Big data from small animals: integrating multi-level environmental data in the Dog Aging Project

D. Xue (1, 2), D. Collins (2, 3), M. Kauffman (2), M. Dunbar (2), K. Crowder (2, 3), S.M. Schwartz (4), Dog Aging Project Consortium & A. Ruple* (5)

(1) Institute for Public Health Genetics, University of Washington, Seattle, Washington, United States of America

(2) Center for Studies in Demography and Ecology, University of Washington, Seattle, Washington, United States of America

(3) Department of Sociology, University of Washington, Seattle, Washington, United States of America

(4) Department of Epidemiology, University of Washington, Seattle, Washington, United States of America

(5) Department of Population Health Sciences, Virginia-Maryland College of Veterinary Medicine, Virginia Tech, Blacksburg, Virginia, United States of America

*Corresponding author: aruple@vt.edu

Summary

Environmental exposures can have large impacts on health outcomes. While many resources have been dedicated to understanding how humans are influenced by the environment, few efforts have been made to study the role of built and natural environment features on animal health. The Dog Aging Project (DAP) is a longitudinal community science study of aging in companion dogs. Using a combination of owner-reported surveys and secondary sources linked through geocoded coordinates, DAP has captured home, yard, and neighbourhood variables for over 40,000 dogs. The DAP environmental dataset spans four domains: 1) Physical and built environment; 2) Chemical environment and exposures; 3) Diet and exercise; and

4) Social environment and interactions. Together with biometric data, measures of cognitive function and behaviour, and medical records, DAP is attempting to use a big data approach to transform our understanding of how the world around us affects the health of companion dogs. Here, we describe the data infrastructure we have developed to integrate and analyse multi-level environmental data that can be used to better understand canine comorbidity and aging. DAP datasets are available upon request at https://dogagingproject.org/open_data_access.

Keywords

Big data – Companion dogs – Environment – Geocoding – One Health – Open science.

Introduction

Advances in technology, social and economic upheaval, and severe effects of climate change have altered the way people interact with each other and the outside world, and will continue to shape social, built, and chemical environments, resulting in immediate and long-term health consequences for humans and animals alike. Exposure to ambient air pollution is associated with higher risk for stroke, heart disease, and respiratory infections [1, 2]. Built environment features including green space and neighbourhood walkability have been shown to be positively related to cognitive function [3, 4]. Neighbourhood density and socio-economic status are associated with multiple facets of health and lifespan [5]. Companion dogs experience many of the same environmental exposures as their owners and develop many comparable health outcomes, yet little is known about how these exposures influence disease and lifespan in the animals. Previous studies that have attempted to capture the effects of environment on disease in companion animals have been relatively small in sample size or restricted to specific geographic locations [6, 7].

The computational advances of the big data era enhance our ability to compile, store, and analyse environmental exposures at the population level. Using owner-reported surveys on household conditions and

publicly available national environmental data sources, our Dog Aging Project (DAP) team has created an extensive household and neighbourhood environmental profile for over 38,000 dogs currently enrolled in the DAP. We describe herein the infrastructure we have developed for compiling data on a comprehensive set of longitudinal environmental factors that can be linked to canine comorbidity and aging.

The Dog Aging Project

The DAP is a large national longitudinal study of aging in companion dogs [8]. The overarching goals of DAP are to understand the biological and environmental factors that shape aging and to discover how we can maximise the healthy lifespan of dogs. In line with a One Health approach [9], a greater understanding of the relationship between the environment and dog aging can also have implications for human health. Companion dogs are an optimal species to study to better understand health and aging in humans: in addition to living in the same environment as their owners, dogs have a robust healthcare system, and they develop many of the same diseases as humans, including cancer and dementia [10, 11]. We can also learn more quickly about how the environment affects aging in dogs because they have shorter lifespans compared to humans. DAP is collecting survey data and electronic veterinary medical records on all dogs included in the project and biospecimen collection occurs for 10,000 dogs who are part of sampled cohorts. All these additional data sources can all be linked to the environmental profile based on the owner-reported residential addresses. Further details on DAP have been previously described [8, 12].

The DAP is built on a community science model. As of this publication, there are 40,689 dogs enrolled in the DAP, and this number is constantly growing. Enrolment in DAP began in 2019 and all dogs are followed longitudinally. Thus, at the time of this publication the dogs enrolled at the time the project was initiated have four years of data associated with them. The enrolled dogs, often referred to as the DAP Pack, represent dogs of almost all breeds registered by the American

Kennel Club as well as mixed breeds, ages, and health conditions located in all 50 United States of America states. Furthermore, DAP data are available to investigators who are unaffiliated with the original study upon request. Environmental variables were carefully selected to fit this Open Science model, with special attention given to ensuring all the environmental variables are publicly available and do not require further permissions to be shared. All of the data collected as part of the DAP data repository are available in a Google cloud-based platform developed in collaboration with the Terra team at the Broad Institute of MIT and Harvard. The cloud-based system allows for seamless harmonisation, analysis, and distribution of the hundreds of terabytes of DAP data. These data and further details including codebooks are available at https://dogagingproject.org/open_data_access.

Multilevel environmental data in Dog Aging Project

For each dog in the DAP Pack, we captured the environment at multiple levels including the home environment, the yard or outside area where the dog is allowed to roam, the neighbourhood, and extra local areas where the dog regularly spends time. Across each of these levels, we focused on four environmental domains: 1) physical and built environment; 2) chemical environment; 3) diet and exercise; and 4) social environment.

We developed an owner-reported Health and Life Experience Survey (HLES) and geocoding system to link the owner-reported addresses to geographically-based, publicly available secondary data sources in order to collect information across environmental domains (Figure 1). Due to the personally identifiable information captured in the survey and linked secondary sources, all dogs are assigned anonymised unique IDs that are universal across all datasets.

Owner-reported survey

The HLES contains eight questionnaires: Dog demographics, Physical activity, Environment, Diet, Medications and preventatives, Health status, and Owner demographics. Survey questions in the HLES were developed based on previous studies of dog health, including the

Golden Retriever Lifetime Study [13] and Darwin's Dogs (DarwinsArk.org), as well as human longitudinal studies [14, 15, 16]. During the design phase of the questionnaire, we identified small samples of dog owners within DAP to test and refine the survey questions using test-retest strategies and conducting qualitative interviews to assess the interpretability and clarity of each question. The final HLES is delivered via Research Electronic Data Capture – REDCap [17, 18] and made available for each participant through a password-protected portal and is administered at the time of enrolment and then annually thereafter.

Questions regarding the four environmental domains are dispersed across the eight HLES categories. The survey responses primarily describe the home and yard environment, with variables including 'Number of people in the household' or 'Surface type of outdoor areas'. Examples of multi-level data across environmental domains are described in Table I.

Geographic information system- and Census-based secondary data

In addition to the eight HLES questionnaires, all owners are asked to provide a primary residential address and secondary addresses of extra local areas where their dogs spend time, if applicable. We then use these addresses as a key to unlock a rich array of secondary data using a sophisticated geocoding procedure designed to ensure a very high proportion of matches between owner-reported residential addresses and geocoded addresses.

We used a three-tiered geocoding strategy. After testing several geocoding software packages on a subset of addresses, we determined that Environmental Systems Research Institute's ArcGIS Business Analyst geocoder [19] was most successful and subsequently designated it as our primary geocoding method. Addresses that are not able to be matched using the primary method (3.1%) are then processed through the Smarty geocoder [20]. Finally, any addresses that remain unmatched are manually reviewed by our team using Google Maps and other readily available web sources such as Zillow

(https://www.zillow.com) or Redfin (https://www.redfin.com). This three-tier method results in successful matches for 99.5% of all submitted primary addresses and 95.8% of secondary addresses. The addresses that we are unable to match are often due to respondent typographical errors, entry of a mailing address rather than residential address (i.e. a PO Box), or entry of a newly built construction.

Addresses that are precisely geocoded are then assigned spatial coordinates and census block-level Federal Information Processing Standards (FIPS) codes, which can be used to link to various publicly available data sources including the American Community Survey (ACS) [21]; the Center for Air, Climate, and Energy Solutions (CACES) [22]; the National Oceanic and Atmospheric Administration (NOAA) Climate Divisional Database [23], and WalkScore (https://www.walkscore.com). From these national databases, we obtained environmentally relevant variables and computed composite indices when appropriate.

Within the secondary databases, we focused on four domains: sociodemographic and economic factors, air pollutants, temperature and precipitation metrics, and walkability indicators. We selected one representative variable from each domain as examples in Figure 2.

Sociodemographic and economic features are collected at the tract-level from ACS. Variables taken directly from the source include descriptive tract information (i.e. population estimate, area), demographics, and economic indicators (i.e. median income, Gini index [24]). We then derived two composite variables based on the ACS variables: the disadvantage index and the stability index. Based on prior publications on neighbourhood effects and the relatedness of certain neighbourhood features, we computed the disadvantage index by averaging the z-scores of:

a) percentage of population below 125% of the poverty line;

b) percentage of working-age unemployed or not in the labour force;

c) percentage of children living in female-led households with no husband present;

d) percentage of population >25 years with less than a Bachelor's degree;

e) percentage of households earning under $100,000 in the last 12 months [25, 26, 27].

Similarly, based on collective efficacy literature, we calculated the stability index by averaging the z-scores of the following variables:

a) percentage of population that was in the same house one year ago;

b) percentage of owner-occupied housing units;

c) percentage of population born in the United States of America [28].

Air pollution data are obtained directly from publicly available estimates developed by CACES using v1 empirical models as described [29]. The most recent data include estimates of outdoor concentrations for six pollutants: ozone, carbon monoxide, sulfur dioxide, nitrogen dioxide, particulate matter 10 micrometres or smaller (PM10), and fine particulate matter 2.5 micrometres or smaller (PM2.5), each linked to geocoded addresses at the tract level.

Temperature and precipitation measures from NOAA are provided at the county level for two time periods. First, annual summaries contain temperature and precipitation averages for each month of every year since collection began. DAP environmental data include annual summaries starting from 2019. Second, 'normals' summaries are provided, which are long-term averages over 30-year periods with ten-year increments. The earliest set of 'normals' data used in DAP is from 1981–2010.

Finally, measures of neighbourhood walkability are derived from both WalkScore and ACS. WalkScore generates a walkability index based

on walking routes to nearby amenities of latitude/longitude spatial coordinates. Using ACS, we included residential density variables at the census tract level to compute a walkability score based on prior findings that residential density is an appropriate proxy for objectively-measured neighbourhood walkability [30].

Longitudinal structure of environmental data

In addition to using unique, universal identifiers for each dog that allow all variables to be merged across surveys and secondary data sources, we have implemented several workflows to accommodate DAP's longitudinal structure. We capture residential moves through an annual check-in survey. Owners are also able to update their residential address on their online DAP user profile at any time. Each address change is time-stamped, and secondary geocoded addresses are also updated annually for all geocoded addresses.

Data protection

Privacy and confidentiality of sensitive, personally identifiable information are a top priority in the DAP. Addresses and geocoded coordinates are stored with extra security, and all survey results and linked secondary data are deidentified before they are released for analysis. The original addresses used to link to secondary data sources will never be shared outside of the environmental data management team. External investigators must complete a data-use agreement before they can access data that have been prepared for public release.

Discussion

Big data will play a pivotal role in the future of both human and animal health. Our team has developed infrastructure to compile data on a comprehensive set of environmental risk factors that will allow us to gain new insights in aging and age-related disease in companion dogs. Here, we have laid out a blueprint for using individual addresses to uncover multi-level environmental data for each dog that spans social, physical, and built environment domains. Rather than relying only on owner-reported individual-level information, environmental big data

offer an opportunity to discover upstream, contextual determinants that can be targeted for population-level improvement of animal health.

The breadth of environmental factors covered through the owner-reported surveys and secondary sources can help us measure the associations of environmental exposures. We can investigate direct associations between environmental variables and disease, such as measuring the association between exposure to particulate matter and cardiovascular disease. In conjunction with other survey responses and biospecimens collected as a part of DAP, we can use the environment data to conduct multi-level studies to understand how variation at the macrolevel modifies associations between individual level factors. For example, we know that cognitive dysfunction is associated with chronological age [31], but adding on the environmental data allows us to investigate whether the level of association between age and cognitive dysfunction differs depending on the number of other animals or humans in the home environment or whether neighbourhood greenspace density modifies the association between age and cognition. Furthermore, big data projects in companion animals like DAP can shift the paradigm in our understanding of the effects of gene-environment interactions throughout the life-course. These effects have been difficult to study in humans, as they require sample sizes of tens of thousands throughout a lifespan. Moreover, while individuals may be unwilling to provide their own genetic, epigenetic, or metabolomic data due to privacy concerns, they may be less concerned about sharing this information for their dogs.

The initial curation of environmental variables was limited to information that is publicly accessible and informative for dogs nationwide, but our data infrastructure is flexible for use in more geographically specific ancillary studies and designed to seamlessly handle additional variables added in the future. Due to the use of universal individual IDs that are consistent across datasets and timestamped geocoding, new variables can be retroactively and prospectively added to our data structure as they become available. As an example, we are currently working on processing and incorporating normalized difference vegetation index (NDVI) values, indicators of

density of vegetation or greenspace in a given area. Global NDVI values are available through the National Centers for Environmental Information and can be added retroactively to the DAP dataset in future releases [32]. DAP's Open Science model welcomes requests for ancillary studies, which provide opportunities for external investigators to add environmental variables that are only available in limited geographic areas to our existing scaffold. Some examples of geographically limited environmental phenomena that could impact dog health include fracking, wildfires, or tick density. These data were not included in DAP's original design, either because they are not widespread across the country or geocoded data on these measures are not publicly available, but these exposures can have ramifications on dog health that are worth investigating in ancillary studies [33, 34, 35, 36].

DAP has some limitations. The DAP environmental data are expansive, covering both social and physical environmental features, but our representation and socioeconomic diversity of owners is limited (Figure 2). For example, the median income category of our participants is $100,000 to $119,999, which is greater than the national average of $67,521 in 2020 [37]. Our study population recruitment is volunteer-based, and as a result, the dogs enrolled in our study are restricted to those who have owners with internet access and sufficient time to complete what can be a multi-hour assessment. Another potential selection bias is that owners who believe their dogs are unhealthy or living with serious medical conditions may be less likely to enrol them into a longitudinal study of aging. Lack of representation limits the generalisability of research findings. Moving forward, DAP will look toward other big data initiatives such as All of Us [38] that have implemented strategies to diversify participation.

The future of veterinary care, and health care as a whole, will increasingly rely on big data methods. Beyond electronic veterinary medicine records and biometric data, DAP goes one step further, linking environmental big data to each participant record. While precision medicine is often thought of as treatments tailored to individual genomic patterns, and environmental or contextual

determinants belonging in the realm of public health, the integration of big data that span genomic, epigenomic, and geospatial variation allows for precision public health – an overlap between two seemingly distinct fields. Owners go to great lengths to ensure the health of their dogs, but they might not be able to detect or control how environmental determinants like air pollution or precipitation affect their dog's health. By simply providing an address, they enable DAP investigators to unearth potentially 'unseen' patterns, which can inform population-level interventions that can enhance the effects of individual-level care.

Acknowledgements

The authors thank Dog Aging Project participants, their dogs, and community veterinarians for their important contributions. The Dog Aging Project is supported by U19 grant AG057377 from the National Institute on Aging, a part of the National Institutes of Health, and by private donations. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

[1] Gold D.R. & Samet J.M. (2013). – Air pollution, climate, and heart disease. *Circulation*, **128** (21), e411–e414. https://doi.org/10.1161/CIRCULATIONAHA.113.003988

[2] Kelly F.J. & Fussell J.C. (2011). – Air pollution and airway disease. *Clin. Exp. Allergy*, **41** (8), 1059–1071. https://doi.org/10.1111/j.1365-2222.2011.03776.x

[3] Besser L.M., Chang L.-C., Hirsch J.A., Rodriguez D.A., Renne J., Rapp S.R., Fitzpatrick A.L., Heckbert S.R., Kaufman J.D. & Hughes T.M. (2021). – Longitudinal associations between the neighborhood built environment and cognition in US older adults: the multi-ethnic study of Atherosclerosis. *Int. J. Environ. Res. Public Health*, **18** (15), 7973. https://doi.org/10.3390/ijerph18157973

[4] Yuchi W., Sbihi H., Davies H., Tamburic L. & Brauer M. (2020). – Road proximity, air pollution, noise, green space and neurologic disease incidence: a population-based cohort study. *Environ. Health*, **19** (1), 8. https://doi.org/10.1186/s12940-020-0565-4

[5] Link B.G. & Phelan J. (1995). – Social conditions as fundamental causes of disease. *J. Health Soc. Behav.*, Extra Issue, 80–94. https://doi.org/10.2307/2626958

[6] Calderón-Garcidueñas L., Mora-Tiscareño A. […] & Engle R.W. (2008). – Air pollution, cognitive deficits and brain abnormalities: a pilot study with children and dogs. *Brain Cogn.*, **68** (2), 117–127. https://doi.org/10.1016/j.bandc.2008.04.008

[7] Packer R.M.A., Davies A.M., Volk H.A., Puckett H.L., Hobbs S.L. & Fowkes R.C. (2019). – What can we learn from the hair of the dog? Complex effects of endogenous and exogenous stressors on canine hair cortisol. *PLoS ONE*, **14** (5), e0216000. https://doi.org/10.1371/journal.pone.0216000

[8] Creevy K.E., Akey J.M., Kaeberlein M., Promislow D.E.L. & The Dog Aging Project Consortium (2022). – An open science study of ageing in companion dogs. *Nature*, **602** (7895), 51–57. https://doi.org/10.1038/s41586-021-04282-9

[9] Zinsstag J., Schelling E., Waltner-Toews D. & Tanner M. (2011). – From "one medicine" to "one health" and systemic approaches to health and well-being. *Prev. Vet. Med.*, **101** (3–4), 148–156. https://doi.org/10.1016/j.prevetmed.2010.07.003

[10] Gray M., Meehan J., Martínez-Pérez C., Kay C., Turnbull A.K., Morrison L.R., Pang L.Y. & Argyle D. (2020). – Naturally-occurring canine mammary tumors as a translational model for human breast cancer. *Front. Oncol.*, **10**, 617. https://doi.org/10.3389/fonc.2020.00617

[11] Urfer S.R., Darvas M., Czeibert K., Sándor S., Promislow D.E.L., Creevy K.E., Kubinyi E. & Kaeberlein M. (2021). – Canine cognitive dysfunction (CCD) scores correlate with amyloid beta 42 levels in dog brain tissue. *GeroScience*, **43** (5), 2379–2386. https://doi.org/10.1007/s11357-021-00422-1

[12] Kaeberlein M., Creevy K.E. & Promislow D.E.L. (2016). – The dog aging project: translational geroscience in companion animals. *Mamm. Genome*, **27** (7–8), 279–288. https://doi.org/10.1007/s00335-016-9638-7

[13] Guy M.K., Page R.L., Jensen W.A., Olson P.N., Haworth J.D., Searfoss E.E. & Brown D.E. (2015). – The golden retriever lifetime study: establishing an observational cohort study with translational relevance for human health. *Philos. Trans. R. Soc. B Biol. Sci.*, **370** (1673), 20140230. https://doi.org/10.1098/rstb.2014.0230

[14] Bild D.E., Bluemke D.A. […] & Tracy R.P. (2002). – Multi-ethnic study of Atherosclerosis: objectives and design. *Am. J. Epidemiol.*, **156** (9), 871–881. https://doi.org/10.1093/aje/kwf113

[15] McGonagle K.A., Schoeni R.F., Sastry N. & Freedman V.A. (2012). – The panel study of income dynamics: overview, recent innovations, and potential for life course research. *Longit. Life Course Stud.*, **3** (2), 188. https://doi.org/10.14301/llcs.v3i2.188

[16] Andersson C., Johnson A.D., Benjamin E.J., Levy D. & Vasan R.S. (2019). – 70-year legacy of the Framingham heart study. *Nat. Rev. Cardiol.*, **16** (11), 687–698. https://doi.org/10.1038/s41569-019-0202-5

[17] Harris P.A., Taylor R., Thielke R., Payne J., Gonzalez N. & Conde J.G. (2009). – Research electronic data capture (REDCap): a metadata-driven methodology and workflow process for providing translational research informatics support. *J. Biomed. Inform.*, **42** (2), 377–381. https://doi.org/10.1016/j.jbi.2008.08.010

[18] Harris P.A., Taylor R., Minor B.L., Elliott V., Fernandez M., O'Neal L., McLeod L., Delacqua G., Delacqua F., Kirby J., Duda S.N. & REDCap Consortium (2019). – The REDCap Consortium: building an international community of software platform partners. *J. Biomed. Inform.*, **95**, 103208. https://doi.org/10.1016/j.jbi.2019.103208

[19] Environmental Systems Research Institute (ESRI) (2022). – ArcGIS Business Analyst. ESRI, Redlands, United States of America. Available at: https://www.esri.com/en-us/arcgis/products/arcgis-business-analyst/overview (accessed on 18 April 2022).

[20] Smarty (2022). – Address verification, validation, and autocomplete. Smarty, Hasbrouck Heights, United States of America. Available at: https://www.smarty.com (accessed on 18 April 2022).

[21] United States Census Bureau (2022). – American Community Survey (ACS). US Census Bureau, Washington, DC, United States of America. Available at: https://www.census.gov/programs-surveys/acs (accessed on 18 April 2022).

[22] Zimmerman N., Li H.Z. […] & Presto A.A. (2020). – Improving correlations between land use and air pollutant concentrations using wavelet analysis: insights from a low-cost sensor network. *Aerosol Air Qual. Res.*, **20** (2), 314–328. https://doi.org/10.4209/aaqr.2019.03.0124

[23] Vose R.S., Applequist S., Squires M., Durre I., Menne M.J., Williams Jr. C.N., Fenimore C., Gleason K. & Arndt D. (2014). – National Oceanic and Atmospheric Administration (NOAA) Monthly U.S. Climate Divisional Database (NClimDiv). NOAA National Climatic Data Center, Asheville, United States of America. https://doi.org/10.7289/V5M32STR

[24] Gastwirth J.L. (1972). – The estimation of the Lorenz curve and Gini index. *Rev. Econ. Stat.*, **54** (3), 306–316. https://doi.org/10.2307/1937992

[25] Crowder K. & South S.J. (2011). – Spatial and temporal dimensions of neighborhood effects on high school graduation. *Soc. Sci. Res.*, **40** (1), 87–106. https://doi.org/10.1016/j.ssresearch.2010.04.013

[26] Krivo L.J., Peterson R.D. & Kuhl D.C. (2009). – Segregation, racial structure, and neighborhood violent crime. *Am. J. Sociol.*, **114** (6), 1765–1802. https://doi.org/10.1086/597285

[27] Sampson R.J. & Graif C. (2009). – Neighborhood social capital as differential social organization: resident and leadership dimensions. *Am. Behav. Sci.*, **52** (11), 1579–1605. https://doi.org/10.1177/0002764209331527

[28] Sampson R.J. (2012). – Great American City: Chicago and the Enduring Neighborhood Effect. University of Chicago Press, Chicago, United States of America, 552 pp. https://doi.org/10.7208/chicago/9780226733883.001.0001

[29] Kim S.-Y., Bechle M., Hankey S., Sheppard L., Szpiro A.A. & Marshall J.D. (2020). – Concentrations of criteria pollutants in the contiguous U.S., 1979–2015: role of prediction model parsimony in integrated empirical geographic regression. *PLoS ONE*, **15** (2), e0228535. https://doi.org/10.1371/journal.pone.0228535

[30] Dalmat R.R., Mooney S.J., Hurvitz P.M., Zhou C., Moudon A.V. & Saelens B.E. (2021). – Walkability measures to predict the likelihood of walking in a place: a classification and regression tree analysis. *Health Place*, **72**, 102700. https://doi.org/10.1016/j.healthplace.2021.102700

[31] Legdeur N., Heymans M.W., Comijs H.C., Huisman M., Maier A.B. & Visser P.J. (2018). – Age dependency of risk factors for cognitive decline. *BMC Geriatr.*, **18** (1), 187. https://doi.org/10.1186/s12877-018-0876-2

[32] Vermote E. (2019). – NOAA Climate Data Record (CDR) of AHVRR Normalized Difference Vegetation Index (NDVI), Version 5. NOAA National Centers for Environmental Information, Asheville, United States of America. https://doi.org/10.7289/V5ZG6QH9

[33] Jackson R.B., Vengosh A., Carey J.W., Davies R.J., Darrah T.H., O'Sullivan F. & Pétron G. (2014). – The environmental costs and benefits of fracking. *Annu. Rev. Environ. Resour.*, **39** (1), 327–362. https://doi.org/10.1146/annurev-environ-031113-144051

[34] Reid C.E., Jerrett M., Tager I.B., Petersen M.L., Mann J.K. & Balmes J.R. (2016). – Differential respiratory health effects from the 2008 northern California wildfires: a spatiotemporal approach. *Environ. Res.*, **150**, 227–235. https://doi.org/10.1016/j.envres.2016.06.012

[35] Xu R., Yu P., Abramson M.J., Johnston F.H., Samet J.M., Bell M.L., Haines A., Ebi K.L., Li S. & Guo Y. (2020). – Wildfires, global climate change, and human health. *N. Engl. J. Med.*, **383** (22), 2173–2181. https://doi.org/10.1056/NEJMsr2028985

[36] Eremeeva M.E. & Dasch G.A. (2015). – Challenges posed by tick-borne rickettsiae: eco-epidemiology and public health implications. *Front. Public Health*, **3**, 55. https://doi.org/10.3389/fpubh.2015.00055

[37] Shrider E.A., Kollar M., Chen F. & Semega J. (2021). – Income and poverty in the United States: 2020. US Census Bureau, Current Population Reports, **P60-273**. US Government Publishing Office, Washington, DC, United States of America, 83 pp. Available at: https://www.census.gov/content/dam/Census/library/publications/2021/demo/p60-273.pdf (accessed on 18 April 2022).

[38] Sankar P.L. & Parker L.S. (2017). – The precision medicine initiative's All of Us research program: an agenda for research on its ethical, legal, and social issues. *Genet. Med.*, **19** (7), 743–750. https://doi.org/10.1038/gim.2016.183

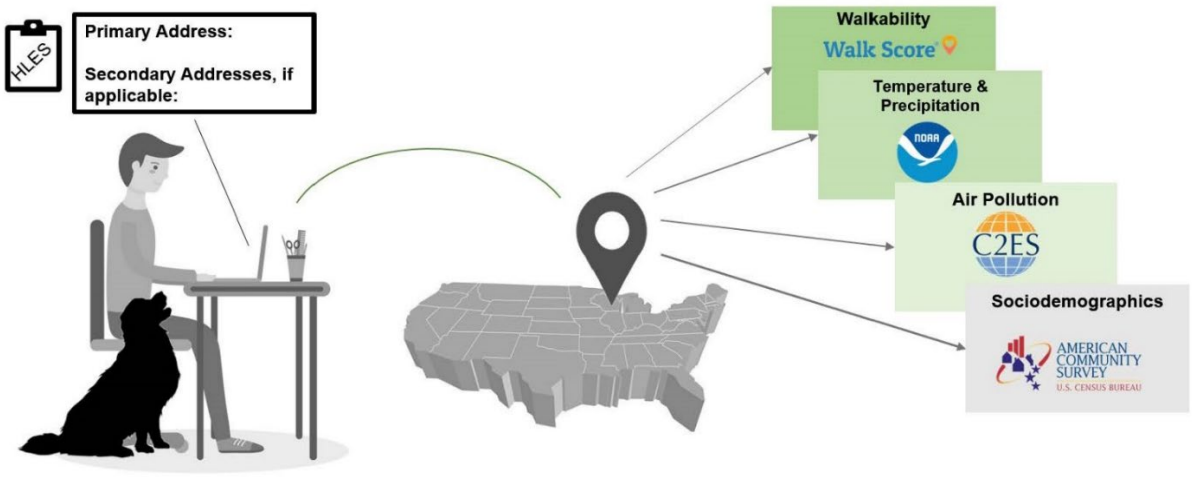
Table I

Variables collected from HLES and secondary data sources

Multi-level environmental data across four domains are collected as part of the DAP data. This table shows selected variants in each category. The full set of available variables can be found at https://dogagingproject.org/open_data_access

From HLES ***From GIS-linked source***		*Level*			
		Home	**Yard**	**Neighbourhood**	**Extra local**
Environmental domain	**Physical and built environment**	Year built Stairs Pipes in home Fencing	Sun exposure Surface types	Sidewalks	Owner's work environment
		Total size		Parks and greenspace *Walkability score* *Average temperature* *Average precipitation*	
	Chemical environment and exposures	Heating sources Cooking fuel Water source Water filtration Smoke exposure Flea and tick preventatives Lead paint Radon	Pesticide application and frequency Herbicide application and frequency	Traffic noise *Nitrogen dioxide* *PM2.5* *PM10* *Ozone*	Exposure to grooming products
	Diet and exercise	No. of times fed per day Primary component of diet Secondary components of diet Treat types and frequency Supplement types and frequency	Activity levels modified by local weather Activities with a lead of leash Activities without a lead of leash Bodies of waters available for swim or play		
	Social environment	Number of animals Number of people Household income		*Median household income* *Stability index*	Accessibility of dog parks Accessibility of doggy day-care
			Interactions with other animals and humans		

DAP: Dog Aging Project
GIS: geographic information system
HLES: Health and Life Experience Survey
PM: particulate matter

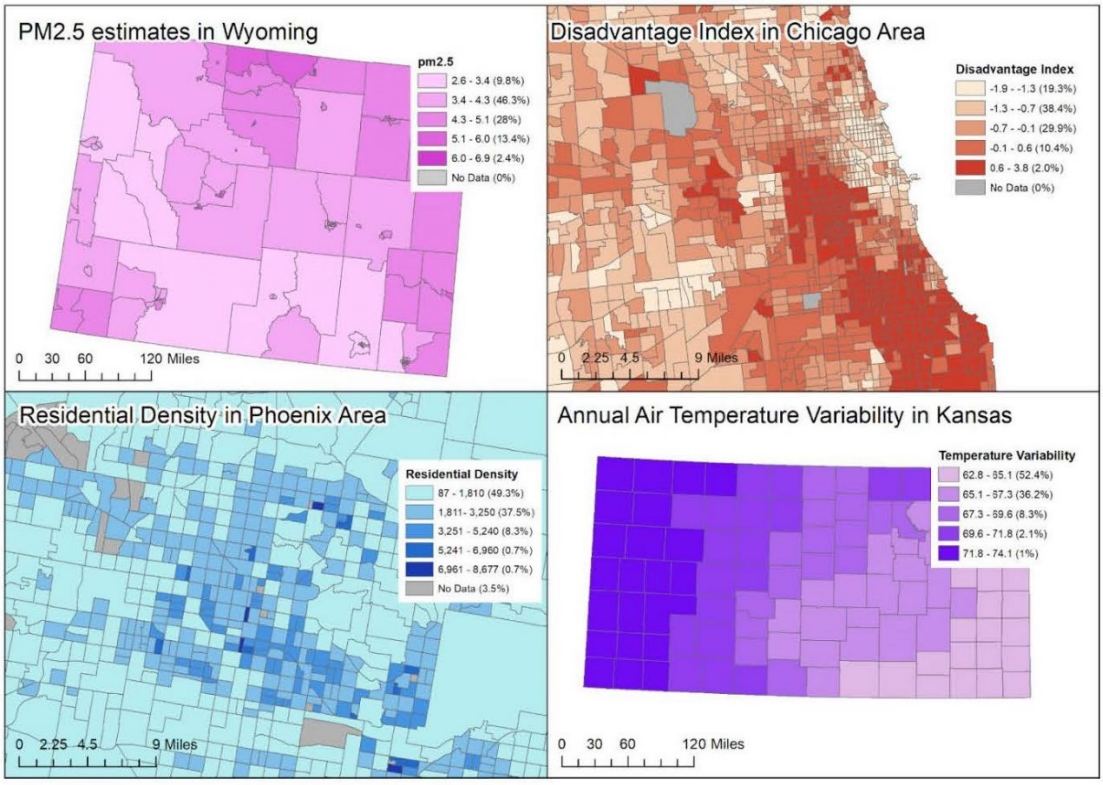


ACS: American Community Survey
CACES: Center for Air, Climate, and Energy Solutions
DAP: Dog Aging Project
FIPS: Federal Information Processing Standards
HLES: Health and Life Experience Survey
NOAA: National Oceanic and Atmospheric Administration

Figure 1

Conceptual figure

DAP participants are asked to provide the primary and secondary addresses where their dog resides or spends time as part of the DAP HLES. We can then geocode these addresses to obtain spatial coordinates and FIPS codes that can be linked to numerous publicly available secondary sources. Thus far, all geocoded addresses have been linked to data from ACS, CACES, NOAA, and WalkScore. These environmental variables can be merged with data from other DAP surveys



DAP: Dog Aging Project
PM: particulate matter

Figure 2

Environmental variability for selected regions and variables

The variability of one variable from each environmental domain is depicted here, within a selected region. Each variable is split into five categories, with the proportion of DAP participants residing in each respective category described in parentheses. PM2.5 represents an estimate of long-term outdoor concentrations of fine particulate matter 2.5 micrometres or smaller. Disadvantage index is composed of the following five socioeconomic variables: 1) Percentage of tract population below 125% of the poverty line; 2) Percentage of tract population ages 16–64 unemployed in the labour force and not in the labour force; 3) Percentage of own children living in households with female householder, no husband present; 4) Percentage of population 25 years or older with less than a bachelor's degree; and 5) Percentage of households earning under $100,000 (inflation-adjusted) in the past 12 months. Residential density is an estimate of the number of housing units in a census tract. Annual air temperature variability is a measure of difference between average highest and lowest recorded temperature in the past year