TITLE: Observational Study Design in Veterinary Pathology, Part 1: Study Design

AUTHORS: Caswell, J L; Bassel, L L; Rothenburger, J L; Gröne, A; Sargeant, J M; Beck, A P; Ekman, S; Gibson-Corley, K N; Kuiken, T; LaDouceur, E E B; Meyerholz, D K; Origgi, F C; Posthaus, H; Priestnall, S L; Ressel, L; Sharkey, L; Teixeira, L B C; Uchida, K; Ward, J M; Webster, J D; Yamate, J

JOURNAL TITLE: Veterinary Pathology

PUBLICATION DATE: 2 August, 2018 (online)

PUBLISHER: SAGE Publications

DOI: 10.1177%2F0300985818785705

Veterinary Pathology

Observational Study Design in Veterinary Pathology. Part 1: Study Design.

Journal:	*Veterinary Pathology*
Manuscript ID	Draft
Manuscript Type:	Commentary
Date Submitted by the Author:	n/a
Complete List of Authors:	Caswell, Jeff; University of Guelph, Bassel, Laura; University of Guelph, Pathobiology Rothenburger, Jamie; University of Calgary Faculty of Veterinary Medicine, Department of Ecosystem and Public Health Gröne, Andrea; Universiteit Utrecht Faculteit Diergeneeskunde, Department of Pathobiology Sargeant, Jan; University of Guelph, Department of Population Medicine Beck, Amanda; Yeshiva University Albert Einstein College of Medicine, Pathology Ekman, Stina; Swedish University of Agricultural Sciences, Pathology; Gibson-Corley, Katherine; University of Iowa, Pathology Kuiken, Thijs; Erasmus MC, LaDouceur, Elise; Joint Pathology Center Meyerholz, David; University of Iowa Carver College of Medicine, Origgi, Francesco; University of Bern, FIWI-ITPA Posthaus, Horst; Institute of Animal Pathology, Vetsuisse Faculty Priestnall, Simon; The Royal Veterinary College, Pathology and Pathogen Biology Ressel, Lorenzo; School of Veterinary Science, University of Liverpool, Section of Veterinary Pathology; Sharkey, Leslie; Tufts University Cummings School of Veterinary Medicine, Department of Clinical Sciences Teixeira, Leandro; University of Wisconsin-Madison, Pathobiological Sciences Uchida, Kazuyuki; The University of Tokyo , Department of Veterinary Pathology Ward, Jerrold Webster, Joshua Yamate, Jyoji; Osaka Prefecture University , Veterinary Pathology
Keywords:	Reproducibility of Results, Research design, Epidemiology, Pathology, Observational studies, Study design, Hypothesis testing
Abstract:	Observational studies are the basis for much of our knowledge of veterinary pathology and are highly relevant to the daily practice of pathology. However, recommendations for conducting pathology-based observational studies are not readily available. In part 1 of this series, we offer advice on planning and conducting an observational study with

	examples from the veterinary pathology literature. Investigators should recognize the importance of creativity, insight and innovation in devising studies that solve problems and fill important gaps in knowledge. Studies should focus on specific and testable hypotheses, questions or objectives. The methodology is developed to support these goals. We consider the merits and limitations of different types of analytic and descriptive studies, and of prospective versus retrospective enrollment. Investigators should define clear inclusion and exclusion criteria and select adequate numbers of study subjects, including careful selection of the most appropriate controls. Studies of causality must consider the temporal relationships between variables, and the advantages of measuring incident cases rather than prevalent cases. Investigators must consider unique aspects of studies based on archived laboratory case material, and take particular care to consider and mitigate the potential for selection bias and information bias. We close by discussing approaches to adding value and impact to observational studies. Part 2 of the series focuses on methodology and validation of methods.

Observational Study Design in Veterinary Pathology. Part 1: Study Design.

Jeff L. Caswell, Laura L. Bassel, Jamie L. Rothenburger, Andrea Gröne, Jan M. Sargeant, Amanda P. Beck, Stina Ekman, Katherine N. Gibson-Corley, Thijs Kuiken, Elise E.B. LaDouceur, David K. Meyerholz, Francesco C. Origgi, Horst Posthaus, Simon L. Priestnall, Lorenzo Ressel, Leslie Sharkey, Leandro B.C. Teixeira, Kazuyuki Uchida, Jerrold M. Ward, Joshua D. Webster, Jyoji Yamate.

Jeff L Caswell, jcaswell@uoguelph.ca, Department of Pathobiology, University of Guelph, Guelph, ON, Canada N1G 2W1

Laura L Bassel, lbassel@uoguelph.ca, Department of Pathobiology, University of Guelph, Guelph, ON, Canada N1G 2W1

Jamie L. Rothenburger, jamie.rothenburger@ucalgary.ca, Department of Ecosystem and Public Health; Canadian Wildlife Health Cooperative (Alberta), Faculty of Veterinary Medicine, University of Calgary, 3280 Hospital Dr. NW, Calgary, AB, Canada T2N 4Z6

Andrea Gröne, a.grone@uu.nl, Department of Pathobiology, Faculty of Veterinary Medicine, Utrecht University, Utrecht, the Netherlands

Jan M. Sargeant, sargeanj@uoguelph.ca, Department of Population Medicine and Centre for Public Health and Zoonoses, University of Guelph, Guelph, ON, Canada N1G 2W1.

Amanda P. Beck, Amanda.beck@einstein.yu.edu, Albert Einstein College of Medicine, 1301 Morris Park Ave, Bronx, NY 10461

Stina Ekman, Stina.Ekman@slu.se, Department of Biomedicine and Veterinary Public Health, Swedish University of Agricultural Sciences, Box 7028, 75007 Uppsala, Sweden

Katherine N. Gibson-Corley, katherine-gibson-corley@uiowa.edu, Department of Pathology, Roy J. and Lucille A. Carver College of Medicine, University of Iowa, Iowa City, IA 52242

Thijs Kuiken, t.kuiken@erasmusmc.nl, Department of Viroscience, Erasmus University Medical Centre, Rotterdam, The Netherlands.

Elise E.B. LaDouceur, elise.e.ladouceur.civ@mail.mil, Joint Pathology Center, 606 Stephen Sitter Ave, Silver Spring, MD 20910

David K. Meyerholz, david-meyerholz@uiowa.edu, University of Iowa Carver College of Medicine; 1165 Medical Laboratories, University of Iowa Carver College of Medicine, Iowa City, Iowa, 52242

Francesco C. Origgi, francesco.origgi@vetsuisse.unibe.ch, Centre for Fish and Wildlife Health, Vetsuisse Faculty, University of Bern, Switzerland

Horst Posthaus, horst.posthaus@vetsuisse.unibe.ch, Institute of Animal Pathology, Vetsuisse-Faculty, University of Bern, Switzerland

Simon L Priestnall, spriestnall@rvc.ac.uk, Dept Pathobiology & Population Sciences, The Royal Veterinary College, Hatfield, United Kingdom

Lorenzo Ressel, Department of Veterinary Pathology and Public Health, Institute of Veterinary Science, University of Liverpool, Liverpool, United Kingdom.

Leslie Sharkey, Leslie.Sharkey@tufts.edu, Department of Clinical Sciences, Cummings School of Veterinary Medicine, Tufts University, 200 Westboro Rd., N. Grafton, MA 01536

Leandro B.C. Teixeira, leandro.teixeira@wisc.edu, Department of Pathobiological Sciences, University of Wisconsin-Madison, 2015 Linden Drive, Madison, WI 53706, USA (608) 262-8089

Kazuyuki Uchida, auchidak@mail.ecc.u-tokyo.ac.jp, Department of Veterinary Pathology, The University of Tokyo, Tokyo 113-8657, Japan

Jerrold M Ward, veterinarypathology@gmail.com, GlobalVetPathology, Montgomery Village, Maryland

Joshua D. Webster, websterj@gene.com, Genentech, 1 DNA Way, South San Francisco, CA 94080

Jyoji Yamate, yamate@vet.osakafu-u.ac.jp, Laboratory of Veterinary Pathology, Graduate School of Life and Environmental Sciences, Osaka Prefecture University, 1-58 Rinku-Ourai-Kita, Izumisano City, Osaka 598-8531, Japan.

Corresponding author:

Jeff L Caswell jcaswell@uoguelph.ca

Department of Pathobiology, University of Guelph, Guelph, ON, Canada N1G 2W1

519-824-4120 ext 54555

Abstract

Observational studies are the basis for much of our knowledge of veterinary pathology and are highly relevant to the daily practice of pathology. However, recommendations for conducting pathology-based observational studies are not readily available. In part 1 of this series, we offer advice on planning and conducting an observational study with examples from the veterinary pathology literature. Investigators should recognize the importance of creativity, insight and innovation in devising studies that solve problems and fill important gaps in knowledge. Studies should focus on specific and testable hypotheses, questions or objectives. The methodology is developed to support these goals. We consider the merits and limitations of different types of analytic and descriptive studies, and of prospective versus retrospective enrollment. Investigators should define clear inclusion and exclusion criteria and select adequate numbers of study subjects, including careful selection of the most appropriate controls. Studies of causality must consider the temporal relationships between variables, and the advantages of measuring incident cases rather than prevalent cases. Investigators must consider unique aspects of studies based on archived laboratory case material, and take particular care to consider and mitigate the potential for selection bias and information bias. We close by discussing approaches to adding value and impact to observational studies. Part 2 of the series focuses on methodology and validation of methods.

Keywords

Reproducibility of Results, Research design, Epidemiology, Pathology, Descriptive studies, Observational studies, Study design, Case-control, Cohort, Hypothesis, Bias, Laboratory medicine

Observational studies are the foundation for most of the current knowledge that veterinary pathologists apply to their daily practice. The published literature contains considerable advice on designing and reporting observational studies, including the recent STROBE-Vet guidelines.[31,39] However, these publications are oriented to epidemiology and often focus on studies of causation, whereas pathology-based studies more often investigate mechanisms or consequences of disease. Moreover, investigations based on archived laboratory case material have unique caveats and limitations that should be recognized in the early phases of study design.

Here, editors and editorial board members of *Veterinary Pathology* and our colleagues present the sequential steps in devising and conducting observational studies in veterinary pathology. We also provide examples from published articles for clarity. This article is not intended as a list of requirements to publish in *Veterinary Pathology* because application of these principles will depend on the study context. Instead, the article describes principles intended to stimulate thinking on effective study design.

This article—the first of a 2-part series—focuses on design and development of observational studies. We discuss devising the study, developing the rationale, and forming a specific hypothesis, question or objective. Next, we consider the details of study design: choosing between descriptive and analytic studies, types of analytic studies, prospective vs retrospective enrollment, study design considerations that pertain to causal inferences, selection and numbers of subjects for the study, and issues of bias, confounding and chance associations. Finally, we consider the need for careful critique of the study design, and approaches to adding value and rigor. The second article of the series[8] addresses methodology and validation of methods.

We should clarify a few terms. Study subjects are the individuals being studied, such as the cases and controls. Studies of causal association measure an exposure and an outcome. The exposure (independent variable) is presumed to precede the outcome (dependent variable). Depending on the study design, the disease could either be the exposure or the outcome. For example, a virus infection could be the exposure and pneumonia is the outcome, or pneumonia could be the exposure and serum fibrinogen levels are the outcome.

Various study types, as defined in Figure 1, can be considered when investigating the hypothesis that panleukopenia virus causes restrictive cardiomyopathy in cats.[29] Panleukopenia virus infection is the exposure, and development of restrictive cardiomyopathy is the outcome. In an *experimental study*, the exposure is manipulated: cats could be challenged with virus or saline control to determine the effect on development of restrictive cardiomyopathy. In contrast, an *observational study* would investigate a population of cats without controlling the exposure. Observational studies come in two flavors: descriptive and analytic. A *descriptive study* could report one or more cases of restrictive cardiomyopathy and indicate how many had evidence of panleukopenia virus infection. Or, a descriptive study could report on cats with natural panleukopenia virus infection, mentioning the number that had concurrent restrictive cardiomyopathy. In contrast, an *analytic study* compares two groups, such as reporting the frequency of panleukopenia virus infection in cats with restrictive cardiomyopathy and in cats without restrictive cardiomyopathy.

Experimental studies sit proudly atop the hierarchy of evidence because exposures can be precisely controlled. But, let us not abandon our respect for observational studies!

Observational studies investigate the very animals that comprise pathologists' routine caseload and are therefore highly relevant to daily practice. Observational studies are essential when experimental studies are impossible or undesirable. They are often easier and less expensive to carry out because study subjects and data may already be available or more easily obtained, and are well-suited to the analysis of conditions that develop over a long period of time. Many risk factors or outcomes can be investigated simultaneously, including interactions among variables. Observational studies usually contribute an early foundation of knowledge, before it becomes possible—if ever—to study the disease experimentally. Finally, observational studies are the most frequent type published within the pages of *Veterinary Pathology* (Figure 2), so it is prudent to optimize the design of these studies, as we continue to welcome them as a key basis for knowledge in veterinary pathology.

Devising an observational study

This earliest step in the study—choosing a topic—shapes its eventual impact. We suggest a formula for devising observational studies that will have value:

1. Identify important problems and gaps in knowledge, and work toward solutions for them.
2. Have an innovative mindset, being open to and actively searching for new possibilities. Consider observations that don't fit with existing knowledge, and what they might mean for alternative understanding. Consider alternative interpretations of existing observations, and what might be done to evaluate differing explanations.
3. Use the scientific method: observations, experiences, knowledge→ clearly formulate a question or identify a problem→ create a hypothesis→ design and conduct an observational study→ critically analyze the results, their inferences and implications→ (communicate findings)→ refine questions/hypotheses and repeat.
4. Apply novel methods to existing problems, if they open new areas of investigation. Novel methods are not enough by themselves; they must lead to new and meaningful knowledge. But, innovative methodologies can offer new ways of probing old problems; a key that opens a previously locked door.
5. Throughout this process, recognize the essential role of creativity. A study is dull and meaningless without the imaginative insights and ideas that have been termed the creative, aha or eureka moments, the happy thought, or the art of discovery.
6. When unexpected but seemingly valid results emerge, resist the tendency to force them into the mold of prior thinking. Exciting advances in knowledge are based on troublesome and unanticipated findings. Let the data speak.

Most studies take unexpected twists and turns as investigators encounter and overcome challenges, and as surprising findings emerge. The initial plan will be modified accordingly: research is an iterative process that requires reflection and critical analysis at each stage of the study (Table 1).

Creating the hypothesis, question or objective

The hypothesis, question or objective is the central pillar of the study that determines the appropriate methodology and frames the anticipated findings (Figure 3). In crafting the manuscript, the Introduction, Methods, Results and Discussion are all built around the hypothesis or question. Studies with a strong hypothesis, question or objective are likely to yield specific findings of interest and can be clearly presented to readers. Studies that are focused on applying a new method or those in which the hypothesis, question or objective

were developed as the manuscript was being written often lack clear findings of value and do not have a strong narrative.

The hypothesis, question or objective must be precise and specific. The aim—if the study proceeds according to plan—is for the results to definitively confirm or refute the hypothesis, or conclusively answer the question, or completely satisfy the objectives. The objectives need not be grandiose or world-changing but must be precisely achievable: vague or unattainable objectives are not of value as a solid basis for a study. Recent studies provide examples of effective, specific and testable hypotheses: "the histologic diagnosis of pectinate ligament dysplasia (PLD) [does] not correlate with the gonioscopic diagnosis of PLD, and PLD cannot be diagnosed solely by routine histological examination in canine globes affected with chronic glaucoma",[3] and "myocardial CPV-2 infection is … associated with cardiac damage in dogs less than 2 years old."[16]

Hypotheses must be specified before the study is conducted. If hypotheses are formed after observation of the data then the study is merely exploratory, and testing the hypothesis in a new population of study subjects would be needed to confirm the hypothesis. When hypotheses are formed as the paper is being written, this simply fits the "hypothesis" to the observed data. This is the reverse sequence—the tail now wags the dog—and thus invalidates the merits of hypothesis testing.

The methodology is not part of the hypothesis, question or objective. The methodology is subservient and developed subsequently (Figure 3). Too often we think of cool methods and only later create a study objective, but this is the reverse of effective study design. Investigations that are not built upon on specific objectives can become an exercise in data collection with the hope of discovering an unexpected association. This may yield interesting data but is highly exploratory, and a confirmatory study would be necessary to validate such an association. In the same way, studies that measure a myriad of parameters generate heaps of information, but can become unfocused and lack statistical power to make valid inferences.

Descriptive vs analytic studies

What study design is most appropriate and practical to address the hypothesis, question or objective of the study? Here, we consider the gritty details of study design: descriptive vs analytic studies, the merits of various types of analytic studies, retrospective vs prospective enrollment, the number of study subjects, validation of study subjects, considerations of causal inferences, and the thorny topics of bias, confounding and chance associations.

Descriptive studies are sometimes dismissed as the poor cousins of designed studies, that provide only weak evidence because unmeasured variables are not controlled and have an unknown impact on the findings. Further, cases represented in laboratory archives are a highly selected population that may differ in important ways from those cases of the same disease that were never sampled. For instance, those dogs whose tumors were biopsied and subsequently archived may have a substantially different clinical outcome from those dogs whose owners did not pursue advanced diagnostic tests. Finally, the lack of a control group leaves readers wondering whether the observed findings might also be seen in some normal animals, particularly for species or tissues not often examined. Microscopic observations in marine invertebrates, inclusion bodies in the ganglia of coatis, and the variety of age-related

lesions in older animals provide examples of "background" findings that might be incorrectly attributed to a disease if controls were not also examined.[2,13,24,32] These issues are particularly pronounced for single-animal case reports, where the relationship between 2 findings might be explained by a host of unmeasured factors.

Despite these limitations, descriptive studies provide undeniable value to the daily practice of veterinary pathology. They focus on communicating objective factual observations, relatively free of inference. As keepers of the archive, pathologists have unique access to a nearly unlimited collection of laboratory samples. For some questions, descriptive studies may be the best approach. For example, in a descriptive cohort study, a single defined population of animals initially free of the outcome is followed over time to determine the incidence of a disease or an outcome of the disease.[38] Examples include the incidence of uterine decidual reaction in mice subjected to a superovulation protocol,[34] and the incidence of recurrence after excision of feline epitheliotropic mastocytic conjunctivitis.[5] Finally, the process of marshalling these cases for a study may identify patterns and generate hypotheses not considered during the routine processing of case material. Much of our knowledge in veterinary pathology is rooted in descriptive studies, and some of our most-downloaded and most-cited articles are descriptive studies of new disease conditions. Veterinary pathologists should not be apologetic about the position descriptive studies occupy on those evidence hierarchies that were designed for evaluating human medical treatments.[11]

Analytic studies offer important advantages over descriptive studies because they formally compare results between two groups that differ with respect to the exposure or the outcome (Table 2). Descriptive studies have no control group, so it is impossible to determine if certain findings are true features of the disease or if they are alternatively due to an unrelated characteristic of the population or the method of acquiring the study subjects. When it is relevant to the study objectives, including a meaningful control group can add considerable value and impact to observational studies (Figure 4). If the objective of your study is to describe or characterize, try changing it to compare for a more powerful study design.

An overview of the classic types of observational studies is provided in Figure 1 and detailed elsewhere.[14,37] The merits and limitations of different analytic study designs are outlined in Table 3.

Prospective vs retrospective enrollment

Retrospective enrollment makes use of existing materials and data, which is easier, faster and less expensive, and generally allows increased numbers of study subjects for greater statistical power. Most studies published in *Veterinary Pathology* involve retrospective enrollment because veterinary pathologists have such easy access to marvelous archives of case material.

Conversely, prospective enrollment allows a standardized approach to sampling and analysis, and the scope of data collection is intentionally designed. Thus, prospective enrollment may avoid bias and reduce variability by minimizing unintentional differences among samples. Furthermore, prospective sampling may be necessary for specialized analyses, such as flow cytometry or analysis of gene expression. Thus, use of prospective studies is one of the main recommendations for improving studies in pathology and laboratory medicine.[28] But, they are far more costly and time-consuming, and it may be impossible to acquire a sufficient number of cases within a reasonable time frame. It is an unstudied marvel of biology, how even common

diseases seemingly disappear once a prospective study is underway.

Study design and causal inferences

Observational studies that focus on causality or pathogenesis require particular attention to study design. In experimental studies, the subjects may be more uniform and there is controlled manipulation of the exposure (i.e. the causative agent, or the earlier event in the pathogenesis). In contrast, these factors are uncontrolled in observational studies making it inherently difficult to show causality. When an observational study reveals an association between two factors, Hill's criteria [21] (Table 4) provide a framework for considering whether the relationship is causal.

The fourth of Hill's criteria—the temporal relationship of cause and effect—can be problematic for studies using single biopsies or samples obtained after death. Specifically, it may be impossible to determine the causal sequence if the two variables are measured at a single point in time. For example, a landmark study[46] identified the association of equine multinodular pulmonary fibrosis (EMPF) and equine herpesvirus 5 (EHV-5) infection. However, case-control or cross-sectional study designs cannot confirm the sequence of causation: does EHV-5 infection cause EMPF, or alternatively does the abnormal tissue environment in EMPF favor infection with or replication of EHV-5? In this example, objective identification of the causal sequence was later supported by an experimental study[47] (Hill's 8th criterion in Table 4) and by comparative studies (Hill's 9th criterion);[45] a cohort study would be an alternative approach in other contexts.

Sometimes, the direction of causality is obvious. In a cross-sectional study of zebrafish that identified an association between the genetic mutation 'smoothened' and the occurrence of endocardiosis, it is not plausible that endocardiosis caused the genetic mutation, but it is plausible that the mutation caused endocardiosis.[12] Similarly, the causal sequence is self-evident when death is the outcome, for example that canine mammary carcinosarcoma confers a poor survival time compared to other types of mammary carcinoma.[33] In other studies it might be reasonable—based on existing knowledge—to infer a causal sequence, for example that systemic hypertension in cats with chronic renal failure led to vasa vasorum arteriopathy, rather than the converse.[23] Nonetheless, the sequence of causality is not always clear in cross-sectional and case-control studies: pancreatic islets of diabetic cats more frequently contain T and B lymphocytes compared to pancreatic islets of control cats, but we can't be sure if the lymphocytes are responding to the pathologic process in the islets, or if they caused the loss of islet cells.[48]

Longitudinal sampling of initially outcome-free animals in a cohort study (or exposure of animals known to be free of the disease, in an experimental study) may be needed to show that exposure precedes outcome. For example, the Golden Retriever Lifetime Study follows dogs that are initially cancer-free over their lifetime, and is expected to identify risk factors for later development of 4 types of cancer.[19] Studies that make use of longitudinal sampling are rare in *Veterinary Pathology*.

Consider also if the study measures new occurrences of a disease (i.e. incident cases) or existing cases in a population (i.e. prevalent cases). For prevalent cases, it may be impossible to determine if the cause (the exposure) preceded development of disease (the outcome). Furthermore, since prevalence is a factor of both incidence and duration of disease, case-control and cross-sectional study designs may not discern whether an exposure causes

development of new cases or increased survival of existing cases. For example, consider a cross-sectional study with the valid observation of a higher prevalence of amyloidosis in captive compared to free-ranging Island foxes.[17] It is plausible that factors related to captivity increase the likelihood that foxes develop amyloidosis, but an alternative explanation is that foxes with amyloidosis survive longer in captivity than in the wild. Thus, cohort studies can be logistically difficult because of the need to identify animals initially free of the outcome and then follow them over time to determine development of the outcome. Nonetheless, cohort studies are considered a stronger study design than case-control and cross-sectional studies because they measure development of new cases rather than existing cases, and confirm that the proposed cause preceded development of the outcome.

Selecting study subjects: ethics and permissions

All research involving live animals or samples obtained for the purpose of the study require approval by an institutional animal care and use committee, which ensures that the study is conducted in accordance with relevant legislation. Permits may be required to obtain or possess samples obtained from threatened species or from free-ranging wildlife. Permission may be necessary to publish findings based on case material owned by other individuals or by an institution. Written informed consent is required if samples are obtained from client-owned animals for the purpose of the study. The situation is less consistent for studies conducted on archived laboratory materials sampled for the purpose of diagnosis. In many jurisdictions, these samples may be considered the property of the laboratory depending on agreements at the time of sample submission, and written informed owner consent is not required. However, these laws vary among jurisdictions and may change over time, and we expect this could become an emerging issue in the future.

Selecting study subjects: unbiased sampling, effective controls, and inclusion and exclusion criteria

When selecting animals to include in the study, choose a contiguous series of subjects in each study group, or a randomly selected subset. It would introduce considerable bias if we included only those cases that were the most interesting, had the most solid diagnosis, or were most memorable. This is an important critique of single-animal case reports—the reported cases are highly selected and thus may not be representative—but the situation is only improved in an analytic study if the subjects are appropriately selected. Many observational studies use all of the available cases, whereas our archives contain far more controls than are necessary for the study. How do we select which controls to include? In general, selection of a subset of study subjects from the larger population should be done by refining the inclusion and exclusion criteria or by a formal random method. Other approaches—purposive, convenience or haphazard sampling—are likely to bias the outcome.

Selecting controls is key to the study design, not an afterthought. Choose controls that offer the best comparison to the population being studied, in the context of the study objectives. Often, the best controls are not normal individuals, but ones with an alternative disease. For example, in a study using calretinin immunohistochemistry to identify the neural tracts affected by equine degenerative myeloencephalopathy, 2 groups of controls were included: normal horses to validate the use of calretinin immunohistochemistry for tracing neural tracts, and horses with "other spinal disease" to show that calretinin-positive spheroids were unique to equine degenerative myeloencephalopathy and not found in other spinal diseases.[15] Similarly,

in a study that determined the sensitivity and specificity of histologically visible cilia-adherent bacteria for diagnosis of *Bordetella bronchiseptica* pneumonia compared to the gold standard of bacterial culture, other forms of bacterial pneumonia were considered to be a more appropriate control instead of normal lung.[41] To measure the specificity of surfactant protein A for diagnosis of pulmonary carcinomas, 113 non-pulmonary neoplasms were used as controls.[4] Finally, unaffected marine invertebrates were important controls, to demonstrate that the histologic findings in those with either spontaneous or experimentally induced copper toxicosis were not simply normal findings in these little-studied species.[24] Choosing the most appropriate controls is a fundamental basis for any analytic study and is completely dependent on the details of the hypothesis, question or objective of the study.

Inclusion and exclusion criteria must be defined for both study groups; that is, for the cases as well as the controls. Inclusion and exclusion criteria are a precisely detailed description of how study subjects were selected from the population and the reasons that some subjects were omitted from the study. The importance of clear inclusion and exclusion criteria is not simply to allow replication of the experimental approach. More importantly, these criteria allow readers (and indeed investigators) to understand potential sources of selection bias that could influence the study outcomes. Effective description of inclusion and exclusion criteria read as poetry to discerning journal editors:

> "A search of the archives between June 2007 and November 2014 was performed [*i.e. the method of selection of a contiguous series of cases and controls*], and cases limited to cats at least 1 year of age were identified using the keywords feline or cat and endomyocardial fibrosis, endocardial fibrosis, endocardial scar, endomyocarditis, or restrictive cardiomyopathy [*i.e. the inclusion criteria for cases*]. We excluded cases having keywords hypertrophic and dilated [*i.e. the exclusion criteria for cases*]. Control cases were identified using keywords describing acute trauma, neoplasia, or other noncardiac causes of sudden death [*i.e. the inclusion criteria for controls*]. A similar age distribution of control cases was selected from the same time period and source [*i.e. the method of matching controls and cases*]."[29]

After the initial round of selecting study subjects, confirm that each of them are assigned to the correct group. Critically evaluate that the cases are really cases and the controls are really controls, and they meet their respective inclusion and exclusion criteria. Validating the study subjects at an early stage avoids later errors introduced by reclassification and recalculation. False positives (erroneously diagnosed cases) are particularly problematic in case-control studies.

Selecting study subjects: unique aspects of archived laboratory material

Consider the target population (eg. all dogs with lymphoma), the source population from which samples were drawn (all dogs that have lymphoma samples in the laboratory archive) and the study population (the dogs entered into the study because they meet the inclusion and exclusion criteria), and how these populations might differ. For example, animals represented in laboratory archives may be more likely to have had a higher level of veterinary care, been treated with antibiotics, be affected by serious disease, and be affected by risk factors for other diseases. How will these factors affect the findings and the external validity of the study—the relevance of the findings to the general population of interest?

Both study groups should be sampled from the same population, but this is troublesome for laboratory-based studies where the archived material is of diverse and ill-defined provenance. The detailed circumstances of these animals' life circumstances are usually unknown and not often considered when selecting study subjects—particularly for the controls. Thus, there is considerable risk that study groups will differ with respect to unmeasured variables such as those shown in Table 5.

Uneven distribution of these variables between the different study groups can introduce bias or confounding. This problem—the possibility that clustering of unmeasured variables might create the false appearance of an association between the exposure and outcome being studied—is perhaps the major limitation of observational analytic studies based on archived laboratory samples. Bias and confounding are considered in more detail below.

When working with archival samples, the process of selecting study subjects is often iterative. Reviewing the details of the initially selected cases and controls usually identifies problems, and it is typical to revise and clarify the inclusion and exclusion criteria, then restart the selection process. Repeating this process is tedious, but it is far better to solidify the study population at the beginning than to make changes after collecting the data.

Selecting study subjects: numbers of study subjects

It is useful to conduct a formal sample size calculation prior to carrying out the study, to determine the number of study subjects required to identify a significant difference between study groups. Online tools are available (e.g., Statulator, http://statulator.com/SampleSize/ss1P.html; and StatCalc-EpiInfo, https://www.cdc.gov/epiinfo/index.html). If the outcome of interest is a proportion (binary scale), the calculation requires desired values for the level of confidence (typically 0.95) and statistical power (typically 0.8), as well as an estimate of the effect size. For binary variables, the effect size can be the odds ratio or risk ratio that the investigator considers to be meaningful, and this is estimated based on the anticipated proportion with the outcome in the exposure-positive and exposure-negative groups. If the outcome of interest is measured on a continuous scale, the calculation requires that investigators estimate a meaningful difference in the outcomes between the exposure groups, as well as the estimated variability in the outcome, and the desired levels for confidence and power. Thus, although the sample size calculation requires estimates for some variables unless a pilot study is done, it can provide an informative estimate of sample numbers to suggest the feasibility of finding a meaningful difference in the outcome between the exposure groups.

Inadequate number of study subjects is a common limitation of studies in pathology and laboratory medicine[28] and is a frequent critique of manuscripts submitted to *Veterinary Pathology*. Conversely, studies with large numbers of study subjects are admired by readers and reviewers. However, even if overall case numbers are large, the tendency for pathologists to be "splitters" rather than "lumpers" leads to low numbers in some categories. This was addressed in studies of canine pulmonary carcinoma and mammary carcinoma, by including sufficiently large numbers of cases—67 and 229 respectively—to permit meaningful analysis of tumor subtypes.[4,33]

Investigators have control over the number of study subjects. Studies of archived cases could cover a broader time period. It may be possible to relax the inclusion criteria and limit the exclusion criteria, and still fulfil the study objectives. Collaboration among institutions is the

most effective way to increase case numbers, and brings added benefits of increasing the external validity, establishing professional relationships, adding expert insights, and fomenting discussion of the study material. For example, an investigation of oxalate nephrosis in cheetahs included cases from Southern Africa, North America and France, and included geographic origin in the statistical analysis.[30] Finally, we should ensure that our laboratory information management systems can be effectively queried, so that a contiguous series of cases can be retrieved in a standardized manner.

Refining the number of study subjects in each group can optimize statistical power. If cases are frequent, aim for a 1:1 ratio of cases and controls. If cases are rare enough that it will be difficult to achieve statistically significant results, increasing the number of controls will increase the statistical power of the study. However, using more than 3 or 4 controls for each case increases the cost of the study without much increase in statistical power. Conversely, having fewer controls than cases would be rarely justified.

Bias, confounding and chance associations

Take a deep breath, intrepid pathologist, as we plumb the final depths of epidemiology. This road is a hard one, but leads to a truth that we all must know.

A statistically significant association between an exposure (e.g. presence of a virus in tissues) and an outcome (e.g. lesions of a particular disease) is a welcome finding in any observational study and cause for celebration. But, before considering that the relationship is causal—that the virus did indeed induce the lesion—some critical analysis is in order. Observational studies are susceptible to spurious associations that are not easy to detect, so investigators must carefully search for alternative explanations of their data.

Consider what factors might differ between the study groups, and how these differences might poison the findings of the study. The study groups obviously differ in ways defined by the inclusion and exclusion criteria, but they might be dissimilar in other ways as listed in Table 5. If the frequency or distribution of 1 of these factors differs between the 2 study groups, this could bias the association between the exposure and the outcome. For example, this might give a false appearance that the exposure was associated with the outcome, or it might lessen or obscure a true association between exposure and outcome.

These factors may be particularly problematic for laboratory data. In designing a clinical study with prospective enrollment, one would never select cases from a referral hospital and controls from a humane society practice, nor process and analyze case samples with one method and control samples with another. But these and other factors are surely variable and largely occult for archived laboratory case material, increasing the likelihood of spurious conclusions as a result of random or systematic differences between study groups. Furthermore, those clinicians, pathologists and laboratorians who originally managed and investigated the cases (and the controls) did so with full knowledge of the clinical details. Consider how this knowledge might have affected the case management or the laboratory investigation, and how these differences between study groups might affect the findings of the study.

Finally, note the importance of the “independence of study subjects”. Using study subjects that are not independent of each other violates the assumptions of many statistical analyses and may introduce bias. For example, if an otherwise heterogeneous study population contained several individuals from the same herd or household, these subjects may not be independent.

At a broader level, clustering of data is common within animal populations because of their population structure, and may involve the exposure variable, the outcome variable, or both. In addition to affecting the statistical analysis, clustering of data may lead to bias if it affects both the exposure and the outcome. Furthermore, statistical methods to control for clustering may reduce the power of study, thus requiring larger sample sizes.

Mitigation of bias, confounding and chance associations

It is important to recognize potential bias and confounding factors because their effects can be minimized by measurement, exclusion, statistical analysis, or matching.

1. Exclusion. Eliminate the effects of confounding by excluding a subset of the study subjects. In the example of selection bias from Table 6, exclude study subjects from primary care clinics, if they are few and if they complicate the association of nodal metastasis and survival.
2. Measurement. As the study is being conducted, collect data on potential sources of bias and confounding, and then compare their frequency in a data table. For example, compare the study groups with respect to factors including those listed in Table 5. Is the distribution of ages the same in cases and controls? Does the proportion of large vs small dog breeds differ between the study groups? If so, consider how the differences might affect the findings of the study. As an example, physeal lesions were studied in bulls raised in the same geographic area with similar husbandry practices. The similar ages and body weights of cases and controls suggested that these were not confounding factors.[26]
3. Analysis. Multivariable analysis or stratified analysis are frequently used to analyze and mitigate the effects of confounding. For example, multivariable analysis was used to control for the effect of age and sex in comparing the prevalence of bacterial infection in St Lawrence belugas in 1983–2002 vs 2003–2012,[25] and would be effective for analysis of the sources of bias shown in Tables 5 and 6.
4. Matching. If potentially confounding variables can be identified at the time the study is designed, study groups could be intentionally matched when study subjects are selected. For example, in a study of X-linked hereditary nephropathy in Navasota dogs, cases and controls were matched during the selection process with respect to their sex.[6] Similarly, in an investigation of the relationship between squamous cell carcinoma and papillomavirus infection, case and control samples were matched with respect to sheep breed and anatomic site.[44] However, factors that are matched cannot be analyzed as risk factors: if subjects are matched based on age, age cannot be analyzed as a risk factor for the outcome. Thus, multivariable statistical analysis may be advantageous in controlling for differences between groups while allowing for assessment of the factor of interest.

Critique the study design

Before starting data collection, it is recommended to write a study proposal and seek peer review. The act of writing forces appraisal of the relevant literature, planning and critical analysis. It tests the coherence of the various elements: the rationale, the hypothesis/question/objective, the study design and methodology, the expected findings, and the anticipated impact (Figure 3). What is our current understanding, and what is the gap in knowledge that the study aims to correct? What is the important problem that the study

addresses? Is the hypothesis, question or objective based on a clear rationale, and is it sufficiently specific? Are the study design and methodology expected to yield results that definitively test the hypothesis or answer the question? Are there conceptual flaws with respect to showing causality? Might unmeasured factors cause bias or confounding? Will the expected findings have the anticipated impact and address the problem or gap in knowledge that was described in the rationale? Revisit the questions posed in Table 1, as an approach to refining the study design and methodology. If doubt that the study results will be definitive or valuable, now is the time to refine the methods or revise the hypothesis, question or objective. A clear and detailed description of the rationale, anticipated findings, and significance of the study might seem as tedious work, but it allows effective critique of the study design, ensures that the study is solidly guided by a strong and specific hypothesis or question, and forms a guide for the decisions that must be made as the study is conducted (Figure 3). Moreover, writing the ensuing manuscript will be a breeze if this structure is in place from the beginning.

Value-added

Adopt a discovery mindset during the various phases of the study. The goal of an observational study is not usually to confirm what is known, but to discover something new. Critically analyze the emerging data: consider alternative interpretations, and what might be done to evaluate the differing possibilities. After analysis of the initial results, consider elements that could be added to give the study more value or impact. Discovery is iterative and it is a mistake to anticipate a simple progression from planning to execution to publication. Initial results beget additional investigations that greatly strengthen the overall work with limited increased effort.

Use insights from a single case as the starting point for a more comprehensive study. A study of *Bordetella bronchiseptica* pneumonia in dogs was initiated by the microscopic observation of bacteria adherent to cilia, but the analytic study yielded information well beyond that of the index case.[41] A novel herpesvirus was identified in a single bottlenose dolphin with benign genital plaques, which stimulated development of a case series, and eventually made use of banked sera from the same animals to show that seroconversion to the virus occurred at the age of onset of sexual behavior.[43] A single case report of a pig with amyloidosis was transformed by bioinformatic analysis of the amyloid amino acid sequences and in vitro testing of amyloid fibril formation to substantially advance the understanding of pathogenesis.[22] Thus, useful observational studies often arise from but go far beyond the observations on a single case.

Finally, consider value-added outcomes that give the study a broader impact. Mechanistic studies may have greater application if the pathologic findings can be related to clinical outcomes. For example, evaluating the survival of dogs with mast cell tumor was essential to the impact of studies on receptor tyrosine kinase expression[42] and cytologic grading.[7] Similarly, morphologic analysis of feline chronic kidney disease was given added clinical relevance by analyzing the relationship to measures of renal function.[9] Alternatively, consider whether an analysis of causes or risk factors could be added to a descriptive study by including an appropriate comparison group. For example, a study of endocardiosis in aging zebrafish described the pathologic findings, but also identified associations with recirculating water systems, commercial diets, and a mutant smoothened gene.[12] Likewise, a description of amyloidosis in island foxes identified increased lesion severity in older, female, and captive foxes as well as between subspecies.[17] Creativity and a discovery mindset are the keys to

identifying such opportunities for added insights. Further examples include adding genetic analysis to a study of age-related spontaneous lesions in mice,[20] comparing young and old animals to increase the value of a study of background lesions and clinical pathology parameters in laboratory beagle dogs,[2] quantitative analysis to validate the concurrence of cardiac fibrosis and chronic renal lesions in aged chimpanzees,[10] and comparing findings in wild and laboratory rats with respect to understanding the pathogenesis of cardiomyopathy in this species.[36]

These ideas are summarized in Figure 5. We hope that veterinary pathologists can apply these principles and use imagination, insight, collaboration, and laboratory archives bursting with samples to transform their daily work into focused observational studies that provide value and impact for advancing our knowledge of animal disease.

Acknowledgements

We thank Lauren Sergejewich, Siobhan O'Sullivan, and David Pearl for their contributions.

Declaration of conflicting interests

Some authors of this commentary are editors of *Veterinary Pathology*. This editorial commentary was not peer-reviewed.

Funding

Aspects of this article were supported by a grant from the Natural Sciences and Engineering Research Council of Canada (RGPIN-2017-03872, J. Caswell).

References

1. Avallone G, Pellegrino V, Roccabianca P, et al. Tyrosine Kinase Receptor Expression in Canine Liposarcoma. Vet Pathol. 2017;54:212–217.

2. Barnes J, Cotton P, Robinson S, Jacobsen M. Spontaneous Pathology and Routine Clinical Pathology Parameters in Aging Beagle Dogs: A Comparison With Adolescent and Young Adults. Vet Pathol. 2016 Mar;53:447–455.

3. Bauer BS, Sandmeyer LS, Philibert H, Feng CX, Grahn BH. Chronic Glaucoma in Dogs: Relationships Between Histologic Lesions and the Gonioscopic Diagnosis of Pectinate Ligament Dysplasia. Vet Pathol. 2016;53:1197–1203.

4. Beck J, Miller MA, Frank C, DuSold D, Ramos-Vara JA. Surfactant Protein A and Napsin A in the Immunohistochemical Characterization of Canine Pulmonary Carcinomas: Comparison With Thyroid Transcription Factor-1. Vet Pathol. 2017;54:767–774.

5. Beckwith-Cohen B, Dubielzig RR, Maggs DJ, Teixeira LBC. Feline Epitheliotropic Mastocytic Conjunctivitis in 15 Cats. Vet Pathol. 2017;54:141–146.

6. Benali SL, Lees GE, Nabity MB, et al. X-Linked Hereditary Nephropathy in Navasota Dogs: Clinical Pathology, Morphology, and Gene Expression During Disease Progression. Vet Pathol. 2016;53:803–812.

7. Camus MS, Priest HL, Koehler JW, et al. Cytologic Criteria for Mast Cell Tumor Grading in Dogs With Evaluation of Clinical Outcome. Vet Pathol. 2016;53:1117–1123.

8. Caswell JL. Observational Study Design in Veterinary Pathology. Part 2: Methodology and validation of methods. 2018;55:XXX–XXX.

9. Chakrabarti S, Syme HM, Brown CA, Elliott J. Histomorphometry of feline chronic kidney disease and correlation with markers of renal dysfunction. Vet Pathol. 2013 Jan;50:147–155.

10. Chilton J, Wilcox A, Lammey M, Meyer D. Characterization of a Cardiorenal-like Syndrome in Aged Chimpanzees (Pan troglodytes). Vet Pathol. 2016 Mar;53:417–424.

11. Concato J. Observational versus experimental studies: what's the evidence for a hierarchy? NeuroRx. 2004 Jul;1:341–347.

12. Cooper TK, Spitsbergen JM. Valvular and Mural Endocardiosis in Aging Zebrafish (Danio rerio). Vet Pathol. 2016 Mar;53:504–509.

13. Cooper TK, Garner MM, Baccon J, Mani H. Coati Bodies: Cytoplasmic Hyaline Globules in the Ganglionic Neurons of Aging Captive Coatis. Vet Pathol. 2017;54:851–854.

14. Dohoo IR, Martin SW, Stryhn, H. Veterinary Epidemiologic Research. Charlottetown: VER Inc; 2009.

15. Finno CJ, Valberg SJ, Shivers J, D'Almeida E, Armién AG. Evidence of the Primary Afferent Tracts Undergoing Neurodegeneration in Horses With Equine Degenerative Myeloencephalopathy Based on Calretinin Immunohistochemical Localization. Vet Pathol. 2016 Jan;53:77–86.

16. Ford J, McEndaffer L, Renshaw R, Molesan A, Kelly K. Parvovirus Infection Is Associated With Myocarditis and Myocardial Fibrosis in Young Dogs. Vet Pathol. 2017 Nov;54:964–971.

17. Gaffney PM, Witte C, Clifford DL, et al. Systemic Amyloid A Amyloidosis in Island Foxes (Urocyon littoralis): Severity and Risk Factors. Vet Pathol. 2016 May;53:637–647.

18. Gaynor AM, Zhu KW, Dela Cruz FN, Affolter VK, Pesavento PA. Localization of Bovine Papillomavirus Nucleic Acid in Equine Sarcoids. Vet Pathol. 2016 May;53:567–573.

19. Guy MK, Page RL, Jensen WA, et al. The Golden Retriever Lifetime Study: establishing an observational cohort study with translational relevance for human health. Philos Trans R Soc Lond, B, Biol Sci. 2015 Jul 19;370.

20. Harbison CE, Lipman RD, Bronson RT. Strain- and Diet-Related Lesion Variability in Aging DBA/2, C57BL/6, and DBA/2xC57BL/6 F1 Mice. Vet Pathol. 2016 Mar;53:468–476.

21. Hill AB. The environment and disease: association or causation? 1965. J R Soc Med. 2015 Jan;108:32–37.

22. Kamiie J, Sugahara G, Yoshimoto S, et al. Identification of a Unique Amyloid Sequence in AA Amyloidosis of a Pig Associated With Streptococcus Suis Infection. Vet Pathol. 2017;54:111–118.

23. Kohnken R, Scansen BA, Premanandan C. Vasa Vasorum Arteriopathy: Relationship With Systemic Arterial Hypertension and Other Vascular Lesions in Cats. Vet Pathol. 2017;54:475–483.

24. LaDouceur EEB, Wynne J, Garner MM, Nyaoke A, Keel MK. Lesions of Copper Toxicosis in Captive Marine Invertebrates With Comparisons to Normal Histology. Vet Pathol. 2016 May;53:648–658.

25. Lair S, Measures LN, Martineau D. Pathologic Findings and Trends in Mortality in the Beluga (Delphinapterus leucas) Population of the St Lawrence Estuary, Quebec, Canada, From 1983 to 2012. Vet Pathol. 2016 Jan;53:22–36.

26. Levi M, Dittmer KE, Gentile A, et al. Growth Plate Lesions of Fattening Bulls. Vet Pathol. 2017;54:437–444.

27. Maeda S, Ohno K, Fujiwara-Igarashi A, Uchida K, Tsujimoto H. Changes in Foxp3-Positive Regulatory T Cell Number in the Intestine of Dogs With Idiopathic Inflammatory Bowel Disease and Intestinal Lymphoma. Vet Pathol. 2016 Jan;53:102–112.

28. Marchevsky AM, Wick MR. Evidence levels for publications in pathology and laboratory medicine. Am J Clin Pathol. 2010 Mar;133:366–367.

29. McEndaffer L, Molesan A, Erb H, Kelly K. Feline Panleukopenia Virus Is Not Associated With Myocarditis or Endomyocardial Restrictive Cardiomyopathy in Cats. Vet Pathol. 2017;54:669–675.

30. Mitchell EP, Church ME, Nemser SM, et al. Pathology and Epidemiology of Oxalate Nephrosis in Cheetahs. Vet Pathol. 2017 Nov;54:977–985.

31. O'Connor AM, Sargeant JM, Dohoo IR, et al. Explanation and Elaboration Document for the STROBE-Vet Statement: Strengthening the Reporting of Observational Studies in Epidemiology-Veterinary Extension. J Vet Intern Med. 2016 Nov;30:1896–1928.

32. Radaelli E, Castiglioni V, Recordati C, et al. The Pathology of Aging 129S6/SvEvTac Mice. Vet Pathol. 2016 Mar;53:477–492.

33. Rasotto R, Berlato D, Goldschmidt MH, Zappulli V. Prognostic Significance of Canine Mammary Tumor Histologic Subtypes: An Observational Cohort Study of 229 Cases. Vet Pathol. 2017;54:571–578.

34. Robles TG, Fernández RAG, García-Palencia P, et al. Hoxa-10 and Cyclin D3 Overexpression in the Decidual Reaction in a Superovulation Protocol in Young Adult C57BL/6J Mice. Vet Pathol. 2017;54:328–335.

35. Roels E, Dourcy M, Holopainen S, et al. No Evidence of Herpesvirus Infection in West Highland White Terriers With Canine Idiopathic Pulmonary Fibrosis. Vet Pathol. 2016;53:1210–1212.

36. Rothenburger JL, Himsworth CG, Treuting PM, Leighton FA. Survey of cardiovascular pathology in wild urban Rattus norvegicus and Rattus rattus. Vet Pathol. 2015 Jan;52:201–208.

37. Sargeant JM, Kelton DF, O'Connor AM. Study designs and systematic reviews of interventions: building evidence across study designs. Zoonoses Public Health. 2014 Jun;61 Suppl 1:10–17.

38. Sargeant JM, O'Connor AM, Cullen JN, Makielski KM, Jones-Bitton A. What's in a Name? The Incorrect Use of Case Series as a Study Design Label in Studies Involving Dogs and Cats. J Vet Intern Med. 2017 Jul;31:1035–1042.

39. Sargeant JM, O'Connor AM, Dohoo IR, et al. Methods and processes of developing the strengthening the reporting of observational studies in epidemiology - veterinary (STROBE-Vet) statement. Prev Vet Med. 2016 Nov 1;134:188–196.

40. Sihvo H-K, Lindén J, Airas N, Immonen K, Valaja J, Puolanne E. Wooden Breast Myodegeneration of Pectoralis Major Muscle Over the Growth Period in Broilers. Vet Pathol. 2017;54:119–128.

41. Taha-Abdelaziz K, Bassel LL, Harness ML, Clark ME, Register KB, Caswell JL. Cilia-associated bacteria in fatal Bordetella bronchiseptica pneumonia of dogs and cats. J Vet Diagn Invest. 2016 Jul;28:369–376.

42. Thompson JJ, Morrison JA, Pearl DL, et al. Receptor Tyrosine Kinase Expression Profiles in Canine Cutaneous and Subcutaneous Mast Cell Tumors. Vet Pathol. 2016 May;53:545–558.

43. van_Elk CE, van de Bildt MWG, de Jong AAW, Osterhaus ADME, Kuiken T. Herpesvirus in bottlenose dolphins (Tursiops truncatus): cultivation, epidemiology, and associated pathology. J Wildl Dis. 2009 Oct;45:895–906.

44. Vitiello V, Burrai GP, Agus M, et al. Ovis aries Papillomavirus 3 in Ovine Cutaneous Squamous Cell Carcinoma. Vet Pathol. 2017;54:775–782.

45. Williams KJ. Gammaherpesviruses and pulmonary fibrosis: evidence from humans, horses, and rodents. Vet Pathol. 2014 Mar;51:372–384.

46. Williams KJ, Maes R, Del Piero F, et al. Equine multinodular pulmonary fibrosis: a newly recognized herpesvirus-associated fibrotic lung disease. Vet Pathol. 2007 Nov;44:849–862.

47. Williams KJ, Robinson NE, Lim A, et al. Experimental induction of pulmonary fibrosis in horses with the gammaherpesvirus equine herpesvirus 5. PLoS ONE. 2013;8:e77754.

48. Zini E, Lunardi F, Zanetti R, et al. Endocrine Pancreas in Cats With Diabetes Mellitus. Vet Pathol. 2016 Jan;53:136–144.

Figure legends

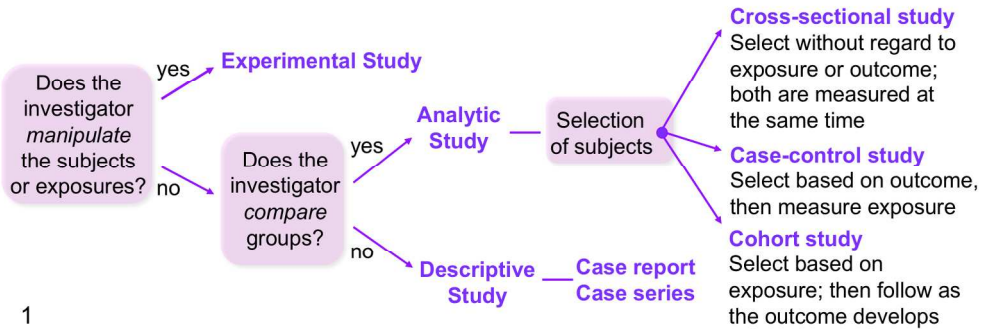
Figure 1. In an *experimental study*, the exposure (independent variable) is controlled and manipulated by the investigator. The 3 classic *observational study* designs differ in whether exposure or outcome defines how study subjects are selected. In cross-sectional studies, study subjects are selected without regard for either the exposure or the outcome, and the outcome and exposure are measured at the same time. In case-control studies, study subjects are selected based on the outcome, and the exposure is compared between groups with differing outcomes. In cohort studies, study subjects known to be free of the outcome are selected based on their exposure to the putative causal factor, then followed over time; development of the outcome is compared in study subjects with differing exposures. Examples of analytic studies are provided in Table 3. It is notable that comparison of diseased and healthy animals (often termed cases and controls by veterinary pathologists) are case-control studies only if subjects are selected based on their disease status and compared with respect to their exposure to a putative causal factor.

Figure 2. Numbers of observational studies (analytic and descriptive) and experimental studies published in *Veterinary Pathology*. Most published articles are observational studies, and most of these are descriptive.

Figure 3. Interrelationships of the various elements of study design. Studies are based on a clear, precisely worded, and specifically testable/answerable hypothesis, question or objective. The hypothesis, question or objective is supported by a clear rationale that identifies the problem or the gap in current knowledge. The study design and methods are developed to serve the hypothesis, question or objectives of the study. The methods are expected lead to an outcome that clearly confirms or refutes the hypothesis, answers the question, or fulfils the study objectives. In so doing, the anticipated results of the study fills the above-mentioned gap in knowledge and thus addresses the rationale of the study.

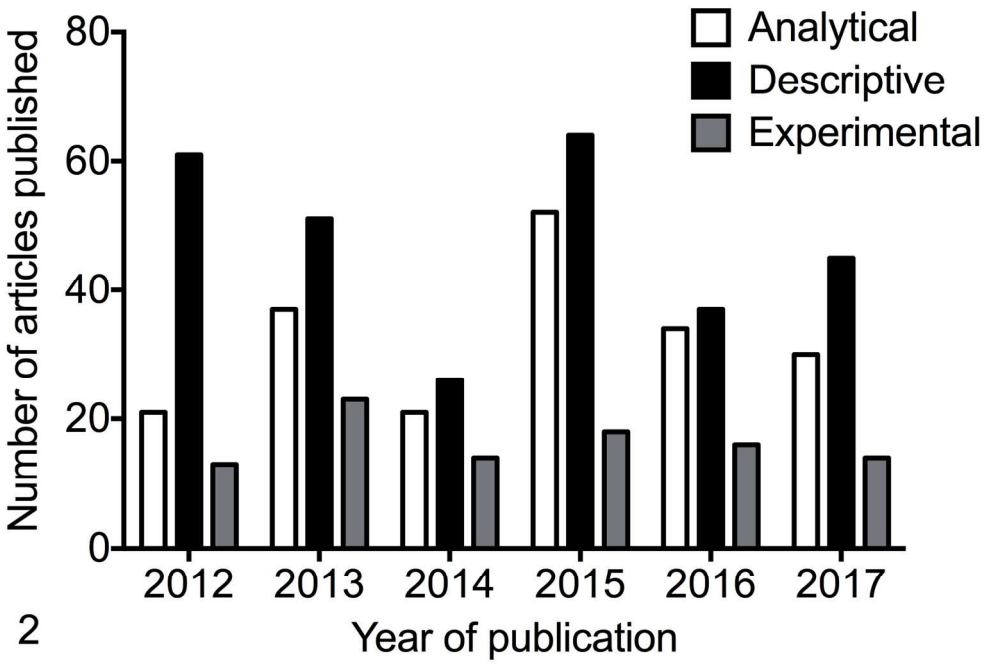
Figure 4. Citations and usage of observational studies (analytic and descriptive) and experimental studies published in *Veterinary Pathology*. The data show the number of citations (panel A) and number of downloads (panel B) per article based on year of publication (mean with 95% confidence interval). Analytic studies tend to be cited and downloaded more often than descriptive studies (* $P<0.05$). Further, analytic observational studies have similar or higher numbers of downloads and citations as experimental studies, even though the latter is classically considered more a robust approach to knowledge discovery.

Figure 5. Considerations for the effective design of observational studies in veterinary pathology.

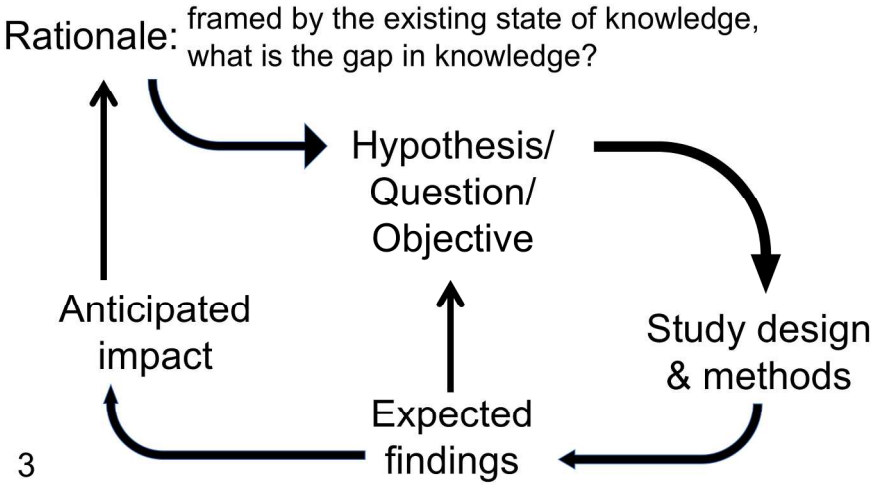


1

294x100mm (207 x 207 DPI)

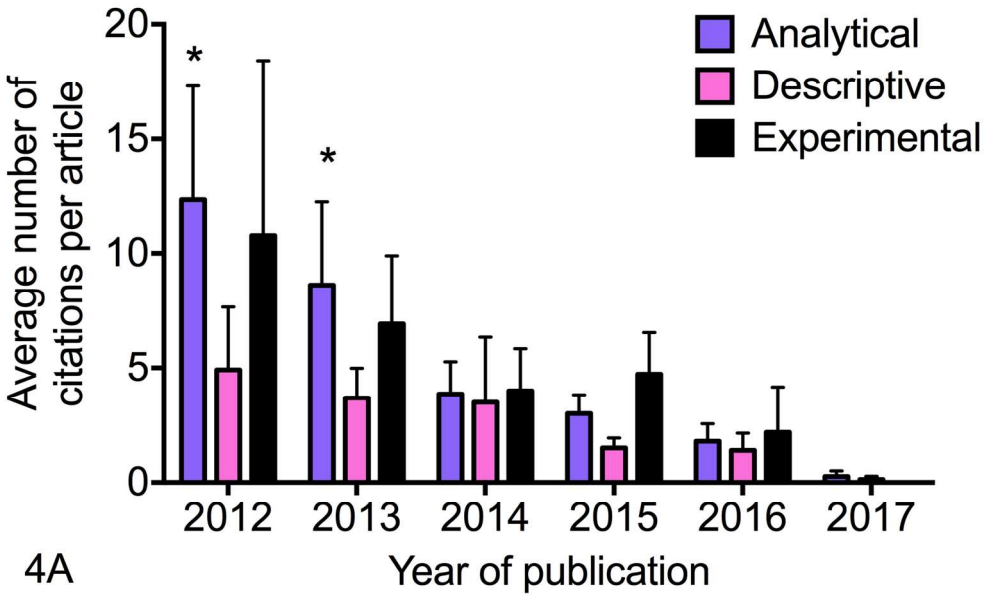


155x107mm (300 x 300 DPI)



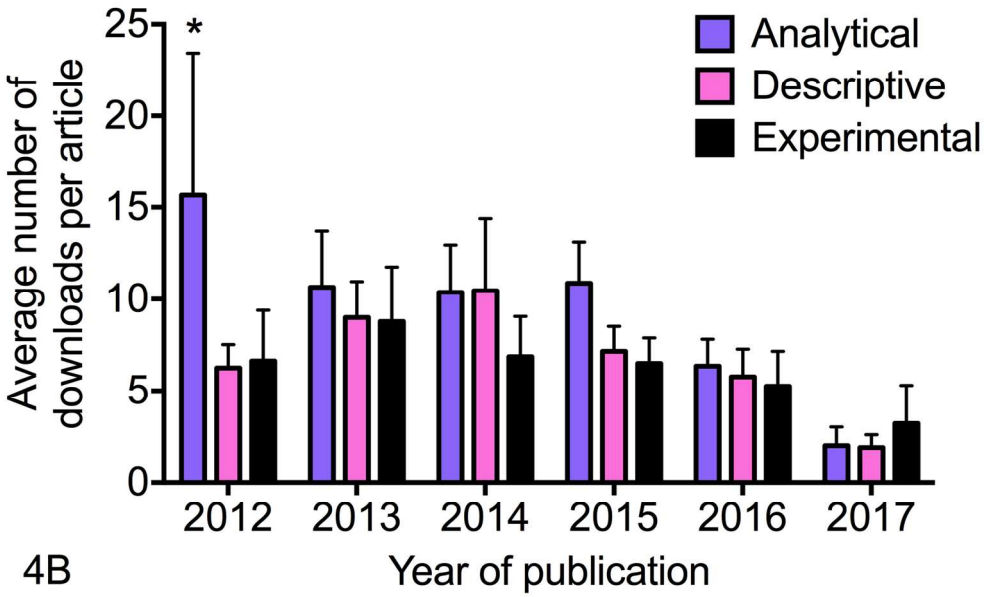
3

254x190mm (240 x 240 DPI)



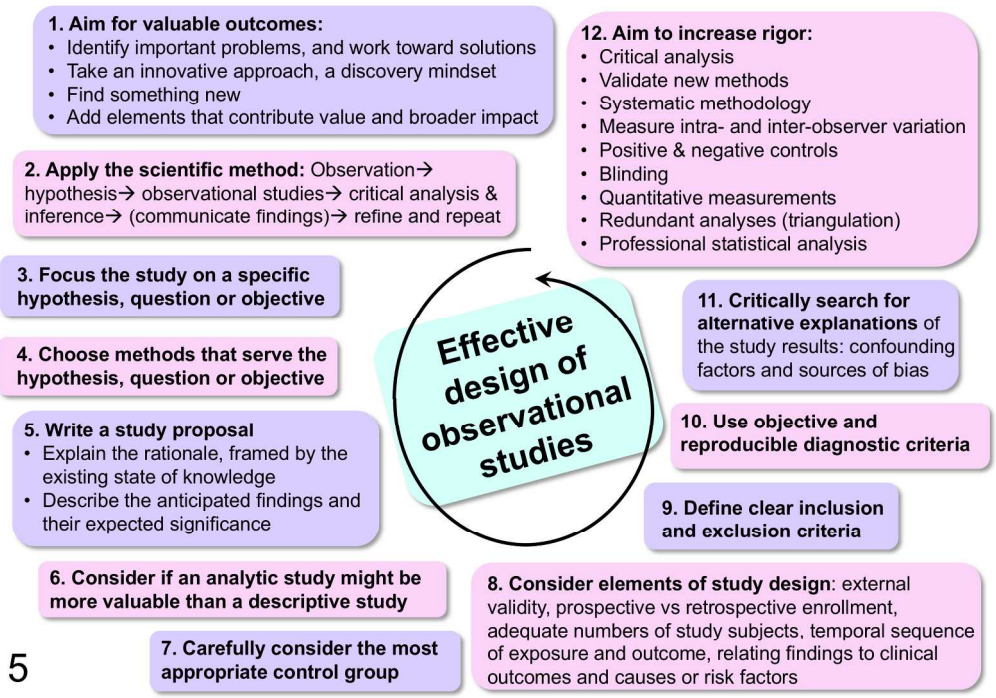
4A

157x99mm (300 x 300 DPI)



4B

157x99mm (300 x 300 DPI)



5

254x190mm (240 x 240 DPI)

Table 1. Questions to revisit at each stage of the study

1. Will the study be a useful contribution to new knowledge, and what can be done now to give it additional value?
2. What critiques will peer reviewers make, and what can be done now to mitigate them?
3. Does the plan aim to conclusively address the hypothesis/ question/ objectives of the study, and what can be done now to ensure this occurs?
4. Are the number of study subjects adequate, given the anticipated variability of the data and the magnitude of the difference between exposure groups that is considered to be meaningful?

Table 2. Some analytic observational studies in *Veterinary Pathology*, 2016-2017. Note that disease may represent either the exposure or the outcome, depending on whether the study investigates the causes or consequences of disease.

Article Title	Exposure	Outcome
Wooden breast myodegeneration of pectoralis major muscle over the growth period in broilers[42]	Different age categories	Morphology, severity and distribution of muscle lesions
Changes in Foxp3-positive regulatory T cell number in the intestine of dogs with idiopathic inflammatory bowel disease and intestinal lymphoma[29]	Inflammatory bowel disease vs intestinal lymphoma	Number of $Foxp3^+$ cells; level of interleukin-10 gene expression
Prognostic significance of canine mammary tumor histologic subtypes: an observational cohort study of 229 cases[35]	Morphologic subtypes of mammary carcinoma	Median survival time
Cytologic criteria for mast cell tumor grading in dogs with evaluation of clinical outcome[7]	High-grade vs low-grade mast cell tumor	2-year survival
Localization of bovine papillomavirus nucleic acid in equine sarcoids[20]	Presence/absence of papillomavirus DNA	Sarcoids vs various non-sarcoid skin samples
Valvular and mural endocardiosis in aging zebrafish (*Danio rerio*)[13]	Water systems, diet, genotype, presence of intestinal carcinoid	Presence/absence of endocardiosis
Feline panleukopenia virus is not associated with myocarditis or endomyocardial restrictive cardiomyopathy in cats[31]	Presence/absence of parvoviral DNA	Presence/absence of endomyocardial disease

Table 3. The classic analytic observational study designs.

Design	Example	Potential advantages	Possible limitations
Cross-sectional study	Select 50 biopsy samples of canine liposarcoma (14 well-differentiated, 7 myxoid, 25 pleomorphic, 4 dedifferentiated); compare high vs low expression of various growth factor receptors (the exposure) among histologic subtypes (the outcome).[1]	Can analyze multiple exposures and multiple outcomes Measures prevalence of the exposure and of the outcome Practical if there is a long interval between exposure and outcome	Consequences of the single sampling: • may not determine if the exposure preceded the outcome, which is important for causal inferences • measures prevalence (not incidence), and thus may not distinguish if an exposure affected development of the disease or alternatively affected the survival of affected animals Limited number of subjects in one group, if one of the exposures or outcomes is rare
Case-control study	Select lung samples from 28 dogs with pulmonary fibrosis and 18 normal controls. Compare the frequency of herpesvirus infection (the exposure)[37] in dogs with and without	Useful if the outcome is rare (e.g. studying causes or risk factors of rare diseases) Practical if there is a long interval between exposure and outcome Can analyze multiple exposures or putative causes	Susceptible to bias if: • the method of selecting subjects for the different study groups affects the likelihood of exposure to the putative cause • the method for measuring the exposure differs between study groups • determination of the exposure is done with knowledge of the outcome or is based on recall • study groups differ in ways other than the outcome that defines the study

	pulmonary fibrosis (the outcome).	Measures prevalence of the exposure in the different study groups	May not determine if the exposure preceded the outcome, which is important for causal inferences Cannot measure incidence or prevalence of the outcome
Cohort study	Select 30 dogs with the rare diagnosis of marginal zone lymphoma and 30 dogs with the frequent diagnosis of diffuse large B cell lymphoma (the exposure); compare with respect to survival time (the outcome). In a beef feedlot, select 300 calves transported long distances and 300 calves transported short distances (the exposure); compare with respect to the later	Useful if the exposure is rare (e.g. studying consequences of rare diseases) Measures incidence (eg. development of new cases) rather than prevalence (eg. presence of existing cases) Establishes the temporal relationship of the exposure and the outcome Can analyze multiple outcomes	Susceptible to bias if: • the method of selecting subjects for the different study groups affects the likelihood of developing the outcome • one study group is more likely to be lost to follow-up • the method for measuring the outcome differs between study groups • study groups differ in ways other than the exposure that defines the study A low dose or short duration of exposure may not induce the outcome May be difficult and expensive to enroll animals free of the outcome and analyze or sample them over time Cannot measure incidence or prevalence of the exposure

	development of respiratory disease (the outcome).		

Table 4. Hill's criteria for evaluating the strength of evidence that an observed association is causal [23]

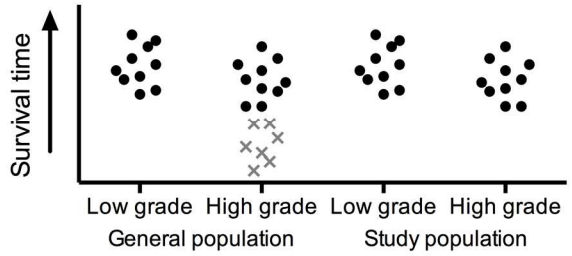
Criterion	Explanation
Strength of the association	Animals exposed to the risk factor are more likely to develop the disease outcome than those not exposed, or the putative cause was significantly more frequent in cases vs controls. However, a statistically weak relationship may nonetheless be causal, as is the case with weak predisposing factors or genetic causes with incomplete penetrance. Thus, investigators should not only report the likelihood that the observed association is due to chance (i.e. the *P* value), but more importantly the precision of the estimate (i.e. 95% confidence intervals) and the strength of the association (i.e. relative risk or odds ratio).
Consistency	The association between exposure and outcome is consistently found in studies of different populations.
Specificity	Although not required, causality is supported by the observation that an exposure induces a specific outcome, such as a unique histologic lesion.
Temporality	The putative cause precedes development of the outcome; for example, the infection precedes disease, or development of the disease precedes changes in serum levels of a biomarker. Temporal relationships are discussed in more detail in the text.
Biologic gradient or dose-response	Progressively higher or more prolonged exposure to the putative cause is associated with a greater likelihood of disease or more severe disease. Such relationships need not be linear or monotonic.
Plausibility	Current understanding of pathogenesis allows for a sequence of events linking the causal exposure and the resulting outcome. In dismissing the absolute requirement for this criterion, Hill quoted Sherlock Holmes: "'when you have eliminated the impossible, whatever remains, however improbable, must be the truth".[23]
Coherence	The causal relationship "should not seriously conflict with the generally known facts of the natural history and biology of the disease".[23] Although superficially similar to plausibility, coherence relates to our broader understanding of biology and related fields.
Experiment	Evidence supported by controlled manipulation of the independent variable provides strong additional support for causation.

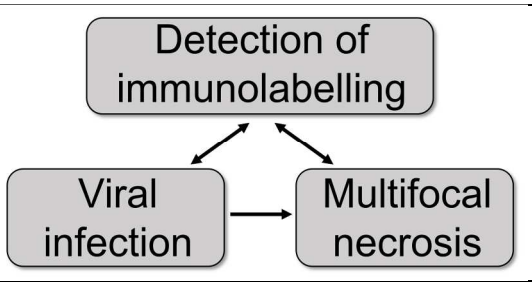
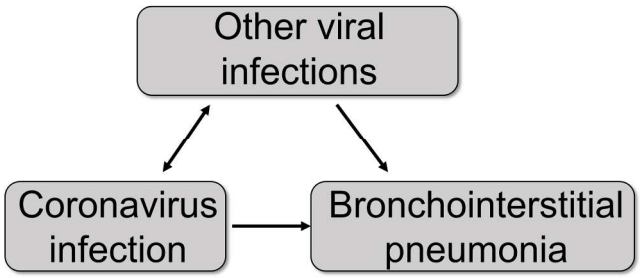
Analogy	There is supporting evidence that a comparable exposure causes a similar outcome.

Table 5. Factors to consider when evaluating the suitability of control or comparison groups. Comparison groups should be similar, except for the factor of interest. Other factors that differ between groups may cause bias or confounding, if their frequency or distribution are not similar between study groups and they are not accounted for by analysis.

- Factors influencing eligibility for entry to the study
- Demographics: age, sex, breed, body weight, geographic origin, diet
- Animal use: types of animal production systems, use for companionship vs performance
- Lifestyle: diet and nutritional status, exercise and fitness level, herd size, type of housing, environmental exposures
- Health: primary vs referral clinics, quality of veterinary care, prevalence of infectious agents, stress, administration of antibiotics or other drugs, frequency of concurrent diseases, details of clinical case management, likelihood of survival
- Details of how samples were acquired, stored, prepared, and analyzed
- Factors that might influence subjective evaluations: blinding of the investigator, different operators, different day of analysis
- Accuracy of case records or recollections of past clinical details
- Other factors affecting the likelihood of errors in diagnosis or histologic scoring, erroneous measurements, or the frequency of false-negative or false-positive tests
- Samples missing from the archive, or loss of animals to follow-up in survival studies. These are problematic if the lost samples differ from the included samples with respect to the exposure and the outcome.

Table 6. Reasons for spurious associations in pathology-based observational studies.

Type of error	Example
Selection bias. Systematic errors in how animals are recruited to the study and assigned to the different study groups.	Biopsy samples of a tumor were retrieved from a laboratory archive. Unexpectedly, the survival time after diagnosis was similar in cases with high-grade vs low-grade tumors. However, some animals with high-grade tumors may have had clinical findings (e.g., ulceration of invasive tumors or detection of metastases) that prompted euthanasia without being biopsied (shown as 'X' in the graph below), whereas the clinical assessment did not similarly influence cases with low-grade tumors. Thus, the clinical findings imposed a selection bias such that only the least clinically aggressive high-grade tumors had biopsies available for study. This falsely reduced the apparent difference in survival between animals with high-grade and low-grade tumors. 
Non-differential information bias. Errors that result in incorrect classification of exposure or outcome, but have the same impact in both study groups (e.g. the same effect in cases and controls).	Animals with high-grade carcinomas have shorter survival than those with low-grade carcinoma. However, because of imprecise grading criteria, or lack of suitable training or experience of the operator, there were errors in grading that increased the variability of the data. As a result, the study failed to identify a statistically significant difference in survival between groups.
Differential information bias. Errors that result in incorrect classification of exposure or outcome, and have differing impacts in different study groups.	In evaluating immunohistochemistry for a viral antigen, brown staining within foci of necrosis was more likely to be noticed (or more likely to be interpreted as positive), whereas it was more likely to be overlooked or interpreted as background staining within areas of normal liver. Thus, immunolabelling falsely appeared more frequent in cases with multifocal hepatic necrosis compared to normal liver.

	
• **Confounding.** A factor that is associated with the exposure and causally influences outcome, but is not part of the causal sequence linking exposure to outcome. A confounding factor is thus a second independent exposure that causes or causally influences the outcome, and is associated with the exposure being studied.	It remains controversial whether bovine coronavirus is a significant cause of bronchointerstitial pneumonia. Beef calves tend to be infected with other viruses in addition to coronavirus if they have been co-mingled with calves from other sources, and these other viruses are known to cause bronchointerstitial pneumonia. Thus, other viral infections confound the association of bovine coronavirus and bronchointerstitial pneumonia. 
Chance. Random differences between study groups	Dogs in a study of lymph node metastasis happen by chance to be younger than those without metastases. Thus, dogs with lymph node metastasis seem to have longer survival, but only because they happen to be younger than those without nodal metastasis.