Challenges in Estimating the Incidence of Disease in the Canine

DR. HARVEY MOHRENWEISER, Standard Schnauzer Club of America, member of Canine Health Committee

Health Committee

Health surveys usually focus on one of two questions. The first is the qualitative question – does a disease (or list of diseases) exist in the population of interest, usually a breed of dog. When a case is identified and diagnosis confirmed, the question is resolved; YES – the disease exists in the breed. Study complete! A new study may address the cause of the disease.

The issues are more complex if no cases are identified. The confidence in the "NO – the disease does not exist in the breed" is dependent upon the number of dogs studied and the extent to which they are a representative sampling of the population. The confidence level for the "NO" answer is low if only 10 dogs were enrolled, and they were all located in Nebraska and were less than a year old.

Other health surveys (epidemiology studies) focus on obtaining a quantitative estimate of the incidence or prevalence of diseases. Estimating the incidence of disease is a challenge as it requires accurately determining two numbers: the number of cases and the overall size of the population from which your samples are obtained.

It is important to note that the picture of health status and disease profile is very dependent upon the population studied. A study at the M. D. Anderson Cancer Center would yield one set of conclusions; very different conclusions would be derived from a study of patients at Kaiser Permanente clinics. Similarly, the picture of canine health from a study that focused on Banfield clinics for example, would differ from that obtained from a veterinary school clinic-based study.

Incidence estimates generally cannot be obtained from hospital/clinic-based studies because obtaining estimates of the size or characteristics of the population from which the enrolled dogs are sampled is not always feasible in hospital/clinic-based studies. Also, the population is not a random representation of the general population. Any sampled at a clinic (especially a vet school clinic) is more likely to be ill.

Two strategies are commonly employed for obtaining estimates of human disease incidence:

1. All cases in a defined geographical location of known population size and character are ascertained. This is the strategy employed by the National Cancer Institute for estimating cancer incidence and the Center for Disease Control for specified infectious diseases (where reporting is mandatory for these diseases); or
2. Individuals within a defined sampling frame (often a geographical area or a defined population) are randomly sampled and assigned to disease/non-disease status.

The challenge with the first strategy is obtaining (nearly) complete ascertainment of cases in the target population. The challenges with the second approach are insuring that the sampled individuals are representative of the total population and that the record of health is complete and accurate.

Obtaining a representative sample is a significant challenge when the population has substructuring (sampling from a non-homogeneous population, such as individuals of different gender, age, etc). This is very critical when the goal of the study is to compare the incidence of disease between different populations. The populations may be individuals residing in different areas, individuals with different lifestyles or jobs, or individuals of differing age or gender. In most epidemiology studies comparing disease incidence in different populations, one strives to enroll individuals with risk factors that may influence the incidence of disease (ethnicity, gender, smoking history, etc) in (nearly) equal numbers.

Population substructuring is the BIG challenge for estimating the incidence of disease in "the canine". Even when studying a single breed, substructure is an issue. Many breeders employ line-breeding strategies and still tend to use sires within a geographical area, even with the availability of artificial insemination. Thus, dogs in different regions of the country are not genetically identical. Also, husbandry practices and risk factors may differ in different areas of the country. These issues negatively impact the confidence in the conclusions derived from a study. Confidence in the data can be increased by sampling a larger fraction of the population and careful attention to the sampling strategy.

Except when focused on a specific breed, the sampled population will be a subset of a population of generally unknown size and character, and will include dogs from potentially over 160 defined breeds (in very unequal numbers) and a literally infinite number of mixed and random-bred breeds.

Among the challenges to obtaining estimates of disease incidence are:

1. Definition of and accurate determination of breed, including mixed and random heritage dogs – "looks like" will not be sufficient; are clinic records sufficiently accurate in determining breed?
2. Strategies for enrolling the number of dogs necessary for deriving statistically valid estimates of disease incidence in a population with extensive substructuring – (should a study be limited to the 10 or 20 most popular purebred breeds plus mixed and random breed dogs?);
3. Strategy for obtaining health data– what diseases are tracked, how is data obtained (Is clinical diagnosis

Cont'd on page 20

Cont'd from page 19

required for disease verification? How is complete history of health obtained? Obtaining age at and cause of death?); How is data on healthy dogs obtained?

Remember, the goal is an estimate of the incidence of disease, thus both numerator and denominator are required.

4. What demographic data (age, gender, spay/neuter status, etc) should be obtained;

5. Estimating size and characteristics of population sampled and the extent to which it is an adequate representation of the total population of canines;

6. How is consistency and accuracy of diagnostic criteria insured; what quality control is necessary to validate data;

7. In addition to the significant challenges at the study design and scientific level, challenges also exist at the operational level. The data storage and information management infrastructure required for a data set of this scale is substantial.

Conducting epidemiology studies requires a substantial investment of resources so that they (1) enroll a sufficient numbers of individuals, especially for heterogeneous populations such as purebred, mixed breed and random breed dogs, and (2) have a sufficiently accurate history of health and disease to derive valid conclusions. Studies that are under-resourced, under-funded and under-powered (fail to enroll sufficient individuals to obtain statistically significant results) rarely provide quality data. The absence of sufficient quality data seriously compromises confidence in the results obtained and conclusions derived. 🐾

This article has been approved by the Health Committee and it's reproduction is encouraged.

2011 AKC DELEGATES MEETINGS

March 7-8 (Monday, Tuesday): New York City area
June 12-13 (Sunday, Monday): Location to be determined
September 12-13 (Monday, Tuesday): New York City area
December 15-16 (Thursday, Friday): Orlando, Florida

Special thanks to Sylvia Arrowwood, issue coordinator for this issue, and to the members of the Perspectives committee. Your efforts are very appreciated.

As always, thanks and appreciation, to the authors who contributed to this issue. Without you, there would be no Perspectives.

Carol Williamson, *Editor*