Epidemiology with epiR :: **CHEAT SHEET**

epiR turns count data into the core measures of epidemiology: frequency, association, test accuracy and sample size.

1 Build the 2×2 table

	Outcome +	Outcome −
Exposure +	**a**	**b**
Exposure −	**c**	**d**

Exposure in rows · outcome in columns · positives first. Cell a = exposed cases.

THREE WAYS TO FEED `epi.2by2()`

A Cell counts, in the order a, b, c, d

```
dat <- c(30, 44, 29, 86)
```

B A table() of two factors (one row per subject)

```
bw <- MASS::birthwt
bw$smk <- factor(bw$smoke, levels = c(1,0))
bw$lbw <- factor(bw$low, levels = c(1,0))
tab <- table(bw$smk, bw$lbw)
```

C A 3-D table: one 2×2 per stratum (see 4)

```
table(bw$smk, bw$lbw, bw$race)
```

> **Most common mistake.** table() sorts 0 before 1, so non-exposed and non-cases come first and every ratio is inverted. Set `levels = c(1, 0)` or reorder with `tab[2:1, 2:1]`.

Outcome in rows? Add `outcome = "as.rows"`. **Labels:** `elab` (exposure) and `olab` (outcome) name the rows and columns of the printout.

TIP: CHECK BEFORE YOU RUN

```
addmargins(tab)
#       1   0 Sum
#  1   30  44  74   <- a = exposed cases
#  0   29  86 115
#  Sum 59 130 189
```

Row and column 1 (the positives) must come first.

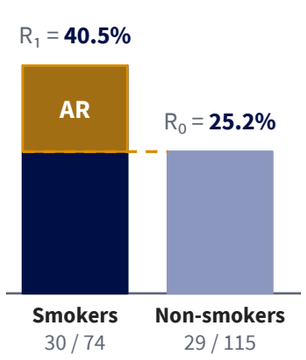
2 Pick the study design

```
epi.2by2(tab, method = "cohort.count")
```

Cohort counts `"cohort.count"`	Risk ratio · odds ratio · attributable risk · attributable fraction (exposed, population) · NNT · E-value
Cohort person-time `"cohort.time"`	Incidence rate ratio · attributable rate · attributable fraction. Enter c(a, time exposed, c, time unexposed)
Case-control `"case.control"`	Exposure odds ratio · estimated attributable fraction · E-value
Cross-sectional `"cross.sectional"`	Prevalence ratio · prevalence odds ratio · attributable prevalence & fraction · E-value

WHAT THE MEASURES MEAN

birthwt: smoking → low birth weight (LBW)



RR = $R_1 \div R_0$ = **1.61** *times the risk*

AR = $R_1 - R_0$ = **15.3** *extra cases per 100*

AF = AR ÷ R_1 = **38%** *of exposed cases*

OR = ad ÷ bc = **2.02** *odds, not risk*

NNH = 1 ÷ AR = **6.5** *exposed per extra case*

Ratios (RR, OR) measure the strength of an association; **impact measures** (AR, AF, NNT) show its public-health weight.

> **TIP** Cross-sectional data with a common outcome? Report the **prevalence ratio**; the prevalence odds ratio drifts away from 1 as prevalence rises.

3 Run it, read it

```
library(epiR)
res <- epi.2by2(tab,
  method = "cohort.count",
  elab = "Smoking", olab = "LBW",
  conf.level = 0.95, units = 100)
res
```

	estimate	95% CI
Inc risk ratio	**1.61**	(1.06, 2.44)
Inc odds ratio	**2.02**	(1.08, 3.78)
Attrib risk exp *	**15.32**	(1.61, 29.04)
Attrib fraction exp (%)	**37.80**	(5.38, 58.95)
Attrib risk pop *	**6.00**	(-4.33, 16.33)
Attrib fraction pop (%)	**19.22**	(10.89, 28.78)
E-value	**2.60**	(1.31, NA)

Uncorrected χ^2 p = 0.026 · Fisher exact p = 0.036
* per 100 population units (`units =`)

Printout abridged; labels shortened.

WRITE IT UP: `interpret = TRUE`

> *"The incidence risk of LBW among those that were smoking positive was 1.61 (95% CI 1.06 to 2.44) times the incidence risk of LBW among those that were smoking negative."*

PULL OUT THE NUMBERS

`res$massoc.summary`
Data frame: var, est, lower, upper.

`res$massoc.detail$RR.strata.score`
One measure with one CI method.

`res$massoc.detail$NNT.strata.wald`
Number needed to treat or harm (6.5 here).

`res$massoc.detail$chi2.strata.fisher`
Fisher exact test.

USEFUL ARGUMENTS

`digits = 2` decimals shown

`conf.level = 0.95` CI width

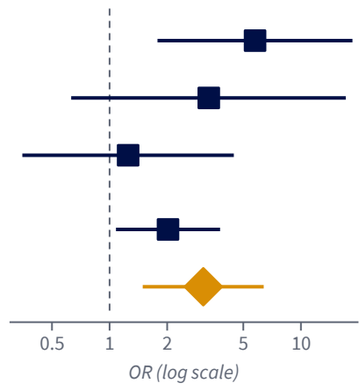
`units = 100` risk per 100 (or 1000…)

4 Adjust for confounding

Give a 3-D table (exposure × outcome × confounder) and get **crude**, **stratum-specific** and **Mantel–Haenszel (M-H) adjusted** estimates.

```
res3 <- epi.2by2(
  table(bw$smk, bw$lbw, bw$race),
  method = "cohort.count")
```

OR BY RACE (SMOKING → LBW)



Strata alike?
Homogeneity test: $\chi^2(2)$ = 2.80, p = 0.25. **Yes** → report M-H. **No** → effect modification: report each stratum.

Estimate moved?
Crude OR 2.02 vs M-H OR 3.09 → race confounds the association (rule of thumb: change > 10%).

How robust?
E-value (M-H) 3.73: a hidden confounder would need RR ≥ 3.73 with both smoking and LBW to explain the M-H risk ratio (2.15) away.

`res3$massoc.detail$OR.mh.wald`
`res3$massoc.detail$wOR.homog`

TIPS: WHICH MEASURE CAN I REPORT?

- **Case-control:** cases and controls are sampled by outcome, so risks can't be estimated. Report the **OR**, never the RR.
- **Rare outcome (< 10%):** OR ≈ RR. Common outcome: the OR exaggerates the RR (here 2.02 vs 1.61).
- **Person-time data:** use `"cohort.time"` and set `units` (e.g. 1000 person-years).

CC BY SA Anthony Romero-Cerdán • anthonyromero.netlify.app • Learn more at `browseVignettes("epiR")` • epiR version 2.0.99 • Updated: 2026-10

Epidemiology with epiR :: CHEAT SHEET

Most functions take counts or summary figures, not raw data, and return estimates with confidence intervals.

5 Measure frequency

`epi.conf()` takes a matrix of (cases, denominator) and returns the estimate with its CI. Multiply by your units.

Prevalence 26 of 200 surveyed

```
dat <- cbind(26, 200)
epi.conf(dat, ctype = "prevalence",
  method = "wilson") * 100
```

→ **13.0** (9.0 – 18.4) per 100

Incidence risk 12 new cases in 300 followed

```
epi.conf(cbind(12, 300),
  ctype = "inc.risk",
  method = "exact") * 100
```

→ **4.0** (2.1 – 6.9) per 100

Incidence rate 136 cases in 22 050 person-years

```
epi.conf(cbind(136, 22050),
  ctype = "inc.rate",
  method = "exact") * 1000
```

→ **6.17** (5.17 – 7.30) per 1000 PY

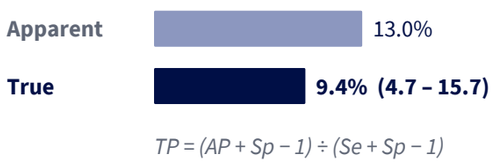
OTHER ctype = ...

`"mean.single"` `"prop.single"` `"odds"` `"smr"` and paired / unpaired versions.

APPARENT → TRUE PREVALENCE

A test with imperfect Se/Sp misstates prevalence. **`epi.prev()`** corrects it (Rogan–Gladen).

```
epi.prev(pos = 26, tested = 200,
  se = 0.90, sp = 0.95, units = 100)$tp
```



TP = (AP + Sp − 1) ÷ (Se + Sp − 1)

COMPARE WITH A STANDARD POPULATION

```
epi.smr(obs = 30, exp = 20, method = "byar")
```

→ **SMR 1.50** (1.01 – 2.14): 50% more deaths than expected

Adjusted rates: **`epi.directadj()`** (direct) · **`epi.indirectadj()`** (indirect).

6 Test accuracy

Test result in rows, true status (gold standard) in columns. **Positives first.**

	Disease +	Disease −
Test +	**a** 670	**b** 202
Test −	**c** 74	**d** 640

Se = a ÷ (a+c)
Sp = d ÷ (b+d)
PPV = a ÷ (a+b)
NPV = d ÷ (c+d)

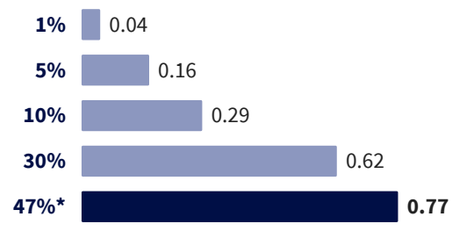
```
dx <- epi.tests(c(670, 202, 74, 640),
  method = "exact")
dx; dx$detail   # all statistics
```

Sensitivity	**0.90** (0.88, 0.92)	*detects the sick*
Specificity	**0.76** (0.73, 0.79)	*clears the healthy*
PPV	**0.77** (0.74, 0.80)	*P(D+ | T+)*
NPV	**0.90** (0.87, 0.92)	*P(D− | T−)*
LR+	**3.75** (3.32, 4.24)	*Se ÷ (1 − Sp)*
LR−	**0.13** (0.11, 0.16)	*(1 − Se) ÷ Sp*
Youden	**0.66** (0.61, 0.71)	*Se + Sp − 1*
Accuracy	**0.83** (0.81, 0.84)	*(a + d) ÷ n*

TIP Se and Sp belong to the test; **PPV and NPV depend on prevalence**. They only transfer to your patients if prevalence is similar. Otherwise, use the likelihood ratios (see →).

SAME TEST, DIFFERENT PREVALENCE

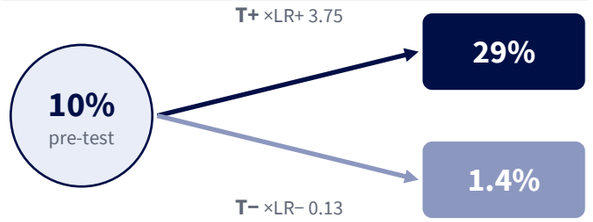
PPV of this test (Se 0.90, Sp 0.76) by prevalence, via Bayes' theorem:



** study prevalence (744 / 1586), as returned by epi.tests()*

FROM PRE-TEST TO POST-TEST PROBABILITY

```
epi.nomogram(pretest.ppos = 0.10,
  se = 0.90, sp = 0.76, verbose = TRUE)
```



Post-test odds = pre-test odds × LR. A positive result raises 10% to 29%; a negative one lowers it to 1.4%.

7 Measure agreement

Two raters (or tests) classifying the same subjects:

```
rat <- matrix(c(40, 5, 8, 47), nrow = 2,
              byrow = TRUE)
k <- epi.kappa(rat)
```

`k$prop.agree` observed 0.87 · expected 0.50

`k$kappa` κ = 0.74 (0.54 – 0.93)

`k$pabak` prevalence/bias-adjusted κ = 0.74

`k$mcnemar` p = 0.41 → no evidence of rater bias

poor	slight	fair	moderate	substantial	almost perfect
<0	≤0.2	≤0.4	≤0.6	≤0.8	≤1

Landis & Koch (1977) scale; κ = 0.74 is substantial agreement.

More than 2 categories? Pass a k × k matrix; epi.kappa() works the same.

CONTINUOUS MEASUREMENTS (LIN'S CCC)

```
x <- c(120, 135, 142, 150, 160, 171)
y <- c(122, 131, 145, 152, 158, 175)
epi.ccc(x, y)$rho.c
```

→ **ρc 0.98** (0.90 – 1.00) · also returns Bland–Altman data

TIP **Kappa paradox:** when one category is very common, κ can be low despite high observed agreement. Report κ together with **PABAK** and the observed agreement.

8 Plan the sample size

Set to NA what you want to compute.
`n = NA, power = 0.80` → sample size
`n = 300, power = NA` → power (0.82, case-control)

Prevalence survey 15% ± 20% (relative) — n = **545**

```
epi.sssimpleestb(Py = 0.15,
  epsilon = 0.20, error = "relative",
  se = 1, sp = 1)
```

Cross-sectional outcome 25% vs 15% — n = **500**

```
epi.ssxsectn(pdexp1 = 0.25,
  pdexp0 = 0.15, n = NA, power = 0.80)
```

Cohort (risks) risk 25% vs 15% — n = **500**

```
epi.sscohortc(irexp1 = 0.25,
  irexp0 = 0.15, n = NA, power = 0.80)
```

Case-control OR 2, 30% exposed controls — n = **282**

```
epi.sscc(OR = 2, p0 = 0.30,
  n = NA, power = 0.80, r = 1)
```

Two means (trial) difference 5, SD 10 — n = **126**

```
epi.sscompc(treat = 5, control = 0,
  sigma = 10, n = NA, power = 0.80)
```

`r` = ratio of groups (e.g. 2 controls per case: n = 320) · **`sided.test`** = 2 · **`conf.level`** = 0.95

CLUSTER SAMPLING

Subjects within a cluster are alike, so n grows by DEF = 1 + ρ(b − 1).

```
epi.ssclus1estb(b = 20, Py = 0.20,
  epsilon = 0.25, error = "relative",
  rho = 0.02)
```

→ **17 clusters × 20 = 340** (DEF 1.38)

epiR warns: at least 25 clusters are recommended.

TIP Expect losses? Divide n by (1 − loss rate): 545 ÷ 0.90 = 606 for 10% non-response.

CC BY SA Anthony Romero-Cerdán • anthonyromero.netlify.app • Learn more at `browseVignettes("epiR")` • epiR version 2.0.99 • Updated: 2026-10