Preprints.org

Concept Paper

How Should We Account for Euthanasia in Veterinary Research? A Proposal to Use Counterfactual Outcome Adjudication

Charles Cummings *

Concept Paper

How Should We Account for Euthanasia in Veterinary Research? A Proposal to Use Counterfactual Outcome Adjudication

Charles Cummings

Graduate School of Biomedical Sciences, Tufts University, Boston, MA 02111, and Animal Care Department, Lincoln Park Zoo, Chicago, IL, 60614, USA; cummings3793@gmail.com

Abstract

While essential for the ethical practice of veterinary medicine, euthanasia profoundly complicates research with a survival outcome. In particular, euthanasia can make it difficult to determine the extent to which a certain clinical sign, lab, or imaging finding is associated with poor prognosis since animals that die while receiving veterinary care are often euthanized rather than dying naturally. The reasons for euthanasia, however, may be dramatically different between patients. Some are euthanized because prognosis is considered extremely poor even with extensive treatment whereas others may be euthanized due to client financial limitations despite a reasonably good prognosis. In addition, when a clinician-scientist veterinarian believes a clinical finding is associated with poor survival, they may consciously or unconsciously influence clients to euthanize their animals. In effect, this could create – or artificially inflate the strength of – an association between that finding and patient survival. In this viewpoint, I will discuss the use of causal inference tools like directed acyclic graphs (DAGs) to identify the treating veterinarian's belief about prognosis as a variable that mediates the effect of clinical findings on the probability of survival. Furthermore, use of counterfactual outcome adjudication committees in observational veterinary research with a survival outcome is proposed to estimate the probability that euthanized patients would have survived had they received additional treatment. By using both DAGs and these estimates of counterfactual survival probability to simulate outcomes for euthanized patients, investigators can estimate the causal effects of different clinical findings on probability of patient survival.

Keywords: causal inference; counterfactual; DAG; prognosis; euthanasia

Research in veterinary medicine is profoundly complicated by euthanasia. Too often, veterinarians in lectures or in print report that some finding – from physical exam, clinical pathology, or imaging- is associated with an increased risk of death. Too little do we question such statements and what data exist to support them.

If as a clinician, I see a pet presenting to the ER with hypoglycemia, a sign I believe might be associated with poor survival, does that influence how I speak with that animal's owner? Does my belief that this animal possesses a poor prognostic indicator (along with my assessments of other findings and the owner's willingness to pursue treatment) either consciously or unconsciously influence an owner into deciding to euthanize that pet? When I go back and evaluate the data on my cases should I be surprised when my beliefs are borne out to be true? The answers are: yes, yes, and no.

Of course a veterinarian's assessment of a clinical scenario as being suboptimal influences their communication with owners, and often this may lead to owners choosing to euthanize their pet. When this occurs in a research setting, however, the story becomes considerably more complicated. This is because studies often aim to determine if certain clinical findings are associated with death and generalize those findings to new populations outside of the study. In human medicine, this tends to be more straightforward than in veterinary medicine as humans typically cannot be euthanized,

whereas animals are often euthanized rather than dying naturally. Owners may have opted to euthanize because of veterinarian-perceived poor prognosis, financial limitations, excessive physical or emotional burden of continued treatment on caregivers, inability to maintain sufficient welfare for the individual, or any combination of these factors. Because of this, it is difficult to know how generalizable the findings are from veterinary research with survival as a principal study outcome.

In this article, I offer an introduction to causal inference methods, such as directed acyclic graphs (DAGs), and their application to veterinary medicine. I conclude this viewpoint by describing the potential for using counterfactual outcome adjudication to make the results of veterinary research with survival outcomes more robust and generalizable.

Introduction to Causal Inference

Anyone with even a passing familiarity with research has heard that 'correlation does not equal causation,' and every year thousands of observational human and veterinary medical studies are published that detail risk factors for a variety of outcomes like cardiovascular disease, cancer, or death. Unfortunately, there is often little to no consideration of whether those risk factors truly have any *causal* relationship to the outcome under study and thus no answers to whether modifying those risk factors would result in any improvement in outcomes. It does not have to be this way.

Causal inference is a discipline that aims to understand the underlying causes of various outcomes by combining data with causal models. Causal models are merely stated assumptions that any researcher or clinician may hold about the potential causal relationships between two or more variables being investigated. These assumptions are based in their areas of expertise and general knowledge within their field of study. These models are often diagrammed in the form of directed acyclic graphs (DAGs) consisting of nodes representing different variables and arrows from one node to another representing how one variable is thought to possibly exert an effect on the other. DAGs describe the variables which can potentially affect or be affected by the primary exposure and outcome variables of interest. In their simplest form, for example, representing an experiment under controlled conditions – a DAG can consist of two nodes, X and Y, representing an exposure and outcome with an arrow pointing from X to Y (Figure 1). This DAG would represent an assumption – based on prior knowledge – that a change in X, such as the vertical force exerted on a ball tossed straight up into the air, causes a change in Y, namely, the peak height of the ball. DAGs describing more complicated real world phenomena, such as the pathophysiology of different diseases, can have many simple and complicated variables in their construction.

Any set of three variables in a DAG – exposure, outcome, and a third variable – can be connected in three ways (Table 1). One connection type is called a chain (Table 1A). In a chain, an exposure exerts an effect on a third variable, and the third variable exerts an effect on the outcome (in such cases, the exposure may also exert an independent direct effect on the outcome under study). Thus, the third variable acts to *mediate* the effect of the exposure on the outcome. The second connection type is called a fork (Table 1B). In a fork, one variable exerts an independent effect on two other variables. When a third variable affects both the exposure of interest and the outcome of interest, it is called a confounder because it makes it unclear whether a causal pathway actually exists between the exposure of interest and the outcome of interest.

As an example, consider we are trying to determine the effect of receiving a cardiac medication on the likelihood of experiencing a myocardial infarction (MI), or heart attack, in the general population of people. In this case, age may confound the relationship because it both affects the likelihood of receiving cardiac medication and the likelihood of experiencing a MI. If we do not account for age in some way, either by stratifying the data by age or controlling for age in a regression analysis, we may erroneously conclude that not receiving the medication is associated with a reduced likelihood of a MI; this would be because the youngest patients, who typically would not be receiving the medicine, had the lowest baseline risk of MI. In middle-aged to older people, however, receiving the cardiac medication may be associated with reduced likelihood of MI. By adjusting for the age

confounder, we could then determine that, for any given age, the risk of MI is lower in the medically treated group.

Lastly, the third connection type is called a collider (Table 1C). In a collider, two variables independently affect a third variable; often, this is the exposure and outcome variables affecting a third variable. When this collider is not recognized or appreciated, it may cause an erroneous understanding of the causal relationships at play. Consider the same example as above, a study of the relationship between receiving a cardiac medication and likelihood of experiencing a heart attack in the general population. Let's assume, out of convenience, we have opted to use hospitalized patients – both receiving and not receiving cardiac medication – for our study. This is potentially problematic, however, because both being on heart medications and having a MI independently affect how likely a person is to be hospitalized. Someone receiving cardiac medications (and possibly other medications) is more likely to be monitored by a clinician and to have comorbidities requiring hospitalization than the average person not receiving cardiac medications. Similarly, a person who has just experienced a MI is more likely to be hospitalized than someone who hasn't. Thus, hospitalization is a collider. When we adjust, or condition, on hospitalization by only including hospitalized patients in our study, the causal relationship between receiving heart medications and the likelihood of MI in the general population is distorted, even when adjusting for confounders. A better strategy would be to sample randomly from the general adult population.

Causal diagrams are especially useful for planning one's data analysis. They promote good statistical practice by requiring investigators to think about the topic of interest and define their hypotheses about causal links *a priori*, which limits both the number of statistical tests that are performed and the typical post-hoc rationalizations given when unexpected associations are found. Causal diagrams also allow identification of non-causal pathways linking the exposure-outcome dyad of interest, which are said to be "open" if confounders (or fork connections) of the exposure-outcome *are not* adjusted for or if collider variables *are* adjusted for, which would result in selection bias and distortion of the causal relationship in an exposure-outcome pair of interest.[1] Mediator variables, the middles of chain connections, are often left unadjusted, because investigators are usually interested in the *total* effect of an exposure on an outcome.

For example, let's return to our study which aims to determine whether receiving a heart medication changes the likelihood of experiencing a myocardial infarction in the general population. According to our DAG (Table 1A), the heart medication has some direct effect on MI incidence, but it also affects blood pressure, which itself affects the likelihood of developing a MI. During the data analysis, if the investigator unwittingly adjusts for blood pressure by including it in a multiple regression analysis as a potential confounding variable, they are only evaluating the *direct* effects of the drug on likelihood of MI, which could be so small as to be inconsequential. If the data are analyzed this way, the investigator might erroneously conclude that the heart medication is useless for their patients. Had the investigator analyzed the data without adjusting for blood pressure, which is more relevant for assessing the total effect of the medication, they might have found a marked reduction in the likelihood of MI due to the mediating effect of lowered blood pressure. This second way of analyzing the data is more relevant for answering the clinical question at hand, but it can be easy to choose the first method if a causal model was not employed.

Euthanasia as a Fundamental Limitation in Veterinary Research

One important difference between clinical research in human and veterinary patients, besides scale of the enterprises, is the widespread use of euthanasia in veterinary medicine. Euthanasia complicates veterinary research at a fundamental level because observational veterinary medical studies with a survival outcome often do not adequately address its impact.[2,3] Because euthanasia is typically considered a preferable outcome to natural death from a welfare perspective, the number of patients that are euthanized is typically far greater than those which die naturally. These euthanized patients are often combined, however, with natural deaths into a single 'nonsurvival'

outcome. This approach can be problematic because clinicians are not perfect predictors of whether an animal would or would not have died with additional treatment.

A recent study found that, while veterinarians are reasonably accurate (81%) at predicting which dogs admitted to a university teaching hospital would survive to discharge, the incorrect predictions tended to be pessimistic and had the potential to lead to discontinuation of care or euthanasia of animals that could have had better outcomes.[2] Similarly, another study found an increased relative risk of euthanasia for dogs with non-traumatic hemoabdomen when they were managed by interns compared with non-interns.[4] A corollary of these studies is that if a clinician believes a particular clinical sign is a negative prognostic indicator and then consciously or unconsciously influences the owners of animals with that clinical sign to opt for euthanasia, they have effectively created or inflated a negative association between the presence of the clinical sign and patient survival.

Despite most 'risk factor' studies failing to employ causal models, investigators often present or interpret measures of risk (e.g., odds ratios, risk ratios) using causal language, such as 'was affected by' or 'influenced'.[5] To readers these measures can seem like an estimate of the direct effect of the risk factor on a survival outcome (Figure 2A). In actuality, these measurements are the *total* effects of the risk factor on the outcome, including the effects that are mediated by the veterinarian's belief about prognosis and the treatment effort.

A veterinarian's assessment of an animal's prognosis, while impacted by a particular clinical sign, is also affected by other clinical signs, prior treatment effort, the perceived ability and willingness of the client to continue treatment, the veterinarian's attitudes, beliefs, and prior experience, and available published evidence (Figure 2B). Likewise, the treatment effort depends on the veterinarian's assessment of prognosis and the client's ability to provide for treatment. Because these factors are highly situation-dependent and dramatically affect each patient's outcome – one client's inability to treat is not necessarily relevant to another case – mediation analysis could be used to estimate the direct effects of a particular clinical sign on survival. Mediation analysis aims to decompose the *total* effects of an exposure on an outcome into the *mediated* effect and *direct* effect.

Ideally, mediation analyses would simply adjust for the treatment received, eliminating the mediated effects of both the treatment received and the veterinarian's belief about prognosis, leaving only the direct effect of the exposure, the clinical variable of interest, on survival (Figure 2C). Because treatment received is not easily measured on a continuous scale, use of an ordinal scale of treatment-effort may be useful. In certain disease processes, like ruptured splenic hemangiosarcoma, there may be obvious treatment-effort strata such as no treatment, supportive care alone, supportive care and surgery, and surgery and chemotherapy. In other disease processes where the treatment plan is more ambiguous, the use of generalized treatment-effort strata could be considered.[3] While this stratified approach could be very useful for understanding disease prognosis with different levels of treatment-effort, it would require large numbers of patients in each stratum to ensure that each strata has a sufficient number of patients to estimate reasonably precise measures of risk, such as the association of hematocrit level with 7-day survival among canine hemangiosarcoma patients receiving different levels of treatment-effort. Unfortunately, large datasets are the exception rather than the rule in veterinary medicine. Ideally, there would be a way to avoid stratifying data, yet still mediate the effects of both the treatment received and the veterinarian belief about prognosis.

Counterfactuals in Euthanasia

One way a mediation analysis can account for a veterinarian's belief about prognosis is to use counterfactuals. Counterfactuals are an essential component of the causal inference toolkit and are basically imagined realities that answer the question "what would have happened had some past action been different?"[6] For euthanasia, the counterfactual question is "would this animal have survived had it not been euthanized?"

Randomized controlled trials (RCTs) are based on counterfactual thinking. In RCTs, we want to know the treatment effect for an average patient, namely what would happen if they received the new intervention versus standard care or a placebo (control group) regimen? Unfortunately, we often

cannot assign an individual to both treatment and control groups and, thus, can only observe the outcome for their assigned condition (treatment or control); the counterfactual condition outcome cannot be directly observed or assessed on an individual level. Fortunately, through randomization and with a sufficient sample size, the average patient in the treatment group is sufficiently similar to the average patient in the control group, so the average control patient provides a surrogate counterfactual outcome for the average treatment patient. This idea is called exchangeability, and, in essence, is an assumption that had the control group instead received treatment and the treatment group received the control regimen, the average treatment effect would remain the same. Without randomization, however, the exchangeability of treated (or otherwise exposed) groups with their non-treated (or unexposed) counterparts is called into question. Individuals in the treated/exposed group may have a different average age, disease severity, or other set of factors that might impact their outcome and differ from those in the control/unexposed group. Because of this, most observational studies lack exchangeability.

Some relatively recent advances in observational studies aim to improve exchangeability by using demographic/clinical/social variables to estimate how likely a patient was to be treated (or exposed). In one of the methods, propensity score-matching, pairs of individuals are often matched based on having similar propensity to be treated but only one of whom actually received treatment. In a second method, inverse probability of treatment weighting, the outcomes of individuals who had a low probability of receiving treatment but did receive treatment are weighted relatively greater than those with a high probability of receiving treatment that actually received treatment; the same is true for the non-treatment groups. Unfortunately, however, these methods require large amounts of data, particularly in propensity score-matching which requires adequate matches for each treated individual to avoid dropping unmatched individuals from the analyses.

Given the typically limited sample sizes in veterinary medicine, adjusting for treatment-effort or using propensity score-matching or inverse probability of treatment weighting is often infeasible (Figure 3A). In such cases, counterfactual outcome adjudication committees (COACs) can be used to estimate counterfactual outcomes for euthanized patients included in a study. In essence, this would require having a panel of three veterinarians – this number being recommended in human medical clinical adjudication committees[7] – who were not involved in the animal's care would review the available clinical data – history and physical examination findings, clinical pathology, diagnostic imaging, biopsy results, and prior treatment – of each euthanized patient and envision a scenario in which the animal received additional treatment (with a defined upper limit of treatment, such as that which can be reasonably provided in a tertiary referral hospital) instead of being euthanized. Each veterinarian in the COAC would then estimate the animal's probability of survival in this counterfactual world, taking into consideration the clinical data, published evidence, and their own prior experience. These estimates would then be averaged and used to simulate a counterfactual outcome.

For example, let's envision a scenario in which COAC members estimate an animal's probability of survival with counterfactual treatment at 40%, 30%, and 20%. In this scenario, a simulated counterfactual outcome will be different from the actual outcome (death) in about 30% of simulations. These data with a simulated outcome for each euthanized patient can then be analyzed within a multivariable adjusted regression analysis using a particular study's, using a particular study's DAG to adjust for other variables that could affect both the exposure of interest and the animal's survival (Figure 3B).

Adjudicating the counterfactual outcomes of only those patients who are euthanized may seem odd to some. However, these are the only patients at risk of having their outcome misclassified as essentially equivalent to natural deaths because withdrawal of all care without euthanasia is very rare. The patients who survive are known to survive, and the patients who died a natural death were essentially all undergoing active treatment for their condition. By analyzing data in this way, researchers can partially account for biased estimates of causal effects of clinical variables on survival arising from situation-specific variables such as veterinarian-held beliefs about prognosis and client

ability to pursue treatment. The unintended inclusion of these variables in analyses is akin to overfitting in that the estimations of risk (e.g., ORs) correspond too closely to that particular set of data (i.e., those patients, those clients, those veterinarians) and may therefore fail to be generalizable. Thus, the use of counterfactual outcomes may be viewed as a method to reduce overfitting similar to how shrinkage regression is used to reduce overoptimism about clinical prediction model performance.

Strengths and Limitations of Counterfactual Outcome Adjudication Committees

One of the biggest strengths of this approach is that it provides estimates of causal effects rather than associations. By understanding how different exposures can *affect* survival rather than simply be statistically associated with survival, clinicians can work to mitigate those exposures to improve the odds of good outcomes. Another strength of this approach is that it borrows a concept common in human medical trials, the clinical adjudication committee, and modifies it to address a pervasive problem in veterinary research. Clinical adjudication committees are relatively common in human clinical trials to review clinical data and classify outcomes or events, such as MI, stroke, or other diagnoses. These committees have also begun to be used in some veterinary studies, including those based on data from the Golden Retriever Lifetime Study.[8]

One of the bigger limitations of this approach is that it requires simulating data, which is foreign to many clinician-scientists and may limit adoption of this proposed methodology. Understandably, some researchers may be reticent to ask other clinicians to review patient data and opine on an outcome that might have been. Those same researchers would be wise to consider that their data already include the treating clinician's beliefs about prognosis baked into them. Along those lines, the counterfactual survival estimates of the COAC are subject to some of the same limitations as the treating clinician's prognosis estimates, albeit without client-specific information like willingness and ability to treat. The other major limitation is one of logistics, and that there may be limited ability to cover expenses associated with a COAC. These costs might include labor – for both recruiting expert clinicians and paying those experts – and data storage and distribution. Additionally, identification of suitable subject matter experts may be difficult depending on the particulars of the study.

Conclusion

The ability to euthanize patients is essential to the practice of ethical veterinary medicine, but it creates considerable difficulties for investigators aiming to produce generalizable results in studies with a survival outcome. Using DAGs, it is possible to identify the treating veterinarian's belief about an animal's prognosis as a variable mediating the effect of an exposure variable of interest on the survival outcome. Because the treating veterinarian rarely estimates the probability of survival with additional treatment numerically (and does not record it in the medical record), I outline a strategy by which counterfactual outcome adjudication committees may review clinical data and provide their own estimate of probability of survival with additional treatment. By simulating the outcomes of euthanized patients according to counterfactual survival probabilities and analyzing the data using the simulated outcomes, investigators may better understand the causal relationships between exposures of interest and survival, reducing the impact of external factors such as financial or emotional limitations of clients.

References

1. Hernán MA, Monge S. Selection bias due to conditioning on a collider. BMJ. 2023; 381:1135. doi: 10.1136/bmj.p1135.

2. Le Gal A, Barfield DM, Wignall RH, Cook SD. Outcome prediction in dogs admitted through the emergency room: accuracy of staff prediction and comparison with an illness severity stratification system for hospitalized dogs. J Vet Emerg Crit Care. 2024; 34(1):69-75. doi: 10.1111/vec.13350.

3. Cummings CO, Krucik DD. Not all euthanasias are alike: stratifying treatment effort to facilitate better prognosis prediction. Vet Rec. 2023;192(2):72–4. doi: 10.1002/vetr.2661.

4. Molitoris A, Pfaff A, Cudney S, Laforcade A de. Early career clinicians euthanize more dogs with nontraumatic hemoabdomen but not gastric dilatation and volvulus than more experienced clinicians. J Am Vet Med Assoc. 2022; 260(12):1514-1517. doi: 10.2460/javma.22.05.0198.

5. Sargeant JM, O'Connor AM, Totton SC, Vriezen ER. Watch your language: an exploration of the use of causal wording in veterinary observational research. Front Vet Sci. 2022; 9:1004801. doi: 10.3389/fvets.2022.1004801.

6. Höfler M. Causal inference based on counterfactuals. BMC Med Res Methodol. 2005; 5:28. doi: 10.1186/1471-2288-5-28.

7. Kahan BC, Feagan B, Jairath V. A comparison of approaches for adjudicating outcomes in clinical trials. Trials. 2017; 18(1):266. doi: 10.1186/s13063-017-1995-3.

8. Labadie J, Swafford B, DePena M, Tietje K, Page R, Patterson-Kane J. Cohort profile: The Golden Retriever Lifetime Study (GRLS). PLOS ONE. 2022; 17(6):e0269425. doi: 10.1371/journal.pone.0269425.