Surveillance of a Vomiting Outbreak In Dogs in the UK Using Owner-Derived And Internet Search Data

Citation for published version:
Woolley, C, Bronsvoort, M, Handel, I, Schoenebeck, J & Clements, D 2021, 'Surveillance of a Vomiting Outbreak In Dogs in the UK Using Owner-Derived And Internet Search Data', *Veterinary Record*, vol. 189, no. 9, e308, pp. 1-12. https://doi.org/10.1002/vetr.308

Digital Object Identifier (DOI):
10.1002/vetr.308

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Veterinary Record

VetRecord

Surveillance of a vomiting outbreak in dogs in the UK using owner-derived and internet search data

Journal:	*Veterinary Record*
Manuscript ID	vetrec-2020-106344.R1
Wiley - Manuscript type:	Original research
Date Submitted by the Author:	09-Feb-2021
Complete List of Authors:	Woolley, Charlotte; The University of Edinburgh The Roslin Institute; The University of Edinburgh Royal Dick School of Veterinary Studies, Bronsvoort, Mark; The University of Edinburgh The Roslin Institute; The University of Edinburgh Royal Dick School of Veterinary Studies Handel, Ian; The University of Edinburgh The Roslin Institute; The University of Edinburgh Royal Dick School of Veterinary Studies Schoenebeck, Jeffrey; The University of Edinburgh The Roslin Institute; The University of Edinburgh Royal Dick School of Veterinary Studies Clements, Dylan; The University of Edinburgh The Roslin Institute; The University of Edinburgh Royal Dick School of Veterinary Studies
Abstract:	Background In early 2020, the Small Animal Veterinary Surveillance Network reported evidence of an outbreak of acute prolific vomiting in dogs in the UK. The aims of this study were to investigate whether there was evidence for a vomiting outbreak in Dogslife and Google Trends data and to describe its characteristics. Methods Incidence of Dogslife vomiting reports and the Google search index for 'dog vomiting' and 'puppy vomiting' between December 2019 to March 2020 was compared to the respective data from the same months in previous years. Risks for dogs vomiting and factors influencing veterinary attendance in Dogslife were identified using multivariable logistic regression. Results This study confirmed a vomiting outbreak was evident in UK dogs between December 2019 and March 2020 using data from Dogslife and Google Trends. The odds of a vomiting incident being reported to Dogslife was 1.51 (95% CI: 1.24 – 1.84) in comparison to previous years. Dogslife data identified differences in owner-decision making when seeking veterinary attention and identified factors associated with dogs at higher odds of experiencing a vomiting episode. Conclusion Owner-derived data including questionnaires and internet search queries should be considered a valid, valuable source of information for veterinary population health surveillance.

Surveillance of a vomiting outbreak in dogs in the UK using owner-derived and internet search data

C S C Woolley,* I G Handel, B M Bronsvoort, J J Schoenebeck, D N Clements

The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Easter Bush Campus, Midlothian EH25 9RG

*Corresponding author. The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Easter Bush Campus, Midlothian EH25 9RG. Email: s1576743@sms.ed.ac.uk

ABSTRACT

Background

In early 2020, the Small Animal Veterinary Surveillance Network reported evidence of an outbreak of acute prolific vomiting in dogs in the UK. The aims of this study were to investigate whether there was evidence for a vomiting outbreak in Dogslife and Google Trends data and to describe its characteristics.

Methods

Incidence of Dogslife vomiting reports and the Google search index for 'dog vomiting' and 'puppy vomiting' between December 2019 to March 2020 was compared to the respective data from the same months in previous years. Risks for dogs vomiting and factors influencing veterinary attendance in Dogslife were identified using multivariable logistic regression.

Results

This study confirmed a vomiting outbreak was evident in UK dogs between December 2019 and March 2020 using data from Dogslife and Google Trends. The odds of a vomiting incident being reported to Dogslife was 1.51 (95% CI: 1.24 – 1.84) in comparison to previous years. Dogslife data identified differences in owner-decision making when seeking veterinary attention and identified factors associated with dogs at higher odds of experiencing a vomiting episode.

Conclusion

Owner-derived data including questionnaires and internet search queries should be considered a valid, valuable source of information for veterinary population health surveillance.

INTRODUCTION

In February 2020, the Small Animal Veterinary Surveillance Network (SAVSNET) reported a potential outbreak of acute prolific vomiting in dogs in the UK (1), which has since been confirmed to be statistically outwith the normal fluctuation of seasonal enteric disease (2). The outbreak was suspected to have started in November 2019 and was confirmed to have subsided by April 2020, with the peak of the outbreak between December 2019 and March 2020 (3). Canine enteric coronavirus (CECoV) was isolated from 60% of 15 faecal samples and 55% of 11 vomit samples collected from suspected prolific vomiting cases collected by SAVSNET, in comparison to 0% of control samples, which has led to the virus being a suspected cause of the outbreak. However, only 21% of 33 oral swabs tested positive for CECoV (2). CECoV can be asymptomatic in the canine population and a cross-sectional survey of 249 UK dogs reported that 2.8% tested positive for the virus without presenting with disease (4). Furthermore, CECoV is generally perceived as causing enteritis with mild diarrhoea (5,6), rather than the prolific vomiting reported by SAVSNET, with at least 5 vomiting episodes in a 12 hour period and prolonged malaise (1). Therefore, it remains uncertain whether CECoV was the cause of the vomiting outbreak or presented as a co-morbidity.

Studies estimating the prevalence of vomiting in UK dogs have had poor concurrence. VetCompass, a large-scale study of the clinical records from domestic animals under primary veterinary care at participating UK clinics, reported prevalence estimates of vomiting in Labrador Retrievers of 3.60 – 3.80% (7,8). Owner-based studies have reported higher estimates, with cumulative incidence of vomiting of 5.7% – 9.0% in Norwegian Labrador Retrievers (9) and prevalence in dogs in Northern Ireland of 21.1% (10) and England of 25% (11). Dogslife reported that within its cohort of over 6000 Labrador Retrievers, owners completed questionnaires for 4678 dogs and recorded vomiting episodes 2601 times, but only took their dog for veterinarian consultation 28% of the time (12). This suggests that vomiting is underreported in primary-care data and might explain the lower prevalences in comparison to owner-reported data.

Owner-based longitudinal cohort studies such as Dogslife provide a unique opportunity to study epidemics within the wider population and thus avoid the underestimation of their scale. Dogslife also provides information that primary-care practices do not collect, such as factors involved in owners seeking veterinary care, details of dogs' demography, morphological factors such as weight and height and lifestyle factors such as diet and exercise (13). These factors can be vital in determining the epidemiological risks in the population and can inform owners, breeders and veterinarians about the best disease prevention methods.

Another potential source of data for the surveillance of veterinary disease that is not reliant on primary-care consultations is internet search queries. In recent years, increasing numbers of people are turning to the web for healthcare advice (14,15), which has led to the development of "infodemiology": the study of the internet to reveal the distribution and determinants of health information (16,17). Google Trends (a source of search query data from Google) has been used to study a range of health conditions including multiple sclerosis (18), depression (19), zika virus (20) and more recently, COVID-19 (21) in humans and tick paralysis in cats and dogs (22). Similar to owner-based questionnaires, search query data is unique in that it is able to provide health insights within the wider population.

It is vital that comprehensive and reliable methods for the surveillance of disease in pet populations are developed so that the public can be accurately advised about outbreaks in a timely fashion to prevent large-scale epidemics and epidemiological risks can be identified. The aims of this study were to investigate whether the vomiting outbreak reported by SAVSNET was evident in Dogslife and Google Trends data, to describe the characteristics of the outbreak period in comparison with data extracted from earlier years, to summarise the clinical presentation of cases in Dogslife, to identify factors that were involved in Dogslife owners seeking veterinary treatment for their dogs vomiting and to identify risk factors associated with an owner reporting a vomiting episode to Dogslife.

MATERIALS AND METHODS

All data analysis was carried out using R statistical software. An example of the code, including the specific packages and functions used for this study, is available at https://github.com/CharlotteWoolley/vomiting_outbreak_uk_dogs. All data are reported to two decimal places except for *p-values*, which are reported to three decimal places. The STROBE (23) reporting guidelines were adhered to in this study.

Collection, data cleaning and processing of Dogslife data

Dogslife is a longitudinal, online study of the health of pedigree UK Kennel Club registered Labrador Retrievers in the UK. Recruitment to Dogslife began in July 2010 and continues at the time of writing, with extensive details of the study design published previously (13). Dogslife was approved by the University of Edinburgh Veterinary Ethical Review Committee (Ref: 7.5.09) and Human Ethical Review Committee (Ref: HERC_161_17). Owners supply demographic and geographic information when they register on the Dogslife website and complete regular online questionnaires about their dog's morphology, lifestyle and illness incidences. Data for this study were collected during the first 10 years of Dogslife from December 2010 to March 2020, via routine online reporting.

Dogslife data were thoroughly cleaned prior to analysis and all variables were assessed for quality using pre-established procedures such as the 'NLME- A' method for cleaning growth variables (24). Postcode data were joined to other demographic factors available from the Office for National Statistics (ONS), such as the rural urban status and geographical region (25). Data from the ONS is licensed under the Open Government Licence V. 3.0. For further detail about data cleaning, please refer to Supplementary File 1. The total number of Dogslife reports between December 2010 to March 2020 after data cleaning was 56411.

When Dogslife owners sought veterinary treatment for their dog due to an illness episode, they were asked questions about their dog's diagnosis and treatments according to the information given to them by their veterinarian. When Dogslife owners did not seek veterinary treatment for their dog due to an illness episode, owners were asked to give a speculative reason for their dog's illness. The total numbers of vomiting and non-vomiting Dogslife reports were grouped into the season of interest (December 2019 to March 2020) and into previous similar seasons in earlier years (December to March in the years 2010 to 2019). Of the vomiting reports recorded by owners in this subset, 96.24% had completed vomiting start dates. Start dates were used where available to remove vomiting episodes that preceded the month of December for each season, leading to the removal of 208 vomiting reports. The mean time lag between start dates and reporting dates in the remaining data was 16.50 (SD: 28.67, 95% CI: 14.94 – 18.06, Median: 10, IQR: 17) days. Table 1 displays the summary statistics of Dogslife data broken down into the subsets used in analyses for this study.

Table 1: Summary statistics of Dogslife data in the season of interest (December 2019 to March 2020) and previous seasons (December to March in the years 2010 to 2019)

Variable	Previous seasons (2010-2019)	Season of interest (2019-2020)
Total reports	17486	1630
Age category		
Under 1 year	7908 (45.22%)	145 (8.90%)
1 to 3.49 years	5351 (30.60%)	315 (19.33%)
3.5 to 6.99 years	3512 (20.08%)	595 (36.50%)
7 to 10.49 years	714 (4.08%)	575 (35.28%)
Missing	1 (0.01%)	0 (0.00%)
Sex		
Female	8558 (48.94%)	832 (51.04%)
Male	8927 (51.06%)	798 (48.96%)
Missing	1 (0.01%)	0 (0.00%)

Extraction and processing of Google Trends data

Google Trends data are publicly available from their website (https://www.google.com/trends). It is an index of search queries; a measure of the proportion of queries within a particular geographical region at a particular time divided by the total number of queries at that region and time. For each search query, the index maximum is scaled at 100. The data is cached daily and in order to average multiple samples, a delay of one day is required to obtain a new sample. Google Trends anonymity threshold returns 0 when a search term is obsolete at a particular location or geographical location (26).

Monthly search query data was extracted from Google Trends using the terms "dog vomiting" and "puppy vomiting" for the time period between December 2010 and March 2020 and combined by taking the average Google hits index of the two queries. To allow for variation in data scrapes, the data were extracted on 3 separate occasions on the 27^{th}, 28^{th} and 29^{th} of May 2020 and the mean value across the scrapes was used. It was evident that there was a gradual increase in searches for "dog vomiting" and "puppy vomiting" from 2010 to 2020, which was independent of seasonal variation. This indicates that proportionally more internet users are using the internet to seek veterinary healthcare information in recent years. To attempt to minimise this potential bias, the linear trend was removed from the data by calculating the residuals of a linear model that predicted hits index by date. Both the original and the 'detrended' data are reported.

Statistical methods used to analyse the vomiting outbreak period in Dogslife and Google Trends data

To assess whether there was a peak in Dogslife vomiting reports indicative of an outbreak in the season of interest, the monthly totals of vomiting episodes, the corresponding vomiting start dates of the episodes and the vomiting incidence rate (the percentage of vomiting reports out of total Dogslife reports per month) were plotted over time from 2010 to 2020. Chi-squared tests were performed to assess whether vomiting incidence was higher than would be normally expected based on the previous seasons and whether there were differences in the frequency of owners taking their dog to a veterinarian, the frequency and duration of symptoms between the season of interest and previous seasons. To account for the aging population of the Dogslife cohort, these tests were performed with and without age stratification (using the Mantel-Haenszel adjustment) by the age groups indicated in Table 1. The age groups were arbitrarily selected to separate dogs under 1 year due to differences in Dogslife reporting, to divide the remaining data within the confines of Dogslife (where the oldest dog was less than 10.5 years old) and to approximately capture the four periods of Labrador Retriever aging as reported by Wang and colleagues (27).

To assess whether there was a peak in the Google Trends search index for 'dog vomiting' and 'puppy vomiting' indicative of an outbreak in the season of interest, the combined data was plotted over time from 2010 to 2020. A normal distribution of the combined Google Trends search index means of all seasons prior to the season of interest was fitted by maximum likelihood. To test whether the Google Trends search index in the season of interest was greater than would be normally expected, a one-tailed statistical test was defined, which estimates the probability of observing the value of the season of interest value or greater under the null distribution based on the combined Google Trends search index means in previous seasons.

The risks for dogs vomiting and factors influencing veterinary attendance in the Dogslife vomiting outbreak indicated were modelled using multivariable logistic regression. All variables included in the models were assessed for collinearity using 'one-hot' dummy variable encoding and those with strong correlations ($r > 0.50$ or $r < -0.50$) were not included in the models. For the vomiting risk model, 1162 Dogslife reports were included; 128 from owners of dogs who experienced a vomiting episode and 1034 from owners of dogs that did not. For the veterinary attendance model, 117 Dogslife reports were included; 37 from owners of dogs who attended the veterinarian for vomiting and 80 from owners of dogs that did not. For details of the model design process, variable selection and model fit and performance tests, please refer to Supplementary File 2 and Table S1.

RESULTS

Confirmation of the peak in Dogslife vomiting incidence and Google Trends search index

The monthly number, start dates and incidence rate of Dogslife vomiting reports and the monthly original and detrended Google Trends search index between December 2010 and March 2020 are shown in Figure 1. There were visual peaks evident in the season of interest in comparison to previous seasons in both the Dogslife and the Google Trends data.

In the season of interest, 146 (8.96%) of Dogslife reports involved a vomiting episode, while 1484 (91.04%) did not. In previous seasons, 1215 (6.95%) of Dogslife reports involved a vomiting episode, while 16271 (93.05%) did not. The peak vomiting incidence in Dogslife data was reached in February 2020 at 13.82%. The incidence of Dogslife vomiting in the season of interest had increased odds of 1.51 (95% CI: 1.24 – 1.84) in comparison with previous seasons after adjustment for age of dogs (as shown in Table 2). There were no differences in the odds of veterinary visit occurrence, the frequency or the duration of vomiting episodes reported to Dogslife in the season of interest in comparison to previous seasons after adjustment for the age of dogs (as shown in Table 2).

Table 2: The odds of vomiting incidence, veterinary visit occurrence, frequency and duration of vomiting episodes reported in Dogslife between December 2019 and March 2020 in comparison to previous seasons between December and March from 2010 to 2019.

Variable of comparison	Original odds ratio with 95% CI	Dog age adjusted (Mantel-Haenszel) odds ratio with 95% CI
Incidence of vomiting episodes	1.32 (1.10 – 1.58)	1.51 (1.24 – 1.84)
Proportion of vomiting episodes that led to a veterinary visit	1.51 (1.04 – 2.18)	1.14 (0.75 – 1.72)
Proportion of vomiting episodes that were prolific (where vomiting occurred more than five times a day)	1.52 (0.98 – 2.37)	1.22 (0.75 – 2.00)
Proportion of vomiting episodes that had a long duration (where symptoms lasted longer than 48 hours)	1.08 (0.68 – 1.73)	0.82 (0.49 – 1.39)

A normal distribution of the original combined Google Trends search index means of previous seasons was fitted based on the mean and standard deviation (49.84 ± 12.83) of the data and compared to the combined Google Trends search index mean (83.77) of the season of interest. The peak original Google Trends search index was reached in January 2020 at 100.00. The Google Trends search index mean of the season of interest was greater ($p = 0.004$) than the Google Trends search index means computed from previous seasons. A normal distribution of the detrended combined Google Trends search index means of previous seasons was fitted based on the mean and standard deviation (-0.12 ± 4.42) of the data and compared to the combined Google Trends search index mean (10.40) of the season of interest. The peak detrended Google Trends search index was reached in January 2020 at 26.83. The Google Trends search index mean of the season of interest was greater ($p = 0.009$) than the Google Trends search index means computed from previous seasons. Figure 2 illustrates the fitted distribution curve and combined Google Trends search index means of the previous seasons and in the season of interest for both the original and detrended data.

Description and analysis of the presentation of vomiting in Dogslife reports between December 2019 and March 2020

The geographical distribution of Dogslife reports between in the season of interest is shown in Figure 3. Geographical location was not associated with vomiting risk during multivariable and univariable analyses. The percentage of veterinary diagnoses, owners' reasons and veterinary treatments for vomiting episodes in Dogslife reports in the season of interest are shown in Figure 4. The percentage of vomiting episodes that led to a veterinary visit, the frequency and duration of vomiting in Dogslife reports in the season of interest and in previous seasons is shown in Figure 5.

In the season of interest, 49 (33.56%) vomiting episodes resulted in a veterinary visit, while 97 (66.44%) did not. In previous seasons, 305 (25.10%) vomiting episodes resulted in a veterinary visit, while 910 (74.90%) did not. In the season of interest, 29 (19.86%) of vomiting episodes were considered prolific, 108 (73.97%) were not prolific and 9 (6.16%) were not reported. In previous seasons, 179 (14.73%) of vomiting episodes were considered prolific, 1016 (83.62%) were not prolific and 20 (1.65%) were not reported. In the season of interest, 24 (16.44%) of vomiting episodes were considered long duration, 112 (76.71%) were short duration and 10 (6.85%) were not reported. In previous seasons, 183 (15.06%) of vomiting episodes were considered long duration, 925 (76.13%) were short duration and 107 (8.81%) were not reported.

In the season of interest, 45 (30.82%) vomiting episodes were reportedly caused by something that the dog ate, 44 (30.14%) as an illness due to a virus or 'bug', 39 (26.71%) were reported as unknown or not reported, 12 (8.22%) as caused by another health reason and 6 (4.11%) as due to environmental or dietary reasons. Of the vomiting episodes that resulted in a veterinary visit, 27 (61.36%) were treated with anti-emetics, 13 (29.55%) with gastrointestinal protectors, 10 (22.73%) with antibiotics, 10 (22.73%) with probiotics, 6 (13.64%) with pain relief, 6 (13.64%) with dietary changes, 6 (13.64%) with fluids, 5 (11.36%) with undefinable treatments, 5 (11.36%) were not given a treatment, 2 (4.55%) with other treatments and 1 (2.27%) with steroids.

Risk factors associated with an owner-reported vomiting episode between December 2019 and March 2020

Nine variables were found to be potentially influential for owners reporting that their dog experienced a vomiting episode in the season of interest. Figure 6 displays a forest plot of the parameter estimates. The area under of the curve (AUC) of the vomiting risk model was 69.68% (95% CI: 64.82% – 74.54%).

The age of the dog, whether the dog had travelled abroad since the owner's last visit to the Dogslife website and whether the dog had been wormed in the last 9 months were not associated with owners reporting that their dog experienced a vomiting episode. Owners of dogs that had bathed their dog since their last visit to Dogslife had increased odds (1.53, 95% CI: 1.04 – 2.25) of reporting that their dog experienced a vomiting episode. The quantity of food that the dog was fed was positively associated with increased odds (4.58, 95% CI: 1.56 – 12.68) of owners reporting that their dog experienced a vomiting episode for each additional kilogram of food their dog ate. Owners of dogs that had experienced another vomiting episode in the last 4 months preceding the current episode had increased odds (2.29, 95% CI: 1.51 – 3.53) of reporting that their dog experienced a vomiting episode. Owners of dogs that were insured had increased odds (2.50, 95% CI: 1.45 – 4.62) of reporting that their dog experienced a vomiting episode. Owners of dogs that were fed titbits had reduced odds (0.55, 95% CI: 0.32 – 0.98) of reporting that their dog experienced a vomiting episode. The weight of the dog was negatively associated with reduced odds (0.95, 95% CI: 0.91 – 0.98) of owners reporting that their dog experienced a vomiting episode for each kilogram that their dog's weight decreased.

Factors influencing Dogslife owners' veterinary attendance when their dog experienced a vomiting episode between December 2019 and March 2020

Ten variables were found to be potentially influential to owner veterinary attendance when their dog experienced a vomiting episode in the season of interest. Figure 7 displays a forest plot of the model output. The AUC of the veterinary attendance model was 91.45% (95% CI: 86.33% – 96.57%).

The age of the dog, whether the dog had been wormed in the last 9 months, geographical area classification, whether there was another dog in the household, frequency of vomiting and insurance status were not associated with owners taking their dog to the veterinarian with a vomiting episode. Owners of dogs that had an increased duration of vomiting (more than 48 hours) had increased odds (9.21, 95% CI: 2.55 – 39.74) of taking their dog to the veterinarian with a vomiting episode. Owners that smoked had increased odds (11.59, 95% CI: 2.22 – 89.36) of taking their dog to the veterinarian with a vomiting episode. Owners that were retired had increased odds (10.54, 95% CI: 2.18 – 64.79) of taking their dog to the veterinarian with a vomiting episode in comparison to families. Owners that allowed their dog to sleep in the same room as another human in the household had increased odds (4.53, 95% CI: 1.27 – 18.81) of taking their dog to the veterinarian with a vomiting episode.

DISCUSSION

This study confirmed that a vomiting outbreak was evident in UK dogs between December 2019 and March 2020 using data from Dogslife and Google Trends. The characteristics of vomiting in the outbreak period in comparison with data extracted from earlier years were successfully described. Factors that were involved in Dogslife owners seeking veterinary treatment for their dogs vomiting and risk factors associated with an owner reporting a vomiting episode to Dogslife were identified. To our knowledge, this is the first time Dogslife data has been used to describe a disease outbreak and the first time that Google Trends has been used for canine health surveillance in the UK.

During the outbreak period, peak incidence rate of vomiting reports in Dogslife reached almost 14% in February 2020, which is over double the peak proportion of consults involving vomiting reported by SAVSNET at approximately 6%, also in February 2020 (1,3). However, Dogslife, Google Trends and SAVSNET data all exhibit peaks in vomiting that are approximately double the rates preceding the outbreak period, demonstrating strong concurrence between the datasets. The peak in Google Trends search index for ‘dog vomiting’ and ‘puppy vomiting’ occurs slightly earlier, in January 2020, which reflects the more immediate nature of owners seeking healthcare information in comparison to reporting their dog’s illness to Dogslife or a veterinary practice. Dogslife vomiting incidence began to decline in March 2020, but the increased odds of 1.51 (95% CI: 1.24 – 1.84) between December 2019 to March 2020 in comparison with previous seasons (after adjustment for age of dogs) are indicative of a substantial outbreak. Only one third of Dogslife vomiting reports led to a veterinary visit during the outbreak period, which may explain why SAVSNET vomiting consultation rates were so much lower than Dogslife vomiting incidence rates. This highlights the importance of owner-based data and the problem of the “symptom iceberg”: the information from the wider population that is missed when collecting healthcare record data alone (28).

Vomiting episodes reported to Dogslife in the outbreak period did not differ in terms of rate of veterinary attendance, the frequency of vomiting occurrence or the length of vomiting duration in comparison with Dogslife vomiting reports in the previous seasons, after adjusting for age. This suggests that the vomiting episodes in the outbreak period were not any more 'severe' than would be normally expected. Furthermore, only 19.86% of vomiting episodes during the outbreak were described as prolific (occurring more than five times a day) and only 16.44% lasted over 48 hours. These results differ from SAVSNET's, who described prolific vomiting with 5 episodes or more in a 12 hour period and prolonged malaise (1). Disparity in the classification of 'prolific' between Dogslife and SAVSNET due to differences in data collection may have led to distinctions in the interpretation of the symptoms. Another explanation is that dogs presented at veterinary clinics with vomiting during the outbreak were the most severe cases. In support of this theory, when a vomiting episode was considered prolific there was a trend of increased odds (2.85, 95% CI: 0.85 – 10.22) and when a vomiting episode lasted longer than 48 hours there were increased odds (9.21, 95% CI: 2.55 – 39.74) of an owner taking their dog to a veterinarian. Delayed onset of prolific vomiting may have contributed towards the owner's decision to consult a veterinarian. Primary care consultation data is limited by owners' perceptions of which diseases require veterinary treatment and are therefore reported (29). An illness outbreak that contains diverse symptom severities may be particularly susceptible to bias and not accurately represent the characteristics of the illness in the general population.

Demographic factors that increased the odds of an owner taking their dog to a veterinarian with a vomiting episode during the outbreak period included if the owner was a smoker, if the owner was retired (in comparison to families) and if the owner allowed their dog to sleep in the same room as another human in the household. Previous research has reported that owners who have a stronger bond with their pet are more likely to be female, on a lower household income, solely responsible for their pet, have lower education levels, have no children under 18 and are more likely to seek veterinary care more often and at a higher level (30). Furthermore, dog owners who sleep in the same room as their dogs have been reported to had higher attachment scores (31). The results of this study agree with this research, although data about the sex, income or education levels of Dogslife owners were not assessed. Dogslife has previously reported misclassification of smokers as non-smokers during validation of the questionnaire due to the perception of smoking as undesirable (32) and it is possible that this 'social desirability bias' may have affected the results of this study. Alternatively, is possible that Dogslife owners who smoke may be more likely to be unemployed or retired and have more time to notice symptoms or take their dog to the veterinarian or have a stronger bond with their dog. In support of this theory, the ONS reported that the proportion of unemployed people that smoke in the UK (29.2%) is higher than that of employed people (15.0%) (33). However, these explanations are purely speculation and many more factors may have had an impact on owner-decision making. Furthermore, the wide confidence intervals of the odds ratios that are reported in the veterinary risk factor model reflect the fact that only sparse data was available to be included (n = 117) and a larger sample size would be required to reduce the variability of data and get more accurate estimates of the odds of the factors reported affecting owner behaviour.

Although the majority of dogs were not reported to have undergone any form of diagnostic test, so the cause of vomiting was speculative, the most common suspected causes or veterinary diagnoses reported by owners were that their dog ate something to cause the episode or that the dog had a virus or 'bug', which jointly attributed over 60% of the possible causes cited. Dogs that engage in scavenging have been reported to have increased susceptibility to gastrointestinal signs (11,34) but it is often difficult to separate whether the cause of the symptoms was due to infectious agents, dietary allergy, poison or physical obstruction. Furthermore, the dogs may have fallen ill from the scavenging alone, even if there was no ongoing vomiting outbreak. Reports of a virus or 'bug' support the theory of an infectious cause of vomiting, but there are many pathogens that can cause vomiting in dogs, including CECoV as SAVSNET have suggested (2), *Campylobacter* spp. (35), giardia (36) and parvovirus (37). A novel, undetected virus could also have caused the vomiting episodes, although this is purely speculation and a more focal geographical distribution might have been expected if a novel virus was the cause. The most common veterinary treatment for vomiting were anti-emetics, with 61.36% of episodes resulting in dogs receiving them. Antibiotics were the joint third most common treatment (22.73%), despite the fact that no specific diagnoses for bacterial infection were reached and a growing effort in recent years to promote the cautious use of antibiotics due to antimicrobial resistance (38-42). The purpose of this paper was not to investigate antimicrobial usage but it is apparent that more research needs to be done to monitor their prescription rates for vomiting episodes in UK dogs.

Factors that increased the risk of an owner-reported vomiting episode during the outbreak period were if the dog had been bathed since the last visit to Dogslife, if the dog was given increasing quantities of food, if the dog had experienced another vomiting episode within the previous four months and if the dog was insured. Factors that reduced the risk of an owner-reported vomiting episode during the outbreak period included if the dog was fed titbits and if the dog had a lower weight. Geographic location was not associated with Dogslife vomiting episodes during the outbreak period, which is in contrast with SAVSNET, who reported possible hot spots in North West England, the Midlands and Southern England (43). The difference in results might be explained by limited SAVSNET data in Scotland and Northern Ireland (44) and a wider spread of geographical data across the UK in Dogslife.

It is unlikely that bathed dogs were more likely to experience a vomiting episode due to an apparent effect of bathing itself. In fact, UK water is considered amongst the safest in the world (45). A more plausible explanation is that owners of dogs who experienced a vomiting episode bathed them in response to their dog vomiting. However, Dogslife do not collect data about the reasons that owners bathe their dogs or the time of dog bathing in relation to illness episodes, so it is impossible to be certain about the direction of causality. Dogslife dogs that had experienced a vomiting episode in the previous four months may have been exposed to infectious, genetic or environmental factors, had an ongoing health issue or experienced changes to their microbiota causing dysbiosis (46), making them more susceptible to another episode. It is not surprising that owners of insured dogs had higher odds of reporting a vomiting episode, as it is logical that owners of dogs with previous health issues would choose the security of insurance. The associations between body weight, dietary factors and vomiting outcome may have numerous different explanations. Canine obesity is associated with increased inflammation and immune dysregulation (47) and dogs with increased bodyweight may have had different diets, exercise regimes or training levels affecting their susceptibility to an infectious agent during the vomiting outbreak, although no associations with these factors were identified. Large quantities of food eaten too quickly is anecdotally associated with rapid vomiting or regurgitation (48), which might explain why owners of dogs that were fed larger quantities had higher odds of reporting a vomiting episode. Finally, titbits are frequently given as a training tool, so it is possible that dogs with higher training levels were less likely to interact with other dogs or scavenge whilst on walks, thus reducing their susceptibility to infectious agents.

As is the case with all research, this study had several limitations. We recommend that readers do not rely on the *p-values* of odds ratios to infer statistical significance, but consider the confidence intervals and other results to interpret our findings (49). Google Trends data is highly processed prior to being made publicly available and should be used with caution when attempting to predict disease trends (50). The composition of Google users has changed over time and interpretation of Google Trends data can be challenging due to demographic and media coverage differences (51). This study aimed to remove the linear increasing trend of ‘dog vomiting’ and ‘puppy vomiting’ searches over time, but it is difficult to estimate how big this effect was. Additionally, the linear trend was estimated over all dates of the data which is likely to have been influenced by increases in the end of the time period and may have underestimated the size of the outbreak. Search query selection can be ambiguous and cause interpretation errors, so the terms ‘dog vomiting’ and ‘puppy vomiting’ were chosen as they were perceived to be the most widely recognised terms, but it is likely that these terms did not fully encapsulate the target data.

Questionnaire data such as Dogslife is limited by social desirability bias and 'recall bias' (the ability of the participants to accurately recall events that happened in the past). Previous Dogslife studies have attempted to account for recall bias by excluding data outside of a certain reporting timeframe, but such methods introduce subjective cut-offs and limit the data available, which was not deemed appropriate in the current study. Dogslife makes no attempt to distinguish between 'regurgitation' and 'vomiting', which may have affected the classification of vomiting by owners. However, regurgitation is less frequently observed as a clinical sign, so it is likely that most of the dogs were vomiting and not regurgitating. Dogslife is a study of Labrador Retrievers and therefore results may not be generalisable to other dog breeds. Furthermore, the behaviour of Dogslife owners may differ from dog owners who do not take part in longitudinal studies. For example, they may be more observant of their dogs, have more access to veterinary information or have other demographic factors that could affect their behaviour. Finally, the cleaning, categorising and coding of data from the questionnaires is subjective and relies on the expertise and opinion of the person performing the task.

Advantages of using Dogslife data were that it described the vomiting outbreak in a wider population of dogs than primary care and used epidemiological data to identify differences in owner-decision making and risk factors for dogs experiencing a vomiting episode. Conversely, veterinary consultation data is not capable of describing cases in the general population that are not presented for medical care and typically does not record details that can be used in such epidemiological analyses. Owner-derived and Google Trends data should be considered a valid and valuable source of information for veterinary population health surveillance.

ACKNOWLEDGEMENTS

The authors are grateful for the participants who contributed and continue to contribute data to Dogslife.

AUTHOR CONTRIBUTIONS

CW is guarantor and was responsible for planning and conceptualisation, funding acquisition, data curation and cleaning, statistical analysis, data visualisation, methodological design and validation, experimental investigation, writing programmatic code and drafting and revising the paper. MB was involved in planning and conceptualisation, funding acquisition, supervision, methodological design and revising the paper. IH was involved in planning and conceptualisation, funding acquisition, supervision, methodological design and revising the paper. JS was involved in planning and conceptualisation, funding acquisition, supervision, methodological design and revising the paper. DC was involved in project administration, obtaining resources, planning and conceptualisation, funding acquisition, supervision, methodological design and revising the paper.

FUNDING

This work was supported by an Institute Strategic Programme Grant from the Biotechnology and Biological Sciences Research Council (https://bbsrc.ukri.org/) to the Roslin Institute [BB/ J004235/1] and the lead author was funded by the Biotechnology and Biological Sciences Research Council under the EASTBIO (http://www.eastscotbiodtp.ac.uk/) doctoral training programme [BB/ J01446X/1 to CW]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

CONFLICT OF INTEREST

None declared.

REFERENCES

1 Singleton DA, Noble PJ, Radford AD, et al. Prolific vomiting in dogs. *Vet Rec* 2020;186(6):191

2 Smith SL, Singleton DA, Noble PJ, et al. Possible cause of outbreak of prolific vomiting in dogs. *Vet Rec* 2020;186(10):324

3 SAVSNET. Latest updates on the outbreak of prolific vomiting in dogs. 2020. https://www.liverpool.ac.uk/savsnet/dog-vomiting-potential-outbreak/latestupdatesontheoutbreakofprolificvomitingindogs/ (accessed 1 June 2020)

4 Stavisky J, Pinchbeck GL, German AJ, et al. Prevalence of canine enteric coronavirus in a cross-sectional survey of dogs presenting at veterinary practices. *Vet Microbiol* 2010;140(1):18–24

5 Tennant BJ, Gaskell RM, Kelly DF, et al. Canine coronavirus infection in the dog following oronasal inoculation. *Res Vet Sci* 1991;51(1):11–8

6 Stavisky J, Pinchbeck GL, German AJ, et al. Prevalence of canine enteric coronavirus in a cross-sectional survey of dogs presenting at veterinary practices. *Vet Microbiol* 2010;140(1–2):18–24

7 O'Neill DG, Church DB, McGreevy PD, et al. Prevalence of disorders recorded in dogs attending primary-care veterinary practices in England. *Vet J* 2014;9(3)

8 McGreevy PD, Wilson BJ, Mansfield CS, et al. Labrador retrievers under primary veterinary care in the UK: demography, mortality and disorders. *Canine Genet Epidemiol* 2018;5(8)

9 Sævik BK, Skancke EM, Trangerud C. A longitudinal study on diarrhoea and vomiting in young dogs of four large breeds. *Acta Vet Scand* 2012;54(8)

10 Wells DL, Hepper PG. Prevalence of disease in dogs purchased from an animal rescue shelter. *Vet Rec* 1999;144(2):35–8

11 Hubbard K, Skelly BJ, McKelvie J, et al. Risk of vomiting and diarrhoea in dogs. *Vet Rec* 2007;161(22):755–7

12 Pugh CA, Bronsvoort BM de C, Handel IG, et al. Incidence rates and risk factor analyses for owner reported vomiting and diarrhoea in Labrador Retrievers – findings from the Dogslife Cohort. *Prev Vet Med* 2017;140:19–29

13 Clements DN, Handel IG, Rose E, et al. Dogslife: a web-based longitudinal study of Labrador Retriever health in the UK. *BMC Vet Res* 2013;9(13)

14 Dutton WH, Blank G, Groselj D. Cultures of the Internet: The Internet in Britain. Oxford Internet Survey 2013. Oxford: Oxford Internet Institute, 2013

15 Kogan LR, Schoenfeld-Tacher R, Viera AR. The Internet and health information: differences in pet owners based on age, gender, and education. *J Med Libr Assoc* 2012;100(3):197–204

16 Fox S, Duggan M. Health online 2013. Washington, DC: Pew Research Center, 2013

17 Eysenbach G. Infodemiology : The Epidemiology of (Mis) information. *Am J Med* 2002;113(9):763–5

18 Moccia M, Palladino R, Falco A, et al. Google Trends: New evidence for seasonality of multiple sclerosis. *J Neurol Neurosurg Psychiatry* 2016;87(9):1028–9

19 Yang AC, Tsai SJ, Huang NE, et al. Association of Internet search trends with suicide death in Taipei City, Taiwan, 2004-2009. *J Affect Disord* 2011;132(1–2):179–84

20 Fortescue-Webb DC, Dimitrov BD. Nowcasting incidence of emergent zika virus infection and its outbreaks using Google Trends data: Examples from Brazil and Colombia. *Int J Infect Dis* 2016;53S:98

21 Walker A, Hopkins C, Surda P. Use of Google Trends to investigate loss-of-smell–related searches during the COVID-19 outbreak. *Int Forum Allergy Rhinol* 2020;10(7):839–47

22 Guernier V, Milinovich GJ, Bezerra Santos MA, et al. Use of big data in the surveillance of veterinary diseases: early detection of tick paralysis in companion animals. *Parasit Vectors* 2016;9(303)

23 von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *PLoS Med* 2007;4(10):e296

24 Woolley CSC, Handel IG, Bronsvoort BM, et al. Is it time to stop sweeping data cleaning under the carpet? A novel algorithm for outlier management in growth data. *PLoS One* 2020;15(1):1–21

25 Office for National Statistics. National Statistics Postcode Lookup UK (February 2020); Contains Ordnance Survey data © Crown copyright and database right 2020. 2020. https://geoportal.statistics.gov.uk/datasets/1951e70c3cc3483c9e643902d858355b/ (Accessed 6 March 2020)

26 Stephens-Davidowitz S, Varian H. A Hands-on Guide to Google Data. Mountain View, CA: Google, Inc, 2015

27 Wang T, Ma J, Hogan AN, et al. Quantitative Translation of Dog-to-Human Aging by Conserved Remodeling of the DNA Methylome. *Cell Syst* 2020; 11(2)

28 Elnegaard S, Andersen RS, Pedersen AF, et al. Self-reported symptoms and healthcare seeking in the general population – exploring “The Symptom Iceberg.” *BMC Public Health* 2015;15(685)

29 O’Neill DG, Church DB, McGreevy PD, et al. Approaches to canine health surveillance. *Canine Genet Epidemiol* 2014;1(2)

30 Lue TW, Pantenburg DP, Crawford PM. Impact of the owner-pet and client-veterinarian bond on the care that pets receive. *J Am Vet Med Assoc* 2008;232(4):531–40

31 Martens P, Enders-Slegers MJ, Walker JK. The emotional lives of companion animals: Attachment and subjective claims by owners of cats and dogs. *Anthrozoos* 2016;29(1):73–38

32 Pugh CA, Summers KM, Bronsvoort BM de C, et al. Validity of internet-based longitudinal study data: The elephant in the virtual room. *J Med Internet Res* 2015;17(4)

33 Office for National Statistics. Adult smoking habits in the UK: 2018. 2019. http://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandlifeexpectancies/bulletins/adultsmokinghabitsingreatbritain/2018/ (accessed 10 October 2020)

34 Stavisky J, Radford AD, Gaskell R, et al. A case-control study of pathogen and lifestyle risk factors for diarrhoea in dogs. *Prev Vet Med* 2011;99(2–4):185–92

35 Acke E. Campylobacteriosis in dogs and cats: a review. *N Z Vet J* 2018;66(5):221–8

36 Ballweber LR, Xiao L, Bowman DD, et al. Giardiasis in dogs and cats: update on epidemiology and public health significance. *Trends Parasitol* 2010;26(4):180–9

37 Bird L, Tappin S. Canine parvovirus: where are we in the 21st Century? *Companion Anim* 2013;18(4):142–6

38 Singleton DA, Noble PJM, Sánchez-Vizcaíno F, et al. Pharmaceutical prescription in canine acute diarrhoea: A longitudinal electronic health record analysis of first opinion veterinary practices. *Front Vet Sci* 2019;6(218)

39 Schmidt VM, Pinchbeck G, McIntyre KM, et al. Routine antibiotic therapy in dogs increases the detection of antimicrobial-resistant faecal Escherichia coli. *J Antimicrob Chemother* 2018;73(12):3305–16

40 Lappin MR, Blondeau J, Boothe D, et al. Antimicrobial use Guidelines for Treatment of Respiratory Tract Disease in Dogs and Cats: Antimicrobial Guidelines Working Group of the International Society for Companion Animal Infectious Diseases. *J Vet Intern Med* 2017;31(2):279–94

41 Weese JS, Blondeau J, Boothe D, et al. International Society for Companion Animal Infectious Diseases (ISCAID) guidelines for the diagnosis and management of bacterial urinary tract infections in dogs and cats. *Vet J* 2019;247:8–25

42 Marks SL, Rankin SC, Byrne BA, et al. Enteropathogenic Bacteria in Dogs and Cats:

Diagnosis, Epidemiology, Treatment, and Control. *J Vet Intern Med* 2011;25(6):1195–208

43 SAVSNET. Outbreak of prolific vomiting in dogs. 2020. https://www.liverpool.ac.uk/savsnet/dog-vomiting-potential-outbreak/ (accessed 1 June 2020)

44 Sánchez-Vizcaíno F, Jones PH, Menacere T, et al. Small animal disease surveillance: Respiratory disease. *Vet Rec* 2015;178(15):361–4

45 Water UK. Water and health. 2020. https://www.water.org.uk/advice-for-customers/water-and-health/ (accessed 5 November 2020)

46 Pilla R, Suchodolski JS. The Role of the Canine Gut Microbiome and Metabolome in Health and Gastrointestinal Disease. *Front Vet Sci* 2020;6(498)

47 Cortese L, Terrazzano G, Pelagalli A. Leptin and immunological profile in obesity and its associated diseases in dogs. *Int J Mol Sci* 2019;20(10)

48 Finlay K. How to Slow Down a Speedy Eater. 2017. https://www.akc.org/expert-advice/health/4-ways-to-slow-your-dogs-eating/ (accessed 1 August 2020)

49 Leek JT, Peng RD. Statistics: P values are just the tip of the iceberg. *Nature* 2015: 520(7549):612

50 Lazer D, Kennedy R, King G, et al. Big data. The parable of Google Flu: traps in big data analysis. *Science* 2014;343(6167):1203–5

51 Alicino C, Bragazzi NL, Faccio V, et al. Assessing Ebola-related web search behaviour: insights and implications from an analytical study of Google Trends-based query volumes. *Infect Dis poverty* 2015;4(54)

Figure 1. a) Monthly Dogslife vomiting episode reports between December 2010 and March 2020 and the corresponding vomiting start dates of the episodes, where reported in the same time period. b) The monthly incidence (%) of Dogslife vomiting episode reports between December 2010 and March 2020, with the months between December and March highlighted between either dotted lines (for seasons prior to 2019-2020) or solid lines (for the most recent season in 2019-2020). c) The combined monthly Google search index for 'dog vomiting' and 'puppy vomiting' between December 2010 and March 2020 in its original format and with the linear temporal trend removed

Figure 2. The Google search index means (grey dots) and fitted normal distribution curve (black line) of the original and detrended combined Google Trends data for ‘dog vomiting’ and ‘puppy vomiting’ based on the means and standard deviations of the Google search index means for previous seasons (December to March, 2010 to 2019), with the combined Google Trends search index mean of the season of interest highlighted (red dashed line)

Figure 3. The geographical distribution of Dogslife reports between December 2019 and March 2020, by whether the report included a vomiting episode or not

Figure 4. a). The percentage of veterinary diagnoses and owners’ reasons for vomiting by whether the dog visited a veterinarian in Dogslife reports between December 2019 and March 2020. b). The treatments for vomiting after visiting a veterinarian in Dogslife reports between December 2019 and March 2020

Figure 5. a) The percentage of vomiting episodes that led to a veterinary visit in Dogslife reports between December 2019 and March 2020 and in previous seasons between December and March from 2010 to 2019. b) The percentage of vomiting episodes that were prolific (where vomiting occurred more than five times a day) in Dogslife reports between December 2019 and March 2020 and in previous seasons between December and March from 2010 to 2019. c) The percentage of vomiting episodes that had a long duration (where symptoms lasted longer than 48 hours) in Dogslife reports between December 2019 and March 2020 and in previous seasons between December and March from 2010 to 2019

Figure 6. Forest plot of a model identifying risk factors associated with an owner-reported vomiting episode between December 2019 and March 2020 in Dogslife data

Figure 7. Forest plot of the model output for owners’ veterinary attendance when their dog experienced a vomiting episode between December 2019 and March 2020 in Dogslife data

Figure 1. a) Monthly Dogslife vomiting episode reports between December 2010 and March 2020 and the corresponding vomiting start dates of the episodes, where reported in the same time period. b) The monthly incidence (%) of Dogslife vomiting episode reports between December 2010 and March 2020, with the months between December and March highlighted between either dotted lines (for seasons prior to 2019-2020) or solid lines (for the most recent season in 2019-2020). c) The combined monthly Google search index for 'dog vomiting' and 'puppy vomiting' between December 2010 and March 2020 in its original format and with the linear temporal trend removed

1666x916mm (72 x 72 DPI)

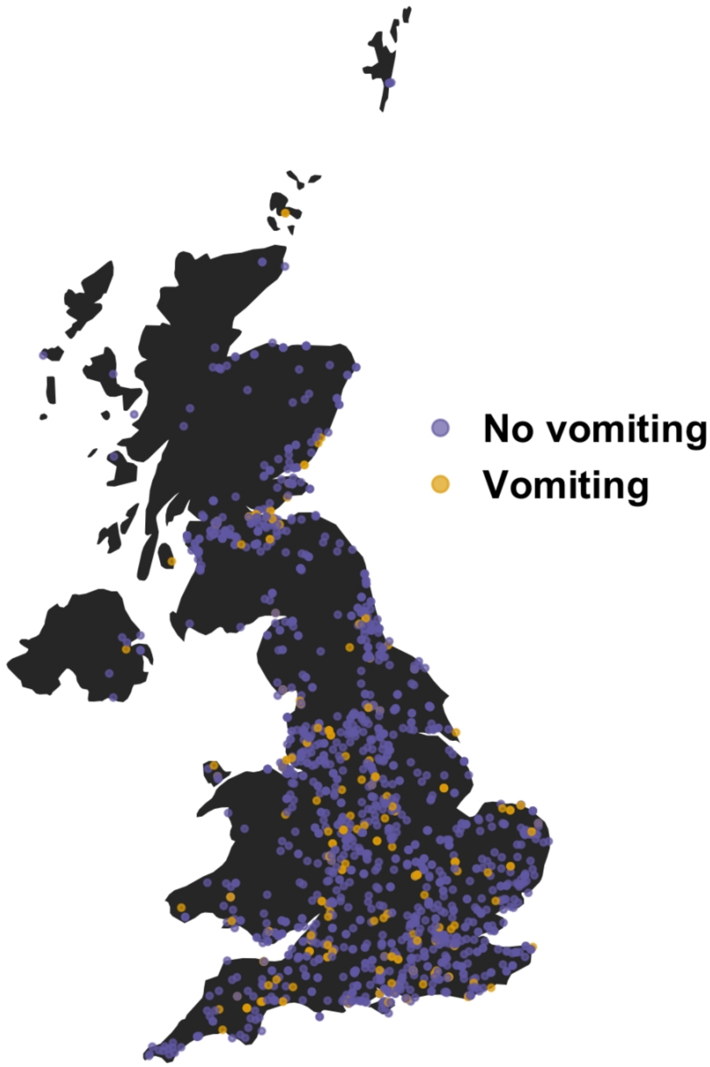

Figure 3. The geographical distribution of Dogslife reports between December 2019 and March 2020, by whether the report included a vomiting episode or not

416x541mm (72 x 72 DPI)

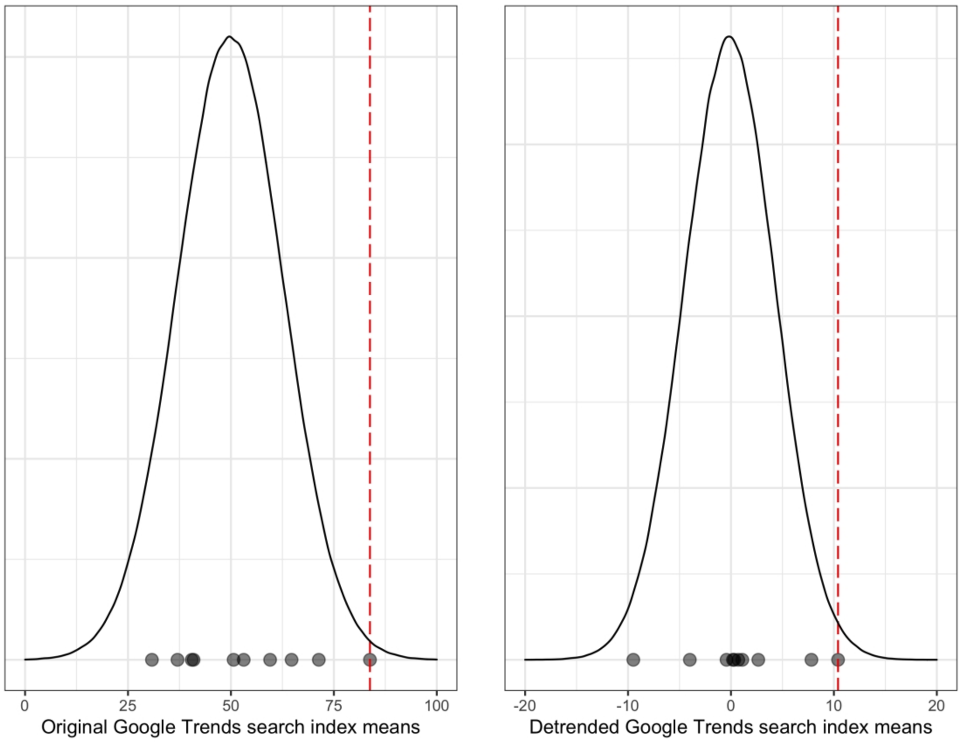

Figure 2. The Google search index means (grey dots) and fitted normal distribution curve (black line) of the original and detrended combined Google Trends data for 'dog vomiting' and 'puppy vomiting' based on the means and standard deviations of the Google search index means for previous seasons (December to March, 2010 to 2019), with the combined Google Trends search index mean of the season of interest highlighted (red dashed line)

833x624mm (72 x 72 DPI)

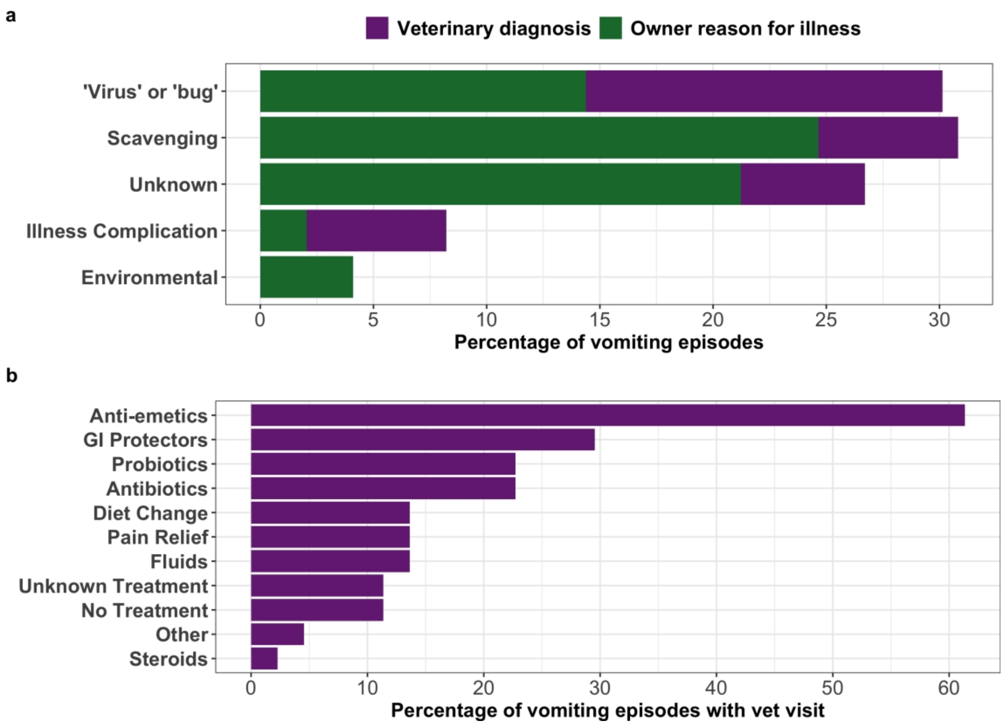

Figure 4. a). The percentage of veterinary diagnoses and owners' reasons for vomiting by whether the dog visited a veterinarian in Dogslife reports between December 2019 and March 2020. b). The treatments for vomiting after visiting a veterinarian in Dogslife reports between December 2019 and March 2020

1041x749mm (72 x 72 DPI)

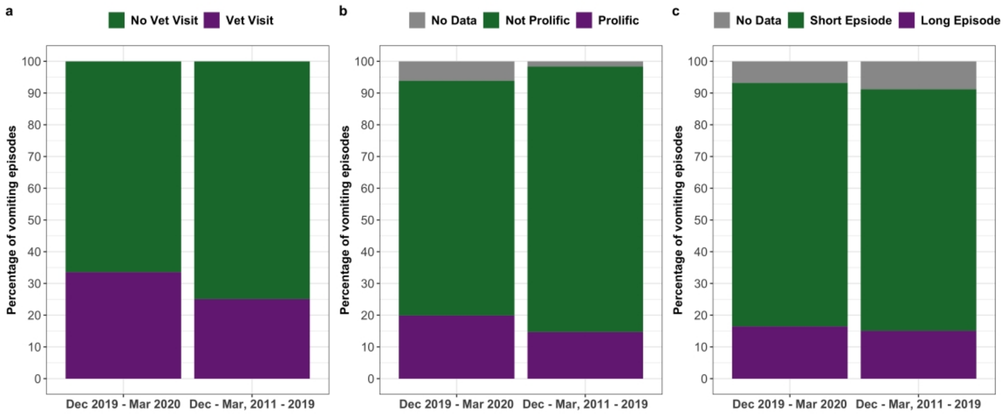

Figure 5. a) The percentage of vomiting episodes that led to a veterinary visit in Dogslife reports between December 2019 and March 2020 and in previous seasons between December and March from 2010 to 2019. b) The percentage of vomiting episodes that were prolific (where vomiting occurred more than five times a day) in Dogslife reports between December 2019 and March 2020 and in previous seasons between December and March from 2010 to 2019. c) The percentage of vomiting episodes that had a long duration (where symptoms lasted longer than 48 hours) in Dogslife reports between December 2019 and March 2020 and in previous seasons between December and March from 2010 to 2019

1458x624mm (72 x 72 DPI)

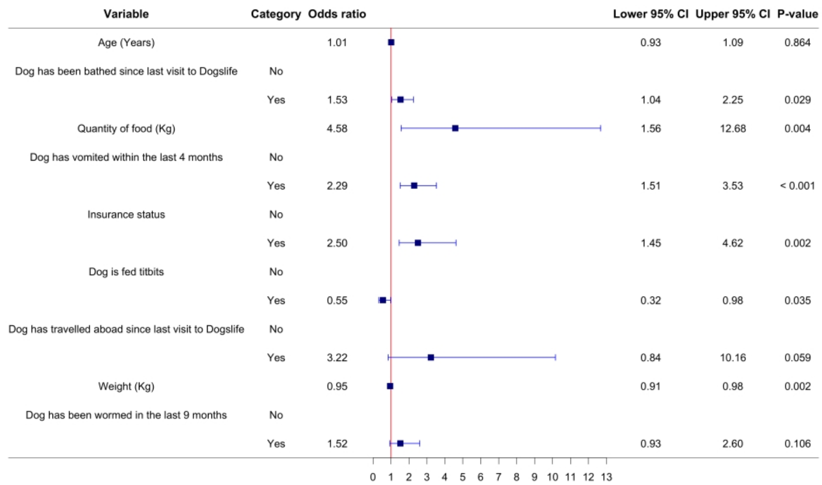

Figure 6. Forest plot of a model identifying risk factors associated with an owner-reported vomiting episode between December 2019 and March 2020 in Dogslife data

2499x1249mm (72 x 72 DPI)

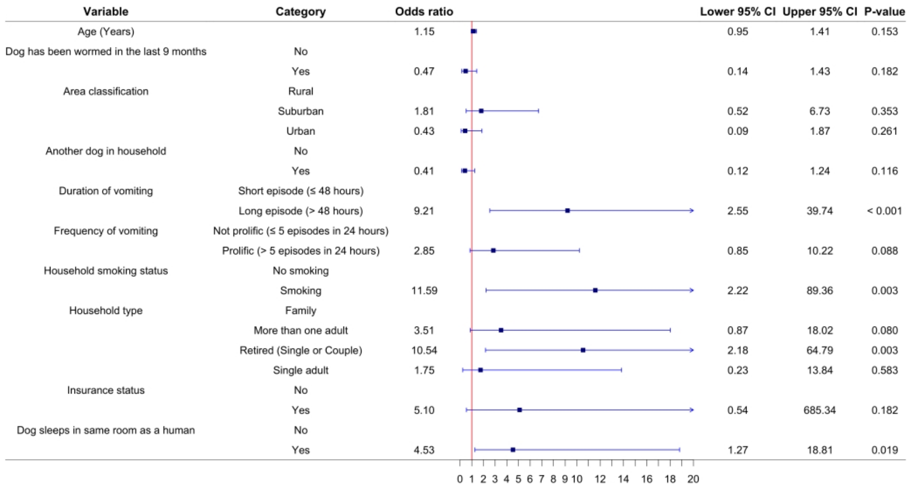

Figure 7. Forest plot of the model output for owners' veterinary attendance when their dog experienced a vomiting episode between December 2019 and March 2020 in Dogslife data

2499x1249mm (72 x 72 DPI)

SUPPLEMENTARY FILE 1: DOGSLIFE DATA CLEANING PROCESS

Duplicates were removed while maximising the information they contained by filling missing information. Quantifying missingness within the Dogslife dataset is complex due to the challenges of handling very large datasets. In many instances, data are not 'missing' in the sense that they are unobserved but instead are absent as relevant events and recorded as 'missing' to create a rectangular data frame. Missingness is also dependent on whether the data entry was duplicated or erroneous and thus removed during data cleaning, whether columns were amalgamated or separated during data cleaning and whether the website auto-filled the data entry as missing as the default setting. These factors make the levels of 'missingness' between variables directly incomparable and make it difficult to report a summary of the missingness in the database before and after data cleaning in a representative manner. For example, in the original uncleaned Dogslife dataset only 3.00% of illness start dates for a vomiting episode were missing, while 71.98% of veterinary visit dates for a vomiting episode were missing, because owners did not always take (or state whether they took) their dog to a veterinarian. Therefore, due to the complexity of reporting these figures, the authors can be contacted directly to obtain this information if needed.

For categorical variables, where data was missing due to the Dogslife website auto-completion settings or where an owner had only partially completed a questionnaire, data was chronologically down-filled from the previous data entry, except when the first data entry and succeeding ones were missing, in which case data was upfilled from the nearest data entry point chronologically. Free text variables were reduced to simple forms using a combination of text mining techniques and manual coding into categories. Event dates were corrected or removed when erroneous and used to categorise data into time periods where possible. Information that could be fitted to growth curves, such as dog weight and height and feeding quantities was cleaned using adapted versions of the published and validated 'NLME-A' method (1). Postcodes were checked for validity against UK geographical information publicly available in a dataset from the Office for National Statistics and joined to other demographic factors available such as the rural urban status and geographical region (2). Data from the Office for National Statistics is licensed under the Open Government Licence V. 3.0.

REFERENCES

1 Woolley CSC, Handel IG, Bronsvoort BM, et al. Is it time to stop sweeping data cleaning under the carpet? A novel algorithm for outlier management in growth data. *PLoS One* 2020;15(1):1–21

2 Office for National Statistics. National Statistics Postcode Lookup UK (February 2020); Contains Ordnance Survey data © Crown copyright and database right 2020. 2020. https://geoportal.statistics.gov.uk/datasets/1951e70c3cc3483c9e643902d858355b/ (Accessed 6 March 2020)

SUPPLEMENTARY FILE 2: MODEL DESIGN PROCESS, VARIABLE SELECTION AND MODEL FIT AND PERFORMANCE TESTS

Preparation of the data for a vomiting risk factor model

To assess the risk factors associated with an owner-reported vomiting episode during the outbreak period, all Dogslife reports between December 2019 and March 2020 (n=1630) were considered for inclusion into a multivariable logistic regression model.

In total, 454 dogs had multiple Dogslife reports submitted by owners, 397 of which had multiple non-vomiting reports, 20 had multiple vomiting reports and 37 had a mixture of vomiting and non-vomiting reports. To avoid including the dog as a random effect in the model, only one vomiting report was included for each dog during the outbreak period. Where dogs had a mixture of vomiting and non-vomiting reports, the vomiting reports were given priority over non-vomiting reports so that all dogs that had experienced a vomiting episode was classified as having done so. This led to the random removal of 383 reports from the dataset.

Thirty-one variables were initially considered for inclusion into this vomiting risk model, assessed for collinearity using 'one-hot' dummy variable encoding and five variables with strong correlations ($r > 0.50$ or $r < -0.50$) were removed from inclusion into the model (see Table S1). A total of 85 reports were removed due to missingness in the remaining 26 variables. Therefore, 1162 reports were included in the model dataset; 128 dogs who experienced a vomiting episode and 1034 dogs that did not.

Preparation of the data for a vomiting veterinary attendance model

To assess what factors influenced owner decision making when considering whether to take their dog to the veterinarian for a vomiting episode during the outbreak period, Dogslife vomiting reports between December 2019 and March 2020 (n=146) were considered for inclusion into a multivariable logistic regression model.

In total, eight owners had submitted two vomiting reports from one dog and two owners had submitted two vomiting reports from two dogs. To avoid including the owner or dog as random effects in the model, only one vomiting report was included for each dog during the outbreak period and only one dog was reported for each owner during the outbreak period. This led to the random removal of ten vomiting reports and two dogs from the dataset.

Twenty-three variables were initially considered for inclusion into this vomiting risk model, assessed for collinearity using 'one-hot' dummy variable encoding and three variables with strong correlations (r > 0.50 or r < -0.50) were removed from inclusion into the model (see Table S1). A total of 19 vomiting reports were removed due to missingness in the remaining 20 variables. Therefore, 117 vomiting reports were included in the model dataset; 37 dogs who attended the veterinarian and 80 dogs that did not.

Variable selection, model design and performance diagnostics

Variable selection and model design were similar for both the vomiting veterinary attendance model and the vomiting risk factor model. Logistic regression models were fitted to both datasets, containing all remaining variables as predictors identified during data preparation. These models contained numerous variables and required refinement into 'final' models. Stepwise variable selection was performed (backwards and forwards) and all variables were examined individually against the outcome variable in univariable models. Variables in the models with the lowest AICs during stepwise selection and/or those identified as having a significant effect on outcome during univariable analysis were selected for inclusion in the final models. Age was additionally included in the final models because it was considered to be a possible confounder.

After the final models were fitted, the standardised residual errors of the data points were inspected to check for outlying values and no data points were identified with an absolute standardised residual above 3. Variable inflation factor (VIF) scores were calculated as an additional step to examine collinearity of the predictor variables after model fitting and no predictor variables were found to have a VIF score greater than 3, which indicates that collinearity was not influencing the output of the model. In the final vomiting veterinary attendance model, perfect separation was evident between a predictor variable (insurance status) and the outcome variable, so Firth's bias-reduced penalized-likelihood logistic regression model was used to minimise bias. Finally, a receiver operating characteristic (ROC) analysis was performed to assess the general performance of the models.

Table S1: Variable selection for vomiting risk and veterinary attendance models during analysis of a vomiting outbreak in UK dogs using data from Dogslife reports between December 2019 and March 2020

Variable	Vomiting risk model			Veterinary attendance model		
	Considered for inclusion	Included after collinearity checks	Included after variable selection	Considered for inclusion	Included after collinearity checks	Included after variable selection
Household smoking status	Yes	Yes	-	Yes	Yes	Yes
Household demography type	Yes	Yes	-	Yes	Yes	Yes
Household urban/rural status	Yes	Yes	-	Yes	Yes	Yes
Household geographical region	Yes	Yes	-	Yes	Yes	-
Household presence of cat	Yes	Yes	-	Yes	Yes	-
Household presence of other dogs	Yes	Yes	-	Yes	Yes	Yes
Insurance status	Yes	Yes	Yes	Yes	Yes	Yes
Veterinary registration status	Yes	-	-	Yes	-	-
Working status	Yes	Yes	-	Yes	Yes	-
Sleeps with another pet	Yes	-	-	Yes	-	-
Sleeps with another human	Yes	Yes	-	Yes	Yes	Yes
Sex	Yes	Yes	-	Yes	Yes	-
Colour	Yes	Yes	-	Yes	Yes	-
Neutered status	Yes	Yes	-	Yes	Yes	-
Neutered in the last 3 months	Yes	-	-	Yes	-	-
Age	Yes	Yes	Yes	Yes	Yes	Yes

Weight	Yes	Yes	Yes	-	-	-
Vaccination status	Yes	Yes	-	Yes	Yes	-
Worming status	Yes	Yes	-	Yes	Yes	Yes
Anti-parasitic status	Yes	Yes	Yes	Yes	Yes	-
History of vomiting	Yes	Yes	Yes	Yes	Yes	-
Bathing status	Yes	Yes	Yes	Yes	Yes	-
Quantity of food	Yes	Yes	Yes	-	-	-
Frequency of feeding	Yes	Yes	-	-	-	-
Time of feeding	Yes	-	-	-	-	-
Titbit feeding	Yes	Yes	Yes	-	-	-
Diet includes wet food	Yes	Yes	-	-	-	-
Diet includes dry food	Yes	-	-	-	-	-
Diet includes raw food	Yes	Yes	-	-	-	-
Exercise frequency	Yes	Yes	-	-	-	-
Travelled abroad status	Yes	Yes	Yes	-	-	-
Frequency of vomiting	-	-	-	Yes	Yes	Yes
Duration of vomiting	-	-	-	Yes	Yes	Yes

Variables were assessed for collinearity using 'one-hot' dummy variable encoding and those with strong correlations (r > 0.50 or r < -0.50) were not included. Variables in the models with the lowest AICs during stepwise selection and/or those identified as having a significant effect on outcome during univariable analysis were selected for inclusion in the final models. Age was additionally included in the final models because it was considered to be a possible confounder.