The views and opinions expressed in this presentation are solely mine and do not represent the position or opinion of ISPE or any other institution.

No conflict of interest to declare.

Lecture material from previous years has provided inspiration for the preparation of this lecture. Thanks to my predecessors.

Thanks to Edeltraut Garbe, Miriam Sturkenboom, and Samy Suissa.

Study Designs – Analytic Studies

- Experimental (interventional)
  - Uncontrolled trials
  - Controlled trials
    - Randomized (RCT)
    - Quasi-randomized
    - Non-randomized

- Quasi-experimental (lacks full control)
  - Cohort
  - Case-Control

- Non-experimental (observational)
  - Case-only
  - Cross-sectional
  - Ecological
Why Conduct a Case-Control Study?

- Efficiency
- Speed
- Cost saving
- Validity
- Precision

- Time-varying exposures
- Multiple exposures
- Rare outcomes
- Risk ratio
- Incidence rate ratio

How are you?
As compared to what?

Case-Control Study: What is the Study Base?
The Source Population

- Primary source population
- “Secondary source population”
Jeanne Dieleman, Silvana Romio, Kari Johanson, Daniel Weibel, Jan Bonhoeffer, Miriam Sturkenboom, and the VAESCO-GBS Case-Control Study Group

<table>
<thead>
<tr>
<th>Country</th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>UK</td>
<td>GPRD</td>
<td>GPRD</td>
</tr>
<tr>
<td>Denmark</td>
<td>National Patient Register</td>
<td>Danish civil registration</td>
</tr>
<tr>
<td></td>
<td></td>
<td>system</td>
</tr>
<tr>
<td>Netherlands</td>
<td>Neurologists. Hospitals</td>
<td>General practice of case</td>
</tr>
<tr>
<td></td>
<td></td>
<td>patients</td>
</tr>
<tr>
<td>France</td>
<td>Neurologists. Hospitals</td>
<td>Trauma unit in same hospital</td>
</tr>
<tr>
<td>Sweden</td>
<td>Neurology assessment labs</td>
<td>Swedish national</td>
</tr>
<tr>
<td></td>
<td></td>
<td>population registry</td>
</tr>
</tbody>
</table>

Selection (Sampling) of Cases and Controls

Controls should be representative of the source population that gave rise to the cases.

Selection (sampling) of cases and controls should be independent of exposure. Otherwise → Selection bias

Index Date

In cases, the index date is the occurrence of the study outcome (we refer to it in exposure classification)

Exposure → Disease Onset → First Symptom → Diagnosis → Time

- Induction period
- Latent period
- (alternative definition in terms of measurable outcome)
Dopamine Agonists and the Risk of Cardiac-Valve Regurgitation

Outcome: Cardiac-valve regurgitation
- Insidious onset → Date of disease onset?
- Unspecific recording → Misclassification (false positives)
- Typical symptoms only/suspected diagnosis → Diagnosis confirmed?

**Background**
Case reports and echocardiographic studies suggest that the ergor-derived dopamine agonists pergoline and cabergoline, used in the treatment of Parkinson's disease and the restless legs syndrome, may increase the risk of cardiac-valve regurgitation.

**Exposure Time Window**
For exposure classification (in the study)

**Case Validation**
Review of records (blinded to exposure)

Cases should be true positives

11,417 patients included in the study cohort

81 with possible new valvular regurgitation

50 excluded
- 8 had myocardial infarction within previous 3 yr
- 2 had preexisting valvular heart disease
- 40 did not have diagnosis confirmed

31 case patients validated and grouped according to history of exposure to a dopamine agonist

**Diagram:**
- Exposure time window
- Index date
- Time
Exposure Risk Window → Exposure Time Window
referring to biological mechanism

The exposure risk window is the time period during which a drug puts the patient at risk for the outcome.
(It is the basis on which we define the study’s exposure time window)

Exposure Risk Window → Exposure Time Window
for exposure classification (in the study)

Index date

Time

Exposure time window

Scenario 1

Scenario 2

Scenario 3

Closed Cohort

Risk Ratio (RR) = \frac{\text{incidence proportion exposed}}{\text{incidence proportion unexposed}} = \frac{\text{a}(a+b)}{\text{c}(c+d)}

RR = 1 \Rightarrow \text{The "null" value of no association}

The risk ratio is easy to interpret ("x-fold increased risk")
Cumulative (Exclusive) Sampling

Outcome

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td>a</td>
<td>c</td>
</tr>
<tr>
<td>Controls</td>
<td>b</td>
<td>d</td>
</tr>
</tbody>
</table>

Odds Ratio (OR) = \( \frac{a/c}{b/d} = \frac{ad}{bc} \)

About the “rare disease” assumption:
If the outcome risk is less than 5% in both exposed and unexposed then
Odds Ratio (with exclusive sampling) ≈ Risk Ratio (from full cohort)

What does the case-control odds ratio estimate?
→ It’s all about the sampling!

<table>
<thead>
<tr>
<th>Sampling scheme</th>
<th>What does odds ratio estimate?</th>
<th>About the “rare disease” assumption</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cumulative (exclusive)</td>
<td>Odds ratio</td>
<td>If “rare disease” then odds ratio = risk ratio (from full cohort)</td>
</tr>
</tbody>
</table>

Case-Cohort Design
(Case-Cohort Sampling, Inclusive Sampling, Case-Base Sampling)

Risk ratio (from full cohort) ≈ \( \frac{a/d}{b/c} \)

If “rare disease” then
Risk ratio = exposure odds ratio

About the “Rare disease” assumption:
If risk is less than 5% in both exposed and unexposed then
Odds ratio (from entire cohort at start of follow-up) = Risk ratio (from full cohort)
Case-Cohort Design
(Case-Cohort Sampling, Inclusive Sampling, Case-Base Sampling)

Case-Cohort Sampling (Inclusive)
(from entire cohort at start of follow-up)

Odds ratio → Risk ratio
No rare disease assumption required

Random sample (sub-cohort) → Controls
Cases

- Hutchinson (1968), Kupper et al. (1975), Miettinen (1982), Prentice (1986)
- Can study multiple outcomes (case series) with one control group

What does the case-control odds ratio estimate?
→ It’s all about the sampling!

<table>
<thead>
<tr>
<th>Sampling scheme</th>
<th>What does odds ratio estimate?</th>
<th>About the “Rare disease” assumption</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cumulative (exclusive)</td>
<td>Odds ratio</td>
<td>If “rare disease” then odds ratio ≈ risk ratio (from full cohort)</td>
</tr>
<tr>
<td>Case-cohort (inclusive, case-base)</td>
<td>Risk ratio</td>
<td>NOT REQUIRED (odds ratio estimates risk ratio without assumption)</td>
</tr>
</tbody>
</table>

Longitudinal Patient Data

Examples for databases of electronic medical records from primary care:
- The Clinical Practice Research Datalink (CPRD), formerly known as the General Practice Research Database (GPRD)
- The Health Improvement Network (THIN)
- >550 general practices (gatekeepers) in the United Kingdom
- >10 mio. patients total (>70 mio. person-years) since 1987
- >3.5 mio. active patients

General practitioners record medical information on computers:
- Diagnoses, symptoms
- Procedures, laboratory tests
- Smoking, alcohol, BMI (height, weight)
- Prescriptions (strength, quantity, dosing)
- Free text
- Specialist letters (hospital discharges, referrals)
Epidemiological Study with Longitudinal Patient Data

Illustration

Risk Set Sampling
Time Axis: Calendar Time

Risk Set Sampling
Time Axis: Follow-Up Time

Same cohort as on previous slide
Figure 1. Percentages of All Newly Treated Parkinson’s Patients who Received Monotherapy Regimens as Initial Antiparkinsonian Treatment by Calendar Year – Age Category 50-64

(Schade and Sturkenboom, 2012)
Risk Set Sampling
Combination of Calendar Time and Follow-Up Time

- Risk set: same calendar year of cohort entry

Matching on time (here 1:2, i.e., randomly 2 controls among risk set)
Conditional logistic regression: Odds ratio → Incidence rate ratio
Nested Case-Control Design

- Mantel (1973), Thomas (1977), Prentice and Breslow (1978)
- Generally: Case-control study that is conducted in a fully enumerated (well-defined) cohort
- More strictly: Risk set (incidence density) sampling within a cohort
  - Context of proportional hazards model (Cox, 1972) → Control of time (by matching on time) → Incidence rate ratio
  - Steps:
    1. Define the cohort: time axis, entry, exit
    2. Select cases in the cohort
    3. Form risk set for each case
    4. Randomly select one or more controls among each risk set
  - Subjects can be selected as control more than once

What does the case-control odds ratio estimate?

→ It’s all about the sampling!

<table>
<thead>
<tr>
<th>Sampling scheme</th>
<th>What does odds ratio estimate?</th>
<th>About the “Rare disease” assumption</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cumulative (exclusive)</td>
<td>Odds ratio</td>
<td>If “rare disease” then odds ratio = risk ratio (from full cohort)</td>
</tr>
<tr>
<td>Case-cohort (inclusive, case-base)</td>
<td>Risk ratio</td>
<td>NOT REQUIRED (odds ratio estimates risk ratio without assumption)</td>
</tr>
<tr>
<td>Risk set (incidence density)</td>
<td>Incidence rate ratio</td>
<td>NOT REQUIRED (odds ratio estimates risk ratio without assumption)</td>
</tr>
</tbody>
</table>

Dopamine Agonists and the Risk of Cardiac-Valve Regurgitation

René Schade, M.D., Frank Andersohn, M.D., Samy Suisse, Ph.D., Wilhelm Havenkamp, M.D., Ph.D., and Edeltraut Garbe, M.D., Ph.D.

**Methods**

We used data from the United Kingdom General Practice Research Database to identify a population-based cohort comprising 11,417 subjects 40 to 80 years of age who were prescribed antiparkinsonian drugs between 1988 and 2005. We conducted a nested case-control analysis within this cohort in which each patient with newly diagnosed cardiac-valve regurgitation was matched with up to 25 control subjects from the cohort, according to age, sex, and year of entry into the cohort. Incidence-rate ratios for cardiac-valve regurgitation with the use of different dopamine agonists were estimated by conditional logistic-regression analysis.
Power and Precision

Number of cases is fixed, while control-to-case ratio can vary

- Gain in power is substantial up to four controls per case; thus, 4-to-1 ratio appropriate in the majority of instances
- Ratio of 10 or more controls per case can be appropriate if (Suissa, 2006):
  - Exposure is infrequent
  - Hypothesized relative risk moved further from the null value
  - Several factors or other drugs assessed simultaneously

Want more? → Multitime case-control design (Suissa et al., 2010)

Case-Crossover Design

- Case-only design → Cases serve as their own controls
- Removes confounding by time-invariant factors
- Prerequisites:
  - Transient exposure (intermittent)
  - Stable exposure prevalence over time
  - Acute outcome
  - Time-window of effect (risk period) determined
- Extension: Case-Time-Control Design (Suissa, 1995)
There is no cookbook recipe for a good epidemiological study. Know your field!

Modern Case-Control Studies

René Schade
Pharmacoepidemiology Unit, Dept. of Medical Informatics
Erasmus University Medical Center
Rotterdam, The Netherlands

r.schade@erasmusmc.nl