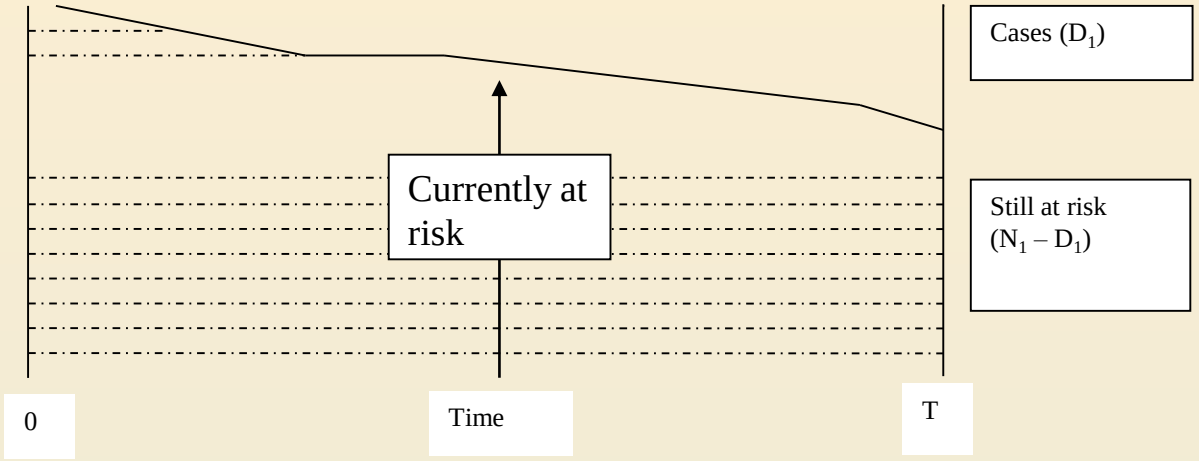
The analysis of case-control studies

"Nested case control and case cohort studies: Estimation of relative risk"

1. CHOICE OF CONTROLS IN CASE-CONTROL STUDIES

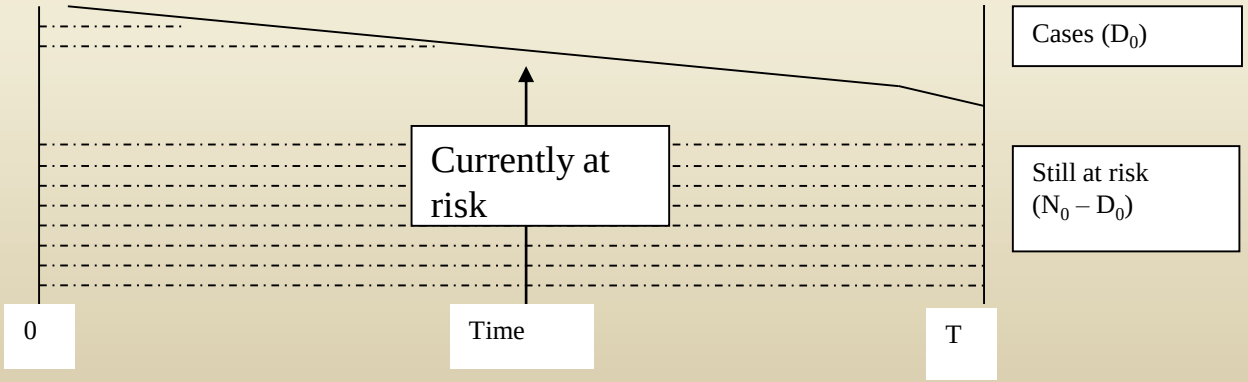
Exposed

Initially at risk (N_1)



Unexposed

Initially at risk (N_0)



2. Measures of relative risk and odds ratio in the underlying population

Measure	Definition	Alternative formulation
Risk Ratio	$\frac{D_1/N_1}{D_0/N_0}$	$\frac{D_1/D_0}{N_1/N_0}$ = people at risk in the beginning
Rate Ratio	$\frac{D_1/Y_1}{D_0/Y_0}$	$\frac{D_1/D_0}{Y_1/Y_0}$ = persons years at risk for the duration of the study
Odds Ratio	$\frac{D_1/(N_1-D_1)}{D_0/(N_0-D_0)}$	$\frac{D_1/D_0}{(N_1-D_1)/(N_0-D_0)}$ = **people still at risk at the end of the study**

Common numerator: exposed/non-exposed **cases** in the population: D_1/D_0

"the odds of exposure among those with disease" - can be estimated from **cases** in a case-control study- represent the underlying population of those with disease from where the cases were drawn!.

Denominators: exposed/non-exposed **individuals** or their **follow-up times**

- **By a suitable sampling scheme, each of these three denominators may be estimated by the exposed/non-exposed controls from a case-control study.**

3.a. CHOICE OF CONTROLS IN A CASE-CONTROL STUDY (1)

Cumulative (Exclusive) sampling: from the disease-free in the end of the study period.

Cumulative (Exclusive) sampling

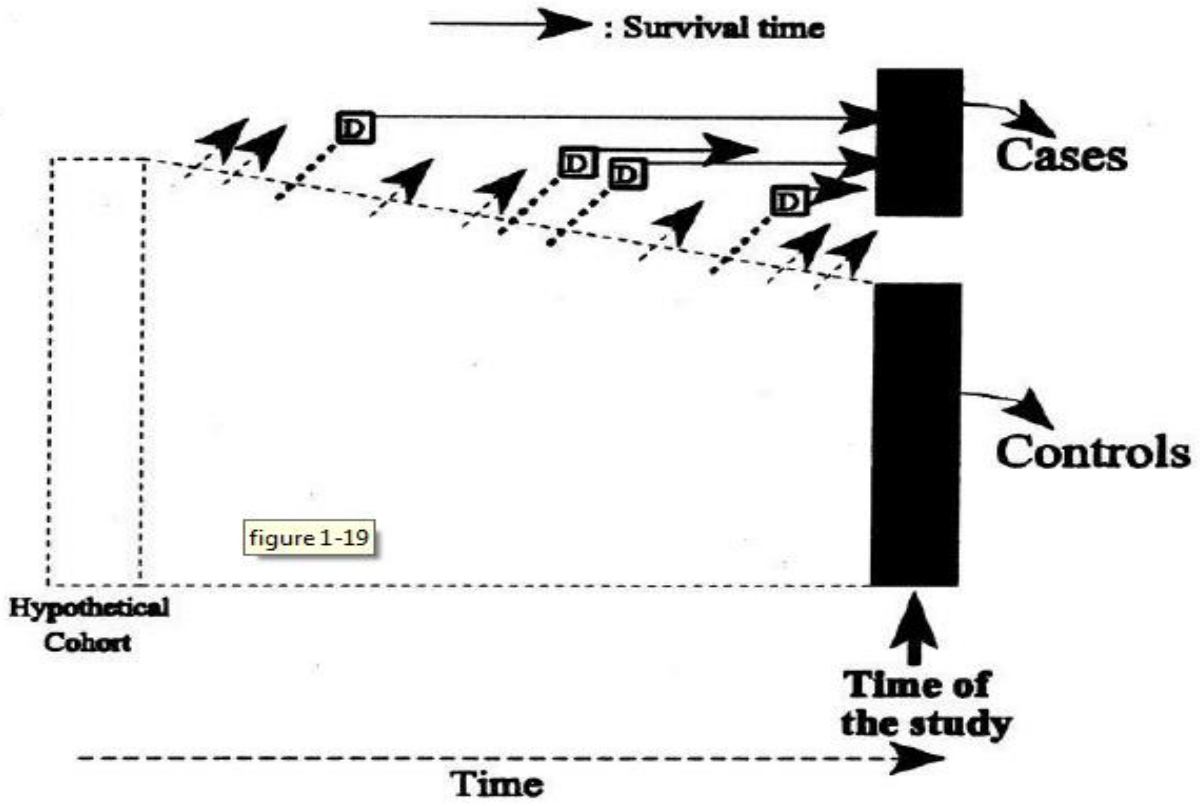


Figure 1–19 Survival bias in a case-based case-control study carried out "cross-sectionally": only cases with long survival after diagnosis (best prognosis) are included in the case group. In this hypothetical example, the horizontal lines starting in the cases' "D" boxes represent survival times; note that only two of the four cases are included in the study. Broken diagonal lines with arrows represent losses to follow-up.

Szklo & Nieto. Epidemiology: beyond the basics. Aspen Publishers, 2000

Odds ratio from case-control =

population odds ratio (without the non-rare disease assumption)

Cumulative (Exclusive) sampling - example:

Case–Control Study of Blood Lead Levels and Attention Deficit Hyperactivity Disorder in Chinese Children

Hui-Li Wang,[1] Xiang-Tao Chen,[1,2] Bin Yang,[3] Fang-Li Ma,[4] Shu Wang,[1] Ming-Liang Tang,[1] Ming-Gao Hao,[5] and Di-Yun Ruan[1]

BACKGROUND: Attention deficit/hyperactivity disorder (ADHD) and lead exposure are high-prevalence conditions among children.

OBJECTIVE: Our goal was to investigate the association between ADHD and blood lead levels (BLLs) in Chinese children, adjusting for known ADHD risk factors and potential confounding variables.

METHODS: We conducted a pair-matching case–control study with 630 ADHD cases and 630 non-ADHD controls 4–12 years of age, matched on the same age, sex, and socioeconomic status. The case and control children were systematically evaluated via structured diagnostic interviews, including caregiver interviews, based on the *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed., revised criteria (DSM-IV-R). We evaluated the association between BLLs and ADHD using the Pearson chi-square test for categorical variables and the Student *t*-test for continuous data. We then performed conditional multiple variables logistic regression analyses with backward stepwise selection to predict risk factors for ADHD.

RESULTS: There was a significant difference in BLLs between ADHD cases and controls. ADHD cases were more likely to have been exposed to lead during childhood than the non-ADHD control subjects, with adjustment for other known risk factors [children with BLLs ≥ 10 μg/dL vs. ≤ 5 μg/dL; OR = 6.0; 95% confidence interval (CI) = 4.10–8.77, $p < 0.01$; 5–10 μg/dL vs. ≤ 5 μg/dL, OR = 4.9; 95% CI = 3.47–6.98, $p < 0.01$]. These results were not modified by age and sex variables.

CONCLUSIONS: This was the largest sample size case–control study to date to study the association between BLLs and ADHD in Chinese children. ADHD may be an additional deleterious outcome of lead exposure during childhood, even when BLLs are < 10 μg/dL.

KEY WORDS: attention deficit hyperactivity disorder, blood lead levels, case–control study. *Environ Health Perspect* 116:1401–1406 (2008). doi:10.1289/ehp.11400 available via *http://dx.doi.org/* [Online 5 June 2008]

ADHD subjects were consecutively recruited from children coming for initial or follow-up assessment from October 2003 to August 2007 in two pediatric Clinics [Note: prevalent cases are included]

The non-ADHD controls were randomly selected from computerized lists of outpatients admitted for acute upper respiratory infection at the same two pediatric medical clinics during the same period

3.b. CHOICE OF CONTROLS IN A CASE-CONTROL STUDY (2)

Case-base, Case-cohort, Inclusive sampling: from the initially (in the beginning of the study) at risk of disease:

In a case-cohort study, cases are defined as those participants of the cohort who developed the disease of interest, but controls are identified before the cases develop.
This means that controls are randomly chosen from all cohort participants regardless of whether they have the disease of interest or not, and that baseline data can be collected early in the study.

Controls are randomly selected from the parent cohort, forming a sub-cohort. **No matching is performed**.

Case-base, Case-cohort, Inclusive sampling

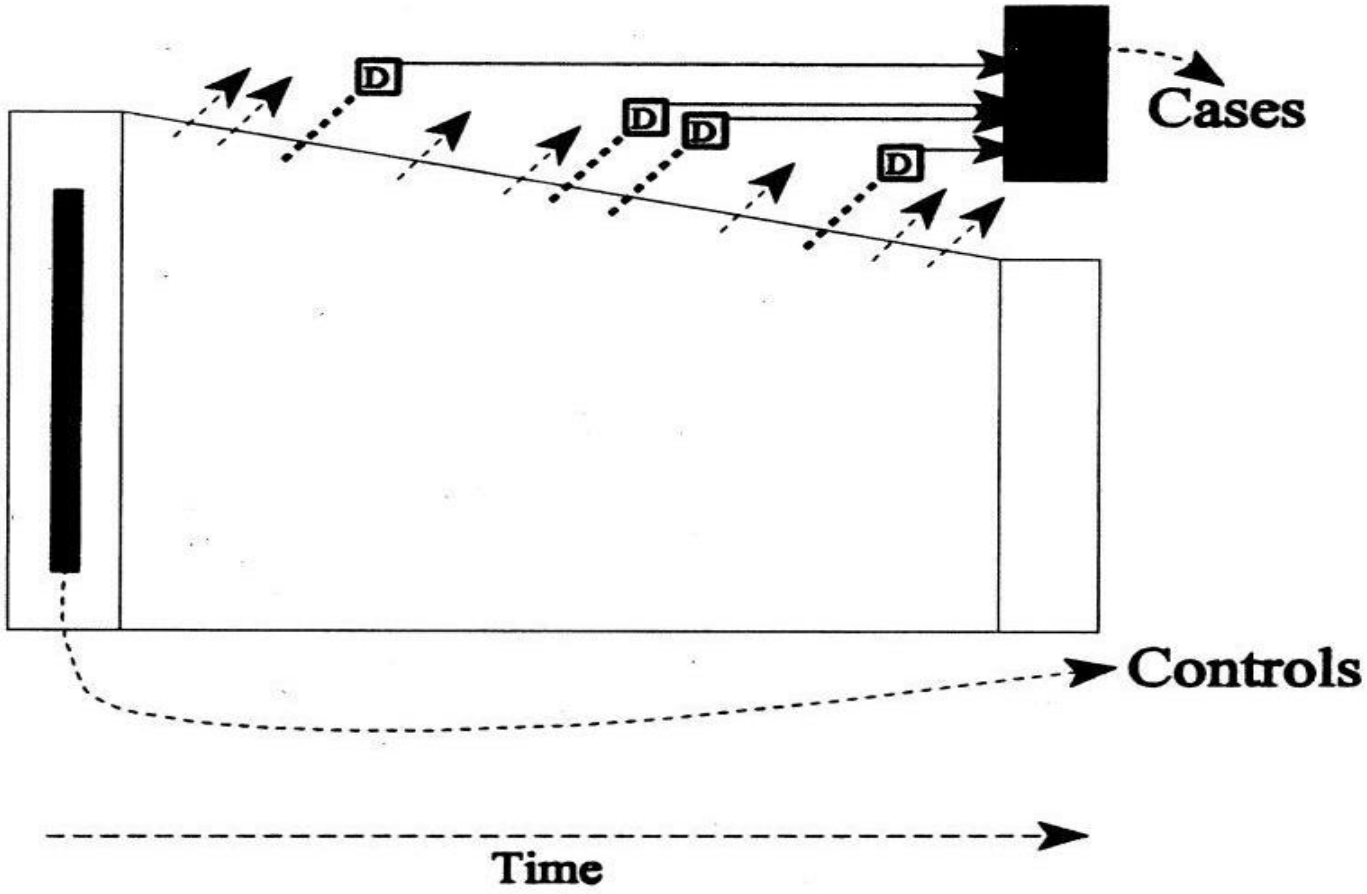
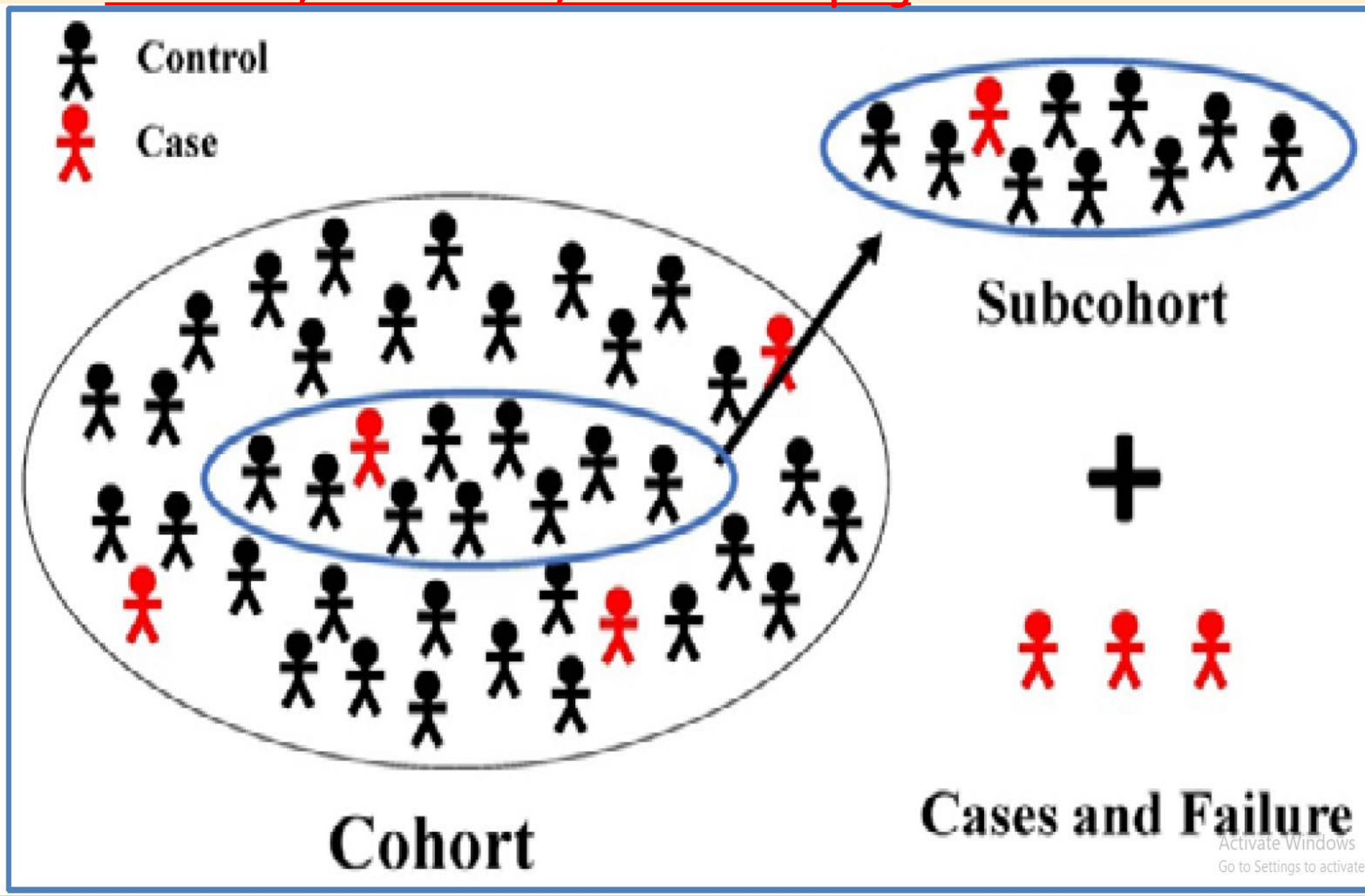


Figure 1–20 Case-control study in which the controls are selected from the baseline cohort (case-cohort study). Cases are represented by "D" boxes. Broken diagonal lines with arrows represent losses to follow-up.

Szklo & Nieto. Epidemiology: beyond the basics. Aspen Publishers, 2000

Case-base, Case-cohort, Inclusive sampling



Case-Cohort Studies vs Nested Case-Control Studies, Soyoung Kim, PhD, Division of Biostatistics, MCW- *Volume 22, Number 1, February/March 2016*

3.b. CHOICE OF CONTROLS IN A CASE-CONTROL STUDY (2)

Case-base, Case-cohort, Inclusive sampling: from the initially (in the beginning of the study) at risk of disease:

- controls can become cases and selected 2 times in the case-control,
- **the control group is a sample of the total population initially at risk of the disease**,
- the odds of exposure in the control group estimates the odds of exposure in the population $(N_1/N)/(N_0/N) = N_1/N_0$
- It can be shown that the odds ratio from case-control = population risk ratio

Explanation

From a cohort study measuring risk of disease in exposed (Ne) and unexposed (Nu) cohorts we can draw the following results table:

Exposure	cases	Population at risk	IP	Risk ratio
Yes	D_1	N_1	D_1 / N_1	$(D_1 / N_1)/(D_0/ N_0)$
No	D_0	N_0	D_0/N_0	

Explanation

- Now assume a case-cohort, where controls are a sample of the source population:

Exposure	Cases	Controls=Sample from source population
Yes	D_1	$N_1/10$
No	D_0	$N_0/10$

- The **risk of disease cannot be estimated** from the above table, since denominators sampled from exposed and unexposed cohorts are only a sampling fraction of these two populations.
- However, **the risk ratio remains the same**. If in the risk ratio calculation we replace the denominators by the 10% samples representing them, we obtain the same value for the risk ratio:
  - » $RR=D_1/(N_1/10)/D_0/(N_0/10)$

Case-cohort (Inclusive) sampling - example:

European Journal of Epidemiology (2007) 22:173–181
DOI 10.1007/s10654-006-9099-5

PERINATAL EPIDEMIOLOGY

Oral clefts and life style factors – A case–cohort study based on prospective Danish data

Camilla Bille[1,8], Jorn Olsen[2], Werner Vach[3], Vibeke Kildegaard Knudsen[4], Sjurdur Frodi Olsen[4], Kirsten Rasmussen[5], Jeffrey C. Murray[1,6], Anne Marie Nybo Andersen[7] & Kaare Christensen[1]

Abstract. This study examines the association between oral clefts and first trimester maternal lifestyle factors based on prospective data from the Danish National Birth Cohort. The cohort includes approximately 100,000 pregnancies. In total 192 mothers gave birth to child with an oral cleft during 1997–2003. Information on risk factors such as smoking, alcohol consumption, tea, coffee, cola, and food supplements was obtained during pregnancy for these and 828 randomly selected controls. We found that first trimester maternal smoking was associated with an increased risk of oral clefts (odds ratio (OR): 1.50; 95% confidence interval (CIs): 1.05, 2.14). Although not statistically significant, we also saw associations with first trimester consumption of alcohol (OR: 1.11; CIs: 0.79, 1.55), tea (OR: 1.31; CIs: 0.93, 1.86), and drinking more than 1 l of cola per week (OR: 1.40; CIs: 0.92, 2.12). Furthermore supplementation with ≥400 mcg folic acid daily during the entire first trimester (OR: 0.75; CIs: 0.46, 1.22) suggested an inverse associated with oral clefts, similar to our results on coffee drinking. No effects were found for smaller doses of folic acid, vitamin A, B6 or B12 in this study. The present study found an association between oral clefts and smoking and, although not conclusive, supports an association of oral cleft with alcohol.

Cases were identified through 2 sources: (1) maternally reported oral clefts in post pregnancy interviews in the birth cohort ; and (2) a discharge diagnosis of oral clefts or an ICD-10 code for reconstructive surgery on lips or palate in The National Patient Register.

Controls were selected randomly among participants at baseline (the first interview) in the birth cohort.

Nested case-control studies

Case-control study nested in cohort:

Use an existing cohort design a case control study nested within the cohort study:

Cases for the case-control study are all cases identified in the cohort in a specific period (usually the whole period of the study)

Controls are usually chosen from those at risk of becoming cases (**risk set**) at the time of diagnosis of each case (incidence density sampling):

- Cost saving: collecting data on controls only (instead of the entire cohort) at risk at each time of diagnosis of cases.

- Collection of data on a sub-sample of subjects for unmeasured covariates in a cohort study.

New Data collection only restricted to Cases and matched Controls.

3.c. CHOICE OF CONTROLS IN A CASE-CONTROL STUDY (3)

Concurrent sampling, incidence density sampling:
from those still at risk of the disease at the time of diagnosis of the case.

Concurrent sampling, incidence density sampling:

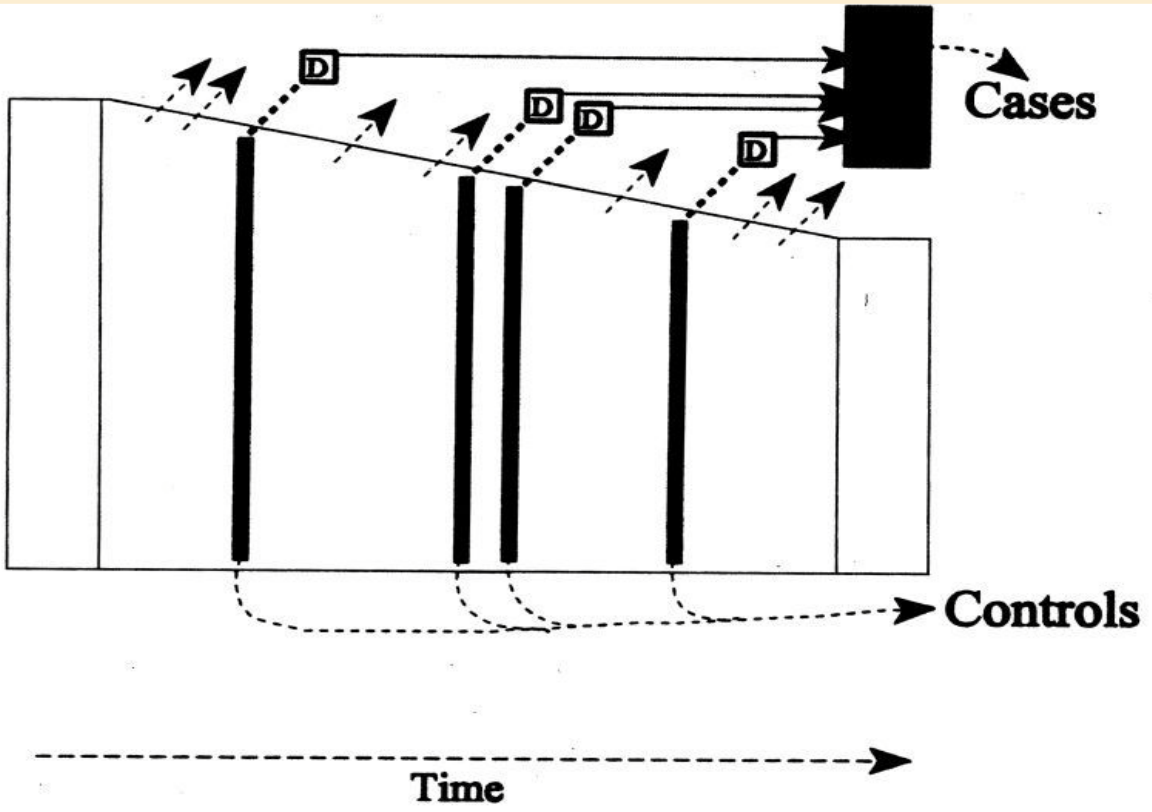


Figure 1–21 Nested case-control study in which the controls are selected at each time when a case occurs (incidence density sampling). Cases are represented by "D" boxes. Broken diagonal lines with arrows represent losses to follow-up.

Szklo & Nieto. Epidemiology: beyond the basics. Aspen Publishers, 2000

3c. CHOICE OF CONTROLS IN A CASE-CONTROL STUDY (3)

Concurrent sampling, incidence density sampling:
from those still at risk of the disease at the time of diagnosis of the case.

- controls can become cases and selected as a cases,
- implies matching for time-at-diagnosis,
- the odds of exposure in the control group provide an estimate of Y_1/Y_0,

Odds ratio from case-control analysed as matched = population rate ratio

Concurrent sampling, incidence density sampling - example:

Nested case-control: example

Risk of acute myocardial infarction and sudden cardiac death in patients treated with cyclo-oxygenase 2 selective and non-selective non-steroidal anti-inflammatory drugs: nested case-control study

David J Graham, David Campen, Rita Hui, Michele Spence, Craig Cheetham, Gerald Levy, Stanford Shoor, Wayne A Ray

Summary

Background Controversy has surrounded the question about whether high-dose rofecoxib increases or naproxen decreases the risk of serious coronary heart disease. We sought to establish if risk was enhanced with rofecoxib at either high or standard doses compared with remote non-steroidal anti-inflammatory drug (NSAID) use or celecoxib use, because celecoxib was the most common alternative to rofecoxib.

Methods We used data from Kaiser Permanente in California to assemble a cohort of all patients age 18–84 years treated with a NSAID between Jan 1, 1999, and Dec 31, 2001, within which we did a nested case-control study. Cases of serious coronary heart disease (acute myocardial infarction and sudden cardiac death) were risk-set matched with four controls for age, sex, and health plan region. Current exposure to cyclo-oxygenase 2 selective and non-selective NSAIDs was compared with remote exposure to any NSAID, and rofecoxib was compared with celecoxib.

Findings During 2 302 029 person-years of follow-up, 8143 cases of serious coronary heart disease occurred, of which 2210 (27·1%) were fatal. Multivariate adjusted odds ratios versus celecoxib were: for rofecoxib (all doses), 1·59 (95% CI 1·10–2·32, p=0·015); for rofecoxib 25 mg/day or less, 1·47 (0·99–2·17, p=0·054); and for rofecoxib greater than 25 mg/day, 3·58 (1·27–10·11, p=0·016). For naproxen versus remote NSAID use the adjusted odds ratio was 1·14 (1·00–1·30, p=0·05).

Interpretation Rofecoxib use increases the risk of serious coronary heart disease compared with celecoxib use. Naproxen use does not protect against serious coronary heart disease.

We assembled a cohort of NSAID-treated patients to undertake a nested case-control study. From Jan 1, 1999, to Dec 31, 2001, we identified all individuals age 18–84 years who filled at least one prescription for a COX2 selective (celecoxib or rofecoxib) or non-selective (all other) NSAID. Those with at least 12 months of health plan coverage before the date of that first NSAID prescription were entered into the cohort if they had no diagnoses of cancer, renal failure, liver failure, severe respiratory disease, organ transplantation, or HIV/AIDS during the screening interval. We followed up cohort members from this entry date until the end of the study period (December, 2001) or until occurrence of an acute myocardial infarction or death, whichever came first.

For every case, we randomly selected four controls from individuals under observation in the study cohort on the date of the case event (index date), and matched them for age (year of birth), sex, and health plan region (north or south).[17] A given cohort member selected as a control for a case on one date could become a control for another case occurring on a later index date, as long as he or she remained in the study cohort and was therefore also at risk of becoming a case. Thus, a control could subsequently become a case. We excluded potential cases and controls if they were not enrolled on the index date and for at least 11 of the 12 preceding months. During the study period, pharmacy benefits persisted for enrolment lapses of up to 1 calendar month.

Nested case-control studies

Sampling from those at risk

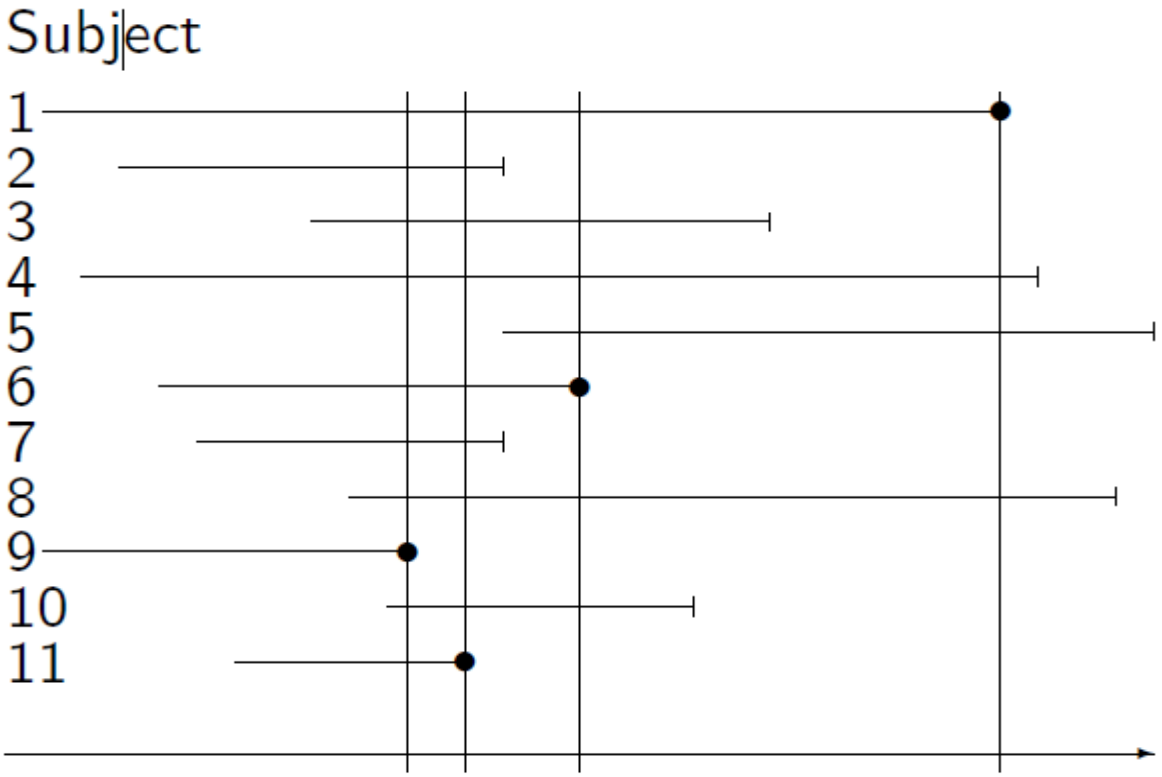
Ideally we would collect information on exposures of interest (or confounders) on all persons **at-risk** at the time of diagnosis of each case (the **risk set**).

Sample the **risk set** instead:

- collect exposure information only for a subsample of those at risk (chosen at random).

- the sampled persons are then used to represent the risk set.

Sampling from the risk set: example



- Case at time (•)
- List all available controls (**risk set**) for each case and draw *n* controls (e.g. n=2) at random for each case (**sample**)

Risk sets and sampling procedures

at each event time of a case (•):

Event	Risk set	Sample
1	1,2,3,4,6,7,8,**9**,10,11	4,1
2	1,2,3,4,6,7,8,10,**11**	2,1
3	1,3,4,5,**6**,8,10	8,3
4	**1**,4,5,8	4,5

- Note that subjects 4, 1 are selected twice as controls, and individual 1 eventually becomes a case.
- This is perfectly OK, since these are at risk at the time where they are selected as controls for the sample of risk set.

Nested case-control studies

Study base/pool = "large" cohort.

- Expensive to measure exposure or confounders of interest for all individuals (e.g. expensive biochemical analyses, DNA analyses).
- Measure exposure/confounders only for cases and time-matched controls
- Usually for each case 1-5 controls are sampled from the risk set — i.e. persons at risk at the time of diagnosis of the case.
- Can use any case-control design (not only incident density sampling)

Study base/pool = entire population.

- Nested case-control study sampled from the entire population (e.g. using registries to define cases and sample the controls)
- Cannot afford to collect exposure/confounders information for the entire population, so only a subsample is used.

- **Can use any case-control design** (not only incident density sampling)

Study base/pool = entire population

Mobile phone use and risk of acoustic neuroma: results of the Interphone case–control study in five North European countries

MJ Schoemaker[*,1], AJ Swerdlow[1], A Ahlbom[2,13], A Auvinen[3,10], KG Blaasaas[4], E Cardis[5], H Collatz Christensen[6], M Feychting[2], SJ Hepworth[7], C Johansen[6], L Klæboe[8], S Lönn[2], PA McKinney[7], K Muir[9], J Raitanen[10], T Salminen[3], J Thomsen[11] and T Tynes[8,12]

[1]Section of Epidemiology, Institute of Cancer Research, Brookes Lawley Building, Sutton SM2 5NG, UK; [2]Institute of Environmental Medicine, Karolinska Institute, Box 210, 171 77, Stockholm, Sweden; [3]STUK-Radiation and Nuclear Safety Authority, 00881 Helsinki, Finland; [4]Norwegian Armed Forces, Bygning 0028A, Sessvollmoen 2058, Norway; [5]International Agency for Research on Cancer, 150 Cours Albert Thomas, 69372 Cedex 08, Lyon, France; [6]Institute of Cancer Epidemiology, Danish Cancer Society, Strandboulevarden 49, 2100 Copenhagen, Denmark; [7]Centre for Epidemiology and Biostatistics, University of Leeds, 30 Hyde Terrace, Leeds LS2 9LN, UK; [8]The Cancer Registry of Norway, Institute of Population-based Cancer Research, Montebello, 0310 Oslo, Norway; [9]Division of Epidemiology and Public Health, University of Nottingham, Nottingham NG7 2UH, UK; [10]Tampere School of Public Health, University of Tampere, Tampere 33014, Finland; [11]Department of Otolaryngology-Head and Neck Surgery, Gentofte Hospital, University of Copenhagen, DK-2900 Heilerup, Denmark; [12]Norwegian Radiation Protection Authority, PO Box 55, 1332 Osteras, Norway

There is public concern that use of mobile phones could increase the risk of brain tumours. If such an effect exists, acoustic neuroma would be of particular concern because of the proximity of the acoustic nerve to the handset. We conducted, to a shared protocol, six population-based case–control studies in four Nordic countries and the UK to assess the risk of acoustic neuroma in relation to mobile phone use. Data were collected by personal interview from 678 cases of acoustic neuroma and 3553 controls. The risk of acoustic neuroma in relation to regular mobile phone use in the pooled data set was not raised (odds ratio (OR) = 0.9, 95% confidence interval (CI): 0.7–1.1). There was no association of risk with duration of use, lifetime cumulative hours of use or number of calls, for phone use overall or for analogue or digital phones separately. Risk of a tumour on the same side of the head as reported phone use was raised for use for 10 years or longer (OR = 1.8, 95% CI: 1.1–3.1). The study suggests that there is no substantial risk of acoustic neuroma in the first decade after starting mobile phone use. However, an increase in risk after longer term use or after a longer lag period could not be ruled out.

British Journal of Cancer (2005) **93,** 842–848. doi:10.1038/sj.bjc.6602764 www.bjcancer.com

Cases were identified through neurosurgery, neuropathology, oncology, neurology and otorhinolaryngology centres in the study areas. Lists of cases were also obtained from the appropriate population-based cancer registries to ensure completeness of ascertainment. Eligible cases were individuals diagnosed with acoustic neuroma between 1 September 1999 and 31 August 2004 (the exact dates within this period vary by centre) at ages 20–69 years in the Nordic countries, 18–59 in Southeast England, and 18–69 in the Northern UK, and resident in the study region at the time of diagnosis.

Controls in the Nordic centres were randomly selected from the population register for each study area, frequency matched to cases on age, sex and region. In the UK, where there is no such accessible population register, controls were randomly selected from general practitioners' practice lists. Controls were subject to the same age and residence criteria as cases and had never been diagnosed with a brain tumour.

Case-cohort vs. nested case control

- In a **case-cohort study**, cases are defined as those participants of the cohort who developed the disease of interest, but **controls are identified before the cases develop**.
- This means that controls are randomly chosen from all cohort participants regardless of whether they have the disease of interest or not, and that baseline data can be collected early in the study.
- The **main difference** between a nested case-control study and a case-cohort study is **the way in which controls are chosen**.
- Generally, the main advantage of case-cohort design over nested case-control design is that the same control group can be used for comparison with different case groups in a case-cohort study.

Encyclopedia of Public Health
Editors: Wilhelm Kirch

Analysis of nested case-control and case-cohort studies

- For both nested case-control and case-cohort designs, inverse probability weighting methods were more powerful than the standard methods.

- However, the difference became negligible when the proportion of failure events was very low (<1%) in the full cohort.

Kim RS.
A new comparison of nested case-control and case-cohort designs and methods.
Eur J Epidemiol. 2015 Mar;30(3):197-207.

Sampling options in case-control studies (summary – references)

Type of sampling	References	What does the case-control odds ratio estimate?
Cumulative sampling (Exclusive)	Cornfield (1951)	Odds ratio
Case-cohort sampling (Inclusive)	Thomas (1972) Kupper et al (1975) Miettinen (1982) Smith et al (1984) Prentice (1986)	Risk ratio
Density sampling (Concurrent)	Sheehe (1962) Miettinen (1976)	Rate ratio

More References

•Rodrigues L, Kirkwood BR. Case-control designs in the study of common diseases: updates on the demise of the rare disease assumption and the choice of sampling scheme for controls. Int J Epidemiol 1990; 19: 205-213.

•Pearce N. What does the odds ratio estimate in a case-control study? Int J Epidemiol 1993; 22: 1189-1192.