# Case-control studies

Madhukar Pai, MD, PhD
Professor
McGill University
Montreal
madhukar.pai@mcgill.ca

McGill

# Acute-onset smell and taste disorders in the context of COVID-19: a pilot multicentre polymerase chain reaction based case–control study

Á. Beltrán-Corbellini[a], J. L. Chico-García[a], J. Martínez-Poles[b,c], F. Rodríguez-Jorge[a], E. Natera-Villalba[a], J. Gómez-Corral[a], A. Gómez-López[a], E. Monreal[a], P. Parra-Díaz[a], J. L. Cortés-Cuevas[d], J. C. Galán[d], C. Fragola-Arnau[e], J. Porta-Etessam[c,f], J. Masjuan[a] and A. Alonso-Cánovas[a]

[a]*Department of Neurology, University Hospital Ramón y Cajal, Madrid;* [b]*Department of Neurology, Hospital La Luz, Madrid;* [c]*Neurological Disorders Study Group, Institute for Sanitary and Biomedical Research Clínico San Carlos (IdISSC), University Hospital Clínico San Carlos, Madrid;* [d]*Department of Clinical Microbiology, University Hospital Ramón y Cajal, Madrid;* [e]*Department of Otorhinolaryngology, University Hospital Ramón y Cajal, Madrid; and* [f]*Department of Neurology, University Hospital Clínico San Carlos, Madrid, Spain*

- Aim: to determine whether new-onset STDs are more frequent amongst COVID-19 patients than influenza patients.
- This was a case–control study including hospitalized patients of two tertiary care centres.
- Cases: 79 consecutive patients positive for COVID-19 PCR
- Controls: 40 patients positive for influenza PCR
- Exposure: A self-reported smell and taste disorders questionnaire.

# Results

| | Case | Control |
|---|---|---|
| STD | 31 | 5 |
| No STD | 48 | 35 |

OR= (31/48) / (5/35) = 4.5

$$\text{OR} = \frac{\text{Odds of STD among cases}}{\text{Odds of STD among controls}}$$

# Introduction to case control designs

- The case control (case-referent) design is really an efficient sampling technique for measuring exposure-disease associations in a cohort that is being followed up or “study base”
- All case-control studies are done within some cohort (defined or not)



Miettinen OS. Eur J Epidemiol. 2010 Oct;25(10):671-5

# Incident cases occur in a study base

Study base is the aggregate of population-time in a defined study Population's movement over a defined span of time [OS Miettinen, 2007]

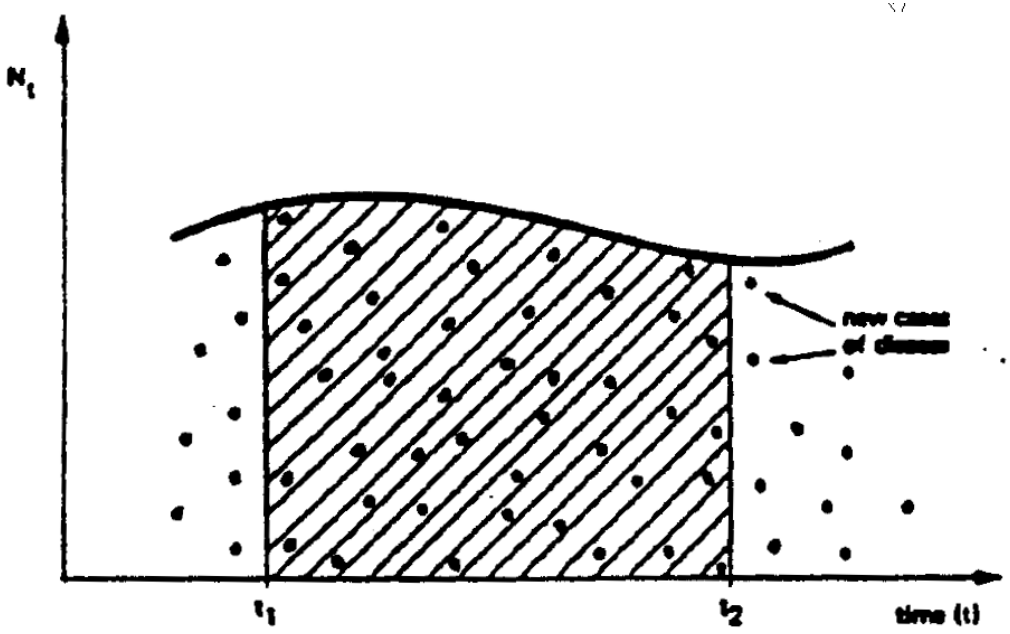



FIGURE 2 *Graphical illustration of the occurrence of new (incident) cases over time in a candidate population (of size $N_t$ at time t)*

Cohort and case-control designs differ in the way cases and the study base are sampled to estimate the incidence density ratio



Morgenstern IJE 1980

# Cohort design

Total cohort (total person-time experience)

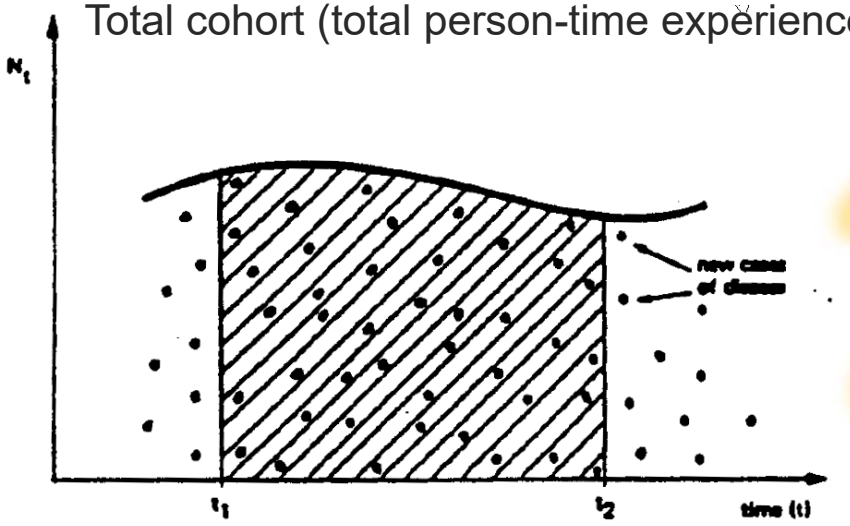
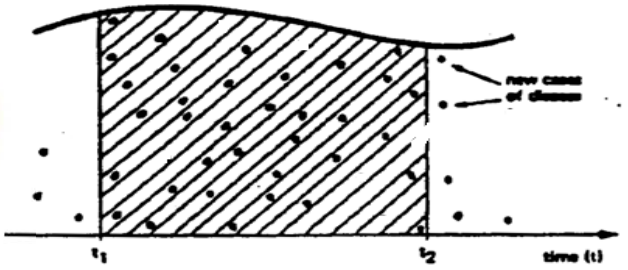
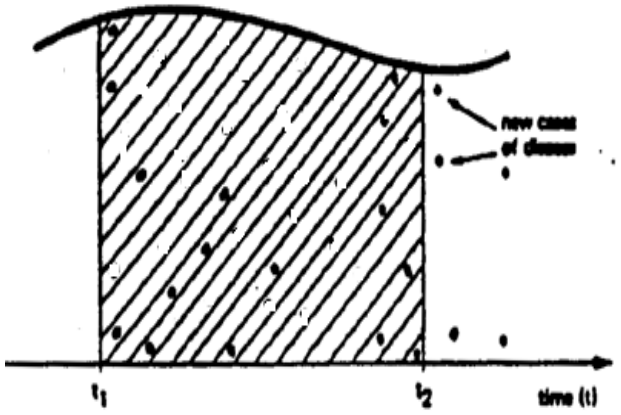



**FIGURE 2** *Graphical illustration of the occurrence of new (incident) cases over time in a candidate population (of size $N_t$ at time t)*




New cases / exposed person-time




New cases / unexposed person-time

$IDR = I_{exp} / I_{unexp}$

The IDR is directly estimated by dividing the incidence density in the exposed person-time by the incidence density in the unexposed person-time

# Case-control design: a more efficient way of estimating the IDR

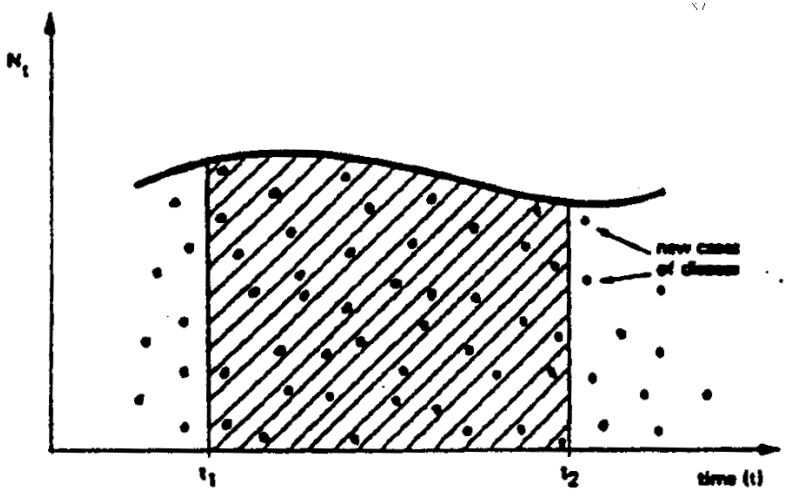



FIGURE 2 *Graphical illustration of the occurrence of new (incident) cases over time in a candidate population (of size $N_t$ at time t)*

* ** **
** ***
*** * *

Obtain all or a fair sample of the cases ("case series") that occur in the study base; estimate odds of exposure in case series

Obtain a fair sample of the study base itself ("base series"); estimate the odds of exposure in the study base [what fraction of the total person-time is exposed? this is a reflection of overall exposure distribution in the source population]

OR = IDR

The exposure odds in the case series is divided by the exposure odds in the base series to estimate the OR, which can be directly interpreted as the IDR (no assumptions required)

# Cases & controls must come from the same study base

- Controls must be an unbiased sample of the study base that generated the cases

# Study base: clear

Published by Oxford University Press on behalf of the International Epidemiological Association *International Journal of Epidemiology* 2006;**35**:1514–1521

doi:10.1093/ije/dyl197

## Risk factors for skin cancers: a nested case–control study within the Nurses' Health Study

Jiali Han,[1,3]* Graham A Colditz[1,2] and David J Hunter[1,2,3]

Accepted 2 August 2006

**Background** Constitutional factors and sun exposure are associated with skin cancer risk. However, these relations are complex and differ according to skin cancer type.

**Methods** We examined the associations of constitutional risk factors and sun exposure with the risks of three types of skin cancer simultaneously and evaluated the interaction between constitutional susceptibility and sun exposure in a nested case–control study within the Nurses' Health Study [200 melanoma, 275 squamous cell carcinoma (SCC), and 283 basal cell carcinoma (BCC) cases, and 804 controls]. Information regarding skin cancer risk factors was obtained from the retrospective supplementary questionnaire.

**Results** Constitutional susceptibility was an independent risk factor for all three types of skin cancer. Sunlamp usage or tanning salon attendance was a risk factor for melanoma after adjusting for potential confounding variables (OR for ever vs never usage, 2.06, 95% CI 1.30–3.26). Higher sun exposure while wearing a bathing suit was an independent risk factor for all three types of skin cancer. We observed a significant interaction between constitutional susceptibility and sun exposure while wearing a bathing suit on melanoma risk (*P*, interaction, 0.03); women with the highest susceptibility and highest exposure had an OR of 8.37 (95% CI 3.07–22.84). This interaction was weaker and non-significant for SCC and BCC.

**Conclusions** These data largely confirm past studies on risk factors for skin cancer but provide evidence of difference on the strength of these risk factors for melanoma compared with SCC and BCC.

# Study base: less clear

## Risk factors for acute myocardial infarction in a rural population of central India: A hospital-based case–control study

SAMIR S. PATIL, RAJNISH JOSHI, GAUTAM GUPTA, M. V. R. REDDY, MADHUKAR PAI, S. P. KALANTRI

*Background.* There is a paucity of data on the relative importance of various traditional risk factors for coronary artery disease among rural Indians. We conducted a prospective case–control study to determine the risk factors for acute myocardial infarction in a rural population of central India.

*Methods.* We recruited 111 consecutive patients admitted to our hospital with a first episode of acute myocardial infarction and 222 age- and sex-matched controls. Demographics, anthropometric measures, lipids, blood glucose, smoking and other lifestyle factors were compared among cases and controls. Multivariate analyses were used to identify the risk factors independently associated with acute myocardial infarction.

*Results.* Elevated fasting blood glucose (odds ratio [OR] 8.9; 95% confidence interval [CI] 4.5, 17.9), abnormal waist–hip ratio (OR 3.0; 95% CI 1.7, 5.4) and income (OR 4.0 and 5.9 for the high- and middle-income categories, compared to the lowest category) were independently associated with the first episode of acute myocardial infarction. Abnormal triglycerides (OR 1.7; 95% CI 0.9, 3.1) and current smoking (OR 1.9; 95% CI 0.9, 4.0) were risk factors but were not statistically significant.

*Conclusion.* Reduction in blood glucose levels and truncal obesity may be important in controlling the burden of coronary artery disease in rural Indians.

Natl Med J India 2004;17:189–94

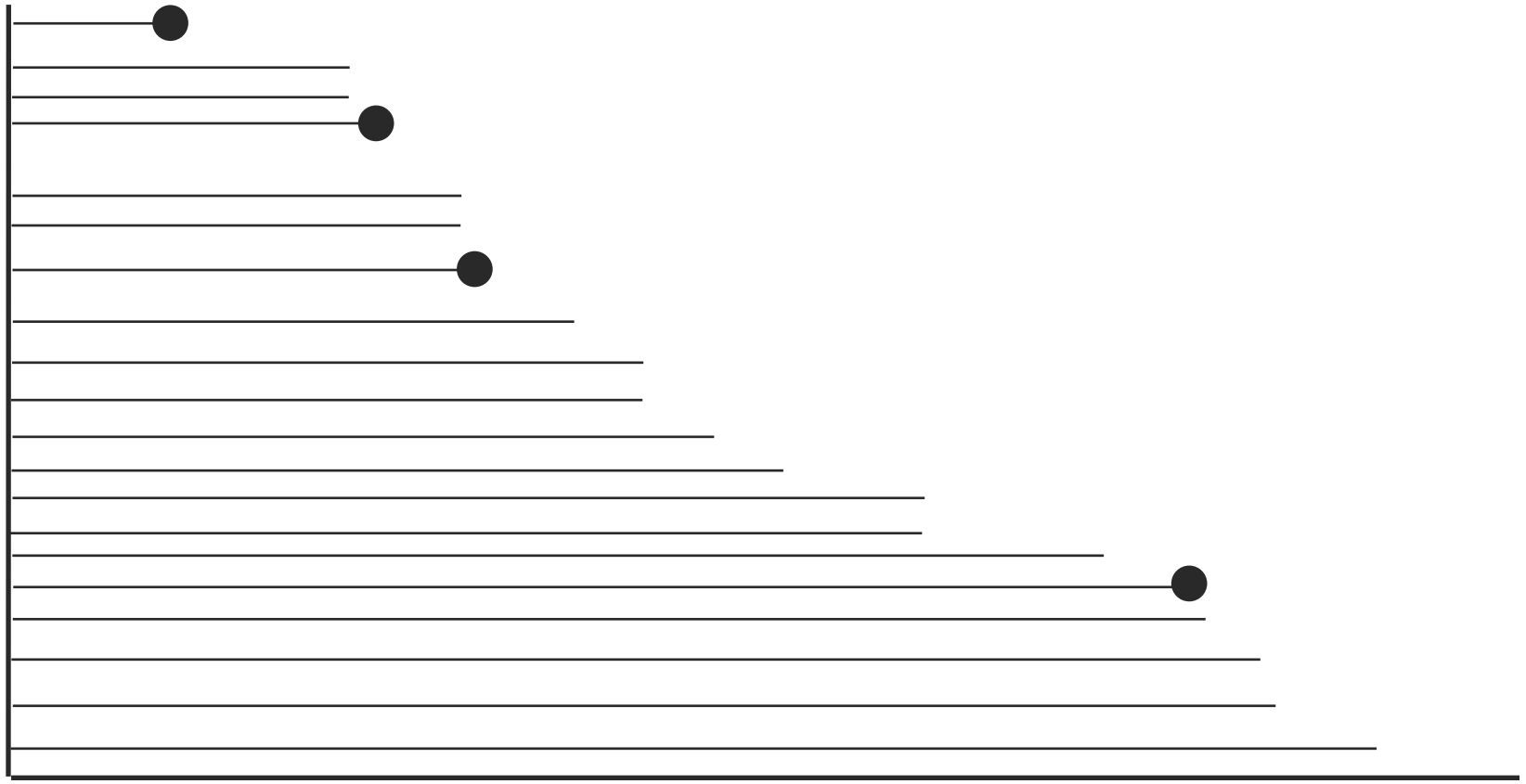
*Cases*

All consecutive patients with AMI admitted to our critical care unit between December 2001 and April 2003 were included in the study. AMI was diagnosed if patients fulfilled 2 of 3 criteria: typical ischaemic chest pain, raised concentrations of creatinine kinase-MB in the serum, and typical ECG findings including development of pathological Q waves.[13] Patients were excluded if they had a past history of myocardial infarction or presented more than 24 hours after the onset of symptoms. Written informed consent was obtained from all patients eligible for inclusion in the study.

*Controls*

We chose two age- and sex-matched controls for each case. Controls were recruited from patients admitted at the same time to the surgical wards for minor surgery such as hernia repair, hydrocele or cataract extraction, or from those visiting the out-patient department for minor complaints related to the eye, ear, nose and throat, and for general health examination. Patients were excluded from the control group if they had clinical evidence of liver disease (jaundice, ascites, oedema or splenomegaly) or if they had had angina or an AMI in the past. We obtained a 12-lead ECG in all controls and excluded those with an abnormal ECG.

# Underlying Hypothetical Cohort

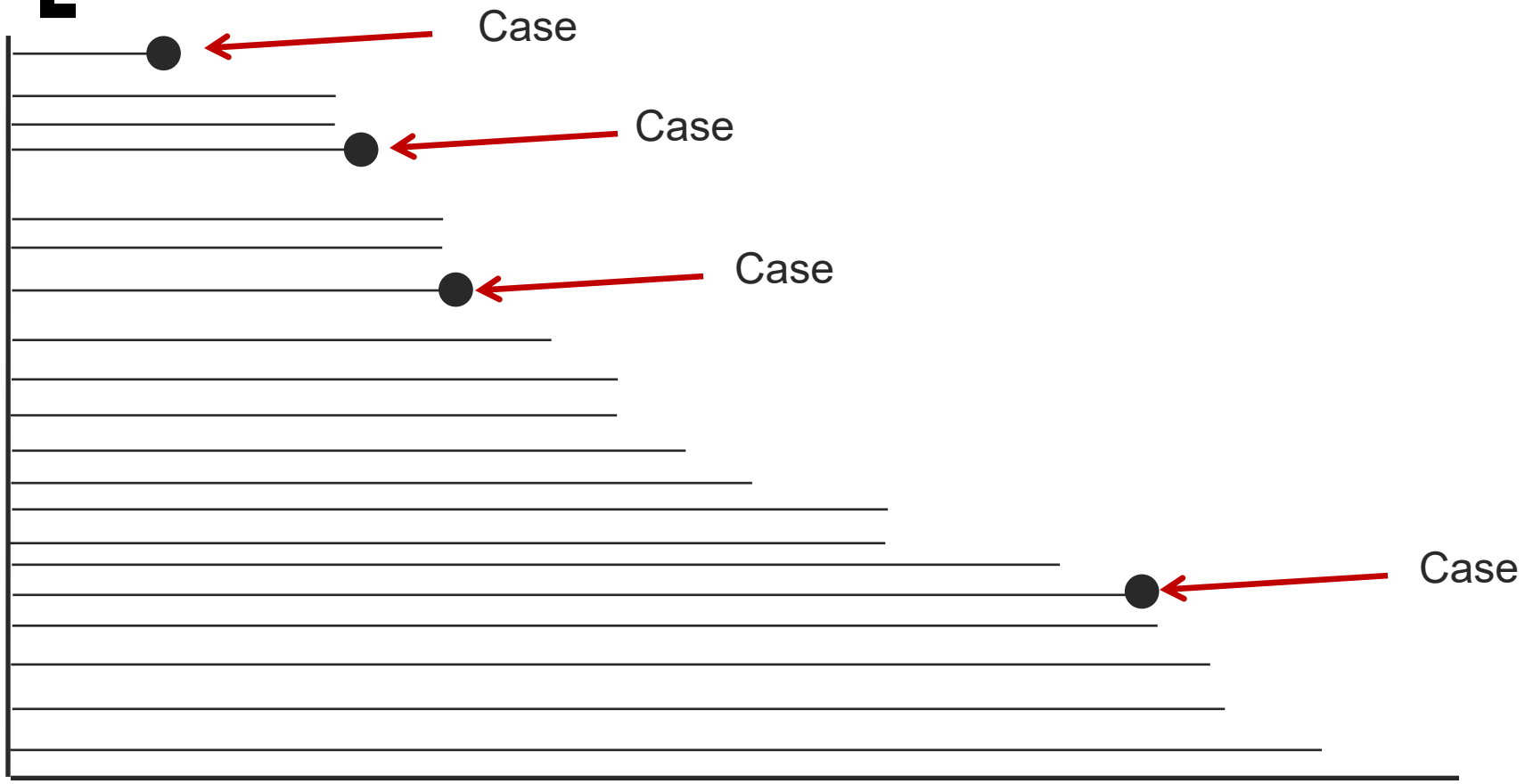


Time



Courtesy: Kris Filion

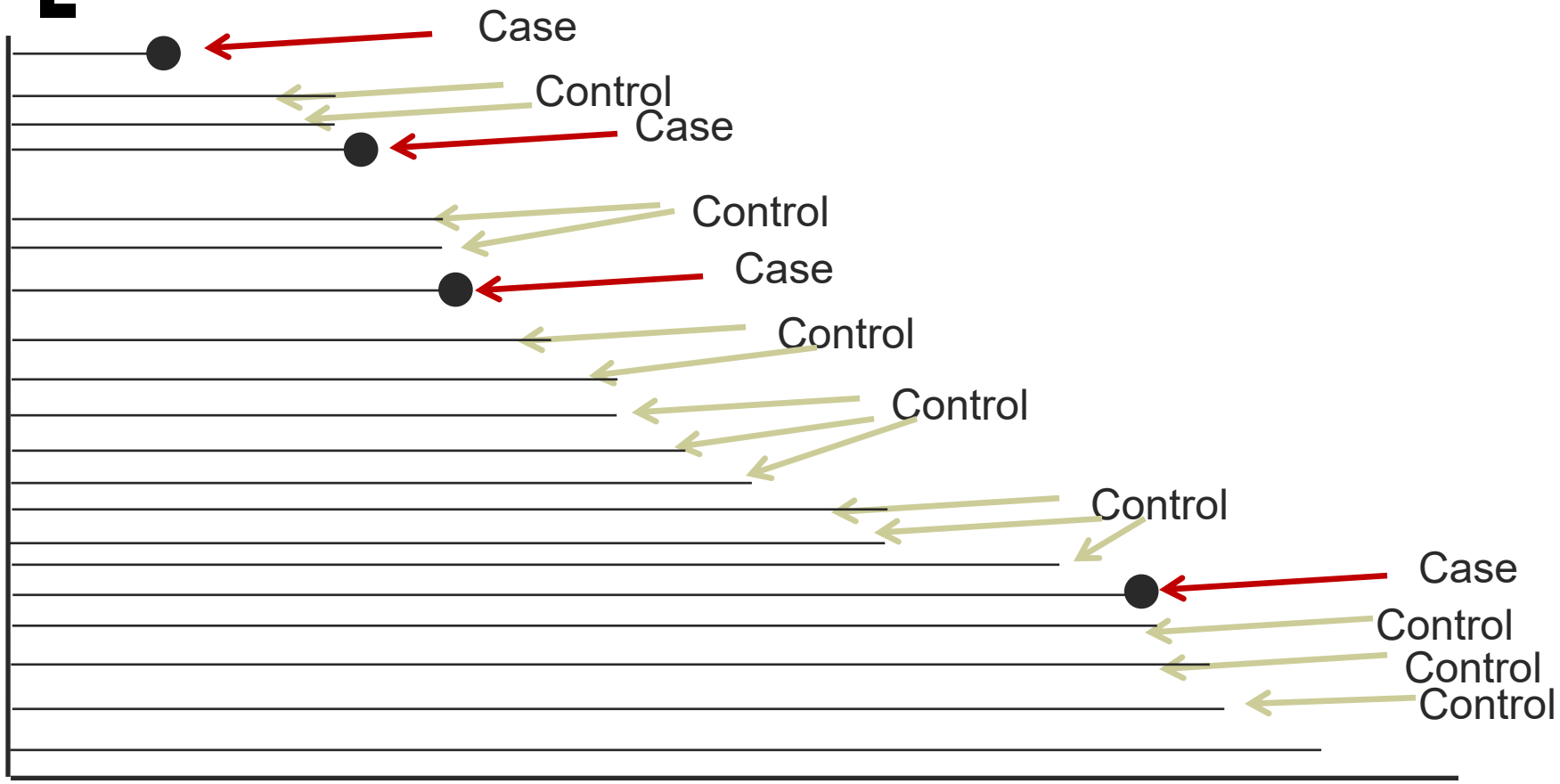
# Cases occur in the cohort






Courtesy: Kris Filion

# Controls are those who could have become cases, but did not

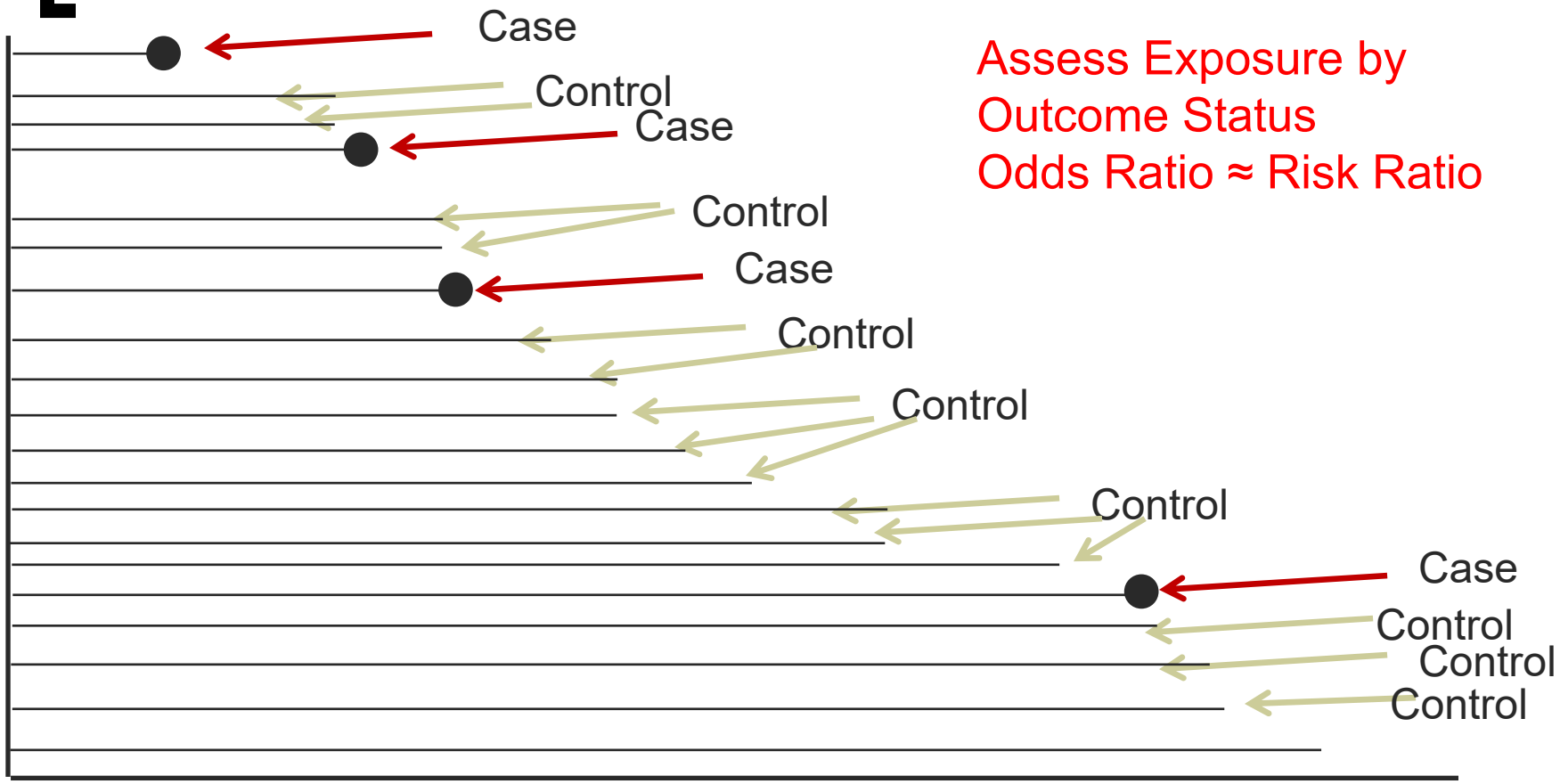



Time



Courtesy: Kris Filion

# We get odds of exposure in cases vs controls and compute the odds ratio




Assess Exposure by Outcome Status

Odds Ratio ≈ Risk Ratio

Time



Courtesy: Kris Filion

# TYPES OF CONTROLS

# Types of controls in case control studies

- Population controls
- Hospital or disease registry controls
- Controls from a medical practice
- Friend controls
- Relative controls

Epidemiology 2

## Compared to what? Finding controls for case-control studies

David A Grimes, Kenneth F Schulz

Lancet 2005; 365: 1429–33

Family Health International, PO Box 13950, Research Triangle Park, NC 27709, USA (D A Grimes MD, K F Schulz PhD)
Correspondence to: Dr David A Grimes dgrimes@fhi.org

Use of control (comparison) groups is a powerful research tool. In case-control studies, controls estimate the frequency of an exposure in the population under study. Controls can be taken from known or unknown study populations. A known group consists of a defined population observed over a period, such as passengers on a cruise ship. When the study group is known, a sample of the population can be used as controls. If no population roster exists, then techniques such as random-digit dialling can be used. Sometimes, however, the study group is unknown, for example, motor-vehicle crash victims brought to an emergency department, who may come from far away. In this situation, hospital controls, neighbourhood controls, and friend, associate, or relative controls can be used. In general, one well-selected control group is better than two or more. When the number of cases is small, the ratio of controls to cases can be raised to improve the ability to find important differences. Although no ideal control group exists, readers need to think carefully about how representative the controls are. Poor choice of controls can lead to both wrong results and possible medical harm.

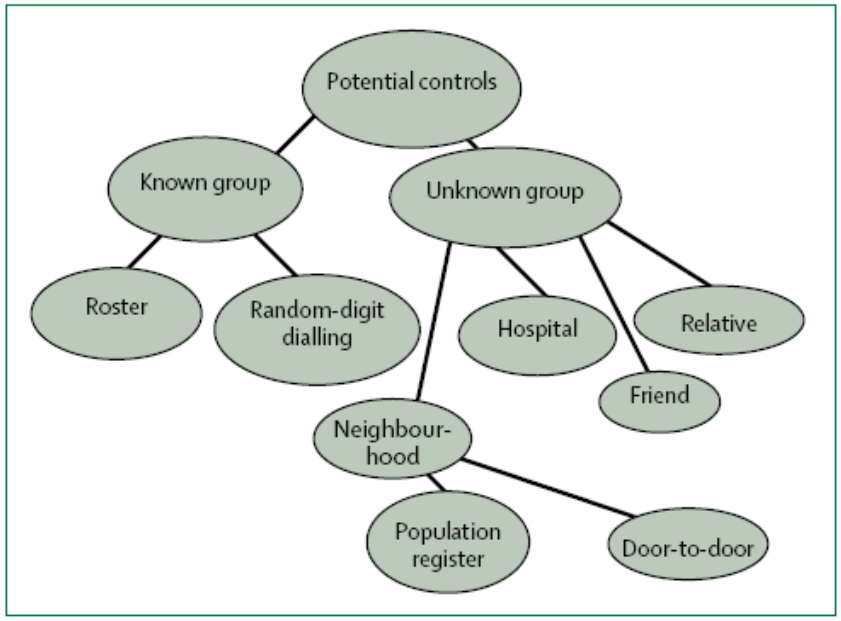



*Figure 2:* Choosing controls with known and unknown group of study participants

# Bias in case-control studies

# Selection bias

- Huge concern in case-control studies
  - Which control group is chosen?
  - How are controls actually recruited?
  - Are controls from the same study base that gave rise to the cases?
  - Are controls chosen independent of the exposure?

# Selection bias (friend controls)

- **Risk factors for menstrual toxic shock syndrome: results of a multistate case-control study.**

  **Reingold AL, Broome CV, Gaventa S, Hightower AW.**

- For assessment of current risk factors for developing toxic shock syndrome (TSS) during menstruation, a case-control study was performed
- Cases with onset between 1 January 1986 and 30 June 1987 were ascertained in six study areas with active surveillance for TSS
- Age-matched controls were selected from among each patient's friends and women with the same telephone exchange
- Of 118 eligible patients, 108 were enrolled, as were 185 "friend controls" and 187 telephone exchange-matched controls



Reingold AL et al. Rev Infect Dis. 1989 Jan-Feb;11 Suppl 1:S35-41

# Selection bias (friend controls)

- Risk factors for menstrual toxic shock syndrome: results of a multistate case-control study

- **Results:**
  - OR when both control groups were combined = 29
  - OR when friend controls were used = 19
  - OR when neighborhood controls were used = 48

- Why did use of friend controls produce a lower OR?
  - Friend controls were more likely to have used tampons than were neighborhood controls (71% vs. 60%)



Reingold AL et al. Rev Infect Dis. 1989 Jan-Feb;11 Suppl 1:S35-41

# Direction of bias

| | | Case | | Control |
|---|---|---|---|---|
| Exposure | Yes | a | ⟺ | b |
| | No | c | | d |

OR = ad / bc

If cases and controls share similar exposures (e.g. friend controls), then a and b will tend to be nearly the same -- this will bias the OR towards 1 (towards null)

# Information bias in case-control studies

Sources:

- Poor recall of past exposures (poor memory; can happen with both cases and controls; so, non-differential)
- Differential recall between cases and controls ("recall bias" or "exposure identification bias" or "exposure suspicion bias")
  - Cases have a different recall than controls
- Differential exposure ascertainment (influenced by knowledge of case status)
  - Interviewer/observer bias (cases are probed or interviewed or investigated differently than controls)

# Poor recall versus recall bias

*Journal of Exposure Science and Environmental Epidemiology* (2009) 19, 369–381

www.nature.com/jes

## Recall bias in the assessment of exposure to mobile phones

MARTINE VRIJHEID[a,b], BRUCE K. ARMSTRONG[c], DANIEL BÉDARD[d], JULIANNE BROWN[c], ISABELLE DELTOUR[a], IVANO IAVARONE[e], DANIEL KREWSKI[d], SUSANNA LAGORIO[f], STEPHEN MOORE[c], LESLEY RICHARDSON[a], GRAHAM G. GILES[g], MARY MCBRIDE[h], MARIE-ELISE PARENT[i], JACK SIEMIATYCKI[j] AND ELISABETH CARDIS[a,b]

[a]*Radiation Group, International Agency for Research on Cancer, Lyon, France*
[b]*Centre for Research in Environmental Epidemiology (CREAL), Municipal Institute of Medical Research (IMIM), Barcelona, Spain*
[c]*Sydney Cancer Centre and School of Public Health, The University of Sydney, Sydney, New South Wales, Australia*
[d]*McLaughlin Centre for population Health Risk Assessment, University of Ottawa, Ottawa, Ontario, Canada*
[e]*Department of Environment & Primary Prevention, Istituto Superiore di Sanità, Rome, Italy*
[f]*National Centre for Epidemiology, Surveillance and Health Promotion, Istituto Superiore di Sanità, Rome, Italy*
[g]*Cancer Epidemiology Centre, The Cancer Council Victoria, Melbourne, Victoria, Australia*
[h]*B.C. Cancer Agency, Vancouver, British Columbia, Canada*
[i]*INRS-Institut Armand-Frappier, Université du Québec, Laval, Quebec, Canada*
[j]*University of Montreal, Montreal, Quebec, Canada*

Most studies of mobile phone use are case–control studies that rely on participants' reports of past phone use for their exposure assessment. Differential errors in recalled phone use are a major concern in such studies. INTERPHONE, a multinational case–control study of brain tumour risk and mobile phone use, included validation studies to quantify such errors and evaluate the potential for recall bias. Mobile phone records of 212 cases and 296 controls were collected from network operators in three INTERPHONE countries over an average of 2 years, and compared with mobile phone use reported at interview. The ratio of reported to recorded phone use was analysed as measure of agreement. Mean ratios were virtually the same for cases and controls: both underestimated number of calls by a factor of 0.81 and overestimated call duration by a factor of 1.4. For cases, but not controls, ratios increased with increasing time before the interview; however, these trends were based on few subjects with long-term data. Ratios increased by level of use. Random recall errors were large. In conclusion, there was little evidence for differential recall errors overall or in recent time periods. However, apparent overestimation by cases in more distant time periods could cause positive bias in estimates of disease risk associated with mobile phone use.
*Journal of Exposure Science and Environmental Epidemiology* (2009) **19,** 369–381; doi:10.1038/jes.2008.27; published online 21 May 2008

# Information bias: example from case-control study of risk factors for suicide in Pakistan

**Table 1 ICD–10 principal diagnosis**

| Diagnosis | Cases (*n*=100) | Controls (*n*=100) |
|---|---|---|
| Moderate depressive episode (F32.1) | 30 | 1 |
| Severe depressive episode (F32.2) | 43 | 0 |
| Severe depressive episode with psychotic symptoms (F32.3) | 6 | 2 |
| Schizophrenia (F20) | 6 | 2 |
| Adjustment disorders (F43.2) | 3 | 0 |
| Acute stress reaction (F43.0) | 6 | 0 |
| Alcohol use (F10.0) | 0 | 0 |
| Substance abuse (F11.0) | 1 | 0 |
| Mental retardation (F79) | 1 | 0 |
| Personality disorder (F60) | 1 | 1 |
| No psychiatric diagnosis | 4 | 94 |

**Table 2 Final multivariable conditional logistic regression model**

| Variable | Adjusted OR (95% CI) |
|---|---|
| Educational attainment | |
| No formal education/primary[a] | 4.9 (0.8–29.8) |
| Secondary and above | 1.0 |
| Marital status | |
| Never married | 1.0 |
| Ever married | 3.6 (0.6–22.3) |
| Depression | |
| No | 1.0 |
| Yes | 208.3 (11.0–3935.2) |

a. Adjusted for employment status.

- Close relatives of 100 suicide cases and 100 live controls were interviewed.
- 79/100 suicide cases were found to have had depression.
- Only 3/100 controls were found to have depression (lower than the population average).
- Due to lack of blinding, quality of interviews may have been lower in controls



Brit J Psych 2008

# Confounding in case-control studies

- Always an issue!
- Can be addressed at the design or analysis stage [usually both]:
  - Design:
    - Matching
    - Restriction
  - Analysis:
    - Multivariable analysis
      - Logistic regression (LR) is the most natural model
      - Results reported as adjusted odds ratios