cdc.gov/fungal/diseases/coccidioidomycosis/statistics.html.
39. Coffey, C. M. et al. Evolving epidemiology of reported giardiasis cases in the United States, 1995–2016. *Clin. Infect. Dis.* **72**, 764–770 (2021).
40. Census Data Table. <http://www.census.gov/programs-surveys/decennial/2020/data/apportionment/population-change-data-table.pdf>.

Acknowledgements

The authors thank the owners and dogs, and veterinarians for their important contributions.

Author contributions

J.O. – writing, data analysis A.M. – writing, review, editing C.D. – data analysis A.R. – writing, review & editing.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-99229-9>.

Correspondence and requests for materials should be addressed to A.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025