

Original Article

# Stratification of Companion Animal Life Stages from Electronic Medical Record Diagnosis Data

Carina Salt, MMath,<sup>1,\*</sup> Emi K. Saito, VMD, MSPH, MBA,<sup>2</sup> Ciaran O'Flynn, MSc,<sup>1</sup> and David Allaway, PhD<sup>1,○</sup>

<sup>1</sup>Waltham Petcare Science Institute, Leicestershire, UK. <sup>2</sup>Banfield Pet Hospital Vancouver, Washington, USA.

\*Address correspondence to: Carina Salt, MMath, Waltham Petcare Science Institute, Freeby Lane, Waltham on the Wolds, Leicestershire LE14 4RT, UK. E-mail: [carina.salt@effem.com](mailto:carina.salt@effem.com)

Received: September 2, 2021; Editorial Decision Date: October 17, 2022

**Decision Editor:** Rozalyn M. Anderson, PhD, FGSA

## Abstract

Aging is a complex, multifactorial process, where different life stages reflect changes in metabolic processes, immune capacities, and genetic/epigenetic repertoires. With accumulating exposure to environmental stresses and deterioration of physiological functions, body systems become more prone to low-grade chronic inflammation and an increasing range of pathologies. We hypothesized that differential susceptibility to diseases across life span reflects phased changes in an organism's physiological capacity that may highlight when interventions may be appropriately used. Furthermore, the number of life stages may vary between species and be impacted by signalment such as breed. We tested this hypothesis using disease diagnoses data from veterinary electronic medical records containing almost 2 million cats and over 4 million dogs. Bi-clustering (on rates of disease diagnoses) and adaptive branch pruning were used to identify age clusters that could be used to define adult life stages. Clustering among diagnoses were then interpreted within the context of each defined life stage. The analyses identified 5 age clusters in cats and 4 age clusters within each of the 4 canine breed size categories used. This study, using population scale data for two species, one with differential size and life expectancies, is the first to our knowledge to use disease diagnosis data to define adult life stages. The life stages presented here are a result of a data-driven approach to age and disease stratification and are intended to support conversations between clinicians and clients about appropriate health care recommendations.

**Keywords:** Co-morbidity, Epidemiology, Health

Health span and life span can be impacted by factors across all life stages from pre-birth, through growth to adulthood and old age. Physiological change can be specific to certain developmental windows (eg, sexual development), and therefore interventions to support health may be time limited. For example, delivery method and weaning may impact immune function (1,2); sexual maturity (and age when neutered in cats and dogs) can impact behavior and growth and are risk factors in specific conditions and diseases (3–5); appropriate growth/weight gain prior to epiphyseal plate closure impact load-bearing and risk factors to osteoarthritis (6,7). At the other end of development, no active programming directly causes aging (8). The physiological criteria are less discrete because they relate more to loss of function and/or dysfunction, and these may be highly dependent on the individuals' exposure to risks, their cumulative experience, and their resilience (9).

We postulated that health conditions may reflect differential physiological capacity (eg, loss of function/dysfunction in later life stages) and that the relative risk of developing diseases through life would segment different developmental stages. These phenotype-derived life stages would then be relevant for both identifying appropriate clinical recommendations and assisting tailored lifestyle changes. The hypotheses under test here were: that the frequency of disease diagnoses would differ across life stage, that animals with different physiologies would also have different frequencies of disease diagnoses across life stage, and that these trends would be visible in a large database of canine and feline disease diagnoses data. The latter was tested by comparing and contrasting 2 companion animal species that exist in similar environments (cat and dog) and also, within the dog, comparing dogs of different body weights. The latter was investigated as size in the dog has a wide-ranging impact, effecting,

for example, mitochondrial function (10,11), disease predisposition (12–14), and life span (15,16). To summarize, the objective of this research was to identify evidence-based canine and feline life stages based on medical phenotypes from electronic medical records.

## Materials and Methods

Data were accessed from proprietary electronic medical records from Banfield Pet Hospitals. As the largest networks of primary care veterinary hospitals in the United States, Banfield operates over 1 000 locations in 42 U.S. states.

Canine and feline patients in the study data set were those that had at least one hospital visit during which the pet was examined by a veterinarian at a Banfield hospital between January 1, 2010 and August 31, 2016; were at least 1 year old but less than 16 years old at this visit; and were recorded in the pet record as spayed/neutered. Feline patients in the study included only domestic shorthair cats, a mixed breed group that accounts for the majority of Banfield feline patients. Canine patients were purebred dogs of 40 popular breeds seen at a Banfield hospital during the study period, with the 10 most popular breeds taken from each of 4 different breed size categories (<6.5 kg, 6.5–9 kg, 9–30 kg, and 30 kg+, which we refer to as “toy,” “small,” “medium,” and “large,” respectively). This set of purebred dogs accounts for over half of all Banfield canine patients. The canine breed size categories were based upon those previously defined (16), with less prevalent groups merged to adjacent groups—namely, the third and fourth groups were combined into the “medium” category, and the fifth and sixth groups combined to make the “large” category. Breeds were assigned to size categories according to the average adult weight of individuals (excluding those recorded as being over- or underweight) between 2 and 10 years old in the patient record database.

For each pet, diagnosis history during the study period was extracted, as well as demographic information on breed, birth date, gender, dates of first and last hospital visits, and last contact with a Banfield hospital. The diagnosis history was constructed from the structured diagnostic codes reported during the study period. The diagnostic codes used at Banfield hospitals include both actual illnesses (eg, “diabetes mellitus”) and—more rarely—descriptive stages (eg, “healthy pet” and “geriatric pet”). The diagnosis history initially included all diagnostic codes received at hospital visits during the study period (totaling 1 261 distinct codes). To exclude very rare codes, data were limited to include only those used at least 10,000 times in either cats or dogs during the study period ( $n = 353$  codes).

Data sets for cats and for each breed size category of dogs were analyzed separately, using the same methodology. For each combination of diagnosis code, breed, and gender, a “diagnosis profile” was constructed, showing the event count per pet-year at-risk for that diagnosis code at each year of age. Whenever a pet received a diagnosis, the age of the patient was calculated from the diagnosis date and date of birth and rounded down to the nearest whole year in order to bin into 1-year age categories. The event count was then calculated as the occurrences of a given diagnostic code within an age bin. This was used as the numerator of the diagnosis profile calculation. The denominator was the number of pets at-risk within each age group. A pet was counted as being “at-risk” from either its age at its first Banfield visit or its age on January 1, 2010, whichever was later. The minimum of the pet’s age on August 31, 2016, and the mean of the pet’s age at last contact and last hospital visit was taken as the last age at which the pet would be counted as being

at-risk. If a pet was not considered to be at-risk for the entire year of age, for example, if its first visit was partway through that year of age, then the fraction of the year it was at-risk was used. If a pet only visited a Banfield hospital once during the study period, then its time at-risk was taken to be the average of all times at-risk for the age group in which it visited and zero for all other age groups. At any point in a profile where the number of at-risk pets was equal to zero, this necessarily implied zero events, and the profile was set to zero at that point.

Individual profiles were then combined over breed (dogs only) and gender to create a single profile per diagnostic code by breed size combination (for dogs) or per diagnostic code (for cats). This was done via the use of weighted means with weights equal to the number of pets at-risk.

Bi-clustering (a technique used to cluster both rows and columns of a data set) and seriation (re-ordering so that similar rows and similar columns are placed close together) (17) were applied to each set of profiles. The bi-clustering analysis provides information on 2 aspects of the data—first, how the 1 year age groups cluster together and then second, how the diagnosis codes group together into sets of codes that have similar patterns of events per pet-year at-risk. This can help us characterize the different age groups in terms of the diagnosis codes that are more or less prevalent during that age range.

Sparse profiles—those where the count of events per pet-year at-risk was less than a parameter “diagnosis threshold” for all age groups—were removed from the data set. Various values of “diagnosis threshold” were tried before a final value was selected; this process is described shortly. The remaining profiles were scaled by their root mean square and stacked one on top of another to create a new data set with rows representing diagnoses and columns representing age groups. Hierarchical clustering (18) using the Euclidean distance metric and complete (or furthest neighbor) linkage (19) was then applied to both the rows and columns, followed by optimal leaf ordering (20) to re-order the age groups and diagnoses.

Adaptive branch pruning, using the dynamic hybrid algorithm (21), was then used to identify major clusters in the age groups and diagnoses. This method detects clusters according to their shape rather than the dendrogram height at which the cluster splits off (as in the traditional height cut-off technique) and is considered to be more flexible (21). The number of clusters detected by this algorithm is governed by a “deep split” parameter, which must be set separately for each group of variables to which the algorithm is to be applied (diagnosis codes and age groups in this case). This was set to zero for the diagnosis codes (to detect only the most major clusters), and 3 different values (described below) were tried for the age groups.

In order to avoid reporting clusters arising as artifacts of particular hyperparameter choices, combinations of “diagnosis threshold” and “age deep split” were evaluated in combination to test the sensitivity of the resulting splits to these parameters. The splits reported for cat and each breed size category of the dog were those that consistently appeared across the different combinations. Values used were between 0.05 and 0.08 diagnoses per pet-year at-risk (in increments of 0.005) for “diagnosis threshold” and values of 0, 0.5, and 1 for “age deep split.” A set of parameter values giving clusters most representative of the common splits was then chosen for the final analysis. Had the results shown extreme sensitivity to the parameter values (as shown by a lack of consistent splits), then the suitability of the methodology would have been re-examined.

Based on these results, a heat map was generated, with each cell shaded according to the (scaled) count of diagnoses per pet-year at-risk. Dendrograms representing the clustering were added to the

top, and left-hand sides of the grid, and the clusters detected by the adaptive branch pruning were additionally shown by white lines in the grid.

Principal component analyses (PCAs) were additionally carried out on each set of species/breed size diagnosis profiles (treating each diagnosis as a variable with centering and scaling to unit variance) to see whether evidence for the existence of dog and cat life stages might be seen in an unguided analysis.

Two-parameter exponential growth models (equivalent in form to the Gompertz equation) were fitted to diagnosis profiles using ordinary least squares, to examine if diseases associated with older life stages showed an exponential growth with age.

All analyses were done using R, version 3.4.1 (R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria, 2017), using libraries “seriation” (17,22), “dynamicTreeCut” (23) for the bi-clustering and cluster identification and “stats” for the PCA.

### Results

The study population consisted of almost 2.0 million cats and over 4.4 million dogs, the latter divided into 1.4 million (30.3%) toy, 0.9 million (20.3%) small, 1.3 million (30.0%) medium, and 0.8 million (19.4%) large breed dogs. Figure 1 shows the study population stratified by species (and by breed and breed size for dogs) and gender. In each canine breed size category, the predominant breeds were the following: chihuahua (54.3% of the toy), dachshund (65.6% of the small), American Staffordshire terrier (32.2% of the medium), and Labrador retriever (62.8% of the large). The total exposure time in the data was 2.9 million pet-years for toy, 1.3 million pet-years for small, 1.9 million pet-years for medium and 1.7 million pet-years for large breed dogs, and 1.7 million pet-years for cats. Within combinations of breed size and year of age, exposure time ranged between 4 308 pet-years for 15-year-old large breed dogs and 333,871 pet-years for 2-year-old toy breeds. For

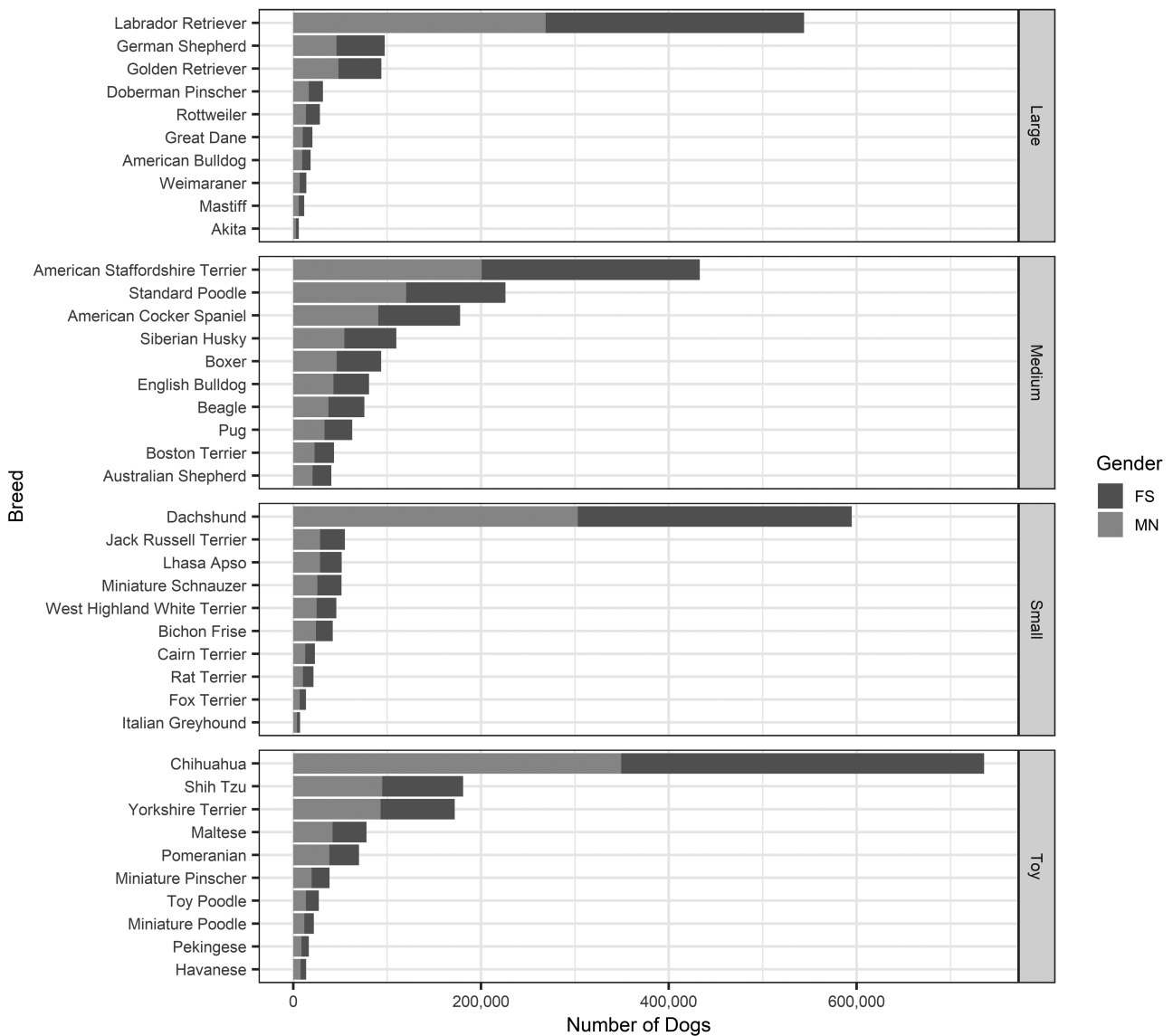
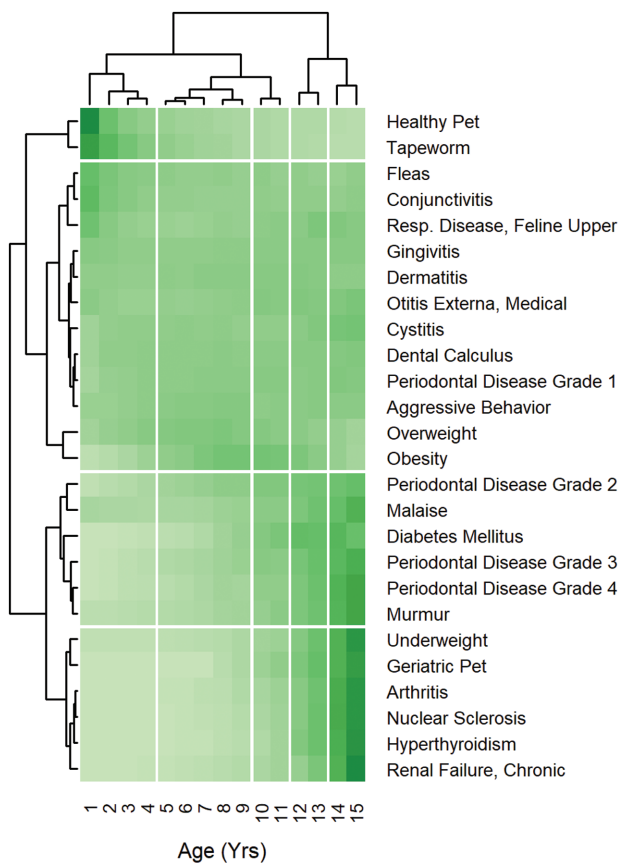


Figure 1. Diagram illustrating the frequency distribution of breed and gender (FS = female spayed; MN = male neutered) within each breed size category for the dog data set.

cat, exposures ranged between 35,829 pet years at 15 years old to 224,321 at 1 year old.

Examination of the sensitivity of the clustering to values of “diagnosis threshold” and “age deep split” showed the age groups remained in chronological order for every combination of values examined, even after the optimal re-ordering was applied. The age clusters found can therefore be described in terms of the ages at which splits occurred between neighboring clusters. [Supplementary Table 1](#) shows how many age clusters were created by each of the 21 different combinations of “diagnosis threshold” and “age deep split,” for cat and each dog size, and [Supplementary Table 2](#) shows where the splits between these clusters were located. For cats, all combinations of parameters resulted in splits at some combination of 4–5, 9–10, 11–12, and 13–14 years, with the most common result being for all of these splits to be present (i.e., 5 clusters). A pair of values for “diagnosis threshold” and “age deep split” which gave these splits was therefore chosen to be used in the final analysis, and these were 0.05 and 0.5, respectively. The resulting diagram produced by the bi-clustering and adaptive branch-pruning algorithms is shown in [Figure 2](#).

For all canine size categories, the most common result was 4 clusters. For both toy and small dog, the 3 most common splits were at 6–7, 11–12, and 13–14 years old, with rarer splits appearing at 7–8 and 9–10 years old. For medium dog, the 3 most common splits



**Figure 2.** Diagram illustrating results produced by the bi-clustering and adaptive branch-pruning algorithms for cat (using the final values of “diagnosis threshold” and “age deep split”). Cells are colored to show the (scaled) count of diagnoses per pet-year with darker shading indicating higher values. Clusters identified by the adaptive branch-pruning algorithm are separated from neighboring clusters by white lines.

were at 5–6, 9–10, and 13–14 years old, with rarer splits appearing at 6–7 years old and 11–12 years old. For large dog, the 3 most common splits were at 5–6, 9–10, and 11–12 years old, with rarer splits appearing at 7–8 and 12–13 years old. No one combination of “diagnosis threshold” and “age deep split” replicated this clustering pattern for all breed sizes, so final values of 0.06 and 0.5, respectively, were chosen for toy, medium and large size categories, and 0.05 and 0.5, respectively for the small category. [Figure 3](#) shows the resulting diagram for each breed size.

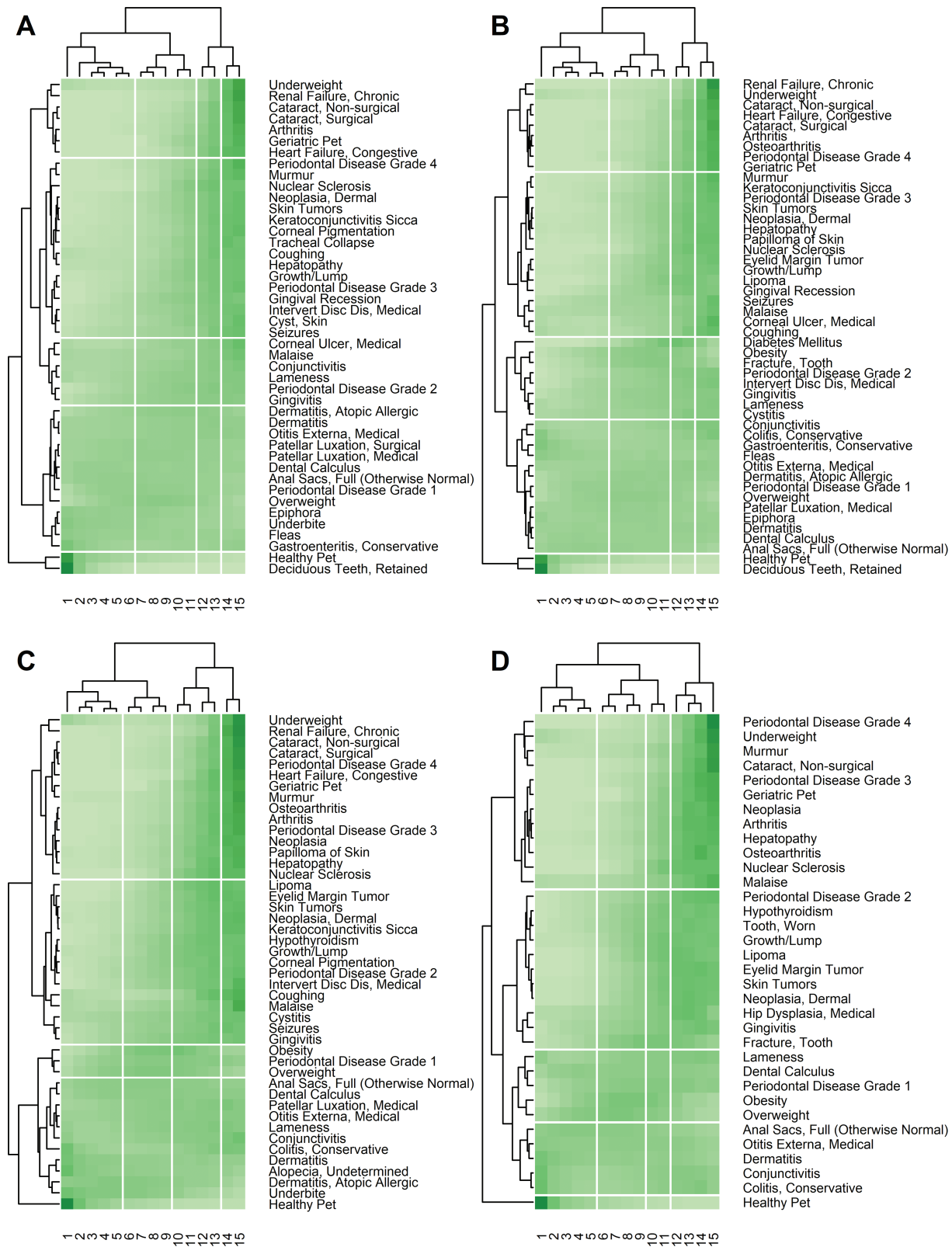
The age groups in [Figures 2](#) and [3](#) were optimally re-ordered into chronological order by the bi-clustering algorithm. The clusters identified by the adaptive branch pruning algorithm for age may therefore be interpreted as different life stages, and this interpretation is shown in [Tables 1](#) and [2](#) and [Supplementary Table 4](#). As an illustration of how the clustering algorithm finds similar ailment profiles, [Supplementary Figure 1](#) shows the profiles for each cluster for cat. [Supplementary Table 5](#) compares how ailments are clustered across species and breed size with an emphasis on the cluster interpretation, which can be roughly ordered from “young” ailments to “old” ailments.

Existing research shows that there are groups of diseases in humans that increase exponentially with age (24). Most of the disease profiles in this analysis belonging to clusters interpreted as being “Diagnoses of senior life stage onwards” or “Diagnoses of senior and (especially) super senior life stage” are fitted well by a 2-parameter exponential growth model—out of the 38 ailments thus classified, 20 have an  $R^2$  of at least 90% in all species/breed size where they feature and a further 8 have an  $R^2$  of at least 80%. Diagnoses of senior and/or super-senior life stages poorly suited to this model (yielding  $R^2$  values of less than 70%) were “Tooth, Fracture,” “Lipoma,” and “Nuclear Sclerosis.” Perhaps unsurprisingly, the diagnoses found to be of “senior and (especially) super senior life stage” tended to have a larger maximum growth rate than those of “senior life stage onwards,” with median fitted values of 0.29–0.36 years<sup>-1</sup> for the former and 0.15–0.20 years<sup>-1</sup> for the latter. Across the 6 “senior” and “super-senior” diagnoses shared between the species, the maximum growth rate appeared to be slightly higher in the dog than a cat. However, there was no marked difference in maximum growth rates between breed sizes of dogs—the maximum growth rate estimate found from the exponential model was not significantly associated with dog breed size, for either of the “senior and (especially) super senior life stage” or “senior life stage onwards” diagnoses, in a model with terms breed size class and diagnosis ( $p$ -values = .14 and .33, respectively), although a decreasing directional trend with increasing breed size was evident in the former.

A potential criticism of clustering-type approaches is that they can force splits in continuous data, even where these are artifacts of the method and not truly justified. It can therefore be useful to compare results with PCA, which is an unguided approach and should therefore be less subject to producing such artificial points of change. Trajectories for ages 1–15 on the first 3 principal components are shown in [Supplementary Figure 2](#) (cat) and [Supplementary Figure 3](#) (dog).

## Discussion

Aging may be considered a consequence of a gradual, progressive deterioration in physiological function/dysfunction, underpinned by cellular processes described as the hallmarks of aging (25). However, the different timing and type of specific age-related pathologies experienced by individuals are likely to be related to both the inherent



**Figure 3.** Diagram illustrating results produced by the bi-clustering and adaptive branch-pruning algorithms for 4 breed-sizes of dog (using the final values of “diagnosis threshold” and “age deep split”), with **A-D** representing toy, small, medium, and large, respectively. Cells are colored to show (scaled) count of diagnoses per pet-year with darker shading indicating higher values. Clusters identified by the adaptive branch-pruning algorithm are separated from neighboring clusters by white lines.

**Table 1.** Age Group Clusters Identified by the Adaptive Branch Pruning Algorithm for Cat

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Youth (years)	Early midlife (years)	Late midlife (years)	Senior (years)	Super-Senior (years)
1–4	5–9	10–11	12–13	≥14

**Table 2.** Age Group Clusters Identified by the Adaptive Branch Pruning Algorithm for Each Breed Size Category of Dog

Breed Size	Cluster 1	Cluster 2	Cluster 3	Cluster 4
	Youth (years)	Midlife (years)	Senior (years)	Super-Senior
Toy	1–6	7–11	12–13	≥14
Small	1–6	7–11	12–13	≥14
Medium	1–5	6–9	10–13	≥14
Large	1–5	6–9	10–11	≥12

functional capacity and the environmental stresses experienced. We considered that differential disease susceptibility with age may represent transition points in physiological functions that might expose differential aging processes and that these could be presented as discrete life stages. If such transitions were observed, we believed it may make possible more targeted preventive care and treatment recommendations throughout life. To identify if such transitions exist we undertook a data-driven stratification of life stages by disease susceptibility using differences in the occurrence of age-related health conditions in species (cat and dog) and within a species with diverse morphologies (breed size in dog), provided by an extensive database of companion animal health records.

While electronic clinical health records have been used to determine the incidence of age-, size-, and breed-related causes of death in dogs (26,27), this analysis is the first we are aware of to utilize diagnosis data from a large canine and feline population data set to establish adult life stages in these species. Although previous life stage definitions included descriptions that indicated that the pets in age groups were at higher risk for certain diseases, they did not include analyses from pet population data that supported the descriptions of disease risk for the stages—details of previously published life stages can be found in [Supplementary Table 3](#). The construction of the life stage, defined in terms of propensity to different disease conditions, as presented in [Tables 1 and 2](#), notably resulted in ages falling in chronological order. These life stages can then be described based on the occurrence of the different groups of diagnostic codes, as shown in [Supplementary Table 4](#).

[Supplementary Table 5](#) demonstrates that most ailments are placed similarly in all clusterings where they were included. However, some trends can be seen, for example, gingivitis and all grades of periodontal disease tend to feature in “younger” clusters in smaller dogs than in their larger counterparts, reflecting existing research showing smaller breeds to be more prone to oral health conditions (28).

Most of the disease profiles belonging to clusters interpreted as being “Diagnoses of senior life stage onwards” or “Diagnoses of senior and (especially) super senior life stage” fit well to a 2-parameter exponential growth model, showing that these diagnoses increase exponentially with age. This is consistent with research on human diseases (24) and suggests that the health span, on average, extends to the onset of the “Senior” life stage. For “exponential” disease clusters in humans,

the corresponding doubling rate has been seen to be numerically close to the mortality rate doubling time from the Gompertz mortality law (24). The exponential model used here is equivalent in form to the Gompertz equation so that the maximum growth rate in our model would be equivalent to the “actuarial aging rate” of the Gompertz equation if applied to mortality data and inversely proportional to the doubling time (29). The maximum growth rate estimates found for “senior and (especially) super senior life stage” and “senior life stage onwards” diagnoses are indeed consistent with the approximate range of actuarial aging rates found for the oldest group of dogs in ref. (30) suggesting that this relationship between doubling time for exponential diseases and doubling time for mortality is also true of dogs.

It has been noted that life stages can be detected visually in PCAs of various markers in humans (31,32). The PCA trajectories showed marked changes in direction and/or gradient at several points, which correspond with the life stages identified using bi-clustering, although changes in PC3 for toy and small dogs around age 4–5 years do not lead to a new life stage according to the bi-clustering. This may be because PC3 is a comparatively minor component, accounting for 3%–4% of the total variance in those breed sizes. The PCA showed transition points despite being an unguided analysis, and these were similar to the life stages we see in the bi-clustering analysis. This corroborates that these identify transition points in aging.

There are some considerations in this study that are worth noting. First, those related to the animal population. The medical records used in the analysis were from one large primary care veterinary hospital network, and it is unknown how representative this pet population is of pets in general—in particular, it can be assumed that seriously unwell patients may transfer to speciality clinics and are potentially under-represented in primary care records. Additionally, the analysis only considers neutered animals, which represents the majority of the pet population in the United States. However, it may not reflect the impact of reproductive capacity and may exaggerate or obscure any impact of neutering on age-related disease occurrence. Also, some of the size categories for the dog were dominated by a single breed (eg, Labrador retriever for the large category and chihuahua for the toy category), and the results of the analysis are strongly skewed toward these breeds. Similarly, this should not markedly impact the utility of the results as they reflect the popularity of these breeds in the pet dog population of the United States (33) and, therefore, relevance to vets and owners. Second, diagnosis profiles were constructed based on diagnostic codes per pet-year at-risk. As a diagnosis is not necessarily a first occurrence for an individual, this quantity has a different interpretation for a chronic condition than it does for a condition that resolves quickly. However, patterns should still be detectable and relevant, regardless of the exact interpretation of the quantity used. Other approaches to creating profiles have been reported (34) where profiles according to age at first onset were used to cluster human diseases, although these were created from questionnaire data rather than medical records. Furthermore, there are a variety of ways that hierarchical clustering can be enacted depending on the combination of distance metric and linkage used. Although the settings here were picked to be appropriate to the application, they do not form a uniquely appropriate choice, and other combinations of metric and linkage could produce different results. Similarly, the dynamic tree-cut method chosen to detect the final clusters is not the only such algorithm. The cut-height method is the most notable alternative, although others creating clusters in human disease (34) used the PAM algorithm (35), which is, like the method used here, a more objective way of deciding on the final clusters.

For cats, this study identified up to 2 additional life stages than those previously published (Supplementary Table 3), resulting in “Senior” and “Super Senior” stages starting, for the most part, at a later age than previously published guidelines. Similarly, compared to previously published charts defining canine life stages by breed size (36,37), even with the differing weights used to define breed size, the results from this study define “Senior” and “Super Senior” at noticeably later ages than previously defined for the equivalent stages (“Senior” and “Geriatric”). We also noted that the smaller-sized dogs classified as senior 2 years later than larger dogs, consistent with them simply living longer (15,16) and also having an extended health span, potentially due to mitochondrial bioenergetics (11). Differences to previous approaches may reflect that we only considered one aspect of aging, that of the development of disease conditions, and not including characteristics described in other life stage guidelines (behavioral/psychologic and physical characteristics, such as lean tissue and activity levels). Despite these differences and the complexity and the many unknowns regarding aging, the findings of this analysis can still be utilized by clinicians to support conversations on the importance of preventive care services and better target recommendations for diagnostic tests and other care recommendations.

We also expect the data will provide insights into stratified care pathways. For example, the data indicate an increased incidence of retained teeth and earlier onset of periodontal disease in smaller dog groups, and also increased incidence of later heart conditions and renal health issues. This association was previously reported as risk issues for these later-onset diseases in dogs, regardless of breed size (38,39). Such insights could lead to relatively easy and early interventions, such as teeth cleaning, focused on those breeds that might have far more valuable health span-related benefits in later life. Additionally, we believe the information may support healthy aging research. For example, one application would be in large-scale studies where the differential onset of age-related disease in specific dog size groups (rather than an “average” dog) could be used as an outcome to identify more personalized care pathways to extend health span.

In summary, we postulated that the type and timing of disease diagnoses derived from an extensive electronic medical records database would enable data-driven stratification of life stages that would distinguish between different physiologies and potentially highlight those biological processes that underpin differential aging. We have developed a tool to support tailored discussions between clinicians and clients regarding care to reduce the risk of age-related diseases, derive insights for researchers, and also enhance understanding of the importance of physiology on the development and progression of age-related diseases.

## Supplementary Material

Supplementary data are available at *The Journals of Gerontology, Series A: Biological Sciences and Medical Sciences* online.

## Funding

This work was supported by Mars, Inc. in the form of salaries for all authors and the funders had no additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Conflict of Interest

C.S., C.O.F., and D.A. are employees of Waltham Petcare Science Institute and E.K.S. is an employee of BANFIELD Pet Hospitals. Both companies both are owned by Mars, Inc.

## Acknowledgments

We are grateful to the more than 1 000 Banfield clinicians who diligently and consistently recorded their observations in a medical record over the 6+ year study period. Without that collective effort this work would not have been possible.

## Author Contributions

All authors participated in review of the literature, authorship of the manuscript, and content editing. In addition, C.S. carried out the project design and data analysis and E.K.S. provided veterinary expertise and information on practices at Banfield Veterinary Hospitals.

## Data Availability

Although the data used in the study is proprietary, data access will be accommodated where possible upon direct request to the corresponding author (carina.salt@effem.com).

## References

- Georgountzou A, Papadopoulos NG. Postnatal innate immune development: from birth to adulthood. *Front Immunol*. 2017;8:957. doi:10.3389/fimmu.2017.00957
- Wampach L, Heintz-Buschart A, Fritz JV, et al. Birth mode is associated with earliest strain-conferred gut microbiome functions and immunostimulatory potential. *Nat Commun*. 2018;9(1):5091. doi:10.1038/s41467-018-07631-x
- Zlotnick M, Corrigan V, Griffin E, Alayon M, Hungerford L. Incidence of health and behavior problems in service dog candidates neutered at various ages. *Front Vet Sci*. 2019;6:334. doi:10.3389/fvets.2019.00334
- Hoffman JM, Creevy KE, Promislow DEL. Reproductive capability is associated with lifespan and cause of death in companion dogs. *PLoS One*. 2013;8(4):e61082. doi:10.1371/journal.pone.0061082
- Hart BL, Hart LA, Thigpen AP, Willits NH. Assisting decision-making on age of neutering for mixed breed dogs of five weight categories: associated joint disorders and cancers. *Front Vet Sci*. 2020;7:472. doi:10.3389/fvets.2020.00472
- Smith GK, Paster ER, Powers MY, et al. Lifelong diet restriction and radiographic evidence of osteoarthritis of the hip joint in dogs. *J Am Vet Med Assoc*. 2006;229(5):690–693. doi:10.2460/javma.229.5.690
- Lopez M, Schachner E. Diagnosis, prevention, and management of canine hip dysplasia: a review. *Vet Med Res Reports* 2015;6:181–192. doi:10.2147/vmrr.s53266
- Kirkwood TBL. Why and how are we living longer? *Exp Physiol*. 2017;102(9):1067–1074. doi:10.1113/EP086205
- Bateson M, Poirier C. Can biomarkers of biological age be used to assess cumulative lifetime experience? *Anim Welf*. 2019;28(1):41–56. doi:10.7120/09627286.28.1.041
- Jimenez AG, Winward J, Beattie U, Cipolli W. Cellular metabolism and oxidative stress as a possible determinant for longevity in small breed and large breed dogs. *PLoS One*. 2018;13(4):e0195832. doi:10.1371/journal.pone.0195832
- Nicholatos JW, Robinette TM, Tata SVP, et al. Cellular energetics and mitochondrial uncoupling in canine aging. *GeroScience* 2019;41(2):229–242. doi:10.1007/s11357-019-00062-6
- Priester WA, Mulvihill JJ. Canine hip dysplasia: relative risk by sex, size, and breed, and comparative aspects. *J Am Vet Med Assoc*. 1972;160(5):735–739.
- Evans KM, Adams VJ. Mortality and morbidity due to gastric dilatation-volvulus syndrome in pedigree dogs in the UK. *J Small Anim Pract*. 2010;51(7):376–381. doi:10.1111/j.1748-5827.2010.00949.x
- Wallis C, Holcombe LJ. A review of the frequency and impact of periodontal disease in dogs. *J Small Anim Pract*. 2020;61(9):529–540. doi:10.1111/jsap.13218
- Deeb BJ, Wolf NS. Studying longevity and morbidity in giant and small breeds of dogs. *Vet Med*. 1996;91:702–713.

16. Michell AR. Longevity of British breeds of dog and its relationships with sex, size, cardiovascular variables and disease. *Vet Rec.* 1999;145(22):625–629. doi:10.1136/vr.145.22.625
17. Hahsler M, Hornik K, Buchta C. Getting things in order: an introduction to the R Package seriation. *J Stat Softw.* 2008;25(3):1–34. doi:10.18637/jss.v025.i03
18. Ward JH. Hierarchical Grouping to Optimize an Objective Function. *J Am Stat Assoc.* 1963;58(301):236–244. doi:10.1080/01621459.1963.10500845
19. Manly BFJ. *Multivariate Statistical Methods: A Primer.* Vol 6; 1986. <http://books.google.be/books?id=bLSEeMjvD9oC>
20. Bar-Joseph Z, Gifford DK, Jaakkola TS. Fast optimal leaf ordering for hierarchical clustering. *Bioinformatics.* 2001;17 Suppl 1:S22–S29. doi:10.1093/bioinformatics/17.suppl\_1.s22
21. Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* 2008;24(5):719–720. doi:10.1093/bioinformatics/btm563
22. Hahsler M, Buchta C, Hornik K, et al. seriation: Infrastructure for Ordering Objects Using Seriation. 2017. Accessed September 24, 2018. <https://cran.r-project.org/package=seriation>
23. Langfelder P, Zhang B, and with contributions from Steve Horvath. dynamicTreeCut: Methods for Detection of Clusters in Hierarchical Clustering Dendrograms. 2016. Accessed September 24, 2018. <https://cran.r-project.org/package=dynamicTreeCut>
24. Zenin A, Tsepilov Y, Sharapov S, et al. Identification of 12 genetic loci associated with human healthspan. *Commun Biol.* 2019;2:41. doi:10.1038/s42003-019-0290-0
25. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. *Cell.* 2013;153(6):1194–1217. doi:10.1016/j.cell.2013.05.039
26. O'Neill DG, Church DB, McGreevy PD, Thomson PC, Brodbelt DC. Longevity and mortality of owned dogs in England. *Vet J.* 2013;198(3):638–643. doi:10.1016/j.tvjl.2013.09.020
27. Fleming JM, Creevy KE, Promislow DEL. Mortality in North American dogs from 1984 to 2004: an investigation into age-, size-, and breed-related causes of death. *J Vet Intern Med.* 2011;25(2):187–198. doi:10.1111/j.1939-1676.2011.0695.x
28. Wallis C, Saito EK, Salt C, Holcombe LJ, Desforges NG. Association of periodontal disease with breed size, breed, weight, and age in purebred client-owned dogs in the United States. *Vet J.* 2021;275:105717. doi:10.1016/j.tvjl.2021.105717
29. Kirkwood TBL. Deciphering death: a commentary on Gompertz (1825) “On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philos Trans R Soc B Biol Sci.* 2015;370(1666):20140379. doi:10.1098/rstb.2014.0379
30. Kraus C, Pavard S, Promislow DEL. The size–life span trade-off decomposed: why large dogs die young. *Am Nat.* 2013;181(4):492–505. doi:10.1086/669665
31. Pyrkov TV, Getmantsev E, Zhurov B, et al. Quantitative characterization of biological age and frailty based on locomotor activity records. *Aging (Albany NY)* 2018;10(10):2973–2990. doi:10.18632/aging.101603
32. Pyrkov TV, Avchaciov K, Tarkhov AE, Menshikov LI, Gudkov AV, Fedichev PO. Longitudinal analysis of blood markers reveals progressive loss of resilience and predicts human lifespan limit. *Nat Commun.* 2021;12(1):2765. doi:10.1038/s41467-021-23014-1
33. American Kennel Club. Most popular dog breeds – full list. Accessed December 12, 2017. <http://www.akc.org/content/news/articles/most-popular-dog-breeds-full-ranking-list/>
34. Dönertaş HM, Fabian DK, Fuentealba M, Partridge L, Thornton JM. Common genetic associations between age-related diseases. *Nat Aging* 2021;1(4):400–412. doi:10.1038/s43587-021-00051-5
35. Kaufman L, Rousseeuw PJ. Partitioning around medoids (program PAM). Finding groups in data: an introduction to cluster analysis 1990;344:68–125. doi:10.2307/2532178
36. Fortney WD. Implementing a successful senior/geriatric health care program for veterinarians, veterinary technicians, and office managers. *Vet Clin North Am Small Anim Pract.* 2012;42(4):823–834. doi:10.1016/j.cvs.2012.04.011
37. Bellows J, Colitz CMH, Daristotle L, et al. Defining healthy aging in older dogs and differentiating healthy aging from disease. *J Am Vet Med Assoc.* 2015;246(1):77–89. doi:10.2460/javma.246.1.77
38. Glickman LT, Glickman NW, Moore GE, Lund EM, Lantz GC, Pressler BM. Association between chronic azotemic kidney disease and the severity of periodontal disease in dogs. *Prev Vet Med.* 2011;99(2-4):193–200. doi:10.1016/j.prevetmed.2011.01.011
39. Glickman LT, Glickman NW, Moore GE, Goldstein GS, Lewis HB. Evaluation of the risk of endocarditis and other cardiovascular events on the basis of the severity of periodontal disease in dogs. *J Am Vet Med Assoc.* 2009;234(4):486–494. doi:10.2460/javma.234.4.486