ELSEVIER
SAUNDERS

VETERINARY
CLINICS
Food Animal Practice

Vet Clin Food Anim 22 (2006) 125–147

Determining Cause and Effect in Herds

John M. Gay, DVM, PhD

Department of Veterinary Clinical Sciences, AAHP Field Disease Investigation Unit, College of Veterinary Medicine, Washington State University, P.O. Box 646610, Pullman, WA 99164-6610, USA

Establishing cause and effect is the critical step of many clinical diagnostic work-ups, particularly for herd problems, because understanding cause provides the basis for control and prevention decisions. The confidence required in one's conclusion about cause and effect depends on the situation. Factors include the comparative costs and risks of the problem and potential interventions and the difficulty of implementing interventions, particularly if changes in ingrained human behavior are required. Because of the complexity of biologic systems and processes, the often long lag between cause and effect, biologic variability within the same animal over time and between animals at the same time, and the inevitable continual changes in husbandry conditions (eg, individual animals proceeding through the production cycle, weather changes, feed changes, and personnel changes), defining cause with certainty sufficient for developing control and prevention interventions and with confidence is challenging, even with a definitive clinical diagnosis.

Although the continuing economic loss from a problem usually forces rapid decisions after the problem is recognized, the available information is often incomplete. The cost and delay of acquiring additional information must be balanced against the value of a more timely intervention and the potential cost of error due to incomplete information. To effectively and efficiently determine cause in diverse livestock production environments, considerable background knowledge and creativity are usually required. Although it is important, a definitive clinical diagnosis alone is often not enough to identify the causes of disease problems that are changeable as part of control or prevention interventions. By their nature, most livestock enterprises are complex, dynamic systems with variously lagged positive and negative feedback loops impacted by continual changes in controlled (eg,

E-mail address: jmgay@vetmed.wsu.edu

0749-0720/06/$ - see front matter
doi:10.1016/j.cvfa.2005.12.004 *vetfood.theclinics.com*

change in feed source) and uncontrolled inputs (eg, weather, seasonal climate), changes in location as animals move through the production cycle, and so on. On extensive livestock operations that have low-technology systems, such as grazing beef cow-calf operations in the intermountain West, information is sparse because of infrequent animal observation and the common lack of individual performance records. The coming of mandated individual animal identification from birth may be an opportunity to remedy this lack of information. On intensive livestock operations that have highly technical, tightly interconnected systems, such as large dry-lot confinement dairies, even when sufficient standard operating procedures are in place, uncertainty results from the continual compromises required due to economic and resource limitations, recurring problems due to procedural and technical system drift, and human nature, particularly when poor communication occurrs between personnel.

The purpose of this article is to present the components of the logical process for determining cause and effect and to list common cognitive errors of the medical decision-making process, the thought being if individuals are aware of these errors, then they will be better able to avoid committing them. The first section provides the concepts used in considering cause and effect relationships, the second section provides a logical basis for evaluating cause and effect relationships, and the third section illuminates potential reasoning errors.

Underlying concepts

The following concepts are important for considering cause and effect; the reader is encouraged to consult a dictionary of epidemiology [1] or a veterinary epidemiology text such as Thrusfield [2] for additional details. The **effect** is the outcome of interest, which may be the occurrence of disease, low production, poor performance, or other event or phenomenon. The **case definition** is the set of criteria (eg, death or sickness with certain clinical signs) used to establish which animals experienced the effect. **Risk factors** are individual attributes or exposures that may be involved in the cause of a specific effect, increasing or decreasing the risk of the effect occurring. An **attribute** is a risk factor that is an intrinsic characteristic of an animal, such as genetic susceptibility, immune status, age, sex, breed, or weight. Some attributes can be changed quickly, such as improving host resistance by vaccination, whereas others require a much longer time frame, such as reducing genetic susceptibility by changing the breeding program. An example is reducing the risk of ocular squamous cell carcinoma in future replacement Hereford heifers by using only breeding bulls with dark pigmentation around their eyes. An **exposure** is a risk factor that is in the environment external to the individual, such as nutrition, housing, husbandry practice, or an infectious or toxic agent. A **risk marker** is a noncausal factor associated sufficiently well with a risk factor so that it can be used as a marker, or

indicator, of exposure to that specific risk factor when detection of the factor itself is difficult and expensive. For example, serologic titer is often used as a marker for previous exposure to an infectious agent when the agent itself cannot be detected; however, cross-reactions with other antigens and vaccination-induced response can make interpretation difficult.

As defined elsewhere in this issue, a **key determinant**, sometimes called a risk determinant, leverage point, or a critical control point, is a specific risk factor that can be modified or eliminated to control or prevent the effect. In terms of the benefit from changes versus the cost of the changes, some key determinants are more important than others. For example, the risk factors for the transmission of the endemic infectious agents involved in calf scours are more often key determinants than simply the presence of common viral agents because these agents are essentially endemic in almost all livestock operations. The key question is why (if the infectious agents are essentially ubiquitous) some livestock operations have the clinical problem when many others do not. The answer is differences in the key determinants. In such circumstances, rather than limiting interventions to the use of pharmaceuticals for treatment or biologics for prevention, the best interventions often involve long-term changes in facilities and human behavior, which is often expensive and difficult. Carrier cows, however, are certainly a risk factor that could be modified by testing the herd, but for such agents, the cost would likely exceed the short-term benefit.

For diseases with a long latency, such as cancer, the exposure may have occurred much earlier in the host's life. For other diseases such as persistent bovine viral diarrhea virus infection, the exposure must have occurred during a specific period in the animal's life. The **induction** or **incubation period** is the time required from exposure to a specific risk factor until initiation of the disease. These periods are usually distributed in somewhat of a normal or bell-shaped curve that is determined by infectious dose and host susceptibility. Generally, the longer the induction period, the more difficult the assessment of risk factor exposure and disease and, thus, the more difficult the evaluation of cause and effect. The **latent period** is the time between biologic onset of the disease process and disease detection (clinical disease: appearance of clinical signs; subclinical disease: positive diagnostic tests). Potential risk factors that acted on a case in less than the necessary induction and latent periods for their action cannot be components of the cause and effect.

The cause of cause and effect is the presence of a combination of risk factors that alone or in combination and in the correct sequence and timing during the animal's life inevitably result in the effect (such as clinical disease or low production) occurring in that individual. A **necessary cause** [3] is a specific risk factor that must have been or must be present at the appropriate time in the causal model for the effect to occur, such as a specific infectious agent that is the etiologic agent in a particular infectious disease. **Sufficient cause** [3] is that set of risk factors in the causal model whose confluence in an animal's life, with appropriate timing, inevitably results in the

effect. Note that if the effect of concern is an infectious clinical disease, many infectious agents are necessary causes but not, by themselves, sufficient causes for the clinical disease for which they are the etiologic agent. **Competing risks** are other sets of risk factors that can cause the condition of concern (such as death or low production) that coexist with the set of factors of interest. These risks are things that cause ''red herring'' cases at a constant, background rate in the group of animals being worked up for a particular problem.

A **causal model** is the scheme of what happened from the start of the problem to the effect in sufficient detail that potential intervention points are identified as the basis of an effective plan for controlling or preventing the problem. J.P. Box is alleged to have said that all models are wrong; some are just more useful than others. A **causal web**, sometimes called a causal pathway or a path model, contains the specific links that connect each individual risk factor (acting across time or together) and result in the effect in an individual animal or group. The first level of factors are those that act directly on the affected animal, the second level of factors are those that act through one or more first-level factors, the third level of factors act on second-level factors, and so on. The goal is to develop this web in sufficient detail so that risk factors that are the best key determinants are identified. A sketch of the model is often useful in explaining to producers why the problem occurred, how the components are linked, and how to control or prevent it in the future. These models are often unique, with different sets of risk factors for animals in different groups, sometimes even on the same premises. A given disease can be caused by more than one set of sufficient causes and, thus, result from different causal pathways in animals contracting the disease in different management situations. What is effective in one situation is often not effective in another because of farm-specific differences and the basic differences in the component risk factors and their linkages (eg, bronchopneumonia in a barn-housed dairy calf versus a hutch-housed dairy calf versus a feedlot calf; Figs. 1 and 2). Although infectious agents may be necessary causes, for the purposes of prevention, very few cause disease by themselves (ie, are a sufficient cause by themselves) and operate more as opportunists, exploiting weaknesses in livestock management systems. Even for agents that, from a pathology perspective, are a sufficient cause by themselves, the causal web includes the risk factors for introduction into a herd and for continued transmission within the herd after the agent is present.

Evidence is information that tends to support or refute that a cause-and-effect situation exits. Depending on how it was obtained, evidence varies greatly in strength. **Empiric evidence** includes the facts of the situation that are obtained by examining, measuring, or counting rather than by reasoning or from impressions or feelings. The strongest empiric evidence is that obtained from a properly designed and executed experiment, such as a randomized, blinded, controlled clinical trial. The weakest empiric evidence is that obtained from a single case, such as a single necropsy in the

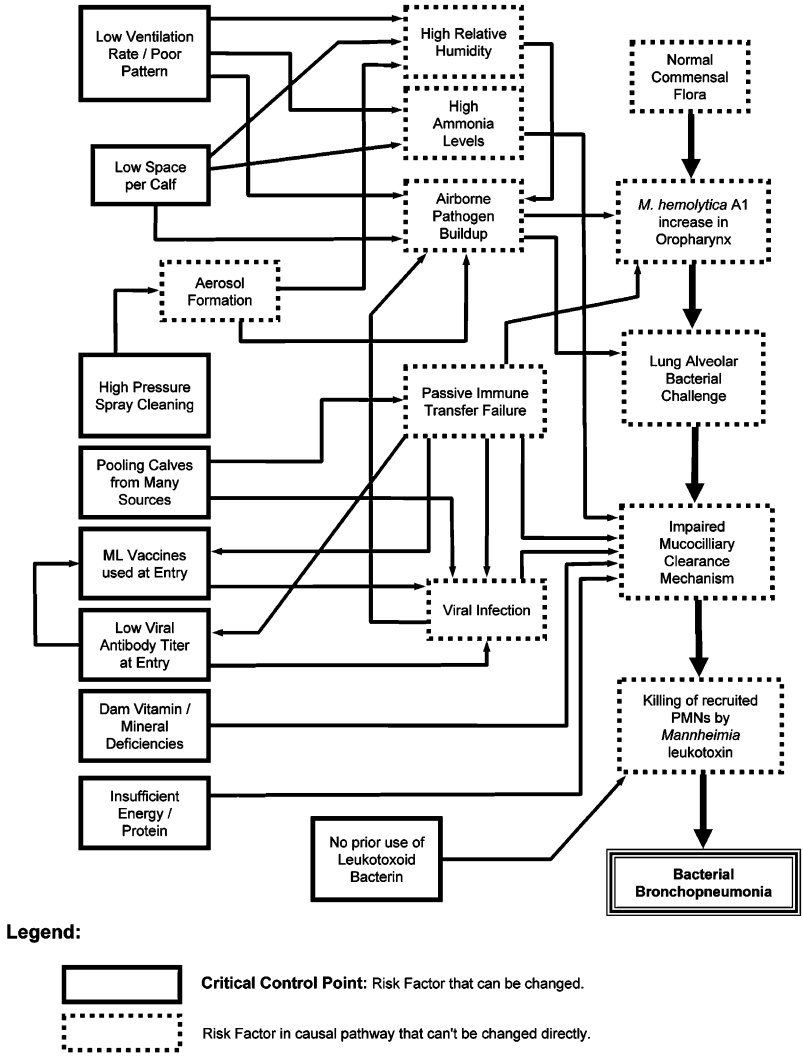

Fig 1. Example of a causal web of risk factors for housed-calf bronchopneumonia.

midst of an outbreak. Empiric evidence for a cause and effect relationship is weakened by the opportunity for other explanations to account for the findings. For example, the necropsied animal may have been one of the sporadic deaths that are at a low but continuous risk of occurrence in groups of animals. For this reason, observational evidence (ie, "an experiment of nature") is inherently weaker than experimental evidence (eg, a randomized, blinded, controlled clinical trial). Stronger observational evidence is obtained from counts, records, or direct observation than from recall, from prospective cohort studies, or from retrospective case-control studies.

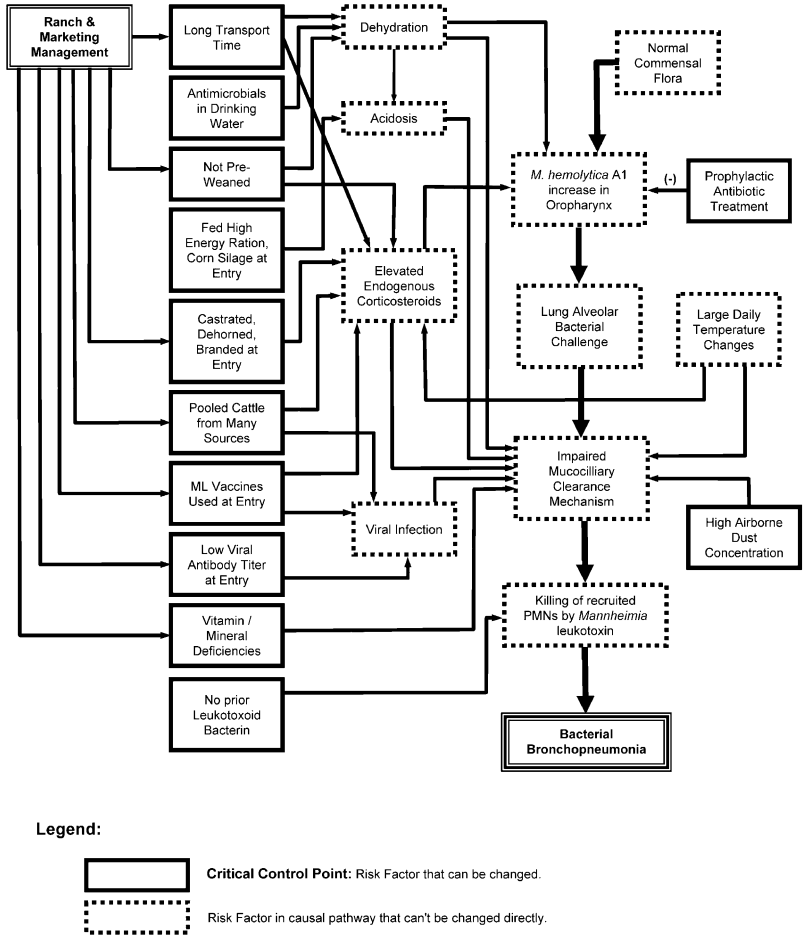

Fig 2. Example of a causal web of risk factors for feedlot-calf bronchopneumonia.

Analogic evidence is the evidence for a cause and effect relationship that is based on reasoning by analogy, which is concluding (from comparing known similarities between two systems) that a relationship shown to exist in one system but unknown in the other system also likely exists in the other. For example, if drug "X" has been shown to be effective against disease "Y" in a species "Z," then perhaps the same relationship exists with a similar drug or a similar disease or a similar species, the other two components of the analogy being identical. Evidence based on analogic reasoning is common in veterinary medicine because it provides a basis for action when the sounder, empiric evidence is lacking. With the broad range of species served by the profession, the many breeds often existing within a species, and the relatively scarce resources for veterinary research, the practice of veterinary medicine is often based on analogic evidence. An example is using a drug in

a minor species based on what is known about the pharmacology of that drug in a somewhat similar major species. Analogic evidence is inherently weaker than empiric evidence because of the likelihood that different, unknown factors are operating in the two systems, weakening or invalidating the analogy. Because analogic reasoning yields an inherently weaker form of evidence than empiric evidence, it is likely to be the source of much unexamined dogma and, when possible, better used as a basis for generating hypotheses that can be empirically evaluated.

The weakest form of evidence is **anecdotal evidence**, which is the evidence from a single event such as the medical recovery of a single case or the necropsy of a single animal in an affected group. The probability of apparently unusual events is often considerably higher than is intuitively expected, and other unrecognized factors (confounders) may have invalidated the initial prediction of disease course, thus making the event not that unusual. For example, that a group of only 23 people is required to have a 50% or better chance that two or more of the group have the same birthday is counterintuitive yet easily shown with probability calculations.

Dogma are unexamined beliefs, which can be right or wrong, that are held as established or put forth as an authoritative or expert opinion but that have little or no supportive empiric evidence. An example is the long-held belief that routine intrauterine infusion of antibiotics during prebreeding examinations improved reproductive performance in dairy cows. This practice made biologic sense but was not supported by empiric evidence from subsequent field studies. Medical dogma is usually derived from unevaluated biologic hypotheses and uncritical observation or experience without recognition of the effects of chance, the complexity of biologic systems, natural biologic variation, and observer bias. It is unfortunate that a significant portion of the veterinary medical knowledge base likely falls into this category. Repetition across sources or the number of people (whatever their qualifications and experience) who hold this belief does not change the status of such information until it has been properly examined in light of sufficient strong empiric evidence obtained from sound studies. Because of time and resource limitations, much information conveyed in instructional settings (before and after attaining a doctorate of veterinary medicine degree) is presented in the form of dogma without the associated information necessary to judge its credibility or strength of evidence. The elimination of dogma is the primary goal of the evidence-based medicine paradigm that emerged in human medicine during the last decade.

The **scientific method**, which is a process for obtaining and assessing empiric evidence, provides a sound logical scheme for considering the evidence for a cause and effect relationship. It has evolved as the strongest procedure, given the inescapable limitations of human observation ability and cognitive function, for developing a valid understanding of nature. The basis of the scientific method logic is the **hypotheticodeductive model**. This model comprises (1) the logical deduction of a prediction, (2) the collection of the

evidence to examine this prediction by an experiment such as a randomized controlled trial or by observation such as a case-control or cohort study, (3) the analysis of this empiric evidence, and (4) the logical induction that the cause and effect relationship is supported or not supported. **Deduction** is reasoning from the general to the specific situation; **induction** is reasoning from the results of the specific circumstance to conclusions about the general situation. A **hypothesis** is a provisional conjecture that further evidence will support or refute. In this sense, a list of differential diagnoses represents a set of hypotheses that guides the process for determining what additional evidence must be obtained in the diagnostic process and for evaluating that evidence (ruling in or ruling out diagnoses) before a level of certainty sufficient for action is reached. Having a list of specific differentials (hypotheses) guides the clinician in selecting laboratory tests (empiric evidence of varying strength) that will be useful in ruling in or ruling out a differential (further directed investigation). When the risk and cost of treatment is sufficiently low and the differential is sufficiently likely, treatment response is often used as empiric evidence. Similarly, a set of working hypotheses about a herd problem guides the clinician in deciding what herd record data to gather, what animals to examine, what farm procedures to watch, who to interview about management practices and recent events, and so on. A logical problem is that although the facts derived from the test of a hypothesis are correct, the broader underlying theory from which the hypothesis was derived can still be wrong (eg, the laboratory value can be abnormal for reasons other than a problem on the list of differentials).

Hypotheses are most useful if set up as a testable "if…is the cause of…then" statement. The proposed cause and the effect are contained in the "if" component, and the predicted outcome of an evaluation of this hypothesis, such the experimental manipulation of one of those variables or the observation of a natural experiment, is contained in the "then" component. Performing the test (experiment or observation) leads to a logical conclusion that supports or refutes the proposed relationship. For example, if transition pen crowding is the cause of the increased risk of displaced abomasums, then the risk of displaced abomasums should be higher in the weekly calving cohorts with more cows than in the cohorts with fewer cows. This hypothesis could be evaluated visually by creating a graph that contains a smoothed plot of displaced abomasum risk in the weekly calving cohorts and a smoothed plot of number of cows calving by week over a weekly timeline. If the peaks and valleys of the risk plot correlate with the peaks and valleys of the density plot, then it is time to talk to the manager. Establishing these "tests" often involves considerable creativity in identifying on-farm information resources and in execution. The better the "test," the stronger the evidence in supporting or refuting the relationship.

Hypotheses are often confused with theories. In nonscientific contexts, the word "theory" is often used to mean a mere hypothesis or

speculation—a much less reliable proposition than is a scientific theory (eg, "I have a theory that it is going to rain today"). A scientific **theory** is the coherent, inter-related structure of scientific propositions and principles derived over time from empiric scientific evidence that explains a class of observed phenomena or facts. Theories enable us to make sense of what we see in nature and to make predictions. For example, a clinician may hypothesize that a down cow has hypocalcemia, but the clinician's treatment of the case and recommendations for prevention in the herd are based on the physiologic theories of calcium regulation in the bovine. A scientific theory must have predictive power (ie, predict phenomena that are observable) and must be testable and falsifiable (ie, the theory is falsified if the predicted phenomena are not observed in the appropriate experiments).

Propositions that are not testable by appropriate scientific research are not theories. Theories are not absolute truth, cannot be verified as such, and are always subject to change due to advancing knowledge and research technology. For example, the physiologic theory about calcium regulation in the bovine has changed over the past several decades as the result of continuing scientific research and, consequently, testing and prevention methods have changed accordingly. Thus, any claim of scientific truth or scientific proof can be immediately disregarded. Any theory that can be tested only by select, unique methods or by certain individuals (usually the promoter of the theory) is highly suspect. A scientific theory must be consistent internally and consistent with broader, more fundamental theories related to other aspects of the phenomenon. For example, a theory of disease physiology cannot be inconsistent with most of the more fundamental and considerably stronger theories of chemistry. It is unfortunate that many of the theories supporting alternative medical practices fail this test. An erroneous underlying theory, however, does not necessarily invalidate a particular treatment because the treatment may represent doing the right thing but for the wrong reason. In these circumstances, empiric evidence derived from the proper application of the scientific method is the only valid basis for evaluating the efficacy of such treatments, and reasoning from the theory should serve only as a source of hypotheses for testing. Depending on relative qualities and amounts of empiric evidence, competing theories explaining the same phenomenon usually differ widely in strength and consistency from more fundamental theories that have broader support. Similar findings replicated by different observers using different but appropriate experimental or observational methods at different locations are regarded as stronger support than is replication by the same investigators or by using identical methods. Although in the end, everything is linked in some way to everything else, the "granularity" or level of detail (eg, atomic, molecular, tissue, organ, body) necessary in a theory depends on the purpose for which it is being used as the basis for prediction.

An **association** exists between two phenomena if a change in one corresponds to a change in the other. An association must be present for a cause

and effect relationship to exist, but the presence of an association alone does not prove that the relationship is causal. A positive association or direct relationship exists if the magnitudes of both variables move up or down together. When the correlation coefficient is positive and when the relationship is causal, higher levels of or greater exposure to the risk factor may cause more of the outcome (a dose-response effect). For example, the number of high-string clinical coliform mastitis cases increases when the number of coliforms in the organic bedding is higher. A negative association or inverse relationship exists when the magnitude of one variable moves in the opposite direction of the associated variable. When the correlation coefficient is negative and when the relationship is causal, higher levels of the risk factor are protective. For example, the number of early postpartum metritis cases declines with higher levels of vitamin E in the prepartum ration.

The strength of the association between a risk factor and the outcome in a cause and effect relationship is measured several ways, including the correlation coefficient, relative risk (RR), odds ratio (OR), and logistic regression coefficients. The RR is how much more likely the disease will occur in animals with exposure to the risk factor than in animals without exposure (the risk of disease in the exposed group divided by the risk of disease in the unexposed groups). OR equals the risk of exposure to a risk factor in cases divided by the risk of exposure to a risk factor in controls.

For RR and OR, a value of 1 or close to 1 indicates that a biologically significant cause and effect relationship is not present. The threshold for biologic significance (how close to 1 is important) depends on the relative costs of cases and of prevention. A value larger than 1 is a positive association; a value less than 1 is a negative (protective) association. In scientific studies, the 95% confidence intervals for measures that include 1 are interpreted as meaning that no cause and effect relationship is present for that risk factor in the studies that had sufficient power to detect clinically significant differences. As noted in the statistics article found elsewhere in this issue, in these situations, the χ^2 p-value is larger than 0.05. If, however, a negative study (eg, statistically insignificant findings) had insufficient power to detect the minimum clinically, biologically, or economically significant difference, then the only valid conclusion is that the study should have been larger. In diagnostic work-ups on herds, the width of the 95% confidence intervals is better used to indicate the precision or "stability" of the estimate rather than as a measure of statistical significance. When the estimate is too imprecise (the confidence interval is too wide) for comfort, more sampling and testing is likely in order.

An example of how to calculate an OR from a two-by-two table is provided in the statistics article found elsewhere in this issue. Many on-line calculators are also available for these measures and can be found using Web search engines to search for terms such as "odds ratio calculator." To check that these on-line calculators are programmed and being used correctly, one

should enter the numbers from a trusted, worked example and compare the answers. Entering different counts into these calculators, such as doubling the number of controls sampled while keeping the proportion constant, provides an intuitive understanding of how changes in sampling strategies changes the precision of the result.

Although the OR is only an approximation of RR and is a better approximation of RR when the disease is rare than when it is common, as noted in the statistics article found elsewhere in this issue, the OR is often the only valid measure of risk. For example, the OR is the only valid risk measure for a case-control study. A general rule is that the OR overestimates the true underlying RR, being further from 1 than the RR and the OR confidence interval being wider than the RR confidence interval when the RR is valid. When the disease is common, the OR overestimation can be an order of magnitude. Although RR is the preferred measure, it is only valid when all the animals in an exposure group have an equal probability of being included in the calculation (eg, when an entire calving cohort is followed through the risk and disease expression periods with few losses to follow-up, when an entire herd is randomly sampled for testing irrespective of case status, or less commonly, when animals are selected for study on the basis of exposure status). RR estimates are valid only when disease incidence estimates are also valid, such as in cohort studies. In situations in which controls are matched to cases, such as one control for each case, or when cases and controls are selected separately, only the OR is valid. When controls are matched to cases, matching more than three controls per case does not add much additional information on risk relative to the expense. All of these measures are potentially biased by the loss of animals to follow-up when the risk of loss is associated with disease status or exposure status and by misclassification of case or exposure status.

Logical basis for causal reasoning

Ceteris paribus—*holding everything else constant*

The reasoning required for determining cause and effect is often not simple or straightforward. The logical principles for establishing cause and effect relationships have been tempered over the centuries as people struggled to develop sound methods for transforming observations from nature and medicine into reliable, repeatable understandings of cause and effect; namely, to make useful decisions about treatment and prevention. One of the first long-lasting principles to emerge is that of Ockam's Razor (William of Ockam, 1285–1349), which is the principle of parsimony, simplicity, or economy. It states that when two causal models explain a phenomenon equally, the simpler model that requires fewer assumptions and explanatory principles is more likely to be true. Francis Bacon (1561–1626) was the first

to develop the methods of inductive reasoning from empiric findings as the basis for developing sound scientific knowledge. His work became the foundation of the modern scientific method.

In the System of Logic, John Stuart Mill (1806–1873) distilled Bacon's methods into several principles of experimental inquiry that provide a logical basis for generating and evaluating hypotheses about cause and effect relationships. The **method of agreement** is that if a risk factor is common to multiple instances of the effect when other factors are dissimilar, then that factor may be a cause. For example, if the same type of feed from the same feed mill is associated with the appearance of the same problem on multiple farms, then that feed is a good candidate to be the cause. The **method of difference** is that if the risk of disease is different when one risk factor is different but other factors are similar, then that risk factor may be a cause. For example, if diarrhea is occurring in pens of horses being fed sweet mix from one bin but not in pens of horses being fed sweet mix from other bins, then the sweet mix in the one bin is a good candidate to be the cause. The **method of concomitant variations**, or simultaneous changes, is that if the risk of disease changes with changes in the level of a risk factor and other factors are more constant, then that risk factor may be a cause. For example, on a dairy where dry cows are kept on a premise separate from the lactating herd, if the risk of displaced abomasum is higher in cows calving a shorter time after being hauled from the dry-cow herd than in cows calving a longer time after being hauled, then hauling is a good candidate to be a cause.

Developed from Mill's methods, the Henle-Koch Postulates (1877) are four sequential criteria for assessing a cause (particular infectious agent) and effect (particular disease) relationship that are familiar to veterinarians. The criteria that must be met before the relationship is accepted as causal are

1. The microorganism must be found in all cases of the disease.
2. It must be isolated from the host and grown in pure culture.
3. It must reproduce the original disease when introduced into a susceptible host.
4. It must be found in the experimental host so infected.

The logical power of these postulates rapidly advanced the germ theory of disease over competing theories of disease cause at the time, such as humors and miasma. Because these criteria focus exclusively on the infectious agent, these postulates are insufficient for identifying the nonagent key determinants involved in infectious or noninfectious diseases or for distinguishing circumstances resulting in subclinical infection from those resulting in clinical disease. Fulfilling the postulates experimentally can be surprisingly difficult, even when the infectious process is well understood and the agent is regarded as a pathogen. Because of these limitations, the Henle-Koch Postulates were superseded by the Hill-Evans Postulates as more generally useful for establishing causality under broader circumstances.

The Hill-Evans Postulates are a set of 9 or 10 criteria, depending on one's interpretation of the original papers, for establishing cause-and-effect relationships. Fulfilling the entire set constitutes very strong evidence for causality but, unlike the Henle-Koch Postulates, the failure to fulfill one or more particular criteria may not significantly weaken evidence for causality. Strong negative evidence for certain criteria, however, logically refutes causality, whereas other criteria provide only weak evidence for cause when fulfilled. Hill [4] proposed the following criteria (Box 1) to evaluate cause and effect relationships for noninfectious diseases, particularly the relationship between smoking and lung cancer.

Based on Hill's [4] criteria, Evan's Postulates [5] provide a direct basis for developing testable hypotheses by constructing comparisons (Box 1).

Creating and evaluating formal hypotheses using these principles provides a powerful basis for establishing cause to the level of certainty needed for action. The unique strength of evaluating cause and effect relationships in herds is that affected animals (clinical and subclinical) can be compared with unaffected animals in a cross-section (at one point in time) and over time to determine the differences and similarities between the animals themselves and the factors affecting them. An expert in systems analysis stated, "Starting with the behavior of the system directs one's thoughts to dynamic, not static analysis—not only to 'what's wrong?' but also to 'how did we get there?' … And finally, starting with history discourages the common and distracting tendency we all have to define a problem not by the system's actual behavior, but by the lack of our favorite solution" [6]. For example, many calf scour agents are ubiquitous and, often, sampling scouring calves only confirms this ubiquity. The critical question is, Why are these individuals on this farm having a problem with this agent when many others are not, even though the infection is most likely present there as well?

Failure to recognize the cause and effect time frame inherent in the type of problem can lead to errors. The three general types of problems are acute, additive or cyclic, and chronic. **Acute problems** are precipitated by a temporally associated management or husbandry error of sufficient magnitude to be the primary cause of the problem. **Additive** or **cyclic problems** are precipitated by a combination of management or husbandry errors over time and the effects of cyclic factors such as season or production cycle stages such that the combination is sufficient to precipitate the problem (eg, the summer coliform mastitis outbreak that is associated with the previous winter change to sawdust bedding). **Chronic problems** are precipitated by the long-term action of management or husbandry errors that require the passage of time before the consequences become of sufficient magnitude to be recognized, such as the slow spread of a contagious mastitis agent or of *Mycobacterium paratuberculosis* (eg, the recognition of a slowly spreading *Staphylococcus aureus* mastitis problem associated with the adoption of a less efficacious teat dipping procedure more than a year previously). For additive, cyclic, or chronic problems, the initial occurrence of the underlying

Box 1. Hill's criteria for causation (1965)

1. Strength of association: the larger the relative effect, the more likely the causal role of the factor. At minimum, a biologically significant association must be present for a cause and effect relationship to be present.
2. Consistency: if similar associations are found in different studies in different populations, then the more likely the causal role of the factor.
3. Specificity: if the effect does not result from other causes, then the more likely the factor is causal.
4. Temporality: risk factor exposure (the cause) must precede the effect. Solid evidence that the effect preceded the exposure to the risk factor indicates, at least, that other risk factors are also causal or, at most, that this factor is not.
5. Dose-response (biologic gradient): if the risk increases with increasing dose of or exposure to the risk factor, then the more likely that a cause and effect relationship exists.
6. Biologic plausibility: given current knowledge, the mechanism is biologically plausible, in that it is consistent with and does not contravene well-established theory.
7. Coherence: associations between the risk factor and the effect are consistent with existing knowledge and do not conflict with the generally known facts of the natural history and biology of the disease.
8. Intervention (experiment): reduction or removal of the risk factor reduces the risk of the effect (the strongest evidence of a cause and effect relationship).
9. Analogy: that a similar but not identical cause and effect relationship has been observed and established elsewhere as causal provides weak evidence for causality.

Evan's Postulates (1976)

1. Prevalence of the disease should be significantly higher in those exposed to the risk factor than in those not exposed.
2. Exposure to the risk factor should be more frequent among those with the disease than among those without.
3. In prospective studies, the incidence of the disease should be higher in those exposed to the risk factor than in those not exposed.
4. The disease should follow exposure to the risk factor with a normal or log-normal distribution of incubation periods.
5. A spectrum of host responses along a logical biologic gradient from mild to severe should follow exposure to the risk factor.

6. A measurable host response should follow exposure to the risk factor in those lacking this response before exposure or should increase in those with this response before exposure. This response should be infrequent in those not exposed to the risk factor.
7. In experiments, the disease should occur more frequently in those exposed to the risk factor than in control subjects not exposed.
8. Reduction or elimination of the risk factor should reduce the risk of the disease.
9. Modifying or preventing the host response should decrease or eliminate the disease.
10. All findings should make biologic and epidemiologic sense.

management or husbandry deficiency is usually not close in time to the recognition of the problem. In many cases, because of the lag between a management change and recognition of the problem, the manager is reluctant to acknowledge that a change precipitated the problem.

Subjective perceptions of employees and managers are valuable sources of hypotheses about risk factors; the clinician's task is to support or refute these and other hypotheses using objective empiric evidence. Unless based on objective evidence (eg, analysis of records), employee and manager perception of the problem may be correct but, more often than not, it is off the mark. The clinician can compare the actual number of cases to the expected number to determine whether the frequency is excessive. The clinician should be very careful of "dangling numerators"; that is, counting the number of cases without considering the number of animals actually at risk of becoming a case during that time period. A large increase or decrease in the number of animals susceptible to a condition causes a corresponding change in the number of cases of that condition, even though the underlying risk remains constant. Because of seasonal effects, few herds maintain a constant number of animals passing through the period of susceptibility year-round. Changes in individual animal performance must be distinguished from changes in total output that are due to changes in the numbers of producing animals.

The strength of herd-focused investigation compared with individual-focused investigation is the opportunity to compare affected animals to unaffected animals. What are the characteristics of affected versus unaffected animals in terms of exposure to potential risk factors, age, production level, stage of production cycle, and source? As noted earlier, because of the spectrum of disease, the clinician must be careful when classifying animals into affected and unaffected groups. Another common error is to overlook the culled or dead animals in the cohort of susceptible animals that entered

the risk period because their records were deleted from those of the remaining animals. Where were the affected versus unaffected animals located during the potential exposure period? Because different groups or pens of animals often have different levels of exposures (eg, different amounts of feed ingredients, different water sources, different housing, different pasture, different origins, different stages of the production cycle) and because a dose-response relationship exists for many etiologic agents, this set of clues is important.

Gathering and analyzing objective data on a herd problem is an examination process that is analogous to using laboratory tests or imaging procedures in the diagnosis in an individual animal. These objective data support or refute clinical impressions of the herd problem much like testing or imaging supports or refutes clinical impressions of the clinical case. The clinician should concentrate on the data that will support or refute hypotheses (differential diagnoses). Computer spreadsheets provide a convenient means to enter, validate, and manipulate the relevant individual and group information. If accessing herd data through a production accounting system, the clinician first evaluates the data for quality by verifying that known, relevant events identified by means other than the records system are present and correct in the records. Outliers and logical inconsistencies in the data are detected by sorting variables into numeric order and looking at the minimums and maximums. Plots of production data with variable smoothing over time are easy to create, allowing trends to be discerned amid the noise of random variation. From count data of the numbers of affected and the numbers of susceptible animals, case morbidity and fatality rates by exposure and RRs can be calculated. For endemic problems, risk over time by cohort group can be plotted. The clinician can construct cohorts of at-risk animals passing through a critical point in the production cycle associated with the problem (eg, calving, weaning) over a time interval (eg, day, week, month) that on average, have enough animals (30 or so) to reduce the effects of natural variation but do not obscure trends over time, with wider intervals being needed for smaller herds. The effects of other factors that vary over time (eg, calving pen density, average of weekly high temperature, sources of animals) on risk of occurrence or production can be examined. Detailed weather data from nearby automated weather stations can be downloaded from on-line government sources into a spreadsheet. If the herd does not have a good production accounting system for animal performance information, then the clinician should not overlook clues found in indirect sources of similar information. For example, the delivery dates and weights on feed invoices can provide approximate information on feed batch disappearance and, thus, approximate information on consumption patterns. On this basis, expected disappearance of feeds can be compared with actual disappearance. Invoices from rendering services may provide information on dates of animal deaths if they have not been recorded. Often, events such as calvings and breedings are written on calendars or in pocket books.

Establishing and executing good tests of hypotheses require a large amount of ingenuity. Based on the previous hypotheses, the clinician can predict what should be found in other animals, such as diagnostic test results or production effects, and can evaluate these predictions. As noted earlier, predictions of the form "if this cause is present, then this finding should be present" can be made. Because many causes have multiple effects, finding more of these multiple effects provides stronger support for the presence of the cause than finding only one. Often, a single effect can result from several different causes. Finding what is predicted supports a hypothesis; not finding what is predicted weakens a hypothesis. The key is figuring out what predictions will provide good tests and are "doable." Just as in individual animal work-ups, "scattershot" sampling and testing should be avoided because doing so without an objective in mind is seldom useful and is expensive monetarily for the client and expensive in credibility and time for the clinician. Ockam's Razor should be applied by asking, What is the simplest set of explanations that covers the most findings?

The following are example predictions:

If this infectious agent is being transmitted between animals in this manner, then these other animals are at risk and some will be infected, whereas these others are not at risk and will not be infected.

If overcrowding in the fresh pen (risk factor) is causing displaced abomasums (disease), then the pattern in the associated data that can be expected is a higher proportion of cows experiencing displaced abomasums in the cohorts with more crowding in the fresh pen compared with cohorts with less crowding.

Common causal reasoning errors

When determining cause and effect, veterinarians and their clients are subject to lapses and biases in reasoning and memory recall that can lead the process astray. The following are several important concepts in considering these errors.

Belief is the mental act or state of mind of an individual after he or she accepts and internalizes an external concept or idea, which then becomes part of subsequent thought processes. Internalized deeply, belief becomes part of intuition, particularly for the expert performing conscious and unconscious pattern recognition. Belief can occur after deliberate, systematic, critical thinking or can occur with immediate, nonreasoned, uncritical acceptance. The problem is that after an erroneous belief is established, accepting a more correct belief becomes considerably more difficult than if no previous belief was held. The nature of human thinking is (1) to give more weight to the information that is consistent with the currently held belief and to ignore or discount discordant information, (2) to have better recall of or to give more weight to the unusual or the more recent than the

common or the more distant, and (3) to limit the search for additional information to that which has the potential of confirming rather than potentially refuting a belief (eg, selective necropsy to confirm a gross diagnosis). Prior belief biases subjective observation (such as occurs during the diagnostic process or during nonblinded measurement) because it subtly and unconsciously changes perception, particularly of vague or ambiguous characteristics. This bias occurs unbeknownst to the observer despite his or her best intentions and is the reason for many of the aspects of epidemiologic study design and execution such as blinding and randomization. By nature, humans tend to develop explanations from limited, incomplete information, whether in social relationships, from observations of the workings of nature, or when operating in a professional capacity as a veterinarian, often without considering the weaknesses in the information or the assumptions being made.

Bias (systematic error) is any effect at any stage of a process (thinking, observing) or study that produces results or conclusions that differ systematically from the truth. Bias can be reduced by critical thinking and proper study design and execution but not by increasing sample size (which only increases precision by reducing the opportunity for random chance deviation from the truth). The critical question is how likely the results are due to the presence of a large bias rather than the true state of nature, thus making the conclusions invalid. Observational study designs are inherently more susceptible to bias than experimental study designs, but well- designed and executed observational studies can provide more solid evidence than poorly designed and executed experimental studies.

Cognitive bias is the distortion of an individual's perception of his or her world that is due to common observational and reasoning errors. The study of cognitive bias is a relatively young but active area of research in psychology and cognitive science (the study of mind and intelligence). The findings have been applied to behavioral economics and to business and political decision making for 3 decades but only recently to medical decision making. Essentially, the scientific method is a process developed to minimize the effects of cognitive bias on the development of scientific knowledge.

Critical thinking is the disciplined ability and willingness to assess evidence and claims, to seek a breadth of contradicting and confirming information, to make objective judgments on the basis of well-supported reasons as a guide to belief and action, and to monitor one's thinking while doing so (metacognition). The thinking process that is appropriate for critical thinking depends on the knowledge domain (eg, scientific, mathematic, historical, anthropologic, economic, philosophical, moral), but the universal criteria are clarity, accuracy, precision, consistency, relevance, sound empiric evidence, good reasons, depth, breadth, and fairness. One of the more entertaining texts on critical thinking is that of Shick and Vaughn [7].

Some 80 different cognitive biases have been identified, several of which are subtle variations on a theme rather than distinct entities and some of

which are opposites. Awareness of these potential errors may reduce their impact. The major cognitive biases that apply to medical diagnostic decision making, selected and modified from Croskerry [8,9], are listed in alphabetical order in Box 2.

The clinician must be careful to detect situations in which the manager has reasoned from a primary problem to what he or she believes is the cause and then presents that conclusion as the primary complaint rather than the original problem. For example, a manager presented a complaint of poor barn ventilation after reasoning that the serious drop in milk production was caused by adult cow pneumonias and further that these pneumonias were caused by poor barn ventilation [10]. In this case, the actual cause of the production loss was the 22% underfeeding of grain due to a miscalibrated scale on a grain auger.

A serious and common error is to jump to generating hypotheses without first developing the quantitative information (the who, when, where counts) beyond vague clinical impressions (eg, these animals seem to be affected more than those) to support or refute a specific hypothesis. This approach is analogous to scattershot ordering of laboratory tests in diagnosing individual animal cases, hoping something will pop up rather than using the laboratory tests to rule in or out specific differentials, and will likely be as unrewarding. After the clinician has communicated a leading hypothesis about cause to a producer, it is difficult for both parties to return later to a more open frame of mind.

The clinician should be careful of "pseudoepidemics" caused by the onset of producer awareness of a more chronic problem or caused by a change in problem definition. For example, changing from detection of fetal loss by visual detection of a conceptus to detection of open cows by repeat palpation post early pregnancy diagnosis will cause a pseudoepidemic of fetal loss in virtually any dairy herd. In one case, a dairy producer invested almost $1,000 in laboratory fees in attempting to establish the etiologic cause of such a pseudoepidemic. The operation had recently switched to a dairy herd records program that classified any cow returning to heat after a positive pregnancy examination as an abortion in addition to those with visible signs of late gestation fetal losses.

The clinician should be aware of what the producer accepts as "normal" (endemic) occurrence. In one high-producing herd, the producer believed that third- or higher parity Holsteins going down with milk fever was a normal occurrence. Thus, he accepted most of his older cows going down with milk fever and did not recognize this situation as abnormal and warranting correction. The producer's veterinarian and nutritionist were not aware of the high incidence of milk fever on the farm. The clinician should be aware of the events that the producer is omitting because of assuming they are not related to the problem of concern. For example, a recent episode of late-term abortions that is not mentioned may be related to the more prevalent metritis problem that is being investigated. One should remember the old

Box 2. Cognitive biases that apply to medical diagnostic decision making

- Aggregate bias (ecologic fallacy, fallacy of composition): the tendency to substitute what is known to be happening in the relationship between group averages of two variables for what is unknown about these two variables at the individual animal level or vice versa. Group aggregate data (eg, bulk tank ship weights, milk composition, pen intakes) is often more readily available than individual data (eg, individual daily milk weights, individual milk composition, individual intakes).
- Anchoring bias (''jumping to conclusions''): the tendency to fixate on limited information too early in the diagnostic process.
- Ascertainment bias: the tendency to allow prior expectations to shape thinking and observation of information, particularly of subtle, vague clues.
- Association as causation fallacy (post hoc, ergo propter hoc fallacy): incorrectly assuming that one event caused another simply because the former was associated with and preceded the latter in a previous occurrence. Many superstitions such as having a lucky token for success in a sporting event are based on this fallacy.
- Availability bias: the tendency to judge things as more likely if they readily come to mind, which tend to be the more recent, the more prevalent, the more striking, or the more readily available.
- Confirmation bias: the strong tendency to look for further confirming evidence to support a weak diagnosis or hypothesis rather than looking for refuting evidence, which is logically more definitive. It is more powerful to ask, What would disprove this hypothesis if found? than to ask, What else would support this hypothesis if found?
- Diagnostic momentum bias: a weak diagnosis may gain momentum without gaining verification, particularly if it is communicated to others without the associated evidence and it induces cognitive bias into their reasoning and recall.
- Framing bias: biased thinking or memory recall that occurs due to being influenced by how the problem is stated, the question is asked, or the information is presented. For example, the question, When did your cow stop eating? is likely to elicit a different response from the 4-H'er than is the question, How is your cow's appetite?

- Fundamental attribution bias: the tendency to take excess credit for one's successes and to deflect responsibility for one's failures while attributing to others insufficient credit for their successes and excess responsibility for their failures.
- Hindsight bias: knowing the outcome profoundly alters interpretation of the events before the outcome, leading to underestimation (illusion of failure) or overestimation (illusion of control) of abilities. In hindsight, events appear to have fit together better and to be explained better than they did at the time.
- Multiple alternatives bias (paralysis by analysis, ''wallpaper phenomenon''): multiple options (eg, multiple differential diagnoses) multiply the conflict and uncertainty compared with fewer options, leading to paralysis of action and irrational decision making. Instead of comparing all competing options with each other, one should compare each with a common benchmark such as the status quo, starting with the more relevant or the more likely.
- Null feedback bias: failing to regularly obtain feedback, positive or negative, on the outcomes of previous work-ups and recommendations after the passage of time, and instead, concluding (in the absence of evidence) that the outcomes were successful.
- Order bias: information communicated at the beginning and at the end of an exchange is remembered better than the information communicated in the middle. This bias can be avoided by recording information during the communication process rather than relying on recall later.
- Overconfidence bias: the tendency to spend too little time gathering and synthesizing information before taking action because of placing too much faith in one's opinions and hunches. One should ask whether information been gathered in a logical, thorough, and logical fashion and whether this information supports one's opinion.
- Premature closure bias: the tendency to accept a diagnosis before it has been sufficiently verified by tests for adequacy, coherence, parsimony, and falsification.
- Satisfying bias: the tendency to stop searching for further information after something is found. The questions to ask are, Is there anything else to be found? and Did I look in the right places?
- Support bias: the tendency to judge a hypothesis that has more detailed information as being more likely.

- Sunk cost bias: the greater the investment of funds, time, and mental energy in a diagnosis, the greater the reluctance to let it go and consider other alternatives.
- Vertical thinking bias: the failure to think laterally or ''outside of the box,'' which is reduced by asking the question, What else might explain this?

aphorism, ''*More mistakes are made from not looking than from not knowing!*''

The clinician should be careful of a red herring. Often the tendency is to necropsy only a few animals; the necropsies are often incomplete (ie, a wide selection of tissues are not submitted from all major organ systems); and the necropsied animals are often not representative, even when large numbers of animals are dying. The mistaken tendency is to select the worst rather than the representative. When taking samples, the clinician should consider the ''regret'' factor versus current cost of attainment. For example, when working up a reproductive problem in a grazed beef herd, taking and holding blood samples from palpated animals only to discard them later when they are found not to be needed may be less expensive than finding that such samples are needed when new hypotheses emerge, necessitating rounding up and corralling the herd again.

The clinician should be careful of shortcuts. Failure to perform complete gross necropsies and to submit a full set of properly handled and preserved tissue samples from the major organ systems and instead submitting only those samples that would confirm a leading differential diagnosis is a common error. In a continuing problem in one large dairy herd, the underlying problem was believed to be a severe respiratory condition, but laboratory findings on the lung samples from partial necropsies were inconclusive. Complete necropsies performed on several euthanized cases revealed a severe uterine condition subsequent to improper but widely applied postparturient treatment. The potential regret cost of missing a major gross diagnosis or failure to obtain a laboratory diagnosis must be balanced against the marginal cost of a more complete necropsy compared with a partial.

The clinician should be careful of the overlooked. During an investigation of a large dairy herd, one third of the retained heifer calves were documented to be dying due to salmonellosis. The producer, however, did not recognize the magnitude of these losses because the calves were dying one by one and were being removed by the rendering service during their daily visits. Only by comparing the current young stock inventory on the farm with the calving events recorded on a calendar did the producer recognize the magnitude of this loss. In another large dairy, the manager knew that 10% of the cows calving during a 2-week period were clinically affected by a problem. When the records of all of the cows that had calved during

this period were reviewed, however, he was surprised to learn that all had been culled from the herd in the intervening period of 2 months. On larger operations, the discrepancies between what the management intends to happen and what employees do, and what or when major changes in procedures occurred as described by employees versus the management are sometimes amazing.

The clinician should remember the iceberg principle (ie, subclinical cases usually outnumber clinical cases several fold and thus represent the greatest loss) and the spectrum of disease (ie, incubating, clinical, and recovered). Overlooking these concepts in searching for clues can lead to comparing clinically affected animals with subclinically affected and recovered animals instead comparing definitely affected animals with definitely unaffected animals. Doing so may lead to erroneous conclusions about the factors involved in the problem, which defeats the strength of herd investigation (ie, comparison between groups of animals over time). More severely affected animals, however, may have experienced higher levels of a common risk factor than lesser or unaffected animals. Even in general herd problems (eg, low milk production), some individuals are affected more severely than others. For example, in a group of pregnant cattle, approximately 10% of all pregnancies diagnosed before 45 days of gestation are lost and approximately 20% of these losses are observed. The analysis and resolution of the problem will be confused by including cases that are due to other problems.

References

[1] Last J. A dictionary of epidemiology. 4th edition. New York: Oxford University Press; 2000.
[2] Thrusfield M. Veterinary epidemiology. 2nd edition. Malden (MA): Blackwell Science; 1995.
[3] Rothman KJ. Causes. Am J Epidemiol 1976;104:587–92.
[4] Hill AB. The environment and disease: association or causation? Proc R Soc Med 1965;58: 295–300.
[5] Evans AS. Causation and disease: the Henle-Koch postulates revisited. Yale J Biol Med 1976;49:175–95.
[6] Meadows DH. Dancing with systems. Whole Earth 2001;106:58–63.
[7] Schick T, Vaughn L. How to think about weird things: critical thinking for a new age. 3rd edition. McGraw-Hill; 2002.
[8] Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. Acad Med 2003;78:775–80.
[9] Croskerry P. Achieving quality in clinical decision making: cognitive strategies and detection of bias. Acad Emerg Med 2002;9:1184–204.
[10] Bradish SK. A practical approach to the diagnosis of a dairy herd problem: a case report. The Bovine Practitioner 1997;31:111–4.