Title
From California sea lions to urban coyotes: Maximizing insights from Leptospira surveillance in coastal California wildlife

Author
Helman, Sarah Kate

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

From California sea lions to urban coyotes:

Maximizing insights from *Leptospira* surveillance in coastal California wildlife

A thesis submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Biology

by

Sarah Kate Helman

2022

ABSTRACT OF THE DISSERTATION

From California sea lions to urban coyotes:

Maximizing insights from *Leptospira* surveillance in coastal California wildlife

by

Sarah Kate Helman

Doctor of Philosophy in Biology

University of California, Los Angeles, 2022

Professor James O. Lloyd-Smith, Chair

Many pathogens, and all zoonotic pathogens, are capable of infecting multiple hosts. Understanding the transmission dynamics of multi-host, generalist pathogens is a major frontier in disease ecology, with far-reaching implications for both animal and human health. A central challenge is that multiple lines of evidence are required to identify maintenance hosts and assess the relative transmission contributions of multiple species. This evidence can unfortunately be difficult to obtain, particularly in wildlife systems, due to resource limitations and low sample accessibility. To address these challenges, I utilize computational tools to maximize the insights gained from limited wildlife data, using the globally significant zoonotic pathogen *Leptospira interrogans* in California's coastal wildlife as a case study.

Leptospira interrogans serovar Pomona presents a unique long-term case study of multi-host, generalist pathogen dynamics in California wildlife. This pathogen has affected the California sea lion (*Zalophus californianus*) population for decades, causing low levels of infection year-round and recurrent cyclical outbreaks of disease every few years. More recently, *Leptospira* has been found to be circulating among terrestrial wildlife in the California Channel Islands as well. Despite multi-year surveillance of the bacteria in California sea lions (Zalophus californianus) and Channel Island wildlife, the mechanisms governing transmission and persistence in this system are still unclear, as are potential connections to circulation of *Leptospira* among mainland wildlife host species. In the following chapters, I investigate the transmission potential and prevalence of *Leptospira interrogans* in California's coastal wildlife.

Within ecological systems, infection prevalence is critical to understanding pathogen dynamics, as it reflects transmission risk to others. However, uncertainty in the accuracy of diagnostic assays makes prevalence estimation difficult, particularly in wildlife where test methods are often not validated, and sample sizes may be low. Bayesian latent class analysis (BLCA) offers a statistical solution to this problem, but research detailing its limitations and usefulness in biological systems is lacking. In my first chapter, I estimate disease prevalence and diagnostic test accuracy using simulations to assess the ability of BLCA to produce accurate estimates across a range of biological conditions. I demonstrate that this method is effective, but has the potential to bias estimates depending on underlying biological system traits (e.g., sample size, test accuracy, and true prevalence). I use the California sea lion system as a case study to assess infection prevalence and test accuracy, describing situations in which this method would be preferable to results from a single high quality diagnostic test. Our findings directly benefit

scientists and veterinary professionals working on the California sea lion system, and, more generally, they validate a statistical tool and show ecologists when this technique may be of use.

In chapter two, I develop models to predict *Leptospira* shedding, indicative of transmission potential, in California sea lions. Shedding can be detected via polymerase chain reaction (PCR) of urine or kidney samples to identify *Leptospira* DNA, but obtaining these samples is difficult, and historical data are limited. Antibody titers were previously identified as predictive of shedding in this species, but antibody results take time and are not always available. I utilized LASSO regression to assess if shedding predictions from antibody titers improve in the presence of additional environmental, clinical, and demographic data. I then exclude antibody results to identify more accessible data that are predictive of shedding in their absence, and show that these predictions are robust to differences in the underlying sample population. Extrapolations to out-of-sample data provide accurate shedding estimates in the broader sea lion population, providing key information for understanding *Leptospira* transmission and persistence in California sea lions.

Understanding multi-host pathogen dynamics requires identification of possible hosts and the assessment of pathogen prevalence and transmission links in relevant host species. A closely related strain of *Leptospira interrogans* serovar Pomona has been identified in California sea lions, Channel Island foxes (*Urocyon littoralis*), and island spotted skunks (*Spilogale gracilis amphiala*), but it is unknown if mainland coastal wildlife play a transmission role in this multi-host pathogen system. In my final chapter, I conduct the first extensive survey of *Leptospira* in Southern California wildlife, using serology to investigate possible links to *Leptospira* in sea lions. Sampling primarily focused on five core species in the greater Los Angeles region: coyotes (*Canis latrans*), raccoons (*Procyon lotor*), Virginia opossums (*Didelphis virginiana*), striped

skunks (*Mephitis mephitis*) and fox squirrels (*Sciuris niger*). Infections were detected in all core species except fox squirrels, and all five species exhibited *Leptospira* exposure and were reactive to serovar Pomona. This evidence of widespread *Leptospira* circulation demonstrates a potential risk to both animal and human health across the Los Angeles region, and animals with primary reactivity to serovar Pomona represent future sampling targets to assess possible transmission links in the broader multi-host system.

Using statistical techniques to analyze multi-year surveillance data from California sea lions and mainland terrestrial mammals, I address critical knowledge gaps in the ecology of *Leptospira* in the coastal California ecosystem. Maximizing the information gained from limited data allows us to better understand the local prevalence and transmission ecology of this globally significant zoonosis, with direct applications for public health and wildlife management. Extending these methods to other systems will empower future multi-host pathogen studies, addressing key challenges in wildlife disease surveillance and ecology.

The dissertation of Sarah Kate Helman is approved.

Bradley H. Shaffer

Noa Pinter-Wollman

Kyle C. Cavanaugh

James O. Lloyd-Smith, Committee Chair

University of California, Los Angeles

2022

To my parents, who gave me the world.
To the animals, for giving it purpose.
To my friends and family, for elevating it.
To Mason, for completing it.
To my children, who shall inherit it,
and who make it infinitely better every day.
My love and immeasurable gratitude to you all.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SUPPLEMENTAL FIGURES

LIST OF SUPPLEMENTARY TABLES

ACKNOWLEDGMENTS

I count myself lucky that so many people have helped me with this accomplishment. A few are mentioned here, but I could write a PhD-length gratitude book in its own right to everyone who helped me on this path. So, to my people everywhere: please know I am grateful, I couldn't have done it without you, and I love you very much. Thank you from the bottom of my heart.

Thank you to my lab…

To my advisor, Jamie Lloyd-Smith (JLS). For taking a chance on a small animal vet in Australia who wanted to research wildlife, and for facilitating all the research that went into this dissertation. For your patience, guidance, mentorship, and countless hours of feedback and support that made it all possible. For the scramble at the end. For being the brilliant scientist, writer and big thinker that you are, and inspiring me to reach for brilliance. Thank you for the doors you have helped me open, and I look forward to continuing to collaborate with and learn from you in the years to come.

To Riley Mummah. You are a coauthor and copilot for this entire dissertation. You have improved both my life and research for the better, and this feat would not have been possible without you. I am eternally grateful for the immeasurable contributions and countless hours that you contributed to this dissertation, both intellectually and personally. Thank you for your hard work, formatting wizardry, and for keeping my tank full and helping me cross the finish line.

To Katie Prager (KP), for being my esteemed mentor and dear friend. For good food and necropsies (not together), pep talks and conference adventures. For taking me both under your wing and out to catch sea lions, and helping me to let it go. I am lucky to know you and work with you, and appreciate your many contributions and everything you've done to make these chapters possible. I look forward to all of the laughs, memories and research still to come.

To Amanda Tokuyama, for being a great friend since we first bonded over necropsies. For joining me in the JLS lab, helping manage the mammal team, and for your countless hours of help and collaboration on this dissertation. For teaching me about worms and scat. For amazing tables. For being a

wonderful lab partner, coauthor, supportive friend and toddler idol. I am so glad we are continuing on our path of friendship and science.

To the Lloyd-Smith village: Jamie, Riley, Katie, Amanda, Michael Buhnerkempe, Katie Gostic, Benny Borremans, Monique Ambrose, Amandine Gamble, Ana Gomez, Christian Mason, Angela Guglielmino, Oshiomah Oyageshio, Celine Snedden, Dylan Morris, Lizzie Blackmore, Dave Daversa, and the Lloyd-Smiths (Katy, Sophie and Liam). Thank you all so much. To all of my fool's gold coauthors, thank you for helping me code Chapter 1 into a reality. To Monique and Katie G., for your inspirational intelligence and positivity, I am so lucky to have shared an office with you. To Ana, for all you do on team lepto, and for supporting me since day one. To Benny, for your optimism and encouragement, for suggesting and facilitating the LASSO, for letting me be a part of your research, and for your ongoing support with mine. To my lab coauthors and the entire lepto team, for all you have done to build these chapters. To everyone, thank you for your friendship, time, lab meetings, feedback, collaborations, help, code, support and encouragement. For helping me take my projects from NoAb to MoAb, and being amazing coauthors and colleagues. Your inspirational collective brilliance has helped me become the scientist I am today. It takes a village and I couldn't have done it without you, and I am grateful to have more time to learn and grow with many of you in the two years to come.

To my broader UCLA supporters and funders…

To my committee, Noa Pinter-Wollman, Kyle Cavanaugh and Brad Shaffer, thank you for your time, support, invaluable feedback and encouragement.

To my SLA research team (Jessica Lynch, Hazel Byrne, Tony Friscia, Chris Kelty, KP and JLS), for the grant that led to exciting local wildlife research and collaborations. Tony and KP, thank you for your help with necropsy trainings. To Jocelyn Yamadera, Tessa Villaseñor, the funding managers and the broader EEB Department for all your behind the scenes academic and financial support. To Humberto Barba, Alberto Gonzalez and everyone who manages TLSB facilities and the loading dock, thank you for your smiles and helping make Chapter 3 possible. To Alan Garfinkel, Nick Wisniewski and Khris

Griffis, for my statistical awakening; thank you for helping me grow the coding confidence and the Bayesian/bootstrapping foundation I needed for Chapters 1 and 2. To the many other coauthors, collaborators, and friendly faces in the UCLA community that my life and career were made better by, I am grateful to you all.

To our research assistants, who made Chapter 3 possible: Caitlin Cox, Isobel Tweedt, Julianne Nussbaum, Nihal Punjabi, Abby Low, Daniel Ribal, Molly Maloney, Reed Munson, Alex Molina, Sarah Nugen, Philip Kim, Brady DeMattei, Renee Crippen, Aonya Montoya, Genesis Navarrette, Ryan Neis, Ally Osterland, Nick Rodgers, Valerie Millette, Macy Ward, Dayanira Cruz, Amanda Hagen, Marcus Weiss, Monica Arnaout, Lauryn Yee, Shefali Kaul, and Danielle Tran, Carlye Rosen and many others. Thank you for being part of our lab and for caring about local wildlife research. Your time, dedication and innumerable contributions to this project won't be forgotten. Keep calm and carrion.

To my funders. This work was supported by the National Science Foundation Awards OCE-1335657, the U.S. Department of Defense Strategic Environmental Research and Development Program Award RC-2635, the John H. Prescott Marine Mammal Rescue Assistance Grant Program and the UCLA Sustainable LA Grand Challenges program. To Brad Shaffer, Gary Bucciarelli, and the others behind Stunt Ranch & La Kretz Center for California Conservation Science, thank you for countless rounds of support as a graduate student, and for supporting me as a soon-to-be postdoc. I am so excited to continue working with you all.

To project contributors beyond UCLA….

Thank you to the wildlife warriors who facilitated the sample collection which made this work possible, especially the wonderful staff, volunteers and biologists at The Marine Mammal Center, the National Marine Mammal Laboratory, the California Wildlife Center (CWC), USDA/APHIS, the LA County Department of Animal Care and Control, the Department of Defense and the National Park Service. To Drs. Emily Beeler and Maria Sabio-Solacito, for helping make Chapter 3 sampling a reality. To Drs. Duane Tom and Lorraine Barbosa, and countless others at CWC for your immense contributions.

To Justin Brown, Seth Riley and NPS, for collaborative carcass collection, facilitating necropsies before I had a lab, inspiring me to ask questions about local coyote ecology, and supporting my efforts to answer more of them as a post-doc. To the Natural History Museum of Los Angeles County, for your passion, inspiration, collaboration and outreach. To Greg Pauly and Kayce Bell and everyone at the Urban Nature Research Center, for your encouragement and support, I'm excited to continue collaborating with you all. To Javier Monzón, for your collaboration and help with the land class analysis in Chapter 3. To the people who analyzed our samples, thank you for your time, especially Renee Galloway. Thank you to all of the agencies who have facilitated this work and helped us learn more about sea lions and CA wildlife, and thank you all for your friendship along the way.

To my beloved family and friends…

To my mom and dad, for filling my life with joy and love, and my childhood with animals, nature and science museums. Those moments shaped my view of the world and started me down the path I am on today. Thank you for keeping me on that path and supporting me always. Thank you both for a million memories, and for sending me out into the world to find my way, no matter how many oceans you had to wave across. Thank you for continuing to support me and my family throughout this academic journey, you helped make this possible. You are both forever loved and I wouldn't be who I am without you, and I am forever grateful.

To Mason Gamble. My wonderful husband, who started as my classmate when I began this dissertation. Thank you for being part of this journey since day one, and for wearing a sea lion shirt that day which gave us an instant connection (though I'm pretty sure we had one without it). For re-teaching me and our entire cohort calculus, and explaining things so thoughtfully and enthusiastically that I realized I wanted to listen to your voice forever. For organizing our wedding because I was so busy writing my prospectus and getting ready for my defense. For being my behind the scenes coauthor, and for your immeasurable help and support. For living environmental impacts on sea lion prey resources when your anchovies vanished, I learned and continue to learn so much from you every day. For staying

so positive dealing with that and other hurdles we have faced. For babirusa, hospital water and strawberry-banana, and always lifting me up. For being my best friend and the most incredible father to our children, who I am blessed with because of you. I could write an entire 4th chapter dedicated to you, but this paragraph will have to do. I love you so much.

To my babies, for filling my life with happiness, and redefining what I know about love. To Isobel Lily, my Eeweebelle, thank you for making me a mother, and filling every day since then with fountains of love, joy and wonder. You're my Zoom-Zoom who inspires me to never to give up and keep running, even when I am tired. Thank you for being so patient while I worked so much to get here, I know it's so hard to wait, and I appreciate how much you did. Thank you for asking about my life and work, believing I am a super hero, "helping me work", and for sending the first email to my committee for me. And to Logan James, my little big man, my sweet smiley guy. You are the ray of sunshine that lights up my whole world. You both inspire me to be a better mom and person, and I am doing what I can to leave the world better than I found it for you. I love you both so much.

To Jim and Katherine Helman. Thank you for helping me navigate childhood (Jim), and continuing to help me navigate family and life. For being a loving aunt and uncle. I treasure your love, support and the funny memories we continue to share, and I am so grateful for you both. I love you and look forward to so many more adventures with you in the years to come.

To my extended family, for loving me and supporting me from afar, and sharing so much love with me and my children regardless of the distance between us. For so many little things and gifts and cards and texts and gatherings that brightened these past years. Thank you all.

To the Gamble family, for the cherished memories that we've shared that made this dissertation period brighter, and for your ever-flowing words of love, support and encouragement. I love you all.
To Sally and Tim, for driving across the country in a pandemic and helping our family for so long, I couldn't have done this without you. You're the most amazing in-laws I could ever have hoped for. Thank you for making my soulmate and shooting the most beautiful arrow out into the world to find me, and for making me a part of your wonderful family.

Riley & Bryant Crubaugh. You are two of my dearest friends, and my heart and life are so much fuller because of you. Thank you for supporting me through all of this. For cross-stiches, baked goods, Thanksgivings, monthly flavors, witches, Harry Potter, beach silhouettes, fun-efficiency and BBFs. For sandcastle cakes and wedding dress ice cream. For being Auntie Raroo and Unkie Ba-Ba. For finding lost treasures. For so many special moments and memories. I am eternally grateful for your love, support, and friendship, and I love and appreciate you both so much.

To countless other friends, thank you for your love and support throughout this degree and my development as a scientist. To Lorraine & Miguel Barbosa, for 'Salsalito', marine mammals, CWC, Cali adventures, baby pics, merengue cats & the macarena; for our girls bonding and our boys who will soon, for driving to stay with us during a pandemic, and for many more memories to come. To my Aussie family, for sharing wondrous wildlife and memories with me, and for teaching me to be a vet. To my Scotland and Dick Vet family, for starting my path in vet med with me and filling it with adventures and love. To my Irish family, for your love and shamrock peanuts, which continues to lift me up whenever I am weighed down. To my WIAG family & mother-light for opening my heart and soul. To my childhood friends, for helping me find myself and keeping me in your hearts. You are all missed and loved.

To Jim Brandenburg and wildlife documentarians everywhere, you enthralled and inspired me as a kid, and planted the seed of the dream I'm now living. Thank you for the inspiration.

To the animals and the land…

To the animals that were sampled during the course of this project: To those of you that were alive, I am sorry we bugged you a bit by catching you and releasing you. To those of you that were dead, may what we learn from you honor your memory and death in some small way. Thank you for making this research possible.

Land Acknowledgment I acknowledge the Gabrielino/Tongva, Chumash, Fernandeño Tataviam, Salinan, Esselen, Ohlone/Costanoan and Coast Miwok peoples as the traditional caretakers of the land where this research took place. As a researcher at a land grant institution, I pay my respects to their descendants and elders, past, present, and emerging.

Chapter 1 is the accepted version of a manuscript, now published as: Helman, SK, Mummah, RO, Gostic, KM, Buhnerkempe, MG, Prager, KC & Lloyd-Smith, JO. (2020). Estimating prevalence and test accuracy in disease ecology: How Bayesian latent class analysis can boost or bias imperfect test results. *Ecology and Evolution* 10, 7221-7232. DOI: 10.1002/ece3.6448. SH, MB, KP and JLS conceived of the idea for the study. SH drafted the manuscript. All authors helped develop the modeling analyses, interpret the results and revise the manuscript. We thank the Lloyd-Smith lab for helpful comments and discussions, and the staff and volunteers at The Marine Mammal Center for data collection.

Chapter 2 is a version of a manuscript currently in preparation: Helman, SK, Mummah, RO, Gomez, ACR, Borremans, B, Galloway, RL, Harris, JD, Orr, AJ, Melin, SR, DeLong, RL, Greig, DJ, Gulland, FMD, Prager, KC & Lloyd-Smith, JO. He Shed She Shed: Predictors of *Leptospira* urinary shedding in California sea lions (*Zalophus californianus*). SH, RM and BB wrote the code for the analyses with support from AG, KP and JLS. SH drafted the manuscript, and RM, BB, AG, KP and JLS helped with result interpretation and revisions. We are grateful to Renee Galloway for microscopic agglutination test results and the Lloyd-Smith lab for many helpful discussions. We thank the staff and volunteers at The Marine Mammal Center and staff at The National Marine Mammal Laboratory for data collection and system insights.

Chapter 3 is a version of a manuscript currently in preparation: Helman, SK, Mummah, RO, Tokuyama, AFN, Gomez, ACR, Borremans, B, Galloway, RL, Prager, KC & Lloyd-Smith, JO. An investigation of *Leptospira* in Southern California wildlife. SH conducted the wildlife sampling with the support of AT, KP and JLS. SH drafted the manuscript and made the maps, and SH, RM, AT and AG made the figures and tables. All authors helped with the interpretations of results, and RM, AT, KP and JLS helped revise the manuscript. We thank the Lloyd-Smith lab for helpful comments and discussions, local wildlife agencies for supplying samples and carcasses, and undergraduate researchers for facilitating data collection and curation.

VITA

2011-2014 Associate Veterinarian
Balaclava Veterinary Surgery, Cairns, QLD, Australia

2006-2010 B.V.M.&S. Bachelor of Veterinary Medicine and Surgery
Royal (Dick) School of Veterinary Studies, Edinburgh, Scotland

2001-2006 B.A. in Biology; Specialization in Ecology & Conservation Biology.
Boston University, Boston, MA, USA

PUBLICATIONS

Snedden C.E., Makanani S.K., Schwartz S.T., Gamble A., Blakey R.V., Borremans B., Helman S.K., Espericueta L., Valencia A., Endo A., Alfaro M.E. and Lloyd-Smith J.O. (2021). SARS-CoV-2: cross-scale insights from ecology and evolution. *Trends in Microbiology*, in press. DOI: 10.1016/j.tim.2021.03.013.

Lin M., Simons A.L., Harrigan R.J., Curd E.E., Schneider F.D., Ruiz-Ramos D.V., Gold, Z., Osborne M.G., Shirazi S., Schweizer T.M., Moore T.N., Fox E.A., Turba R., Garcia-Vedrenne A.E., Helman S.K., Rutledge K., Mejia M.P., Marwayana O., Munguia Ramos M.N., Wetzer R., Pentcheff D., McTavish E.J., Dawson M.N., Shapiro B., Wayne R.K. and Meyer R.S. 2021. Landscape analyses using eDNA metabarcoding and Earth observation predict community biodiversity in California. *Ecological Applications* 31(6): e02379. DOI:10.1002/eap.2379.

Borremans B., Gamble A., Prager K.C., Helman S.K., McClain A.M., Cox C., Savage V. and Lloyd-Smith J.O. (2020). Quantifying antibody kinetics and RNA shedding during early-phase SARS-CoV-2 infection. *eLife* 9, e60122. DOI: 10.7554/eLife.60122.

Helman, S.K., Mummah, R.O., Gostic, K.M., Buhnerkempe, M.G., Prager, K.C, Lloyd-Smith, J.O. (2020). Estimating prevalence and test accuracy in disease ecology: How Bayesian latent class analysis can boost or bias imperfect test results. *Ecology and Evolution* 10, 7221-7232. DOI: 10.1002/ece3.6448.

Helman, S. (2010). A case of basilar empyema in a Texel ewe lamb. UK Vet Livestock, 15: 41–48.

Received: 1 November 2019 | Accepted: 9 March 2020

DOI: 10.1002/ece3.6448

ORIGINAL RESEARCH

Ecology and Evolution Open Access WILEY

Estimating prevalence and test accuracy in disease ecology: How Bayesian latent class analysis can boost or bias imperfect test results

Sarah K. Helman[1] | Riley O. Mummah[1] | Katelyn M. Gostic[1] | Michael G. Buhnerkempe[1,2] | Katherine C. Prager[1,3] | James O. Lloyd-Smith[1,3]

[1]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, CA, USA

[2]Department of Internal Medicine, Southern Illinois University School of Medicine, Springfield, IL, USA

[3]Fogarty International Center, National Institutes of Health, Bethesda, MD, USA

Correspondence
Sarah K. Helman, Department of Ecology and Evolutionary Biology, University of California, Los Angeles, 610 Charles E Young Dr S, Los Angeles, CA 90095, USA.
Email: sarahkh@ucla.edu

Funding information
U.S. National Science Foundation, Grant/Award Number: OCE-1335657 and DEB-1557022; the Strategic Environmental Research and Development Program, Grant/Award Number: RC-2635; the John H. Prescott Marine Mammal Rescue Assistance Grant Program; the Hellman Family Foundation; UCLA Sustainable L.A. program; Science and Technology Directory; Department of Homeland Security; Fogarty International Center; National Institutes of Health, Grant/Award Number: F31AI134017 and T32-GM008185

Abstract

1. Obtaining accurate estimates of disease prevalence is crucial for the monitoring and management of wildlife populations but can be difficult if different diagnostic tests yield conflicting results and if the accuracy of each diagnostic test is unknown. Bayesian latent class analysis (BLCA) modeling offers a potential solution, providing estimates of prevalence levels and diagnostic test accuracy under the realistic assumption that no diagnostic test is perfect.
2. In typical applications of this approach, the specificity of one test is fixed at or close to 100%, allowing the model to simultaneously estimate the sensitivity and specificity of all other tests, in addition to infection prevalence. In wildlife systems, a test with near-perfect specificity is not always available, so we simulated data to investigate how decreasing this fixed specificity value affects the accuracy of model estimates.
3. We used simulations to explore how the trade-off between diagnostic test specificity and sensitivity impacts prevalence estimates and found that directional biases depend on pathogen prevalence. Both the precision and accuracy of results depend on the sample size, the diagnostic tests used, and the true infection prevalence, so these factors should be considered when applying BLCA to estimate disease prevalence and diagnostic test accuracy in wildlife systems. A wildlife disease case study, focusing on leptospirosis in California sea lions, demonstrated the potential for Bayesian latent class methods to provide reliable estimates under real-world conditions.
4. We delineate conditions under which BLCA improves upon the results from a single diagnostic across a range of prevalence levels and sample sizes, demonstrating when this method is preferable for disease ecologists working in a wide variety of pathogen systems.

KEYWORDS
Bayesian latent class, California sea lion, diagnostic test, disease, infection, prevalence, sensitivity, specificity

1 | INTRODUCTION

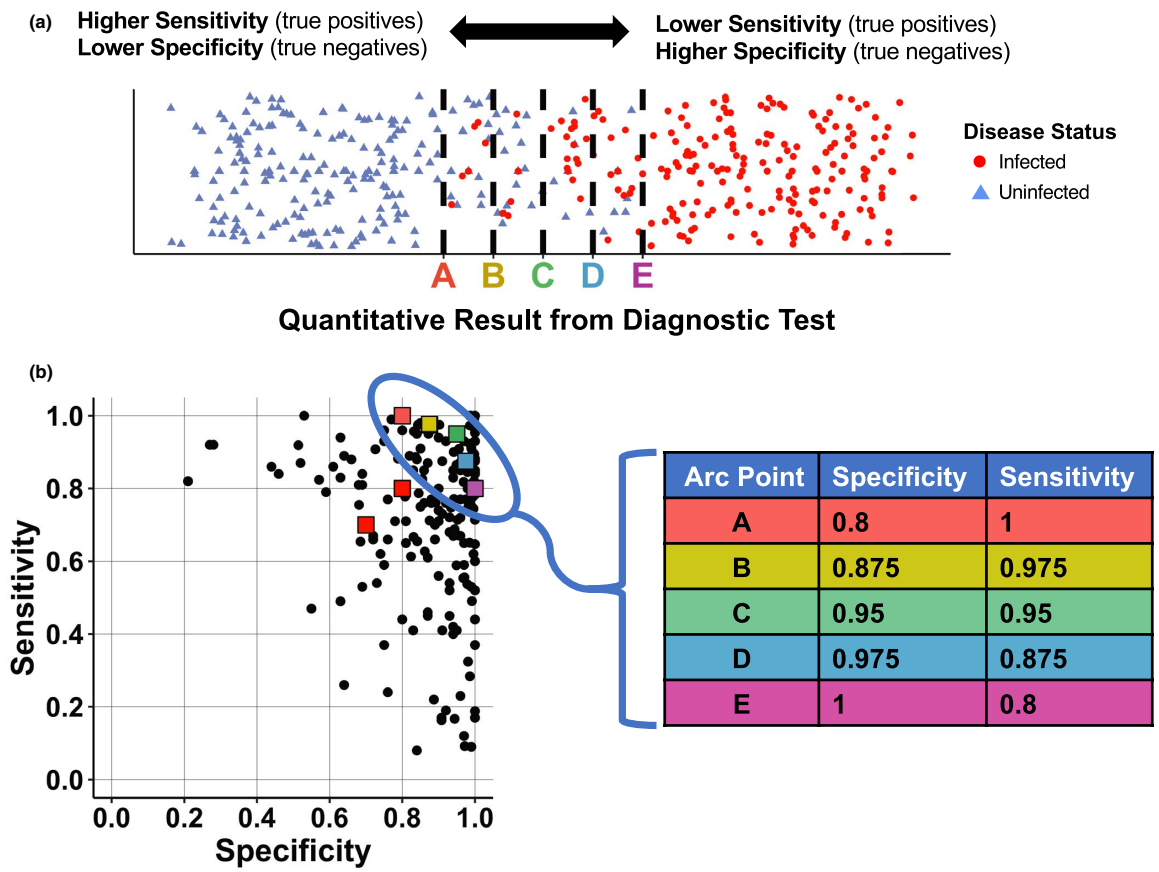
Infection prevalence, or the fraction of individuals in a population that are infected with a pathogen at a given time, is a crucial metric of pathogen dynamics within ecological systems (Buhnerkempe et al., 2015; Haydon, Cleaveland, Taylor, & Laurenson, 2002; Viana et al., 2014). Knowledge of infection prevalence can elucidate disease dynamics in a system, providing data to health professionals aiming to mitigate disease risk and to scientists seeking to understand key mechanisms. The true infection prevalence within an ecological system is usually impossible to measure exactly but can be estimated by testing representative subsets of a population. However, it can be difficult to obtain large representative data sets to estimate disease prevalence in wildlife populations. Limitations including funding, personnel, regulatory restrictions, and the availability of tests appropriate to a specific study species typically determine which diagnostic tests can be used in a given wildlife system and how many individuals can be tested. Wildlife studies face additional challenges, as they are often restricted to the use of diagnostic tests whose accuracy may not be known if the tests have been validated in domestic animals, rather than the host species of interest (Moreno-Torres, Wolfe, Saville, & Garabed, 2016).

While diagnostic test accuracy is sometimes overlooked in favor of more immediate challenges such as obtaining representative samples, it can have substantial impacts on disease prevalence estimates. Diagnostic tests vary in their sensitivity (probability of detecting true positives) and specificity (probability of detecting true negatives), so both individual diagnostic results and population-level prevalence estimates will vary depending on the tests used in a given system. Furthermore, a set of imperfect diagnostic tests may show conflicting results in the same individual (e.g., due to differences in test accuracy or what disease state the tests are measuring). Assessing the true infection status of individuals from imperfect information and using this information to estimate population prevalence is a challenge facing epidemiologists and disease ecologists worldwide.

To complicate matters further, when considering a test with continuous quantitative output, users must choose a diagnostic threshold that separates negative test results from positive results. A trade-off exists between sensitivity and specificity, such that this threshold can be lowered to make the test more sensitive (limiting the number of false-negative results) or raised to make the test more specific (limiting the number of false-positive results; Florkowski, 2008). Many tests that are conventionally viewed as binary, such as serology or even polymerase chain reaction (PCR), actually have underlying quantitative thresholds that could be tuned to maximize sensitivity or specificity, but not both. Disease ecologists and epidemiologists routinely use different thresholds for diagnostic assays, depending on their research aims and system characteristics (Almberg, Cross, Dobson, Smith, & Hudson, 2012).

In situations where careful choice of diagnostic threshold is not itself sufficient to improve prevalence estimates, a statistical method called Bayesian latent class analysis (BLCA) has been applied to facilitate estimates of infection prevalence and diagnostic test accuracy (Gonçalves et al., 2012; Limmathurotsakul et al., 2012; Muma et al., 2007; Pan-ngum et al., 2013). When applying this technique, an individual's true clinical infection status is assumed to be a latent unobserved process. BLCA does not explicitly categorize each individual as infected or uninfected. Rather, each tested individual has a probability of being infected or uninfected, given their observed combination of test outcomes and the accuracy of each test. The model integrates probabilistic information about the true infection status of all tested individuals to simultaneously estimate overall infection prevalence, along with the sensitivity and specificity of each test, under the realistic assumption that no diagnostic test is perfect (Rindskopf & Rindskopf, 1986). Traditionally, BLCA methods assume conditional independence of test results, given the disease status of a tested individual. Recent research has addressed the issue of identifiability and potential for biases due to the underlying dependence structure among test results, as well as approaches to modeling conditional dependence and adding random effects to address these challenges (Albert & Dodd, 2004; Dendukuri & Joseph, 2001; Hadgu & Qu, 1998; Jones, Johnson, Hanson, & Christensen, 2010; Pepe & Janes, 2006; Qu, Tan, & Kutner, 1996). Since higher-order information (e.g., longitudinal sampling) is unlikely to be available for diagnostic tests in wildlife hosts, here we analyze the performance of BLCA under the assumption of conditional independence (Wang & Hanson, 2019). This assumption is reasonable when diagnostic tests measure distinct biological processes that are not expected to be substantially correlated (e.g., the presence of a pathogen in urine vs the antibody response to a pathogen in the bloodstream; Kostoulas et al., 2017), and this study assesses the application of BLCA in systems where this assumption is valid.

Bayesian latent class analysis has been used primarily to estimate disease prevalence and test accuracy in domestic animals (Basso et al., 2013; Boelaert, Aoun, Liinev, Goetghebeur, & Van der Stuyft, 1999; Hartnack et al., 2013; Mathevon, Foucras, Falguières, & Corbiere, 2017; Muma et al., 2007; Nielsen, Toft, & Ersbøll, 2004) or humans (Gonçalves et al., 2012; Limmathurotsakul et al., 2012; Pan-ngum et al., 2013; Schumacher et al., 2016), but it has also been applied sparsely in wildlife systems (Bronsvoort et al., 2008; Moreno-Torres et al., 2016; Verma-Kumar et al., 2012). The limitations and biases from test sensitivity and specificity, and situations where BLCA improves upon single test estimates, have not previously been explored in the context of wildlife. Our study assesses the accuracy and potential for bias across a range of biologically



Arc Point	Specificity	Sensitivity
A	0.8	1
B	0.875	0.975
C	0.95	0.95
D	0.975	0.875
E	1	0.8

FIGURE 1 Infection status for a group of individuals relative to the sensitivity and specificity of test thresholds (top), and the values we chose for simulations relative to levels reported in the literature. (a) The true infection status (red circles = infected, blue triangles = uninfected) is plotted for hypothetical test results in a group of individuals. The *x*-axis represents a range of quantitative test results, with lower test results on the left and higher test results on the right. A threshold must be chosen, above which value a test result is considered positive. Thresholds that correspond to points A-E in b are shown as dashed black lines, demonstrating the trade-off between sensitivity (True Positives/(True Positives + False Negatives)) and specificity (True Negatives/(True Negatives + False Positives)). (b) Diagnostic test sensitivities and specificities previously reported in the literature (Alberg et al. 2004; Maxim et al. 2014), shown as black circles. For data simulations, the sensitivity/specificity values of test 1 and test 2 were set at 70%/70% and 80%/80%, respectively (shown as red squares). The fixed specificity provided to the Bayesian model was selected from points A-E (table on right)

realistic sample sizes and prevalence levels by applying BLCA to simulated data. When using BLCA models, the specificity of the most accurate test is typically fixed at or close to 100% (Hartnack et al., 2013; Limmathurotsakul et al., 2012; Mathevon et al., 2017; Pan-ngum et al., 2013; Schumacher et al., 2016), which is often not the case in real-world conditions, particularly when dealing with wildlife. Our analysis relaxes this assumption, simulating diagnostic test data using multiple diagnostic test ensembles to investigate BLCA efficacy as fixed test specificity decreases from 100% to 80%. In doing so, we also provide actionable guidance for situations where the investigators can choose the diagnostic threshold to tune the specificity of their fixed test.

To demonstrate the application of this method in a wildlife system, we apply BLCA to *Leptospira* surveillance data from California sea lions (*Zalophus californianus*). The bacteria *Leptospira interrogans* serovar Pomona is one of the primary causes of strandings in this species, having caused cyclical outbreaks since the mid-1980s that are associated with high morbidity and mortality (Greig, Gulland, & Kreuder, 2005; Lloyd-Smith et al., 2007; Prager et al., 2013). Animals with the disease, known as leptospirosis, present with clinical signs associated with *Leptospira*-induced kidney dysfunction (Cameron et al., 2008). While detection of *Leptospira* DNA in the urinary tract (Polymerase Chain Reaction - PCR) is the definitive diagnosis, obtaining samples to test via PCR is often impossible, so high antibody titers (Microscopic Agglutination Test - MAT) and serum chemistry markers indicative of *Leptospira*-induced kidney dysfunction are also utilized to detect clinical infections. We used BLCA to estimate the prevalence of clinical infections in

stranded California sea lions, along with test sensitivity and specificity, using results from these three diagnostic tests. We then simulated data using the parameter estimates from the sea lion data to assess BLCA estimates for prevalence and test accuracy under real-world sample sizes and testing conditions. Finally, we compared prevalence estimates from the BLCA model to what would be estimated from a single diagnostic test, to understand the circumstances under which the BLCA method is most worthwhile. In combination, analyses of the simulated data and results of the wildlife case study provide insights into the use and limitations of BLCA in disease ecology.

2 | METHODS

The Bayesian latent class model used in this analysis requires binary test outcomes. Thus, tests that yield results on a continuous scale (e.g., serological titers or quantitative PCR Ct values) must be classified as positive or negative, based on whether the test result falls above or below a diagnostic threshold. This classification threshold can be selected to maximize sensitivity (Se) or specificity (Sp) and must be chosen carefully for each test (Figure 1a). We simulated results from three diagnostic tests, using a hypothetical diagnostic test ensemble by selecting pairs of sensitivities and specificities from a range of previously reported values for 193 medical tests in the literature (Alberg, Park, Hager, Brock, & Diener-West, 2004; Maxim, Niebo, & Utell, 2014; Figure 1b). Two of the tests (tests 1 and 2) were assigned lower Se/Sp combinations (Figure S1b,c), representative of more mediocre diagnostic tests reported in the literature. The remaining test (test 3) in the ensemble was assigned properties chosen across an arc of Se/Sp values from 100%/80% to 80%/100% (Figure 1b: points A-E), which corresponded to the highest Se/Sp combination of the three tests. This range was chosen because nearly one third of tests in the literature survey (n = 63/193) had both sensitivity and specificity > 80%, so we assumed biologists would usually have at least one diagnostic test that fell within this range. In each simulation, the specificity of test 3 (which was always the highest specificity of all three tests) was fixed in the BLCA model.

2.1 | Parameter selection

We chose a range of biologically realistic parameter sets, using each one to simulate diagnostic test data that were then analyzed using BLCA. Each parameter set included the sample size, sensitivity and specificity values for three hypothetical diagnostic tests and the "true" underlying infection prevalence in a hypothetical sample population. Tests 1 and 2 had Se/Sp fixed at 70%/70% and 80%/80%, respectively (red squares in Figure 1b), and these values remained constant for all data simulations. Test 3 was selected sequentially from points A-E, such that the fixed specificity provided to the model decreased from 100% to 80% (A-E in Figure 1b). We simulated datasets using these five initial diagnostic test selections ((test1, test2) x (test A-E)), seven sample sizes (n = 20, 40, 80, 160, 320, 640 and 1,280) and three prevalence levels spanning a broad range of ecological scenarios (10%, 50%, 90%), resulting in 105 unique parameter sets. To assess whether observed patterns were influenced by the initial choices for tests 1 and 2 (which had Se:Sp ratios of 1:1), the following two alternate selections for these tests were used: Se/Sp for tests 1 and 2 set to 90%/70% and 70%/90%, respectively (Figure S1b), and the Se/Sp for tests 1 and 2 set to 80%/60% and 50%/90%, respectively (Figure S1c). These alternative scenarios explore different Se:Sp ratios as well as different overall quality of tests 1 and 2.

2.2 | Data simulation

For each parameter set (i.e., sample size, prevalence, and hypothetical test combination), a number of individuals (equal to sample size*prevalence) were assigned the status infected, and all remaining individuals in the population were assigned the status uninfected (Figure 2a). Once infection status was set, a series of Bernoulli trials was used to simulate the outcome of each hypothetical diagnostic test. Among infected individuals, the probability of a positive result was equal to test sensitivity, and the probability of a negative result was equal to (1-Se). Among uninfected individuals, the probability of a positive result was equal to (1-Sp), and the probability of a negative result was equal to test specificity. These simulations generated a set of three binary test outcomes for each individual, assuming independence among tests, with eight possible combinations of positive and negative test results (ranging from all negative to all positive). The number of individuals that fell into each of the eight possible test result combinations was counted (test profiles a-h; Figure 2b), and this vector was saved to input in the BLCA model. Test results were simulated 1,000 times for each set of parameters.

2.3 | Bayesian latent class analysis

Bayesian latent class analysis is a likelihood-based statistical method that estimates the prevalence of particular class types within a population sample. Here, individuals fall into one of eight observed classes (a-h), based on the profile of their diagnostic test outcomes (Limmathurotsakul et al., 2012; Figure 2b). Our model assumes the outcome of each diagnostic test is independent of the others, conditional on the individual's underlying (and unknown) state with respect to pathogen infection and disease. Thus, the probability of obtaining a given diagnostic profile depends on the probability that an individual was truly infected (equal to population prevalence) and on the outcome of each diagnostic test given the underlying infection status. As sensitivity is defined as the probability of detecting true positives and specificity is defined as the probability of detecting true negatives, the probability of three negative test outcomes (diagnostic profile a), is:

$$p(a) = \pi\left(1-\mathrm{Se}_1\right)\left(1-\mathrm{Se}_2\right)\left(1-\mathrm{Se}_3\right) + \left(1-\pi\right)\left(\mathrm{Sp}_1\right)\left(\mathrm{Sp}_2\right)\left(\mathrm{Sp}_3\right).$$

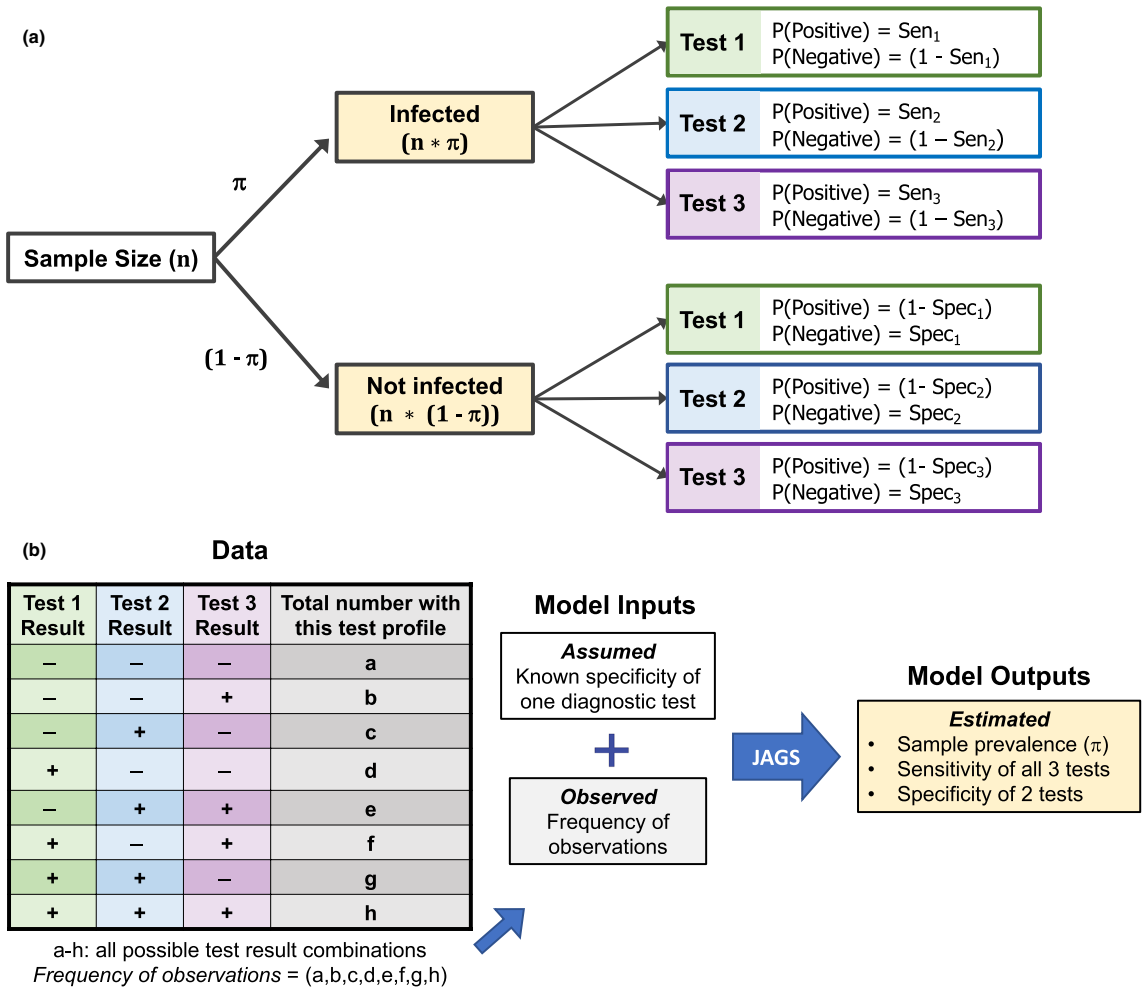


FIGURE 2 Possible infection categories and test results for a sample population (a), and the workflow for assessment in Bayesian latent class analysis (b). (a) The relationship between sample size, prevalence, and the probability of a positive or negative result for three different diagnostic tests. For infected individuals, the probability of a given test result is proportional to the sensitivity (Se) of that test (top right). For uninfected individuals, the probability of a given test result is proportional to the specificity (Sp) of that test (bottom right). (b) Workflow diagram for Bayesian latent class analysis, taking results from data along with the fixed specificity of one test to obtain posterior probability estimates for all unknown (latent) parameters

where π denotes prevalence, Se_1 denotes the sensitivity of test 1, Sp_1 denotes the specificity of test 1, and so on. The first term in this expression represents the probability of being infected and having a false-negative result for all three tests, while the second term represents the probability of being uninfected and having a true-negative result for all three tests. Similar logic can be used to find the probability of each diagnostic profile (b-h, Figure S2), and the observed distribution of diagnostic profiles can be modeled by a multinomial likelihood, with probabilities for each class given by $\{p(a),p(b),...p(h)\}$ (Rindskopf & Rindskopf, 1986).

2.4 | Parameter estimation

We estimated parameters in a Bayesian framework using Markov chain Monte Carlo (MCMC). We ran three chains for 10,000 iterations each, with the first 5,000 steps discarded as burn-in. Uninformative priors (uniform distributions on [0,1]) were assumed for the prevalence, sensitivity of tests 1–3, and specificity of tests 1 and 2 (Figure S2). The fixed specificity for test 3 (one value from points A-E; Figure 1b) and the frequency of each test profile type (frequency of observations) were used as model inputs (Figure 2b). We modified Bayesian inference code (WinBUGS (Lunn, Thomas, Best,

& Spiegelhalter, 2000)) from a previous study (Limmathurotsakul et al., 2012), and JAGS (Plummer, 2003) model estimation was performed using the package R2jags (R2jags, Su, & Yajima, 2015) in R (R Foundation for Statistical Computing, 2016; version 3.3.2). We checked that the Gelman and Rubin statistic was < 1.1 to verify convergence of all MCMC chains (Gelman, Carlin, Stern, & Rubin, 2003) and reported the median and marginal composite 95% credible interval (CrI) for all estimated parameters. Prevalence estimates and 95% CrI were computed for all hypothetical test sets (left panels of Figure 3, S3 and S4), and residuals for all estimated parameters were computed for the fixed test assuming the sensitivity and specificity combination at arc point C (Se_3 = 0.95/Sp_3 = 0.95; right panels of Figure 3, S3 and S4). We ran additional simulations using informed priors to determine how an investigator's prior knowledge or suspicion of low, medium or high prevalence levels in a system would affect the estimates of prevalence obtained from BLCA (π ~ beta(2,9), beta(9,9) and beta(9,2) for low, medium, and high prevalence, respectively; Figure S5). Results were compared to the original estimates obtained using uninformative priors (Figure S6).

2.5 | Wildlife case study

To assess BLCA in a wildlife dataset, we analyzed results of three different tests used to determine *Leptospira* infection status in California sea lions admitted to The Marine Mammal Center (TMMC). TMMC is a marine mammal rehabilitation center that maintains a detailed database of health and medical diagnostic records for individual marine mammals stranding along the California coast. Clinical *Leptospira* infections are diagnosed by clinicians at TMMC using the following diagnostic criteria: high serum MAT antibody titers (>1:3,200) against serovar Pomona, *Leptospira* DNA present in

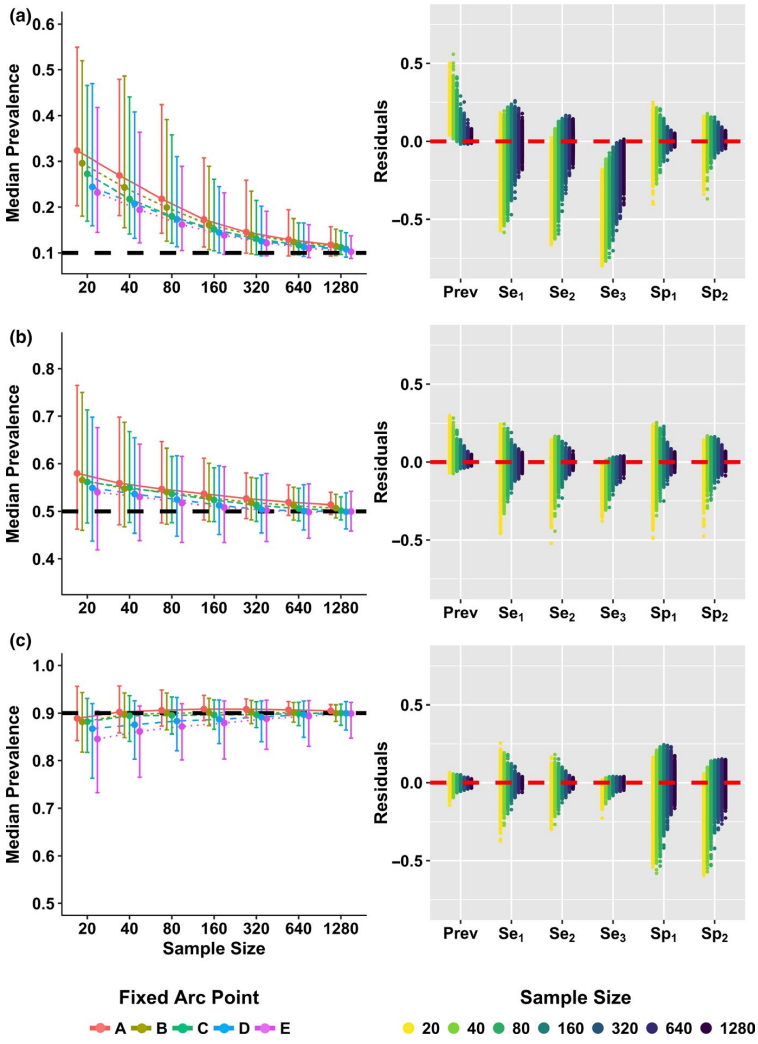


FIGURE 3 Parameter estimates at three true prevalence levels (10%, 50%, and 90%). *Left:* Median prevalence estimates and 95% credible intervals (CrI) are shown for points A-E at a true prevalence of 10% (a), 50% (b), and 90% (c), with true prevalence shown as dashed black lines (*y*-axes scaled equally). *Right:* Residuals for all parameter estimates (prevalence, sensitivities for tests 1–3, specificities for tests 1 and 2) using simulated samples (n = 1,000) generated with fixed arc point C (test 3 fixed Se = 95% and Sp = 95%), with zero shown as dashed red line

TABLE 1 BLCA median parameter estimates and 95% CrIs obtained from three *Leptospira* diagnostic test results in California sea lions (left)

	BLCA Estimates from California sea lion data	Values selected for CSL simulated data	CSL simulated data
Prevalence (π)	20.2% (15.6–25.5%)	20%	20.6% (15.8–26.2%)
Sensitivity – MAT (Se_1)	64.4% (52.0–78.1%)	65%	64.0% (50.5–76.6%)
Sensitivity – SC (Se_2)	61.1% (48.2–74.3%)	61%	60.0% (46.7–72.7%)
Sensitivity – PCR (Se_3)	96.0% (86.4–99.9%)	96%	93.9% (90.3–99.6%)
Specificity – MAT (Sp_1)	98.1% (95.8–99.6%)	98%	98.0% (95.3–99.7%)
Specificity – SC (Sp_2)	93.2% (89.6–96.3%)	93%	92.9% (89.0–99.7%)
Specificity – PCR (Sp_3)	NA (fixed at 97.2%)	Fixed at 97.2%	NA (fixed at 97.2%)

Note: These estimated values were chosen as set values for a CSL data simulation (middle). BLCA parameter estimates were then calculated from this simulated CSL data to see how well the model performed (right).

TABLE 2 Comparison of prevalence estimates from BLCA versus a single test (sample size, n = 320)

	BLCA Fixed Point A	Only Point A	BLCA Fixed Point C	Only Point C	BLCA Fixed Point E	Only Point E
Low (10%) Prevalence	15.0% (10.0-26.0%)	28.0% (23.0-33.0%)	13.1% (9.6-21.5%)	13.8% (11.6-16.6%)	12.0% (0.09-0.19)	8.0% (6.6-9.3%)
Moderate (50%) Prevalence	53.0% (48.5-58.0%)	60.0% (56.8-63.1%)	51.3% (47.9-57.0%)	50.0% (47.5-52.2%)	50.2% (43.7-57.9%)	40.0% (36.5-43.1%)
High (90%) Prevalence	91.0% (89.0-93.0%)	92.0% (91.0-93.0%)	89.7% (86.9-92.4%)	85.9% (83.4-88.1%)	90.0% (84.0-92.5%)	72.0% (68.0-76.0%)

Both BLCA & single test estimates within 5% of true prevalence

Only BLCA estimate within 5% of true prevalence

Note: Both BLCA and single test estimates within 5% of true prevalence. Only BLCA estimate within 5% of true prevalence. The BLCA estimates were obtained using the original test 1 (Se_1 = 70%/Sp_1 = 70%) and test 2 (Se_2 = 80%/Sp_2 = 80%) settings, along with point A (left; Se_3 = 100%/Sp_3 = 80%), point C (middle; Se_3 = 95%/Sp_3 = 95%), or point E (right; Se_3 = 80%/Sp_3 = 100%). Single test estimates and 95% CI were obtained using 1,000 Bernoulli trials weighted by the test Se/Sp for test A, C, or E alone. Scenarios where both BLCA and single test estimates were within 5% of the true value are shown in yellow, while scenarios where BLCA alone was within 5% of the true prevalence are shown in green.

urine or kidney samples (tested via PCR; Wu et al., 2014), or serum chemistry markers indicative of kidney dysfunction (BUN > 100 mg/dl, creatinine > 2 mg/dl, sodium > 155 meq/L and phosphorus > calcium; Colagross-Schouten, Mazet, Gulland, Miller, & Hietala, 2002; Greig et al., 2005). In this system, we judged that conditional independence among tests was a reasonable assumption, due to the different biological systems targeted by these three diagnostic tests (humoral immune response, presence of pathogen DNA in the urinary tract, and measures of renal function, respectively). To minimize the effects of clinical treatment on test outcomes, we selected only California sea lions at TMMC that had test results for all three *Leptospira* diagnostics from samples collected within one week of admission (n = 290; years: 2006–2016). We summed the total number of animals with each test result profile (the frequency of observations) and fixed the specificity of test 3 (PCR) to 97.2% based on a recent estimate for *Leptospira* in humans (Limmathurotsakul et al., 2012). While the PCR method utilized here was previously reported with 100% analytic specificity in CSL urine or kidney tissue (Wu et al., 2014), we chose this slightly more conservative specificity level to reflect the possibility that sample contamination could lead

to rare false positives. Parameter estimation was conducted as described above using R2jags, yielding median estimates with 95% CrIs for all unknown parameters.

To test model performance under these estimated real-world conditions, we simulated CSL data (n = 300) using our best parameter estimates as known parameter values ("Values Selected for CSL Simulated Data" in Table 1), then used BLCA on the simulated CSL data to see how accurate model estimates were across 1,000 simulations (Table 1). To assess when BLCA prevalence estimates would be preferable to those obtained using the single best diagnostic test, we compared BLCA estimates from our initial hypothetical test set (Figure 1b) to results generated solely from the single best test (points A, C, and E), which were simulated by Bernoulli trials as described above (Tables 2 and S2).

3 | RESULTS

3.1 | Simulation study

For all simulated scenarios (i.e., all prevalence levels and all hypothetical 3-test combinations), BLCA prevalence estimates converged on the correct value as the sample size grew (Figure 3). There was some directional bias in prevalence estimates, particularly at low sample sizes, that varied depending on the true infection prevalence. Prevalence of infection was consistently overestimated when infections were rare (true prevalence = 10%) and to a lesser degree when infections were moderately common (true prevalence = 50%). At these prevalence levels, when we varied the specificity of fixed test 3 according to arc points A-E (Figure 1b), tests with higher specificity returned more accurate estimates at lower sample sizes, although credible intervals across these tests largely overlapped (Figure 3, S3 and S4; Table S1). These patterns were reversed when infections were common (true prevalence = 90%), with prevalence being slightly underestimated and higher sensitivity tests returning more accurate estimates at lower sample sizes, although again credible intervals across these tests largely overlapped (Figure 3, S3 and S4; Table S1).

When infections were rare (true prevalence = 10%), the 95% CrIs for prevalence did not contain the true value until sample size was relatively large ($n \geq 160$; 95% CrIs for points D & E). At the largest sample sizes ($n > 320$) the true value was contained within the 95% CrIs for all points, and median prevalence estimates were within 3% of the true value (in absolute terms). When true prevalence was moderate (50%), the true value was contained in the 95% CrIs at all sample sizes, and median prevalence estimates were within 8% of the true value at all sample sizes and within 2% at the highest sample sizes (n = 640 & n = 1,280; Table S1B. In contrast, at higher true prevalence (90%) where prevalence was underestimated at lower sample sizes, the 95% CrIs always contained the true value and prevalence estimates converged quickly to the true prevalence value across all hypothetical test sets (Figure 3c, S3C and S4C; Table S1C).

As with the prevalence estimate, the BLCA estimates of the sensitivity and specificity of each test became more precise and accurate as sample sizes increased (right panels of Figure 3, S2 and S3). However, there were directional biases in these estimates, which exhibited more complex structure than the biases of prevalence estimates. Test sensitivity tended to be underestimated when true prevalence was low, while specificity was underestimated at high prevalences (Figure 3, S2 and S3). When infections were rare (true prevalence = 10%), specificity estimates were more accurate and precise across all sample sizes than sensitivity estimates, while sensitivity estimates were more accurate and precise than specificity estimates when infections were common (true prevalence = 90%; Figure 3, S2 and S3). The residuals of both sensitivity and specificity estimates were generally symmetric, indicating little bias, when infection level was moderate (true prevalence = 50%; Figure 3, S2 and S3).

Considering the potential trade-off between sensitivity and specificity of a given test (i.e., from tuning the threshold value used to classify a result as positive; Figure 1), we found that the optimal parameters of the best test depend on infection prevalence. When true prevalence is low (10%) or moderate (50%), a fixed specificity of 1.0 of the best test (Point E) yields the most accurate estimate of prevalence (Figure 3a,b). However, when prevalence is high (90%), a fixed sensitivity of 1.0 of the best test (Point A) is preferable (Figure 3c). The influence of this trade-off is greatest at low prevalence (10%) and weakest at high prevalence (90%) where any Point (A-E) gives a reasonable prevalence estimate (Table 2).

These broad patterns remained the same regardless of the hypothetical test set used. That is, as the parameters of the two lower-quality tests change (Figure S1), the patterns of prevalence, sensitivity, and specificity estimation did not vary qualitatively (Figures S3 and S4). However, the quantitative results were noticeably worse (i.e., larger residuals and larger sample sizes needed for accuracy) when these two tests had lower sensitivity and specificity (Figure S4). When we used informative priors in the BLCA to represent investigator knowledge of the prevalence level, estimates of prevalence improved if the prior was close to the true prevalence level, but worsened if the prior was not close to the true prevalence value (Figure S6). Adjusting the prevalence prior did not qualitatively alter the estimates of other parameters.

3.2 | Wildlife case study

Results from the sea lion case study were concordant with our analyses of the broader simulated data. Although disease prevalence was low in the sea lion system, our sample size was well within the range at which BLCA could produce accurate prevalence estimates using simulated CSL data. The estimated prevalence of clinical *Leptospira* infections in this sample of California sea lions was 20.2% (95% CrI, 15.6%–25.5%; Table 1). Estimates of PCR, MAT and serum chemistry relative test accuracy were broadly consistent with expert knowledge (Table 1). Marine mammal veterinarians consider PCR the best diagnostic test for leptospirosis in sea lions, whereas MAT and serum chemistry are known to be less sensitive and typically used as second-line tests when urine samples cannot be obtained for PCR.

We also simulated data to verify that BLCA was accurate when using parameters and sample sizes consistent with the best estimates returned by real CSL data. Using these simulated data, the median BLCA prevalence estimate was 20.6% (95% CrI, 15.8%–26.2%; Table 1; Figure S7B), and differed negligibly from the true input prevalence (20%). Sensitivity and specificity values were slightly underestimated, but always within 2.1% of the true value (Table 1). Although this test of simulated data returned very accurate estimates, the direction and magnitude of observed errors were consistent with the error structures reported above for data simulated using other parameters.

3.3 | Comparing BLCA to results of a single diagnostic test

The BLCA prevalence estimate for the California sea lion sample (20.2%) was very similar to the crude estimate obtained from PCR alone (62/290 positive; 21.4%). This prompted us to consider the marginal value of BLCA and whether it was worth the additional effort. In particular, we explored the circumstances under which the 3-test BLCA prevalence estimates would improve upon results from a single best test, exploring the influence of the trade-off between sensitivity and specificity of the best test by considering points A, C, and E from our simulation analysis. At a sample size similar to our CSL case study (n = 320), prevalence estimates obtained using BLCA and using the single best test (points A, C, or E alone) were comparable in most cases, but in several scenarios BLCA was clearly superior (Table 2). A single test at point A (Se_3 = 0.8/Sp_3 = 1) alone overestimated prevalence at low and mid true prevalence levels, while a single test at point E (Se_3 = 1/Sp_3 = 0.8) alone underestimated prevalence at mid and high true prevalence levels; in contrast, BLCA was accurate in both these scenarios (Table 2). Thus, the BLCA method can yield stabilizing estimates that are more robust to fluctuations in prevalence than estimates produced by any single test with unbalanced sensitivity and specificity (Tables 2 and S2). These stabilizing effects of BLCA would be particularly useful in a system with cyclical outbreaks.

When comparing BLCA to single test estimates across all sample sizes, these broad patterns held for larger sample sizes, but differed at lower sample sizes (n < 320; Table S2). Regardless of sample size, a balanced high-quality test with very high sensitivity and specificity (test C) is comparable to BLCA. However, at high sample sizes BLCA converges on the true value at all prevalence levels whereas test C alone converges on over- or under-estimates of prevalence. At 50% prevalence, test C converges on the true estimate, but this is due to canceling of symmetric errors from its identical sensitivity and specificity values.

Bayesian latent class analysis also usually outperformed estimates from a single test when test specificity or sensitivity was low (closer to points A or E; Figure 1b), but neither method worked well when prevalence, test specificity, and sample size were all low. Thus, in circumstances when sample size and disease prevalence are both low, we recommend choosing a diagnostic test threshold that optimizes test specificity, as this can improve the performance of both BLCA and of the single-test method (Table S2).

4 | DISCUSSION

Estimating infection prevalence is challenging in wildlife disease systems, where researchers are often confronted with limited sample sizes and imperfect diagnostic tests that lack species-specific validation. Here, we have explored the utility of Bayesian latent class analysis (BLCA) as a technique to improve estimates of prevalence and of diagnostic test sensitivity and specificity. We have assumed conditional independence among test results, which is reasonable for the biological system we examined due to differences in the biological systems targeted by our diagnostic assays and our lack of longitudinal sampling (Kostoulas et al., 2017; Wang & Hanson, 2019), but in situations where this is not the case the conditional dependence structure should be considered (Albert & Dodd, 2004; Dendukuri & Joseph, 2001; Hadgu & Qu, 1998; Jones et al., 2010; Pepe & Janes, 2006; Qu et al., 1996). Using simulated data and a case study to explore the utility of BLCA, we demonstrate that the accuracy of prevalence estimates depends on multiple factors: the sample size being tested, the true prevalence in the study system and the sensitivity/specificity of the diagnostic tests being used. We compare BLCA prevalence estimates to those from a single test, demonstrating the stabilizing effects of the BLCA method under different sample sizes and prevalences. In addition, recognizing that many diagnostic tests have an intrinsic trade-off between sensitivity and specificity (which can be tuned by altering the threshold value used to define a positive test result), we show how the accuracy of prevalence estimates can be optimized depending on the epidemiological context.

The precision and accuracy of parameter estimates increased with sample size across all simulations, providing accurate estimates at large sample sizes regardless of the true infection prevalence (Figure 1, S2, S3, S7). The use of informed priors has the potential to further improve prevalence estimates, highlighting the potential for this Bayesian framework to incorporate expert knowledge from the field. However, in the absence of accurate prior information the use of uninformed priors provides more stable prevalence estimates (Figure S6). We observed directional biases in the prevalence, sensitivity, and specificity estimates depending on whether infections are common (high prevalence) or rare (low prevalence; Greiner & Gardner, 2000). For example, when sample size is relatively low, an overrepresentation of false positives can elevate prevalence estimates when diseases are rare. Conversely, an overrepresentation of false negatives can bias prevalence estimates downward when diseases are common.

Our work demonstrates the potential to improve the accuracy of prevalence estimates by altering the threshold for positivity for the highest quality test (Figure 1a). If results from the best test (the test with the fixed specificity provided to the BLCA model) are quantitative, choosing a threshold that maximizes specificity will improve

prevalence estimate accuracy when infections are rare, while maximizing sensitivity will improve estimates when infections are common (Figure 3, S2 and S3; Table S2). This choice follows naturally, because higher specificity allows you to detect more true negatives, which are prevalent when infections are rare. In contrast, higher sensitivity allows you to detect more true positives, which are prevalent when infections are common. In addition, this logic can guide the choice of single tests (or the choice of a threshold for a single test) to use for estimation of prevalence without the need for advanced statistical analysis: our work shows that maximizing test sensitivity for common diseases, or test specificity for rare diseases, can produce single test estimates of comparable accuracy to BLCA.

Analyzing our wildlife case study of *Leptospira interrogans* in California sea lions, we report new estimates for the sensitivity and specificity of key diagnostic tests in this system to explore the statistical power of BLCA for a given sample size. The samples used in this study span a ten-year period across a range of different epidemiologic conditions, so here our prevalence estimates reflect the prevalence in the sample of stranded animals rather than the prevalence in the wild population at any point in time. At a titer cutoff 1:3,200, our estimates for the sensitivity and specificity of MAT were 64.4% (95% CrI: 52%-78.1%) and 98.1% (95% CrI: 95.8%-99.6%), respectively, which differ from previous CSL estimates using this titer cutoff obtained from known positive and negative animals (Se = 100% and Sp = 100%; Colagross-Schouten et al., 2002). These previous estimates were likely idealized due to small samples and the study design (Greiner & Gardner, 2000), as the negative controls were born in captivity with no possibility of residual titers from a previous exposure, and the positive animals were selected based on clear clinical signs and renal lesions indicating leptospirosis. Diagnosis in wild animals is likely to be complicated by residual titers from previous exposures, or by chronic infections that are no longer associated with a high titer (Buhnerkempe et al., 2017). Due to these and other complicating factors, sensitivity and specificity are unlikely to be perfect in stranded wild animals.

This contrast highlights the influence of the underlying study population and the importance of considering system-specific characteristics and ecological context when utilizing BLCA. Test sensitivity and specificity estimates likely vary with underlying prevalence and sample size due to the probability of sampling individuals that are truly infected or truly uninfected, which in turn modulates the ratio of false positives to false negatives in the data. For example, at 90% true prevalence, most individuals will be true positives, so testing regimes will have the potential to pick up more true positives/false negatives and fewer true negatives/false positives, leading to a higher estimated sensitivity and lower estimated specificity.

Broadly, we demonstrate that BLCA works well for estimating prevalence and test accuracy, but some caution is warranted because its performance does not always beat that of the single best available test. In particular, there are scenarios with low sample size and low-to-moderate prevalence where a single test with high specificity can yield more accurate prevalence estimates than BLCA. A rule of thumb, apparent in Table S2, is that this can happen when the expected number of infected individuals in the sample is ≤10. When the best test has lower specificity (e.g., Test A in Table S2), neither approach worked well if the expected number of infections is ≤20. In all other situations, prevalence estimates from BLCA are comparable to or better than estimates from a single diagnostic test, and this performance advantage increases as the highest quality diagnostic test decreases in sensitivity or specificity (i.e., moving toward points A or E). Furthermore, prevalence estimates made using BLCA will be more robust to changes in prevalence across cyclical epidemics than estimates made using a single test. Our data simulations provide quantitative insight into the relative performance of these approaches, to help researchers assess whether the additional effort of BLCA is worthwhile. In many circumstances, the BLCA method provides more accurate estimates than researchers would otherwise be able to obtain, making it a worthwhile tool that addresses many challenges faced by disease ecologists.

ACKNOWLEDGMENTS

This work was supported by the U.S. National Science Foundation (OCE-1335657 and DEB-1557022), the Strategic Environmental Research and Development Program (SERDP, RC-2635) of the U.S. Department of Defense, the John H. Prescott Marine Mammal Rescue Assistance Grant Program, and the Hellman Family Foundation. SKH was supported by The UCLA Sustainable L.A. program. MGB, KCP, and JOL-S were supported by the Research and Policy for Infectious Disease Dynamics (RAPIDD) program of the Science and Technology Directory, Department of Homeland Security, and Fogarty International Center, National Institutes of Health. KG was supported by the National Institutes of Health (F31AI134017, T32-GM008185). The content of the information does not necessarily reflect the position or the policy of the U.S. government, and no official endorsement should be inferred. Special thanks to the entire TMMC Veterinary Science team and volunteers for generating the wildlife data for this study.

CONFLICT OF INTEREST

The authors declare no competing interests.

AUTHOR CONTRIBUTION

Sarah K Helman: Conceptualization (lead); Data curation (lead); Formal analysis (lead); Investigation (lead); Methodology (lead); Project administration (lead); Writing-original draft (lead); Writing-review & editing (lead). Riley O. Mummah: Data curation (supporting); Formal analysis (supporting); Investigation (supporting); Methodology (supporting); Project administration (supporting); Writing-review & editing (supporting). **Katelyn M Gostic:** Data curation (supporting); Formal analysis (supporting); Investigation (supporting); Methodology (supporting); Writing-review & editing (supporting). **Michael Buhnerkempe:** Conceptualization (equal); Data curation (supporting); Formal

analysis (supporting); Investigation (supporting); Methodology (supporting); Writing-review & editing (supporting). **Katherine C Prager:** Conceptualization (supporting); Data curation (supporting); Formal analysis (supporting); Investigation (supporting); Methodology (supporting); Writing-review & editing (supporting). **James O Lloyd-Smith:** Conceptualization (equal); Data curation (supporting); Formal analysis (supporting); Investigation (supporting); Methodology (supporting); Project administration (supporting); Writing-review & editing (supporting).

DATA AVAILABILITY STATEMENT

All CSL data and the code used to generate additional datasets during this study are included in this published article (and its supplementary information files).

ORCID

Sarah K. Helman https://orcid.org/0000-0002-7292-1533
Riley O. Mummah https://orcid.org/0000-0002-4542-3483
Katelyn M. Gostic https://orcid.org/0000-0002-9369-6371
Michael G. Buhnerkempe https://orcid.org/0000-0002-5996-8927
Katherine C. Prager https://orcid.org/0000-0003-0669-0754
James O. Lloyd-Smith https://orcid.org/0000-0001-7941-502X

REFERENCES

Alberg, A. J., Park, J. W., Hager, B. W., Brock, M. V., & Diener-West, M. (2004). The use of "overall accuracy" to evaluate the validity of screening or diagnostic tests. *Journal of General Internal Medicine*, *19*(5 Pt 1), 460–465. https://doi.org/10.1111/j.1525-1497.2004.30091.x

Albert, P. S., & Dodd, L. E. (2004). A cautionary note on the robustness of latent class models for estimating diagnostic error without a gold standard. *Biometrics*, *60*(2), 427–435. https://doi.org/10.1111/j.0006-341X.2004.00187.x

Almberg, E. S., Cross, P. C., Dobson, A. P., Smith, D. W., & Hudson, P. J. (2012). Parasite invasion following host reintroduction: A case study of Yellowstone's wolves. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences*, *367*(1604), 2840–2851. https://doi.org/10.1098/rstb.2011.0369

Basso, W., Hartnack, S., Pardini, L., Maksimov, P., Koudela, B., Venturini, M. C., … Deplazes, P. (2013). Assessment of diagnostic accuracy of a commercial ELISA for the detection of Toxoplasma gondii infection in pigs compared with IFAT, TgSAG1-ELISA and Western blot, using a Bayesian latent class approach. *International Journal for Parasitology*, *43*(7), 565–570. https://doi.org/10.1016/j.ijpara.2013.02.003

Boelaert, M., Aoun, K., Liinev, J., Goetghebeur, E., & Van der Stuyft, P. (1999). The potential of latent class analysis in diagnostic test validation for canine Leishmania infantum infection. *Epidemiology and Infection*, *123*(3), 499–506.

Bronsvoort, B. M. D. C., Parida, S., Handel, I., McFarland, S., Fleming, L., Hamblin, P., & Kock, R. (2008). Serological survey for foot-and-mouth disease virus in wildlife in Eastern Africa and estimation of test parameters of a nonstructural protein enzyme-linked immunosorbent assay for buffalo. *Clinical and Vaccine Immunology*, *15*(6), 1003. https://doi.org/10.1128/CVI.00409-07

Buhnerkempe, M. G., Prager, K. C., Strelioff, C. C., Greig, D. J., Laake, J. L., Melin, S. R., … Lloyd-Smith, J. O. (2017). Detecting signals of chronic shedding to explain pathogen persistence: Leptospira interrogans in California sea lions. *Journal of Animal Ecology*, *86*(3), 460–472. https://doi.org/10.1111/1365-2656.12656

Buhnerkempe, M. G., Roberts, M. G., Dobson, A. P., Heesterbeek, H., Hudson, P. J., & Lloyd-Smith, J. O. (2015). Eight challenges in modelling disease ecology in multi-host, multi-agent systems. *Epidemics*, *10*, 26–30. https://doi.org/10.1016/j.epidem.2014.10.001

Cameron, C. E., Zuerner, R. L., Raverty, S., Colegrove, K. M., Norman, S. A., Lambourn, D. … Gulland, F. M. (2008). Detection of pathogenic Leptospira bacteria in pinniped populations via PCR and identification of a source of transmission for zoonotic leptospirosis in the marine environment. *Journal of Clinical Microbiology*, *46*(5), 1728–1733. https://doi.org/10.1128/JCM.02022-07

Colagross-Schouten, A. M., Mazet, J. A. K., Gulland, F. M. D., Miller, M. A., & Hietala, S. (2002). Diagnosis and seroprevalence of leptospirosis in California sea lions from coastal California. *Journal of Wildlife Diseases*, *38*(1), 7–17.

Dendukuri, N., & Joseph, L. (2001). Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics*, *57*(1), 158–167. https://doi.org/10.1111/j.0006-341X.2001.00158.x

Florkowski, C. M. (2008). Sensitivity, specificity, receiver-operating characteristic (ROC) curves and likelihood ratios: Communicating the performance of diagnostic tests. *The Clinical Biochemist Reviews*, *29*(Suppl 1), S83–S87.

Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2003). *Bayesian data analysis* (2 nd ed.). Boca Raton, FL: Chapman & Hall/CRC Texts in Statistical Science.

Gonçalves, L., Subtil, A., de Oliveira, M. R., & do Rosário, V., Lee, P.-W., & Shaio, M.-F. (2012). Bayesian latent class models in malaria diagnosis. *PLoS One*, *7*(7), e40633. https://doi.org/10.1371/journal.pone.0040633

Greig, D. J., Gulland, F. M. D., & Kreuder, C. (2005). A decade of live California sea lion (*Zalophus californianus*) strandings along the central California coast: Causes and trends, 1991–2000. *Aquatic Mammals*, *31*(1), 11–22. https://doi.org/10.1578/AM.31.1.2005.11

Greiner, M., & Gardner, I. A. (2000). Epidemiologic issues in the validation of veterinary diagnostic tests. *Preventive Veterinary Medicine*, *45*(1), 3–22. https://doi.org/10.1016/S0167-5877(00)00114-8

Hadgu, A., & Qu, Y. (1998). A biomedical application of latent class models with random effects. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *47*(4), 603–616. https://doi.org/10.1111/1467-9876.00131

Hartnack, S., Budke, C. M., Craig, P. S., Jiamin, Q., Boufana, B., Campos-Ponce, M., & Torgerson, P. R. (2013). Latent-class methods to evaluate diagnostics tests for Echinococcus infections in dogs. *PLoS Neglected Tropical Diseases*, *7*(2), e2068. https://doi.org/10.1371/journal.pntd.0002068

Haydon, D. T., Cleaveland, S., Taylor, L. H., & Laurenson, K. M. (2002). Identifying reservoirs of infection: A conceptual and practical challenge. *Emerging Infectious Diseases*, *8*(12), 1468–1473.

Jones, G., Johnson, W. O., Hanson, T. E., & Christensen, R. (2010). Identifiability of models for multiple diagnostic testing in the absence of a gold standard. *Biometrics*, *66*(3), 855–863. https://doi.org/10.1111/j.1541-0420.2009.01330.x

Kostoulas, P., Nielsen, S. S., Branscum, A. J., Johnson, W. O., Dendukuri, N., Dhand, N. K., … Gardner, I. A. (2017). STARD-BLCM: Standards for the reporting of diagnostic accuracy studies that use bayesian latent class models. *Preventive Veterinary Medicine*, *138*, 37–47. https://doi.org/10.1016/j.prevetmed.2017.01.006

Limmathurotsakul, D., Turner, E. L., Wuthiekanun, V., Thaipadungpanit, J., Suputtamongkol, Y., Chierakul, W., … Peacock, S. J. (2012). Fool's gold: Why imperfect reference tests are undermining the evaluation of novel diagnostics: A reevaluation of 5 diagnostic tests for leptospirosis. *Clinical Infectious Diseases*, *55*(3), 322–331. https://doi.org/10.1093/cid/cis403

Lloyd-Smith, J. O., Greig, D. J., Hietala, S., Ghneim, G. S., Palmer, L., St Leger, J., … Gulland, F. M. D.. (2007). Cyclical changes in seroprevalence of leptospirosis in California sea lions: Endemic and epidemic disease in one host species? *BMC Infectious Diseases*, *7*, 125. https://doi.org/10.1186/1471-2334-7-125

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS-a Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, *10*(4), 325–337.

Mathevon, Y., Foucras, G., Falguières, R., & Corbiere, F. (2017). Estimation of the sensitivity and specificity of two serum ELISAs and one fecal qPCR for diagnosis of paratuberculosis in sub-clinically infected young-adult French sheep using latent class Bayesian modeling. *BMC Veterinary Research*, *13*(1), 230–230. https://doi.org/10.1186/s12917-017-1145-x

Maxim, L. D., Niebo, R., & Utell, M. J. (2014). Screening tests: A review with examples. *Inhalation Toxicology*, *26*(13), 811–828. https://doi.org/10.3109/08958378.2014.955932

Moreno-Torres, K., Wolfe, B., Saville, W., & Garabed, R. (2016). Estimating Neospora caninum prevalence in wildlife populations using Bayesian inference. *Ecology and Evolution*, *6*(7), 2216–2225.

Muma, J. B., Toft, N., Oloya, J., Lund, A., Nielsen, K., Samui, K., & Skjerve, E. (2007). Evaluation of three serological tests for brucellosis in naturally infected cattle using latent class analysis. *Veterinary Microbiology*, *125*(1–2), 187–192. https://doi.org/10.1016/j.vetmic.2007.05.012

Nielsen, L. R., Toft, N., & Ersbøll, A. K. (2004). Evaluation of an indirect serum ELISA and a bacteriological faecal culture test for diagnosis of Salmonella serotype Dublin in cattle using latent class models. *Journal of Applied Microbiology*, *96*(2), 311–319. https://doi.org/10.1046/j.1365-2672.2004.02151.x

Pan-ngum, W., Blacksell, S. D., Lubell, Y., Pukrittayakamee, S., Bailey, M. S., de Silva, H. J., … Limmathurotsakul, D. (2013). Estimating the true accuracy of diagnostic tests for dengue infection using bayesian latent class models. *PLoS One*, *8*(1), e50765. https://doi.org/10.1371/journal.pone.0050765

Pepe, M. S., & Janes, H. (2006). Insights into latent class analysis of diagnostic test performance. *Biostatistics*, *8*(2), 474–484.

Plummer, M. (2003). *JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling*, Vol. *124*. Vienna, Austria: 3rd International Workshop on Distributed Statistical Computing (DSC 2003).

Prager, K. C., Greig, D. J., Alt, D. P., Galloway, R. L., Hornsby, R. L., Palmer, L. J., … Lloyd-Smith, J. O. (2013). Asymptomatic and chronic carriage of Leptospira interrogans serovar Pomona in California sea lions (*Zalophus californianus*). *Veterinary Microbiology*, *164*(1–2), 177–183. https://doi.org/10.1016/j.vetmic.2013.01.032

Qu, Y., Tan, M., & Kutner, M. H. (1996). Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics*, *52*(3), 797–810. https://doi.org/10.2307/2533043

R Core Team (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org

R2jags, Su, Y-S, & Yajima, M (2015). *R2jags: Using R to Run 'JAGS'*. R package version 0.5–7. Available from https://CRAN.R-project.org/package=R2jags

Rindskopf, D., & Rindskopf, W. (1986). The value of latent class analysis in medical diagnosis. *Statistics in Medicine*, *5*(1), 21–27. https://doi.org/10.1002/sim.4780050105

Schumacher, S. G., van Smeden, M., Dendukuri, N., Joseph, L., Nicol, M. P., Pai, M., & Zar, H. J. (2016). Diagnostic test accuracy in childhood pulmonary tuberculosis: A Bayesian latent class analysis. *American Journal of Epidemiology.*, *184*(9), 690–700. https://doi.org/10.1093/aje/kww094

Verma-Kumar, S., Abraham, D., Dendukuri, N., Cheeran, J. V., Sukumar, R., & Balaji, K. N. (2012). Serodiagnosis of tuberculosis in Asian elephants (Elephas maximus) in Southern India: A latent class analysis. *PLoS One*, *7*(11), e49548. https://doi.org/10.1371/journal.pone.0049548

Viana, M., Mancy, R., Biek, R., Cleaveland, S., Cross, P. C., Lloyd-Smith, J. O., & Haydon, D. T. (2014). Assembling evidence for identifying reservoirs of infection. *Trends in Ecology & Evolution*, *29*(5), 270–279. https://doi.org/10.1016/j.tree.2014.03.002

Wang, C., & Hanson, T. E. (2019). Estimation of sensitivity and specificity of multiple repeated binary tests without a gold standard. *Statistics in Medicine*, *38*(13), 2381–2390. https://doi.org/10.1002/sim.8114

Wu, Q., Prager, K. C., Goldstein, T., Alt, D. P., Galloway, R. L., Zuerner, R. L., … Schwacke, L. (2014). Development of a real-time PCR for the detection of pathogenic *Leptospira* spp. in California sea lions. *Diseases of Aquatic Organisms*, *110*(3), 165–172. https://doi.org/10.3354/dao02752

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Helman SK, Mummah RO, Gostic KM, Buhnerkempe MG, Prager KC, Lloyd-Smith JO. Estimating prevalence and test accuracy in disease ecology: How Bayesian latent class analysis can boost or bias imperfect test results. *Ecol Evol*. 2020;10:7221–7232. https://doi.org/10.1002/ece3.6448

1.7 Supplement

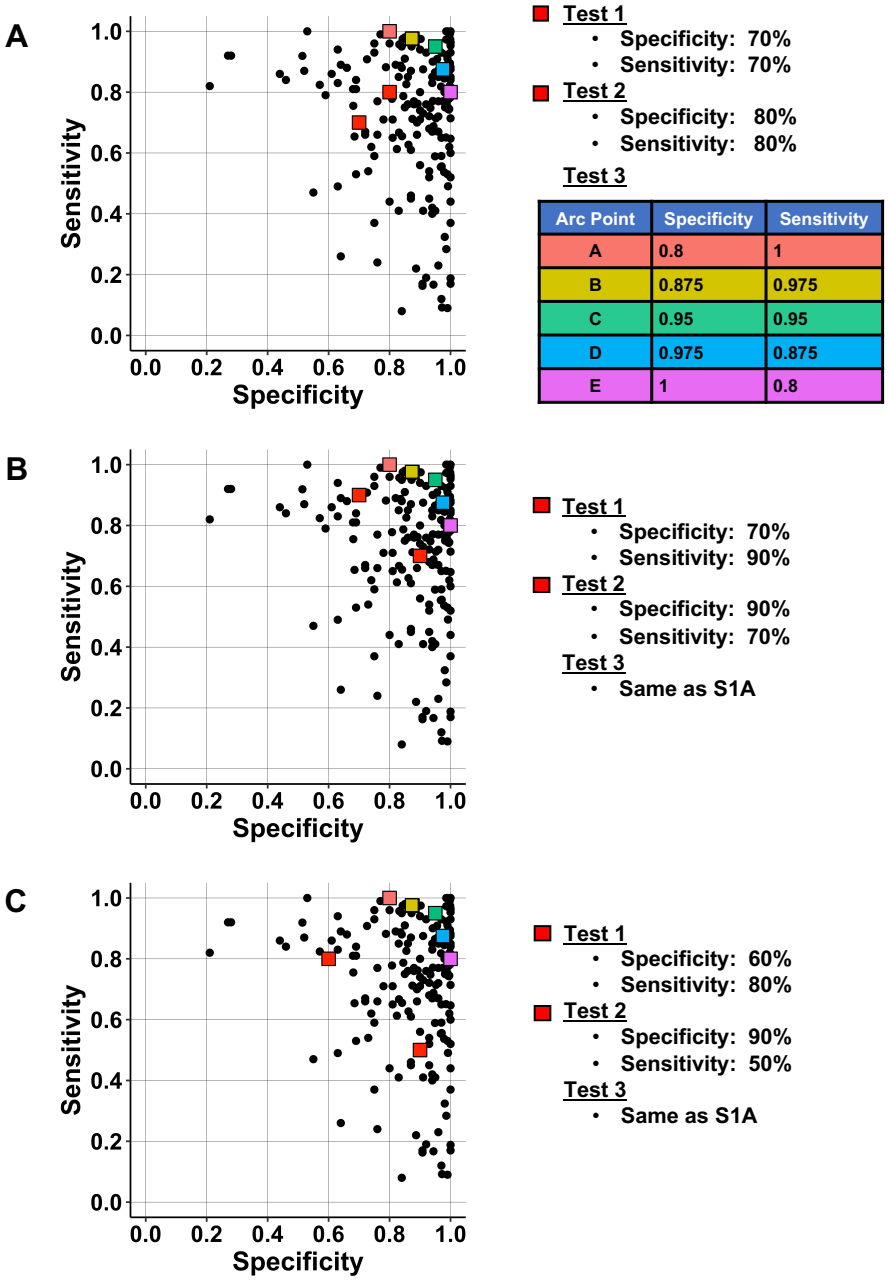


Figure S 1.1: Sensitivity and specificity values for test 1 & test 2 (red points) that were used in conjunction with Test 3 (fixed arc points A-E) for data simulations. (A) Original test 1 and test 2 values, corresponding to the simulation results in Figure 3. (B) Alternate test 1 and test 2 values, corresponding to the simulation results in Figure S3. (C) Alternate test 1 and test 2 values, corresponding to the simulation results in Figure S4.

A. JAGS Model

```
model{
#total<-sum(freqobs[1:8])

#Likelihood
freqobs[1:8]~ dmulti(p[1:8],total)

for (i in 1:8){

p[i]<- prev*(positive[i]) + (1-prev)*(negative[i])

positive[i]  <- (Se[1]) * test1[i] + (1-Se[1]) * (1-test1[i])) *
                (Se[2]) * test2[i] + (1-Se[2]) * (1-test2[i])) *
                (Se[3]) * test3[i] + (1-Se[3]) * (1-test3[i]))

negative[i] <- ((1-Sp[1]) * test1[i] + Sp[1] * (1-test1[i])) *
               ((1-Sp[2]) * test2[i] + Sp[2] * (1-test2[i])) *
               ((1-Sp[3]) * test3[i] + Sp[3] * (1 -test3[i]))}

#Priors
prev~dbeta(1,1)
Se[3]~dbeta(1,1)
Sp[3]<- XX

for (j in 1:2){
s[j]~dbeta(1,1)
x[j]~dbeta(1,1) }

}
```

B. JAGS Model Explanation

freqobs: frequency of observations for each test profile (1:8 refers to profiles a-h)

Sum(freqobs[1:8]): total individuals tested

Likelihood

p[i]: probability of each test profile ($\sum$ p[i]=1)

positive[i]: probability that an individual has test profile [i] given the individual is truly positive

negative[i]: probability that an individual has test profile [i] given the individual is truly negative

Tests: JAGS model input designating test results (0=negative, 1=positive). These test vectors jointly represent all possible test profiles.

test1 <- c(0,0,1,1,0,0,1,1)
test2 <- c(0,1,0,1,0,1,0,1)
test3 <- c(0,0,0,0,1,1,1,1)

Priors

Prev: Prevalence (noninformative)
Se[3]: Sensitivity of test 3 (noninformative)
Sp[3]: Specificity of test 3 **(fixed value=XX)**

Se[1:2]: Sensitivity of test 1 & 2 (noninformative)
Sp[1:2]: Specificity of test 1 & 2 (noninformative)

C. Probability Table

Test 1 Result	Test 2 Result	Test 3 Result	Total number with this test profile	Expected Total Number for Each Test Profile $n\pi$: total true positives $n(1-\pi)$: total true negatives
-	-	-	a	$n\pi * (1-Se_1)(1-Se_2)(1-Se_3) + n(1-\pi) * (Sp_1Sp_2Sp_3)$
-	+	-	b	$n\pi * (1-Se_1)Se_2(1-Se_3) + n(1-\pi) * Sp_1(1-Sp_2)Sp_3$
+	-	-	c	$n\pi * Se_1(1-Se_2)(1-Se_3) + n(1-\pi) * (1-Sp_1)Sp_2Sp_3$
+	+	-	d	$n\pi * Se_1Se_2(1-Se_3) + n(1-\pi) * (1-Sp_1)(1-Sp_2)Sp_3$
-	-	+	e	$n\pi * (1-Se_1)(1-Se_2)Se_3 + n(1-\pi) * Sp_1Sp_2(1-Sp_3)$
-	+	+	f	$n\pi * (1-Se_1)Se_2Se_3 + n(1-\pi) * Sp_1(1-Sp_2)(1-Sp_3)$
+	-	+	g	$n\pi * Se_1(1-Se_2)Se_3 + n(1-\pi) * (1-Sp_1)Sp_2(1-Sp_3)$
+	+	+	h	$n\pi * Se_1Se_2Se_3 + n(1-\pi) * (1-Sp_1)(1-Sp_2)(1-Sp_3)$

Figure S 1.2: JAGS Model and probability functions for each test result profile (a-h).
(A) Conditional independence model used in JAGS, assuming that each individual's test results are independent given true disease status. (B) Explanation of the JAGS model. (C) Probabilities for the number of individuals expected in each test profile category for a given sample size (**n**) and prevalence ($\boldsymbol{\pi}$).

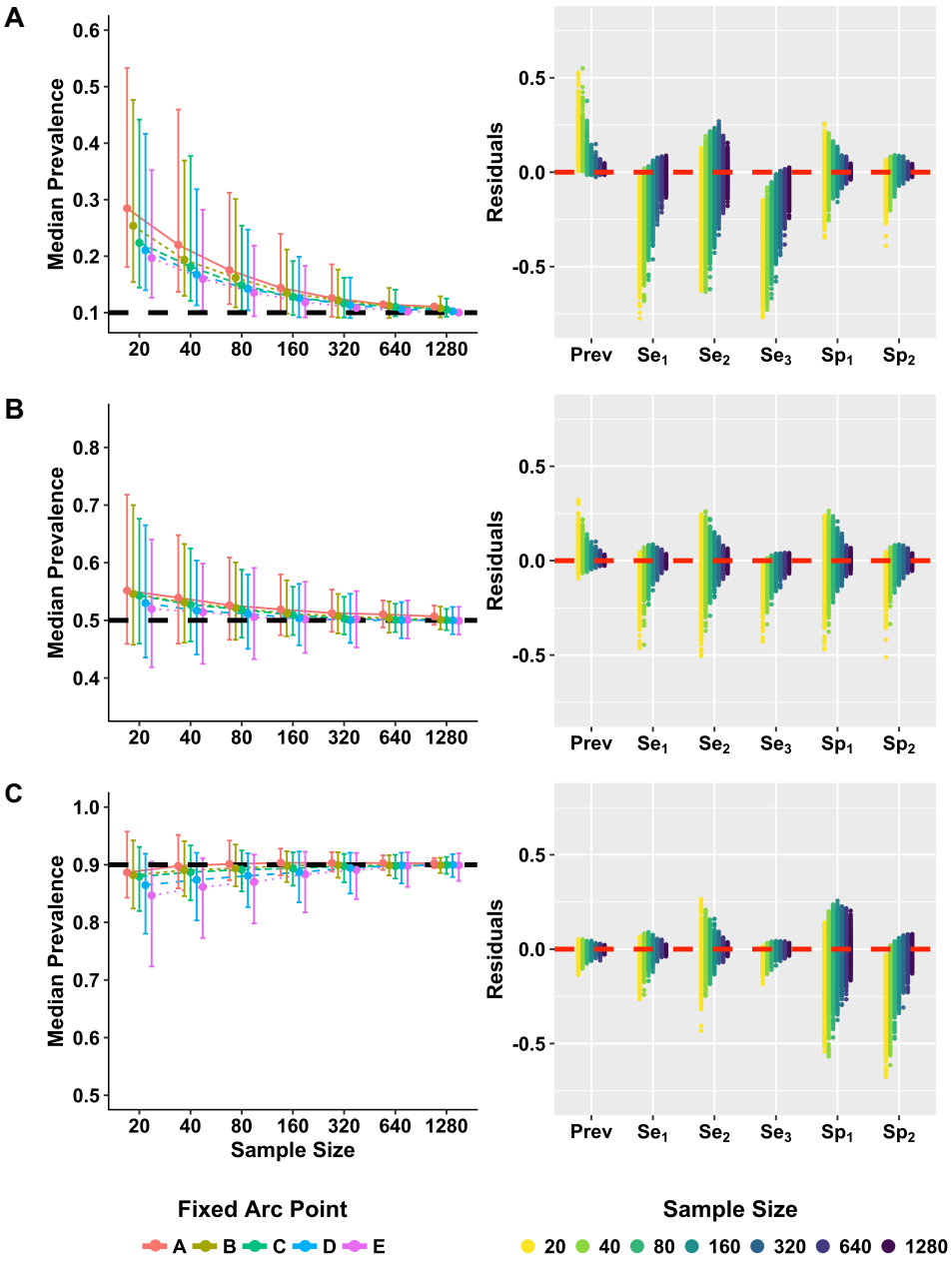


Figure S 1.3: Parameter estimates at three true prevalence levels (10%, 50% & 90%) using the first alternate scenario test 1 (Se1=0.9 and Sp1=0.7) and test 2 (Se2=0.7 and Sp2=0.9). Left: Median prevalence estimates and 95% credible intervals (CrI) are shown for points A-E at a true prevalence of 10% (A), 50% (B) and 90% (C), with true prevalence shown as dashed black lines (y-axes scaled equally). Right: Residuals for all parameter estimates (prevalence, sensitivities for tests 1-3, specificities for tests 1 & 2) using simulated samples (n=1000) generated with fixed arc point C (test 3 fixed Se=95% and Sp=95%), with zero shown as dashed red line.

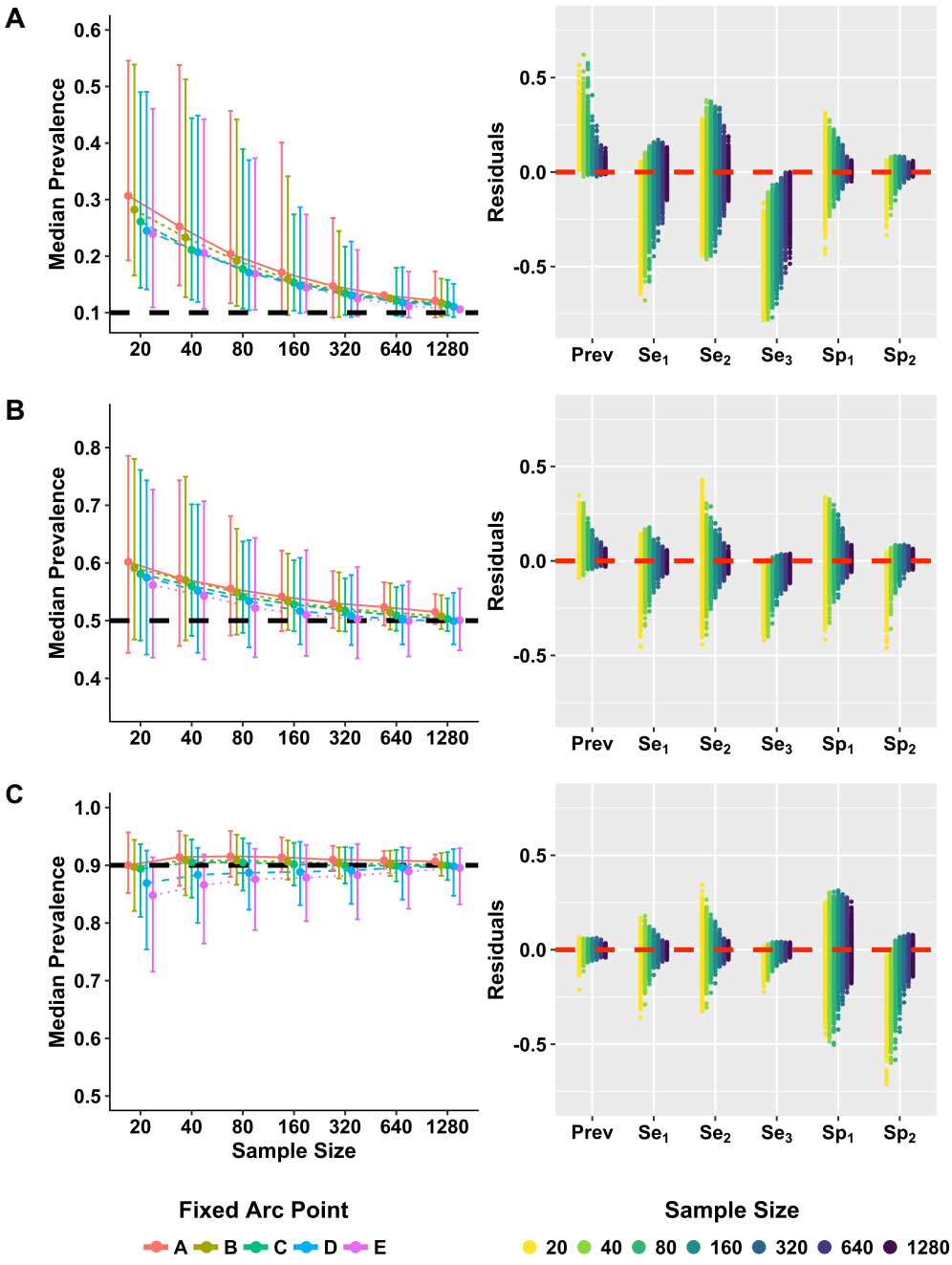


Figure S 1.4: Parameter estimates at three true prevalence levels (10%, 50% & 90%) using the second alternate scenario for test 1 (Se_1=0.8 and Sp_1=0.6) and test 2 (Se_2=0.5 and Sp_2=0.9). *Left:* Median prevalence estimates and 95% credible intervals (CrI) are shown for points A-E at a true prevalence of 10% (A), 50% (B) and 90% (C), with true prevalence shown as dashed black lines (y-axes scaled equally). *Right:* Residuals for all parameter estimates (prevalence, sensitivities for tests 1-3, specificities for tests 1 & 2) using simulated samples (n=1000) generated with fixed arc point C (test 3 fixed Se=95% and Sp=95%), with zero shown as dashed red line.

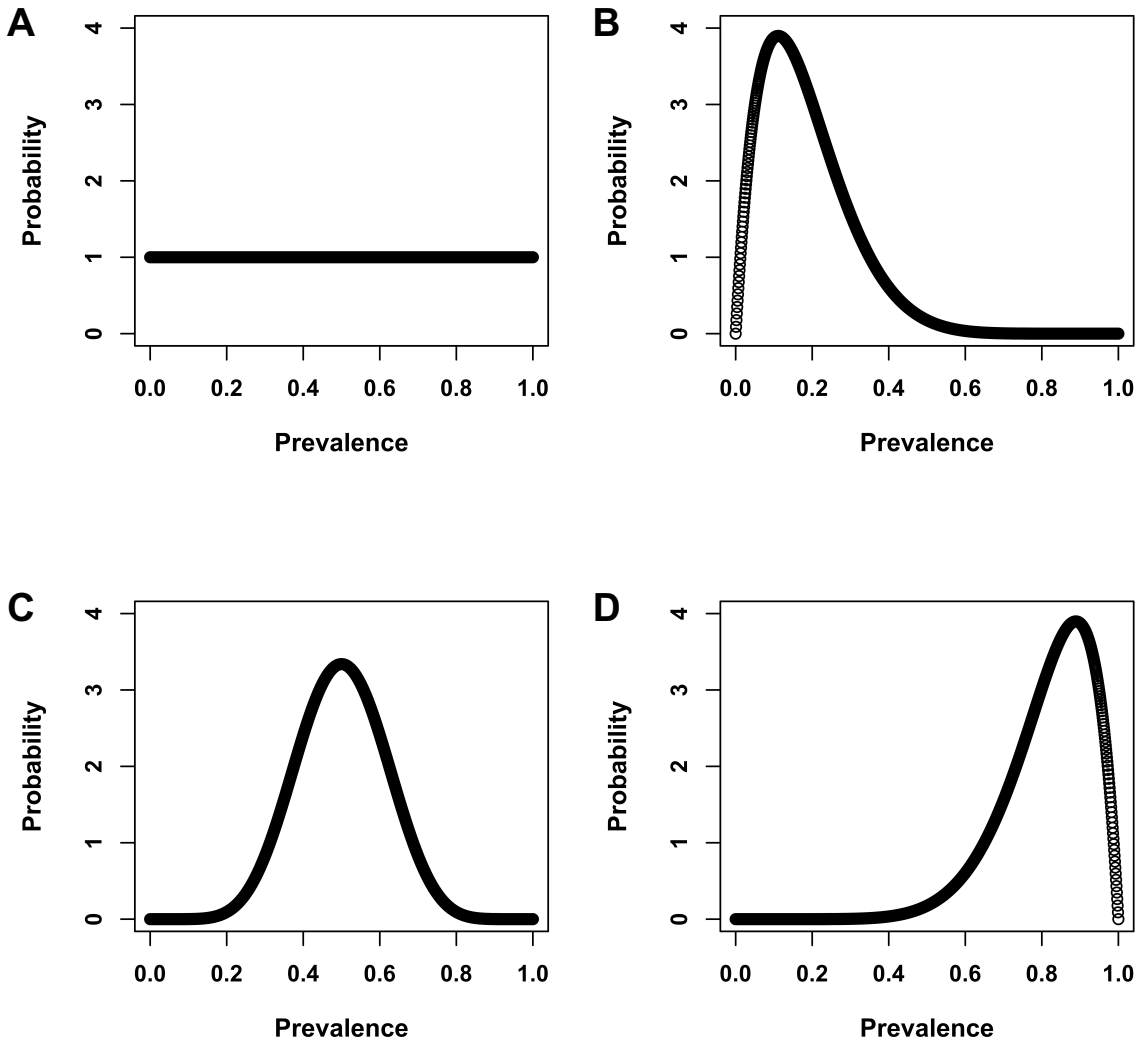


Figure S 1.5: Prior distributions used to estimate prevalence. Panel A shows an uninformed prior (uniform distribution on [0,1]). Panels B-D show informed prevalence priors (prevalence ~ beta(2,9), beta(9,9) and beta(9,2) for suspected low, medium and high prevalence respectively).

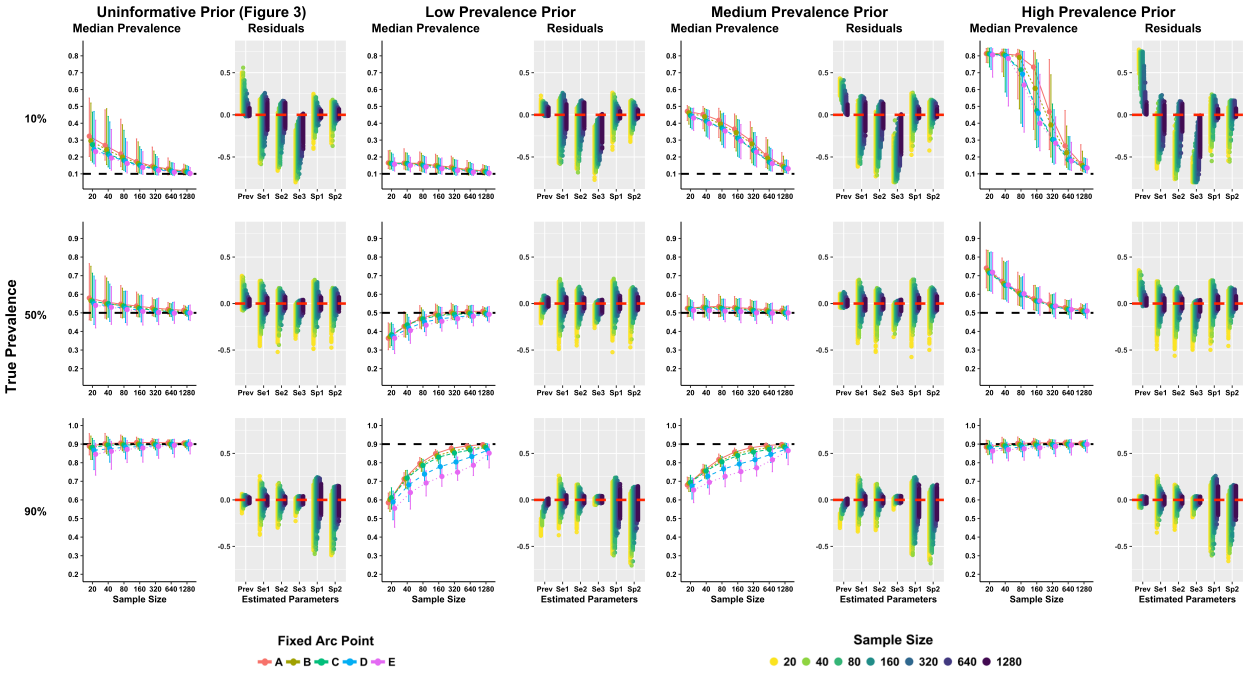


Figure S 1.6: Comparison of simulation results using uninformative prevalence priors versus informed priors with suspected low, medium and high prevalence. Each row represents results for a known true prevalence: 10% (top), 50% (middle) or 90% (bottom). Each pair of columns represents the prevalence prior: uninformed (left), low (left middle), medium (right middle) and high (right). Within each column pair, the left column represents the median prevalence by sample size (true prevalence shown as dashed black line; y-axes scaled equally), and the right column illustrates the residuals for all estimated parameters (zero shown as dashed red line).

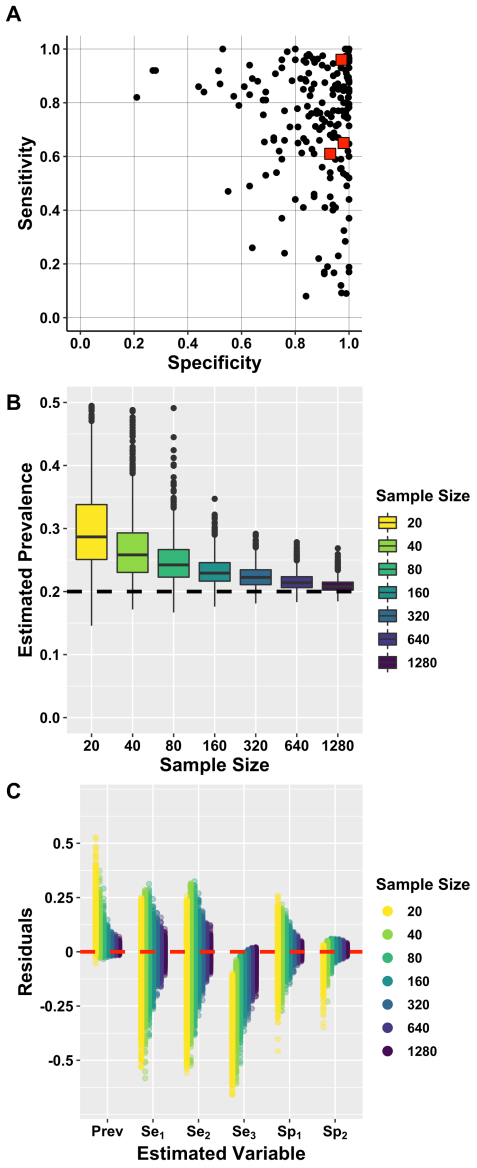


Figure S 1.7: Simulation results from model estimates using California sea lion (CSL) data. (A) Diagram showing the location of the CSL test estimates (shown in red) relative to tests previously reported in the literature (shown in black). (B) Model estimates of prevalence by sample size, using simulated data based on CSL tests, when true prevalence = 20% (dashed black line). (C) Residuals for all parameter estimates, with the true values at zero (dashed red line).

Table S 1.1: Median prevalence estimates with 95% CrI for all arc points (A-E), shown at increasing sample sizes (n=20, 40, 80, 160, 320, 640, 1280), when true prevalence is 10% (top), 50% (middle) and 90% (bottom).

Table S1A: True prevalence = 10%.

Sample Size	Prevalence *Point A Test 3: Sp=0.8/Se=1*	Prevalence *Point B Test 3: Sp=0.875/Se=0.975*	Prevalence *Point C Test 3: Sp=0.95/Se=0.95*	Prevalence *Point D Test 3: Sp=0.975/Se=0.875*	Prevalence *Point E Test 3: Sp=1/Se=0.8*
n=20	0.323 (0.203-0.549)	0.296 (0.18-0.52)	0.273 (0.169-0.466)	0.244 (0.159-0.47)	0.232 (0.145-0.418)
n=40	0.269 (0.181-0.479)	0.243 (0.155-0.486)	0.218 (0.141-0.44)	0.207 (0.132-0.408)	0.194 (0.122-0.363)
n=80	0.218 (0.143-0.424)	0.199 (0.125-0.391)	0.18 (0.121-0.358)	0.173 (0.113-0.31)	0.162 (0.105-0.289)
n=160	0.173 (0.113-0.307)	0.161 (0.107-0.29)	0.151 (0.105-0.261)	0.144 (0.1-0.245)	0.138 (0.097-0.231)
n=320	0.145 (0.099-0.259)	0.14 (0.098-0.235)	0.131 (0.096-0.215)	0.125 (0.094-0.202)	0.121 (0.094-0.191)
n=640	0.129 (0.093-0.194)	0.124 (0.096-0.175)	0.117 (0.096-0.166)	0.113 (0.092-0.165)	0.11 (0.09-0.162)
n=1280	0.118 (0.093-0.157)	0.115 (0.094-0.152)	0.112 (0.097-0.148)	0.108 (0.091-0.144)	0.103 (0.088-0.137)

Table S1B: True prevalence = 50%.

Sample Size	Prevalence *Point A Test 3: Sp=0.8/Se=1*	Prevalence *Point B Test 3: Sp=0.875/Se=0.975*	Prevalence *Point C Test 3: Sp=0.95/Se=0.95*	Prevalence *Point D Test 3: Sp=0.975/Se=0.875*	Prevalence *Point E Test 3: Sp=1/Se=0.8*
n=20	0.58 (0.462-0.765)	0.566 (0.46-0.75)	0.562 (0.475-0.713)	0.549 (0.438-0.698)	0.54 (0.419-0.676)
n=40	0.559 (0.472-0.698)	0.548 (0.467-0.687)	0.55 (0.476-0.668)	0.536 (0.454-0.655)	0.53 (0.438-0.641)
n=80	0.547 (0.475-0.647)	0.541 (0.473-0.633)	0.536 (0.476-0.615)	0.525 (0.448-0.617)	0.518 (0.435-0.615)
n=160	0.537 (0.482-0.612)	0.529 (0.478-0.601)	0.524 (0.478-0.591)	0.513 (0.451-0.596)	0.508 (0.434-0.594)
n=320	0.527 (0.485-0.58)	0.518 (0.482-0.571)	0.513 (0.479-0.57)	0.504 (0.454-0.576)	0.502 (0.437-0.579)
n=640	0.519 (0.491-0.556)	0.512 (0.482-0.551)	0.506 (0.48-0.548)	0.501 (0.46-0.556)	0.498 (0.444-0.558)
n=1280	0.514 (0.493-0.54)	0.507 (0.486-0.536)	0.502 (0.482-0.53)	0.499 (0.463-0.539)	0.50 (0.458-0.542)

Table S1C: True prevalence = 90%.

Sample Size	Prevalence *Point A Test 3: Sp=0.8/Se=1*	Prevalence *Point B Test 3: Sp=0.875/Se=0.975*	Prevalence *Point C Test 3: Sp=0.95/Se=0.95*	Prevalence *Point D Test 3: Sp=0.975/Se=0.875*	Prevalence *Point E Test 3: Sp=1/Se=0.8*
n=20	0.889 (0.842-0.956)	0.882 (0.818-0.943)	0.882 (0.817-0.931)	0.867 (0.763-0.92)	0.846 (0.733-0.904)
n=40	0.902 (0.858-0.957)	0.897 (0.849-0.942)	0.894 (0.84-0.937)	0.875 (0.803-0.926)	0.861 (0.765-0.915)
n=80	0.906 (0.873-0.948)	0.9 (0.864-0.942)	0.895 (0.856-0.934)	0.883 (0.821-0.933)	0.872 (0.801-0.92)
n=160	0.909 (0.884-0.937)	0.901 (0.873-0.931)	0.896 (0.866-0.928)	0.887 (0.836-0.927)	0.879 (0.803-0.925)
n=320	0.908 (0.891-0.93)	0.902 (0.878-0.926)	0.897 (0.869-0.924)	0.892 (0.84-0.925)	0.888 (0.824-0.927)
n=640	0.907 (0.895-0.924)	0.901 (0.881-0.922)	0.899 (0.873-0.923)	0.897 (0.849-0.925)	0.894 (0.83-0.926)
n=1280	0.905 (0.896-0.918)	0.9 (0.882-0.917)	0.9 (0.874-0.918)	0.9 (0.864-0.924)	0.899 (0.847-0.922)

Table S 1.2: Comparison of prevalence estimates using BLCA to those obtained with a single diagnostic test. BLCA results (using fixed point A) versus test A alone (left two columns), BLCA results (using fixed point C) versus test C alone (middle two columns), and BLCA results (using fixed point E) versus test E alone (right two columns). Results are shown for all sample sizes when true prevalence is 10% (S2A), 50% (S2B) and 90% (S2C). Sensitivity and specificity for tests A, C and E were Se_3=1.0/Sp_3=0.8, Se_3=0.95/Sp_3=0.95 and Se_3=0.8/Sp_3=1.0 respectively.

A

Sample Size	Prevalence	BLCA - Fixed A	Test A Only	BLCA - Fixed C	Test C Only	BLCA - Fixed E	Test E Only
20	0.1	0.32 (0.2-0.55)	0.25 (0.15-0.45)	0.27 (0.17-0.47)	0.15 (0.05-0.25)	0.23 (0.14-0.42)	0.1 (0-0.1)
40	0.1	0.27 (0.18-0.48)	0.28 (0.15-0.4)	0.22 (0.14-0.44)	0.12 (0.08-0.22)	0.19 (0.12-0.36)	0.08 (0.05-0.1)
80	0.1	0.22 (0.14-0.42)	0.28 (0.2-0.36)	0.18 (0.12-0.36)	0.14 (0.1-0.19)	0.16 (0.11-0.29)	0.09 (0.05-0.1)
160	0.1	0.17 (0.11-0.31)	0.28 (0.22-0.34)	0.15 (0.1-0.26)	0.14 (0.11-0.18)	0.14 (0.1-0.23)	0.08 (0.06-0.09)
320	0.1	0.15 (0.1-0.26)	0.28 (0.24-0.32)	0.13 (0.1-0.21)	0.14 (0.12-0.16)	0.12 (0.09-0.19)	0.08 (0.07-0.09)
640	0.1	0.13 (0.09-0.19)	0.28 (0.25-0.31)	0.12 (0.1-0.17)	0.14 (0.12-0.16)	0.11 (0.09-0.16)	0.08 (0.07-0.09)
1280	0.1	0.12 (0.09-0.16)	0.28 (0.26-0.3)	0.11 (0.1-0.15)	0.14 (0.13-0.15)	0.1 (0.09-0.14)	0.08 (0.07-0.09)

B

Sample Size	Prevalence	BLCA - Fixed A	Test A Only	BLCA - Fixed C	Test C Only	BLCA - Fixed E	Test E Only
20	0.5	0.58 (0.46-0.76)	0.6 (0.5-0.75)	0.56 (0.48-0.71)	0.5 (0.4-0.6)	0.54 (0.42-0.68)	0.4 (0.25-0.5)
40	0.5	0.56 (0.47-0.7)	0.6 (0.52-0.7)	0.55 (0.48-0.67)	0.5 (0.42-0.57)	0.53 (0.44-0.64)	0.4 (0.32-0.48)
80	0.5	0.55 (0.48-0.65)	0.6 (0.55-0.66)	0.54 (0.48-0.62)	0.5 (0.45-0.55)	0.52 (0.43-0.62)	0.4 (0.32-0.46)
160	0.5	0.54 (0.48-0.61)	0.6 (0.56-0.64)	0.52 (0.48-0.59)	0.5 (0.47-0.53)	0.51 (0.43-0.59)	0.4 (0.36-0.44)
320	0.5	0.53 (0.48-0.58)	0.6 (0.57-0.63)	0.51 (0.48-0.57)	0.5 (0.48-0.52)	0.5 (0.44-0.58)	0.4 (0.37-0.43)
640	0.5	0.52 (0.49-0.56)	0.6 (0.58-0.62)	0.51 (0.48-0.55)	0.5 (0.48-0.52)	0.5 (0.44-0.56)	0.4 (0.38-0.42)
1280	0.5	0.51 (0.49-0.54)	0.6 (0.58-0.61)	0.5 (0.48-0.53)	0.5 (0.49-0.51)	0.5 (0.46-0.54)	0.4 (0.38-0.42)

C

Sample Size	Prevalence	BLCA - Fixed A	Test A Only	BLCA - Fixed C	Test C Only	BLCA - Fixed E	Test E Only
20	0.9	0.89 (0.84-0.96)	0.9 (0.9-1)	0.88 (0.82-0.93)	0.85 (0.75-0.95)	0.85 (0.73-0.9)	0.75 (0.55-0.85)
40	0.9	0.9 (0.86-0.96)	0.92 (0.9-0.95)	0.89 (0.84-0.94)	0.88 (0.8-0.92)	0.86 (0.77-0.91)	0.72 (0.6-0.82)
80	0.9	0.91 (0.87-0.95)	0.91 (0.9-0.95)	0.9 (0.86-0.93)	0.86 (0.8-0.9)	0.87 (0.8-0.92)	0.72 (0.64-0.8)
160	0.9	0.91 (0.88-0.94)	0.92 (0.9-0.94)	0.9 (0.87-0.93)	0.86 (0.82-0.89)	0.88 (0.8-0.93)	0.72 (0.66-0.78)
320	0.9	0.91 (0.89-0.93)	0.92 (0.91-0.93)	0.9 (0.87-0.92)	0.86 (0.83-0.88)	0.89 (0.82-0.93)	0.72 (0.68-0.76)
640	0.9	0.91 (0.9-0.92)	0.92 (0.91-0.93)	0.9 (0.87-0.92)	0.86 (0.84-0.88)	0.89 (0.83-0.93)	0.72 (0.69-0.75)
1280	0.9	0.9 (0.9-0.92)	0.92 (0.91-0.93)	0.9 (0.87-0.92)	0.86 (0.85-0.87)	0.9 (0.85-0.92)	0.72 (0.7-0.74)

Neither test within 5% of the true prevalence
Both tests within 5% of the true prevalence
Only test within 5% of the true prevalence

2 He Shed She Shed: Predictors of *Leptospira* shedding in California sea lions (*Zalophus californianus*)

2.1 ABSTRACT

Pathogen shedding is a critical component of host-pathogen dynamics and disease transmission, but direct measurements of shedding are often challenging to obtain, particularly in wildlife. However, pathogen shedding can be assessed indirectly if predictors can be identified, drawing on biological knowledge of a given system to inform statistical modeling of surveillance data. In this chapter, we explore *Leptospira interrogans* in California sea lions (*Zalophus californianus*) to investigate drivers of shedding in an endemic pathogen of concern. Antibody titers against *Leptospira interrogans* serovar Pomona, measured using microscopic agglutination testing (MAT), were previously identified as an important predictor of shedding. Unfortunately, obtaining this type of data is expensive and difficult, so it is often unavailable in wildlife systems. Utilizing biological samples from stranded and wild-caught sea lions sampled between 2010-2018, we used a regression approach to investigate possible intrinsic and extrinsic drivers of *Leptospira* shedding in California sea lions to see: 1) if we could improve upon predictions made using antibody titer data alone, and 2) if we could accurately predict shedding using other, more readily available data. We found that we can predict shedding with 91% accuracy using antibody titers, and that additional clinical, demographic and environmental information does not significantly improve these estimates. Encouragingly, when antibody titer data were excluded, other significant predictors (season, sex, and indicators of kidney function) were still able to estimate shedding status with 87% accuracy, and results were robust to differences in underlying sample populations. Both models maintained their predictive accuracy when tested against out-of-sample data. The models were then used to project shedding

prevalence in the broader sea lion population, yielding patterns that aligned with independent epidemiological data. This work demonstrates an approach to leverage biological information to estimate pathogen shedding from more readily available data, generating useful insights at individual and population scales which address this universal challenge in infectious disease ecology.

2.2 INTRODUCTION

Pathogen transmission is a fundamental component of host-pathogen dynamics. Understanding and controlling infectious diseases hinges on being able to detect infectious agents in individuals and estimate levels of transmission within and between populations (Caley & Hone, 2004; Charleston et al., 2011). Though transmission cannot generally be observed directly in natural settings, the potential for transmission can be measured by detecting pathogen shedding in infectious individuals. Unfortunately, it is difficult to collect the relevant tissues or excretions during the correct window of time, and pathogen detection in these samples can encounter technical barriers. This challenge is exacerbated in wildlife populations, where logistic hurdles and cost make it difficult or impossible to obtain the necessary samples and conduct diagnostic testing at sufficient scale to study population processes. Improving our ability to predict pathogen shedding from more readily obtained sample types and surveillance data is therefore critical for estimating transmission potential when direct measurements are not available.

2.2.1 Past Efforts to Predict Shedding

Efforts to predict pathogen shedding have been influential in humans (Hagan et al., 2016; Munywoki et al., 2015), domestic animals (Harkin et al. 2003; Schares et al. 2016; Huston et al.

2002) and peridomestic rodents (Costa, Wunder, et al., 2015). In humans, predicted shedding patterns have been used for understanding transmission risks and informing public health interventions. Recent efforts include using more readily available data to predict COVID-19 shedding, employing various machine learning models to predict patient diagnosis as a contribution to large scale pandemic response efforts (Zhang et al., 2021). Predictions of shedding in domestic animals can highlight environmental risk factors and transmission risk to humans and other contact animals, with direct implications for animal welfare, public health and economics in the case of food animals. For example, multivariate analysis has been used to identify dairy herd size as a predictor of *Salmonella* shedding (Huston et al. 2002), as well as management practices (e.g., flooring type) as predictors of *Cryptosporidium* shedding on dairy farms (Trotz-Williams et al. 2008).

Investigations into drivers of shedding in wildlife are less common and typically focus on population scale dynamics (Becker et al., 2021; Blanco & Díaz de Tuesta, 2021; Hernandez et al., 2016). Variation in individual shedding (e.g., duration and intensity) can arise due to heterogeneities in host biology and the environment (Germeraad et al., 2019; Siva-Jothy & Vale, 2021). If identified, factors associated with pathogen shedding can potentially be used as proxies for shedding and transmission potential, which is highly beneficial in cases when comprehensive pathogen testing is unavailable. Increased knowledge of wildlife shedding drivers can also direct management of pathogens in wildlife populations. In Australia for example, habitat loss was identified as linked to Hendra virus shedding in bats, motivating habitat restoration efforts that reduced shedding along with the risk of cross-species transmission (Becker et al., 2021). In wildlife rehabilitation centers, identifying risk factors for shedding (e.g., seasons when patient

density and shedding prevalence of specific pathogens may be highest) can encourage enhanced hygiene and protective measures that may reduce the risk of transmission to humans and other animals (Siembieda et al., 2011). These are prime examples demonstrating that increased knowledge of shedding drivers in wildlife, particularly in the case of multi-host pathogens, has the potential to inform risk assessments and direct disease control and prevention across scales.

2.2.2 Leptospira interrogans

As a case study to investigate the drivers of shedding in an endemic wildlife pathogen, we examine leptospirosis in California sea lions (*Zalophus californianus*; CSL). Leptospirosis, the disease caused by pathogenic bacteria in the genus *Leptospira,* is the most widespread zoonotic disease in the world and affects virtually all species of mammals (Adler & de la Peña Moctezuma, 2010). With one million cases and 59,000 global deaths estimated annually in humans (Costa, Hagan, et al., 2015), and many more cases likely undiagnosed, it is surprising that *Leptospira* pathogen dynamics remain understudied (Lloyd-Smith et al., 2009). Transmission typically occurs through urinary excretion, known as shedding, following bacterial colonization of the kidneys that can lead to renal pathology (Adler & de la Peña Moctezuma, 2010; Haake & Levett, 2015). Past exposure to *Leptospira* bacteria is typically assessed by detection of anti-*Leptospira* antibodies in serum using microscopic agglutination testing (MAT), and infections with active shedding are diagnosed based on polymerase chain reaction (PCR) detection of *Leptospira* DNA in urine or kidney samples. Clinical signs can vary, with asymptomatic infections often associated with maintenance hosts (a.k.a reservoir hosts), and symptomatic infections associated with accidental (a.k.a. spillover) hosts, though these classifications are not absolute (Ko et al., 2009; Lloyd-Smith et al., 2007). In symptomatic

clinical infections in humans and other species, clinical signs are often associated with primary renal pathology, though they can range from mild to severe, including kidney failure and death (Williams & Barker, 2001). Susceptible hosts contract infections through direct contact with contaminated host tissues, or via indirect contact with environments contaminated by urinary shedding (Casanovas-Massana et al., 2018). Transmissibility and environmental survival of the bacteria can vary with ecological conditions (Barragan et al., 2017), such as rainfall and flooding which have both been identified as environmental drivers of leptospirosis outbreaks (Lau et al., 2010; Nakata et al., 2007).

2.2.3 Study System: Leptospira interrogans in California Sea Lions

Leptospira interrogans serovar Pomona has circulated endemically in sea lions since the early 1980s, with seasonal peaks in the fall and periodic outbreaks every few years (Lloyd-Smith et al., 2007). Clinical signs in infected sea lions range from asymptomatic to severe kidney failure and death associated with primary renal disease, with mortality in stranded sea lion infections approaching 60-70% (Greig et al., 2005; Gulland et al., 1996; Prager et al., 2013). This demonstration of both endemic and epidemic dynamics, along with clinical signs ranging from non-existent to fatalities, make sea lions a prime example of why the maintenance and accidental host paradigm does not universally apply (Lloyd-Smith et al., 2007; Prager et al., 2013). Higher disease prevalence and seroprevalence have been reported in males, putatively due to the differences in migratory patterns and contact rates (Greig et al., 2005; Gulland et al., 1996). Environmental conditions have known associations with sea lion prey abundance and demography (DeLong et al., 1991, 2017; Melin et al., 2008, 2010), and links between the environment and sea lion leptospirosis incidence and susceptibility have long been suspected

(Gulland et al., 1996). While early theories centered on El Nino effects mediated by demographic impacts, recent work has refined our knowledge of environmental influence on this system, with correlations between leptospirosis outbreak intensity and local oceanographic factors (sea surface temperature and upwelling) demonstrated even after accounting for demographic effects (Borremans et al., In Preparation). The underlying causes of these correlations with environmental factors are unknown, but effects mediated by individual level shedding (via changes in individual condition, due to trophic stress) are a leading hypothesis.

Decades of past research efforts have provided extensive information about this disease process in sea lions, including stranding records dating back to the 1980s (Greig et al., 2005), long-term time series including serology and diagnostics (Lloyd-Smith et al., 2007), targeted clinical studies (Prager et al., 2015, 2020) and extensive field research on the wild sea lion population since 2010 (DeLong et al., 2017; Melin et al., 2012; Prager et al., 2020). Analyses of these data point to the possible roles of demography and environmental factors in this pathogen system (Buhnerkempe et al. 2017), in addition to identifying biomarkers such as high antibodies and serum chemistry alterations that are associated with recent cases of leptospirosis (Prager et al., 2020). Acute cases typically present with high MAT titers, and long-term clinical data highlights the strong, predictive correlation between *Leptospira interrogans* serovar Pomona antibody titers and sea lion shedding status (Prager et al., 2020). Leveraging the wealth of existing information in this complex host-pathogen system can provide insights into additional factors associated with leptospirosis infections, which could be used to improve predictions of *Leptospira* shedding status. For example, alterations in serum biochemistry are associated with renal compromise in clinical leptospirosis cases (Greig et al., 2005; Gulland et al., 1996), and may be a better

indicator of very recent infections than high antibody titers (Prager et al., 2020). Additional information such as serum chemistry values may therefore be predictive of *Leptospira* shedding status, and, if used in conjunction with antibody titers, could improve upon individual shedding predictions generated from titer data alone.

Accurate predictions of sea lion shedding status from readily available clinical, demographic or environmental data would advance our understanding of *Leptospira* infections, elucidating this key component of pathogen transmission in this system. This is particularly true since obtaining direct evidence of shedding is not easy, as PCR requires kidney samples (collected during necropsy) or sterile urine samples (collected under anesthesia, which presents major challenges under field conditions). Further, while a clear link has been established between antibody titers and shedding status, antibody titer data is not always available. Assessing antibodies using MAT is resource intensive, requiring the expertise of trained professionals at external laboratories (which can take several days), making it less useful for real-time risk assessments in clinical settings than more readily available clinical data such as serum chemistry. Consequently, marine mammal rehabilitation centers do not conduct MAT as part of routine testing in most California sea lions. Finding ways to estimate shedding from more readily available data would therefore enable individual and population-level shedding predictions in a wider array of circumstances.

In this study, we probe the drivers of *Leptospira* shedding in California sea lions. Using established knowledge of sea lion and *Leptospira* biology, we analyze 21 possible predictors from four broad categories, representing classes that are biologically relevant to cases of leptospirosis: *Leptospira* covariates (MAT titers and leptospirosis cases), kidney pathology

covariates (serum chemistry results indicative of kidney function), demographic covariates (age class and sex) and environmental covariates (season, upwelling, sea surface temperature and spring transition). We develop predictive models of shedding, assessing whether additional information can improve upon shedding predictions made from antibody titer data alone. We then investigate whether shedding can be predicted in the absence of antibody titers using more commonly accessible data. We quantify the accuracy of each predictive model within training data and out-of-sample test data, and then demonstrate the application of our methods by predicting shedding frequency across the broader sea lion population. By leveraging surveillance data to gain a more in-depth understanding of *Leptospira* shedding in California sea lions, we demonstrate an effective method to elucidate drivers of transmission in this complex host-pathogen system.

2.3 DATA

2.3.1 Study Animals

2.3.1.1 Stranded Sea Lions

California sea lions strand along California's central coast due to injury or illness, including pneumonia, neoplasia, malnutrition, trauma, domoic acid toxicity and leptospirosis (Greig et al., 2005; Gulland et al., 1996). The Marine Mammal Center (TMMC; Sausalito, CA, USA) is a marine mammal rehabilitation center that rescues and rehabilitates animals that strand between Mendocino and Santa Barbara counties. TMMC veterinary staff routinely collect serum as part of clinical examinations. Urine is collected when possible, such as during incidental procedures that involve anesthesia, or routine necropsies following the death of an animal. Only sea lions

with samples collected within a week of admission were included in this study to minimize alterations in test results due to treatment (Table 2.1).

2.3.1.2 Wild Sea Lions

Seemingly healthy wild sea lions (Table 2.1) were captured in the field along the California coast, anesthetized using isoflurane gas, and sampled under the authorization of Marine Mammal Protection Act Permit No. 932-1905-00/MA-009526. Wild-caught sea lions were sampled along the southern (San Miguel Island) and central (Año Nuevo Island and Pier 39 in San Francisco) California coast (Table 2.1). Sea lions from San Miguel Island were sampled in March and between August and October, on Año Nuevo Island between September and November, and between August and September at Pier 39. Venipuncture was performed using the caudal gluteal vein (Gili et al., 2018). Serum from centrifuged blood was transferred to cryovials and frozen until use, and sterile urine was collected via catheterization or cystocentesis as detailed in Prager et al. (2020). All samples were collected at the time of capture while animals were under isoflurane anesthesia.

2.3.1.3 Training and Test Datasets

Stranded and wild sea lion samples collected from 2010-2018 were used as training data for our predictive models; our research program on leptospirosis in sea lions intensified in 2010, greatly increasing the number of urine and kidney samples available during this period (Table 2.1). However, only data from years with endemic *Leptospira* circulation were included as part of the training data (2010-2012 and 2017-2018). Data from 2013-2016 were excluded due to an unprecedented pathogen fadeout event, during which time *Leptospira* was not present in the

population (Prager et al., In Preparation). The complete training dataset, composed of both stranded and wild sea lion samples, will be referred to as the aggregate dataset. Below we describe specific subsets of this dataset that were also analyzed separately. The test dataset consisted of sea lion samples collected from 2006-2009 (Table 2.1), and these samples were used to assess the out-of-sample predictive accuracy of our models. During this period, the majority of samples came from stranded animals with a slightly older age distribution than the training data (Table 2.1).

2.3.2 Ethics Statement

Samples from California sea lions were authorized for collection by Marine Mammal Protection Act Permits No. 932-1905-00/MA-009526 and No. 932-1489-10 (issued by the National Marine Fisheries Service - NMFS), and NMFS Permit Numbers 17115–03, 16087–03, and 13430 as detailed in Prager et al. (2000). Approval for the collection protocol was granted by the Institutional Animal Care and Use Committees (IACUC) of the University of California Los Angeles (UCLA ARC # 2012-035-12), The Marine Mammal Center (TMMC protocol # 2008–3), and the National Marine Mammal Laboratory (NMML Alaska Northwest 2013–1 and 2013–5). The University of California Los Angeles is accredited by AAALAC International, and all agencies (TMMC, NMML and UCLA) adhere to the United States Public Health Service Policy on the Humane Care and Use of Laboratory Animals and the United States Department of Agriculture (USDA) Welfare Act.

Table 2.1: Descriptive characteristics of California sea lion data by shedding status. Characteristics are shown by *Leptospira* PCR status for training data (left) and test data (right). Within group percentages are proportions of column totals for each category.

	TRAINING DATA			TEST DATA		
	NEG (n=430)	POS (n=187)	Total (n=617)	NEG (N=120)	POS (N=86)	Total (N=1828)
Sex						
Female	204 (47.4%)	23 (12.3%)	227 (36.8%)	35 (29.2%)	3 (3.5%)	733 (40.1%)
Male	226 (52.6%)	164 (87.7%)	390 (63.2%)	85 (70.8%)	83 (96.5%)	1095 (59.9%)
Age						
Adult	48 (11.2%)	11 (5.9%)	59 (9.6%)	57 (47.5%)	4 (4.7%)	452 (24.7%)
Subadult	18 (4.2%)	39 (20.9%)	57 (9.2%)	21 (17.5%)	55 (64%)	543 (29.7%)
Juvenile	206 (47.9%)	96 (51.3%)	302 (48.9%)	23 (19.2%)	26 (30.2%)	345 (18.9%)
Yearling	158 (36.7%)	41 (21.9%)	199 (32.3%)	19 (15.8%)	1 (1.2%)	488 (26.7%)
Antibody Status						
Seronegative	407 (94.7%)	36 (19.3%)	443 (71.8%)	91 (75.8%)	4 (4.7%)	857 (46.9%)
Seropositive	23 (5.3%)	151 (80.7%)	174 (28.2%)	29 (24.2%)	71 (82.6%)	464 (25.4%)
PCR Sample Type						
All	4 (0.9%)	4 (2.1%)	8 (1.3%)	-	-	-
Kidney	7 (1.6%)	17 (9.1%)	24 (3.9%)	-	-	-
Pellet	285 (66.3%)	68 (36.4%)	353 (57.2%)	-	-	-
Urine	110 (25.6%)	26 (13.9%)	136 (22%)	120 (100%)	86 (100%)	206 (11.3%)
Urine + Kidney	24 (5.6%)	15 (8%)	39 (6.3%)	-	-	-
Urine + Pellet	0 (0%)	57 (30.5%)	57 (9.2%)	-	-	-
Sample Population						
Strand	87 (20.2%)	126 (67.4%)	213 (34.5%)	120 (100%)	86 (100%)	1744 (95.4%)
Wild	343 (79.8%)	61 (32.6%)	404 (65.5%)	0 (0%)	0 (0%)	84 (4.6%)
Location						
Central Coast (Wild)	152 (35.3%)	55 (29.4%)	207 (33.5%)	0 (0%)	0 (0%)	32 (1.8%)
Southern Coast (Wild)	191 (44.4%)	6 (3.2%)	197 (31.9%)	0 (0%)	0 (0%)	52 (2.8%)
TMMC (Strand)	87 (20.2%)	126 (67.4%)	213 (34.5%)	120 (100%)	86 (100%)	1744 (96.6%)
Year						
2018	53 (12.3%)	58 (31%)	111 (18%)	-	-	-
2017	135 (31.4%)	29 (15.5%)	164 (26.6%)	-	-	-
2012	95 (22.1%)	29 (15.5%)	124 (20.1%)	-	-	-
2011	93 (21.6%)	55 (29.4%)	148 (24%)	-	-	-
2010	54 (12.6%)	16 (8.6%)	70 (11.3%)	-	-	-
2009	-	-	-	17 (14.2%)	5 (5.8%)	544 (29.8%)
2008	-	-	-	57 (47.5%)	17 (19.8%)	366 (20%)
2007	-	-	-	21 (17.5%)	30 (34.9%)	544 (29.8%)
2006	-	-	-	25 (20.8%)	34 (39.5%)	374 (20.5%)

2.3.3 Leptospira PCR Data (Dependent Variable)

Leptospira shedding was assessed using real-time polymerase chain reaction (PCR) targeting the *LipL32* gene to detect *Leptospira* DNA in urine or kidney samples (Wu et al., 2014). Urine was collected aseptically by catheterization or cystocentesis, and kidney samples were collected aseptically at necropsy (Table 2.1). Urine pellets were collected in cases where at least 4 ml of urine was available: 2 ml were aliquoted into a 2 ml cryovial tube, the remaining 2+ ml were centrifuged into a pellet and resuspended in supernatant or phosphate buffer saline (PBS). When available, urine pellets were prioritized for PCR testing as described in Prager et al. (2013). Of the animals that had both kidney and urine samples available (n=47), paired PCR results were 100% consistent between the two sample types (19 positive and 28 negative). This was also true for the subset of those animals which had all three (kidney, urine and urine pellets) sample types available (n=8; 4 positive and 4 negative). Results from all sample types were therefore considered equivalent with regards to *Leptospira* shedding and infectiousness.

2.3.4 Details of Covariates (Independent Variables)

2.3.4.1 *Leptospira* Covariates

2.3.4.1.1 Serum Antibody Titers

Past exposure to *Leptospira* was assessed via (MAT) conducted at the Centers for Disease Control and Prevention (CDC; Atlanta, Georgia, USA) or the California Animal Health and Food Safety Laboratory (CAHFS; Davis, California, USA). This test detects the presence of anti-*Leptospira* antibodies in serum, which clump when samples are combined with live cultures of *Leptospira* reference strains (Levett, 2001). Serum samples were assessed at doubling dilutions. Endpoint titers were the highest dilution that achieved 50% agglutination using the reference

strain being tested and are presented on a log-transformed scale, where zero corresponds to seronegativity and each unit increase corresponds to a two-fold dilution starting at 1:100 (Faine et al., 1999). Titers are higher following acute clinical infections (median of 1:102,400) and decay rapidly for 1-6 months, with lower titers persisting for as long as 3 years after an infection in some animals (Prager et al., 2020). Although samples were tested against a panel of potential serovars, only antibody titers against *Leptospira interrogans* serovar Pomona were used for this analysis because Pomona is the only serovar ever isolated from California sea lions (Zuerner & Alt, 2009). Hereafter, the phrase antibody titers will refer to results from the MAT diagnostic test. Seropositive and seronegative refers to animals that were positive or negative for antibody titers, respectively, using the conventional definition of titers less than 1:100 as negative (Table 2.1).

2.3.4.1.2 Leptospirosis Cases

Current evidence shows that most symptomatic animals shed for 2-4 weeks, while others with chronic infections can shed for much longer (Buhnerkempe et al., 2017; Prager et al., 2013, 2015, 2020). As a proxy for the incidence of leptospirosis in the CSL population during the prior 3 months (acute) and 12 months (chronic), we use the total number of leptospirosis cases at TMMC recorded during those periods. Leptospirosis cases at TMMC are designated by clinical staff and defined based on a combination of clinical signs, serum chemistry results indicative of kidney failure (BUN>100 mg/dl, creatinine >2 mg/dl, sodium >155 meq/L, and phosphorus > calcium), kidney pathology, serum antibody titers (MAT) to *Leptospira interrogans* serovar Pomona (>1:3200), or positive PCR or culture results (Greig et al., 2005; Gulland et al., 1996).

2.3.4.2 Environmental Covariates

Both season and oceanographic indices have been associated with leptospirosis in California sea lions. Due to seasonal outbreaks that typically span the second half of the year with peaks in the fall (Buhnerkempe et al., 2017; Gulland et al., 1996), season was incorporated on a quarterly basis (Winter: Jan-Mar; Spring: April-June; Summer: July-Sept; Fall: Oct-Dec). We also incorporated a limited set of oceanographic indices, based on associations with productivity in the California Current, impacts on sea lion prey sources, and effects on *Leptospira* outbreak intensity in the sea lion population (Borremans et al., In Preparation): sea surface temperature (SST), upwelling index (UPW), and the timing of spring transition (ST) of upwelling intensity (Bograd et al., 2009; DeLong et al., 2017; Melin et al., 2008, 2010). Sea surface temperatures can affect the abundance and location of sea lion prey, and have also been linked to sea lion foraging and survival (Melin et al., 2010; DeLong et al., 2017). Upwelling can also impact sea lion survival and movement through prey abundance and location, as upwelling conditions drive marine productivity via transport of micronutrients to the coastal ecosystem which influence prey resources, and an earlier start to the upwelling season (known as spring transition) can result in longer periods of nutrient availability and ecosystem productivity (Chavez et al., 2003; Bograd et al., 2009). Variations in sea lion prey abundance and distributions can impact sea lion condition, with the potential to affect sea lion health and immunity.

Upwelling data are freely available from the Pacific Fisheries Environmental Laboratory (www.pfeg.noaa.gov), and sea surface temperature data are freely available from the National Data Buoy Center (www.ndbc.noaa.gov). In addition to providing a measure of the upwelling index, data from these buoys were used to assess spring transition timing, calculated as the day

between January and June when the cumulative upwelling index reaches a minimum value (Bograd et al., 2009). Upwelling data were measured at 36N latitude to represent the prevailing conditions on the central California coast, where the majority of our data were collected. Sea surface temperature data were obtained from a combination of buoys since no single buoy had data available for the full study period: 46026 (37.755N, 122.839W), 46028 (35.712N, 121.858W), and 46042 (36.785N, 122.398W). Monthly anomalies were calculated for SST and UPW by subtracting the mean value for each calendar month across all years from observed monthly values. Monthly anomalies were then averaged across the previous 3 and 12 months to approximate potential impacts at two time scales: short-term (reflecting direct effects on sea lions and their prey) and longer term (mediated by primary productivity effects which take longer to cascade up the food chain and impact sea lions). We also considered two possible indices for the spring transition of upwelling intensity, by calculating spring transition anomalies for both the current year and the year prior, as timing can be associated with reduced marine productivity the following year (Bograd et al., 2009).

2.3.4.3 Demographic Covariates

Age and sex were both included in the model, and were determined by experienced field biologists, veterinarians and personnel. Age classes were defined as follows : yearling males and females (12 - 23 months old), juvenile males (2 - 3 years old), juvenile females (2 - 4 years old), subadult males (4 - 7 years old), and adults (females 5+ years old and males 8+ years old; Greig, Gulland, and Kreuder 2005). Pups (less than 12 months old) were excluded from this study due to the very low number of animals (n=4) that met the data inclusion criteria.

2.3.4.4 Kidney Pathology Covariates

Clinical leptospirosis cases are associated with primary renal pathology, leading to alterations in serum chemistry values that relate to kidney function (Gulland et al., 1996; Marcondes et al., 1996; Prager et al., 2020). Following serum collection, serum chemistry was analyzed using an ACE Clinical Chemistry System (Alfa Wassermann, Inc., West Caldwell, New Jersey, USA), assessing concentrations of blood urea-nitrogen (BUN), calcium, chloride, creatinine, phosphorus, potassium and sodium. In stranded sea lions serum was collected within one week of admission to TMMC, and in wild sea lions serum was collected at the time of capture.

2.4 STATISTICAL ANALYSIS

2.4.1 Training Data

2.4.1.1 Variable Selection with LASSO Regression

To identify variables that are predictive of *Leptospira* shedding in CSL, the aggregate training dataset (n=617; Table 2.1) was analyzed using a least absolute shrinkage and selection operator (LASSO) logistic regression. This method aims to identify the most parsimonious predictors of an outcome (here, *Leptospira* shedding) by minimizing the residual sum of squares subject to a shrinkage parameter (λ) which restricts the coefficients of less important variables to zero (Tibshirani, 1996). This approach can yield more interpretable models when the number of predictors is large (Rasmussen & Bro, 2012), and has been applied in similar epidemiological contexts (Corsi et al., 2016; Zhang et al., 2021). LASSO is particularly well-suited for dealing with collinearity between predictor variables, which is useful in this analysis since oceanographic variables at different timescales were correlated (UPW 3 & 12 months: Spearman's rho=0.61; SST 3 & 12 months: Spearman's rho=0.71). Using the *R* package

'cv.glmnet', the shrinkage parameter (λ) was selected through 10-fold cross-validation using binomial deviance as the cross-validated error measure (Friedman et al., 2010). We selected the largest lambda value with error within 1 standard error of minimum ('Lambda.1se'), which is a standard approach to balance accuracy and parsimony (T. Hastie et al., 2009; Krstajic et al., 2014). All analyses were done in *R* using version 3.6.1 (R Core Team, 2021).

2.4.1.2 Estimating Parameter Uncertainty

To assess uncertainty in the model parameters, we utilized the bootstrap approach introduced by Hastie et al. (2016). For each run of the bootstrap (n=1000), the aggregate training dataset was resampled with replacement to construct an equivalent sized dataset (n=617), after which the LASSO regression was used for variable selection as outlined above. This generated a set of 1000 bootstrap coefficients for each covariate. We report the median coefficients with 95% confidence intervals (CI), which represent the central 95% of each bootstrapped coefficient distribution. We considered covariates to be *significant predictors* if they had non-zero median coefficients and did not contain zero in their 95% CI. Covariates are referred to as *potential predictors* if they had non-zero median coefficients but did contain zero in their 95% CI; as these variables were retained by the LASSO at least 50% of the time, they were considered to have some potential influence on shedding (which could be clarified with further research). For ease of interpretation, coefficients for our logistic regression models were presented as odds ratios (OR; Table S 2.1).

2.4.1.3 Final Model Construction

The LASSO bootstrap was first performed on the aggregate training dataset (all data from 2010-2018, excluding the period between 2013 and 2016) using all covariates (n=21), including anti-*Leptospira* antibody data. To identify variables predictive of shedding in situations where antibody data may not be available, we excluded the antibody data and repeated the LASSO bootstrap using the remaining covariates (n=20). In each scenario, significant predictors from the LASSO bootstrap were incorporated into a predictive model of shedding. This resulted in two models: one model incorporating significant predictors when considering all available covariates (termed the Full model) and the other incorporating significant predictors when antibody data were excluded (termed the NoAb model).

2.4.1.4 Assessment of Prediction Accuracy

To generate shedding predictions from each selected model, along with associated uncertainties, we utilized the bootstrap distribution of model coefficients (n=1000). For each bootstrap run, the set of coefficients for significant variables was used to calculate the predicted probability of each animal being PCR positive using the logit equation. This yielded a bootstrap distribution of 1000 predicted probabilities for each animal. We calculated the predicted shedding probability for each animal by taking the mean of this distribution, in order to average over model uncertainty while accounting for the covariance structure among the coefficient estimates.

To assess the accuracy of each model at the individual level, we must choose a threshold for the predicted shedding probability above which an animal is predicted to be PCR positive. For each model, classification accuracy was assessed across different threshold probability levels to

identify the optimal predicted probability threshold that maximizes predictive accuracy ('ROCit' package in *R*). Using this optimal threshold, we calculated overall model accuracy based on the proportion of animals that were predicted correctly (accuracy = (true positives + true negatives) / total animals), and then resampled the data 1000 times to estimate 95% confidence intervals.

2.4.1.5 Analysis in Data Subsets

To assess what impact the sample population may have on predictors of shedding, we repeated the bootstrap LASSO regression using all covariates in three subsets of the training data (Table 2.1). First, the aggregate dataset was divided into wild (n=404) and stranded (n=213) sea lions. Previous work identified occasional cases of shedding in seronegative animals, therefore we analyzed the subset of seronegative individuals (n=443) in a separate analysis to identify possible covariates beyond antibody titers that might predict shedding status in these unusual animals (Prager et al., 2020). In cases where significant predictors of shedding were present, we calculated accuracy using the models developed from the aggregate dataset (Full and NoAb models).

2.4.2 Test data

2.4.2.1 Out-of-sample Shedding Predictions

To test the out-of-sample predictive accuracy of our models, the Full model and NoAb model (with coefficients estimated from the training data) were applied to the test dataset from 2006-2009. To assess the Full model, data were selected from sea lions that had antibody titer data available during the 2006-2009 time period (n=805). The Full model and coefficients from the LASSO bootstrap were then applied to this dataset, yielding 1000 predicted shedding

probabilities, which were then averaged to get a predicted probability of shedding for each sea lion. Using the subset of animals in this dataset with PCR results (n=132), we calculated the out-of-sample predictive accuracy of the model, and the dataset was resampled to calculate 95% confidence intervals. This process was repeated for the NoAb model using all animals that had serum chemistry results available (n=1074), with accuracy assessed as outlined above using the subset of those animals with PCR results (n=77).

Results were then extrapolated to the broader population by plotting the monthly shedding predictions for all animals with suitable data in this time period, for each model (yielding n=805 predictions from the Full model, and n=1074 predictions from the NoAb model). To assess the realism of these population-scale-predictions, we compiled estimates into monthly indices of the predicted prevalence of shedding in the sampled sea lion population, and compared these projections (with 95% binomial confidence intervals) to the time series of leptospirosis cases admitted at The Marine Mammal Center.

2.5 RESULTS

2.5.1 Training Data

2.5.1.1 Variable Selection and Model Fitting - All Covariates

To develop the Full model, we considered all 21 covariates, including antibody titer data. Antibody titers were the strongest predictor of shedding (Figure 2.1, Table S 2.1), with an odds ratio of 42.21 (95% CI: 20.35 - 121.86). This was the only significant variable and, therefore, the only one included in the Full model. Sex (male) and season (fall) were both potential predictors, though they had a much smaller influence on shedding, with odds ratios of 1.53 (95% CI: 1.00 -

3.48) and 1.36 (95% CI: 1.00 - 4.26) respectively. With a significant odds ratio approximately 30-fold higher than both sex and season, antibody titer was the most important predictor of shedding within the full set of covariates considered here.

2.5.1.2 Variable Selection and Model Fitting - Antibody Data Excluded

To develop the NoAb model, we considered all covariates with the exception of antibody titer data. Both sex (male) and season (fall) were identified as significant predictors (Figure 2.1; Table S 2.1), with odds ratios of 2.96 (95% CI: 1.69-5.77) and 2.31 (95% CI: 1.2-5.08), respectively. Two additional variables indicative of kidney dysfunction also emerged as significant predictors of shedding: BUN (OR = 35.48, 95% CI: 12.20-384.64) and potassium (OR = 0.43, 95% CI: 0.10-0.86). Additionally, upwelling anomalies (short and longer term) and spring transition were identified as potential predictors, indicating that they may have some additional but lesser influence on *Leptospira* shedding.

2.5.1.3 Assessment of Prediction Accuracy

To assess the accuracy of the Full model and the NoAb model, we needed to select a threshold for predicted shedding probability, above which an animal is predicted to be PCR positive. For each model, overall classification accuracy was assessed at different probability thresholds in order to identify the predicted probability threshold that maximizes predictive accuracy using the 'ROCit' package in *R* (Figure S 2.1). Accuracy (true positives and true negatives / total samples) is maximized at a predicted shedding probability threshold of 0.37 for the Full model and 0.35 for the NoAb model, yielding overall classification accuracy of 91.4% (95% CI : 88.9 - 93.5%) and 87.2% (95% CI : 84.4 - 89.8%), respectively (Figure 2.2; Table S 2.2). Adding potential

predictors to the Full model did not improve shedding predictions, resulting in an accuracy of 91.9% (95% CI : 89.5 - 93.9%). Though we chose to optimize predictions by selecting the predicted probability threshold that maximized overall model accuracy, we note that a range of thresholds, including the intuitive threshold of 0.5 (i.e., 50% likelihood of shedding), would produce comparable accuracy for each model (Figure S 2.1). This is likely because the predicted shedding probabilities are clustered at extreme values, with concentrations of points found near probabilities of zero and one, and fewer points at intermediate values (Figure 2.2). This clustering was mostly driven by high-confidence predictions that were correct, but one exception is the group of seronegative animals that were shedding (i.e., those with negative titers and positive PCR results). Since the Full model is based entirely on antibody titer, it predicts a very low shedding probability for animals with no titer value, resulting in a cluster of false negatives for these seronegative shedders (Figure 2.2).

2.5.1.4 Variable Selection and Prediction Accuracy in Data Subsets

2.5.1.4.1 Variable Selection in Stranded and Wild Sea Lions

When evaluating the full set of covariates, predictors of shedding in wild (n=404) and stranded (n=213) sea lions mirrored those from the aggregate dataset. Antibody titers remained the only significant predictor of shedding in both wild (OR = 52.47, 95% CI: 1.49 - 1549.81) and stranded sea lions (OR = 24.41, 95% CI: 9.40 - 220.88; Table S 2.1). Potential predictors in stranded animals included sex (male), BUN and potassium, while season (fall) and sex (male) were potential predictors in wild sea lions. These potential predictors also echoed results from the aggregate dataset, where these variables were all significant and included in the NoAb model.

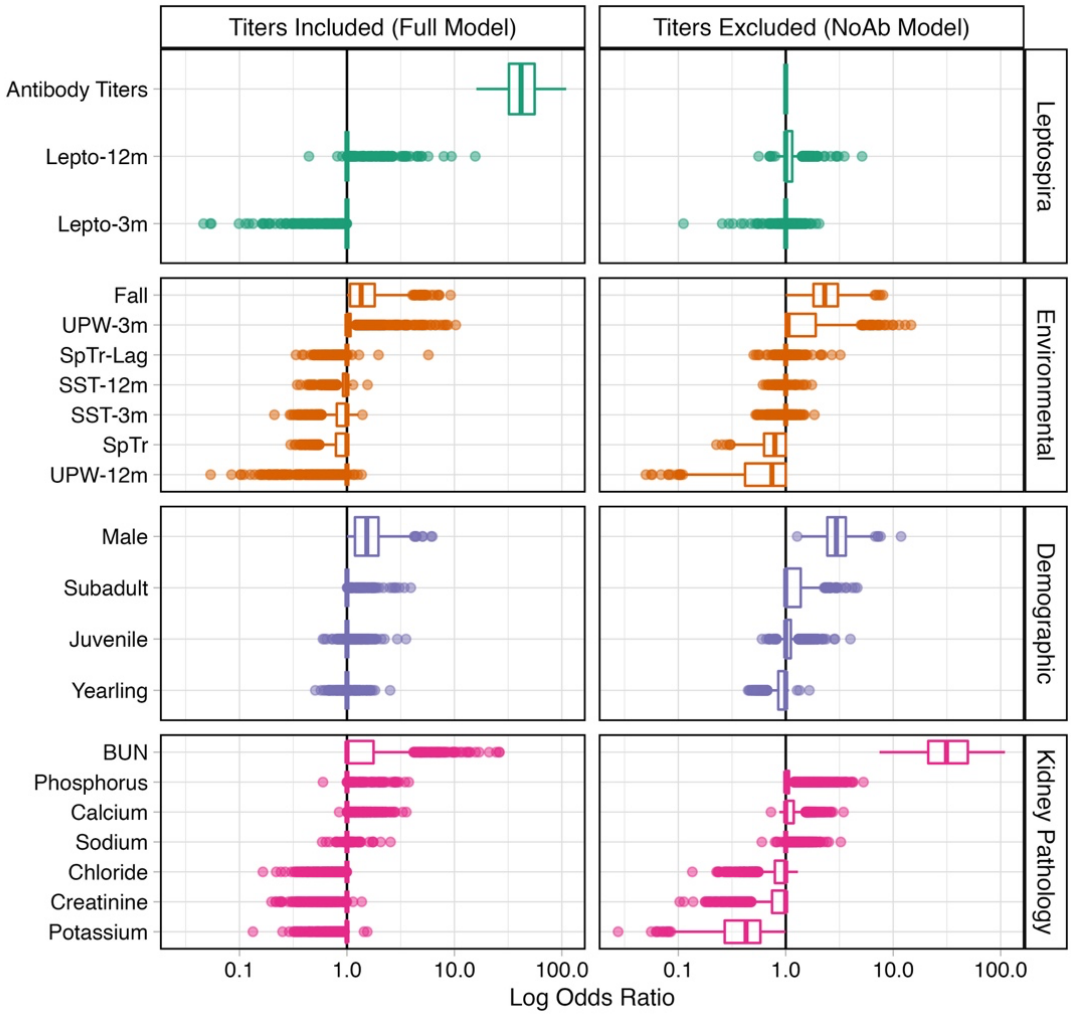


Figure 2.1 Predictors of *Leptospira* shedding in California sea lions, under models with and without antibody data available. Odds ratio distributions from a bootstrap LASSO regression (n=1000) are shown for all covariates including antibody titers (left), and all covariates excluding antibody data (right). Covariates were grouped into four broad classes: those that related to *Leptospira* (green), environment (orange), demographics (purple), and kidney pathology (pink). An odds ratio of 1 equates to a median coefficient of zero, indicating that the variable was not significant in that LASSO run.

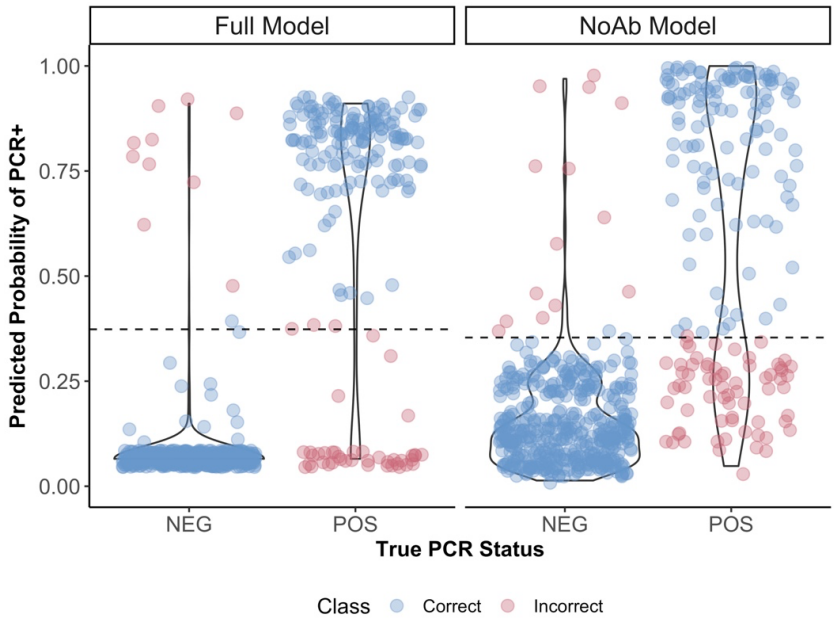


Figure 2.2: Distributions of predicted shedding probabilities relative to true PCR test results for models with and without antibody data available. The left panel shows predicted shedding probabilities from the model developed considering all covariates (Full model), and the right panel shows predicted shedding probabilities from the model developed with antibody titer data excluded (NoAb model). Predicted probability of shedding (y-axis) is shown for all animals relative to true PCR status (x-axis), with correctly classified individuals shown in blue and misclassified individuals shown in red. Predicted probability thresholds were selected to maximize model accuracy, above which animals were considered PCR positive. The optimal threshold estimated for the Full model was 0.37, while the optimal threshold for the NoAb model was 0.35 (horizontal dashed lines).

When analyzing the wild and stranded subsets with antibody titers excluded, we identified the same four significant predictors as for the aggregate dataset, but they were split with two significant predictors for each subset. In stranded sea lions, BUN and potassium were both identified as significant predictors, with odds ratios of 12.80 (95% CI: 3.34-157.68) and 0.3 (95% CI : 0.05 - 0.72), respectively. In wild sea lions, season (fall) and sex (male) were both significant, with odds ratios of 11.52 (95% CI : 2.52 - 49.66) and 2.48 (95% CI : 1.24 - 5.21), respectively. Potential predictors also echoed results from the aggregate training data, with sex (male) and BUN identified in stranded and wild sea lions, respectively.

2.5.1.4.2 Prediction Accuracy in Stranded and Wild Sea Lions

The accuracy of the Full model and the NoAb model (as parameterized against the aggregate dataset) did not differ significantly in these subsets (Table S 2.2). When antibody data were included, accuracy of the Full model was 93.9% (95% CI : 89.8 - 96.7%) in stranded sea lions and 90.1% (95% CI: 86.8 - 92.8%) in wild sea lions, which was comparable to accuracy across the aggregate dataset (Table S 2.2). When antibody data were excluded, the accuracy of the NoAb model was 87.3% (95% CI : 82.1-91.5%) in stranded sea lions and 86.9% (95% CI: 83.2 - 90.0%) in wild sea lions, which was again comparable to accuracy in the aggregate dataset (Table S 2.2). However, accuracy in wild sea lions appears to be driven largely by correct predictions for all of the wild animals that were PCR negative (n=343/404; 84.9%). Of the animals that were PCR positive, the model predicted 15% (n=9/61) correctly, while all of the others (n=52) had predicted probabilities of shedding that were non-zero but below the positivity threshold of 0.35 (median = 0.24; range 0.05-0.34).

2.5.1.4.3 Variable Selection in Seronegative Sea Lions

When seronegative animals (n=443) were assessed using all covariates, no significant predictors of shedding were identified. Only 36/443 (8%) of the seronegative animals were PCR positive, so the sample size of seronegative shedders may have been too small to detect any relationships that do exist. Season (fall) was the only potential predictor, though the magnitude of the median odds ratio was low (OR = 1.16, 95% CI : 1.00 - 2.31), so there is no evidence of a strong effect of season in these animals. Since no significant predictors were identified, prediction accuracy was not assessed in these animals.

2.5.2 Test Data

2.5.2.1 Out-of-sample Prediction Accuracy

Out-of-sample predictive accuracy was calculated by applying the Full model and NoAb model (both trained on the aggregate dataset from 2010-2018) to sea lion data from 2006-2009, comparing shedding predictions to known individual shedding status. Predictive accuracy of the Full model, calculated from animals that had antibody titer data available and a PCR test result (n=132/805), was 92.9% (95% CI 85.3-97.4%; Table S 2.2). Predictive accuracy for the NoAb model, calculated from animals that had serum chemistry and PCR test results available (n=77/1074), was 88.3% (95% CI 79.0-94.5%). For each model, out-of-sample predictive accuracy closely matched the predictive accuracy from the training dataset (Table S 2.2).

2.5.2.2 Extrapolation to Population Scale

After determining that the two models retained accuracy in the out-of-sample period, we applied them to back-project population-scale shedding prevalence in sea lions from 2006-2009. For each model, we estimated monthly shedding prevalence by compiling all shedding predictions from the broader test data, now including animals that did not have PCR results available. Despite using different covariates, the two models predicted similar population-scale patterns across seasons and across years (Figure 2.3A & B). Predictions from both models captured the known seasonality of leptospirosis outbreaks despite different seasonality in overall sample numbers, most notably the surges in 2007 and 2009, which were dominated by malnourished yearlings (Figure 2.3A & B). As a rough assessment of the accuracy of these population scale extrapolations, we compared them to data on monthly strandings due to leptospirosis at TMMC (Figure 2.3C). Predictions of both models aligned closely with the surveillance data, and both

accurately capture the 2009 season when total stranding numbers were high, but leptospirosis cases were low, with a corresponding lull in predicted shedding.

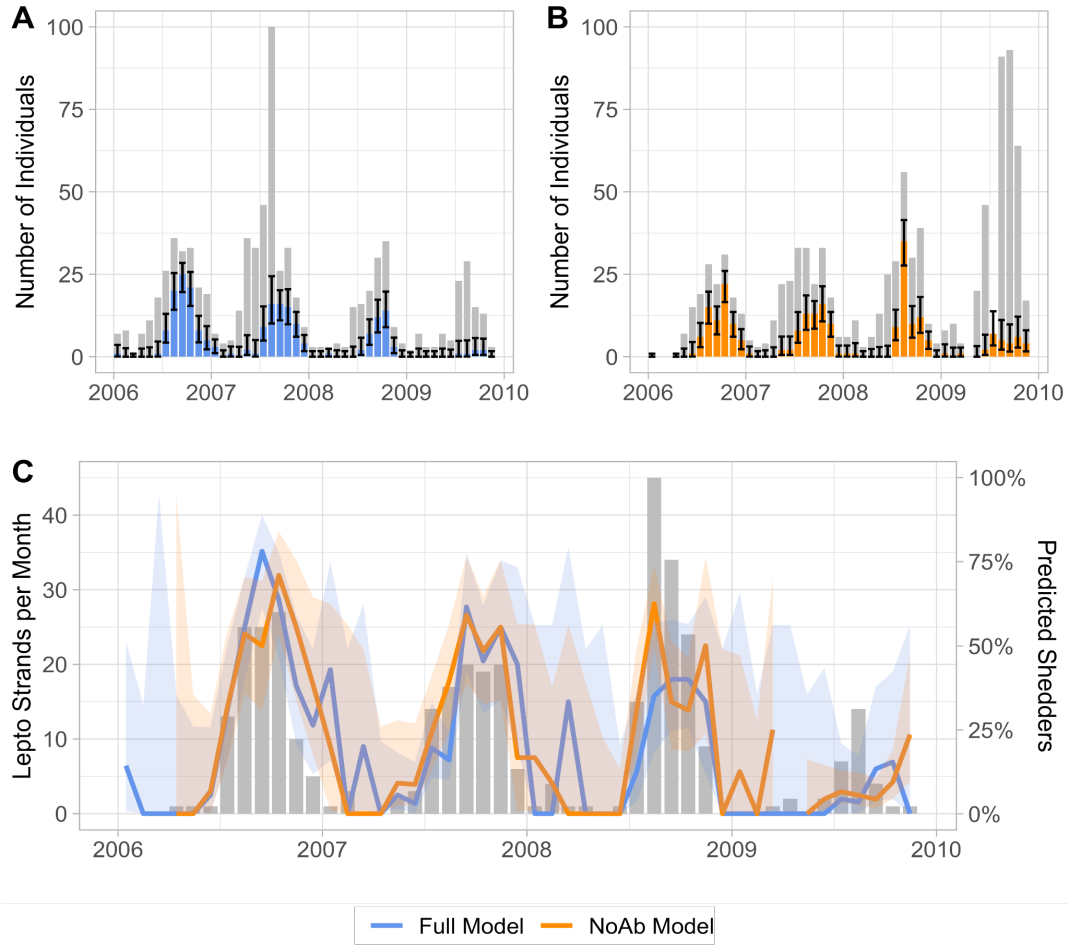


Figure 2.3: Out-of-sample predictions of monthly shedding prevalence in the sea lion population, 2006-2009. Using the Full model for animals with antibody titers available (A; n=805) and the NoAb model for animals with serum chemistry available (B; n=1074), we estimated the number of shedding individuals in each out-of-sample dataset. Animals predicted to be shedding by each model are shown as blue (A) and orange (B) filled bars, respectively, and monthly error bars denote the estimated 95% binomial confidence intervals. In these two panels (A & B), the filled gray bars show the number of samples tested by each model. The bottom panel (C) illustrates monthly strandings attributed to leptospirosis (gray bars; left y-axis) with simultaneous comparison of the proportion of shedders as predicted by the Full and NoAb models (blue and orange lines, respectively; right y-axis). The transparent colored ribbons denote the 95% binomial confidence interval for each model. The classifications from both models broadly track monthly sea lion leptospirosis strands at The Marine Mammal Center (n=382) during this time period.

2.6 DISCUSSION

Pathogen shedding is a critical component of infectious disease transmission but understanding drivers of shedding is a complex challenge that involves disentangling multiple host traits and environmental factors that influence host-pathogen interactions. We leveraged a long-term intensive study of *Leptospira* in California sea lions to demonstrate the potential to develop predictive models of shedding even in free-ranging wildlife hosts. Our analysis found that highly accurate predictions of individual shedding status can be made across a range of datasets and available candidate predictors. We confirmed that serum antibody titers are strong predictors of *Leptospira* shedding, which can yield >90% accurate predictions without the need for additional information. For the common circumstance when antibody data are not available, we demonstrate that almost equally accurate predictions of shedding status can be achieved using more readily available data. Remarkably, both models (with and without antibody titers) maintained their high accuracy when predicting shedding in out-of-sample data despite some differences in underlying data structure (Table 2.1). Encouraged by this success, we projected shedding estimates across the broader California sea lion population, obtaining patterns that aligned with epidemiological trends observed across a four-year period.

The significant predictors of *Leptospira* shedding identified in this study align with established knowledge of this host-pathogen system. We confirmed that antibody titers are highly accurate (91.4%) predictors of *Leptospira* shedding in sea lions (Prager et al., 2020), and that the addition of other covariates examined here did not significantly improve predictions. Indeed, the effect of this relationship is so strong that it wasn't until antibody titers were excluded that other covariates emerged as significant. When antibody data were excluded, significant predictors of

shedding (sex, season, BUN and potassium) were consistent with known patterns in leptospirosis cases, which have known associations with sex and season, and often experience serum chemistry alterations due to renal pathology (Greig et al., 2005; Gulland et al., 1996; Marcondes et al., 1996; Prager et al., 2020). Though antibody titers were far and away the most significant predictor of shedding, we were still able to produce highly accurate (87.2%) predictions by modeling this alternate set of significant covariates (Table S 2.2).

Despite known associations of oceanographic indices with marine productivity, sea lion demography and leptospirosis outbreak intensity (Borremans et al., In Preparation), season was the only environmental covariate that emerged as a significant driver of individual shedding (Table S 2.1). Season (fall) was a potential predictor when antibody titer data were included, and became significant in the absence of titer data; this makes sense in light of the strong seasonality of leptospirosis incidence (Buhnerkempe et al., 2017), and the fact that shedding is most prevalent during the initial weeks of infection. In the absence of antibody titer data, upwelling and spring transition arose as potential predictors of shedding in the aggregate dataset, and the full suite of oceanographic covariates were potential predictors for wild sea lions. Oceanographic perturbations disrupt marine productivity, and may cause negative impacts on sea lion body condition that contribute to shedding in some, but not all, animals (e.g., perhaps only in those that remain chronically infected). However, our results do not reveal any consistent impact of oceanographic conditions on shedding at the individual scale. The impact of oceanographic conditions on leptospirosis outbreak intensity (Borremans et al., In Preparation) may instead be mediated by population-level effects such as alterations in sea lion movement, mixing, or contact rates, or by impacting individual susceptibility to infection rather than pathogen shedding. Future

work is needed to investigate the mechanisms underlying environmental impacts on leptospirosis outbreak intensity in sea lions, and additional data may help to clarify the role of these potential predictors.

Our data for this analysis arose from animals sampled from two very different sources, so to assess the impact of sample population on our findings we examined predictors of shedding for stranded and wild sea lions separately. Significant coefficients were consistent with the findings from the aggregate training dataset, and antibody titers remained the most important predictor of shedding across all subgroups (Table S 2.1). When antibody data were excluded, the four variables that emerged as significant in the aggregate training dataset were evenly split between stranded and wild sea lions, with variables indicative of kidney disease (BUN and potassium) influencing shedding in stranded animals, and other variables (sex and season) influencing shedding in wild caught sea lions. This separation of predictors makes sense in light of selection conditions imposed on both groups. Stranded sea lions were affected by an injury or illness (e.g., clinical leptospirosis) severe enough to be admitted to TMMC, while wild sea lions were only caught, anesthetized and sampled if deemed apparently healthy on visual inspection. It was therefore unsurprising that kidney function was not a significant predictor of shedding in the apparently healthy wild sea lions. Conversely, signs of renal dysfunction were more predictive of *Leptospira* shedding in stranded animals (Prager et al., 2020), and it is possible that the heavy influence of kidney disease in stranded animals overshadowed the seasonal effect seen in the wild and aggregate datasets. Encouragingly, despite these differences, shedding prediction accuracy remained high in these subgroups (87 - 94%), demonstrating that these predictive models are broadly robust to differences in underlying sample populations.

Our study corroborated that a small proportion of seronegative sea lions were found to be shedding *Leptospira* (Table 2.1; Prager et al. 2020). Their lack of antibody titers makes it hard to predict shedding status in these animals, and indeed, this is the major source of prediction failures for our model that includes antibody data (Figure 2.2). Despite exploring a suite of possible shedding drivers in this group, no significant predictors were found, and it remains unclear what factors influence shedding in these animals. Previous work hypothesized that seronegative shedders may be either chronic carriers whose titer levels declined below detection, or animals in the early stages of infection whose immune systems hadn't yet mounted a robust antibody response (Prager et al. 2020). The chronic carrier theory is consistent with the fact that 64% of these animals (n=23/36) were sampled in 2012, the year following a large leptospirosis outbreak, though this explanation does not align well with age data since most of these animals were yearlings (n=13/23). The early infection theory is consistent with the fact that, though there were no significant predictors of seronegative shedding, season (fall) was a potential predictor in this group. This corresponds with the timing of incident leptospirosis cases, and therefore may represent a true effect if some of these individuals are in early stages of infection that have not yet mounted an antibody titer response (Table S 2.1). The hidden drivers of seronegative shedding may be uncovered by future work that increases the sample size of this group or focuses on other possible explanations, such as genetic or immune system markers that may be associated with chronic carriage and seronegative shedding.

This analysis has some limitations, pertaining to both the data and modeling methods. With regard to the data, the composition of the sampled population was not split evenly between stranded (n=213/617; 35%) and wild-caught (n=404/617; 65%) sea lions. In addition to these

datasets being more or less likely, respectively, to have symptomatic infections than a true random sample of the wild population, the stranded data were further biased since 60% of all stranded animals with PCR data (n=129/213) were clinical leptospirosis cases. These differences likely contributed to the differences in significant predictors between the two datasets, and had differences been more extensive this would have affected the accuracy of the aggregate models. Fortunately, our models were quite robust to these differences, as evidenced by the high accuracy obtained when testing the aggregate models against the data subsets. Though accuracy in the wild animals was driven largely by correct predictions of the non-shedding majority, improved identification of shedding in wild animals could potentially be achieved by adjusting the positivity threshold in these animals, so it is important to take your sample population into consideration when making these selections. There were also limitations in the way we developed the models. We examined a targeted set of 21 possible covariates (to avoid data dredging), but this list was not comprehensive, and the specific definitions used for some variables may have masked associations. For example, the quarterly classification we used for season separates late summer cases from the fall cluster, and a broader seasonal classification dividing the year in two (January - June and July - December) might be more likely to capture earlier seasonal peaks of leptospirosis in stranded animals (Buhnerkempe et al., 2017; Gulland et al., 1996). Environmental covariates reflected conditions at certain spatial coordinates which may not capture variability across the system, and time lags for environmental data and leptospirosis cases, which were incorporated in an attempt to reflect short and long-term timescales, may have missed underlying spatiotemporal patterns. While alternative choices could reveal associations not uncovered in this analysis, these selections were made in a best effort to capture patterns without overfitting by using all possible covariate divisions. Future work should

continue to consider data patterns and sample population characteristics in order to minimize potential bias in model inferences.

This study leveraged a long-term multidisciplinary research effort to develop predictive models of shedding for the globally significant zoonotic pathogen, *Leptospira interrogans.* Models generated highly accurate predictions of shedding status at the individual scale, which enabled shedding projections in broader sample populations that align with independently measured epidemiological patterns. This capacity for shedding prediction opens new opportunities to gain knowledge about shedding and transmission potential across both space (from sea lion rehabilitation centers with more limited resources) and time (from periods when research funding and shedding data were more limited). Understanding the drivers of shedding in host-pathogen systems is critical to disentangling transmission mechanisms, yielding benefits that are both fundamental and applied, from individual, population and ecosystem scales. In this system for example, these models could be used by clinicians to predict the shedding risk of individual animals in clinical settings to inform decisions on co-housing, or by ecologists or population managers at the population scale to direct future leptospirosis surveillance efforts or inform studies of epidemic dynamics in this species. More broadly, efforts to understand pathogen shedding have implications for cross-species transmission risk, as shedding is a direct contributor to the pathogen pressure that can cause spillover transmission to humans and other sympatric species (Plowright et al., 2017). This work builds on a growing body of research working to understand and predict pathogen shedding, highlighting the potential for the quantitative analysis of long-term ecological studies to obtain greater insights into transmission, host-pathogen dynamics, and disease control and prevention.

2.7 SUPPLEMENT

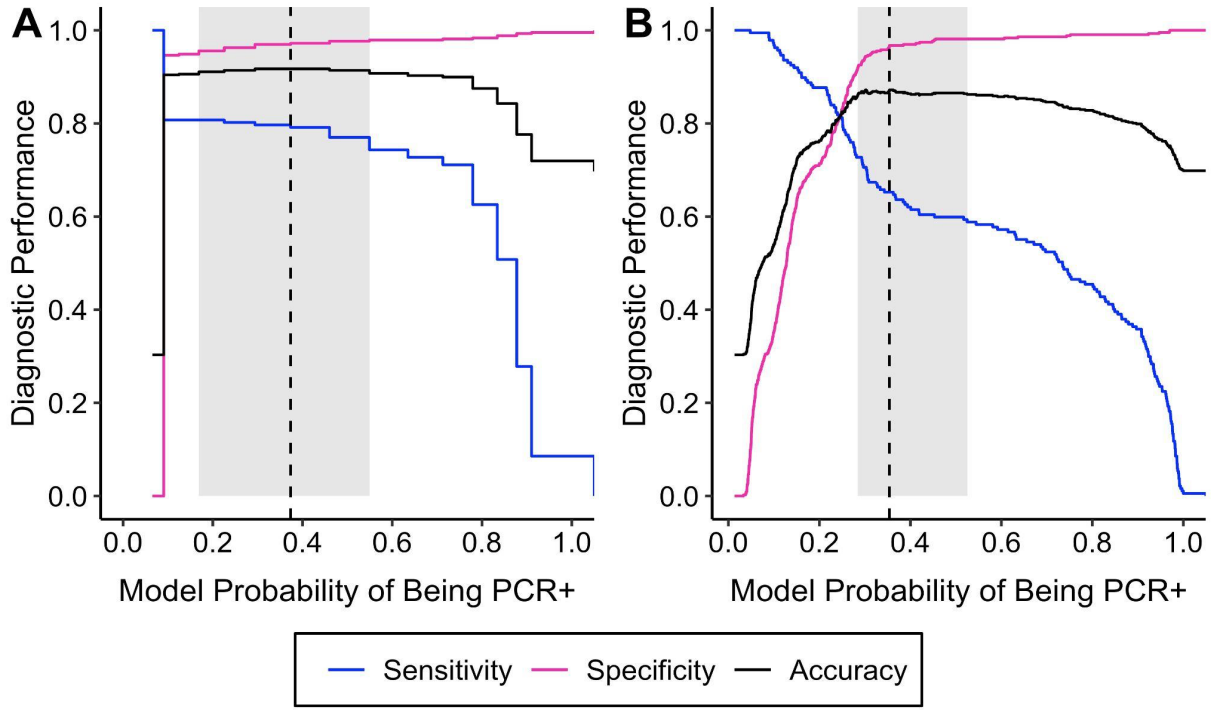


Figure S 2.1: Test diagnostic performance estimates vary depending on the probability threshold for PCR classification. The sensitivity (blue), specificity (pink), and accuracy (black) of each model varies depending on the probability threshold chosen for PCR classification. Accuracy (true classifications / total samples) is maximized at a threshold of 0.37 for the Full model (A) and 0.35 for the NoAb model (B), denoted by vertical dashed lines. The gray bars surrounding the thresholds represent predictive accuracy within 1% of that reached at the optimum (dashed vertical line), indicating a wide range of probability thresholds that would produce comparable results.

Table S 2.1: Estimated odds ratios from a LASSO regression of *Leptospira* shedding outcome in the training data. Estimated odds ratios from bootstrap LASSO regression coefficients (n=1000) for all covariates. Values were calculated using all training data with and without titers included (A & B), in addition to the following data subsets: stranded (C & D), wild (E & F), and seronegative (G) sea lions. All values were converted to odds ratios for ease of interpretation, and an odds ratio of 1 equates to a median coefficient of zero. Median odds ratio values and 95% CI are presented in the table for each coefficient distribution. Significant covariates (those with a median OR≠1 that did not include one in 95% CI) are shown in bold with a gray background.

		All Training Data		Training Data Subsets				
		A. Full Model Titers Included	B. NoAb Model Titers Excluded	C. Strands Titers Included	D. Strands Titers Excluded	E. Wild Titers Included	F. Wild Titers Excluded	G. Seronegatives Titers Included
Leptospira	Antibody Titers	**42.21 (20.35, 121.86)**	*NA*	**24.41 (9.40, 220.88)**	*NA*	**52.47 (1.49, 1549.81)**	*NA*	1.00 (1.00, 1.00)
	Lepto-3m	1.00 (0.32, 1.00)	1.00 (0.84, 1.33)	1.00 (1.00, 1.00)	1.05 (1.00, 2.68)	1.00 (0.15, 1.00)	1.00 (0.29, 1.00)	1.00 (0.07, 1.00)
	Lepto-12m	1.00 (1.00, 2.42)	1.00 (1.00, 1.72)	1.00 (1.00, 7.33)	1.00 (1.00, 2.64)	1.00 (0.27, 1.00)	1.00 (0.14, 1.00)	1.00 (1.00, 7.99)
Environmental	Fall	1.36 (1.00, 4.26)	**2.31 (1.20, 5.08)**	1.00 (0.36, 1.00)	1.00 (0.65, 1.50)	3.96 (1.00, 28.80)	**11.52 (2.52, 49.66)**	1.16 (1.00, 7.37)
	UPW-3m	1.00 (1.00, 3.58)	1.05 (1.00, 5.19)	1.00 (1.00, 1.00)	1.00 (1.00, 1.76)	1.10 (0.27, 40.12)	6.33 (0.04, 97.17)	1.00 (1.00, 2.31)
	UPW-12m	1.00 (0.22, 1.00)	0.74 (0.14, 1.00)	1.00 (0.29, 1.00)	1.00 (0.23, 1.00)	1.00 (0.05, 2.12)	0.13 (0.01, 4.68)	1.00 (0.49, 1.33)
	SST-3m	1.00 (0.43, 1.00)	1.00 (0.72, 1.14)	1.00 (0.26, 1.00)	1.00 (0.21, 1.00)	1.00 (0.45, 4.78)	2.08 (0.55, 12.24)	1.00 (0.51, 1.00)
	SST-12m	1.00 (0.60, 1.00)	1.00 (0.82, 1.00)	1.00 (1.00, 2.12)	1.12 (1.00, 5.58)	0.67 (0.25, 1.00)	0.84 (0.21, 1.00)	1.00 (0.53, 1.00)
	SpTr	1.00 (0.47, 1.00)	0.79 (0.41, 1.00)	1.00 (0.74, 1.00)	1.00 (0.89, 1.68)	1.00 (0.14, 1.00)	0.32 (0.09, 1.00)	1.00 (0.44, 1.00)
	SpTr-Lag	1.00 (0.63, 1.00)	1.00 (0.86, 1.27)	1.00 (0.42, 1.24)	1.00 (0.67, 1.18)	1.00 (0.06, 1.00)	0.67 (0.02, 1.00)	1.00 (0.52, 1.00)
Demographics	Male	1.53 (1.00, 3.48)	**2.96 (1.69, 5.77)**	1.02 (1.00, 5.49)	1.95 (1.00, 5.93)	1.35 (1.00, 3.26)	**2.48 (1.24, 5.21)**	1.00 (1.00, 2.02)
	Juvenile	1.00 (0.95, 1.47)	1.00 (1.00, 1.71)	1.00 (1.00, 2.61)	1.21 (1.00, 4.47)	1.00 (1.00, 1.76)	1.00 (0.97, 2.27)	1.00 (0.87, 1.00)
	Subadult	1.00 (1.00, 1.56)	1.00 (1.00, 2.40)	1.00 (1.00, 2.86)	1.00 (1.00, 2.58)	1.00 (1.00, 1.00)	1.00 (0.68, 1.00)	1.00 (1.00, 1.00)
	Yearling	1.00 (0.82, 1.21)	1.00 (0.57, 1.00)	1.00 (0.30, 1.00)	1.00 (0.48, 1.00)	1.00 (0.91, 1.42)	1.00 (0.40, 763.01)	1.00 (1.00, 1.05)
Kidney Pathology	BUN	1.00 (1.00, 7.84)	**35.48 (12.20, 384.64)**	1.85 (1.00, 18.25)	**12.80 (3.34, 157.68)**	1.00 (1.00, 1.00)	26.16 (1.00, 57881.60)	1.00 (1.00, 3.89)
	Phosphorus	1.00 (1.00, 1.72)	1.00 (1.00, 2.83)	1.00 (1.00, 5.20)	2.70 (1.00, 20.26)	1.00 (0.48, 1.00)	1.00 (0.15, 1.00)	1.00 (0.39, 1.00)
	Creatinine	1.00 (0.37, 1.00)	1.00 (0.25, 1.00)	1.00 (0.30, 1.00)	1.00 (0.31, 1.00)	1.00 (0.58, 1.00)	1.00 (0.36, 1.00)	1.00 (1.00, 1.00)
	Calcium	1.00 (1.00, 1.94)	1.00 (1.00, 1.88)	1.00 (1.00, 4.78)	2.58 (1.00, 9.24)	1.00 (0.16, 1.00)	0.95 (0.12, 1.54)	1.00 (1.00, 1.39)
	Sodium	1.00 (1.00, 1.07)	1.00 (1.00, 1.64)	1.00 (1.00, 1.56)	1.00 (1.00, 2.73)	1.00 (0.30, 5.46)	1.00 (0.32, 4.12)	1.00 (1.00, 1.00)
	Potassium	1.00 (0.48, 1.00)	**0.43 (0.10, 0.86)**	0.81 (0.20, 1.00)	**0.30 (0.05, 0.72)**	1.00 (1.00, 16.44)	1.00 (0.25, 8.31)	1.00 (1.00, 1.95)
	Chloride	1.00 (0.44, 1.00)	1.00 (0.38, 1.00)	1.00 (0.25, 1.00)	1.00 (0.23, 1.00)	1.00 (1.00, 264.68)	2.09 (1.00, 20.90)	1.00 (0.78, 1.00)

Table S 2.2: Model Accuracy in training and test data. Total accuracy ((true positives + true negatives) / total animals) is shown for the model developed with antibody titers included (Full model; left) and without antibody titers (NoAb model; right). The overall accuracy for the Full model and the NoAb model were based on predicted shedding probability thresholds of 0.37 and 0.35 respectively.

	Full Model	**NoAb Model**
Training Data (2010-2018)	**Accuracy (95% CI)**	**Accuracy (95% CI)**
Full (n=617)	91.4% (88.9-93.5%)	87.2% (84.4-89.8%)
Wild (n=404)	90.1% (86.8-92.8%)	86.9% (83.2-90.0%)
Strands (n=213)	93.9% (89.8-96.7%)	87.3% (82.1-91.5%)
Test Data (2006-2009)		
Animals with Antibody Data (n=805)	92.9% (85.3-97.4%)	
Animals with Serum Chemistry Data (n=1074)		88.3% (79.0-94.5%)

2.8 REFERENCES

Adler, B., & de la Peña Moctezuma, A. (2010). *Leptospira* and leptospirosis. *Veterinary Microbiology*, *140*(3–4), 287–296. https://doi.org/10.1016/j.vetmic.2009.03.012

Barragan, V., Olivas, S., Keim, P., & Pearson, T. (2017). Critical Knowledge Gaps in Our Understanding of Environmental Cycling and Transmission of *Leptospira* spp. *Applied and Environmental Microbiology*, *83*(19), e01190-17. https://doi.org/10.1128/AEM.01190-17

Becker, D. J., Eby, P., Madden, W., Peel, A. J., & Plowright, R. K. (2021). *Ecological conditions experienced by bat reservoir hosts predict the intensity of Hendra virus excretion over space and time* [Preprint]. Ecology. https://doi.org/10.1101/2021.08.19.457011

Blanco, G., & Díaz de Tuesta, J. A. (2021). Seasonal and spatial occurrence of zoonotic Salmonella serotypes in griffon vultures at farmland environments: Implications in pathogen pollution and ecosystem services and disservices. *Science of The Total Environment*, *758*, 143681. https://doi.org/10.1016/j.scitotenv.2020.143681

Bograd, S. J., Schroeder, I., Sarkar, N., Qiu, X., Sydeman, W. J., & Schwing, F. B. (2009). Phenology of coastal upwelling in the California Current. *Geophysical Research Letters*, *36*(1), 1–5. https://doi.org/10.1029/2008GL035933

Borremans, B., Pepin, K. M., Prager, K. C., Greig, D. J., Gomez, A. C., Melin, S. R., Laake, J. L., Lowry, M. S., DeLong, R. L., Grenfell, B. T., Gulland, F. M. D., & Lloyd-Smith, J. O. (In Preparation). *Environment and immunity drive disease outbreaks in California sea lions*.

Buhnerkempe, M. G., Prager, K. C., Strelioff, C. C., Greig, D. J., Laake, J. L., Melin, S. R., DeLong, R. L., Gulland, F. M., & Lloyd-Smith, J. O. (2017). Detecting signals of chronic shedding to explain pathogen persistence: *Leptospira interrogans* in California sea lions. *Journal of Animal Ecology*, *86*(3), 460–472. https://doi.org/10.1111/1365-2656.12656

Caley, P., & Hone, J. (2004). Disease transmission between and within species, and the implications for disease control: Disease transmission in ferrets. *Journal of Applied Ecology*, *41*(1), 94–104. https://doi.org/10.1111/j.1365-2664.2004.00867.x

Casanovas-Massana, A., Pedra, G. G., Wunder, E. A., Diggle, P. J., Begon, M., & Ko, A. I. (2018). Quantification of *Leptospira interrogans* Survival in Soil and Water Microcosms. *Applied and Environmental Microbiology*, *84*(13), e00507-18. https://doi.org/10.1128/AEM.00507-18

Charleston, B., Bankowski, B. M., Gubbins, S., Chase-Topping, M. E., Schley, D., Howey, R., Barnett, P. V., Gibson, D., Juleff, N. D., & Woolhouse, M. E. J. (2011). Relationship Between Clinical Signs and Transmission of an Infectious Disease and the Implications for Control. *Science*, *332*(6030), 726–729. https://doi.org/10.1126/science.1199884

Chavez, F. P., Ryan, J., Lluch-cota, S. E., C, M. Ñ., & C, M. N. (2003). From anchovies to sardines and back: Multidecadal change in the Pacific ocean. *Science*, *299*(5604), 217–221. https://doi.org/10.1126/science.1075880

Corsi, S. R., Borchardt, M. A., Carvin, R. B., Burch, T. R., Spencer, S. K., Lutz, M. A., McDermott, C. M., Busse, K. M., Kleinheinz, G. T., Feng, X., & Zhu, J. (2016). Human and Bovine Viruses and Bacteria at Three Great Lakes Beaches: Environmental Variable Associations and Health Risk. *Environmental Science & Technology*, *50*(2), 987–995. https://doi.org/10.1021/acs.est.5b04372

Costa, F., Hagan, J. E., Calcagno, J., Kane, M., Torgerson, P., Martinez-Silveira, M. S., Stein, C., Abela-Ridder, B., & Ko, A. I. (2015). Global Morbidity and Mortality of Leptospirosis: A Systematic Review. *PLOS Neglected Tropical Diseases*, *9*(9), e0003898. https://doi.org/10.1371/journal.pntd.0003898

Costa, F., Wunder, E. A., De Oliveira, D., Bisht, V., Rodrigues, G., Reis, M. G., Ko, A. I., Begon, M., & Childs, J. E. (2015). Patterns in *Leptospira* Shedding in Norway Rats (Rattus norvegicus) from Brazilian Slum Communities at High Risk of Disease Transmission. *PLOS Neglected Tropical Diseases*, *9*(6), e0003819. https://doi.org/10.1371/journal.pntd.0003819

DeLong, R. L., Antonelis, G. A., Oliver, C. W., Stewart, B. S., Lowry, M. C., & Yochem, P. K. (1991). Effects of the 1982–83 El Niño on Several Population Parameters and Diet of California Sea Lions on the California Channel Islands. In F. Trillmich & K. A. Ono (Eds.), *Pinnipeds and El Niño* (Vol. 88, pp. 166–172). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-76398-4_18

DeLong, R. L., Melin, S. R., Laake, J. L., Morris, P., Orr, A. J., & Harris, J. D. (2017). Age- and sex-specific survival of California sea lions (*Zalophus californianus*) at San Miguel Island, California. *Marine Mammal Science*, *33*(4), 1097–1125. https://doi.org/10.1111/mms.12427

Faine, S., Adler, B., Bolin, C., & Perolat, P. (1999). *Leptospira and leptospirosis* (2nd ed.). MediSci. https://research.monash.edu/en/publications/leptospira-and-leptospirosis

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, *33*(1). https://doi.org/10.18637/jss.v033.i01

Germeraad, Sanders, Hagenaars, Jong, Beerens, & Gonzales. (2019). Virus Shedding of Avian Influenza in Poultry: A Systematic Review and Meta-Analysis. *Viruses*, *11*(9), 812. https://doi.org/10.3390/v11090812

Gili, C., Meijer, G., & Lacave, G. (2018). *EAZA and EAAM Best Practice Guidelines for Otariidae and Phocidae (Pinnipeds)*. https://www.eaza.net/assets/Uploads/CCC/EAZA-EAAM-PinnipedGuidelines-approved.pdf

Greig, D. J., Gulland, F. M. D., & Kreuder, C. (2005). A Decade of Live California Sea Lion (*Zalophus californianus*) Strandings Along the Central California Coast: Causes and Trends, 1991-2000. *Aquatic Mammals*, *31*(1), 11–22. https://doi.org/10.1578/am.31.1.2005.11

Gulland, F. M. D., Koski, M., Lowenstine, L. J., Colagross, A., Morgan, L., & Spraker, T. (1996). Leptospirosis in California sea lions (*Zalophus californianus*) stranded along the central California coast, 1981-1994. *Journal of Wildlife Diseases*, *32*(4), 572–580. https://doi.org/10.7589/0090-3558-32.4.572

Haake, D. A., & Levett, P. N. (2015). Leptospirosis in Humans. In B. Adler (Ed.), *Leptospira and Leptospirosis* (Vol. 387, pp. 65–97). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-45059-8_5

Hagan, J. E., Moraga, P., Costa, F., Capian, N., Ribeiro, G. S., Wunder, E. A., Felzemburgh, R. D. M., Reis, R. B., Nery, N., Santana, F. S., Fraga, D., dos Santos, B. L., Santos, A. C., Queiroz, A., Tassinari, W., Carvalho, M. S., Reis, M. G., Diggle, P. J., & Ko, A. I. (2016). Spatiotemporal Determinants of Urban Leptospirosis Transmission: Four-Year Prospective Cohort Study of Slum Residents in Brazil. *PLOS Neglected Tropical Diseases*, *10*(1), e0004275. https://doi.org/10.1371/journal.pntd.0004275

Hastie, P., Tibshirani, R., & Wainwright, M. (2016). *Statistical Learning with Sparsity: The Lasso and Generalizations*. CRC press.

Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed). Springer.

Hernandez, S. M., Welch, C. N., Peters, V. E., Lipp, E. K., Curry, S., Yabsley, M. J., Sanchez, S., Presotto, A., Gerner-Smidt, P., Hise, K. B., Hammond, E., Kistler, W. M., Madden, M., Conway, A. L., Kwan, T., & Maurer, J. J. (2016). Urbanized White Ibises (*Eudocimus albus*) as Carriers of Salmonella enterica of Significance to Public Health and Wildlife. *PLOS ONE*, *11*(10), e0164402. https://doi.org/10.1371/journal.pone.0164402

Ko, A. I., Goarant, C., & Picardeau, M. (2009). *Leptospira:* The dawn of the molecular genetics era for an emerging zoonotic pathogen. *Nature Reviews Microbiology*, *7*(10), 736–747. https://doi.org/10.1038/nrmicro2208

Krstajic, D., Buturovic, L. J., Leahy, D. E., & Thomas, S. (2014). Cross-validation pitfalls when selecting and assessing regression and classification models. *Journal of Cheminformatics*, *6*(1), 10. https://doi.org/10.1186/1758-2946-6-10

Lau, C. L., Smythe, L. D., Craig, S. B., & Weinstein, P. (2010). Climate change, flooding, urbanisation and leptospirosis: Fuelling the fire? *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *104*(10), 631–638. https://doi.org/10.1016/j.trstmh.2010.07.002

Levett, P. N. (2001). Leptospirosis. *Clinical Microbiology Reviews*, *14*(2), 296–326. https://doi.org/10.1128/CMR.14.2.296-326.2001

Lloyd-Smith, J. O., George, D., Pepin, K. M., Pitzer, V. E., Pulliam, J. R. C., Dobson, A. P., Hudson, P. J., & Grenfell, B. T. (2009). Epidemic Dynamics at the Human-Animal Interface. *Science*, *326*(5958), 1362–1367. https://doi.org/10.1126/science.1177345

Lloyd-Smith, J. O., Greig, D. J., Hietala, S., Ghneim, G. S., Palmer, L., St Leger, J., Grenfell, B. T., & Gulland, F. M. (2007). Cyclical changes in seroprevalence of leptospirosis in California sea lions: Endemic and epidemic disease in one host species? *BMC Infectious Diseases*, *7*, 125–136. https://doi.org/125 10.1186/1471-2334-7-125

Marcondes, M., Seguro, A. C., Abdulkader, R. C. R. M., Burdmann, E. A., & Malheiro, P. S. (1996). Peculiar Electrolytic and Hormonal Abnormalities in Acute Renal Failure Due to Leptospirosis. *The American Journal of Tropical Medicine and Hygiene*, *54*(1), 1–6. https://doi.org/10.4269/ajtmh.1996.54.1

Melin, S. R., DeLong, R. L., & Siniff, D. B. (2008). The effects of El Niño on the foraging behavior of lactating California sea lions (*Zalophus californianus*) during the nonbreeding season. *Canadian Journal of Zoology*, *86*, 192–206. https://doi.org/10.1139/Z07-132

Melin, S. R., Laake, J. L., DeLong, R. L., & Siniff, D. B. (2012). Age-specific recruitment and natality of California sea lions at San Miguel Island, California (No. 4). 28(4), 776.

Melin, S. R., Orr, A. J., Harris, J. D., Laake, J. L., DeLong, R. L., Gulland, F. M. D., & Stoudt, S. (2010). Unprecedented mortality of California sea lion pups associated with anomalous oceanographic conditions along the central California coast in 2009. *California Cooperative Ocean and Fisheries Investigations Report 2009*, *51*, 182–194.

Munywoki, P. K., Koech, D. C., Agoti, C. N., Kibirige, N., Kipkoech, J., Cane, P. A., Medley, G. F., & Nokes, D. J. (2015). Influence of age, severity of infection, and co-infection on the duration of respiratory syncytial virus (RSV) shedding. *Epidemiology and Infection*, *143*(4), 804–812. https://doi.org/10.1017/S0950268814001393

Nakata, M., Effler, P. V., Gaynor, K., Katz, A. R., Park, S. Y., & Clark, T. A. (2007). LEPTOSPIROSIS ON OAHU: AN OUTBREAK ASSOCIATED WITH FLOODING OF A UNIVERSITY CAMPUS. *The American Journal of Tropical Medicine and Hygiene*, *76*(5), 882–886. https://doi.org/10.4269/ajtmh.2007.76.882

Plowright, R. K., Parrish, C. R., McCallum, H., Hudson, P. J., Ko, A. I., Graham, A. L., & Lloyd-Smith, J. O. (2017). Pathways to zoonotic spillover. *Nature Reviews Microbiology*, *15*(8), 502–510. https://doi.org/10.1038/nrmicro.2017.45

Prager, K. C., Alt, D. P., Buhnerkempe, M. G., Greig, D. J., Galloway, R. L., Wu, Q., Gulland, F. M. D., & Lloyd-Smith, J. O. (2015). Antibiotic Efficacy in Eliminating Leptospiruria in California Sea Lions (*Zalophus californianus*) Stranding with Leptospirosis. *Aquatic Mammals*, *41*(2), 203–212. https://doi.org/doi: 10.1578/AM.41.2.2015.203

Prager, K. C., Borremans, B., Mummah, R. O., Gomez, A. C. R., Helman, S. K., Galloway, R. L., Alt, D. P., Wu, Q., DeLong, R. L., Melin, S., Harris, J. D., Orr, A. J., Burco, J., Brown, M., Laake, J. L., Gulland, F. M. D., Greig, D. J., & Lloyd-Smith, J. O. (In Preparation). *Fadeout and reemergence of an endemically circulating pathogen in a wildlife host*.

Prager, K. C., Buhnerkempe, M. G., Greig, D. J., Orr, A. J., Jensen, E. D., Gomez, F., Galloway, R. L., Wu, Q., Gulland, F. M. D., & Lloyd-Smith, J. O. (2020). Linking longitudinal and cross-sectional biomarker data to understand host-pathogen dynamics: *Leptospira* in California sea lions (*Zalophus californianus*) as a case study. *PLOS Neglected Tropical Diseases*, *14*(6), e0008407. https://doi.org/10.1371/journal.pntd.0008407

Prager, K. C., Greig, D. J., Alt, D. P., Galloway, R. L., Hornsby, R. L., Palmer, L. J., Soper, J., Wu, Q., Zuerner, R. L., Gulland, F. M. D., & Lloyd-Smith, J. O. (2013). Asymptomatic and chronic carriage of *Leptospira interrogans* serovar Pomona in California sea lions (*Zalophus californianus*). *Veterinary Microbiology*, *164*(1–2), 177–183. https://doi.org/10.1016/j.vetmic.2013.01.032

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rasmussen, M. A., & Bro, R. (2012). A tutorial on the Lasso approach to sparse modeling. *Chemometrics and Intelligent Laboratory Systems*, *119*, 21–31. https://doi.org/10.1016/j.chemolab.2012.10.003

Siembieda, J. L., Miller, W. A., Byrne, B. A., Ziccardi, M. H., Anderson, N., Chouicha, N., Sandrock, C. E., & Johnson, C. K. (2011). Zoonotic pathogens isolated from wild animals and environmental samples at two California wildlife hospitals. *Journal of the American Veterinary Medical Association*, *238*(6), 773–783. https://doi.org/10.2460/javma.238.6.773

Siva-Jothy, J. A., & Vale, P. F. (2021). Dissecting genetic and sex-specific sources of host heterogeneity in pathogen shedding and spread. *PLOS Pathogens*, *17*(1), e1009196. https://doi.org/10.1371/journal.ppat.1009196

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*(1), 267–288. JSTOR.

Trotz-Williams, L. A., Martin, S. W., Leslie, K. E., Duffield, T., Nydam, D. V., & Peregrine, A. S. (2008). Association between management practices and within-herd prevalence of Cryptosporidium parvum shedding on dairy farms in southern Ontario. *Preventive Veterinary Medicine*, *83*(1), 11–23. https://doi.org/10.1016/j.prevetmed.2007.03.001

Williams, E., & Barker, I. (2001). *Infectious Diseases of Wild Mammals* (3rd ed.). Iowa State University Press.

Wu, Q., Prager, K., Goldstein, T., D.P., A., Galloway, R. L., Zuerner, R. L., Lloyd-Smith, J., & Schwacke, L. (2014). Development of a real-time PCR for the detection of pathogenic

Leptospira spp. In California sea lions. *Diseases of Aquatic Organisms*, *110*(3), 165–172. https://doi.org/10.3354/dao02752

Zhang, J., Jun, T., Frank, J., Nirenberg, S., Kovatch, P., & Huang, K. (2021). Prediction of individual COVID-19 diagnosis using baseline demographics and lab data. *Scientific Reports*, *11*(1), 13913. https://doi.org/10.1038/s41598-021-93126-7

Zuerner, R. L., & Alt, D. P. (2009). Variable Nucleotide Tandem-Repeat Analysis Revealing a Unique Group of *Leptospira interrogans* Serovar Pomona Isolates Associated with California Sea Lions. *Journal of Clinical Microbiology*, *47*(4), 1202–1205. https://doi.org/10.1128/JCM.01639-08

3 An investigation of *Leptospira* in Southern California wildlife

3.1 ABSTRACT

This study contributes to cross-species surveillance efforts for the multi-host zoonotic pathogen *Leptospira interrogans* in the wildlife of coastal California. *Leptospira interrogans* serovar Pomona has affected California sea lions (*Zalophus californianus*) and Channel Island terrestrial mammals for years, and mainland terrestrial wildlife species remain a candidate source for observed pathogen introductions in local marine and island ecosystems. Though broad *Leptospira* surveys have been conducted in mainland mammals in coastal and northern California, less is known about *Leptospira* prevalence in southern California. To address this knowledge gap, we assess prevalence levels and risk factors for *Leptospira* in five common mammal species across the greater Los Angeles region: striped skunks (*Mephitis mephitis*), Northern raccoons (*Procyon lotor*), coyotes (*Canis latrans*), Virginia opossums (*Didelphis virginiana*), and fox squirrels (*Sciuris niger*). Three of these species are native to the Los Angeles region (striped skunks, raccoons and coyotes), while two are introduced (opossums and fox squirrels). Additional species and regions were sampled whenever possible. All five species exhibited exposure to *Leptospira* with seroprevalences ranging from 5-60%, and infections were detected in all species except fox squirrels, with infection prevalences ranging from 0.8-15.2%. Individuals from all host species were serologically reactive to serovar Pomona, with skunks, raccoons and coyotes reacting most strongly to this serovar, indicating that these mesocarnivores may contribute to a broader multi-host reservoir for this pathogen in coastal California. Sample distribution across a heterogeneous landscape provided us with the additional opportunity to assess potential risk factors for *Leptospira* in each species. None of the covariates explored here (age, sex, rainfall, or land class distribution within individual home range) were significant

drivers of exposure. This finding, together with *Leptospira* exposure or infection detected in all species of mammals sampled, indicates that *Leptospira* is present across the southern California landscape, and therefore poses a risk to both animal and human health in this region.

3.2 INTRODUCTION

Leptospirosis, the disease caused by pathogenic bacteria from the genus *Leptospira,* is the most widespread zoonotic pathogen in the world (Adler and de la Peña Moctezuma 2010; Fouts et al. 2016). Though many pathogens, and all zoonotic pathogens, are capable of infecting multiple hosts, studying complex generalist pathogen systems remains a challenge, particularly in wildlife since multiple lines of data are required but resources and sample access are often limited (Viana et al., 2014). *Leptospira* is among the most generalist pathogens known to science, and has been detected in almost every species of mammal that has been tested (Adler & de la Peña Moctezuma, 2010). This bacteria is listed as an emerging pathogen of concern for humans (Levett, 2001), and is estimated to cause a million cases and 60,000 human deaths annually (Costa et al., 2015; Munoz-Zanzi et al., 2020). Transmission typically occurs indirectly when the mucus membranes of a susceptible host come in contact with a urine-contaminated environment (Monahan et al., 2008). Alternate routes of transmission also occur, including vertical or pseudo-vertical transmission (in utero or through milk), sexual transmission, or direct contact with contaminated host tissues or urine (Ellis, 2015; Harrison & Fitzgerald, 1988; Minter et al., 2017). Humans may be more at risk if they work in close contact with animals (e.g., abattoir workers and veterinarians), and prevention measures such as protective clothing and rodent control have been emphasized as awareness increases surrounding this veterinary and public health concern (Hartskeerl et al., 2011).

Leptospires are motile aerobic spirochetes comprising more than 12 pathogenic species, which are further classified into over 250 pathogenic serovars based on the expression of cell surface antigens and serologic reactivity (Adler & de la Peña Moctezuma, 2010). Antigenically related serovars were traditionally grouped into serogroups, but neither serovar nor serogroup are reliable predictors of *Leptospira* species (Levett, 2015). Serovars are said to be adapted to specific maintenance host species that are typically less susceptible to severe disease (acquiring asymptomatic or flu-like infections), and maintenance host individuals can become chronic carriers which excrete leptospires in urine for extended periods (e.g., up to 60 weeks in cattle; Leonard et al., 1992). In contrast, when susceptible species referred to as accidental hosts acquire the disease, more severe infections can occur with symptoms ranging from asymptomatic to severe kidney failure and death. Current infection status can be tested from urine or kidney samples, using either polymerase chain reaction (PCR) to detect *Leptospira* DNA or cultures to detect live infectious leptospires. Past exposure to *Leptospira* (i.e., evidence of anti-*Leptospira* antibodies) is assessed using serum microscopic agglutination testing (MAT), the most widely used diagnostic test for *Leptospira* spp. (Adler & de la Peña Moctezuma, 2010; Faine et al., 1999), with MAT panels typically including multiple serovars. Within an individual host, cross-reacting antibodies can be detected by MAT, manifesting as detection of antibody titers against multiple different serovars included in an MAT panel (Blanco et al., 2016). The highest MAT titer from a given individual is often taken to reflect the serovar (or broader serogroup) of the causative infection, but this relationship is known to be imperfect (André-Fontaine & Triger, 2018). Further work is needed to better understand the relationship between serology and infection, particularly in wildlife (Pedersen et al., 2018).

In the United States, roughly 100-150 human cases of leptospirosis are reported annually, with the majority occurring in Puerto Rico and Hawaii (Centers for Disease Control and Prevention, 2018). Livestock, domestic dogs, and horses are also routinely vaccinated for, tested for, and diagnosed with *Leptospira* infections and leptospirosis, with many infections associated with wildlife-domestic animal spillover (Blessington et al., 2020; Davis et al., 2008; Gautam et al., 2010). There is increasing awareness of leptospirosis in domestic dogs, who, like humans, are often infected by wildlife (Moore et al., 2006; Smith et al., 2021; Ward et al., 2004; White et al., 2017). A reduction of urinary shedding remains a key objective of canine vaccination efforts (Schreiber et al., 2005), which have reduced the risk of human transmission and severe cases in children in the United States (Brown & Prescott, 2008). A previous survey conducted by the United States Department of Agriculture (USDA) demonstrated that *Leptospira* titers are common in a variety of wildlife species across the country, though more regional work is needed to understand the risk this poses to the health of humans, domestic pets and livestock (Pedersen et al., 2018).

Extensive work has been done exploring *Leptospira* in coastal California wildlife, where both marine and terrestrial island mammals have been significantly impacted (Buhnerkempe et al., 2017; Colagross-Schouten et al., 2002; Greig et al., 2005; Gulland et al., 1996; Lloyd-Smith et al., 2007; Mummah, 2021; Prager et al., 2020; Zuerner et al., 2009). *Leptospira interrogans* serovar Pomona has been circulating in California sea lions (*Zalophus californianus)* since the mid-1980s (Greig et al., 2005; Gulland et al., 1996). Sea lions present with symptoms ranging from asymptomatic chronic carriage to severe acute cases and fatalities (Lloyd-Smith et al., 2007; Prager et al., 2013, 2015; Buhnerkempe et al., 2017; Prager et al., 2020), demonstrating

characteristics of both reservoir and accidental hosts with their endemic and epidemic cycles (Lloyd-Smith et al., 2007; Zuerner et al., 2009). A strain of *Leptospira interrogans* serovar Pomona, closely related to that seen in sea lions, is known to circulate in Channel Island foxes (*Urocyon littoralis*) and spotted skunks (*Spilogale gracilis amphiala*; Lloyd-Smith, 2021; Mummah, 2021). Phylogenetic analyses of *Leptospira* genomes isolated from sea lions, island foxes and spotted skunks show evidence of repeated introductions of new strains of serovar Pomona into the broader coastal ecosystem, but it remains unclear to what degree mainland and terrestrial wildlife contribute to past or ongoing cross-ecosystem transmission in this multi-host pathogen system (Borremans et al., 2019; Mummah, 2021).

In mainland California, cases of leptospirosis are reported in humans, domestic animals, and wildlife annually, with an average of 7 human cases per year between 2017-2019 (California Department of Public Health, 2020; Gulland et al., 1996; Hennebelle et al., 2014; Meites et al., 2004). It has been considered a potentially reemerging human pathogen in the region, and while reported human cases often originate out of state, it is likely underdiagnosed (Meites et al., 2004). A broad historical wildlife survey reported infections or evidence of prior exposure in multiple species, including bobcats (*Lynx rufus*), coyotes (*Canis latrans*), Northern raccoons (*Procyon lotor*), and striped skunks (*Mephitis mephitis*; Cirone et al., 1978), and other surveys have detected *Leptospira* in individual wildlife species such as black bears (*Ursus americanus*; Ruppanner et al., 1982) and wild pigs (*Sus scrofa*; Clark et al., 1983). *Leptospira* has been reported in southern California animals, including dogs (Greene, 1941), deer (Roug et al., 2012), and wild felids (Straub et al., 2021), though surveillance has been more thorough in more northern regions of California. Spatiotemporal patterns in northern domestic dogs revealed

spatial clusters of *L. interrogans* serovar Pomona, suggesting that wildlife contact is an important route of exposure (Hennebelle et al., 2013, 2014). A recent large-scale survey of wildlife in northern California detected *Leptospira interrogans* serovar Pomona in numerous host species (Straub & Foley, 2020). Mesocarnivores were recently demonstrated to play a substantial role in *Leptospira* circulation in California, and suggested that skunks and raccoons are potential reservoir hosts (Straub et al., 2020). Despite the comprehensive studies of *Leptospira* in coastal and northern California, there are many outstanding questions regarding the ecology of this pathogen, particularly concerning the prevalence levels in different host species of southern California wildlife.

Disease ecology in urban ecosystems presents a particular set of challenges, owing to distinct ecological pressures on host communities, and higher human densities offering opportunities for zoonotic spillover. Urbanization affects wildlife in a variety of ways, including changes in resource use, higher exposure to toxicants, and alterations in community structure, contact rates, and movement barriers (Riley et al., 2014). All of these factors have the potential to influence pathogen dynamics and, crucially, the risk of transmission for zoonotic pathogens. Globally, urbanization has been highlighted as a potential risk factor for leptospirosis. Increased urbanization and climate change can lead to increased flood risk and more favorable environmental conditions for *Leptospira* survival and transmission (Lau et al., 2010). Past work on *Leptospira* in cities has often focused on rodents, which are known to be key hosts of specific serovars, in high-density urban centers and informal settlements (Boey et al., 2019; Costa et al., 2015; Ko et al., 1999; Minter et al., 2018). There has been much less attention on other urban wildlife species, and connections to landscapes surrounding cities, despite the fact that

mesocarnivores and other urban-adapted species may play a significant role in the epidemiology of this multi-host pathogen (Straub et al., 2020; Straub & Foley, 2020). With more than half of the global population currently residing in urban areas (Neiderud, 2015), urbanization is becoming increasingly pervasive, making it crucial to understand and mitigate the impacts of anthropogenic change on wildlife health. Ongoing work to understand the impacts of land use modifications on public health risks is particularly important. The diverse landscape of greater Los Angeles features a range of environments, from natural areas and agricultural land to a dense urban center. In addition to the ten million humans that live in the region (United States Census Bureau, 2020), many sympatric wildlife species are also potential carriers of *Leptospira* bacteria, making it imperative to investigate the prevalence of this pathogen across host species in greater Los Angeles.

We conducted the first in-depth surveillance of *Leptospira* in wildlife in the greater Los Angeles area. We assessed the prevalence of *Leptospira* exposure and active infections in mesocarnivores and rodents represented by five common mammal species, with additional non-target species and regions sampled opportunistically to complement existing knowledge of this pathogen in California. We examine patterns in serologic reactivity across species to gain insights into circulating serovars, with emphasis on illuminating the broader reservoir of *L. interrogans* serovar Pomona in southern California. The greater Los Angeles region additionally provides a unique opportunity to assess this multi-host pathogen, and potentially multiple infecting serovars, across a complex urban landscape. We explore a variety of possible predictors for *Leptospira* exposure, including land use, allowing us to evaluate how landscape may impact the prevalence of this bacteria in individual species throughout a diverse urbanization matrix.

Assessing the prevalence of and risk factors for this multi-host zoonotic pathogen in Los Angeles wildlife has implications for the public health of over ten million people and will additionally inform both veterinary and wildlife management agencies that manage the health of animals across the region.

3.3 METHODS

3.3.1 Study Animals

This study focused primarily on wildlife from the greater Los Angeles region in Southern California. Opportunistic sample collection was approved by the California Department of Fish and Wildlife under scientific collecting permits SC-13267 & SC-13700 and took place from September 2015 to June 2020. For the purposes of this study, the greater Los Angeles region refers to Los Angeles County and surrounding counties: Orange, Riverside, San Bernardino, and Ventura. Sample collection focused on the following five common mammals in the Los Angeles region: striped skunk, Northern raccoon, coyote, Virginia opossum (*Didelphis virginiana*), and fox squirrel (*Sciuris niger*). These five species will be referred to as the core five (Table 3.1, Figure 3.1). Collaborating agencies that donated carcasses or existing samples include the California Wildlife Center (CWC), the Los Angeles County Department of Animal Care and Control (LADACC), the United States Department of Agriculture Animal and Plant Health Inspection Service (USDA APHIS), the Department of Defense (DOD) and the National Park Service (NPS). The majority of samples came from animals that were deceased following vehicle collisions, planned wildlife removal, or animals that were euthanized by animal control or rehabilitation agencies due to illness or injury. Carcasses were necropsied immediately or frozen at -20ºC and thawed in a refrigerator prior to necropsy. Animal measurements and

demographic information were collected at the time of necropsy, with age class (adult or juvenile) determined using a combination of animal size and tooth wear (Grau et al., 1970).

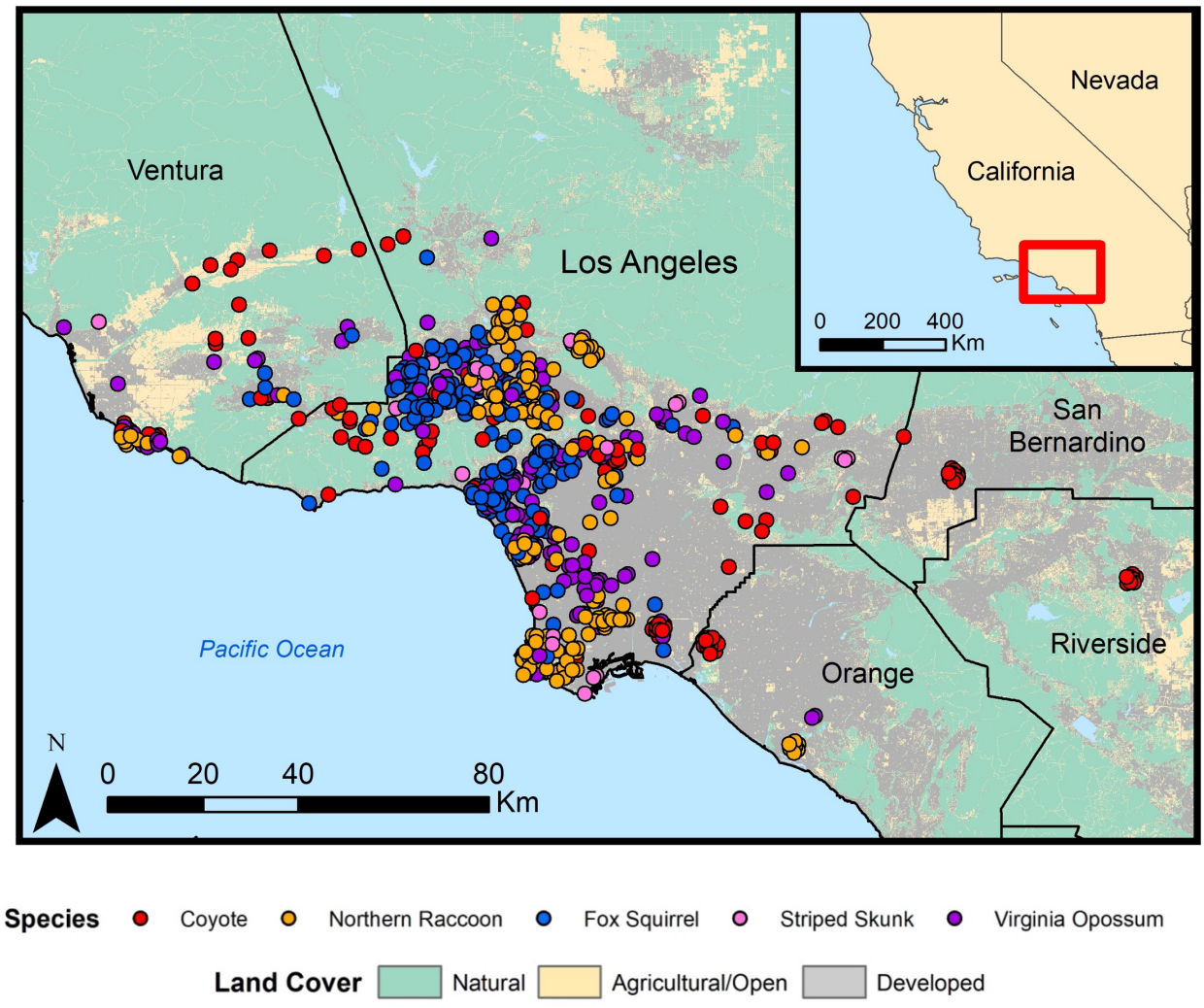


Figure 3.1: Distribution of land classes and *Leptospira* sample locations for the core five species of wildlife in the greater Los Angeles region. Land classification data was obtained from the National Land Cover Database (2019). Data collection occurred from 2015-2020.

Table 3.1: Descriptive characteristics of the core five species of wildlife. Within group percentages are proportions of column totals.

	Coyote (N=137)	Northern Raccoon (N=172)	Squirrel (Fox) (N=187)	Striped Skunk (N=40)	Virginia Opossum (N=171)	Total (N=707)
Age Class						
Adult	59 (43.1%)	133 (77.3%)	148 (79.1%)	20 (50%)	132 (77.2%)	492 (69.6%)
Juvenile	29 (21.2%)	32 (18.6%)	24 (12.8%)	13 (32.5%)	27 (15.8%)	125 (17.7%)
Unknown	49 (35.8%)	7 (4.1%)	15 (8%)	7 (17.5%)	12 (7%)	90 (12.7%)
Sex						
Female	57 (41.6%)	83 (48.3%)	79 (42.2%)	16 (40%)	86 (50.3%)	321 (45.4%)
Male	59 (43.1%)	80 (46.5%)	75 (40.1%)	13 (32.5%)	69 (40.4%)	296 (41.9%)
Unknown	21 (15.3%)	9 (5.2%)	33 (17.6%)	11 (27.5%)	16 (9.4%)	90 (12.7%)
County						
Los Angeles	81 (59.1%)	152 (88.4%)	182 (97.3%)	38 (95%)	147 (86%)	600 (84.9%)
Orange	15 (10.9%)	9 (5.2%)	0 (0%)	0 (0%)	3 (1.8%)	27 (3.8%)
Riverside	8 (5.8%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	8 (1.1%)
San Bernardino	11 (8%)	0 (0%)	0 (0%)	0 (0%)	1 (0.6%)	12 (1.7%)
Ventura	22 (16.1%)	11 (6.4%)	5 (2.7%)	2 (5%)	20 (11.7%)	60 (8.5%)
Season						
Dry	60 (43.8%)	53 (30.8%)	81 (43.3%)	24 (60%)	73 (42.7%)	291 (41.2%)
Wet	76 (55.5%)	119 (69.2%)	105 (56.1%)	13 (32.5%)	96 (56.1%)	409 (57.9%)
Unknown	1 (0.7%)	0 (0%)	1 (0.5%)	3 (7.5%)	2 (1.2%)	7 (1%)

3.3.2 Sample Collection

Serum and urine samples from external agencies were analyzed and included when available. In fresh carcasses, intracardiac blood was collected into serum separation tubes, then kept in a cooler with an ice pack until centrifugation using an Ample Scientific Champion E-33 benchtop centrifuge (1350 x g for 10-15 minutes). Kidney samples were collected from all animals that underwent a necropsy, and urine was collected when available using cystocentesis. The largest possible kidney sample that would fit in a 58 ml Whirl-Pak® was excised (approximate size: entire kidney from smaller mammals such as squirrels or half a kidney from larger mammals such as coyotes) and homogenized in the sealed Whirl-Pak® using manual pressure. Serum and

urine samples were transferred into cryovials prior to storage, and all cryovials and Whirl-Paks® were stored at -20ºC or -80ºC prior to testing (-80ºC preferred when space was available).

3.3.3 Leptospira Serology

Past exposure to *Leptospira* was assessed by using microscopic agglutination testing (MAT). In this test, dark-field microscopy is used to assess the presence of anti-*Leptospira* antibodies in serum by evaluating agglutination (i.e., clumping) when samples are combined with live cultures of *Leptospira* species (Faine et al., 1999). Serum samples are tested at doubling dilutions, and the reported endpoint titers represent the highest dilution that achieved a 50% agglutination using the reference strain being tested. Samples from 2015-2017 were run at the California Animal Health and Food Safety Laboratory (CAHFS; Davis, California, USA) using their panel of 6 common serovars: Bratislava, Canicola, Grippotyphosa, Hardjo, Icterohaemorrhagiae, and Pomona. Samples from 2017-2019 were analyzed at the Centers for Disease Control and Prevention (CDC; Atlanta, Georgia, USA) using an expanded 20 serovar panel: Alexi, Australis, Autumnalis, Ballum, Bataviae, Borinca, Bratislava, Canicola, Celledoni, Cynopteri, Djasiman, Georgia, Grippotyphosa, Icterohaemorrhagiae, Javanica, Mankarso, Pomona, Pyrogenes, Tarassovi, Wolffi. To assess consistency between these two laboratories, a subset of wildlife samples was tested at both laboratories (n=469), demonstrating 98.3% agreement in seropositivity and minor quantitative differences between titers.

3.3.4 Leptospira PCR Analysis

Leptospira infections were assessed using a quantitative polymerase chain reaction (qPCR) assay targeting the LipL32 gene as detailed in Wu et al. (2014). Pathogenic *Leptospira* DNA was

assessed in sterile urine (collected by cystocentesis) and homogenized kidney samples. Testing was conducted at either the Hollings Marine Laboratory in Charleston, South Carolina, USA or Colorado State University Veterinary Diagnostic Laboratory in Denver, Colorado, USA. Samples run at Colorado State University were additionally tested using the VetMAX^{TM} qPCR Master Mix kit, using VetMAX^{TM} suggested reagent volumes in conjunction with the primer and probe concentrations as specified by Wu et al. (2014). Samples that had a cycle threshold less than 37 were considered positive.

3.3.5 Data Analysis

Prevalence was estimated for *Leptospira* exposure and infections. For *Leptospira* serology, individuals could be reactive to multiple serovars, so seroprevalence was calculated by species in two ways: proportion positive against any serovar, and proportion positive within each specific serovar. All 95% binomial confidence intervals were estimated using package 'PropCIs' in program *R* using version 3.6.1 (R Core Team, 2021). Additional analyses were done in *R* using version 3.6.1 (R Core Team, 2021), and maps were created using ArcGIS version 10.8.2 (ESRI Inc., 2017).

Annual rainfall at each sample location was estimated from high resolution (1km^2) precipitation data (Fick and Hijmans, 2017). Land cover analysis was conducted as detailed in Adducci II et al. (2020) using freely available land cover data from the National Land Cover Database (US. Geological Survey (USGS), 2019). The localities of our samples were spread across a landscape gradient, ranging from natural vegetation and open space to the urban core of the city (Figure 3.1)To account for land cover variation within home ranges, we extracted 2019 land cover data

(30 x 30m resolution) from home range buffers around each georeferenced sampling locality using the 'raster' and 'rgdal' packages in *R* (R Core Team, 2021). Buffer size varied by species based on home range estimates previously reported in the literature: 5 km^2 for coyotes (Adducci II et al., 2020), 2 km^2 for raccoons and striped skunks (Šálek et al., 2015), 1 km^2 for opossums (Wright et al., 2012), and 0.5 km^2 for fox squirrels (Prince et al., 2014). As detailed in Adducci II et al. (2020) we grouped land cover classifications into three composite categories: urban/suburban land (20-100% impervious surface cover), agricultural/open land (made up of agricultural and open development zones with <20% impervious surface cover), and natural land (zones dominated by shrublands, forest, grassland and wetland). We then calculated the relative proportions of these three land classes for each individual buffer, and used ternary plots to show *Leptospira* exposure (i.e., presence of antibodies) relative to land class categories (Figure S 3.1).

We used logistic regression to explore potential predictors of *Leptospira* exposure (as indicated by an antibody titer of 1:100 or higher to any serovar). The following covariates were considered: age class, sex, annual rainfall, season (wet Nov-April; dry May-Oct), and composite land class category (developed vs. agricultural/open vs. natural). Antibody titers have been identified as significant predictors of PCR status in California sea lions (Prager et al., 2020). We therefore explored this association in raccoons, the only species with a sufficient number of paired antibody-PCR samples available (n=81). Since this data exhibited complete separation (i.e., a clear distinction between the two outcome levels), we applied a Firth's bias-reduced logistic regression (Firth, 1993) using the 'logistf' package in *R* to assess the association between antibody titers and PCR data. A Firth's regression is a penalized maximum likelihood approach, which is effective in the presence of data separation (Firth, 1993).

3.3.6 Additional Species & Regions

On an opportunistic basis, we also tested samples provided by other agencies that included non-target regions of California and non-target species, including ground squirrels (*Otospermophilus beecheyi*), desert cottontails (*Sylvilagus audubonii*), feral pigs, bobcats, gray foxes (*Urocyon cinereoargenteus*), and red foxes (*Vulpes vulpes*). A total of 255 additional samples were collected from September 2015 thru May 2020. This included non-target species in the greater Los Angeles region (n=93), and all species from the following additional counties: Monterey (n=86), San Luis Obispo (n=61), Santa Barbara (n=15; Table S 3.1).

3.4 RESULTS

3.4.1 Leptospira Serology to Detect Exposure

We detected evidence of *Leptospira* exposure in all five core species sampled in the greater Los Angeles region (Figure 3.2, Table 3.2). We first considered overall seroprevalence in each species, calculated as the proportion positive against any serovar. Fox squirrels had the highest overall seroprevalence at 60.6% (n=66/109, 95% CI: 50.7-69.8; Table 3.2), though titer levels in this species were typically low, and 65.2% (n=43/66) of seropositive squirrels had maximum titers measured at 1:100 or 1:200 (Table 3.2). Other species had lower seroprevalence levels; positivity in raccoons was 32.6% (n=31/95, 95% CI: 23.4-43.0), followed by striped skunks at 28.6% (n=6/21; 95% CI: 11.3-52.2), coyotes at 25.5% (n=14/55, 95% CI: 14.7-39.0), and opossums at 5.2% (n=5/97; 95% CI: 1.7-11.6). These levels were similar to seroprevalence levels reported in northern California by Straub and Foley (2020), who detected antibodies in 42% of fox squirrels (n=15/36; 95% CI: 27-58), 44% of raccoons (n=52/119; 95% CI: 35-53), 38% of striped skunks (n=78/206; 95% CI: 32-45), 30% of coyotes (n=6/20; 95% CI :15-52),

and 6% of opossums (n=2/32; 95% CI: 2-20). To gain insight into the strains giving rise to these overall levels of exposure, we first considered only the focal serovars which were tested for all animals in the study (serovars Bratislava, Canicola, Grippotyphosa, Icterohaemorrhagiae, and Pomona). We examined all positive MAT results to determine the full range of titer magnitude, the frequency of antibody detection against specific serovars, and the presence of antibody cross-reactivity within each of the five core host species (Figure 3.3, Table S 3.2). All host species were serologically reactive to multiple serovars, and serovar Pomona was detected in all species (Figure 3.3). We then determined the serovar with the maximum MAT titer in each individual as the best available indication of infecting serovar (Table 3.3; André-Fontaine & Triger, 2018). Antibody titers against serovar Pomona were most frequently the highest in skunks, raccoons, and coyotes (100%, 85%, and 57%; Table 3.3), consistent with recent work in northern California that identified serovar Pomona as predominant in these species (Straub & Foley, 2020). Fox squirrels exhibited a clearly distinct pattern, with highest titers most often to serovar Icterohaemorrhagiae (71%), which was not highly reactive in any of the other species. Of the two opossums that were reactive to this panel, one individual (the only opossum in this study with an active infection) had a maximum titer to serovar Pomona (1:12800). We emphasize that these maximum titer patterns do not give definitive information on the infecting serovar; confirmation would require genetic analysis of a culture isolate.

We then considered MAT results for serovars that were not tested at both laboratories, so results are available for only a subset of animals (Table S 3.2). The serovar with the highest proportion of positive titers varied by species, with skunks, raccoons, coyotes, opossums and fox squirrels respectively testing positive most often to serovars Autumnalis (42.9%; n=3/7), Pomona (42.5%;

n=54/127), Autumnalis (27.9%; n=12/43), Hardjo (9.3%; n=7/75) and Hardjo (66.7%; n=34/51). We noted that titers against serovars Pomona and Autumnalis were strongly correlated in coyotes (Spearman's rho = 0.81), which is consistent with a similar correlation identified between these serovars in domestic dogs (Moore et al., 2006). Though Autumnalis was most frequently positive in skunks and coyotes, Pomona was not significantly lower and comprised a larger sample size (and total that tested positive) in both species, with 29.2% (n=7/24) and 19.7% (n=14/71) of animals testing positive, respectively. Furthermore, in Channel Island foxes we have found that animals infected with serovar Pomona (as confirmed by genetic analysis of an isolate) frequently have higher MAT titers against serovar Autumnalis than against serovar Pomona (Mummah et al., In Preparation). In aggregate, the data from coyotes and skunks are consistent with serovar Pomona being the major strain causing their infections. An additional seven opossums were reactive to serovar Hardjo, making this the most commonly positive serovar overall in this species, though Pomona was still the highest peak titer in this species (1:12800) in the animal with an active infection (Table S 3.2). Aside from the low titer reactions to serovar Hardjo that more than doubled the overall seroprevalence of opossums, the majority of our conclusions did not change with the consideration of the expanded serovar panel, with Icterohaemorrhagiae still predominant in squirrels and Pomona still considered predominant in all other species (Table S 3.2; Mummah et al., In Preparation).

Table 3.2: *Leptospira* exposure and infection results by species. *Leptospira* antibody (MAT) and DNA (PCR) results in the five core species sampled in the greater Los Angeles region. Antibody results include seropositives to all serovars tested.

	***Leptospira* Exposure (Antibodies)**			***Leptospira* Infections (PCR)**		
Species	**POS**	**n**	**%POS (95% CI)**	**POS**	**n**	**%POS (95% CI)**
Striped Skunk	6	21	28.6 (11.3-52.2)	5	33	15.2 (5.1-31.9)
Northern Raccoon	31	95	32.6 (23.4-43.0)	14	162	8.6 (4.8-14.1)
Coyote	14	55	25.5 (14.7-39)	4	107	3.7 (1.0-9.3)
Virginia Opossum	5	97	5.2 (1.7-11.6)	1	131	0.8 (0.0-4.2)
Fox Squirrel	66	109	60.6 (50.7-69.8)	0	148	0 (0.0-2.5)

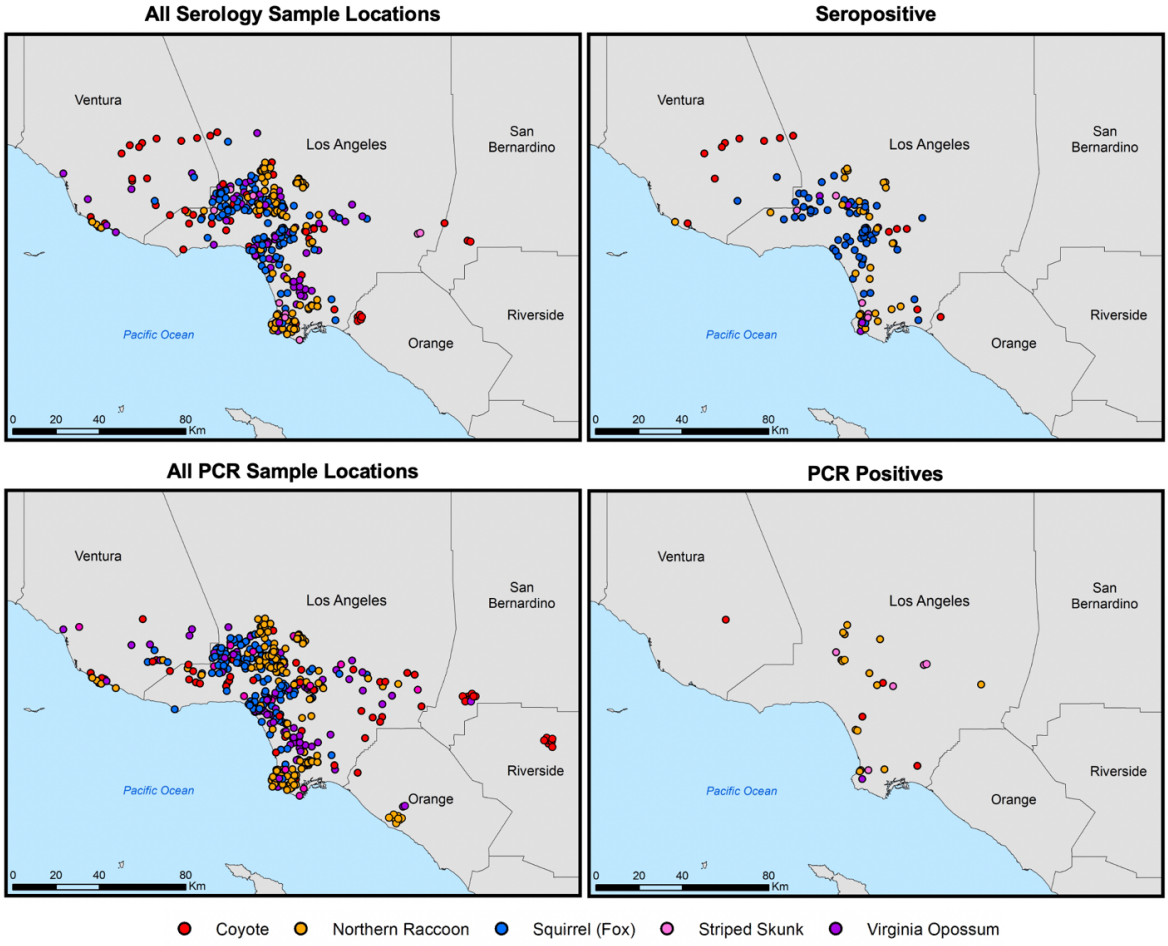


Figure 3.2: Location of sample collection and associated *Leptospira* exposure and infection results for the five core wildlife species. Host sample locations are indicated in the left-hand panels, with the locations of positive results shown in maps on the right. Samples tested by MAT for serum antibodies are shown in the top row, and those tested by PCR for *Leptospira* DNA are shown on the bottom row.

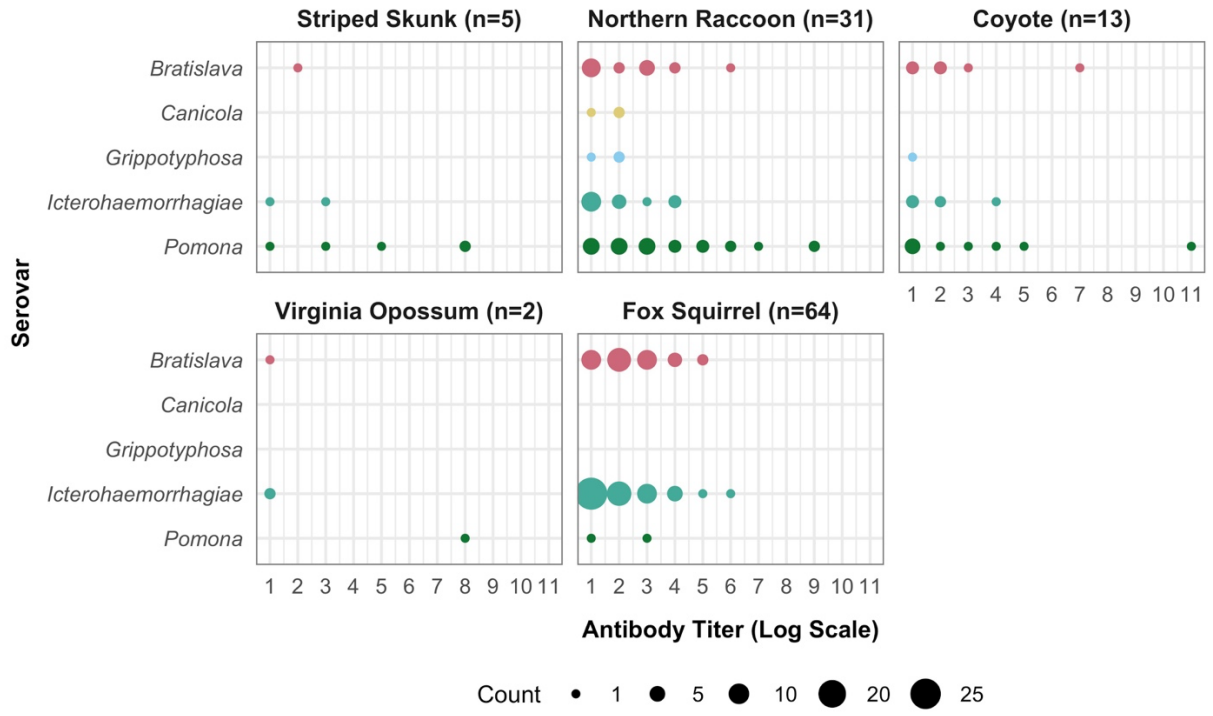


Figure 3.3: Samples positive for antibodies shown by serovar, species and titer level. Positive antibody results (MAT) for each of the five serovars that were tested at both laboratories. Antibody titer (x-axis) is shown on a log scale (1:100 equivalent to 1, 1:200 equivalent to 2, etc.).

Table 3.3: Maximum antibody titers for the five core wildlife species. Maximum antibody titers (MAT) for each of the five serovars tested at both laboratories (serovars Bratislava, Canicola, Grippotyphosa, Icterohaemorrhagiae, and Pomona), reported for the five core species sampled in the greater Los Angeles region. Serovars that are not shown never had a maximum titer. In cases where there were ties for maximum titer, both serovars were counted in the table.

Common Name	Scientific Name	Serovar	100	200	400	800	1600	3200	≥6400	# Positive/ Total	Percentage (95% CI)
Striped Skunk	*Mephitis mephitis*	Pomona	1		1		1		2	5/5	100 (47.8-100)
Northern Raccoon	*Procyon lotor*	Bratislava	1			1				2/27	7.41 (0.9-24.3)
		Icterohaemorrhagiae	1			1				2/27	7.41 (0.9-24.3)
		Pomona	5	3	5	3	2	2	3	23/27	85.19 (66.3-95.8)
Coyote	*Canis latrans*	Bratislava	1	1						2/7	28.57 (3.7-71)
		Icterohaemorrhagiae		1						1/7	14.29 (0.4-57.9)
		Pomona		1	1		1		1	4/7	57.14 (18.4-90.1)
Virginia Opossum	*Didelphis virginiana*	Icterohaemorrhagiae	1							1/2	50 (1.3-98.7)
		Pomona							1	1/2	50 (1.3-98.7)
Fox Squirrel	*Sciurus niger*	Bratislava		2	5	3	1			11/38	28.95 (15.4-45.9)
		Icterohaemorrhagiae	19	2	2	3		1		27/38	71.05 (54.1-84.6)

3.4.2 Leptospira PCR to Detect Active Infections

Infections were detected in all species except fox squirrels, with the prevalence of infection within species consistently lower than corresponding seroprevalence levels (Table 3.2, Figure 3.2). The lack of active infections detected in fox squirrels was surprising because seroprevalence was highest in this species. Infection prevalence ranged from 0.8% in opossums (n=1/132; 95% CI : 0.0-4.2) to 15% in skunks (n=5/34; 95% CI : 5.1-31.9), with coyotes and raccoons both intermediate at 3.7% (n=4/107; 95% CI : 1.0-9.3) and 8.6% (n=14/163; 95% CI : 4.8-14.1), respectively. Infection levels detected here were lower than reported in northern California by Straub and Foley (2020), who detected positive results in 28% of striped skunks (n=40/141; 95% CI: 22-36), 26% of raccoons (n=23/87; 95% CI : 18-37), 100% of coyotes (n=2/2; 95% CI : 34-100), 17% of opossums (n=1/6; 95% CI : 3-56), and 13% fox squirrels (n=4/31; 95% CI : 5-29).

3.4.3 Data Analysis

Leptospira exposure was detected throughout the sampled ranges of each host species (Figure 3.2). When we evaluated sample location relative to the composite land cover classes, we could see indications that different species were using the landscape in different ways. For instance, fox squirrel and opossum samples were clustered around areas with higher levels of human development, providing evidence for increased use of urban and suburban regions in these non-native species (Figure S 3.1). In contrast, coyotes and raccoons were found across all land classes. When we evaluated *Leptospira* exposure data in light of these land classes, no clear patterns emerged to distinguish the locations of positive and negative samples (Figure 3.2, Figure S 3.1), indicating that *Leptospira* circulates throughout the sampled range of each host species.

To probe how exposure patterns are influenced by landscape type alongside seasonal and demographic factors, we used logistic regressions to assess possible correlates for *Leptospira* exposure in all species except skunks, which were excluded due to small overall sample size (n=21). Of the covariates explored here (age, sex, rainfall, county, season, and composite land class), none exhibited significant correlations with *Leptospira* exposure in any of the core species (all p-values > 0.12 in univariate analysis). The lack of correlation between *Leptospira* exposure and composite land classification aligns with what is seen in the ternary plots (Figure S 3.1), supporting that *Leptospira* is distributed throughout the sampled range of these species.

In Chapter 2, I demonstrated that MAT titers can be effective predictors of *Leptospira* shedding in California sea lions, which is consistent with previous findings (Prager et al., 2020). Based on this finding, we used paired titer-PCR results to analyze the association between maximum antibody titer against any serovar and active infection, for which only raccoons had a sufficient sample size (n=81). Of the 81 animals that had paired samples available, 29 were seropositive, and 86% of seropositive animals (n=25/29) had maximum titers to serovar Pomona. We found that individual maximum titers were significant predictors of shedding status in this species (Firth's logistic regression; p-value = 2.09×10^{-9}). Individuals with titers above log 5 (1:1600) were predicted to be highly likely to be actively shedding *Leptospira* (Figure 3.4).

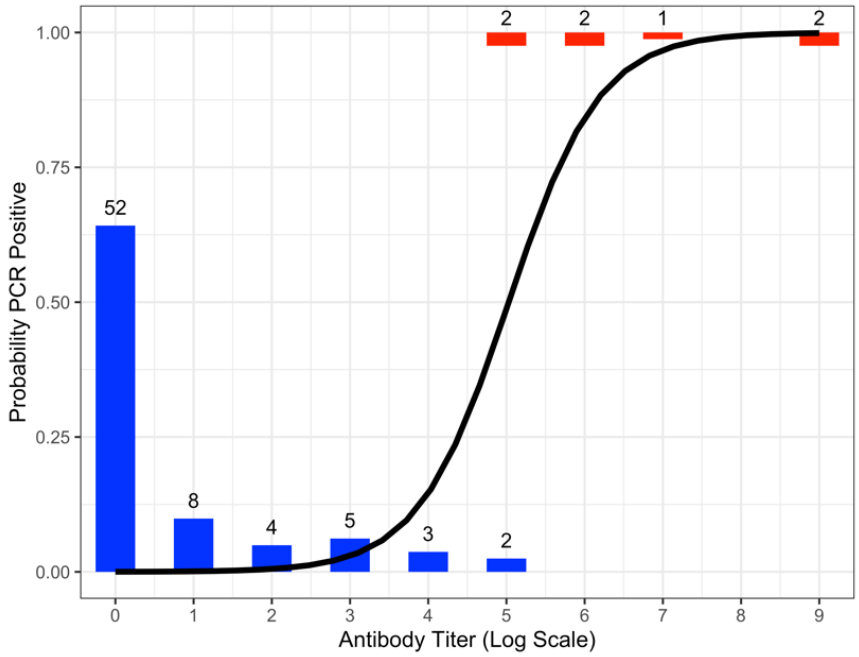


Figure 3.4: The predicted probability of raccoons being PCR positive relative to maximum antibody titer to any serovar. When paired antibody-PCR data were fit with a Firth's logistic regression (black line), individuals with titers greater than 1:1600 (Log 5) are predicted to be at least 80% likely to be PCR positive.

3.4.4 Additional Species & Regions

The non-core dataset (n=241 from 12 species) was comprised of samples from non-target counties (n=153) and species (n=143). Though none of the 72 animals tested by PCR were positive, antibodies were detected in 31.6% of non-core animals (n=61/19), with seroprevalence levels in individual species ranging from 0% to 100% (Table S 3.1). Of the non-target species in the greater Los Angeles region, 10% (n=3/38) of desert cottontails were seropositive, with maximum titers (ranging from 1:100 to 1:800) split between serovars Georgia, Icterohaemorrhagiae, and Pomona (Table S 3.2). Of the bobcats tested, 45% (n=5/11) were seropositive, with titers to Pomona and Bratislava most frequently positive (Table S 3.2). Wild pigs, all from San Luis Obispo (Table S 3.1), were more frequently reactive to serovars

Bratislava (20%; n=8/40), Autumnalis (12.5%; 5/40), Djasiman (10%; 4/40) and Pomona (7.5%; 3/40; Table S 3.2).

3.5 DISCUSSION

We conducted a large-scale survey of *Leptospira interrogans* in mainland terrestrial mammals in California, focusing on the understudied region of southern California, with two goals: 1) to identify the prevalence, potential risk factors and public health risks associated with this bacteria in the greater Los Angeles region, and 2) to assess serological patterns to inform our knowledge of broader multi-host circulation of *Leptospira* in coastal California wildlife. We identified *Leptospira* exposure in all target species sampled and detected active infections in most species. Widespread evidence of exposure, along with the fact that exposure patterns were not correlated with any tested risk factors, highlight that this pathogen is endemic and circulating throughout the sampled range of these wildlife species.

In our core five species (skunks, raccoons, coyotes, opossums and fox squirrels) sampled in the greater Los Angeles area, we detected low to moderate levels of *Leptospira* infection (0-15%) and markedly higher seroprevalence levels (5-60%; Table 3.2). As serovar Pomona has been commonly detected in the broader coastal California ecosystem, we focused on identifying hosts that were reactive to this serovar of interest. Seroreactivity patterns detected here provide further evidence that serovar Pomona is common in California wildlife (Mummah, 2021; Zuerner et al., 2009), particularly mesocarnivores (Straub et al., 2020; Straub & Foley, 2020), indicating that these terrestrial species may be playing a role in the better-studied circulation of serovar Pomona in the broader coastal ecosystem (Lloyd-Smith et al., 2007; Mummah, 2021; Prager et al., 2013).

Antibody titers to serovar Pomona were detected in all species, though it is likely that some of these detections reflect cross-reactivity among serovars in the MAT assay. We focused on which serovar had the highest titer, as a rough, though not foolproof, proxy for the serovar causing infection (Table 3.3; André-Fontaine & Triger, 2018). Among serovars tested for all animals, serovar Pomona was the most common maximum titer in coyotes, raccoons, and skunks, consistent with findings in northern California where serovar Pomona predominated in these mesocarnivores (Straub & Foley, 2020). Conversely, squirrels showed minimal reactivity to serovar Pomona and typically had low maximum titers to serovar Icterohaemorrhagiae, which was not highly reactive in other species.

Though the prevalence of infection was lowest in Los Angeles opossums and fox squirrels, interesting patterns emerged in the results of these species. Only one opossum tested had an active infection, and their seroprevalence was the lowest of all species in our study (Table 3.2). This is consistent with smaller *Leptospira* surveys of opossums in California which have found negative (Krueger et al., 2016) or low prevalence results (Straub & Foley, 2020). Some opossums fail to mount strong antibody responses to some *Leptospira* (Reilly, 1970), and the only opossum found to be shedding by Straub & Foley (2020) was seronegative. However, the one infected opossum detected in this study exhibited a high (1:12800) antibody response to serovar Pomona, indicating that strong serologic responses are possible in this species. Prior findings of seronegative shedders may represent later phases of infection, after antibody responses decline, or could represent a subset of individuals who do not mount antibody responses. Fox squirrels exhibited the highest seroprevalence of all our core hosts, but no active infections were detected in this species (Table 3.2). This could be due to a shorter duration of

shedding, longer duration of titer decay (and hence seropositivity), or potentially an alternate route of transmission (e.g., sexual) and associated tissue distribution which could explain the lack of detection in the urinary tract. *Leptospira spp.* have been isolated from fox squirrels before in cases where antibody titers were low, though high titers in this species are possible. *Leptospira interrogans* serovar Grippotyphosa was isolated from the kidney of a fox squirrel that had a low serum antibody titer (1:50) against its own isolate (Diesch et al., 1967). A more recent study of Colorado fox squirrels found evidence of *L. interrogans* infections with associated renal disease, and antibody titers against serovars Grippotyphosa, Hardjo, Icterohaemorrhagiae and Pomona ranging from 1:100 to 1:102,400 depending on the serovar (Dirsmith et al., 2013). These prior studies in fox squirrels show that *Leptospira* has been isolated from kidneys and that Icterohaemorrhagiae titers can be as high as 1:102,400, but active infections in this species may have low titers or be seronegative using conventional MAT.

The prevalence of infection that we detected in urban mammals of the greater Los Angeles was lower than reported for the same species in northern California (Straub & Foley, 2020). This may reflect true regional differences, though comparisons across studies should be considered carefully, as differences between laboratory assay protocols can impact PCR results. For example, Straub et al. (2020) used a Ct cycle threshold of 45 to define positivity in their PCR assay, which would have higher sensitivity than the cycle threshold of 37 used in this study. However, we expect quantitative differences to be relatively minor, and qualitative comparisons should still be valid across studies. Regional differences in leptospirosis incidence could also be partially attributable to environmental differences between northern and southern California. Landscape and environmental factors are known to impact *Leptospira*, as wet environmental

conditions facilitate the bacteria's survival and transmission. Higher rainfall is associated with higher *Leptospira* incidence in domestic dogs (Ward, 2002), including the wetter region of northern California (Adin & Cowgill, 2000). Though Straub and Foley (2020) did not detect a seasonal pattern in wildlife data, neither this nor previous studies compared extensive results from both regions, and a broader survey across the state may be more likely to detect environmental patterns in the future.

Across all regions and species, the majority of seropositive animals were reactive to multiple serovars (65%; n=120/184), which aligns with typical patterns of cross-reactivity in MAT data, though infections with multiple serovars could also be possible in some cases. Distinguishing these outcomes and making sense of serological patterns and associated PCR data is a field of ongoing research in *Leptospira* ecology. Serovar reactivity patterns in several species in our dataset echo patterns observed in the terrestrial mammals on the Channel Islands. Namely, titers against serovars Pomona and Autumnalis were found to be positively correlated in coyotes, mirroring the association commonly observed in island foxes (Mummah et al., In Preparation). These patterns suggest that the same (or closely related) strains of *L. interrogans* serovar Pomona may be circulating in coastal mesocarnivores and island foxes, but this is not definitive evidence since the same study showed clearly that MAT reactivity profiles can vary by host species (Mummah et al., In Preparation).

Genetic analyses of *Leptospira* isolates from different host species are needed to establish strong evidence regarding the relatedness of strains, and hence to assess whether a given strain is circulating among multiple host species. Though it is not yet possible to genetically identify

Leptospira strains from PCR data, future work looking at conserved gene regions may provide finer scale identifications. The strongest confirmation would require *Leptospira* isolates from mainland terrestrial species, which would enable whole genome sequencing for comparison with sequences from other wildlife species along the California coast (and ideally, from other terrestrial isolates in the western United States). Such isolates, obtained through targeted trapping efforts or in the course of species removals for animal control, would yield key insights into multi-host dynamics of serovar Pomona including possible transmission links between terrestrial, marine and island ecosystems (Borremans et al., 2019). Efforts to obtain isolates would be best directed at raccoons, striped skunks and coyotes in the greater Los Angeles region, or other species identified as reactive to serovar Pomona in coastal northern California (Straub & Foley, 2020).

In 2021, the Los Angeles area experienced an outbreak of leptospirosis in domestic dogs, caused by *L. interrogans* serovar Canicola (LA County Department of Public Health (LADPH), 2022). No wildlife species in our study had predominant maximum titers to serovar Canicola, and few individuals showed any reactivity at all against this serovar. This evidence supports the Los Angeles County Department of Public Health's conclusion that this outbreak did not originate from local wildlife, highlighting the importance of longitudinal wildlife surveillance in determining the source (and ruling out possible sources) of outbreaks caused by multi-host pathogens. However, since these surveillance efforts were finite and only five species were sampled extensively, there could still be unobserved host species contributing to *Leptospira* persistence and transmission in the greater Los Angeles region. Identifying signatures of these cryptic contributors, or 'epidemiological dark matter', remains an ongoing frontier in disease

ecology, and emphasizes the need for ongoing research to understand multi-host pathogen dynamics (Buhnerkempe et al., 2015).

The Los Angeles region provides a unique opportunity to investigate this multi-host zoonotic pathogen, and possibly the co-circulation of multiple infecting serovars, in a complicated urban landscape. In an effort to better understand the animal and human health risks arising from this pathogen, we investigated potential correlates of exposure (location, age, sex, land class and season). Positive test results were detected across the region in no clear spatial pattern with regard to geographic space (Figure 3.2) or land use types (Figure S 3.1). We were unable to identify any significant patterns among the set of landscape, seasonal, and demographic covariates considered here. *Leptospira* appears to be widespread across this landscape with no clear risk factors or geographic hot-spots, and likely presents a potential risk to both animal and human health throughout the region.

Though we were unable to identify predictors of exposure in the region, we proceeded to investigate the association between antibody titers and PCR status, as critical knowledge gaps remain regarding the relationship between serology and pathogen shedding in wildlife species (Pedersen et al., 2018). Motivated by Chapter 2 and other work indicating that antibody titers are predictive of infection status (PCR) in California sea lions (Prager et al., 2020), we performed a regression of paired antibody-PCR results in raccoons, which revealed a positive correlation in this species as well (Figure 3.4). Despite the potential value of tools to predict individual shedding status, investigations into predictors of shedding status in wildlife are not common and often focus on population level host-pathogen dynamics. Efforts to better understand individual-

level infection and shedding status represent a current frontier in disease ecology. This relationship could be explored further in raccoons and other species, and if confirmed, could be a useful screening tool for wildlife in situations where PCR results are not available or feasible to obtain.

Results from non-target species and regions support the fact that *Leptospira* is widespread and common in multiple wildlife species in California (Cirone et al., 1978; Gulland et al., 1996; Mummah, 2021; Pedersen et al., 2018; Straub & Foley, 2020). We found evidence of exposure in all non-target species except gray foxes (Table S 3.1), and confirmed that additional species, including wild pigs and desert cottontails, serve as carriers of this pathogen in California (Table S 3.2). Animals were often seroreactive to multiple serovars (Table S 3.2). Our serologic findings identified wild pigs as an additional species with a profile of MAT reactivity (high titers to serovars Pomona, Bratislava, Autumnalis and Djasiman) similar to that found in terrestrial mammals on the Channel Islands. This makes wild pigs a species of interest, and potential sampling target, when considering possible mainland terrestrial reservoirs within the broader multi-host serovar Pomona system.

While this study contributes a wealth of knowledge regarding the status of *Leptospira* in southern California wildlife, it wasn't without limitations. We were dependent on salvaged and incidentally collected samples for pathogen testing, primarily from opportunistic sampling by collaborating agencies. This study design led to limited sample sizes, with a degree of spatial clustering around collaborator facilities (particularly for raccoons), which could have reduced our ability to detect spatial patterns. Samples may also have been biased towards more

developed areas, as higher numbers of roads could have increased the numbers of traffic related deaths, and higher human density could also increase the likelihood of a sick or dead animal being reported. Additionally, the composite land use categories used here may have masked finer scale spatial associations, and our use of circular home range buffers could overlook important behavior patterns in habitat use. Further investigations into individual land use categories or other metrics of urbanization (e.g., population density) could reveal additional spatial relationships undetected in these analyses.

This study provides the first in-depth look at *Leptospira* ecology in terrestrial wildlife across the greater Los Angeles area. Expanded knowledge of this pathogen in southern California, including comparisons of prevalence levels and serological patterns across host species, provide insights into multi-host pathogen dynamics and the potential for cross-species transmission. Evidence of *Leptospira* circulation in Los Angeles wildlife has been lacking, contributing to the perception that it does not pose a major risk in the area. Our study found evidence consistent with circulation of at least two serovars of *Leptospira* among our core five species in the region. Primary serovars differ between mesocarnivore species and squirrels, and high levels of exposure and wide geographic distribution indicate that this pathogen is ubiquitous across the region. While active infection rates appear lower than in northern California, they are still substantial enough to warrant concern and encourage domestic dog owners to vaccinate against this disease. An expanded understanding of *Leptospira* ecology in California wildlife is critical to the management of this widely circulating pathogen, highlighting the need for systematic, broad-scale research efforts that continue to monitor this pathogen in wildlife, domestic pets and humans in the greater Los Angeles region.

3.6 SUPPLEMENT

Table S 3.1: *Leptospira* test results from non-target species and regions.

	Black Bear (N=2)	Boar (N=40)	Bobcat (N=11)	Coyote (N=20)	Desert Cottontail (N=38)	Fox Squirrel (N=1)	Gray Fox (N=9)	Ground Squirrel (N=38)	Northern Raccoon (N=40)	Red Fox (N=5)	Striped Skunk (N=5)	Virginia Opossum (N=32)	Total (N=241)
Antibodies (MAT)													
Positive	2 (100%)	11 (27.5%)	5 (45.5%)	7 (35%)	3 (7.9%)	1 (100%)	0 (0%)	2 (5.3%)	24 (60%)	2 (40%)	2 (40%)	2 (6.3%)	61 (25.3%)
Negative	0 (0%)	29 (72.5%)	6 (54.5%)	12 (60%)	27 (71.1%)	0 (0%)	4 (44.4%)	2 (5.3%)	16 (40%)	3 (60%)	3 (60%)	30 (93.8%)	132 (54.8%)
Not Tested	0 (0%)	0 (0%)	0 (0%)	1 (5%)	8 (21.1%)	0 (0%)	5 (55.6%)	34 (89.5%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	48 (19.9%)
Infection (PCR)													
Negative	0 (0%)	0 (0%)	1 (9.1%)	2 (10%)	27 (71.1%)	0 (0%)	6 (66.7%)	36 (94.7%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	72 (29.9%)
Not Tested	2 (100%)	40 (100%)	10 (90.9%)	18 (90%)	11 (28.9%)	1 (100%)	3 (33.3%)	2 (5.3%)	40 (100%)	5 (100%)	5 (100%)	32 (100%)	169 (70.1%)
Sex													
Female	1 (50%)	22 (55%)	1 (9.1%)	5 (25%)	16 (42.1%)	0 (0%)	1 (11.1%)	13 (34.2%)	1 (2.5%)	2 (40%)	0 (0%)	5 (15.6%)	67 (27.8%)
Male	1 (50%)	18 (45%)	2 (18.2%)	8 (40%)	10 (26.3%)	1 (100%)	5 (55.6%)	23 (60.5%)	2 (5%)	2 (40%)	3 (60%)	3 (9.4%)	78 (32.4%)
Unknown	0 (0%)	0 (0%)	8 (72.7%)	7 (35%)	12 (31.6%)	0 (0%)	3 (33.3%)	2 (5.3%)	37 (92.5%)	1 (20%)	2 (40%)	24 (75%)	96 (39.8%)
Age Class													
Adult	2 (100%)	28 (70%)	7 (63.6%)	14 (70%)	36 (94.7%)	1 (100%)	6 (66.7%)	25 (65.8%)	38 (95%)	4 (80%)	5 (100%)	31 (96.9%)	197 (81.7%)
Juvenile	0 (0%)	12 (30%)	2 (18.2%)	4 (20%)	0 (0%)	0 (0%)	2 (22.2%)	9 (23.7%)	2 (5%)	1 (20%)	0 (0%)	0 (0%)	32 (13.3%)
Unknown	0 (0%)	0 (0%)	2 (18.2%)	2 (10%)	2 (5.3%)	0 (0%)	1 (11.1%)	4 (10.5%)	0 (0%)	0 (0%)	0 (0%)	1 (3.1%)	12 (5%)
County													
Ventura	2 (100%)	0 (0%)	3 (27.3%)	0 (0%)	14 (36.8%)	0 (0%)	1 (11.1%)	1 (2.6%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	21 (8.7%)
San Luis Obispo	0 (0%)	40 (100%)	0 (0%)	9 (45%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (2.5%)	4 (80%)	3 (60%)	4 (12.5%)	61 (25.3%)
Los Angeles	0 (0%)	0 (0%)	2 (18.2%)	0 (0%)	24 (63.2%)	0 (0%)	5 (55.6%)	36 (94.7%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	67 (27.8%)
Monterey	0 (0%)	0 (0%)	5 (45.5%)	7 (35%)	0 (0%)	1 (100%)	2 (22.2%)	1 (2.6%)	39 (97.5%)	1 (20%)	2 (40%)	28 (87.5%)	86 (35.7%)
Santa Barbara	0 (0%)	0 (0%)	1 (9.1%)	4 (20%)	0 (0%)	0 (0%)	1 (11.1%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	6 (2.5%)
Season													
Dry	2 (100%)	8 (20%)	3 (27.3%)	8 (40%)	16 (42.1%)	0 (0%)	3 (33.3%)	28 (73.7%)	13 (32.5%)	2 (40%)	0 (0%)	6 (18.8%)	89 (36.9%)
Wet	0 (0%)	26 (65%)	8 (72.7%)	11 (55%)	22 (57.9%)	1 (100%)	6 (66.7%)	10 (26.3%)	27 (67.5%)	3 (60%)	5 (100%)	26 (81.3%)	145 (60.2%)
Unknown	0 (0%)	6 (15%)	0 (0%)	1 (5%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	7 (2.9%)

Table S 3.2: *Leptospira* serovars and antibody titer levels for all California wildlife samples.

Common Name	Scientific Name	n	Serovar	Titer 100	200	400	800	1600	3200	>=6400	% Positive (95% CI)
Boar	*Sus scrofa*	40	Autumnalis	3	1	1					12.5 (4.2-26.8)
		40	Bratislava	3	4	1					20 (9.1-35.6)
		40	Canicola	1							2.5 (0.1-13.2)
		40	Cynopteri	1							2.5 (0.1-13.2)
		40	Djasiman	2		2					10 (2.8-23.7)
		40	Georgia		1						2.5 (0.1-13.2)
		40	Pomona	2		1					7.5 (1.6-20.4)
		40	Pyrogenes		1						2.5 (0.1-13.2)
		40	Tarassovi	1							2.5 (0.1-13.2)
Bobcat	*Lynx rufus*	7	Autumnalis						1		14.29 (0.4-57.9)
		7	Bataviae	1							14.29 (0.4-57.9)
		11	Bratislava	2	1		1				36.36 (10.9-69.2)
		11	Canicola					1			9.09 (0.2-41.3)
		7	Cynopteri	1							14.29 (0.4-57.9)
		7	Djasiman						1		14.29 (0.4-57.9)
		11	Grippotyphosa					1			9.09 (0.2-41.3)
		11	Icterohaemorrhagiae		1				1		18.18 (2.3-51.8)
		7	Mankarso					1			14.29 (0.4-57.9)
		11	Pomona		2				1		27.27 (6-61)
		7	Pyrogenes					1			14.29 (0.4-57.9)
		7	Wolffi	1							14.29 (0.4-57.9)
Coyote	*Canis latrans*	42	Australis		1						2.38 (0.1-12.6)
		43	Autumnalis	3	4	2	1		2		27.91 (15.3-43.7)
		71	Bratislava	5	3	1	1			2	16.9 (9-27.7)
		71	Canicola		1						1.41 (0-7.6)
		43	Cynopteri	1							2.33 (0.1-12.3)
		43	Djasiman	2	3	1	2		2		23.26 (11.8-38.6)
		71	Grippotyphosa	2	2	1					7.04 (2.3-15.7)
		38	Hardjo	1	1						5.26 (0.6-17.7)
		71	Icterohaemorrhagiae	4	3	1	1			1	14.08 (7-24.4)
		43	Mankarso	1	3	1	1				13.95 (5.3-27.9)
		71	Pomona	5	1	3	1	2		2	19.72 (11.2-30.9)
		42	Tarassovix	1							2.38 (0.1-12.6)
Desert Cottontail	*Sylvilagus audubonii*	13	Georgia				1				7.69 (0.2-36)
		16	Icterohaemorrhagiae	1							6.25 (0.2-30.2)
		15	Pomona		1						6.67 (0.2-31.9)
Red Fox	*Vulpes vulpes*	4	Autumnalis			1					25 (0.6-80.6)
		4	Djasiman		1						25 (0.6-80.6)
		5	Icterohaemorrhagiae	1							20 (0.5-71.6)
		4	Mankarso	1							25 (0.6-80.6)
		5	Pomona		1						20 (0.5-71.6)
		4	Pyrogenes	1							25 (0.6-80.6)
Northern Raccoon	*Procyon lotor*	10	Autumnalis	2		1					30 (6.7-65.2)
		127	Bratislava	11	3	9	4	3	3	2	27.56 (20-36.2)
		127	Canicola	2	4	2	2	1			8.66 (4.4-15)
		10	Celledoni	1							10 (0.3-44.5)
		9	Cynopteri	1							11.11 (0.3-48.2)
		10	Djasiman	1	1	1					30 (6.7-65.2)
		127	Grippotyphosa	1	5	4	1				8.66 (4.4-15)
		121	Hardjo	11	5	2	4	3			20.66 (13.8-29)
		127	Icterohaemorrhagiae	13	9	5	5	1	2	1	28.35 (20.7-37)
		9	Mankarso			1					11.11 (0.3-48.2)
		127	Pomona	6	10	8	9	7	3	11	42.52 (33.8-51.6)
		9	Pyrogenes	1							11.11 (0.3-48.2)
		10	Tarassovi			1					10 (0.3-44.5)
Fox Squirrel	*Sciurus niger*	39	Australis	2							5.13 (0.6-17.3)
		40	Autumnalis	2	2						10 (2.8-23.7)
		78	Bratislava	9	15	9	4	2			50 (38.5-61.5)
		37	Celledoni	1							2.7 (0.1-14.2)
		51	Hardjo	8	16	7	2	1			66.67 (52.1-79.2)
		93	Icterohaemorrhagiae	29	15	9	5	1	1		64.52 (53.9-74.2)
		36	Javanica	1							2.78 (0.1-14.5)
		36	Mankarso	1	1	1					8.33 (1.8-22.5)
		78	Pomona	1		1					2.56 (0.3-9)
		37	Pyrogenes	2							5.41 (0.7-18.2)
Ground Squirrel	*Otospermophilus beecheyi*	4	Bratislava	2							50 (6.8-93.2)
		4	Hardjo		1						25 (0.6-80.6)
Striped Skunk	*Mephitis mephitis*	7	Autumnalis	2			1				42.86 (9.9-81.6)
		24	Bratislava		1						4.17 (0.1-21.1)
		6	Djasiman	1		1					33.33 (4.3-77.7)
		24	Icterohaemorrhagiae	2		1					12.5 (2.7-32.4)
		6	Mankarso	1							16.67 (0.4-64.1)
		24	Pomona	2		2		1		2	29.17 (12.6-51.1)
Virginia Opossum	*Didelphis virginiana*	110	Bratislava	1							0.91 (0-5)
		75	Hardjo	7							9.33 (3.8-18.3)
		111	Icterohaemorrhagiae	2							1.8 (0.2-6.4)
		110	Pomona							1	0.91 (0-5)

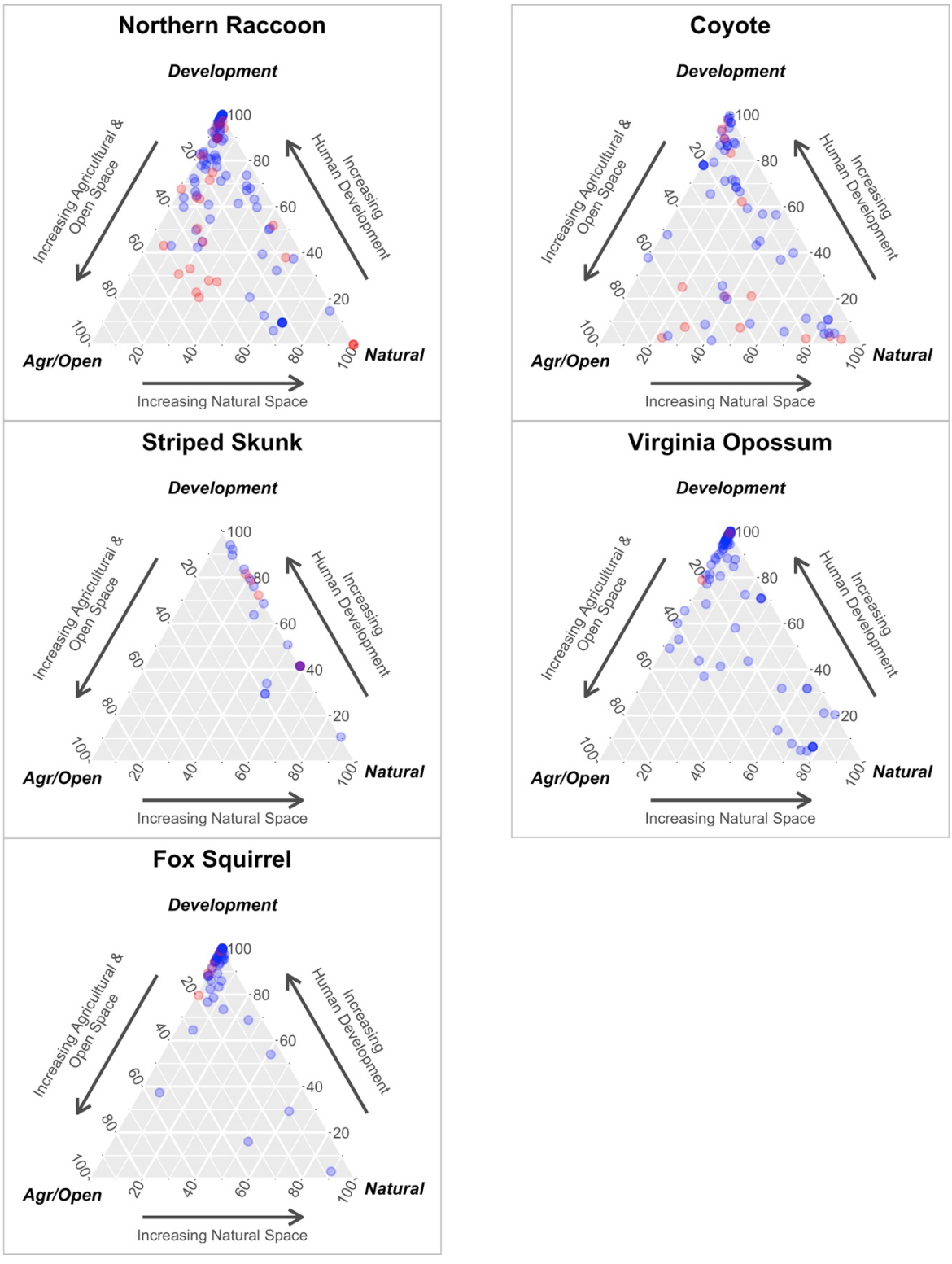


Figure S 3.1: *Leptospira* antibody results relative to land classification. Positive (red) and negative (blue) results are shown for each individual species, relative to the proportion of natural, agricultural/open and developed land calculated within the home range of each species.

3.7 REFERENCES

Adducci II, A., Jasperse, J., Riley, S., Brown, J., Honeycutt, R., & Monzón, J. (2020). Urban coyotes are genetically distinct from coyotes in natural habitats. *Journal of Urban Ecology*, *6*(1), juaa010. https://doi.org/10.1093/jue/juaa010

Adin, C. A., & Cowgill, L. D. (2000). Treatment and outcome of dogs with leptospirosis: 36 cases (1990-1998). *Journal of the American Veterinary Medical Association*, *216*(3), 371–375. https://doi.org/10.2460/javma.2000.216.371

Adler, B., & de la Peña Moctezuma, A. (2010). *Leptospira* and leptospirosis. *Veterinary Microbiology*, *140*(3–4), 287–296. https://doi.org/10.1016/j.vetmic.2009.03.012

André-Fontaine, G., & Triger, L. (2018). MAT cross-reactions or vaccine cross-protection: Retrospective study of 863 leptospirosis canine cases. *Heliyon*, *4*(11), e00869. https://doi.org/10.1016/j.heliyon.2018.e00869

Blanco, R. M., dos Santos, L. F., Galloway, R. L., & Romero, E. C. (2016). Is the microagglutination test (MAT) good for predicting the infecting serogroup for leptospirosis in Brazil? *Comparative Immunology, Microbiology and Infectious Diseases*, *44*, 34–36. https://doi.org/10.1016/j.cimid.2015.12.003

Blessington, T., Schenck, A. P., & Levine, J. F. (2020). Frequency of Animal Leptospirosis in the Southern United States and the Implications for Human Health. *Southern Medical Journal*, *113*(5), 240–249. https://doi.org/10.14423/SMJ.0000000000001093

Boey, K., Shiokawa, K., & Rajeev, S. (2019). *Leptospira* infection in rats: A literature review of global prevalence and distribution. *PLOS Neglected Tropical Diseases*, *13*(8), e0007499. https://doi.org/10.1371/journal.pntd.0007499

Borremans, B., Faust, C., Manlove, K. R., Sokolow, S. H., & Lloyd-Smith, J. O. (2019). Cross-species pathogen spillover across ecosystem boundaries: Mechanisms and theory. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1782), 20180344. https://doi.org/10.1098/rstb.2018.0344

Brown, K., & Prescott, J. (2008). Leptospirosis in the family dog: A public health perspective. *Canadian Medical Association Journal*, *178*(4), 399–401. https://doi.org/10.1503/cmaj.071097

Buhnerkempe, M. G., Prager, K. C., Strelioff, C. C., Greig, D. J., Laake, J. L., Melin, S. R., DeLong, R. L., Gulland, F. M., & Lloyd-Smith, J. O. (2017). Detecting signals of chronic shedding to explain pathogen persistence: *Leptospira interrogans* in California sea lions. *Journal of Animal Ecology*, *86*(3), 460–472. https://doi.org/10.1111/1365-2656.12656

Buhnerkempe, M. G., Roberts, M. G., Dobson, A. P., Heesterbeek, H., Hudson, P. J., & Lloyd-Smith, J. O. (2015). Eight challenges in modelling disease ecology in multi-host, multi-agent systems. *Epidemics*, *10*, 26–30. https://doi.org/10.1016/j.epidem.2014.10.001

California Department of Public Health. (2020). *CDPH INFECTIOUS DISEASES BRANCH (IDB) YEARLY SUMMARIES OF SELECTED COMMUNICABLE DISEASES IN CALIFORNIA, 2011–2019*. https://www.cdph.ca.gov/Programs/CID/DCDC/CDPH%20Document%20Library/YearlySummariesofSelectedCommDiseasesinCA.pdf

Centers for Disease Control and Prevention. (2018). *LEPTOSPIROSIS - Fact Sheet for Clinicians*. https://www.cdc.gov/leptospirosis/pdf/fs-leptospirosis-clinicians-eng-508.pdf

Cirone, S. M., Riemann, H. P., Ruppanner, R., Behymer, D. E., & Franti, C. E. (1978). Evaluation of the hemagglutination test for epidemiologic studies of leptospiral antibodies in wild mammals. *Journal of Wildlife Diseases*, *14*(2), 193–202. https://doi.org/10.7589/0090-3558-14.2.193

Clark, R. K., Jessup, D. A., Hird, D. W., Ruppanner, R., & Meyer, M. E. (1983). Serologic survey of California wild hogs for antibodies against selected zoonotic disease agents. *Journal of the American Veterinary Medical Association*, *183*(11), 1248–1251.

Colagross-Schouten, A. M., Mazet, J. A. K., Gulland, F. M. D., Miller, M. A., & Hietala, S. (2002). Diagnosis and seroprevalence of leptospirosis in California sea lions from coastal California. *Journal of Wildlife Diseases*, *38*(1), 7–17. https://doi.org/10.7589/0090-3558-38.1.7

Costa, F., Hagan, J. E., Calcagno, J., Kane, M., Torgerson, P., Martinez-Silveira, M. S., Stein, C., Abela-Ridder, B., & Ko, A. I. (2015). Global Morbidity and Mortality of Leptospirosis: A Systematic Review. *PLOS Neglected Tropical Diseases*, *9*(9), e0003898. https://doi.org/10.1371/journal.pntd.0003898

Davis, M. A., Evermann, J. F., Petersen, C. R., VanderSchalie, J., Besser, T. E., Huckabee, J., Daniels, J. B., Hancock, D. D., Leslie, M., & Baer, R. (2008). Serological Survey for Antibodies to *Leptospira* in Dogs and Raccoons in Washington State. *Zoonoses and Public Health*, ???-??? https://doi.org/10.1111/j.1863-2378.2008.01137.x

Diesch, S. L., Crawford, R. P., McCulloch, W. F., & Top, F. H. (1967). Human Leptospirosis Acquired from Squirrels. *New England Journal of Medicine*, *276*(15), 838–842. https://doi.org/10.1056/NEJM196704132761504

Dirsmith, K., VanDalen, K., Fry, T., Charles, B., VerCauteren, K., & Duncan, C. (2013). Leptospirosis in Fox Squirrels (Sciurus niger) of Larimer County, Colorado, USA. *Journal of Wildlife Diseases*, *49*(3), 641–645. https://doi.org/10.7589/2012-10-265

Ellis, W. A. (2015). Animal Leptospirosis. In B. Adler (Ed.), *Current Topics in Microbiology and Immunology* (Vol. 387, pp. 99–137). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-45059-8_6

ESRI Inc. (2017). *ArcGIS Desktop* (10.6.0.8321) [Computer Software]. ESRI Inc. https://desktop.arcgis.com/en/

Faine, S., Adler, B., Bolin, C., & Perolat, P. (1999). *Leptospira and leptospirosis* (2nd ed.). MediSci. https://research.monash.edu/en/publications/leptospira-and-leptospirosis

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, *80*(1), 27–38. https://doi.org/10.1093/biomet/80.1.27

Gautam, R., Wu, C.-C., Guptill, L. F., Potter, A., & Moore, G. E. (2010). Detection of antibodies against *Leptospira* serovars via microscopic agglutination tests in dogs in the United States, 2000–2007. *Journal of the American Veterinary Medical Association*, *237*(3), 293–298. https://doi.org/10.2460/javma.237.3.293

Grau, G. A., Sanderson, G. C., & Rogers, J. P. (1970). Age Determination of Raccoons. *The Journal of Wildlife Management*, *34*(2), 364. https://doi.org/10.2307/3799023

Greene, M. R. (1941). A survey for leptospirosis in Southern California. *American Journal of Epidemiology*, *34-SectionB*(2), 87–90. https://doi.org/10.1093/oxfordjournals.aje.a118739

Greig, D. J., Gulland, F. M. D., & Kreuder, C. (2005). A Decade of Live California Sea Lion (*Zalophus californianus*) Strandings Along the Central California Coast: Causes and Trends, 1991-2000. *Aquatic Mammals*, *31*(1), 11–22. https://doi.org/10.1578/am.31.1.2005.11

Gulland, F. M. D., Koski, M., Lowenstine, L. J., Colagross, A., Morgan, L., & Spraker, T. (1996). Leptospirosis in California sea lions (*Zalophus californianus*) stranded along the central California coast, 1981-1994. *Journal of Wildlife Diseases*, *32*(4), 572–580. https://doi.org/10.7589/0090-3558-32.4.572

Harrison, N. A., & Fitzgerald, W. R. (1988). Leptospirosis—Can it be a sexually transmitted disease? *Postgraduate Medical Journal*, *64*(748), 163–164. https://doi.org/10.1136/pgmj.64.748.163

Hartskeerl, R. A., Collares-Pereira, M., & Ellis, W. A. (2011). Emergence, control and re-emerging leptospirosis: Dynamics of infection in the changing world. *Clinical Microbiology and Infection*, *17*(4), 494–501. https://doi.org/10.1111/j.1469-0691.2011.03474.x

Hennebelle, J. H., Sykes, J. E., Carpenter, T. E., & Foley, J. (2013). Spatial and temporal patterns of *Leptospira* infection in dogs from northern California: 67 cases (2001–2010). *Journal of the American Veterinary Medical Association*, *242*(7), 941–947. https://doi.org/10.2460/javma.242.7.941

Hennebelle, J. H., Sykes, J. E., & Foley, J. (2014). Risk Factors Associated with Leptospirosis in Dogs from Northern California: 2001–2010. *Vector-Borne and Zoonotic Diseases*, *14*(10), 733–739. https://doi.org/10.1089/vbz.2014.1624

Ko, A. I., Reis, M. G., Dourado, C. M. R., Johnson, W. D., & Riley, L. W. (1999). Urban epidemic of severe leptospirosis in Brazil. *The Lancet*, *354*(9181), 820–825. https://doi.org/10.1016/S0140-6736(99)80012-9

Krueger, L., Bai, Y., Bennett, S., Fogarty, C., Sun, S., Kosoy, M., Maina, A., Nelson, K., Platzer, E., Osikowicz, L., Richards, A., L., Shariar, F., Trinidad, A., Cummings, A., & Cummings, R. (2016). Identification of Zoonotic and Vector-borne Infectious Agents Associated with Opossums (Didelphis virginiana) in Residential Neighborhoods of Orange County, California. *Proceedings of the Vertebrate Pest Conference*, *27*. https://doi.org/10.5070/V427110386

LA County Department of Public Health (LADPH). (2022, March 11). *Leptospirosis in Dogs in Los Angeles County in 2021*. http://www.publichealth.lacounty.gov/vet/Leptospirosis2021.htm

Lau, C. L., Smythe, L. D., Craig, S. B., & Weinstein, P. (2010). Climate change, flooding, urbanisation and leptospirosis: Fuelling the fire? *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *104*(10), 631–638. https://doi.org/10.1016/j.trstmh.2010.07.002

Leonard, F., Quinn, P., Ellis, W., & O'Farrell, K. (1992). Duration of urinary excretion of leptospires by cattle naturally or experimentally infected with *Leptospira interrogans* serovar hardjo. *Veterinary Record*, *131*(19), 435–439. https://doi.org/10.1136/vr.131.19.435

Levett, P. N. (2001). Leptospirosis. *Clinical Microbiology Reviews*, *14*(2), 296–326. https://doi.org/10.1128/CMR.14.2.296-326.2001

Levett, P. N. (2015). Systematics of Leptospiraceae. In B. Adler (Ed.), *Leptospira and Leptospirosis* (Vol. 387, pp. 11–20). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-45059-8_2

Lloyd-Smith, J. O. (2021). *Leptospirosis in endangered island foxes and California sea lions: Outbreak prediction and prevention in a changing world* (Final RC-2635; Issue RC-2635). University of California, Los Angeles.

Lloyd-Smith, J. O., Greig, D. J., Hietala, S., Ghneim, G. S., Palmer, L., St Leger, J., Grenfell, B. T., & Gulland, F. M. (2007). Cyclical changes in seroprevalence of leptospirosis in California sea lions: Endemic and epidemic disease in one host species? *BMC Infectious Diseases*, *7*, 125–136. https://doi.org/125 10.1186/1471-2334-7-125

Meites, E., Jay, M. T., Deresinski, S., Shieh, W.-J., Zaki, S. R., Tompkins, L., & Smith, D. S. (2004). Reemerging Leptospirosis, California. *Emerging Infectious Diseases*, *10*(3), 406–412. https://doi.org/10.3201/eid1003.030431

Minter, A., Diggle, P. J., Costa, F., Childs, J., Ko, A. I., & Begon, M. (2017). Evidence of multiple intraspecific transmission routes for *Leptospira* acquisition in Norway rats (*Rattus norvegicus*). *Epidemiology and Infection*, *145*(16), 3438–3448. https://doi.org/10.1017/S0950268817002539

Minter, A., Diggle, P. J., Costa, F., Childs, J., Ko, A. I., & Begon, M. (2018). A model for leptospire dynamics and control in the Norway rat (Rattus norvegicus) the reservoir host in urban slum environments. *Epidemics*, *25*, 26–34. https://doi.org/10.1016/j.epidem.2018.05.002

Monahan, A. M., Callanan, J. J., & Nally, J. E. (2008). Proteomic Analysis of *Leptospira interrogans* Shed in Urine of Chronically Infected Hosts. *Infection and Immunity*, *76*(11), 4952–4958. https://doi.org/10.1128/IAI.00511-08

Moore, G. E., Guptill, L. F., Glickman, N. W., Caldanaro, R. J., Aucoin, D., & Glickman, L. T. (2006). Canine Leptospirosis, United States, 2002–2004. *Emerging Infectious Diseases*, *12*(3), 501–503. https://doi.org/10.3201/eid1203.050809

Mummah, R. O. (2021). *Leptospira in the coastal California ecosystem: Challenges and solutions for analyzing complex wildlife disease data* [Dissertation]. University of California Los Angeles.

Mummah, R. O., Gomez, A. C. R., Guglielmino, A., Borremans, B., Galloway, R. L., Prager, K. C., & Lloyd-Smith, J. O. (In Preparation). *MAT is a minefield: A cautionary tale of a single Leptospira serovar in multiple host species*.

Munoz-Zanzi, C., Groene, E., Morawski, B. M., Bonner, K., Costa, F., Bertherat, E., & Schneider, M. C. (2020). A systematic literature review of leptospirosis outbreaks worldwide, 1970–2012. *Revista Panamericana de Salud Pública*, *44*, 1. https://doi.org/10.26633/RPSP.2020.78

Neiderud, C.-J. (2015). How urbanization affects the epidemiology of emerging infectious diseases. *Infection Ecology & Epidemiology*, *5*(1), 27060. https://doi.org/10.3402/iee.v5.27060

Pedersen, K., Anderson, T. D., Maison, R. M., Wiscomb, G. W., Pipas, M. J., Sinnett, D. R., Baroch, J. A., & Gidlewski, T. (2018). *Leptospira* Antibodies Detected in Wildlife in the USA and the US Virgin Islands. *Journal of Wildlife Diseases*, *54*(3), 450–459. https://doi.org/10.7589/2017-10-269

Prager, K. C., Alt, D. P., Buhnerkempe, M. G., Greig, D. J., Galloway, R. L., Wu, Q., Gulland, F. M. D., & Lloyd-Smith, J. O. (2015). Antibiotic Efficacy in Eliminating Leptospiruria in California Sea Lions (*Zalophus californianus*) Stranding with Leptospirosis. *Aquatic Mammals*, *41*(2), 203–212. https://doi.org/doi: 10.1578/AM.41.2.2015.203

Prager, K. C., Buhnerkempe, M. G., Greig, D. J., Orr, A. J., Jensen, E. D., Gomez, F., Galloway, R. L., Wu, Q., Gulland, F. M. D., & Lloyd-Smith, J. O. (2020). Linking longitudinal and cross-sectional biomarker data to understand host-pathogen dynamics: *Leptospira* in

California sea lions (*Zalophus californianus*) as a case study. *PLOS Neglected Tropical Diseases*, *14*(6), e0008407. https://doi.org/10.1371/journal.pntd.0008407

Prager, K. C., Greig, D. J., Alt, D. P., Galloway, R. L., Hornsby, R. L., Palmer, L. J., Soper, J., Wu, Q., Zuerner, R. L., Gulland, F. M. D., & Lloyd-Smith, J. O. (2013). Asymptomatic and chronic carriage of *Leptospira interrogans* serovar Pomona in California sea lions (*Zalophus californianus*). *Veterinary Microbiology*, *164*(1–2), 177–183. https://doi.org/10.1016/j.vetmic.2013.01.032

Prince, A., DePerno, C. S., Gardner, B., & Moorman, C. E. (2014). Survival and Home-Range Size of Southeastern Fox Squirrels in North Carolina. *Southeastern Naturalist*, *13*(3), 456. https://doi.org/10.1656/058.013.0305

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Reilly, J. R. (1970). The susceptibility of five species of wild animals to experimental infection with *Leptospira* grippotyphosa. *Journal of Wildlife Diseases*, *6*(4), 289–294. https://doi.org/10.7589/0090-3558-6.4.289

Riley, S. P. D., Serieys, L. E. K., & Moriarty, J. G. (2014). Infectious Disease and Contaminants in Urban Wildlife: Unseen and Often Overlooked Threats. In R. A. McCleery, C. E. Moorman, & M. N. Peterson (Eds.), *Urban Wildlife* (pp. 175–215). Springer US. https://doi.org/10.1007/978-1-4899-7500-3_10

Roug, A., Swift, P., Torres, S., Jones, K., & Johnson, C. K. (2012). Serosurveillance for Livestock Pathogens in Free-Ranging Mule Deer (Odocoileus hemionus). *PLoS ONE*, *7*(11), e50600. https://doi.org/10.1371/journal.pone.0050600

Ruppanner, R., Jessup, D. A., Ohishi, I., Behymer, D. E., & Franti, C. E. (1982). Serologic survey for certain zoonotic diseases in black bears in California. *Journal of the American Veterinary Medical Association*, *181*(11), 1288–1291.

Šálek, M., Drahníková, L., & Tkadlec, E. (2015). Changes in home range sizes and population densities of carnivore species along the natural to urban habitat gradient. *Mammal Review*, *45*(1), 1–14. https://doi.org/10.1111/mam.12027

Schreiber, P., Martin, V., Najbar, W., Sanquer, A., Gueguen, S., & Lebreux, B. (2005). Prevention of renal infection and urinary shedding in dogs by a *Leptospira* vaccination. *Veterinary Microbiology*, *108*(1–2), 113–118. https://doi.org/10.1016/j.vetmic.2005.03.007

Smith, A. M., Stull, J. W., Evason, M. D., Weese, J. S., Wittum, T. E., Szlosek, D., & Arruda, A. G. (2021). Investigation of spatio-temporal clusters of positive leptospirosis polymerase chain reaction test results in dogs in the United States, 2009 to 2016. *Journal of Veterinary Internal Medicine*, *35*(3), 1355–1360. https://doi.org/10.1111/jvim.16060

Straub, M. H., Church, M., Glueckert, E., & Foley, J. E. (2020). Raccoons (*Procyon lotor*) and Striped Skunks (*Mephitis mephitis*) as Potential Reservoirs of *Leptospira* spp. In California. *Vector-Borne and Zoonotic Diseases*, *20*(6), 418–426. https://doi.org/10.1089/vbz.2019.2528

Straub, M. H., & Foley, J. E. (2020). Cross-sectional evaluation of multiple epidemiological cycles of *Leptospira* species in peri-urban wildlife in California. *Journal of the American Veterinary Medical Association*, *257*(8), 840–848. https://doi.org/10.2460/javma.257.8.840

Straub, M. H., Rudd, J. L., Woods, L. W., Clifford, D. L., & Foley, J. E. (2021). *Leptospira* prevalence and its association with renal pathology in mountain lions (*Puma concolor*) and bobcats (*Lynx rufus*) in California, USA. *Journal of Wildlife Diseases*, *57*(1). https://doi.org/10.7589/JWD-D-20-00070

United States Census Bureau. (2020). *Los Angeles County, California Profile*. https://www.census.gov/search-results.html?q=quick+facts+los+angeles&page=1&stateGeo=none&searchtype=web&csp=SERP&_charset_=UTF-8

US. Geological Survey (USGS). (2019). *NLCD 2019 Land Cover Conterminous United States*. U.S. Geological Survey. https://doi.org/10.5066/P937PN4Z

Viana, M., Mancy, R., Biek, R., Cleaveland, S., Cross, P. C., Lloyd-Smith, J. O., & Haydon, D. T. (2014). Assembling evidence for identifying reservoirs of infection. *Trends in Ecology & Evolution*, *29*(5), 270–279. https://doi.org/10.1016/j.tree.2014.03.002

Ward, M. P. (2002). Seasonality of canine leptospirosis in the United States and Canada and its association with rainfall. *Preventive Veterinary Medicine*, *56*(3), 203–213. https://doi.org/10.1016/S0167-5877(02)00183-6

Ward, M. P., Guptill, L. F., & Wu, C. C. (2004). Evaluation of environmental risk factors for leptospirosis in dogs: 36 cases (1997–2002). *Journal of the American Veterinary Medical Association*, *225*(1), 72–77. https://doi.org/10.2460/javma.2004.225.72

White, A. M., Zambrana-Torrelio, C., Allen, T., Rostal, M. K., Wright, A. K., Ball, E. C., Daszak, P., & Karesh, W. B. (2017). Hotspots of canine leptospirosis in the United States of America. *The Veterinary Journal*, *222*, 29–35. https://doi.org/10.1016/j.tvjl.2017.02.009

Wright, J. D., Burt, M. S., & Jackson, V. L. (2012). Influences of an Urban Environment on Home Range and Body Mass of Virginia Opossums (*Didelphis virginiana*). *Northeastern Naturalist*, *19*(1), 77–86. https://doi.org/10.1656/045.019.0106

Wu, Q., Prager, K., Goldstein, T., D.P., A., Galloway, R. L., Zuerner, R. L., Lloyd-Smith, J., & Schwacke, L. (2014). Development of a real-time PCR for the detection of pathogenic *Leptospira* spp. In California sea lions. *Diseases of Aquatic Organisms*, *110*(3), 165–172. https://doi.org/10.3354/dao02752

Zuerner, R. L., Cameron, C. E., Raverty, S., Robinson, J., Colegrove, K. M., Norman, S. A., Lambourn, D., Jeffries, S., Alt, D. P., & Gulland, F. (2009). Geographical dissemination of *Leptospira interrogans* serovar Pomona during seasonal migration of California sea lions. *Veterinary Microbiology*, *137*(1–2), 105–110. https://doi.org/10.1016/j.vetmic.2008.12.017