PAPER

Spatiotemporal patterns and agroecological risk factors for cutaneous and renal glomerular vasculopathy (Alabama Rot) in dogs in the UK

Kim B Stevens,[1,8] Rosanne Jepson,[2] Laura Phillipa Holm,[3] David John Walker,[3] Jacqueline Martina Cardwell[1]

The annual outbreaks of cutaneous and renal glomerular vasculopathy (CRGV) reported in UK dogs display a distinct seasonal pattern (November to May) suggesting possible climatic drivers of the disease. The objectives of this study were to explore disease clustering and identify associations between agroecological factors and CRGV occurrence. Kernel-smoothed maps were generated to show the annual reporting distribution of CRGV, Kuldorff's space–time permutation statistic used to identify significant spatiotemporal case clusters and a boosted regression tree model developed to quantify associations between CRGV case locations and a range of agroecological factors. The majority of diagnoses (92 per cent) were reported between November and May while the number of regions reporting the disease increased between 2012 and 2017. Two significant spatiotemporal clusters were identified—one in the New Forest during February and March 2013, and one adjacent to it (April 2015 to May 2017)—showing significantly higher and lower proportions of cases than the rest of the UK, respectively, for the indicated time periods. A moderately significant high-risk cluster (P=0.087) was also identified in the Manchester area of northern England between February and April 2014. Habitat was the predictor with the highest relative contribution to CRGV distribution (20.3 per cent). Cases were generally associated with woodlands, increasing mean maximum temperatures in winter, spring and autumn, increasing mean rainfall in winter and spring and decreasing cattle and sheep density. Understanding of such factors may help develop causal models for CRGV occurrence.

Introduction

Cutaneous and renal glomerular vasculopathy (CRGV)—also known as 'Alabama Rot'—is a disease of unknown aetiology variably associated with clinically relevant acute kidney injury (AKI). CRGV cases present with ulcerated skin lesions, most often affecting the distal limbs, progressing within 1–10 days to the development of AKI in some, but not all cases. Skin lesions have also been found to affect the face, nasal planum, oral cavity, tongue, ventrum and flanks. Additional biochemical and haematological findings commonly reported include mild to moderate hyperbilirubinaemia, anaemia and moderate to severe thrombocytopenia.[1]

A study by Holm *et al* which reported on the renal histopathology of CRGV cases confirmed the lesions to be compatible with a thrombotic microangiopathy (TMA).[1] In human medicine, TMAs are considered a complex group of diseases which can involve both hereditary and acquired contributing factors to the development of clinical disease.[2] Hereditary factors that have been identified include mutations in ADAMTS13 which result in the condition known as thrombotic thrombocytopenic purpura, complement factors, metabolic factors (methylmalonic aciduria and homocystinuria type C protein) and diacylglycerol kinase-ε—an abnormality of which results in a prothrombotic state. Autoantibody inhibition of ADAMTS13, Shiga toxin exposure (Shiga

Veterinary Record (2018) doi: 10.1136/vr.104892

[1]Department of Pathobiology and Population Sciences, Royal Veterinary College, Hatfield, UK
[2]Department of Clinical Science and Services, Royal Veterinary College, Hatfield, UK
[3]Anderson Moores Veterinary Specialists, Bunstead Barns, Winchester, UK
[8]Kimene Analytics Ltd, London, UK

E-mail for correspondence: kstevens@rvc.ac.uk

Provenance and peer review Not commissioned; externally peer reviewed.

Received February 13, 2018
Revised May 23, 2018
Accepted July 1, 2018

toxin-haemolytic uraemic syndrome), drug, toxin or complement immune-mediated acquired forms of TMA also occur.[2] Preliminary investigations evaluating the existence of underlying infectious or toxic exposure (eg, Shiga toxin) have so far been unsuccessful.[1]

There has been much speculation in the general and non-peer-reviewed veterinary press on the possible existence of an association between CRGV occurrence and either specific habitats or weather conditions since the majority of early cases occurred in the New Forest in south-eastern England. However, it is unclear whether this apparent connection is simply the result of the coincident locale of the referral veterinary centre (LPH and DJW) that initially raised awareness of CRGV as a disease entity, or a true association. In addition, the UK outbreaks have so far displayed a distinct seasonal pattern with cases generally reported between November and May. Such cyclical occurrence of a disease often signifies the involvement of climatic factors, and the objectives of this study were to therefore explore associations between a range of agroecological factors and CRGV locations, as well as map and explore the distribution of cases between 2012 and 2017. The results of this study may help develop causal models for CRGV, assist with validation of current and future proposed pathogenic mechanisms and play a role in identifying the aetiology of the disease.

Materials and methods

All CRGV cases diagnosed between November 2012 and May 2017 were included in the analysis. Although one case was reported from Northern Ireland within this time period, it did not have locational data and was therefore excluded from the spatial, but retained for the temporal analysis.

Identification of cases

Cases were compiled by two investigators (DJW and LPH) and comprised 70 (68 per cent) from first-opinion practice and 33 (32 per cent) from referral centres. A diagnosis of CRGV was based on the presence of compatible clinical signs (including skin lesions), laboratory diagnostics (including AKI±oligoanuria, progression to azotaemia, hyperbilirubinaemia, anaemia and thrombocytopenia) and renal histopathology findings compatible with TMA. Renal histopathology was available either in isolation or as part of a full postmortem examination in all cases, and in most cases dermal pathology was also available.

The residential postcode of all CRGV cases was available together with the postcode of where the dog had been recently walked, if markedly different from the residential postcode (eg, owners had been on holiday in the New Forest area yet normally resided in a different part of the country). Where the residential and walked postcodes differed (n=5), both postcodes were included in the data set creating a data set of 107 postcodes for inclusion in the spatial analysis. Postcodes were converted to Easting and Northing Cartesian coordinates and the British National Grid projection used for all spatial analyses.

Agroecological data

As nothing is known about the environmental epidemiology of CRGV a broad general selection of agroecological predictors was identified for initial

TABLE 1: Descriptors of the spatial agroecological predictor variables selected to model the distribution of cases of cutaneous and renal glomerular vasculopathy in dogs in the UK

Predictor name	Descriptor	Data source
CattleDens	Density of cattle (heads/km^2)	Gridded Livestock of the World (http://www.fao.org/ag/againfo/resources/en/glw/home.html)
SheepDens	Density of sheep (heads/km^2)	
PigDens	Density of pigs (heads/km^2)	
Habitat	Expected habitat for the soiltype	NATMAP SoilScapes map for England and Wales (1:250,000) (http://www.landis.org.uk/data/nmsoilscapes.cfm)
Landcover	Expected landcover for the soiltype	
SoilDrain	Soil drainage characteristics	
SoilFert	Soil fertility characteristics	
AvTemp	Mean temperature of the spring (Sp), summer (Su), autumn (Au) and winter (Wi) months (°C)	UK Met Office 5 km x 5 km gridded data sets (https://www.metoffice.gov.uk/research/climate/climate-monitoring/ukcp09/register)
AvMaxTemp	Mean maximum temperature of the spring (Sp), summer (Su), autumn (Au) and winter (Wi) months (°C)	
AvMinTemp	Mean minimum temperature of the spring (Sp), summer (Su), autumn (Au) and winter (Wi) months (°C)	
AvRain	Mean rainfall of the spring (Sp), summer (Su), autumn (Au) and winter (Wi) months (mm)	
AvRainDays1	Mean number of days with a rainfall of >1 mm in the spring (Sp), summer (Su), autumn (Au) and winter (Wi) months (days)	
AvRainDays10	Mean number of days with a rainfall of >10 mm in the spring (Sp), summer (Su), autumn (Au) and winter (Wi) months (days)	
GrndFrostDays	Mean number of days with ground frost in the spring (Sp), summer (Su), autumn (Au) and winter (Wi) months (days)	
AirFrostDays	Mean number of days with air frost in the spring (Sp), summer (Su), autumn (Au) and winter (Wi) months (days)	

inclusion in the model and the necessary digital spatial data layers sourced as detailed in table 1. Soil drainage, fertility, habitat and land cover were extracted from the 1:250,000 NATMAP SoilScapes map for England and Wales. There is no such map available for Scotland and therefore the spatial modelling was confined to England and Wales, and all other predictor data were clipped to this extent. Cattle, sheep and pig densities were extracted from Gridded Livestock of the World (http://www.fao.org/ag/againfo/resources/en/glw/home.html), and climate data extracted from the UK's Met Office gridded land surface climate observations (monthly climate variables at 5 km resolution) held by the Centre for Environmental Data Analysis (http://catalogue.ceda.ac.uk/uuid/87f43af9d02e42f483351d79b3d6162a).

For the purpose of analysis, the original soil drainage, habitat and landcover categories were retained, but the 12 original soil fertility categories were collapsed into six as follows: high; moderate to high; moderate; low (low + very low); lime rich (lime rich + lime rich to moderate + lime rich to very low + low to lime rich); and mixed (low to high + low to moderate).

Climatic variables downloaded included monthly data for mean temperature, maximum temperature, minimum temperature, rainfall, rain days 1 mm, rain days 10 mm, air frost and ground frost. These data were downloaded for the period September 2011 to December 2016 (2017 data were unavailable). Although the first cases were recorded in November 2011, climate data for the preceding two months were included to allow for the creation of the autumn 2011 variables (September to November), resulting in six years of autumn data but only five years of data for the remaining seasons. The variables snow-falling and snow-lying would have been included in the analysis but the data were not available after 2011. As reporting of CRGV cases has displayed a strong seasonal pattern with cases occurring primarily in winter and spring, rather than use monthly or annual data, monthly climatic variables were aggregated to create seasonal versions of each variable on the following basis: spring (March to May), summer (June to August), autumn (September to November) and winter (December to February). For each of the three months comprising a season, the relevant monthly raster maps were summed and divided by 18 (autumn) or 15 (spring, summer, winter) to create a mean seasonal version of each climatic variable. The final climatic variables included in the model for each season were: mean temperature, mean maximum temperature, mean minimum temperature, mean rainfall, mean number of days with rainfall more than 1 mm, mean number of days with rainfall more than 10 mm, mean number of days experiencing ground frost and mean number of days experiencing air frost (table 1).

All layers were resampled to a resolution of 1 km^2 and clipped to the England-Wales extent. ArcGIS software V.10.5.1 was used to extract values of each predictor variable to the case and background data points to create the complete data set, which was then randomly divided into training, validation and test sub-data sets comprising 60, 20 and 20 per cent of the data points, respectively.

Mapping the spatiotemporal distribution of cases

A heat map was created using the R tidyverse package[3 4] to illustrate the temporal reporting pattern of CRGV cases between 2012 and 2017 by both month and year. Kernel-smoothed maps were generated for individual years and for the study period as a whole to show the spatial distribution of cases. Optimum bandwidth was estimated using the quartic approximation of a true Gaussian kernel function. A bandwidth of 20 km was used for all maps with an output cell size of 1 km^2. All maps were produced using ArcGIS V.10.5.1.

Cluster detection

Kuldorff's space–time permutation statistic (implemented in SaTScan V.9.5) was used to identify spatiotemporal clusters as this statistic requires only case data (spatial location and time for each case), with no information needed about controls or the population at risk. The number of observed cases in a cluster is compared with what would have been expected if the spatial and temporal locations of all cases were independent of each other so that there is no space–time interaction. That is, there is a cluster in a geographical area if, during a specific time period, that area has a higher proportion of its cases in that time period compared with the remaining geographical areas. Cartesian coordinates of all cases were used as the spatial inputs and month of reporting was used to indicate the timing of each case. The data were analysed for the study period as a whole. Clustering and cluster detection tests are viewed as complimentary as they test different hypotheses, and a simulation study by Waller *et al*[5] indicated that it is possible to have a significant cluster, but no overall significant clustering. For this reason, tests for clustering were not run prior to implementing the space–time permutation statistic.

CRGV suitability modelling

The suitability models were generated using boosted regression trees (BRT), a robust machine learning method with the ability to account for non-linearity and complex relationships between the dependent and predictor variables.[6] BRTs differ from the traditional regression methods commonly used in epidemiological studies in that rather than producing a single 'best' model, they optimise predictive performance by using the technique of boosting to adaptively combine large numbers of relatively simple tree models.[6] As well as being more easily interpreted than other machine learning methods such as support vector machines or random forest models, BRTs have been shown to generally outperform more conventional approaches, such as logistic regression, in general species distribution modelling studies.[7]

Background data points

As calibration of the BRT model used to identify associations between agroecological risk factors and CRGV distribution required both presence and absence records, 2000 background points were randomly generated within the confines of the England/Wales boundary in order to characterise the agroecological conditions existing within. The number of background points was a trade-off between adequately characterising the variability in the environment while maintaining a sufficiently high prevalence so as to not suffer from possible bias linked to artificially induced prevalence.[8]

Calibration and evaluation of the BRT model

The BRT algorithm was implemented using the gbm package (V.1.6–3) in R V.3.3.1[3] together with the k-fold cross-validation stagewise function available from Elith *et al*.[6] Pairwise combinations of a range of potential *lr* and *tc* were trialled to determine the best combination for identifying the optimal number of trees (a tree complexity of 4 and learning rate of 0.005). This optimum combination should result in more than 1000 trees[6] while allowing the model to converge. A Bernoulli error structure was specified and stochasticity was maintained through a bag fraction of 50 per cent. As there was considerable collinearity between the two variables habitat and land cover (habitat nested within land cover), models were run with either habitat or land cover (keeping all other predictor variables constant) to determine which of the two predictors had the higher relative contribution to CRGV distribution and this variable was retained in the model while the other was dropped. In order to determine whether any variables were best omitted from the model, variables were removed in turn, starting with those having the smallest relative influence, and average change in predictive deviance calculated. Variables for which this value exceeded the model's original estimated standard error were excluded from the model.

Relative influence or contribution of the predictor variables to the response was calculated using formulae developed by Friedman[9] and implemented in the gbm package. These measures are based on the number of times a variable is selected for splitting, weighted by the squared improvement to the model as a result of each split, and averaged over all trees. The relative contribution of each variable was scaled so that together they summed to 100 with higher numbers indicating a stronger contribution to the response. Partial dependence plots describing relative probability of CRGV presence in relation to the range of values of each predictor variable were generated after accounting for the average effects of all other variables in the model. The predictive power of the model was evaluated using the test data set and area under the receiver operating characteristic curve (AUC) computed for the binary classifier.

Results

Temporal pattern of CRGV case reporting

The first known cases of CRGV were reported in 2012 (November/December: 4 per cent; n=4) with a slight increase in the following year (7 per cent; n=7). Number of reported cases peaked in 2014 with a third of all cases reported in this year (33 per cent; n=35) and decreased gradually thereafter (January 2017 to May: 17 per cent; n=18) (figure 1). Seasonally, CRGV cases were reported largely between December and May (winter/spring) with a third of all cases diagnosed in the first three months of the year (January to March). Only 7 per cent of cases (n=7) were reported in the summer months (June to August) with no cases reported in October (figure 1).

Spatial distribution of CRGV cases

The kernel density maps in figure 2 show the density of CRGV cases (cases/km^2) with darker brown areas exhibiting a higher reporting density of cases and lighter brown areas a lower (or no) reporting density of cases. Although the four initial cases of CRGV in 2012 were distributed randomly throughout England, in subsequent years reporting of the disease showed a tendency to cluster in certain areas (figure 2). In 2013, cases were located around the New Forest on the southern coast of England and 2014 saw the expansion of this CRGV hotspot of reporting together with the development of a second area of high reporting density in the Manchester region of Northern England. These two main high-density reporting areas (New Forest and Manchester) persisted through to 2017 although the New Forest hotspot was not apparent in 2016, replaced instead by an area of high reporting density around Greater London and a smaller area of activity on the south coast of Wales. In 2017, distribution of cases was the most diffuse of all five years. In all years, the areas with a high reporting density of cases were generally accompanied by a few localised cases of CRGV scattered throughout England (figure 2).

A map of the reporting density (cases/km^2) of all cases aggregated over the five-year period shows the north-east of England and the New Forest region of south England to have the highest five-year density of CRGV cases (figure 3). A diffuse triangular area covering a large part of south-central England showed a medium to high reporting density of cases.

Kuldorff's spatiotemporal permutation statistic identified three spatiotemporal clusters. The cluster locations are shown in figure 3 and details of each are provided in table 2. The most likely cluster occurred

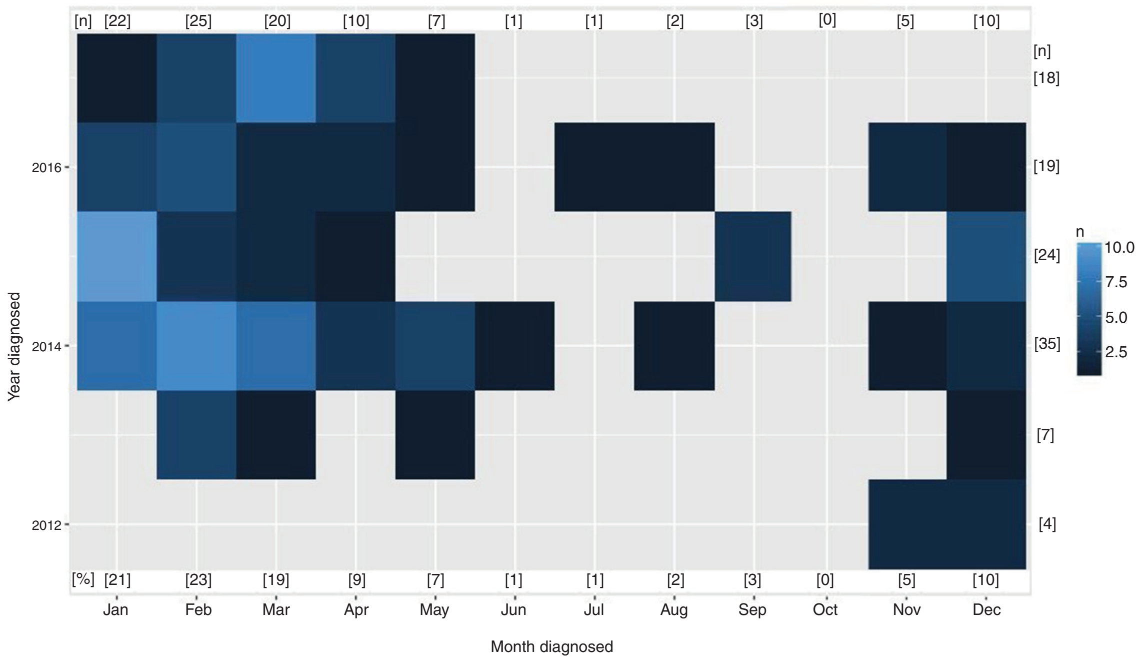

FIG 1: Heat map illustrating the temporal distribution of 107 cases of cutaneous and renal glomerular vasculopathy (CRGV) in dogs in the UK, divided by month and year (November 2012 to May 2017, inclusive). Months are shown on the x axis and years on the y axis. The shading of the blue blocks represents the frequency of CRGV cases reported that month (lighter shading = higher frequency). The grey background is visible when no cases were reported in a month.

between April 2015 and May 2017 and included the area immediately to the right of the New Forest. This region reported a significantly (P=0.002) lower proportion of CRGV cases than the rest of the UK (figure 3). Between February and March 2013, the New Forest region on the south coast of England exhibited a significantly (P=0.004) higher proportion of cases than the rest of the UK while between January and April 2014 the area around Manchester reported a moderately significantly (P=0.087) higher proportion of cases than the rest of the UK (figure 3).

Agroecological factors associated with CRGV distribution

As habitat explained a greater proportion of the variability in CRGV distribution than land cover (20.3 vs. 16 per cent), it was retained in the final model. The final predictive model contrasting CRGV case locations with background points had a good accuracy, with an AUC of 0.903 when evaluated against the model calibration data set and an AUC of 0.884±0.022 when evaluated with cross-validation as implemented by Elith *et al.*[6] The suitability map (figure 4) highlights areas of predicted high suitability for CRGV case occurrence and resembles the aggregated kernel-density map of CRGV case distribution for 2012–2017 (figure 3). Areas with the highest predicted suitability for CRGV occurrence include West Sussex, southern Dorset and southern Hampshire in the south of England, and the greater Manchester area in the north of England, together with the eastern regions of South Glamorgan and western Gwent in Wales. In addition, there are small localised areas of high suitability dotted throughout England, specifically in the counties of Somerset, West Midlands and Nottinghamshire. Most of southern England, apart from south-west England, is classified as moderately suitable. Broad regions of low suitability include North and Central Wales, East Anglia, most of East Midlands, North Yorkshire, North East England and the northern half of North West England (figure 4).

Four variables (AirFrostDays_Su, SoilDrain, RainDays10_Su and GrndFrostDays_Su) were removed from the model on simplification leaving 34 variables. The relative contribution of each of these predictor variables is presented in figure 5 and can be divided into roughly four groups based on their relative influence on CRGV distribution: important, moderate, low and negligible contributors. Habitat was the only important predictor in the model accounting for 20.3 per cent of the variation in CRGV distribution. AvMaxTemp_Wi (8.8 per cent), AvRain_Wi (6.4 per cent), SheepDens (6.3 per cent), CattleDens (6.1 per cent) and AvTemp_Sp (5.5 per cent) were moderate contributors together accounting for an additional 33.1 per cent of the variation in disease distribution. These variables, together with AvRain_Sp (4.9 per cent), AvMaxTemp_Sp (4.0 per cent), AvMaxTemp_Au (3.8 per cent) and PigDens (2.5 per cent), accounted for 68.4 per cent of the variation in CRGV distribution. Predictors with a negligible impact on CRGV distribution included soil fertility, number of days of ground (Au, Wi/Sp) or air frost (Wi/Sp) and number of days with more than 1 (Wi/Sp) or more than 10 mm of rain (Au/Wi/Sp) (figure 5).

Dependency profiles for the first 10 predictors are shown in figure 6. The dependency profile for the predictor of primary importance (Habitat) shows that four habitat types are specifically associated with CRGV distribution (in decreasing order of importance): '*mostly lowland dry heath communities*', '*wet acid*

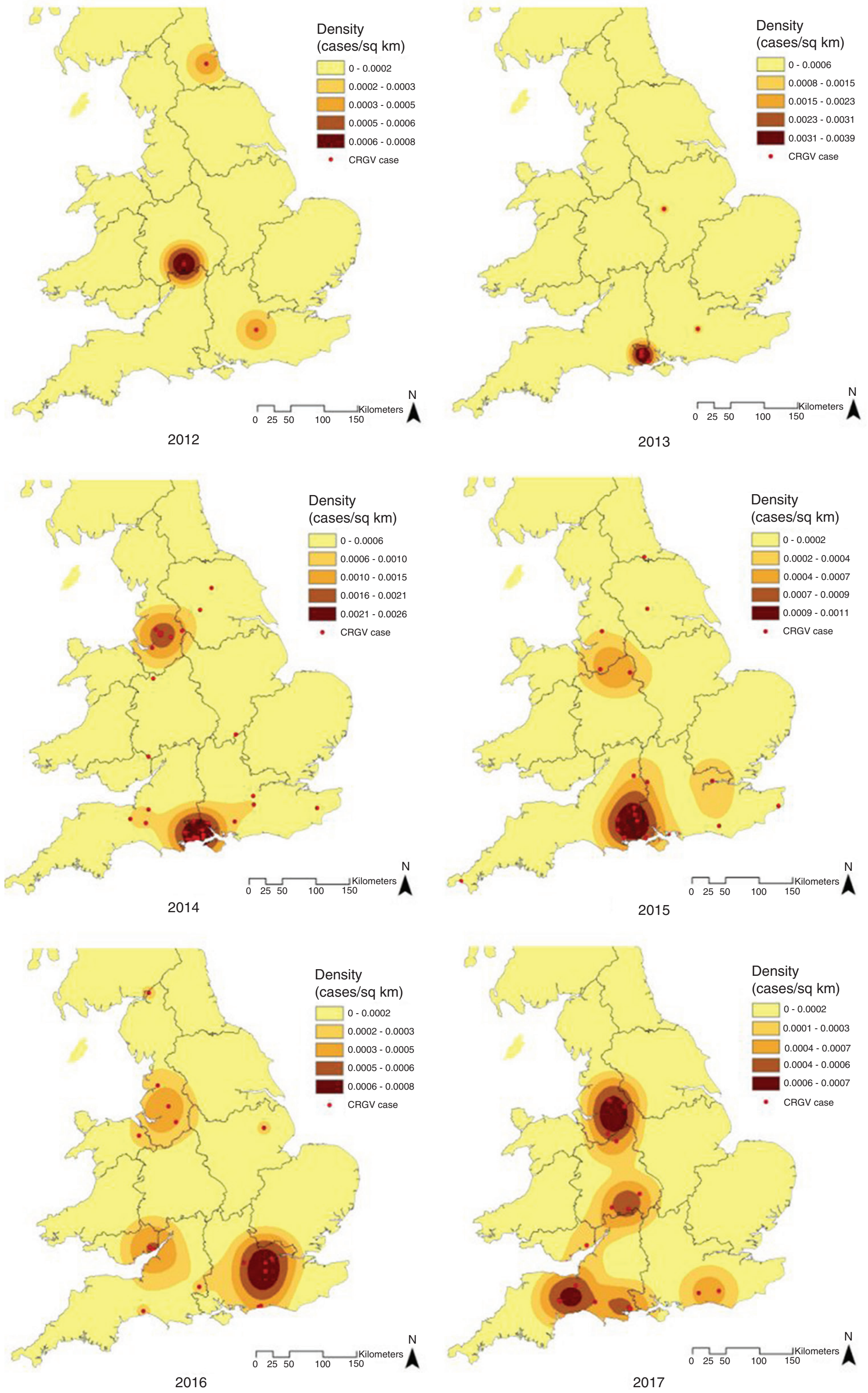

FIG 2: **Maps showing annual location and kernel-smoothed density of cases of cutaneous and renal glomerular vasculopathy (CRGV) in dogs in the UK between January 2012 and May 2017 (inclusive).**

meadows and woodland', '*wet flood meadows with wet carr woodlands in old river meanders*' and '*acid dry pastures*; *acid deciduous and coniferous woodland*; *potential for lowland heath*'. Woodland was a common descriptor in all but the most important habitat ('*mostly lowland dry heath communities*'). Habitat types least likely to be associated with CRGV occurrence included '*base-rich pastures and classic chalky boulder clay ancient woodlands; some wetter areas and lime-rich flush vegetation*', '*base-rich pastures and deciduous woodlands*', '*steep acid upland pastures dry heath and moor; bracken gorse and oak woodlands*' and '*wet brackish coastal flood meadows*'. Pasture was a common descriptor in all these apart from the '*wet brackish coastal flood meadows*' habitat. Dependency profiles for the remaining nine predictors showed that, in addition to associations with specific habitat types, increasing relative probability of CRGV presence was associated

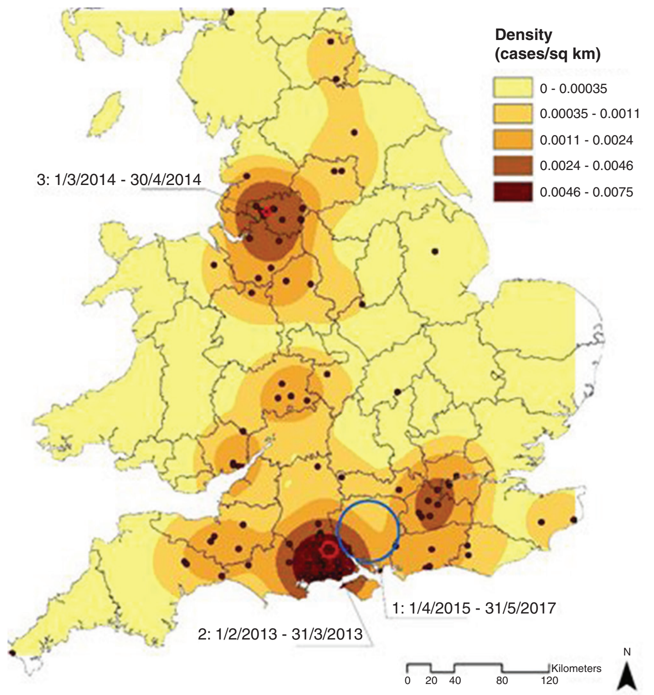

FIG 3: Map showing location and kernel-smoothed density of cases of cutaneous and renal glomerular vasculopathy (CRGV) in dogs in the UK (January 2012 to May 2017) together with the location of two spatiotemporal clusters exhibiting a significantly higher proportion of cases (o), and one spatiotemporal cluster exhibiting a significantly lower proportion of cases (o), than the remainder of the UK. Clusters were identified using Kuldorff's space–time permutation statistic.

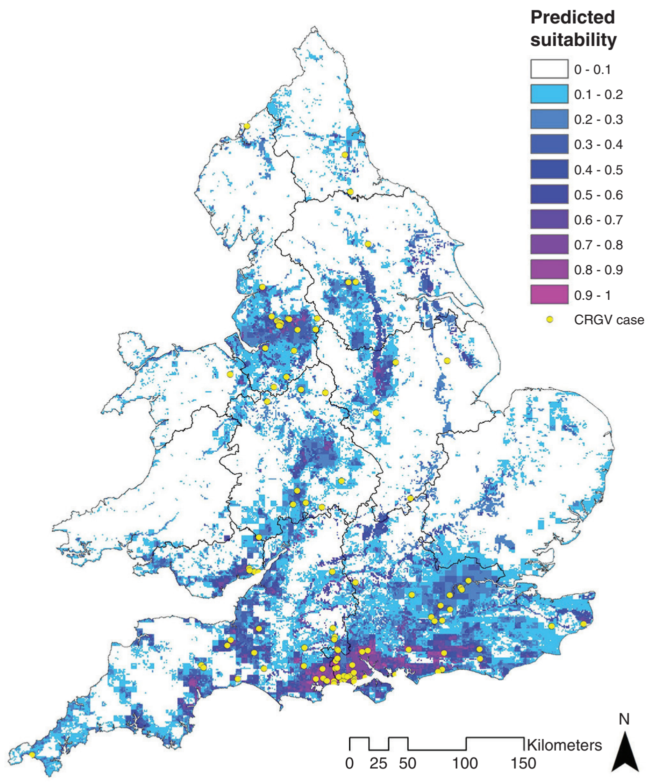

FIG 4: Map showing predicted suitability of England and Wales for the occurrence of cases of cutaneous and renal glomerular vasculopathy (CRGV) in dogs. Yellow dots represent the location of reported CRGV cases (n=107).

with increasing mean maximum temperatures in winter, spring and autumn, increasing mean rainfall in winter and spring, increasing mean temperature in spring, decreasing cattle and sheep density and variable pig density.

There was a mild interaction (strength: 10) between the variables Habitat and AvMaxTemp_Wi with increased probability of CRGV occurrence in three Habitats—'*mostly lowland dry heath communities*', '*wet acid meadows and woodland*', and '*wet flood meadows with wet carr woodlands in old river meanders*'—associated with increasing mean maximum winter temperatures (figure 7).

Discussion

The first known cases of CRGV in UK dogs were reported in 2012 and although initial numbers were very low (2012: n=3) annual frequency of reported cases showed a general increase, although exhibiting occasional year-on-year variation. Diseases that '*have newly appeared in a population or have existed previously but are rapidly increasing in incidence or geographic range*' are defined as 'emerging',[10] and can be further divided into those that are 'newly emerging' (ie, not previously recognised) or 're-emerging/resurging' (ie, diseases that were a major problem before declining dramatically, and then increasing again). The outbreak pattern of CRGV in the UK accords with the definition of a newly emerging disease as no cases were reported prior to 2012. However, that does not mean that the disease was completely unknown in the country as it may simply not have been recognised owing to a very low incidence in the population prior to 2012. A thorough search of practice records is needed to definitely rule out the absence of potential CRGV diagnoses in UK dogs before 2012.

Newly emerging infections are often the result of microbial, host and environmental factors interacting to create opportunities for infectious agents to evolve into new ecological niches. Factors that can contribute to this emergence/re-emergence include changing ecosystems, climate and weather, and microbial adaptation and change.[10] Our BRT model identified the highest relative probability of CRGV occurrence to be associated with a range of agroecological factors specifically, woodland and heath habitats, decreasing cattle and sheep densities, increasing maximum temperatures in winter and, to a lesser extent, spring and autumn, and higher mean rainfall in winter and spring.

TABLE 2: Characteristics of the high- and low-risk clusters of cases of cutaneous and renal glomerular vasculopathy (CRGV) in dogs in the UK (January 2012 to May 2017) as identified by Kuldorff's space–time permutation statistic

Cluster ID	Risk level	Date	Expected cases	Observed cases	P values
1	Low	April 1, 2015 to May 31, 2017	10	0	0.002
2	High	February 1, 2013 to March 31, 2013	0	4	0.004
3	High	January 31, 2014 to April 30, 2014	1	5	0.087

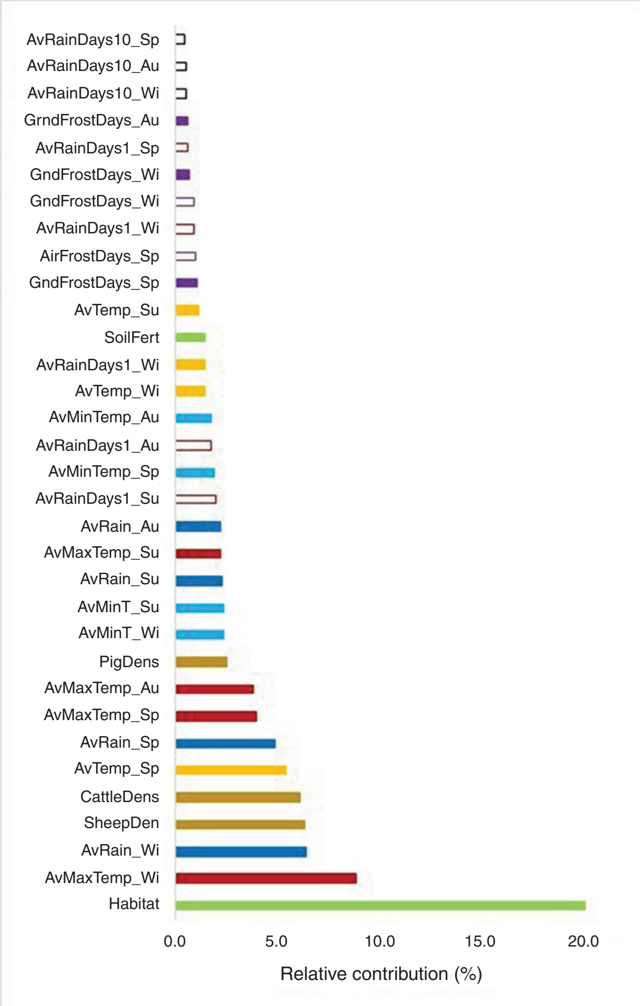

FIG 5: Relative contribution of the 34 predictor variables modelling the spatial distribution of cutaneous and renal glomerular vasculopathy (CRGV) in dogs in the UK (2012–2017). Relative influence (or contribution) of each variable is scaled so that the sum adds to 100, with higher numbers indicating stronger influence on the model outcome. Colours refer to category of predictor variable.

Habitat, in particular woodlands and lowland dry heath communities, was the variable identified by the BRT model to have the highest relative contribution to CRGV occurrence (20.3 per cent). However, UK woodlands are not a unified entity. Ranging from the ancient trees and woodland pasture of the New Forest's old hunting grounds where CRGV clustered in 2013, to the ash woodland of the Derbyshire Dales and Peak District, the lime woods of the East Midlands and the beech woods in the Wye Valley, Cotswolds and Chilterns, the woodlands of the UK are highly diverse, each characterised by different types of trees largely influenced by geology, soils, climate and history (https://www.woodlandtrust.org.uk/visiting-woods/trees-woods-and-wildlife/woodland-habitats/exploring-woodland-habitats/; accessed January 14, 2017). They also provide a rich habitat for a wide range of wildlife, plants and fungi and this diversity makes it very difficult to isolate a single pathogen that might be the cause of CRGV. Lowland heath communities are also highly varied. Pastures were the habitat least associated with CRGV occurrence which, combined with the decreasing domestic livestock densities, suggests it is unlikely CRGV is the result of a livestock-related pathogen to which dogs are exposed while walking across pastures, either from contact with the livestock themselves or their excretions, or from the practice of applying slurry to pastures.[11] The lack of an association with pasture habitats is supported by the decreasing relative probability of CRGV presence with increasing sheep and cattle densities.

Although habitat was the main contributor to the BRT model, a range of climatic variables were identified to be of moderate importance in CRGV occurrence. CRGV cases were more likely to be diagnosed under milder (increasing AvMaxTemp_Wi/Sp/Au), wetter (increasing AvRain_Wi/Sp) conditions in the colder months as typified by the south and west of the country. However, the fact that Wales and most of south-west England (the most extreme of these) were two of the regions predicted to be the least suitable for CRGV occurrence as illustrated in the risk map (figure 4) suggests that appropriate climatic conditions on their own are insufficient; the concomitant presence of suitable habitats appears to be essential for CRGV occurrence (Wales and most of south-west England are dominated by pastures). This hypothesis is supported by the interaction identified by the BRT model between habitat and AvMaxTemp_Wi. Similarly, those years in which the disease was not reported in the New Forest region may have lacked the necessary climatic conditions (eg, colder winters) despite the habitat being suitable. By the same token, it is possible that the low-risk cluster adjacent to the New Forest area lacks either optimal climatic conditions or suitable habitat for disease occurrence.

It is interesting to note that disease distribution was associated with maximum seasonal temperatures (autumn, winter and spring) while the effect of minimum seasonal temperatures on CRGV distribution was negligible. A study of changing climate extremes associated with warming has shown that daily minimum and maximum temperatures have both been increasing globally, although the former more than the latter.[12] Climate is mostly a factor in diseases caused by pathogens that spend part of their life cycle outside the host, exposed to the environment.[13] Increasing maximum temperatures during the colder months in the UK may have provided a favourable habitat for an evolving organism or a new ecological niche for a pathogen that had always been present in the environment but was previously unable to flourish in the comparatively cooler conditions of previous decades. Isolating those climatic factors that might have played a role in the emergence of the disease (pre-2012) may assist in the development of causal models for CRGV and help identify the aetiology of the disease.

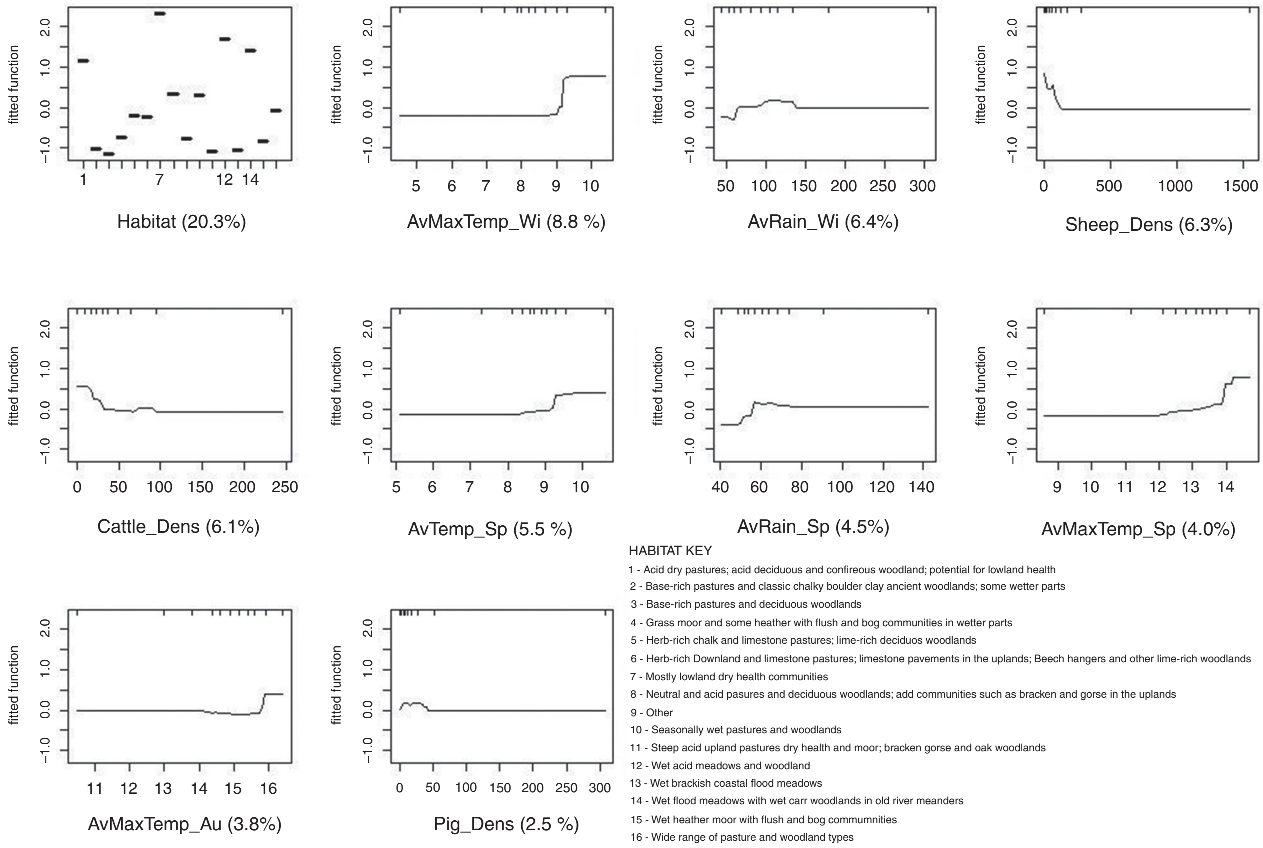

FIG 6: Partial dependence plots or boosted regression tree (BRT) profiles for the top 10 predictor variables modelling the spatial distribution of cutaneous and renal glomerular vasculopathy (CRGV) in dogs in the UK (2012–2017). Partial dependence plots show the predicted dependence between the dependent variable of the BRT model on the y axis (probability of CRGV presence) versus each predictor variable on the x axis. The top 10 predictor variables were included in this figure: Habitat, AvMaxTemp_Wi (°C), CattleDens (heads/km²), SheepDens (heads/km²), AvRain_Wi (mm), AvTemp_Sp (°C), AvRain_Sp (mm), AvMaxTemp_Sp (°C), AvMaxTemp_Au (°C) and PigDens (heads/km²). Relative contribution of each predictor variable is given in brackets and a key provided for habitat types. Habitat types in bold (1, 7 12, 14) are those associated with CRGV presence.

Limitations

CRGV was initially reported largely in the New Forest area of England resulting in an increased interest and awareness of the disease in this region which may have biased the habitat results towards woodland. However, since its inception in 2012, CRGV has been reported in other parts of the UK. In addition, the disease has been widely publicised in national and local media so that increased awareness may no longer be confined to the New Forest area and therefore any potential habitat-related biases arising from the regional focus are likely to have been mitigated over time. Only five (5 per cent) of the 103 cases provided a walking postcode that differed substantially from their residential postcode as a result of the affected dogs having accompanied their owners on holiday to, for example, the New Forest region. However, the walking postcodes of other animals may also differ from their residential postcode, especially in terms of habitat. However, this bias will have been mitigated to some extent as the resolution of all agroecological variables used in this study was 1 km² and therefore, provided dogs were walked within 1 km of their residential postcode there would have been no difference between the values of their residential and walking postcodes. However, for dogs walked greater than 1 km² from their residential postcode, there may have been a difference and therefore for future studies of ecological risk factors it is important to obtain both the walking and residential postcodes.

Cluster detection tests typically require some estimate of the population at risk in order to allow for

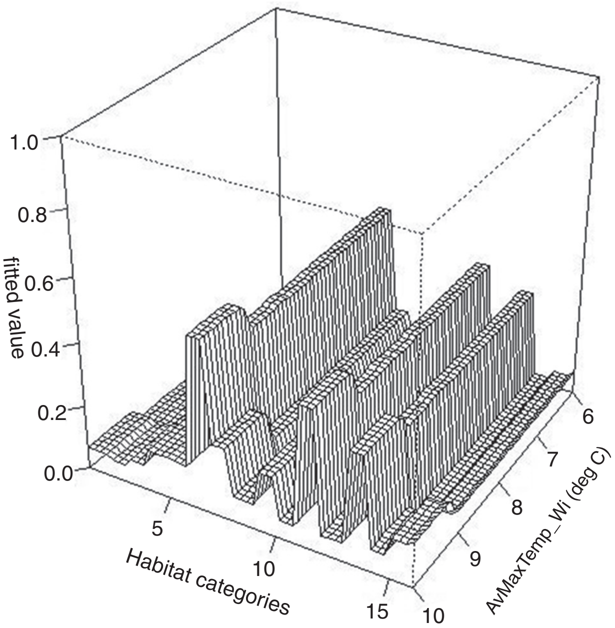

FIG 7: Interaction between the variables Habitat and AvMaxTemp_Wi in the boosted regression tree (BRT) model for the spatial distribution of cutaneous and renal glomerular vasculopathy (CRGV) in dogs in the UK (2012–2017). Interaction plots show the predicted dependence between the dependent variable of the BRT model on the y axis (probability of CRGV presence) versus the combined effect of each of the two interaction predictor variables on the x and y axes. The two predictor variables included in the interaction shown in the plot are Habitat and AvMaxTemp_Wi (°C).

identification of areas with a higher risk of disease while simultaneously compensating for the uneven distribution of the population. As this study lacked control or population-at-risk data, Kuldorff's space–time permutation statistic was implemented instead of the more commonly used space–time scan statistic. While the traditional scan statistic seeks to identify significant excess of cases within a specific space–time window and provides a measure of how unlikely it would be to encounter the observed excess of cases in a larger comparison region, the permutation statistic, on the other hand, seeks to identify areas with a higher proportion of cases compared with the remaining geographical regions of the study area. However, an important limitation of the permutation statistic is that without population-at-risk data it is not possible to determine whether identified clusters are due to an increased risk of disease, or to different geographical population distributions at different times (eg, an influx of tourists and their pets to coastal resorts during the summer months), especially when the study covers more than a single year. However, CRGV cases have generally been reported during the colder months when tourism generally falls off, mitigating the effect of this limitation to some extent and making it more likely that the identified clusters are due to increased disease risk rather than different geographical population distributions.

As the two clusters identified in south-eastern England were reasonably close to the southern boundary of the study area as defined by the physical barrier of the sea, it is necessary to acknowledge the possible existence of edge effects. Although edge effects may be negligible when dealing with large-scale effects, they can be considerable when estimating small-scale effects close to the boundary. Edge effects are usually dealt with either by using a weighting system that gives less weight to those observations near the boundary, or through the use of guard areas.[14] Unfortunately, Kuldorff's space–time permutation statistic (as implemented in SaTScan V.9.5) does not allow for the use of a weighting system. However, as none of the identified clusters intersect with a coastal boundary, and are in fact some distance inland, it is unlikely that edge effects will have substantially distorted the estimates of the space–time permutation technique in this instance.

Similarly, calibration of the BRT model also requires both disease presence and absence data. However, when lacking absence data for species distribution modelling alternatives exist in the form of pseudoabsence or background data. Background data are sampled from the whole study area in order to characterise the environmental conditions existing within it.[15] It can be argued that the use of background data has advantages over that of disease absence data as the latter can be problematic making it difficult to distinguish between absence of disease and lack of observation or reporting of disease events in an area. Alternatively, the disease species may be absent, even though the habitat is suitable for its occurrence, due to a geographical or man-made barrier preventing its spread into the area.[16] These situations can be considered 'false absences', biasing study results. Lobo *et al*[17] identified three types of absence data typically occurring in primary data sets—environmental, contingent and methodological—and insisted that to optimise prediction from species distribution models all absences should ideally be environmental ones; contingent and methodological absences being deemed 'noise'. The use of background data to characterise the environment of the study area therefore largely removes the biases associated with false absences and mimics the environmental absences required to optimise prediction from species distribution models.

In this study, we used fixed seasons although it could be argued that such an approach is not appropriate if, as reported, spring and autumn are becoming shorter in duration.[18] However, the data in this study cover a five-year period making it difficult to account for the official start of each season each year. Furthermore, the start of each season will occur over a period of weeks across the country and therefore a fixed approach in defining the seasons gives a benchmark for a unified analysis of data from different regions and different years.

Conclusion

The results of this study provide owners with a broad overview of when and where their dogs are likely to be most at risk of developing CRGV in the UK. Outbreaks displayed a distinct seasonal pattern with more than 90 per cent of cases reported between November and May while the area from which cases have been reported has expanded since 2012 to encompass most of the western and southern regions of England. The eastern parts of the country—East Anglia in particular—appear to have a decreased risk of disease. These factors, together with the association identified between disease occurrence and specific habitats (CRGV occurrence was most frequently associated with woodlands and lowland dry heath and least associated with pastures), provide dog owners with an indication of when to be most vigilant for symptoms of the disease, as early identification and treatment is critical. Further research into factors differentiating high and low-risk regions—especially the adjacent high and low-risk clusters identified in south-eastern England—has the potential to provide further information central to the epidemiology of this disease.

Contributors KBS performed all analyses and wrote the first draft of the paper. LPH and DJW compiled the case data set. All authors contributed substantially to the interpretation of data, drafting of the final manuscript and critical revision for important intellectual content. All authors approved the final version of the manuscript for submission.

Funding This research was generously funded by the Alabama Rot Research Fund (ARRF) and New Forest Dog Owners Group (NFDog).

Competing interests None declared.

References

1 Holm LP, Hawkins I, Robin C, *et al*. Cutaneous and renal glomerular vasculopathy as a cause of acute kidney injury in dogs in the UK. *Vet Rec* 2015;176:384.
2 George JN, Nester CM. Syndromes of Thrombotic Microangiopathy. *N Engl J Med Overseas Ed* 2014;371:654–66.
3 R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, 2011.
4 Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag, 2016.
5 Waller LA, Hill EG, Rudd RA. The geography of power: statistical performance of tests of clusters and clustering in heterogeneous populations. *Stat Med* 2006;25:853–65.
6 Elith J, Leathwick JR, Hastie T. A working guide to boosted regression trees. *J Anim Ecol* 2008;77:802–13.
7 Elith J, H. Graham C, P. Anderson R, *et al*. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 2006;29:129–51.
8 Barbet-Massin M, Jiguet F, Albert CH, *et al*. Selecting pseudo-absences for species distribution models: how, where and how many? *Methods Ecol Evol* 2012;3:327–38.
9 Friedman J. Greedy function approximation: a gradient boosting machine. *Ann Stat* 2001;29:1189–232.
10 Morens DM, Folkers GK, Fauci AS. The challenge of emerging and re-emerging infectious diseases. *Nature* 2004;430:242–9.
11 Rankin JD, Taylor RJ. A study of some disease hazards which could be associated with the system of applying cattle slurry to pasture. *Vet Rec* 1969;85:578–81.
12 Alexander LV, Zhang X, Peterson TC, *et al*. Global observed changes in daily climate extremes of temperature and precipitation. *J Geophys Res* 2006;111.
13 Baylis M. Potential impact of climate change on emerging vector-borne and other infections in the UK. *Environ Health* 2017;16:112.
14 Pfeiffer DU, Robinson TP, Stevenson M, *et al*. Spatial data. Spatial analysis in epidemiology. OxfordUniversity Press, Oxford, 2008:15.
15 Peterson AT, Soberón J, Pearson RG. Species' occurrence data. In: Levin SA, Horn HS, eds. Ecological niches and geographic distributions. Princeton and Oxford: Princeton University Press.
16 Hirzel AH, Hausser J, Chessel D, *et al*. Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data? *Ecology* 2002;83:2027–36.
17 Lobo JM, Jiménez-Valverde A, Hortal J. The uncertain nature of absences and their importance in species distribution modelling. *Ecography* 2010;33:103–14.
18 Jones MR, Fowler HJ, Kilsby CG, *et al*. An assessment of changes in seasonal and annual extreme rainfall in the UK between 1961 and 2009. *International Journal of Climatology* 2013;33:1178–94.