Meeting 5: Sampling

Reading

- Dohoo pp 27 – 52 (sampling).
- Cameron A (1999) Survey Toolbox, available from http://www.ausvet.com.au/resources/LiveToolbox(en).pdf Part 1 Chapter 3 'Sampling' pp 37-47.
- Cannon, R and Roe, R 'Livestock Disease Surveys: A field Manual for Veterinarians', Australian Bureau of Animal Health, Canberra 1982 pp 2-9 (surveys and sampling methods)
- Baldock FC. (1998) What constitutes freedom from disease in livestock? AVJ 76: 544-545.
- Stevenson MA (2005). Introduction to Veterinary Epidemiology: 227.407 Study Guide. Massey University, Palmerston North, New Zealand. pp 53 – 64.

Presentations

1. Probability and non-probability sampling methods
2. Creating sampling frames and sampling techniques
3. Sources of error when sampling and strategies to reduce error
4. I'm designing a study - how many samples should I take?
5. Estimating population parameters on the basis of a sample

Review and comment on the sampling protocol to declare Belize free from classical swine fever at http://www.oirsa.org/Publicaciones/PREFIP/Publicacion-03/ProtocolodeMuestreo-01.htm

Exercises

1. Carefully read the following examples and then choose your preferred sampling unit. Explain your choice.
    1. You wish to estimate the economic losses arising from lameness in sheep in a region. You decide to conduct a survey to estimate the prevalence of lameness in sheep in the region.
    2. You wish to conduct a survey to determine the incidence of foot-and-mouth disease (FMD) outbreaks in pigs from an intensively farmed endemic region of a country during the last year.
    3. You believe that poor stockyards and overcrowding of cattle in abattoirs before slaughter is contributing to carcass bruising. You plan to conduct a study to investigate this proposed risk factor.

2. It is frequent in animal production systems to divide animals into separate groups. Dairy farms (for example) manage lactating and non-lactating stock as separate groups. These divisions can make the collection of a representative sample from a population difficult. Give three examples from livestock enterprises where free mixing of animals is prevented. List all of the sub-groups that may be present in each. How would you obtain a representative sample from each enterprise?
3. Suppose you wish to determine the prevalence of disease within the pig population of a region. Previous surveys have indicated that 70% of the region's pigs are located in very large, intensive specialised pig farms, 20% of pigs are found within smaller farming units (frequently as a secondary industry on large dairy farms), and 10% of pigs are kept singly within small plots around towns (by people whose major occupation is not farming). With proportional stratification, a sample would be selected at random from within each stratum such that the aggregated sample would consist of 70% pigs obtained from the large intensive farms, 20% pigs obtained from the smaller pig farms, and 10% pigs obtained from small plots near towns.

   Explain why it is important for each stratum of pigs to be represented in this sample for the prevalence survey.
4. Assume that the disease that you are investigating is leptospirosis: combine your knowledge of leptospirosis with the description of the farming systems. Is the epidemiology of leptospirosis likely to vary between the different strata?
5. It is decided to do a survey to estimate the prevalence of disease X in a population of cattle. Three experts are asked for their opinions about the expected prevalence and they reply: 75%, 50% and 25%. Assuming that there are 1 million head of cattle in the study area, a desired precision of 5% and a desired confidence level of 95%.

   Calculate the needed sample size according to the three expert opinions.

   When prevalence is unknown and you have absolutely no idea about its expected value, what prevalence estimate should you use for the sample size calculation?
6. Serological surveillance for a disease of poultry is to be conducted in a population of 15,000 villages. Each village contains between 10 and 2100 eligible birds. The mean number of birds per village is 750. The requirement is to be 95% certain of declaring a village positive for disease if the within-village prevalence is greater than or equal to 5% and the between-village prevalence is greater than or equal to 1%. If all birds were tested in sampled villages, how many villages would need to be sampled to achieve the required probability of detection?

Example examination questions

1. Using examples, write brief notes on sampling methods used to select participants in epidemiological studies (2001 written).
2. Briefly describe the essential features and application of stratified random sampling (2002 written).

Additional reading/resources

- Stevenson MA (2005). Sampling. In: 195.721 Analysis and interpretation of animal health data – study guide. Massey University, Palmerston North, New Zealand.
- Bennett S, Woods T, Liyanage W, Smith D (1991) A simplified general method for cluster-sample surveys of health in developing countries. World Health Statistics Quarterly 44: 98 - 106.
- Cameron A (1999) Survey Toolbox, available from http://www.ausvet.com.au/resources/LiveToolbox(en).pdf Part 1 Chapter 3 'Sampling' pp 48-81.
- Noordhuizen, J., Frankena, K., van der Hoofd, C., Graat, E., 1997, Application of Quantitative Methods in Veterinary Epidemiology. Wageningen Pers, Wageningen pp 31 - 62.
- Kelsey, J., Thompson, W., Evans, A., 1986, Methods in Observational Epidemiology. Oxford University Press, New York. pp 254 - 284.

Meeting 11: Questionnaires, data management, data analysis 1

Reading

- Thrusfield 2nd Edn, pp 191-198 (questionnaires); pp 151-166 (data management) or 3rd Edn, pp 188-204 (questionnaires and data management)
- Thrusfield 2nd Edn, pp 129-131 or 3rd Edn, pp 152-156 (data types);
- Thrusfield 2nd Edn, pp 167-171 or 3rd Edn, 214-216, 220-224 (descriptive statistics, including confidence intervals)
- Thrusfield 2nd Edn, pp 171-174 or 3rd Edn, 217-220 (statistical distributions);
- Thrusfield 2nd Edn, pp 174-177 or 3rd Edn, pp 224-227 (presenting numerical data)
- Dohoo et al, pp 53-63 (questionnaires), 581-586 (data management); pp 586-589 (data analysis)
- Petrie and Watson pp 153-156 (transformation)

Presentations

1. Design and conduct of questionnaires. What different types are there, and what are the strengths and weaknesses of each?
2. Methods of data management. Common data problems, what effect they can have on the interpretation of the data, and how to avoid them
3. Summarise the different types of data. For each type provide a definition and example.
4. Displaying data: measures of centrality and dispersions, confidence intervals.
5. Data distributions and transformations – When is transformation indicated, and how do you do it?

Exercises

1. Compare and contrast methodologies used for the following questionnaires, concentrating on the principles of design and conduct:

- Baldock FC et al. (2003). Estimated and predicted changes in the cat population of Australian households from 1979 to 2005. AVJ 81: 289-292.
- Buckley P et al. (2004). Owners' perceptions of the health and performance of Pony Club horses in Australia. Prev Vet Med 63: 121-133.
- Heller J et al. (2005). Snake envenomation in dogs in New South Wales. AVJ 83: 286-292.
- Wraight MD et al (2000) Compliance of Victorian diary farmers with current calf rearing recommendations for control of Johne's disease. Veterinary Microbiology 77:429-442.

Example examination questions

1. Using examples, write brief notes on data types (2000 written).

2. Write brief notes to demonstrate your understanding of when it is appropriate to use postal questionnaires to conduct epidemiological surveys and how to address the potential weaknesses of this method (1998 written).

3. Write brief notes using examples about paired versus independent data (1997 written).

Additional reading/resources

- Epidemiological Skills in Animal Health, PGFVS Proceedings 143; pp 151-158 (questionnaires), 215-222 (data types and descriptive statistics).
- Cameron A. et al. (2004) Data management for animal health. AusVet Animal Health Services, Brisbane, Australia.
- Salman MD et al. (1990). Data description. Journal of American Veterinary Medical Association 197: 36-38.
- Shott S (1990). Confidence intervals. Journal of American Veterinary Medical Association 197: 576-578.

Meeting 7: Bias, interaction

Reading

- Thrusfield 2nd Edn, pp 230-234 or 3rd Edn, pp 276-280 (bias)
- Thrusfield 2nd Edn pp 75-78, 229-230 or 3rd Edn, pp 92-97, 274-275 (interaction)
- Dohoo et al, pp 207-234 (bias); 235-253 (confounding); pp 253-258 (interaction)

Presentations

1. Bias – definition, types, causes and control options. Alternatively, this could be split into three separate presentations on selection bias, misclassification bias and confounding.
2. Interaction – definition, types and examples.

Exercises

1. You are interested in the effects of age on the occurrence of clinical parvovirus disease in dogs. You go through all practice records for the past 12 months and categorise all dogs on record as either having/not having clinical parvovirus disease over that period. You also record the age of each dog. All of the following factors are associated with age:
   - Breed
   - Sex
   - Location (a high proportion of older dogs are found in the north of your practice area)
   - Vaccination history
   - Number of visits to vet prior to diagnosis

   Which of these factors should be considered as potential confounders of the relationship between age and parvovirus disease and why/why not?
2. Relative to younger cows, older cows are at increased risk of milk fever close to and soon after calving. You wish to estimate the effect of production ability on milk fever incidence. In a cohort study in one large dairy herd, high producing cows (more than 6,000 L of milk produced last lactation) were compared to low producers (less than 4,000 L). Under which of the following circumstances is the observed relationship likely to be confounded by age? Why/why not?

   Study 1: Age was not considered when selecting high and low production groups. Consequently, cows of all ages are included in the study. As it turned out, there were a higher proportion of old cows in the high production group, relative to the low production group.

   Study 2: Only old cows were eligible for inclusion in the study.

   Study 3: Age was not considered when selecting high and low production groups. Consequently, cows of all ages were included in the study. On checking the results, you note that the average age of cows in the high production group was similar to that in the low production group.

3. In an observational study to identify risk factors for neonatal septicaemia in calves, a number of putative risk factors were measured on the day of birth. The outcome (occurrence of neonatal septicaemia) was then determined. Risk factors measured included age and breed of dam, calving ease, colostrum quality (gamma globulin concentration in colostrum), calf suckling vigour soon after birth and calf gamma globulin concentration in serum at 24 to 48 hours after birth (the optimal time to measure these).

   Only 2 factors were associated with neonatal septicaemia occurrence - suckling vigour (vigorous sucklers had a lower incidence) and calf gamma globulin concentration in serum (calves with low concentrations were at increased risk of neonatal septicaemia). Calf gamma globulin concentration was also associated with suckling vigour (vigorous sucklers had higher concentrations).

   You are interested in the relationship between suckling vigour and neonatal septicaemia. Is calf gamma globulin a potential confounder of this relationship?

4. Leptospirosis is contracted by dairy farmers in the dairy, primarily from exposure to urine of infected cows. Herd vaccination is thought to reduce the risk of exposure. This is based on experimental studies where the period of urinary shedding after challenge was much shorter in vaccinated animals relative to non-vaccinates. Supporting this, a cross-sectional study has reported a negative association between leptospirosis serological status of dairy farmers and herd vaccination status. The prevalence of seroconversion was lower in people milking vaccinated herds. Which of the following factors would you seek more information about before you agree with the researchers' conclusion that vaccination reduces the risk of seroconversion in people milking dairy herds? Why?

   - Use of splash guards in the dairy
   - Breed of the herd
   - Wearing of aprons in the dairy
   - Gender of the people milking the herd
   - Attitude and behaviour of people milking the herd to exposure to urine splash
   - Prevalence of urinary shedding amongst cows in the herd

5. For the scenarios listed below, under what conditions would you suspect selection bias? Why?

   Study 1: In the course of practice, dairy practitioners use bacteriology to confirm diagnoses for some cases of suspected salmonellosis in dairy cows. Some dairy practitioners are concerned about the effects of magnesium oxide feeding on susceptibility of cows to salmonellosis and it is possible that suspected cases are sampled for bacteriology more frequently if they are being fed magnesium oxide, relative to suspected cases not being fed magnesium oxide. A case-control study is conducted to investigate risk factors for salmonellosis in dairy cows. Bacteriologically confirmed cases of salmonellosis in dairy cows are compared with controls (the next dairy cow seen by the same vet). Magnesium oxide feeding is again identified as a risk factor for salmonellosis.

   Study 2: Many Western Victorian wool growers consider that ovine dermatophilosis (lumpy wool) is a common problem in weaners sired by rams from a prominent SA merino stud. The disease is readily recognised and accurately diagnosed by most producers. In a case-control study assessing effects of sire source on susceptibility to lumpy wool, case and control flocks (flocks with high and low prevalence of lumpy wool in weaners) are selected from respondents to a series of media items.

Study 3: Following release of the new teat spray “Udderwise”, there were extensive concerns amongst producers and vets that it had contributed to epidemics of clinical mastitis in Australian dairy herds. In an effort to investigate this putative association, you survey all of the practice's dairy clients asking about use of “Udderwise” and clinical mastitis frequency over the past month. Thirty four percent of clients respond, about half of which are using “Udderwise” and half are using other teat sprays. “Udderwise” respondents reported a substantially higher incidence of clinical mastitis than respondents using other teat sprays.

6. Which type of misclassification bias (differential or non-differential) is likely in the following studies?

   Study 1: In 11 dairy herds participating in a herd health program, farmers fail to detect and/or record 50% of cows affected by retained foetal membranes (RFM). However, where records show that RFM occurred, this is invariably true. Not-in-calf cows are identified using whole herd pregnancy testing. A case-control study is then conducted with each not-in-calf cow matched with 5 in-calf cows of similar age and calving date. Records are then examined for a series of potential risk factors for non-pregnancy including occurrence of RFM and other disorders at calving, milk production in early lactation, etc. Effects of RFM on not-in-calf rates are estimated and a strong association observed.

   Study 2: In a study looking at effects of dam Johne's disease (JD) status on progeny JD status amongst dairy cows, pathologists take particular care when histologically assessing tissue from the ileum and caecum from progeny of known-infected dams.

Example examination questions

1. Write brief notes to demonstrate your understanding of measurement error (misclassification) (2005 written).
2. Using examples, write brief notes on handling confounding at the design and analysis stages of a study (2005 written).
3. Using examples, write brief notes on selection bias in epidemiological studies (2002 written).
4. You have been asked to consider lamb mortalities on a property divided by a roadway into two blocks – one block is hilly and the other is flatter and more prone to flood. The owner believes that lambs raised on the flatter block are more likely to die than lambs raised on the hilly block. Because he believes them to be hardier, the owner tends to put more wethers on the flatter block than ewe lambs.

   Part A. The owner has a total of 100 lambs, evenly split between the two blocks. Of these, he has observed 15 dead on the flatter block and 10 dead on the hilly block. Using these data and a contingency table, calculate the relative risk of mortality for lambs born on the flatter (versus hilly) block.

   Part B. Explain how stratified analysis can be used to determine whether a third dichotomous variable (for example, sex of lamb) might confound or otherwise modify the effect of one dichotomous variable on another. Include in your answer, how you would use stratified analysis to delineate between confounding and effect modification. If only confounding is occurring, how might the data be re-analysed?

Part C. The owner now tells you that 30 of his 55 wethers weer sent to the flatter block. Of these, 6 subsequently died, whilst only 4 wethers from the hilly block died. He also tells you that 9 of the ewe lambs on the flatter block died. Use stratified analysis and the data above to determine whether the sex of the lambs is likely to be confounding or otherwise modifying the effect of block on lamb mortality (2003 written)

Additional reading/resources

- Epidemiological Skills in Animal Health, PGFVS Proceedings 143; pp 67-74 (confounding and interaction)
- Multicausality: Effect modification. Available: www.epidemiolog.net/evolving/Multicausality_EffectModification.pdf