Equine Epidemiology: Counting for Something in Equine Practice

Noah D. Cohen, VMD, MPH, PhD, Diplomate ACVIM

As a veterinary student, I had the impression that epidemiology was the purview of those interested in either veterinary public health or the health of herds of livestock. A short period of time in private equine practice revealed to me that equine practitioners also deal routinely with groups of horses congregated at barns, farms, racetracks, and show-grounds. During graduate school, I was exposed to the concept of applying principles of epidemiology not only to populations but also to treatment of individual patients through the then-emerging discipline referred to as clinical epidemiology. Clinical epidemiologists subsequently formulated principles and methods for practicing evidence-based medicine as an approach for improving patient care, relying on epidemiological measures and methods for interpreting results of clinical studies. My career has involved the application of epidemiological methods and principles to equine health disorders. The purpose of this presentation will be to familiarize practitioners with epidemiologic methods and how they can be applied in practice, with emphasis on interpreting results of clinical studies. The importance of private practice–based epidemiological studies also will be highlighted. Author’s address: Large Animal Clinical Sciences, Texas A&M University, 4475 TAMU, College Station, TX 77843. e-mail: ncohen@tamu.edu. © 2011 AAEP.

1. Introduction
The field of epidemiology is broad, and the discipline has many applications and diverse classifications. It is thus no more feasible to comprehensively and meaningfully review the field of epidemiology in a few hours than it would be to discuss theriogenology, surgery, or internal medicine in the same time frame. This presentation will focus on the motivations, basic methods, and applications of epidemiology to equine medicine.

Epidemiology is a science that largely involves counting events or characteristics of individual members in populations. The ultimate purpose for quantifying events or attributes is to improve our understanding of determinants of disease and health. In this sense, then, epidemiology is a matter of counting for something. My hope is that this presentation will foster greater interest in epidemiology and enhance awareness of the relevance of epidemiological methods to clinical practice.

2. Motivations
There are two compelling arguments in favor of the epidemiologic approach for understanding disease and disease outcomes. First, they are advantageous from the standpoint of welfare because only spontaneous disease is studied: No disease needs to be experimentally created or induced in horses or other animals. Second, evidence-based medicine places a premium on patient-based studies. The term “patient” here is used broadly to include healthy individuals receiving vaccines or other preventive strategies in a routine setting (as opposed to
recruited volunteers or privately owned horses such as teaching herds, etc.). The rationale for this prioritization of populations of patients or others receiving healthcare and prevention is 3-fold: (1) experiments in other species often are not directly relevant to the target species; (2) experimental models of disease rarely (if ever) are accurate mimics of spontaneous disease; and (3) experimental subjects and conditions do not accurately reflect health care in the “natural” setting.

Acceptance of the importance of epidemiology as a clinical science for equine medicine has met resistance, primarily attributable to two obstacles. First, there has been a long-standing tradition of empiricism in equine medical education and practice, with an emphasis on authority, expert opinion, and deductive reasoning. Second, equine clinical science has traditionally been primarily experimental in approach. Both empiricism and experimentation, however, have limitations. Regarding empiricism, the opinions and perceptions of experts can be wrong, and logical deduction can be misleading. An oft-cited example is the recommendation by Benjamin Spock that babies not be allowed to lie on their backs (supine position) because they might aspirate and suffocate in this position and might flatten their heads. As a result, the prone position (lying on their stomachs) was promoted for infants. The prone position has been identified as a strong risk factor for sudden infant death syndrome. A relevant example from equine medicine pertains to folate supplementation in horses receiving anti-folate treatments for equine protozoal myeloencephalitis. Although logical deduction suggests that folate supplementation would be rational when anti-folates are administered, evidence emerged that this practice was harmful to the foals of mares that received such supplementation.

The principal scientific advantage of the experimental approach is the ability of investigators to control conditions such that the variables being evaluated can be reduced to a minimum number. Experimental science is also convenient within the temporal constraints of clinical training programs in which academic clinicians are spawned because primary data collection from patients can be time-consuming (particularly for prospective study designs), and it may be difficult to accrue sufficient cases within the training period.

Experimental disease, however, is generally a poor mimic of spontaneous disease and the complex interaction of the diversity of causal factors encountered in the “real world” that generate spontaneous disease. The reductionist approach of experimental conditions enables the isolation of effects to a limited number of possible causes; however, the assumption that a complex system (such as a disease) is equal to the sum of its components and that results of constituents can be extrapolated to the complex system may be simplistic and unrealistic. Environmental conditions in a laboratory or at a research farm will differ from those of nonexperimental settings and will not reflect the diversity of environmental exposures or conditions that occur naturally. Moreover, when we extrapolate results from non-equine laboratory animals to horses, we are crossing relatively large phylogenetic spans. Studying disease in more out-bred species with greater heterosis is important. As mentioned, there are ethical and welfare concerns when disease is created in animals that are obviated by studying spontaneous disease.

Increasing awareness of the limitations of empiricism and experimental models of disease has provided impetus for epidemiologic studies. The study of spontaneous cases of disease using epidemiologic studies thus has the advantage of providing evidence that is of a relatively higher clinical relevance while negating the need to induce disease in animals. As previously mentioned, the principles of evidence-based medicine place a priority on studies of spontaneous disease and patient populations. Clearly, motivation exists for clinicians to consider epidemiology as a basic science for clinical investigation and education.

3. Methods

Epidemiological Measures

Epidemiology is a science of observation of the distribution of disease and determinants of disease in populations. The fundamental observations of epidemiologists are measures of disease frequency, primarily incidence rates and prevalences. Incidence refers to the number of new cases occurring within a defined period of time in a population of individuals at risk for developing the disease. Incidence measures events (occurrences), is a true rate, and is the fundamental measure of risk. Prevalence of a disease refers to the proportion of animals in a group that have the disease. Prevalence is a proportion and not a rate, and consequently it is not a measure of risk. Whereas incidence focuses on events, the focus of prevalence is on the disease status of individuals.

The principal objective of most epidemiological studies is to detect and estimate measures of association, principally relative risks (RRs) and odds ratios (ORs). Incidence rates, prevalences, and other measures of frequency are the building blocks by which measures of association derived from epidemiological studies are constructed. The purpose of measuring associations is to attempt to identify causal determinants of disease. When interpreting epidemiological studies, it is important to recall that not all associations are causal; it can be challenging
to differentiate associations that are causal from those that are not. There are a number of sources of difficulty. Some of this challenge can be attributed to the fact that epidemiology involves clinical scientists making observations without being able to control the occurrence of either disease or the determinants of disease (exposures, treatments, etc.). In laboratory experiments, investigators attempt to create circumstances in which the individual units of analysis (principal and their controls) differ only by a single factor. Even in the laboratory setting, such control is not always possible to achieve. In the clinical setting, where we wish to study our patients, only minimal control over extraneous factors is possible. The core of epidemiologic methodology is in the design of studies to allow for clinically meaningful inferences to be derived about exposures and outcomes from observations of patients and other populations. In this document, the term exposure will be used to define some characteristic or attribute of an individual. An exposure may be an invariant characteristic of an individual horse, such as its sex or breed, or a variable characteristic such as diet or treatment. The term outcome and disease will be used interchangeably; however, other health outcomes also may be studied using epidemiological methods. Epidemiological studies involve studying exposures, outcomes, and their association(s).

Epidemiological Study Design

General Considerations and Principles

Epidemiological study design is rooted in the design of other scientific experiments: Observations are made, and hypotheses are tested on the basis of refutationist philosophy. Refutationist philosophy is based on comparing observations to what might be expected under a specified hypothesis. Scientific inquiry thus entails testing a hypothesis to refute it rather than to support it. The hypothesis tested under refutationist philosophy is the null hypothesis. For example, we do not test the hypothesis that two treatments result in a different clinical response; rather, we test the hypothesis that there is no difference between the clinical response to the two treatments. Two important consequences of the refutationist paradigm merit consideration. First, a single valid study is sufficient to refute or reject a null hypothesis. Second, failure to refute a hypothesis is evidence in favor of that hypothesis, but it does not prove that hypothesis is true. Indeed, it is impossible to definitively prove any hypothesis, according to refutationist philosophy. As a consequence, the hypothesis-testing paradigm allows us to generate our best estimation (in the form of a hypothesis) of what our observations explain, and ideas (hypotheses) continuously emerge to further refine our understanding of the world.

Ideally, epidemiological studies are designed to test hypotheses articulated a priori. It is not uncommon, however, to conduct studies that involve hypotheses tested post hoc (i.e., after the study has been designed and the data have been collected and analyzed). In general, hypotheses identified after statistical testing (e.g., identification of several factors significantly associated with increased odds of horses developing colic among a long list of factors being compared between cases of colic and controls) are accorded less credence than those that are identified as significant following hypotheses declared a priori (e.g., feeding >5 kg of concentrate daily is associated with significantly increased odds of colic). This is because refutationist philosophy is based on construction of a hypothesis that would predict observations, and, consequently, observations cannot precede the hypothesis (much as a cause cannot occur after effects attributed to that cause). Although hypothesis-testing studies are considered to provide a higher quality of evidence, hypothesis-generating studies (such as risk factor studies) are nevertheless important for stimulating inquiry and advancing science. Indeed, eminent epidemiologists have questioned the relegation of risk factor studies to lesser status simply because a hypothesis was not stated in advance of data collection and analysis.

The objective of an epidemiological study is to precisely estimate a valid measure of association (e.g., an OR) or some other epidemiological measure (e.g., an incidence rate). Precision and validity are separate concepts, distinguished by differing types of error. Precision results from the system used for measurement, and lack of precision is considered to occur randomly; put another way, the extent of random error in a measurement tool (e.g., a device for quantifying blood pressure) dictates precision of an estimated parameter. The typical (i.e., mean) value for random error is 0, with dispersion about the mean value. Validity is not a function of random error; rather, validity results from an absence of systematic error. Unlike random error, systematic error has a typical (i.e., mean) value that is non-zero and is not attributable to imprecision of measurement. Rather, systematic errors result from flaws in the design, conduct, or analysis of studies.

Measurement error can occur in epidemiological studies as a result of use of a poor instrument. The instrument may be a mechanical or electronic device but also may be a questionnaire or form. Poorly worded questions may lead to imprecise answers because of ambiguity and misinterpretation. A large random error will negatively affect the statistical power of a study and result in wide confidence intervals that indicate greater uncertainty in the findings (because of imprecision); note that imprecision (i.e., wide confidence intervals) also may result from a small sample size.
Systematic error in epidemiological studies occurs in one of three forms: (1) selection bias; (2) information bias; and (3) confounding bias. Selection bias occurs as a result of the subjects of the study being selected in a manner such that they are systematically different from the reference/target population with respect to their response to the exposure or exposure history. Ideally, horses included in epidemiologic studies should be selected randomly from the reference or target population. Nonrandom sampling is a form of selection bias that results in the relationship between exposure (such as a treatment) and outcome (disease) being distorted among those who participate in the study relative to the population to which we wish to generalize results (i.e., sampling bias influences external validity of a study). For example, if we select for inclusion in a study of colic only those horses with more severe forms of colic (e.g., those referred to a surgical center for further evaluation), our findings of an association between an exposure (lactate concentrations) and an outcome (need for surgical treatment) may be skewed. Another type of selection bias is indication bias, wherein cause and effect become confused because exposure is dependent on treatment. For example, the finding that administration of *Rhodococcus equi* hyperimmune plasma is significantly associated with the odds of a farm having foals affected by *R. equi* pneumonia does not indicate that transfusion of hyperimmune plasma causes the increased incidence of disease; rather, it is an indication bias because affected farms are more likely to transfuse foals to try to prevent the disease.

Information bias refers to a systematic error in the collection of information about study participants. A common and important cause of information bias is error in classification of either exposure or outcome. Misclassification error can be either differential or nondifferential. It is important to distinguish whether misclassification is differential or nondifferential. Differential misclassification refers to the circumstance in which misclassification is more likely to occur in one group of study participants than in another. For example, consider a study designed to determine whether foaling at pasture is less likely to result in foals developing *R. equi* pneumonia. Consider further that foals born in stalls and maintained in barns may be followed more closely for signs of developing pneumonia, such that *R. equi* pneumonia is more likely to be detected in such foals. In this scenario foals born in stalls may be more likely to be have pneumonia detected, and thus a spurious association between foaling at pasture and reduced risk of *R. equi* pneumonia might be observed (N.B., the author does not believe that the observed association between reduced risk of *R. equi* pneumonia and foaling at pasture is attributable to misclassification bias). Differential misclassification of exposure also can occur, as exemplified by so-called recall bias, wherein cases are more likely to recall an exposure than controls (or vice versa). For example, if we are conducting a study of equine protozoal myeloencephalitis (EPM), owners of horses with EPM may be more likely than owners of control horses to recall a historic event (such as seeing wildlife in or near the barn, the horse having had a stressful event during the preceding 60 days, etc.) than healthy horses examined for routine procedures selected as a control group. A spurious association may result from the differential reporting of information: Note that the association may be spurious in qualitative terms (an association is observed that is not real) or quantitative terms (the magnitude of a true association is overestimated or underestimated).

Nondifferential misclassification of either exposure (e.g., a treatment) or outcome (e.g., disease) occurs when the proportion of individuals misclassified for disease does not vary by exposure status or when the proportion of exposed individuals does not vary, depending on the outcome status. For example, if some foals with streptococcal pneumonia are misclassified as having rhodococcal pneumonia in a clinical trial assessing azithromycin to prevent *R. equi* pneumonia, the misclassification could be expected to be nondifferential (i.e., to occur in similar proportions in the azithromycin-treated and placebo-treated groups). In general—though not always—nondifferential misclassification results in a bias toward the null (i.e., toward no association between exposure and outcome). Thus, the impact of disease misclassification in such a study probably would result in underestimation of the effectiveness of the intervention. "Confounding" is a bias that results in distortion of a measure of association between two factors (for example, a disease or health disorder [racing injury] and an exposure [racetrack surface]). A commonly cited example from human medicine is the establishment of whether coffee drinking is a risk factor for lung cancer. Smoking is causally associated with lung cancer, and smokers consume more coffee than nonsmokers. Thus, failing to account for smoking status may result in a confounding bias of the association between coffee drinking and lung cancer. For an equine example, suppose we wish to conduct a study of whether a breed of horse is associated with the risk of colic. Suppose further that we know that consuming increased amounts (i.e., kilograms fed per day) of concentrates results in increased risk of colic. We might observe in a study that Thoroughbreds have a higher incidence of colic than other breeds of horses. Because many Thoroughbreds are involved in intensive exercise training, they are fed and consume larger amounts of concentrate than horses of many other breeds that are less active. Thus, the association between Thoroughbred breed and colic might be confounded (distorted) by failing to account for another factor (amount of concentrate fed daily) that is associated both with the "exposure" (i.e., breed) and the outcome (i.e., colic).
Study Designs

The design of standard epidemiologic studies has been reviewed elsewhere\(^1,10–12,18\) and will be reviewed only briefly. Epidemiological study designs can be characterized by whether one first determines exposure status and then follows the group forward in time to determine outcome such as disease occurrence (i.e., cohort or clinical trial designs) or whether one first determines disease status and exposures are then determined from historic information (i.e., case-control, case series, and case-only designs); alternatively, both exposure and disease may be determined simultaneously (cross-sectional study design). Furthermore, the design of epidemiological studies may be either experimental or observational (Table 1 and Figure 1). In experimental epidemiological studies, investigators control and determine exposure assignments to test a hypothesis, whereas in observational studies, exposure is determined by study participants (self-selection) without either consideration of a research objective or influence of the study investigators.

### Experimental Epidemiological Studies

In experimental studies, researchers randomly assign study participants to two or more groups receiving a treatment or other intervention. Study design is characterized by how horses are assigned to the study (random assignment to group, nonrandom assignment to group, or no group assignment [i.e., case series or case reports]; item 1). Studies in which horses are assigned randomly to two or more groups receiving a treatment or other intervention are referred to as randomized, controlled trials (RCTs). Among horses whose assignment is not random, the study design is then determined by whether horses are sampled (item 2) first on the basis of exposure and then followed forward in time (cohort study) or on the basis of disease and then queried to determine historic exposure (case-control study), or whether exposure and disease are determined simultaneously (cross-sectional study). In general, the quality of evidence is relatively reduced among designs more to the right (dotted line at the bottom of the figure).

---

**Table 1. Summary of Epidemiological Study Designs**

<table>
<thead>
<tr>
<th>Experimental designs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Randomized, controlled trial</td>
</tr>
<tr>
<td>Population-based</td>
</tr>
<tr>
<td>N of 1 randomized, controlled trial</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nonexperimental (observational) study designs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohort study</td>
</tr>
<tr>
<td>Prospective</td>
</tr>
<tr>
<td>Retrospective/Nonconcurrent</td>
</tr>
<tr>
<td>Self-controlled case series (case-only design)</td>
</tr>
<tr>
<td>Case-control study</td>
</tr>
<tr>
<td>Case-control</td>
</tr>
<tr>
<td>Case-cohort</td>
</tr>
<tr>
<td>Nested case-control</td>
</tr>
</tbody>
</table>

| Case-crossover (case-only design)          |
| Cross-sectional study                      |
| Reports of case series                     |
| Reports of individual cases                |

---

**Fig. 1. Schematic of epidemiological study designs.** Study design is characterized by how horses are assigned to the study (random assignment to group, nonrandom assignment to group, or no group assignment [i.e., case series or case reports]; item 1). Studies in which horses are assigned randomly to two or more groups receiving a treatment or other intervention are referred to as randomized, controlled trials (RCTs). Among horses whose assignment is not random, the study design is then determined by whether horses are sampled (item 2) first on the basis of exposure and then followed forward in time (cohort study) or on the basis of disease and then queried to determine historic exposure (case-control study), or whether exposure and disease are determined simultaneously (cross-sectional study). In general, the quality of evidence is relatively reduced among designs more to the right (dotted line at the bottom of the figure).
Nonexperimental Epidemiological Studies

Nonexperimental epidemiological studies are commonly referred to as observational epidemiological studies. In these nonexperimental designs, exposure is self-determined by the individual horses being studied (or by those caring for the study horses, e.g., diet). The observational epidemiological study designs encountered most commonly in equine research reports are cohort, case-control, cross-sectional, and case series designs.

Cohort studies first define the exposure of interest (a common characteristic of a group of horses, such as activity or treatment) and then follow the experiences of the group of horses that share this characteristic over time and measure the incidence of disease in the group. The group sharing the exposure of interest is termed a cohort, from whence the name of the study design is derived. Commonly, two or more cohorts are followed and the ratio of the incidences in the two groups is determined; this ratio of risks is known as the RR and is a fundamental measure of epidemiological association. Cohort designs can be prospective or retrospective. In a prospective cohort study, the cohorts are determined in the present and followed forward in time. For example, we might determine whether or not mares are treated for EPM during pregnancy in the coming spring and follow the subsequent outcomes of those pregnancies to determine whether the incidence of fetal loss or adverse health effects of the foals is associated with treatment. In a retrospective cohort study, exposure is determined at some point in the past and the individuals are followed forward in time from that point. For example, we might obtain last year’s Thoroughbred racing results and determine whether horses racing on turf (one cohort) had a higher incidence of injury during racing than horses racing on dirt.

Cohort studies are well suited to identifying risk factors for disease, studying the outcome of an intervention, and examining the natural history of disease. Prospective cohort studies are generally superior to retrospective studies because exposure data are historic for the latter. Information bias is much less likely in cohort studies than in case-control studies. Some degree of selection bias is to be expected in any cohort study because cohorts are chosen by investigators. Consequently, it is important to scrutinize the criteria for including and excluding cases to understand who the exposed and unexposed groups represent. Cohort studies should be designed to account for known risk factors for the outcome of interest to avoid potential confounding effects. When variables known to be associated with the disease of interest are not accounted for, study results must be interpreted with caution because the results may be spurious due to confounding (please see the examples described above for confounding). Generally, the extent of the impact of confounders that are

...
unaccounted for will be proportional to the extent of the association of the confounding variable with the outcome of interest: Factors strongly associated with the outcome will be strong confounders, and variables weakly associated with the outcome will be weak confounders. Clearly, investigators cannot always collect data on all variables previously associated with the outcome of interest (i.e., potential confounders). Failure to account for known confounders diminishes the strength of but does not vitiate the value of a study. Readers of a study in which known potential confounders have not been accounted for should recognize that it will be necessary for the results of that study to be substantiated by other studies, because the magnitude and statistical significance of observed associations in the report of interest may be changed after accounting for confounding. All epidemiological studies should be assumed to have some degree of confounding by unmeasured variables.

Advantages of cohort studies include: (1) one can quantify absolute risk of an outcome such as a disease because incidences are measured; (2) one can examine multiple outcomes for a given exposure; and (3) the purported cause (the exposure of interest) can be clearly demonstrated to precede the outcome (effect) of interest. Cohort studies have the disadvantages of being relatively expensive and administratively intensive (particularly when they involve multiple centers), they are not well suited for rare outcomes (e.g., catastrophic racing injuries), they can require long periods of time to conduct, and there can be substantial losses to follow-up. Retrospective cohort studies can reduce the time required to conduct a cohort study because much if not all of the data will have already been collected; however, problems with information bias and confounding are increased.

In case-control studies, investigators first establish the disease status of the study participants (either cases of the disease or horses selected as controls) and then determine the historic exposure of case and control horses. The objective of this study design is to identify factors associated with being a case by comparing the frequency of exposures of interest in cases with those in controls. Like cohort studies, case-control studies may be prospective or retrospective. In the prospective design, cases and controls are identified beginning in the present time and continue to be collected in the future; this design is also referred to as an incident case-control design. In retrospective case-control designs, cases and controls are identified from records that have already been identified and recorded. Case-control studies are most efficient for studying rare diseases and can be useful for hypothesis-generating studies for prognostic factors (e.g., risk factors for survival among foals with sepsis).

Advantages of the case-control design include: (1) they are relatively inexpensive to conduct; (2) they may be completed in a relatively brief period of time when a retrospective design is used; and (3) one can explore multiple exposures that are associated with a given disease, provided that the data have been collected and are available to the investigators. Disadvantages of the case-control design are that (1) historic information is used, and the quality of these data can be limited and data may be missing for some individuals; (2) the temporal relationship between exposure and disease is less clear because of the historic nature of the data collection; (3) data were not expressly collected for the purpose of the study; (4) identification of the appropriate control group(s) is extremely difficult; and (5) there is no true measure of risk derived from the study.

Risk cannot be measured in a case-control study because the frequency of disease is determined by the study design (investigators determine the number of cases to be studied), rather than disease occurring spontaneously in a defined group at risk observed for a specified period of time. Instead of an RR, the measure of association in case-control studies is the OR. This is the ratio of the odds of exposure among cases to those among controls; conveniently, this is mathematically equivalent to the ratio of the odds of disease among the exposed to the odds of disease among the nonexposed. When the disease being studied is rare, and when cases and controls are representative of cases and non-cases in the reference population, the OR will approximate the RR. However, a true measure of risk cannot be established.

The collection of historic exposure data renders case-control studies subject to a number of biases. Health-related information generally is not collected from patients with future scientific studies in mind, so retrospective data may be either missing or of relatively poor quality. Selection bias may result because of limitations in the representativeness (to the reference population) of cases and controls; this is especially problematic for the control group. Finally, confounding can be a problem in case-control studies because data regarding important confounding variables may be either missing or of poor quality. As for cohort studies, case-control studies should be designed to account for known risk factors to avoid potential confounding effects.

There are some modifications of case-control designs that are sometimes referred to as hybrid designs. These include case-cohort, nested case-control, and case-only designs (Table 1 and Figure 1). In case-cohort and nested case-control designs, the cases are identified from a well-defined cohort; controls are either identified by random selection at the beginning of the study period (case-cohort design) or at the time each case is identified (nested case-control design). Case-only designs included case-crossover studies and self-controlled case series. These case-only designs examine exposure during a period preceding an outcome (e.g., myocardial infarction, racing injury, etc.) and compare
these with exposures during periods before or (when possible and appropriate) after the outcome.

In cross-sectional studies, both exposure and outcome (disease) status are determined at the same time; often, exposure and outcome pertain to the same time point. Cross-sectional studies can provide estimates of prevalence of exposure and outcomes. These studies may be conducted relatively easily and expeditiously and work well for studying the association of outcomes with exposures that are invariant (e.g., sex or breed). The principal disadvantage of a cross-sectional study is that the relationship between cause and effect is unclear as a result of the simultaneous determination of exposure and outcome status.

Case-series are descriptive epidemiological studies. They are generally stronger sources of clinical evidence than individual case reports because the series of cases reveals more information about the variation among individual patients with regard to clinical signs, diagnostic test results, clinical course, and prognosis. Case series bring to attention new diseases or manifestations of diseases and novel approaches to therapy. They can be helpful for describing the clinical course of a disease and may be the only data available for rare disorders. The principal and critical limitations of case series are that they can be biased in terms of selection of cases and by the subjective interpretations of findings, and they lack a control group.

The objectives of epidemiologic studies are to accurately describe the distribution of disease or determinants of disease and to precisely and validly estimate the association of exposures with disease. Regarding associations of exposures with disease, the principal objective is to identify associations that are causal or suggest causal mechanisms. Various study designs may be used to accomplish these objectives, and each design has its strengths and limitations. All epidemiological studies have potential for confounding and other sources of bias. That being known, it is important not to develop a nihilistic approach to epidemiological studies: Recognizing the limitations and uncertainty in results of studies should not cause one to dismiss the meaningfulness of reported observations. Although conducting and interpreting results of patient-based or other population-based studies present challenges because of the lack of randomization and control of other extraneous factors, results of such studies are generally more relevant to our patients than those derived from carefully controlled experiments of induced disease and regulated exposures.

4. Applications
Thus far, we have largely considered how epidemiologists quantify events and associations. It is important to articulate why quantifying events and associations is relevant to clinical practice, that we are in fact counting for some clinical purpose. Describing the Frequency of Disease

Describing the frequency of occurrence disease (morbidity) and death (mortality) is valuable for many reasons. It allows one to understand the magnitude of the burden of specific diseases, to monitor temporal trends in diseases, and to identify individuals or groups at increased risk. Such information is critical for planning for control and prevention of disease and for identifying emerging problems. Indeed, incidence and prevalence of disease represent the common language of epidemiologists. Obtaining meaningful incidence and prevalence data in equine populations is challenging. There are problems with defining the occurrence of disease that is needed for the numerator of these measures as well as with defining the denominator. Validity of diagnosis can influence numerator results, including whether the disease was diagnosed by a veterinarian or a horse owner, variability in the criteria for diagnosis, and differences in how cases are identified (e.g., whether the identification of cases is active or passive) or defined (e.g., there are no pathognomonic findings for mare reproductive loss syndrome). The denominator in rates and proportions used to describe disease frequency must be well defined. In some instances, this is easily established (e.g., all racehorses at a given racetrack during a specified period of time). In other instances, however, we only have data on a sample of horses. When that sample of horses is selected systematically (such as selecting a probability sample of horses at risk of developing the disease[s] of interest), inferences to the reference population can be made with some confidence. Often, the group of horses from which the rate is determined is a convenience sample, which limits the extent to which the results can be assumed to reflect some other more general group of horses. All those concerned with equine health are indebted to those federal veterinarians and epidemiologists who designed, conducted, and reported the data from the two National Animal Health Monitoring and Surveillance studies and to the horse owners who reported disease data for those projects. Despite the strengths of these seminal data, there remain important gaps in our knowledge of the incidence and prevalence of many diseases in horse populations.

Outbreak Investigations
Perhaps nothing causes greater concern for equine veterinarians and horse owners than the occurrence of outbreaks at farms and clinics. Recent experiences with abortions equine herpes virus myeloencephalopathy, salmonellosis, and Streptococcus equi subsp. equi infections are notable examples of the impact of outbreaks. In essence, an outbreak investigation is simply a combination of a descriptive epidemiological study and a case-control study. Both for descriptive and analytical epidemiological purposes, the first step in
an outbreak investigation entails defining the disease that is occurring. It is imperative that a case definition is rapidly established to be able to identify cases; this will require clinical and clinico-pathological evaluations and the expertise of the clinical veterinarian(s) on site working in collaboration with pathologists, microbiologists, and scientists from other relevant disciplines (e.g., toxicologists, entomologists, etc.). As with other epidemiologic diagnoses, the case definition may not necessarily correspond to a definitive diagnosis. For example, in studying an outbreak of clostridial diarrhea among neonatal foals at a farm, we may include all foals <7 days of age with diarrhea as cases, even in the absence of microbiological evidence of clostridial infection.

Once the case definition is established, it is possible to monitor the incidence and prevalence of disease during the outbreak, thereby allowing for a clear understanding of the absolute magnitude of the problem. Cases should then be described in terms of time (temporality), place/location, and individual characteristics. The temporality of the disease occurrence can suggest whether the outbreak is from a common vehicle/source or whether the outbreak is being propagated through serial transfer from horse to horse. However, when other levels of temporality are superimposed (e.g., such as the seasonal distribution of foaling in an outbreak of abortion in mares), the interpretation of the temporal distribution of cases becomes more complex. Describing the spatial relationship of cases may help to identify risk factors for disease. For example, a common source of water (such as a pond) might suggest the source of the outbreak, or proximity of cases to one another in both space and time may suggest a common infection.

Finally, cases should be compared with unaffected horses from the same environment in an effort to find exposures that are disproportionately distributed. This is in essence a case-control study whose aim is to identify factors common to cases that may suggest a cause or source of the outbreak.

A principal lesson that need not be relearned with each outbreak is that their control and investigation is best accomplished by the cooperative efforts of a team, including clinicians, epidemiologists, and other scientists. It also is imperative to recognize that actions for control (such as quarantining, establishing isolation protocols where feasible, etc.) may need to be implemented before a diagnosis is established or optimal control methods are known. With the continuing emergence of new infections and the evolution of horses from predominately grazing to predominately intensively managed and housed animals, it should be expected that equine practitioners will continue to be plagued by the difficulties of managing outbreaks at areas where horses are grouped and co-mingled.

Interpreting Clinical Data

Our principal activities as clinical veterinarians are: (1) interpreting results of anamnesis, physical examination, and diagnostic testing; (2) selecting treatments; (3) evaluating responses to treatment; (4) making prognostic decisions; and (5) preventing and controlling disease. Because our expertise is based on incorporating what we or others have learned from treating groups of similar patients, epidemiologic methods and principles are integral to every clinical activity in which we are engaged. Although a review of even one of these areas (e.g., diagnostic testing) is beyond the scope of this lecture, some aspects of how epidemiology influences each of these clinical activities will be considered.

Interpreting Test Results

The interpretation of diagnostic tests (which may be broadened to include not only laboratory testing but also evaluating anamnesis and physical examination findings) entails understanding the principles of sensitivity, specificity, predictive values (both positive and negative), and likelihood ratios (both positive and negative). These topics will not be reviewed here in great depth because they have been previously presented in various formats and venues at prior AAEP Conventions and have been reviewed elsewhere.

Before a test can be considered valuable for clinical use, there must be evidence that the test can differentiate between those with the disease and those without the disease for which the test is directed. The two fundamental properties used to characterize this ability to accurately differentiate patients are sensitivity and specificity. Sensitivity refers to the proportion of horses with disease that have a positive test result. Specificity refers to the proportion of horses without disease that have a negative test result. There are two inter-related concepts regarding test interpretation that are important to practitioners: (1) clinicians are primarily concerned with predictive values; and (2) case-control studies are weak evidence for evaluating the clinical usefulness of a diagnostic test.

Although tests with low sensitivity, specificity, or both generally will perform poorly in a clinical setting because they do not adequately distinguish between those with disease and those without, it does not follow that tests with both high sensitivity and specificity will perform well clinically. This is because as clinicians we are much less concerned with whether those with disease have a positive test (sensitivity) or that those without disease have a negative test (specificity). Rather, we are concerned with whether (i.e., the probability that) a positive test result for an animal indicates that the animal has disease and whether (i.e., the probability that) a negative test result indicates that the animal does not have the disease. These probabilities are known as the positive and negative predictive values, respectively. Whereas the sensitivity and
specificity of a test are fixed values (i.e., do not change with conditions), the predictive value of a test is influenced by the probability that an animal has the disease. This probability of disease at the population level is the prevalence of disease (e.g., the prevalence of *R. equi* pneumonia in the population of foals at a farm). At the level of the individual horse, the probability of the test is the pre-test likelihood that the horse has the disease. For example, a horse with signs of asymmetrical ataxia, evidence of both upper and lower motor neuron involvement, and cranial nerve deficits might be considered to have a high pre-test probability of having EPM.

Because we use predictive values to inform us in the clinical setting, it is possible for a test to have very high sensitivity and specificity and yet still perform poorly for diagnosis. This is because, even with a high sensitivity and specificity, when prevalence of disease (or pre-test probability of disease) is low, most positive test results will be negative. For example, few if any elevated concentrations of adrenocorticotropic hormone (ACTH) among 2-year-old horses in race training will indicate Cushing disease (i.e., most or all positive results will be assumed to be false-positives), whereas our interpretation of an elevated ACTH concentration in an 18-year-old horse with signs of polyuria and polydypsia will suggest to us that the horse has Cushing disease. The probabilistic basis of test results can be unsettling to some: The prevalence or pre-test probability of disease is generally ill-defined, and this introduces subjectivity into the diagnostic process that we assume (or wish to assume) to be completely objective.

The second issue concerning interpreting testing is that although case-control studies may be a useful step in assessing the clinical usefulness of a diagnostic test, they are a weak source of evidence for assessing performance in the clinical setting. It is relatively common in the development of a diagnostic test to assess the performance of the test to differentiate severe forms of disease (for example, necropsy-confirmed cases of EPM) and horses free of signs (and pathological lesions) of the disease. Such case-control studies can be useful for initial evaluation of the sensitivity and specificity of tests, but they are of limited value for assessing how the test will perform in the clinical setting. Useful evaluation of diagnostic tests will require assessing performance across the full spectrum of disease to which we plan to apply the test, including patients with milder as well as more severe forms of the disease, in early as well as late stages of disease, and among both treated and untreated patients. Thus, case-control studies are generally weak sources of evidence for evaluating diagnostic tests. The best sources of evidence for diagnostic tests are prospectively designed cohort studies of consecutive patients undergoing prespecified diagnostic testing criteria against a reference standard that is consistently applied. When patients are not consecu-

Interpreting Treatment Data
Epidemiologic studies are our best source of information regarding the effectiveness of treatments. From the standpoint of individual study designs, the gold standard for evaluating treatment is the RCT. However, cohort studies also may yield results that approximate RCTs in terms of validity and precision. Both RCTs and cohort studies are subject to the previously described biases, and the onus is thus on practitioners to critically appraise reports to assess the quality of the data presented. Given the paucity of RCTs for evaluating equine therapeutics, we rely heavily on observational (nonexperimental) epidemiological studies as sources of evidence for therapeutic preferences. Because observational studies are generally more prone to selection, information, and confounding biases, it is all the more important for practitioners to have a firm grounding in epidemiology.

There are a variety of ways to express the results of RCTs and cohort studies to evaluate treatments. These include the RR, the OR, the absolute difference in risk (the incidence of disease in the control group minus the incidence in the principal group), and the number needed to treat, and the RR reduction (RRR). The number needed to treat is the inverse of the absolute difference in risks and is interpreted as the number of horses that need to receive a treatment to prevent one additional case that would have occurred during a specified period of time, had the comparison treatment been used: The lower the number needed to treat, the better the relative effectiveness of the treatment. The RRR is calculated by subtracting from the RR of disease in the principal treatment group from the RR in the control treatment group and dividing this difference by the RR of disease in the control treatment group: The RRR is expressed as a proportion and is analogous to the proportion of efficacy for a vaccine.

As mentioned earlier, there are two types of error that can influence study results: random error related to reliability of the technique for measurement(s) and systematic error in design or analysis. Often, the emphasis on interpreting results of clinical research reports is placed on the p value of the study. The p value associated with an association or comparison is a reflection of the precision of the study. Precision is a desirable feature of research studies: For example, we would not want a valid epidemiological study to be ignored because the results are too imprecise to establish clinical meaning. Although precision is important for clinical research results, there are limitations of p values. From the standpoint of interpreting results of studies comparing treatments and other exposures of interest, the principal shortcoming is that p values are frequently
taken as a measure of the strength of evidence of an association. For example, it is often suggested that a p value of 0.004 is much stronger evidence than a p value of 0.040. This interpretation has been widely propagated, including by scientific journals; however, the interpretation of a p value is not that straightforward. Because the p value is a reflection of random error, it is influenced by a number of factors, including sampling and sample size.

Sampling is often ignored in studies of horses but is an important source of sampling error in all epidemiological studies. Indeed, even if an entire population (e.g., all the horses at a race meet) are included in an epidemiological study, sampling error should be assumed because the study population will be considered by readers to be representative of a broader population or referent to the same setting at another time point when the population will have changed. More importantly, sample size influences the p value such that the p value cannot be interpreted as a fixed measure of strength and thus cannot be interpreted as a fixed measure of evidence. A very small effect in a large study and a very large effect in a small study may have the same value of P; when dealing with moderate-sized studies, the issue becomes particularly unclear, and modest- to moderate-sized studies are common in equine clinical reports.

It is also common that results yielding nonsignificant results are interpreted to mean that the treatment is ineffective and that the results are clinically unimportant. This interpretation ignores the size and direction of the observed effect of treatment (or other intervention). A very small p value can be associated with a clinically insignificant result: The p value does not inform us about the magnitude of the effect but only about the precision with which the effect is estimated. If we believe that the size of an effect is important, then a potentially effective new treatment may be ignored by considering only statistical significance in interpreting results. For this reason, some epidemiologists, statisticians, and clinical scientists advocate using a summary statistic and its associated confidence interval to summarize results of clinical studies. Confidence intervals convey more information by describing the range of possible true values that are consistent with the data rather than focusing on a single value and whether or not it is significantly different from the stated null hypothesis (as is the case for the significance testing approach). For example, consider a report regarding the incidence of laminitis in horses with colic. Using the p value as the principal source of clinical evidence (i.e., statistical significance testing), one would conclude that there was no significant difference (p = 0.138 by Fisher exact test; p = 0.227 using the Yates continuity correction for the χ² test) among horses that received heparin and those that did not. Indeed, the authors’ concluded that “the lack of obvious benefit makes one question the use of the drug in clinical cases of laminitis,” based on this interpretation of the significance testing results. This conclusion probably was highly influential in the subsequent diminished use of heparin as prophylaxis for laminitis.

As an alternative to significance testing, one might have looked at the magnitude of the estimated RR or OR and its 95% confidence interval. For this study, the odds of laminitis in horses treated with heparin relative to horses not treated with heparin was 0.33 (i.e., a 3-fold reduction in the odds of laminitis). The 95% confidence interval for this OR ranged from 0.09 to 1.27. Given the modest sample size and other sources of bias that could have influenced this retrospective study, these results suggested that heparin administration appeared to reduce the incidence of laminitis and that values favoring a beneficial effect were more compatible with the true RR than were values favoring lack of benefit. This interpretation could (indeed, should) have prompted subsequent prospective studies to evaluate the therapeutic benefits of heparin.

This example raises other important aspects of interpreting results of clinical epidemiological studies of treatments. First, because retrospective studies are more prone to bias, prospective cohort studies are better—and RCTs are best—for evaluating treatment. Second, expectations often are that therapeutic or prophylactic methods will be 100% effective. This is rarely if ever the case, and it is possible for treatments or preventatives to exert a salutary effect even when effectiveness is modest. For conditions such as laminitis, in which our understanding of treatment and prevention remain quite limited, even modest benefits (e.g., a 25% reduction in incidence of laminitis) could be clinically valuable.

Confidence intervals are not without limitations. As an alternative approach, likelihood ratios have been proposed as a more effective tool for assessing the evidence in clinical studies. A review of likelihood ratios is beyond the scope of this article, but the advantages of this approach are manifold and likely to become more commonly used for assessing clinical studies.

Another problem with interpreting p values from significance testing is the fact that failure to reject the null hypothesis is not equivalent to proving the null hypothesis is true: Absence of evidence of a difference is not evidence of absence of a difference. It is not unusual for clinical studies to claim equivalency for treatments on the basis of failure to detect a statistically significant difference. This is not desirable because the conclusion may be erroneous due to factors that influence p values, including sample size. Methodologies for equivalency testing exist, ranging from relatively simple to relatively complex methods. Succinctly, the issue boils down to the hypothesis being tested. To demonstrate that a difference exists between two treatments, we test the null hypothesis
that there is no difference between two treatments: if the treatments are significantly different, we reject the hypothesis that the two treatments are the same (however, we do not conclude that they are equivalent). For equivalency testing, the null hypothesis is stated as the difference (either the difference in absolute risk or RR) between two treatments is greater than or equal to a value determined prior to the conduct of the study by investigators. One approach for equivalency testing is that if the 95% confidence interval of the observed difference includes values that are greater than the prespecified difference, the treatments are considered not to be equivalent; if the confidence interval for the difference does not exceed the mean specified difference, the treatments are considered equivalent. It is important that readers do not confuse absence of an association as equivalent to saying that two treatments are either the same or equivalent or that scientific investigators claim that two groups may be considered the same because no significant difference in the distribution of a given parameter was observed between groups.

**Interpreting Risk Factor Data**

Observations from well-designed epidemiologic studies of horses can provide important information substantiating and suggesting hypotheses about pathways and processes that lead to disease. Thus, the process of identifying risk factors for disease is well entrenched in equine clinical research. Advances in molecular biology and new technologies and techniques have increased the opportunities to study risk factors for disease (or prognosis) by enriching the pool of factors that may be investigated to improve our understanding of the basis of disease development. The identification of a factor as being statistically significantly associated with disease or disease outcome does not indicate that the factor will be effective as a means for accurately classifying animals as diseased or not diseased. This is because a risk factor must have a much stronger association with a disease than is typically observed in etiologic risk factor studies to be effective either for diagnostic or screening test purposes.48–51 Nevertheless, many investigators and readers assume that because a risk factor is significantly associated with a disease that the factor will be useful in a clinical setting to distinguish horses that will develop the disease from those that will not (screening) or to differentiate those that have the disease from those that do not (diagnosis). Failure to recognize the limitations of risk factors as diagnostic or prognostic tools has led to disappointment or disenchantment with risk factor studies because a factor (e.g., plasma lactate concentration) may be significantly associated with a medical disorder (e.g., presence of a strangulating intestinal lesion or poor prognosis for survival for neonatal sepsis) but will not be useful in the clinical setting as a screening test or prognostic test for the disorder. This principle will be graphically represented during the presentation.

**Prospects**

The role of epidemiology in equine clinical research and clinical practice will be strengthened in the future. Increasing awareness of the importance of epidemiological studies for advancing clinical knowledge, particularly as a result of the emerging emphasis on evidence-based medicine, is helping to diminish the dominant roles of empiricism and laboratory experimentation in veterinary research. Rapid advancements in molecular biology, genetics and genomics, and computational biology provide new approaches and challenges for epidemiological studies. Genetic and genomic research is increasingly using case-control and other epidemiological study designs to identify the association between genotype(s) and phenotype(s), including disease. New genetic tools and better annotated resources for genetic data and metadata (such as gene ontology resources) are creating a surfeit of opportunities for equine genetic epidemiologic research. Similar opportunities for proteomic and other “omics” also are burgeoning.

The objective of epidemiological science remains the same: to obtain valid and reliable estimates of measures of disease or determinants of disease derived from groups to further our understanding of the causes of disease. Developments in and challenges for this scientific discipline will be determined by methodological and technological advances, such as novel study designs (e.g., case-only studies) and newer measures of exposure (e.g., genome-wide array data).

The limitations of the epidemiological approach result principally from the difficulties of deriving inferences about determinants of health and disease by observing horses in their natural environments. This source of limitation is also the root of the discipline’s greatest strengths: (1) The results of epidemiological studies are directly relevant to the “real world” setting in which the data were collected; and (2) the epidemiological approach does not require the creation of any new pain or suffering because only spontaneous disease is studied and thus has an advantage from a welfare perspective.

In closing, I wish to emphasize two important ways in which epidemiology is relevant to clinical practice. First, patient-based epidemiological studies are the best source of information on which to base our clinical decisions, including selecting and interpreting findings of diagnostic tests, assessing optimal treatments and preventative such as vaccines, and making statements about prognosis. Thus, epidemiology is a basic science of fundamental importance to clinical practice. Second, epidemiology is a science in which practitioners can participate, contributing knowledge that is of direct relevance to practice. The vast majority of equine disease is observed in private practice. Practitio-
ners are thus in a position to contribute observations to epidemiological studies and to formulate etiologic hypotheses about disease(s) that can be evaluated using epidemiological study designs. It is my belief and hope that the American Association of Equine Practitioners can continue to be a powerful alliance of practicing veterinarians who will contribute to epidemiological investigations. The mission statement of the AAEP is, “To improve the health and welfare of the horse, to further the professional development of its members, and to provide resources and leadership to the benefit of the equine industry.” Thus, AAEP member–based studies are consistent with the organization’s mission statement: Practitioner-based studies will provide information directly relevant to improving the health and welfare of our patients, will contribute to the professional development of members by participating in the conduct and interpretation of study results, will offer opportunity to provide an invaluable resource (data from one’s patients), and will demonstrate leadership and team-building as a member of practice-based networking.

References and Footnotes