

Epidemiology, Biostatistics and Evidence-Based Medicine

Curated preclinical study track (not a medical degree, no licensure claim; synthetic datasets, fictional abstracts). Original explanations and practice only, following the reviewed course handout.

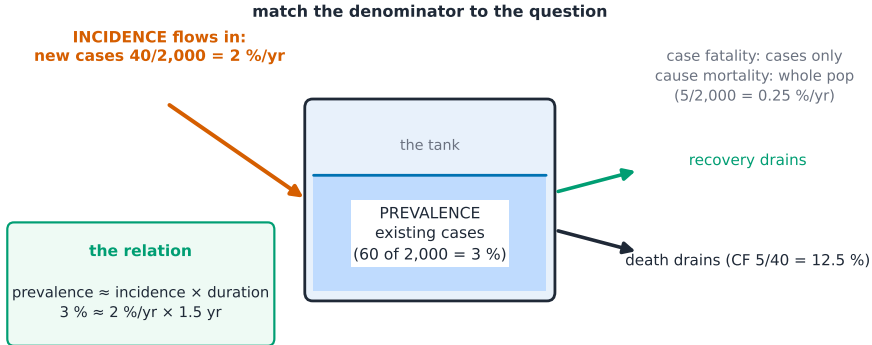
From questions to numbers

- 12.1-12.3 measure and compare; 12.4-12.6 quantify; 12.7 applies.
- Every dataset is synthetic; every abstract fictional.

Measuring disease: incidence and prevalence

- 1 **Flow Incidence** $40/2,000 = 2 \text{ \%/yr}$ (new cases, people at risk, time).
- 2 **Level Prevalence** $60/2,000 = 3 \text{ \%}$ on survey day; $P \approx I \times \text{duration}$ ($2 \text{ \%/yr} \times 1.5 \text{ yr}$).
- 3 **Denominators Case fatality** $5/40 = 12.5 \text{ \%}$ (cases); cause-specific $5/2,000 = 0.25 \text{ \%/yr}$ (population); standardise before comparing crude rates.

The stock-and-flow of cases

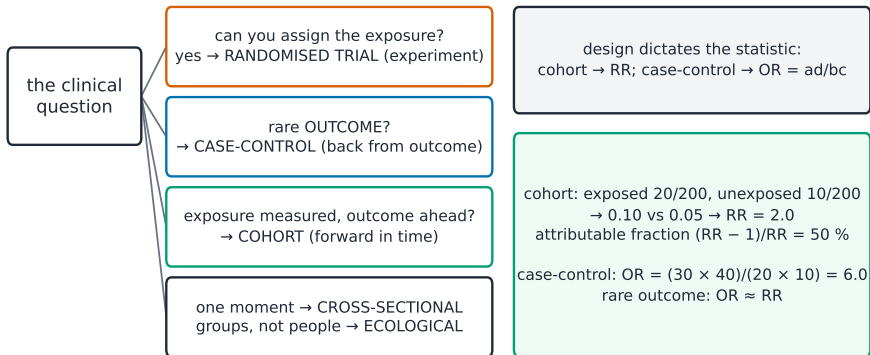


Incidence fills the tank; cure and mortality drain it - **prevalence** is the water level.

Study designs: the ladder and the statistic

- 1 **Ladder** Trial $\rightarrow$ cohort (by exposure, RR) $\rightarrow$ case-control (by outcome, OR = ad/bc) $\rightarrow$ cross-sectional $\rightarrow$ ecological.
- 2 **Worked** Cohort 20/200 vs 10/200 $\rightarrow$ 0.10/0.05 $\rightarrow$ RR 2.0, AF 50 %;
case-control 30 · 40/10 · 20 = OR 6.0.
- 3 **Why** Sampling by outcome makes arm risks artefacts; rare-disease OR $\approx$ RR.

The design decision tree



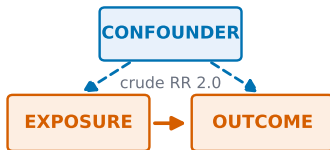
Rare outcome → case-control; exposure ahead → cohort; both at once → trial.

Bias, confounding and causation

- 1 **Failures** **Selection bias** (who counts), information/recall **bias** (what is remembered), **confounding** (third variable on both paths).
- 2 **Proof** Crude RR 2.0 collapsing to 1.0 in both strata = the third variable was the confounder.
- 3 **Causation** Hill's viewpoints with temporality first: strength, consistency, dose-response, plausibility; randomisation prevents confounding prospectively.

The confounding triangle and its cure

the triangle



a third variable on BOTH paths —
linked to exposure AND outcome

the stratification test

	exposed	unexposed	RR
crude	20/200	10/200	2.0
smokers	18/120	6/40	1.0
non-smokers	2/80	4/160	1.0

RR = 1.0 inside both strata: the exposed group
held more smokers (120 vs 40) — smoking confounded

bias vs confounding — different failures, different cures

SELECTION bias: who gets into the study
INFORMATION/recall bias: what is remembered
systematic error in design or measurement

confounding: a real third variable — fix by
randomisation (prospective), stratification,
or design; temporality is Hill's non-negotiable

The confounder sits on the back path - stratify and the spurious association collapses.

Diagnostic tests: columns vs rows

- 1 **Columns** $Se = a/(a+c)$, $Sp = d/(b+d)$ - 90 % each in the worked example.
- 2 **Rows** 10 % **prevalence**: $PPV = 90/180 = 50\%$, $NPV = 810/820 = 98.8\%$; at 1 %, $PPV = 9/108 = 8.3\%$.
- 3 **Portable** $LR+ = Se/(1-Sp) = 0.9/0.1 = 9$ - multiplies pre-test odds anywhere.

The 2×2 table and its natural-frequency tree

the 2×2 table

	diseased	healthy
test +	a = 90 TP	b = 90 FP
test -	c = 10 FN	d = 810 TN

Se = $90/100 = 90\%$ · Sp = $810/900 = 90\%$
columns: test properties

natural-frequency tree: 1,000 people, 10 % prevalence



rows: population properties

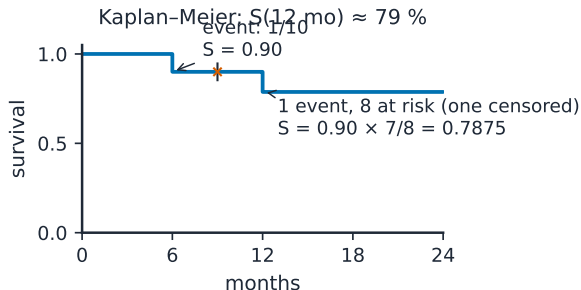
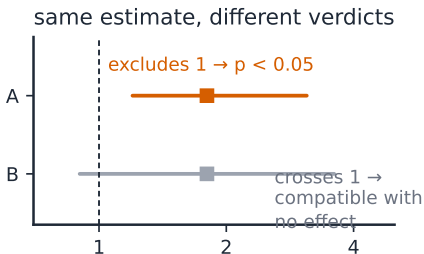
PPV $90/180 = 50\%$ · NPV $810/820 = 98.8\%$
at 1 % prevalence: PPV $9/108 = 8.3\%$
LR+ = $Se/(1-Sp) = 0.9/0.1 = 9$ — travels with you

Sensitivity and **specificity** live in the columns; **predictive values** live in the rows.

Statistical inference for clinical data: intervals and steps

- 1 Interval** RR 1.8 (1.2–3.1) excludes 1; RR 1.8 (0.9–3.6) is compatible with no effect - same estimate, different verdicts.
- 2 Tests** Counts χ^2 · two means t (paired if twice-measured) · several means ANOVA · skewed rank · correlation.
- 3 Survival** Kaplan–Meier multiplies $(1 - \text{events/at-risk})$: $S = 0.90$ then $0.90 \times 7/8 = 0.7875$; median = the 50 % crossing.

CI forest and Kaplan–Meier plots



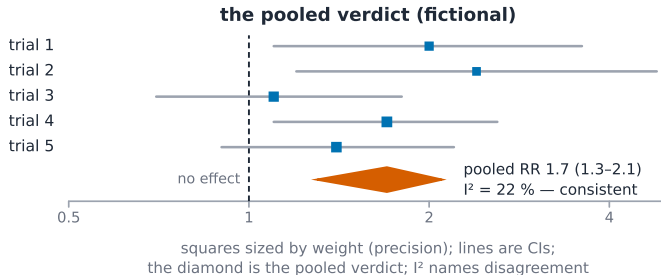
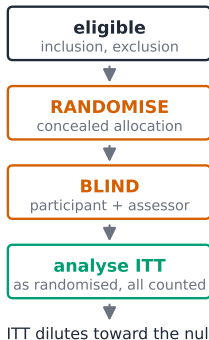
p-value: $P(\text{data this extreme} \mid \text{null})$ — surprise, not size · CI carries both · tests: χ^2 · t · ANOVA · rank · correlation

The interval carries size and precision; the survival curve steps down at events.

Clinical trials and systematic reviews: spine and pool

- 1 **Spine** Eligible → randomise (concealed) → blind → **intention to treat** (as randomised; dilutes toward null).
- 2 **NNT** Control 20 % vs treated 12 % → ARR 8 % → **NNT** 12.5 → 13; baseline 5 %, same RR → ARR 2 % → NNT 50.
- 3 **Pool** Forest: squares weighted by precision, diamond pooled, I^2 the disagreement gauge (22 % fine, 78 % investigate).

Trial flow into the forest plot



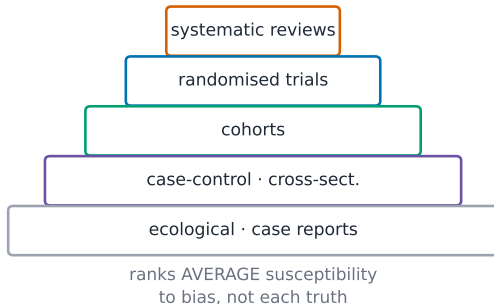
NNT = 1/ARR: 20 % vs 12 % → ARR 8 % → NNT 12.5 → 13
 baseline 5 %, same RR → ARR 2 % → NNT 50:
 relative effects travel, absolute ones are local

Randomise, blind, analyse as randomised - then pool into the **forest plot**.

Critical appraisal and applying evidence

- 1 **Spine** PICO frame; validity (**bias?**); size (RR 0.75, 0.60–0.94 in the fictional abstract); applicability (was my patient enrolled?).
- 2 **GRADE** Trials start high, observational low; downgrade for bias, inconsistency, indirectness, imprecision; upgrade for large effects or a dose-response gradient.
- 3 **Apply** Recompute absolutes: baseline 5 % with the same RR turns ARR 8 % into 2 % and **NNT** 13 into 50.

The evidence hierarchy with caveats



GRADE: the pyramid is only the starting point

randomised evidence starts HIGH, observational LOW

DOWNGRADE for

risk of bias · inconsistency · indirectness
imprecision · publication bias

UPGRADE for

large effect · dose-response gradient

certainty: high · moderate · low · very low
appraise: PICO → validity → size → applicability

The pyramid ranks average susceptibility to **bias**, not the truth of each study.

Check yourself

- 40 new cases in 2,000 people-year...? (2 % per year.)
- $Se = Sp = 90\%$ at 1 % **prevalence**: PPV...? (8.3 %.)
- $ARR\ 8\% \rightarrow$ **NNT**...? ($12.5 \rightarrow 13$.)
- The Kaplan–Meier rule after 1/8 at risk...? ($\times 7/8$.)