Case control studies

Giuseppe Traversa
Case-control studies

- Do not be mislead by the word “control”
  - All etiologic studies have “control” groups
  - The comparison is always aimed at contrasting the frequency of the event of interest in a group of subject presenting the exposure/treatment of interest and the group without the exposure/treatment (or with an alternative exposure/treatment)
The analogy with cohort studies

- A case-control study should be conceived as a cohort study in which:
  - “cases” represent the events (outcome) of interest that would have occurred in a hypothetical cohort, i.e. the numerator of the incidence rates
  - “controls” represent a sample of the source population of cases, i.e. the denominator of incidence rates (subjects or person-time)
Case-control studies

<table>
<thead>
<tr>
<th></th>
<th>Exposed</th>
<th>Not exposed</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cases (or events)</strong></td>
<td>$a_1$</td>
<td>$b_0$</td>
</tr>
<tr>
<td><strong>Controls</strong></td>
<td>$n_1$</td>
<td>$n_0$</td>
</tr>
</tbody>
</table>

Odds Ratio = \( \frac{a_1}{n_1} / \frac{b_0}{n_0} = \frac{a_1 \cdot n_0}{b_0 \cdot n_1} \)
Case-control studies: 
the odds ratio

- The **measure of effect** estimated in case-control studies is the **odds ratio** (OR)
  - The interpretation is the same as for the relative risk and the rate ratio: it indicates how greater/lower the incidence of events is among exposed in comparison with the unexposed subjects (e.g. users of drug A vs non-users or users of drug B)

- **Odds**: ratio between the probability that the event of interest occurs to the probability that it does not
Case-control studies: the question

- Did subjects with the event of interest (cases) experienced something unusual in comparison with those who were similar to the cases but did not develop the event (controls)? e.g.
  - Tobacco smoking and lung cancer
  - Vaccination X and Guillain-Barré syndrome
  - Postmenopausal estrogen use and endometrial cancer
  - Anorexiant drugs and Primary Pulmonary Hypertension
  - Oral contraceptive and thromboembolic events
Case-control studies: when fit best?

- Case-control studies are particularly indicated to assess a potential causal relation in which the **event** is:
  - **rare** and/or
  - occur after a **long induction period**

- Example: in utero exposure to diethilstilbestrol and occurrence of adenocarcinoma of the vagina after almost 20 years

- As for a cohort study the investigated relationship is always “exposure ------> event”
Case-control studies: why a sample of the population of origin?

- Data collection can be very expensive, e.g.: 
  - Interview
  - Review of clinical records (e.g. ascertainment of confounding factors), history of occupational exposure …
  - Biological samples

- There is no need to acquire information on the entire population
From cohort to case-control studies:
an example from Jick et al, Lancet 1995:1589-1593

Risk of idiopathic cardiovascular death and nonfatal venous thromboembolism in women using oral contraceptives with differing progestagen components

Hershel Jick, Susan S Jick, Victor Gurewich, Marian Wald Myers, Catherine Vasilakis

The question:

- Do third generation Oral Contraceptives (OC) carry an additional risk of thromboembolic events (TE) in comparison with second generation ones?
From cohort to case-control studies:
an example from Jick et al, Lancet 1995:1589-1593

Three hypothetical studies, from the cohort to the case-control study

A. The hypothetical study conducted in the entire population of the region

B. The hypothetical study conducted in a sample of the denominator of the population of the region

C. The hypothetical case-control study
From cohort to case-control studies:
an example from Jick et al, Lancet 1995:1589-1593

- Assume that a population of 4 million inhabitants is followed up during several years. Also assume to know the use of oral contraceptive (OC) and the hospitalisations for thromboembolism (TE)

- Around 300,000 women, age 18-40 years, have used OC. In terms of duration of use:
  - 143,000 person-years of 2\textsuperscript{nd} gen OC
  - 180,000 person-years of 3\textsuperscript{rd} gen OC

- 75 hospitalisations for thromboembolism occurred:
  - 23 during the use of 2\textsuperscript{nd} gen OC
  - 52 during the use of 3\textsuperscript{rd} gen OC
A) The results in the cohort study

<table>
<thead>
<tr>
<th>Contraceptives</th>
<th>3&lt;sup&gt;rd&lt;/sup&gt;</th>
<th>2&lt;sup&gt;nd&lt;/sup&gt;</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Events (cases TE)</td>
<td>52</td>
<td>23</td>
<td>75</td>
</tr>
<tr>
<td>Person time (years)</td>
<td>180,000</td>
<td>143,000</td>
<td>323,000</td>
</tr>
</tbody>
</table>

\[
I_{oc3} = \frac{52}{180,000} = 28.9/100,000 \text{ person-years}
\]

\[
I_{oc2} = \frac{23}{143,000} = 16.1/100,000 \text{ person-years}
\]

\[
RR = \frac{I_{oc3}}{I_{oc2}} = \frac{28.9}{100,000} : \frac{16.1}{100,000} = 1.8
\]
B) The results of the study with a sample of the denominator (of the cohort)

The estimate of the RR would have been the same (apart from the statistical precision) had we included in the study:

- All women hospitalised for TE
- A sample of the denominator, the source population of our events, let say, 323 women (1% of the users of oral contraceptives)
B) The results in the study with a sample of the denominator

<table>
<thead>
<tr>
<th></th>
<th>Contraceptives</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3rd</td>
</tr>
<tr>
<td>Events (cases TE)</td>
<td>52</td>
</tr>
<tr>
<td>Person time (years)</td>
<td>180</td>
</tr>
</tbody>
</table>

\[
I_{oc3} = \frac{52}{(180 \times 1,000)} = 28.9/100,000 \text{ person-years}
\]

\[
I_{oc2} = \frac{23}{(143 \times 1,000)} = 16.1/100,000 \text{ person-years}
\]

\[
RR = \frac{I_{oc3}}{I_{oc2}} = \frac{28.9}{100,000} : \frac{16.1}{100,000} = 1.8
\]
C) The results of the study when the cohort of origin is not identified

- Assume we have included in the study all events of hospitalisation for TE, but it is unknown:
  - the dimension of the source population of users of 2\textsuperscript{nd} and 3\textsuperscript{rd} generation OC

- We may interview, with regard to previous use of OC:
  - Women hospitalised for TE (cases)
  - A sample of the source population (controls), i.e. users of oral contraceptives who, had developed the event, would have been included as cases
C) The results of the study when the cohort of origin is not identified

<table>
<thead>
<tr>
<th>Events (cases TE)</th>
<th>3rd</th>
<th>2nd</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>52</td>
<td>23</td>
<td>75</td>
</tr>
</tbody>
</table>

| Person time (years) | 180 | 143 | 323 |

\[
\text{I}_{oc3} = \frac{52}{180 \times \text{?}}
\]

\[
\text{I}_{oc2} = \frac{23}{143 \times \text{?}}
\]

\[
\frac{\text{I}_{oc3}/\text{I}_{oc2}}{\text{OR}} = \frac{52/180 \times \text{?}}{23/143 \times \text{?}} = \frac{52}{23} : \frac{180}{143} = 1,8
\]
Case-control design: the case (subjects)

**Identification and selection of case subjects**

- A case-control study can be carried out on any population of patients. As for other study designs:
  - Definition of the event/disease of interest
  - Validity of the diagnosis
  - Reliability/Reproducibility of the diagnosis
Case-control design: the control (subjects)

Control group: who? what they represent?

- They represent subjects who, had developed the disease would have been included as cases: the population that gave origin to the cases.
- The control group provides the estimate of the relative magnitude of the exposure (eg, drug A-drug B; exposed-unexposed) in the source population.
- The selection of the control group should be independent from exposure status.
- Possible control groups: general population, hospitalisations etc.
Nested case-control studies

- All case-control studies should be considered as nested in a cohort
- “Nested” (generally) refers to situations in which all subjects in the source population (of cases) are identified
Nested case-control studies

- For each case-subject, one or more controls are selected among subjects who have not developed the event by the time it occurred in the case
  - The risk-set of the case subject: subjects who are at risk of becoming cases
  - Cases occurring later in the follow up are eligible to become controls for earlier cases
  - It is possible to estimate the rate ratio (incidence density sampling)
**Nested case-control studies**

**Figure 1–21** Nested case-control study in which the controls are selected at each time when a case occurs (incidence density sampling). Cases are represented by “D” boxes. Broken diagonal lines with arrows represent losses to follow-up.

*Szklo & Nieto. Epidemiology. Beyond the basics.*
GANGLIOSIDES AND GUILLAIN-BARRÉ SYNDROME

R. RASCHETTI, M. MAGGINI, P. POPOLI, B. CAFFARI, R. DA CAS, F. MERRITT-IPOPOLO, S. SPILA-ALEGIANI, AND G. TRAVERSA

1Department of Epidemiology and Biostatistics, and 2Department of Pharmacology, Istituto Superiore di Sanità, Rome, Italy.

(Received in revised form 8 December 1994)

Abstract—Cases of Guillain-Barré syndrome (GBS) associated with parenteral use of gangliosides have been reported in several European countries. To evaluate the hypothesis of association between ganglioside exposure and occurrence of GBS, a case-control study was conducted. GBS cases discharged during 1989 from public and private hospitals in three Italian provinces were identified: 42 GBS cases and 420 controls matched on age and gender were enrolled. Data of onset of symptoms of GBS was taken from clinical records. Exposure status of subjects was ascertained through the regional computerized drug prescription monitoring system. The odds ratio of association between ganglioside use, in the 30 days prior to onset of symptoms, and GBS was 9.1 (95% confidence interval 2.8–29.4). Although there are formidable difficulties in distinguishing prodromal therapy of GBS from drug causation, the association with ganglioside therapy is strong and supportive of the hypothesis of a role of ganglioside preparations in the occurrence of GBS.
GANGLIOSIDES AND GUILLAIN-BARRÉ SYNDROME

- Cases: residents of Rome who were hospitalised (in 1989) with a diagnosis of GBS
- Controls: for each case, 10 residents of the area of Rome were selected matched by age, sex
- Exposure: prescription of gangliosides in the 2 months preceding the onset of GBS symptoms (for cases and matched controls)
- Exposure information: NHS monitoring system
- Confounding factors: concomitant prescriptions; previous prescriptions
APPETITE-SUPPRESSANT DRUGS AND THE RISK OF PRIMARY PULMONARY HYPERTENSION

LUCIEN ABENHAIM, M.D., YOLA MORIDE, Ph.D., FRANÇOIS BRENOT, M.D.*, STUART RICH, M.D., JACQUES BENOCHOU, M.D., XAVIER KURZ, M.D., TIM HIGEBOTTAM, M.D., CELIA OAKLEY, M.D., EMIL WOUTERS, M.D., MICHEL AUBIER, M.D., GÉRALD SIMONNEAU, M.D., AND BERNARD BÉGAUD, M.D., FOR THE INTERNATIONAL PRIMARY PULMONARY HYPERTENSION STUDY GROUP†

ABSTRACT

Background Recently in France, primary pulmonary hypertension developed in a cluster of patients exposed to derivatives of fenfluramine in appetite suppressants (anorexic agents), which are used for weight control. We investigated the potential role of anorexic agents and other suspected risk factors for primary pulmonary hypertension.

Methods In a case–control study, we assessed 95 patients with primary pulmonary hypertension from 35 centers in France, Belgium, the United Kingdom, and the Netherlands and 355 controls recruited from general practices and matched to the patients’ sex and age.

Results The use of anorexic drugs (mainly derivatives of fenfluramine) was associated with an increased risk of primary pulmonary hypertension (odds ratio with any anorexic-drug use, 6.3; 95 percent confidence interval, 3.0 to 13.2). For the use of anorexic agents in the preceding year, the odds ratio was 10.1 (95 percent confidence interval, 3.4 to 29.9). When anorexic drugs were used for a total of more than three months, the odds ratio was 23.1 (95 percent confidence interval, 6.9 to 77.7). We also confirmed an association with several previously identified risk factors: a family history of pulmonary hypertension, infection with the human immunodeficiency virus, cirrhosis, and use of cocaine or intravenous drugs.

Conclusions The use of anorexic drugs was associated with the development of primary pulmonary hypertension. Active surveillance for this disease should be considered, particularly since the use of anorexic drugs is expected to increase in the near future. (N Engl J Med 1996;335:609-16.)

©1996, Massachusetts Medical Society.

PRIMARY pulmonary hypertension is a rare, often fatal disease that tends to occur with particular frequency in women during their third or fourth decade.1,2 The factors leading to its development remain enigmatic. The occurrence of familial primary pulmonary hypertension suggests a genetic susceptibility.3 Reports have also suggested that portal hypertension4,5 and recent pregnancy6 may have causative roles. Exogenous factors have been suspected as well, including cocaine use,7 infection with the human immunodeficiency virus (HIV),8 oral contraceptives,9,10 and the use of anorexic agents.11,12 In the 1960s, there was an epidemic of primary pulmonary hypertension in Switzerland, Germany, and Austria in association with a particular anorexic agent, aminorex fumarate.11 In the early 1990s, French investigators reported a cluster of cases among patients who had used derivatives of fenfluramine.13 Dextfenfluramine,
- Cases: patients hospitalised for PPH (between Sept 1992 - Sept 1994); 220 centers in 4 countries
- Controls: for each case, 4 controls (possibly same GP) matched by age, sex, number of visits
- Exposure: use of anorexant drugs before onset of symptoms (for cases and matched controls), also categorised in relation to recency of use
- Exposure information: interview
- Confounding factors: weight (and history of loosing weight) concomitant and previous use of medicines, drug addiction, pregnancy ...
### Table 3. Use of Appetite Suppressants and Adjusted Odds Ratios for the Risk of Primary Pulmonary Hypertension.

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>CASE PATIENTS (N = 95)</th>
<th>CONTROLS (N = 355)</th>
<th>ADJUSTED ODDS RATIO (95% CI)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definite use of appetite</td>
<td>30 (31.6)</td>
<td>26 (7.3)</td>
<td>6.3 (3.0–13.2)</td>
</tr>
<tr>
<td>suppressants</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration of use</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤3 mo</td>
<td>7 (7.4)</td>
<td>12 (3.4)</td>
<td>1.8 (0.5–5.7)</td>
</tr>
<tr>
<td>&gt;3 mo</td>
<td>18 (19.0)</td>
<td>5 (1.4)</td>
<td>23.1 (6.9–77.7)</td>
</tr>
<tr>
<td>Indeterminate</td>
<td>5 (5.3)</td>
<td>9 (2.5)</td>
<td>2.6 (0.5–12.6)</td>
</tr>
<tr>
<td>Products reported as used†</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dextfenfluramine</td>
<td>18 (18.9)</td>
<td>22 (6.2)</td>
<td>—</td>
</tr>
<tr>
<td>Fenfluramine</td>
<td>6 (6.3)</td>
<td>4 (1.1)</td>
<td>—</td>
</tr>
<tr>
<td>Diethylpropion</td>
<td>3 (3.2)</td>
<td>2 (0.6)</td>
<td>—</td>
</tr>
<tr>
<td>Clofazoxone</td>
<td>3 (3.2)</td>
<td>6 (1.7)</td>
<td>—</td>
</tr>
<tr>
<td>Fenproporex</td>
<td>2 (2.1)</td>
<td>1 (0.3)</td>
<td>—</td>
</tr>
<tr>
<td>Phenmetrazine</td>
<td>2 (2.1)</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>Compounds</td>
<td>7 (7.4)</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>Possible use</td>
<td>3 (3.2)</td>
<td>2 (0.6)</td>
<td>—</td>
</tr>
<tr>
<td>Use after index date</td>
<td>3 (3.2)</td>
<td>17 (4.8)</td>
<td>—</td>
</tr>
<tr>
<td>Timing of use‡</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recent</td>
<td>14 (21.5)</td>
<td>7 (3.0)</td>
<td>10.1 (3.4–29.9)</td>
</tr>
<tr>
<td>Past</td>
<td>7 (10.8)</td>
<td>14 (6.0)</td>
<td>2.4 (0.7–8.2)</td>
</tr>
</tbody>
</table>
Case-cohort studies

- Each person in the population (cohort) of origin of the events of interest has the same probability of being selected in the control group.
- All subjects are classified with regard to exposure (e.g., exposed-unexposed) at the beginning of their risk period.
- It is possible to estimate the ratio of cumulative incidences (relative risk).
Case-cohort studies

Figure 1–20 Case-control study in which the controls are selected from the baseline cohort (case-cohort study). Cases are represented by “D” boxes. Broken diagonal lines with arrows represent losses to follow-up.

Szklo & Nieto. Epidemiology. Beyond the basics.
Innate left handedness and risk of breast cancer: case-cohort study

Made K Ramadhani, Sjoerd G Elias, Paulus A H van Noord, Diederick E Grobbee, Petra H M Peeters, Cuno S P M Uiterwaal

Among the proposed origins of breast cancer are ascertainment of vital status until 1 January 2000 in a

Among the proposed origins of breast cancer are intrauterine influences, such as exposure to sex hormones.¹ Such exposure may also influence cerebral lateralisation, with hand preference being one of its manifestations. We know only of case-control studies

Participants, methods, and results

In a breast cancer screening study in Utrecht, the Netherlands, 12 178 women born between 1932 and 1941 and recruited between 1982 and 1985 (participation rate 40%) had baseline questionnaire

The random sample comprised 165 (11.6%) left handed women. Mean age at baseline was similar for left and non-left handed women (47.4 (range 41.6-53.1) and 47.0 (41.0-53.1) years respectively). These groups did not differ in anthropometry, socioeconomic status, smoking habits, family history of

Follow-up for adequate information about the person years lived for all 12 178 women would have been costly and time consuming, so we
Control selection and rare disease assumption

Which strategy is used to sample the control series?

- Rare disease assumption
  - If sampling occurs among subjects still free of the disease/event of interest at the end of follow up, the odds ratio will overestimate the rate ratio (for positive exposure-events relation)

- No need for rare disease assumption
  - Case-cohort design: estimate of the relative risk
  - Density sampling: estimate of the rate ratio
Control selection and rare disease assumption

Figure 1–18 Hypothetical case-based case-control study, assuming that cases and controls are selected from a hypothetical cohort, as in Figure 1–13. The case group is assumed to include all cases that occurred in that hypothetical cohort up to the time when the study is conducted (“D” with horizontal arrows ending at the “case” bar): that is, they are assumed to be all alive and available to participate in the study; controls are selected from among those without the disease of interest (noncases) at the time when the cases are identified and assembled. Broken diagonal lines with arrows represent losses to follow-up.

Szklo & Nieto. Epidemiology. Beyond the basics.
Case-control design: the exposure

Exposure ascertainment: as in a cohort study

- Should precede the occurrence of the event (including preliminary symptoms)
- Same quality for cases and controls (e.g. recall bias)
- Not only exposed-unexposed: eg, in a study on ADRs consider dose, duration, previous use, concomitant use of other drugs, potential induction period etc.
Case-control studies: synonym for retrospective studies?

- Sometimes, prospective is considered as synonym for cohort studies as opposed to retrospective case-control studies.
- However we can have retrospective cohort and prospective case-control studies.
- It is preferable to use “prospective/retrospective” to describe the timing of the study in relationship with the occurrence of the events.
Cohort and case-control studies

Cohort study

- Complete source population (denominator) experience
- Can calculate incidence rates or risks and their differences and ratios
- Usually very expensive
- Convenient for studying many diseases
- Can be prospective or retrospective

Case-control study

- Sampling from source population
- Can usually calculate only the ratio of incidence rates or risks
- Usually less expensive
- Convenient for studying many exposures
- Can be prospective or retrospective

KJ Rothman. Epidemiology – An Introduction.