Case–control studies and the odds ratio

Applied epidemiology · 90-minute practical

Practical Class 12 · Dental Public Health · 2026/2027

Chief Assistant Professor Kostadin Kostadinov, MD, PhD

Department of Social Medicine and Public Health · Medical University of Plovdiv

Explain the sampling before the odds ratio

Learning outcomes

Define the source population and select cases and controls. Calculate exposure odds and an unmatched OR. Explain invalid sampled risks. Detect selection bias, recall bias and confounding. Distinguish cumulative and risk-set sampling.

90 minutes: population and sampling; odds and OR; validity; independent analysis.

Vandenbroucke & Pearce (2012). doi:10.1093/ije/dys147

Preparation and practice

Materials

Bring a calculator. State the outcome, denominator and observation period in each answer.

15-minute reading

Five tasks with model answers

Slides PDF

CONTENTS

1. **The source population and sampling**
2. Odds and the odds ratio
3. What the OR estimates
4. Selection, recall, matching and confounding
5. Independent analysis and critical appraisal

Begin with the population that could become cases

Hypothetical oral-cancer study in Plovdiv

Include eligible new diagnoses among residents of a defined catchment during a defined period. Specify anatomical sites, diagnostic criteria and residency before recruitment.

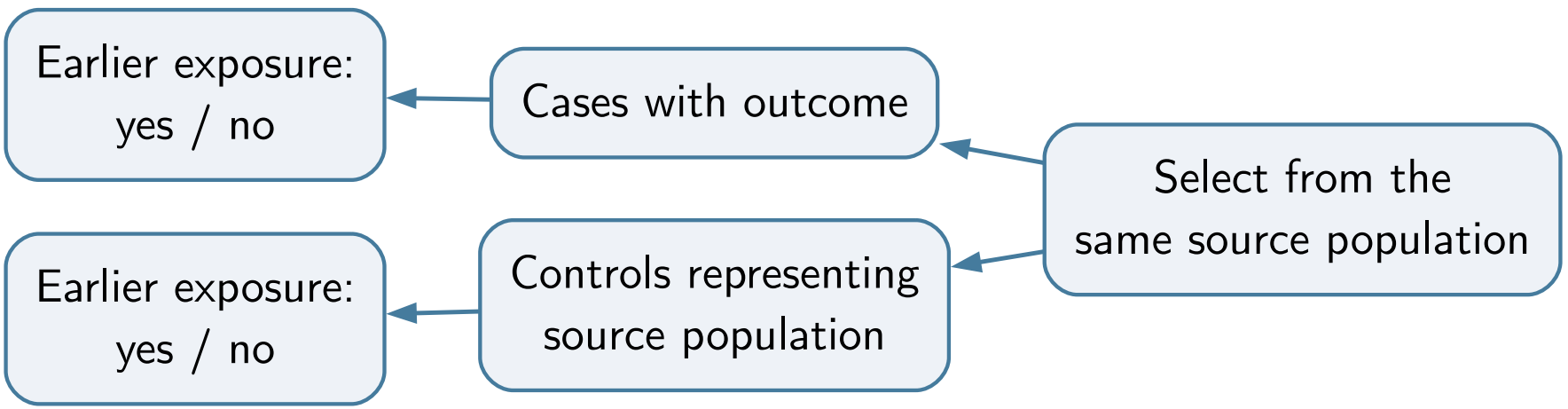
The key control question

If this control developed the qualifying disease during the study period, would that person be identified as a case under the same eligibility and ascertainment rules?

The case and control sampling frames must correspond; the source need not be an entire country.

Vandenbroucke & Pearce (2012). doi:10.1093/ije/dys147

Sampling begins with outcome, not with recall



Exposure may come from interviews, records or samples collected before diagnosis. Asking about the past does not define the design. The defining feature is outcome-based sampling.

Vandenbroucke & Pearce (2012). doi:10.1093/ije/dys147

Cases need consistent definition and a defined period

Decision	Why it matters
New diagnoses	Avoid selecting only long-term survivors
Same disease	Do not silently combine oral-cavity and oropharyngeal cancers
Catchment and time	Define who can enter
Ascertainment	Referral and diagnostic access may affect inclusion

Survivor selection

If exposure affects survival, sampling surviving prevalent cases may distort its association with disease occurrence. A large sample of survivors does not solve this.

Vandenbroucke & Pearce (2012). doi:10.1093/ije/dys147

Controls need not be perfectly healthy

Possible sources

Population registers; underlying cohort; patients from the same system if their reason for attendance does not distort exposure sampling.

Convenience can bias exposure

Controls treated for tobacco-related conditions may have unusually high smoking exposure. Dental-clinic attenders can differ from non-attenders in income and behaviour.

Select controls for their relationship to the source population. Exposed people may be controls; excluding them builds the association into the design.

Vandenbroucke & Pearce (2012). doi:10.1093/ije/dys147

CONTENTS

1. The source population and sampling
2. **Odds and the odds ratio**
3. What the OR estimates
4. Selection, recall, matching and confounding
5. Independent analysis and critical appraisal

Build an unmatched case–control table

Invented study

100 cases and 200 independently sampled controls. Smoking is assessed over a predefined interval before a comparable index date. These are not real cancer estimates.

Smoking	Cases	Controls
Yes	80 = a	100 = b
No	20 = c	100 = d
Total	100	200

The investigator selected numbers of cases and controls. The sample case fraction is not population disease frequency.

Hypothetical teaching data / Условни учебни данни

Probability and odds have different denominators

Among	Exposure proportion	Exposure odds
Cases	80/100 = 0.80	80/20 = 4
Controls	100/200 = 0.50	100/100 = 1

Odds = p/(1 − p)

At p = 0.80: 0.80/0.20 = 4. Odds of 4 mean four exposed people per unexposed person, not probability 400%.

The proportion ratio 0.80/0.50 = 1.6 is not the odds ratio.

Hypothetical teaching data / Условни учебни данни

The cross-product ratio compares exposure odds

OR = (a/c)/(b/d) = ad/bc

$(80 \times 100)/(100 \times 20) = \mathbf{4}$.

Direct interpretation

Exposure odds were four times as high among sampled cases as among sampled controls. The population association estimated depends on how the groups were sampled.

OR = 1 is the null. Reversing one comparison gives 0.25; reversing both axes preserves the cross-product ratio.

Vandenbroucke & Pearce (2012). doi:10.1093/ije/dys147

Sample row totals do not give population risks

Tempting but invalid

80/(80+100) = 44.4% is the case fraction among sampled exposed people. It is not their population risk of cancer. Similarly 20/(20+100) = 16.7% is not the unexposed population risk.

Why?

The sample deliberately has 100 cases and 200 controls. Changing the control count changes these fractions without changing the disease process.

This ordinary sample table does not directly provide absolute risk, RD, NNT or population incidence.

Vandenbroucke & Pearce (2012). doi:10.1093/ije/dys147

Doubling controls preserves this OR

Exposure	Cases	Original controls	Doubled controls
Yes	80	100	200
No	20	100	200

Same control exposure distribution

OR = (80×200)/(200×20) = 4. The apparent exposed risk becomes 80/280 = 28.6%, instead of 44.4%. Neither is population risk.

Precision

A new random control sample may change OR by chance. More independent controls can improve precision, but cannot repair biased selection.

Hypothetical teaching data / Условни учебни данни

CONTENTS

1. The source population and sampling
2. Odds and the odds ratio
3. **What the OR estimates**
4. Selection, recall, matching and confounding
5. Independent analysis and critical appraisal

Rarity brings disease OR closer to RR

Population risks for illustration, not calculated from a case–control sample.

Exposed risk	Reference risk	RR	Disease OR
2%	1%	2	2.02
40%	20%	2	2.67

Calculate with odds

OR = $[R_e/(1 - R_e)]/[R_0/(1 - R_0)]$. At 40% versus 20%: $(0.4/0.6)/(0.2/0.8) = 2.67$.

With controls sampled among non-cases at the end, disease OR approximates RR when outcome risks are small. The sample case fraction does not define rarity.

Vandenbroucke & Pearce (2012). doi:10.1093/ije/dys147

Risk-set sampling has a different interpretation

Incidence-density sampling

At each case occurrence, sample controls from people still at risk then. A control may later become a case and may be sampled again under an appropriate scheme.

With valid risk-set sampling and appropriate analysis, OR estimates an incidence rate ratio without a rare-outcome assumption.

Identify the scheme

This does not turn every case–control OR into a risk ratio. Check time-specific sampling, matching and the analysis.

Vandenbroucke & Pearce (2012). doi:10.1093/ije/dys147

Nested studies may use prediagnostic measurements

Step	Within a defined cohort
Baseline	Collect smoking history and store samples
Follow-up	Identify incident cases
Sampling	Select eligible controls using a stated scheme
Analysis	Compare earlier exposure accounting for the scheme

Practical benefit

Testing expensive biomarkers only for selected cases and controls saves resources. Prediagnostic measurement helps temporality but does not automatically remove confounding or measurement error.

Vandenbroucke & Pearce (2012). doi:10.1093/ije/dys147

CONTENTS

1. The source population and sampling
2. Odds and the odds ratio
3. What the OR estimates
4. **Selection, recall, matching and confounding**
5. Independent analysis and critical appraisal

Selection can change OR without changing biology

Keep 80 exposed and 20 unexposed cases fixed; change only selected controls.

Controls	Original selection	Distorted selection
Exposed	100	150
Unexposed	100	50
OR	4	$80\times50/(150\times20) = 1.33$

Selection bias

When exposure affects control inclusion, its distribution may not represent the required population. Here excess exposed controls attenuate OR; other mechanisms can distort it in other directions.

Hypothetical teaching data / Условни учебни данни

Recall bias is differential exposure measurement

Invented example

True exposure: 40/100 cases and 40/100 controls $\longrightarrow$ OR $= 1$. Reported exposure: 60/100 cases and 40/100 controls $\longrightarrow$ OR $= (60\times60)/(40\times40) =$ **2.25**.

Response	**Limit**
Prediagnostic records	Completeness may differ
Same structured interview	Cannot guarantee equal recall
Interviewer masking	Does not erase knowledge of diagnosis
Comparable reference dates	Avoid lifetime versus current exposure

Information bias also includes recording and diagnostic suspicion. Its direction is not universally towards the null.

CDC, Field Epidemiology Manual, ch. 8 (2025)

Matching must be reflected in analysis

Purpose

Matching can improve comparability of selected factors and efficiency. It does not automatically eliminate confounding and is not randomisation.

Feature	Analysis consequence
Individual pairs	Preserve pairing and use a matched method
Frequency matching	Account for matching factors
Overmatching	Inappropriate factors may lose information or introduce bias

Unmatched shortcut

The cross-product OR from collapsed margins is not generally the matched-pair estimate. Discordant pairs matter.

CDC, Field Epidemiology Manual, ch. 8 (2025)

Confounding differs from selection and recall

Problem	Hypothetical example	Response
Selection	Controls attend a smoking-related clinic disproportionately	Defensible control frame; assess non-response
Information	Cases report smoking more fully	Comparable measurement and dates
Confounding	Age influences both smoking and cancer occurrence	Measure common causes; justified stratification or adjustment

Correcting one problem does not correct the others. A large OR alone cannot distinguish them from causation.

CDC, Field Epidemiology Manual, ch. 8 (2025)

An interval adds precision information, not bias protection

Original 80/20 and 100/100 table

OR = 4; approximate 95% CI **2.28–7.02**, using a large-sample log-OR method for independent unmatched observations.

Read estimate and design together

Report source population, control sampling, exposure period and adjustment. The interval describes sampling uncertainty under analysis assumptions.

Limits

A narrow interval around bias remains biased. Including 1 does not establish equivalence. Matched, clustered and sparse data may require other methods.

Greenland et al. (2016). doi:10.1007/s10654-016-0149-3

CONTENTS

1. The source population and sampling
2. Odds and the odds ratio
3. What the OR estimates
4. Selection, recall, matching and confounding
5. **Independent analysis and critical appraisal**

Group task: OR and a false risk

Hypothetical oral-cancer study

120 cases and 240 unmatched controls. Heavy drinking before comparable index dates: 72 cases and 60 controls report exposure.

Work independently

Fill all cells. Calculate OR. Explain why 72/(72+60) is not cancer risk. Double both control cells and recalculate OR. State one requirement for a valid population association.

Do not infer rarity from the 1:2 case–control ratio.

Hypothetical teaching data / Условни учебни данни

Group task: distinguish three threats

Scenario	Question
Controls volunteer from a smoking-cessation clinic	What is the selection problem?
Cases get a long prompted alcohol interview, controls a short question	Which measurement problem arises?
Heavy drinkers smoke more, and smoking contributes to cancer	Why might the alcohol OR be confounded?

For each scenario

Name the problem, explain distortion and propose a design or analysis response. Under risk-set sampling, must controls remain disease-free to the end?

Hypothetical teaching data / Условни учебни данни

Model answers follow the sampling scheme

Task	Answer
Table	72, 60, 48, 180; OR = (72×180)/(60×48) = 4.5
Doubled controls	OR = (72×360)/(120×48) = 4.5
False risk	72/132 = 54.5% is a selected-sample case fraction
Selection	Control exposure may not represent the source population
Information	Use comparable interviews and reference dates
Confounding	Measure smoking; use justified stratified or adjusted analysis
Risk sets	A control must be at risk when sampled; may later become a case

Hypothetical teaching data / Условни учебни данни

A defensible report names population and OR

Complete conclusion

In this hypothetical unmatched study, exposure odds were four times as high among cases as controls (OR 4, approximate 95% CI 2.28–7.02). Interpretation depends on control sampling, comparable ascertainment and confounding control.

Before accepting it	Check
Selection	Would controls have been identified as cases if they developed disease?
Measurement	Exposure precedes diagnosis; comparable reference periods
Analysis	Sampling, matching, confounding and uncertainty
Claim	OR differs from absolute risk and causation

Hypothetical teaching data / Условни учебни данни

Sources and further reading

Methods

CDC. Field Epidemiology Manual, chapter 8 (2025). Official online chapter; no chapter DOI.

Cochrane Handbook, chapter 15, version 6.5 (2024). Official online chapter; no separate chapter DOI.

Articles

Vandenbroucke JP, Pearce N. Int J Epidemiol. 2012;41:1480–1489. doi:10.1093/ije/dys147.

Greenland S et al. Eur J Epidemiol. 2016;31:337–350. doi:10.1007/s10654-016-0149-3.

All numerical datasets are hypothetical. Source checks and arithmetic: context/risk-assessment/sources.md.

Thank you for your attention!