Cohort studies and risk assessment

Applied epidemiology · 90-minute practical

Practical Class 11 · Dental Public Health · 2026/2027

Chief Assistant Professor Kostadin Kostadinov, MD, PhD

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

From design to a defensible risk estimate

Learning outcomes

Define population, event and time. Build a cohort table. Calculate risk, RR, risk difference and attributable measures. Interpret preventive effects, uncertainty, confounding and loss to follow-up.

A 90-minute practical: cohort assembly; comparisons; prevention and validity; group calculations.

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

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

Classes 8–12 form one epidemiological argument

Class	Question
8 · Frequency	How many people and over what time?
9 · Dental indices	How is disease measured consistently?
10 · Designs	How were comparison groups assembled?
11 · Cohort and risk	How large is the difference in new disease?
12 · Case–control	What do exposure odds tell us?

Apply the design logic from class 10 to calculate and scrutinise risk.

Course sequence / Последователност на курса

CONTENTS

1. **Assembling the cohort**
2. Relative and absolute comparisons
3. Prevention, follow-up and validity
4. Independent calculations and discussion

Define the event before choosing who is at risk

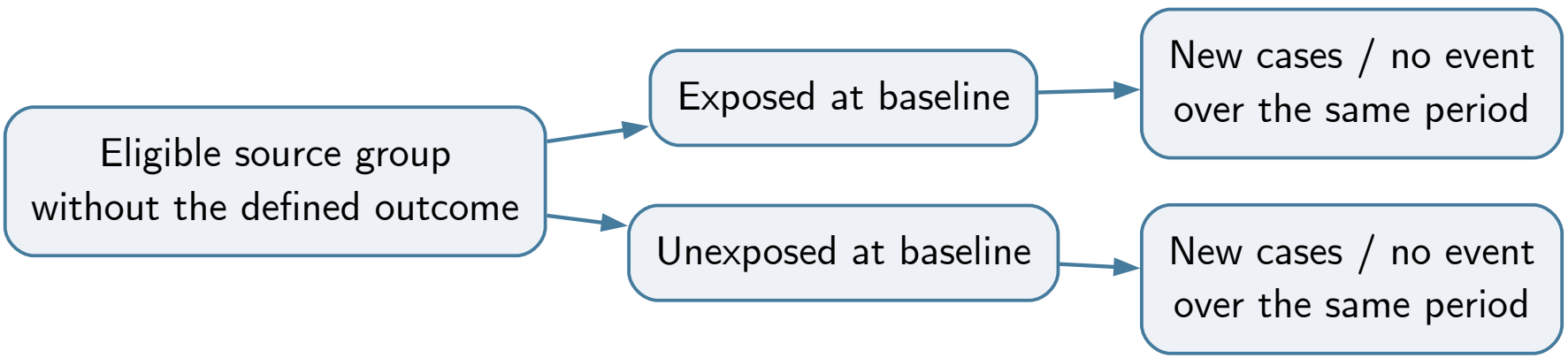
Hypothetical question

Among dentate adults initially without periodontitis, is baseline diabetes associated with first periodontitis over five years?

Element	Specify before follow-up
Population	Common eligibility and recruitment rules
Exposure	Diabetes at baseline, measured by a defined protocol
Event	First periodontitis at a prespecified diagnostic threshold
Time and unit	Five years; one person is one first case

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

A cohort starts with exposure groups



The researcher observes rather than assigns exposure. A historical cohort reconstructs this sequence from records; it does not start by sampling cases. Ascertain outcomes comparably.

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

Eligibility follows the endpoint

Event	Initially eligible	Different endpoint
First caries experience	DMFT = 0 at the chosen threshold	New lesion on an affected tooth
First periodontitis	People without that condition	Progression of existing disease
Becoming edentulous	At least one natural tooth	Losing one tooth while others remain

Keep the question

A progression study may include existing disease. Define its denominator for progression, rather than first onset.

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

Construct the five-year cohort table

Teaching assumptions

Independent adults, known five-year outcomes for everyone, and no competing deaths in this model. All numbers are invented, not estimates from real patients.

Baseline	New periodontitis	No new periodontitis	Total
Diabetes	60	240	300
No diabetes	30	570	600
Total	90	810	900

Row totals are the initially at-risk groups, not numbers selected after diagnosis.

Hypothetical teaching data / Условни учебни данни

CONTENTS

1. Assembling the cohort
2. **Relative and absolute comparisons**
3. Prevention, follow-up and validity
4. Independent calculations and discussion

Start with two risks and a time horizon

With diabetes

$60/300 = 0.20 =$ **20% over five years**.

Without diabetes

$30/600 = 0.05 =$ **5% over five years**.

Overall risk

$90/900 = 10\%$. The unweighted mean of 20% and 5% is 12.5%, which ignores group size. Twenty cases per 100 over five years is not twenty per year.

Hypothetical teaching data / Условни учебни данни

RR compares risks multiplicatively

$RR = R_e/R_0$

$0.20/0.05 = \mathbf{4}$. The diabetes group had four times the observed five-year risk.

RR	Meaning
1	Equal observed risks
Above 1	Higher risk among exposed people
Below 1	Lower risk among exposed people

Association

Four times the risk is a 300% relative increase. This crude association does not establish how much disease diabetes caused.

Hypothetical teaching data / Условни учебни данни

Risk difference expresses the absolute excess

RD = R_e − R_0

0.20 − 0.05 = **15 percentage points**: 15 additional cases per 100 over five years, comparing observed risks.

Scale	Result	Question
Relative	RR = 4	How many times as high?
Absolute	RD = 15 pp	How many additional cases per 100?

Reverse the reference

RR becomes 0.25; RD becomes −15 percentage points. Specify the group order.

Hypothetical teaching data / Условни учебни данни

Attributable fraction among exposed people: 75%

$AF_e = (R_e - R_0)/R_e$

$(0.20 - 0.05)/0.20 =$ **75%**. At the reference risk, 300 exposed people would have 15 rather than 60 cases; the contrast is 45/60.

Causal assumptions

Calling this preventable requires an unbiased causal comparison and a defined exposure change. Removing diabetes does not necessarily immediately restore the reference risk.

Terminology

RD is sometimes called attributable risk; AF is a proportion. State the formula.

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

Population impact depends on exposure frequency

Measure	Same hypothetical population
Population risk	90/900 = 10%
Reference risk	30/600 = 5%
Population risk difference	5 percentage points
PAF	(10% – 5%)/10% = 50%

Count check

900 × 5% = 45 reference-risk cases; 90 – 45 = 45 excess; 45/90 = 50%. AF among exposed people is 75%: a different denominator.

Scope

Crude contrasts. Causal PAF needs valid causal risks and a defined intervention. Do not transfer it to a population with different exposure frequency.

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

CONTENTS

1. Assembling the cohort
2. Relative and absolute comparisons
3. **Prevention, follow-up and validity**
4. Independent calculations and discussion

Prevention needs absolute and relative effects

New hypothetical trial

Individually randomised, independent children; all followed three years. Additional programme versus standard prevention. Endpoint: at least one new cavitated lesion.

Group	New cases / total	Risk
Additional programme	20/400	5%
Standard prevention	40/400	10%

Two effects

RR = 0.5; relative risk reduction = 1 − RR = **50%**. ARR = 0.10 − 0.05 = **5 percentage points**: five fewer affected children per 100 over three years.

Cochrane Handbook, ch. 15 (2024)

NNT includes an outcome, comparator and time

NNT = 1/ARR

1/0.05 = **20**. On average, 20 children need the additional rather than standard programme for three years to prevent one additional child developing the endpoint, if this effect applies.

Check	Reason
Proportions	1/0.05, not 1/5
Period	Three-year NNT is not annual NNT
Endpoint	Children with lesions differ from lesion counts
Uncertainty	Not a guarantee for each group of 20

Cochrane Handbook, ch. 15 (2024)

Equal RR can mean different absolute benefit

Setting	Standard risk	Programme risk	ARR	NNT
Lower baseline risk	2%	1%	1 pp	100
Higher baseline risk	20%	10%	10 pp	10

Assuming the same three-year RR = 0.5, absolute benefit depends on baseline risk.

Decision

Consider harms, acceptability, resources and equity. The table assumes transferable relative effects; it does not establish that assumption.

Cochrane Handbook, ch. 15 (2024)

Unequal follow-up calls for person-time

Group	New cases	Time at risk	Incidence rate
Exposed	18	450 PY	4 per 100 PY
Unexposed	12	600 PY	2 per 100 PY

Rate comparison

Rate ratio = (18/450)/(12/600) = **2**. Rate difference = **2 cases per 100 person-years**.

Units

This is not a two-year risk ratio. Time accrues until the event, loss or end; person-time does not automatically remove informative-loss bias.

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

A missing outcome is not disease-free

Hypothetical cohort

100 start; 80 have a known five-year outcome, including 8 cases; 20 outcomes are unknown.

Calculation	Meaning
$8/80 = 10\%$	Risk among known outcomes; may be selected
$8/100 = 8\%$	Minimum full-cohort risk if none missing are cases
$(8+20)/100 = 28\%$	Maximum if all missing are cases

Full-cohort risk is not identified: crude bounds are 8–28%. Assess reasons for loss and analysis assumptions.

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

Confounding can change a crude comparison

Separate invented cohort: first tooth loss over five years; everyone initially has natural teeth.

Age	Exposure +	Exposure −	RR
Younger	$10/100 = 10\%$	$40/400 = 10\%$	1
Older	$160/400 = 40\%$	$40/100 = 40\%$	1
All	$170/500 = 34\%$	$80/500 = 16\%$	2.125

Different age composition

The exposed group is older. In this model age affects exposure and outcome. Equal age weights give 25% in both groups. Stratification does not guarantee removal of other confounding.

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

Confidence intervals describe sampling uncertainty

Return to 60/300 versus 30/600

RR = 4; approximate 95% CI **2.64–6.06**, using a large-sample log-RR method for independent observations.

Interpretation

Under the statistical assumptions, the data are compatible with a range of risk ratios. A wide interval means limited precision.

Limits

It does not correct confounding, misclassification or selective loss; it does not contain 95% of individual risks. It is not a 95% probability that the fixed parameter lies in this particular interval.

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

CONTENTS

1. Assembling the cohort
2. Relative and absolute comparisons
3. Prevention, follow-up and validity
4. **Independent calculations and discussion**

Group task: analyse a new cohort

Task 1 · Hypothetical Plovdiv cohort

Five-year first periodontitis: 40/200 adults with diabetes and 48/600 without diabetes. Initially all are free of the endpoint; all outcomes are known.

Calculate and interpret

Build the full table. Calculate risks, RR, RD, AF among exposed people and PAF for these 800 people. Write one descriptive conclusion and one condition required for a causal interpretation.

Hypothetical teaching data / Условни учебни данни

Group tasks: prevention and missing outcomes

Task 2

Invented three-year individually randomised trial: endpoint in 15/300 programme children and 30/300 standard-care children. Calculate RR, ARR, relative risk reduction and NNT.

Task 3

Another cohort: 100 start, 80 outcomes are known, 8 are cases. Explain 8/80 versus 8/100 and derive the possible full-cohort risk range.

Discussion

Would a larger sample alone solve selective loss or confounding? Explain.

Hypothetical teaching data / Условни учебни данни

Model answers: denominators determine measures

Task	Checked answer
1 · Table	40 cases, 160 non-cases; 48 cases, 552 non-cases
1 · Comparisons	20% vs 8% over five years; RR = 2.5; RD = 12 pp
1 · Impact	AF = 60%; population risk = 88/800 = 11%; PAF = (11–8)/11 = 27.3%
2	5% vs 10%; RR = 0.5; ARR = 5 pp; relative reduction 50%; NNT = 20 over three years
3	10% among known outcomes; full-cohort bounds 8–28%

More observations improve precision under suitable sampling assumptions; they do not repair systematic distortion.

Hypothetical teaching data / Условни учебни данни

Report event, time, absolute contrast and uncertainty

Complete result

In this hypothetical cohort, five-year risks were 20% and 5%; crude RR 4 (approximate 95% CI 2.64–6.06), RD 15 percentage points. Assess confounding and missing outcomes before causal interpretation.

Measure	Always add
Risk	Population, event and period
RR / RD	Reference group, absolute risks and uncertainty
AF / PAF	Denominator and causal assumptions
NNT	Comparator, endpoint and time

Next: case–control sampling, where ordinary row totals do not give these risks.

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!