**READ ME file**

**DESCRIPTION OF DATASET**

*from* Supplementary Information: Patterns and processes of pathogen exposure in gray wolves across North America

Ellen E. Brandell et al. 2021 *Scientific Reports*

**pop**: ‘population’ or study area ID = Alaska Peninsula (AK PEN), Denali National Park (DENALI), central-eastern Alaska (INT AK), Yukon-Charley National Preserve (YUCH), Ellesmere Island (ELLESMERE), North Slave Northwest Territories (N NWT), South Slave Northwest Territories (SS NWT), British Columbia (BC), southeastern Alaska (SE AK), Banff & Jasper National Parks (BAN JAS), Montana (MT), Yellowstone National Park (YNP), Grand Teton National Park (GTNP), Mexican wolves (MEXICAN), Ontario (ONT), Superior National Forest (SNF), and the Upper Peninsula of Michigan (MI).

**year:** biological year; year starts according to the birth month in each study area; April = British Columbia/Banff/Jasper National Parks and south, May = SE Alaska and north).

**age.cat**: age category as pup [0,1), subadult [1,3), and adult [3+].

**sex:** F female, M male, or NA.

**color:** B black, G gray, or NA.

**lat:** latitude at study area centroid.

**long:** longitude at study area centroid.

**habitat**\***:** a proxy for the presence of carnivore hosts, and was a continuous variable calculated as the product of: percent forest cover1, percent area with slope ≤20º 2, and density of hard edges (e.g., cutblocks, pipeline cuts, forest edges; R package *landscapemetrics*3). These habitat characteristics were selected because they were considered positive predictors of carnivore presence, such as grizzly bears, lynx, bobcat, coyotes, with a focus on wolves 4,5,14–16,6–13. While this proxy for carnivore presence is imperfect as carnivore distributions varied over our sampling distribution, and carnivores may select for different landscape features at different scales, it captures important features where wolves and other carnivores may interact, and therefore where cross-species pathogen transmission may occur.

**human**\***:** the number of people per 1000-km2 17, and was used as a proxy for the presence of unvaccinated dogs and synanthropic animals18.

**pop.density:** average population density (wolves/1000-km2/year). This was calculated as a basic average across years, but for populations with >1 density estimate per year, density was first averaged by year, resulting in one density estimate per year.

**pack.size:** average annual pack size (mean number of wolves/pack/year); pack sizes were averaged within a year, then averaged across years to provide one average annual pack size estimate per study area.

**standard.habitat:** standardized **habitat** column.

**standard.human:** standardized **human** column.

**standard.pop:** standardized **pop.density** column.

**standard.packsize:** standardized **pack.size** column.

**standard.latitude:** standardized **lat** column.

**standard longitude:** standardized **long** column.

***pathogen*.binary:** 0 negative, 1 positive using the titer cutoffs in Table S3.

*pathogen:*

cav = canine adenovirus

chv = canine herpesvirus

cdv = canine distemper virus

cpv = canine parvovirus-2

neo = *Neospora caninum*

toxo = *Toxoplasma gondii*

\*Study areas are assumed to be the same size: a circle with radius 200-km2 (total area = 125664-km2) based on typical wolf territory size and dispersal distance. Centroids were the approximate center of samples, verified by resident expert biologists.

**Literature Cited**

1. North American Land Change Monitoring System 30m, 2010-2015 (Landsat). *Commission of Environmental Cooperation* (2015). Available at: http://www.cec.org/north-american-land-change-monitoring-system/.

2. USGS EROS Archive - Digital Elevation - Global 30 Arc-Second Elevation (GTOPO30). *Earth Resources Observation and Science (EROS) Center* (1996).

3. Hesselbarth, M. H. K., Sciaini, M., With, K. A., Wiegand, K. & Nowosad, J. landscapemetrics: an open‐source R tool to calculate landscape metrics. *Ecography.* **42**, 1648–1657 (2019).

4. Poole, K. G., Wakelyn, L. A. & Nicklen, P. N. Habitat selection by lynx in the Northwest Territories. *Can. J. Zool.* **74**, 845–850 (1996).

5. Nielsen, S. E., Boyce, M. S., Stenhouse, G. B. & Munro, R. H. M. Modeling grizzly bear habitats in the yellowhead ecosystem of Alberta: Taking autocorrelation seriously. *Ursus* **13**, 45–56 (2001).

6. Kittle, A. M. *et al.* Landscape-level Wolf space use is correlated with prey abundance, ease of mobility, and the distribution of prey habitat. *Ecosphere* **8**, (2017).

7. Morin, S. J., Bowman, J., Marrotte, R. R. & Fortin, M. J. Fine-scale habitat selection by sympatric Canada lynx and bobcat. *Ecol. Evol.* **10**, 9396–9409 (2020).

8. O’Neil, S. T., Vucetich, J. A., Beyer, D. E., Hoy, S. R. & Bump, J. K. Territoriality drives preemptive habitat selection in recovering wolves: Implications for carnivore conservation. *J. Anim. Ecol.* **89**, 1433–1447 (2020).

9. Arjo, W. M. & Pletscher, D. H. Coyote and wolf habitat use in northwestern Montana. *Northwest Sci.* **78**, 24–32 (2004).

10. Oakleaf, J. K. *et al.* Habitat selection by recolonizing wolves in the northern Rocky Mountains of the United States. *J. Wildl. Manage.* **70**, 554–563 (2006).

11. Hebblewhite, M. & Merrill, E. Modelling wildlife-human relationships for social species with mixed-effects resource selection models. *J. Appl. Ecol.* **45**, 834–844 (2008).

12. Roever, C. L., Boyce, M. S. & Stenhouse, G. B. Grizzly bears and forestry II: Grizzly bear habitat selection and conflicts with road placement. *For. Ecol. Manage.* **256**, 1262–1269 (2008).

13. Houle, M., Fortin, D., Dussault, C., Courtois, R. & Ouellet, J. P. Cumulative effects of forestry on habitat use by gray wolf (*Canis lupus*) in the boreal forest. *Landsc. Ecol.* **25**, 419–433 (2010).

14. Mayor, S. J., Schneider, D. C., Schaefer, J. A. & Mahoney, S. P. Habitat selection at multiple scales. *Ecoscience* **16**, 238–247 (2009).

15. Milakovic, B. *et al.* Habitat selection by a focal predator (*Canis lupus*) in a multiprey ecosystem of the northern Rockies. *J. Mammal.* **92**, 568–582 (2011).

16. Kittle, A. M. *et al.* Wolves adapt territory size, not pack size to local habitat quality. *J. Anim. Ecol.* **84**, 1177–1186 (2015).

17. NASA Socioeconomic Data and Applications Center. Gridded Population of the World (GPW), v4. *EarthData* (2015). Available at: https://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-count-rev11/data-download.

18. Millán, J. *et al.* Patterns of exposure of Iberian wolves (*Canis lupus*) to canine viruses in human-dominated landscapes. *Ecohealth* **13**, 123–134 (2016).

**DESCRIPTION OF STATISTICAL MODELS**

*from* Patterns and processes of pathogen exposure in gray wolves across North America

Ellen E. Brandell et al. 2021 *Scientific Reports*

Table 2. A list of variables considered for inclusion in generalized linear mixed models predicting pathogen and parasite exposure. Variable descriptions and rationales or predictions are provided; a \* indicates the variable was included in the final *complete model*, a + indicates the variable was included in the *geographic model.*

|  |  |  |
| --- | --- | --- |
| **Variable name** | **Description** | **Rationale for inclusion / Prediction** |
| *latitude*+ | Latitude at study area centroid. | Latitude may capture geographic variation in pathogen infections; we predicted that seroprevalence decreases as latitude increases. |
| *longitude*+ | Longitude at study area centroid. | Longitude may capture geographic variation in pathogen infections. |
| *age class\**+ | Estimate of wolf age class: pup (<1), subadult (1-2), and adult (≥3). | As individuals age, they have more time to be exposed to pathogens, thus older wolves will have higher seroprevalence. Age category is less error-prone than numerical age estimates. |
| *year\** | Biological year, birth month = first month. | Pathogen exposure may be predictable by year (i.e., endemics), or unpredictable (i.e., epidemics). |
| *study area\** | Study area abbreviation. | Study area may describe variation in pathogen exposure, not accounted for by other variables. |
| *habitat quality\** | Index for habitat quality based on land cover type and topography. | A continuous estimate of the habitat quality of the study area, this covariate considers habitat characteristics that carnivores, especially wolves, positively select. This is a proxy for the presence of sympatric carnivore hosts. Prediction: seroprevalence increases with habitat quality. |
| *human density\** | Number of people/1000-km2. | Provides information about how urban the area is, and thus the potential for contact between unvaccinated dogs or synanthropic species (e.g., rodents, coyotes, raccoons, skunks, cats) and wolves. Prediction: seroprevalence increases with human density. |
| *wolf density\** | Number of wolves/1000-km2; mean annual density results in one estimate per study area. | Population density is related to direct transmission rates and environmental contamination. Prediction: seroprevalence increases with wolf density. |
| *pack size\** | Mean annual pack size; one estimate per study area. | This tells us about the daily contacts of a wolf, which differs from contact rate at the population-level. Prediction: seroprevalence increases with pack size. |
| *sex\** | Male or Female. | There is evidence that males have higher pathogen prevalence than females across many taxa and pathogens – we predict males have higher seroprevalence. |
| *coat color\** | Gray or Black. | The locus that confers black coat color in wolves is linked to beta-defensin genes, which increases the responsiveness of the innate immune system. We assume gray = missing k-locus, black = presence of k-locus. Prediction: black wolves have higher seroprevalence. |
| *age* | Estimate of wolf age; integer to two decimal places. | As individuals age, they have more time to be exposed to pathogens, thus we predicted older wolves have higher seroprevalence. |
| *social status* | Breeder or non-breeder. | Breeders typically have higher stress levels and energetic demands than non-breeders, which we predict increases seroprevalence. |
| *prey species* | Top two primary prey species. | *N. caninum* or *T. gondii* may be more prevalent in different intermediate hosts. Prediction: seroprevalence is higher where white-tailed deer are a primary prey species. |
| *pack membership* | Name of the pack the wolf was a member of when sampled. | There may be heterogeneities in pathogen exposure based on pack membership. |
| *pack density* | Number of packs/1000-km2; mean annual density results in one estimate per study area. | Contact among wolves from different packs is likely influenced by the number of packs in the population. Prediction: seroprevalence increases with pack density. |

We constructed and analyzed models predicting the probability that a wolf was exposed to a given pathogen using R v3.6.31. We tested how well geography (i.e., latitude and longitude) explained and predicted pathogen exposure compared with mechanistic predictor variables. Two models were constructed for each pathogen: a *complete model* and a *geographic model* (Eq. 1). The *geographic model*, which acted as a null/uninformative model, contained latitude and longitude, and controlled for the effect of age. The *complete model* contained selected predictor variables (Table 2). Both models included random effects (generalized linear mixed model, ‘GLMM’). Models were fit with a complementary-log-log link and a Bernoulli error distribution using the function *glmer* in the package *lme4*2. In the *complete model*, *year* and *study area* were both considered as random effects, where *year* was nested within *study area* because we posited that the effect of year differed within each study area. Nesting year within study area gave us a random effect for *study area* alone, as well as *study area\*year*. *Study area* was the only random effect considered in the *geographic model*. The form of our GLMMs was:

*Yijk =* Bernoulli(*pijk*) (Eq. 1)

*f* (*pijk*) *= 𝛽0 + 𝛽1x1ijk + … + 𝛽nxnijk + 𝜶j + ɣjk + εi*

*𝜶j* ~ Normal(0, σ2)

*ɣjk* ~ Normal(0, σ2)

where *Yijk* is the seropositive result for the *njk* trial from the *i*th individual from the *j*th study area in year *k*; *pijk* is the probability of exposure from the *i*th individual from the *j*th study area in year *k*; *xnijk* is the *i*th value of the *j*th study area in the *k*th year for the *n*th predictor; *𝛽n* are the estimated predictor coefficients; *𝜶j* is the study area-specific effect; *ɣjk* is the effect of year within that study area; *εi* is the remaining error in seropositivity. The *year* effects, including *ɣjk*, did not appear in the *geographic model*. The link function (*f*) applied is the complementary-log-log.

All metadata were collected specifically for this project such that we determined our hypotheses *a priori*3 (Table 2, Supplementary Table S1). All variables considered were expected to influence pathogen exposure. Table 2 displays variables considered for inclusion in the models, descriptions, and rationales or predictions. Each sample was assumed to be unique, given that <7% of the data were recollared wolves. If multiple age estimates were given (e.g., 3 or 4 years old), we randomly selected one age estimate. Some variables were removed prior to model building due to lack of sufficient data, including *pack membership*, *social status*, and *pack densit*y (Supplementary Table S4). *Prey species* was not included because primary prey species were too similar across study areas (e.g., a combination of elk, deer spp., moose, caribou), and after exploratory plotting, did not appear to provide additional information above *study area* and *habitat quality*. Prey species also are likely reflected in wolf density and pack size4–7. We included *age class* instead of *age* in our models because *age* was based on tooth wear and body size, and is an error-prone estimate especially for older ages8. We used coat color as a proxy for the presence of the K-locus allele, which is supported by Anderson et al. (2009)9, who found that >98% of wolves from Yellowstone and western Canada classified as ‘black’ did indeed have the K-locus genotype.

We also considered *wolf density*, *pack size*, *human density*, *habitat quality*, and *sex* as potentially important predictors of pathogen exposure (Table 2). Wolves were counted in all study areas, including annual population counts and pack size estimates. These data were typically collected during aerial or ground tracking surveys in the winter. If more than one estimate was available per year within a study area, which was common for pack sizes, they were averaged to create one annual *wolf density* (number of wolves/1000-km2/year) and one annual mean *pack size* (mean number of wolves/pack/year) value per study area. To estimate *human density* and *habitat quality*, we first had to determine how large of an area should be considered, as most areas did not have clearly defined boundaries or isolated wolf populations. We considered a range of area sizes (radius 50-km to 300-km from study area centroids), and selected a 200-km radius because *human density* and *habitat quality* were less variable in comparison with small or large radii, and it is more congruent with wolf dispersal distance10,11. *Human density* was considered to be the number of people per 1000-km2 12, and was used as a proxy for the presence of unvaccinated dogs and synanthropic animals13. *Habitat quality* was a proxy for the presence of carnivore hosts, and was a continuous variable calculated as the product of: percent forest cover14, percent area with slope ≤20º 15, and density of hard edges (e.g., cutblocks, pipeline cuts, forest edges; R package *landscapemetrics*16). These habitat characteristics were selected because they were considered positive predictors of carnivore presence, such as grizzly bears, lynx, bobcat, coyotes, with a focus on wolves17–29. While this proxy for carnivore presence is imperfect as carnivore distributions varied over our sampling distribution, and carnivores may select for different landscape features at different scales, it captures important features where wolves and other carnivores may interact, and therefore where cross-species pathogen transmission may occur. Finally, *sex* (male or female) was recorded during captures.

Before building the *complete model*, all variables were screened for collinearity using Spearman’s correlation coefficient (⍴). Human density and wolf density were highly correlated (⍴ = 0.62; Supplementary Fig. S3, S4) and thus were not included in the same model; however, as we were interested in the effects of both wolf and human density on pathogen dynamics, we ran the complete model both ways (i.e., with either wolf density or human density). All variables other than latitude and longitude were retained (i.e., correlation <0.4). Latitude was highly correlated with human density (⍴ = -0.79) and moderately correlated with wolf density (⍴ = -0.36) and habitat quality (⍴ = -0.33, Fig. 2, S3). Longitude was moderately correlated with human density (⍴ = 0.37), habitat quality (⍴ = 0.30), and proportion of black wolves (⍴ = -0.33, Fig. 2, Supplementary Fig. S3). Our models were as follows (note that the divider between *year* and *study area* denotes the nested structure *study area* + *study area\*year*):

*Complete model*:

Probability(exposure) ~ *wolf density* or *human density* *+ habitat type + pack size+*

*age class + sex + color +* (*study area | year*)

*Geographic model*:

Probability(exposure) ~ *latitude + longitude + age class + study area*

Continuous variables were standardized prior to model implementation (subtract the variable mean and divide by the standard deviation, Gelman & Hill 2007, Menard 2011). This centers all variables (mean = 0), and deviations from the mean are represented in standard deviations. Standardizing puts all continuous variables on the same scale, allowing for direct comparisons and simplifying interpretation. All models converged using the bobyqa optimizer.

Models were evaluated by root mean square error (RMSE) and area under the receiver-operator curve (AUC). RMSE and AUC provide different, important model evaluation. RMSE is a measure of model fit as it calculates the error between the observed data and the fitted model, whereas AUC provides a measure of the classification accuracy of the model; both criteria use model fixed effects. To calculate AUC, the false positive rate (1 – specificity) is plotted against the true-positive rate (sensitivity); AUC = 0.5 indicates no discrimination, AUC > 0.5 indicates that the true positive rate is higher than the false-positive rate, and AUC > 0.8 indicates excellent discrimination30. We compared the testing set and training set RMSE and AUC using four-fold cross validation31 (see Supplementary Information for training and testing group information). Supplementary Figure S5 and Table S6 display the mean RMSE and AUC across the four datasets (training and testing) per pathogen and model.

Model fit assessments included: training and testing set RMSE and AUC estimates, pseudo-R2 values (calculated with fixed effects only), Maximum Likelihood estimator convergence, and p-values (i.e., hypothesis testing, Table 2). Predictor variables were considered statistically significant at an alpha value of ≤0.05. The geographic and complete models, parameter estimation, and their evaluations used all (non-missing) data.

**Literature Cited**

1. Team, R. C. R: A Language and Environment for Statistical Computing. (2019).

2. Bates, D., Maechler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).

3. Gelman, A. & Loken, E. The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time. *Psychol. Bull.* **140**, 1272–1280 (2013).

4. Fuller, T. K. & Murray, D. L. Biological and logistical explanations of variation in wolf population density. *Anim. Conserv.* **1**, 153–157 (1998).

5. Fuller, T. K. & Sievert, P. R. Carnivore demography and the consequences of changes in prey availability. in *Conservation biology series - Cambridge* 163–178 (2001).

6. MacNulty, D. R., Tallian, A., Stahler, D. R. & Smith, D. W. Influence of group size on the success of wolves hunting bison. *PLoS One* **9**, 1–8 (2014).

7. Barber-Meyer, S. M., Mech, L. D., Newton, W. E. & Borg, B. L. Differential wolf-pack-size persistence and the role of risk when hunting dangerous prey. *Behaviour* **153**, 1473–1487 (2016).

8. Gipson, P. S., Ballard, W. B., Nowak, R. M. & Mech, L. D. Accuracy and precision of estimating age of gray wolves by tooth wear. *J. Wildl. Manage.* **64**, 752 (2000).

9. Anderson, T. M. *et al.* Molecular and evolutionary history of melanism in North American gray wolves. *Science (80-. ).* **323**, 1339–1343 (2009).

10. Fuller, T. K., Mech, L. D. & Cochrane, J. F. Wolf Population Dynamics. in *Wolves: Behavior, Ecology, and Conservation* (eds. Mech, L. D. & Boitani, L.) 161–191 (2003).

11. Jimenez, M. D. *et al.* Wolf dispersal in the Rocky Mountains, Western United States: 1993–2008. *J. Wildl. Manage.* **81**, 581–592 (2017).

12. (SEDAC), N. S. D. and A. C. Gridded Population of the World (GPW), v4. *EarthData* (2015). Available at: https://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-count-rev11/data-download.

13. Millán, J. *et al.* Patterns of exposure of Iberian wolves (Canis lupus) to canine viruses in human-dominated landscapes. *Ecohealth* **13**, 123–134 (2016).

14. System, N. A. L. C. M. Land Cover Change 30m, 2010-2015 (Landsat). *Commission of Environmental Cooperation* (2015). Available at: http://www.cec.org/north-american-land-change-monitoring-system/.

15. USGS. USGS EROS Archive - Digital Elevation - Global 30 Arc-Second Elevation (GTOPO30). *Earth Resources Observation and Science (EROS) Center* (1996).

16. Hesselbarth, M. H. K., Sciaini, M., With, K. A., Wiegand, K. & Nowosad, J. landscapemetrics: an open‐source R tool to calculate landscape metrics. *Ecography (Cop.).* **42**, 1648–1657 (2019).

17. Poole, K. G., Wakelyn, L. A. & Nicklen, P. N. Habitat selection by lynx in the Northwest Territories. *Can. J. Zool.* **74**, 845–850 (1996).

18. Nielsen, S. E., Boyce, M. S., Stenhouse, G. B. & Munro, R. H. M. Modeling grizzly bear habitats in the yellowhead ecosystem of Alberta: Taking autocorrelation seriously. *Ursus* **13**, 45–56 (2001).

19. Kittle, A. M. *et al.* Landscape-level wolf space use is correlated with prey abundance, ease of mobility, and the distribution of prey habitat. *Ecosphere* **8**, (2017).

20. Morin, S. J., Bowman, J., Marrotte, R. R. & Fortin, M. J. Fine-scale habitat selection by sympatric Canada lynx and bobcat. *Ecol. Evol.* **10**, 9396–9409 (2020).

21. O’Neil, S. T., Vucetich, J. A., Beyer, D. E., Hoy, S. R. & Bump, J. K. Territoriality drives preemptive habitat selection in recovering wolves: Implications for carnivore conservation. *J. Anim. Ecol.* **89**, 1433–1447 (2020).

22. Arjo, W. M. & Pletscher, D. H. Coyote and wolf habitat use in northwestern Montana. *Northwest Sci.* **78**, 24–32 (2004).

23. Oakleaf, J. K. *et al.* Habitat selection by recolonizing wolves in the northern Rocky Mountains of the United States. *J. Wildl. Manage.* **70**, 554–563 (2006).

24. Hebblewhite, M. & Merrill, E. Modelling wildlife-human relationships for social species with mixed-effects resource selection models. *J. Appl. Ecol.* **45**, 834–844 (2008).

25. Roever, C. L., Boyce, M. S. & Stenhouse, G. B. Grizzly bears and forestry II: Grizzly bear habitat selection and conflicts with road placement. *For. Ecol. Manage.* **256**, 1262–1269 (2008).

26. Houle, M., Fortin, D., Dussault, C., Courtois, R. & Ouellet, J. P. Cumulative effects of forestry on habitat use by gray wolf (Canis lupus) in the boreal forest. *Landsc. Ecol.* **25**, 419–433 (2010).

27. Mayor, S. J., Schneider, D. C., Schaefer, J. A. & Mahoney, S. P. Habitat selection at multiple scales. *Ecoscience* **16**, 238–247 (2009).

28. Milakovic, B. *et al.* Habitat selection by a focal predator (Canis lupus) in a multiprey ecosystem of the northern Rockies. *J. Mammal.* **92**, 568–582 (2011).

29. Kittle, A. M. *et al.* Wolves adapt territory size, not pack size to local habitat quality. *J. Anim. Ecol.* **84**, 1177–1186 (2015).

30. Hosmer, D. W. & Lemeshow, S. *Applied logistic regression*. (John Wiley and Sons Inc., 2000).

31. Hastie, T., Tibshirani, R. & Friedman, J. *The elements of statistical learning: data mining, inference, and prediction*. (Springer Science & Business Media, 2009).