Study design in Epidemiology

Kristen Reyher
Centre for Veterinary Epidemiological Research
Atlantic Veterinary College
University of Prince Edward Island
The theory behind study design

• We want to design studies to
  – Identify causal factors for disease, so that we can
    • Focus on target points to work towards prevention
    • Minimise harmful effects of treatments or management changes
Study types

Descriptive
- Case Report
- Case Series
- Survey

Explanatory (Analytical)
- Observational
- Experimental
  - Laboratory
  - Controlled Trials

Cross-sectional
- Cohort
- Case-control
Descriptive vs. Analytical

• Descriptive
  – Describe characteristics
  – Do not make comparisons
  • case report
  • case series
  • survey

• Analytical (explanatory)
  – Seek to make comparisons
    • inference about exposures (risk factors, treatments) and outcomes (disease, death, production)
  – Experimental vs. observational
Descriptive studies 1

• Case study
  – Report on one or a few cases
  – Usually a rare condition
  – Limited to ‘real world’ conditions?
  – Any conclusions about cause or outcome are author’s conjecture
Descriptive studies 2

• Case series
  – Describe (often unusual) clinical course of condition of interest
  – Might provide information about prognosis if cases are representative of all cases
  – Again, no direct data but features might help build hypotheses
Descriptive study 3

• Survey
  – Estimate the frequency and distribution of outcomes
  – Provides some data (say about disease in a population)
  – Need to take care re: sampling (Signe) and design of questions
  – Surveys including exposures and outcomes = cross-sectional analytic studies!
Analytical studies

- Experimental
  - Investigator can allocate study subjects
  - Advantages
    - stronger evidence of causation
    - control of confounders through randomisation
  - Disadvantages
    - limited range of hypotheses
    - may not be “do-able”
Analytical studies

• Observational
  – No allocation of study subjects
    • Do not confuse random sampling with random allocation!
    • Observation in a real-world setting
  – Advantage
    • Complex web of causation might not be otherwise reproducible
      – practically
      – ethically
      – economically
Observational studies

• Prospective vs. Retrospective
  – Has outcome occurred before study starts?
    • yes = retrospective
    • no = prospective
  – Advantage of prospective
    • data quality
    • better able to study incidence
Observational studies

- Classified by subject selection
  - Cross-sectional studies
  - Cohort studies
  - Case-control studies
Observational study 1

• **Cross-sectional**
  – Most frequent study design in vet epi = straightforward
  – Random sample of subjects from a population
    • Try to represent population in sample
  – Non-directional in time = ‘snapshot’
  – Simultaneously classify according to
    • Disease status (or outcome)
    • Study factor or risk factor
      – determinant
      – exposure
Cross-sectional

Study Population

Past  Present  Future

Time
Cross-sectional studies

• Limitations
  – Only suitable for chronic conditions occurring at a moderate level in the population
  – Only quantifies prevalence of exposure and outcome
    • May over-represent factors affecting incidence and duration
    • Can confuse protective risk factors
  – Reverse-causation
    • Best for time-invariant exposures (sex, breed, housing)
    • Can confuse procedures implemented in response to disease
Cross-sectional studies

• Example:
  – 100 dairy herds selected randomly from the Dairy Herd Improvement (DHI) register to answer a questionnaire assessing treatment of dry cows
  – Each farm is classified according to exposure (which dry cow therapy used) and according to outcome (>30% fresh cow mastitis or <30% fresh cow mastitis)
Cross-sectional

100 farms

- Used DryClox / >30% mastitis
- Used DryClox / <30% mastitis
- Used CefaDri / >30% mastitis
- Used DryClox / <30% mastitis

Past | Present | Future
Cross-sectional studies

• **Pros**
  – Representative of population
  – Potentially efficient
  – Low cost
  – Rapid

• **Cons**
  – Must verify that risk factor came before the disease
Observational study 2

• Cohort
  – Identify subjects
    • with exposure
    • without exposure
  – Follow the groups through time to determine if disease develops
    • usually prospective
Cohort

Time
Past  Present  Future

Exposed

D+
D-

Non-exposed

D+
D-
Cohort studies – special case

- Single cohort = longitudinal study
  - Follows an entire population through time
  - Record all exposures of interest
    - Investigate multiple exposures at once
  - Record all outcomes of interest
    - Outcomes must follow exposure!
  - Useful in measuring incidence of disease
Cohort Studies

• Example:
  – 500 healthy cows (D-) from farms using dry cow therapy (E+) were randomly selected from a list of cows in DHI databases across Canada
  – 500 healthy cows (D-) from farms not using dry cow therapy (E-) were also randomly selected from the same list of cows
  – Followed for two years
  – Assessed for clinical mastitis during that time
Cohort

- **Past**
  - Farms not using dry cow therapy
    - Mastitis
    - No mastitis

- **Present**
  - Farms using dry cow therapy
    - No mastitis
  - Farms not using dry cow therapy
    - Mastitis
    - No mastitis

- **Future**

**Time**

Past  Present  Future
Cohort studies

• Pros:
  – Less susceptible to bias compared to case-control
  – More control over quality of data
  – No confusion on time order of exposure and disease

• Cons:
  – Expensive
  – Time-consuming
  – Potential losses to follow-up
  – Only works for diseases common in a population
Observational study 3

• **Case-Control**
  – Identify subjects
    • with disease
    • without disease
  – Compare histories of risk factor (exposure)
    • Usually retrospective
Case-control

- D+
  - exposed
  - non-exposed
- D-
  - exposed
  - non-exposed

Time:
- Past
- Present
- Future
Case-control studies

- Used for rare diseases
- Relatively quick and inexpensive (if quality data is accessible)
Case-control studies

• Limitations
  – Finding source of cases
  – Defining a case
  – Appropriate controls are often difficult to identify. These should be animals that would have been cases if they got the disease (but not always as straightforward as that sounds!)
Case-control

• Example:
  – A rare mastitis is being studied.
  – 50 farms in Quebec has confirmed cases of this type of mastitis (D+)
  – 50 comparable farms in Quebec with no confirmed cases (D-) are also identified for the study
  – All 100 farmers are asked about management practices (type of dry cow therapies) used on their farms (exposure)
Case-control

50 farms with mastitis

Used CefaDri

Used DryClox

50 without mastitis

Used CefaDri

Used DryClox

Past

Present

Future

Time
Case-control studies

• **Pros:**
  – Rare diseases
  – Potentially efficient
  – Low cost
  – Potential for rapid completion

• **Cons:**
  – Highly susceptible to bias related to selection of controls
Summary of observational studies

<table>
<thead>
<tr>
<th>Study type</th>
<th>Cross-sectional</th>
<th>Cohort</th>
<th>Case-control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimal cost</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Short time (little to no follow-up)</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Control selection difficult</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Representative of population</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Good for rare disease</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Good for rare exposure</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Time sequence known</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>
Experimental studies

- Laboratory-based
- Randomised controlled trials
Experimental study 1

• Laboratory-based
  – Carried out under strictly controlled conditions
  – Investigator has almost complete control over experimental conditions
  – Evidence of association of exposure and outcome is the best evidence of causation
  – Relavance to ‘real-world’ conditions often doubtful
Experimental study 2

• Randomised controlled trials
  – Covered by Signe next!
## Characteristics of various study types

<table>
<thead>
<tr>
<th>Type of study</th>
<th>Level of difficulty</th>
<th>Level of investigator control</th>
<th>Strength of “proof” of causal association</th>
<th>Relevance to “real-world” situations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Descriptive</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case report</td>
<td>very easy</td>
<td>very low</td>
<td>n/a</td>
<td>low to high</td>
</tr>
<tr>
<td>Case series</td>
<td>easy</td>
<td>very low</td>
<td>n/a</td>
<td>low to high</td>
</tr>
<tr>
<td>Survey</td>
<td>moderate</td>
<td>moderate</td>
<td>n/a</td>
<td>high</td>
</tr>
<tr>
<td><strong>Explanatory - experimental</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laboratory</td>
<td>moderate</td>
<td>very high</td>
<td>very high</td>
<td>low</td>
</tr>
<tr>
<td>Controlled trial</td>
<td>moderate</td>
<td>high</td>
<td>very high</td>
<td>high</td>
</tr>
<tr>
<td><strong>Explanatory - observational</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cross-sectional</td>
<td>moderate</td>
<td>low</td>
<td>low</td>
<td>moderate</td>
</tr>
<tr>
<td>Cohort</td>
<td>difficult</td>
<td>high</td>
<td>high</td>
<td>high</td>
</tr>
<tr>
<td>Case-control</td>
<td>moderate</td>
<td>moderate</td>
<td>moderate</td>
<td>high</td>
</tr>
</tbody>
</table>

*n/a = not applicable*