Predictive values — PPV & NPV

Se and Sp condition on the **truth** and are read down the columns. PPV and NPV condition on the **test result** and are read across the rows. Bayes is the operation that turns one into the other, and prevalence is the only extra ingredient it needs.

BLOCK 01 Definitions

π $\pi = \Pr(D^+)$
Prevalence. Marginal — conditional on nothing. For one patient: the pre-test probability, the prior.

Se $\mathrm{Se} = \Pr(T^+ \mid D^+)$
Of the diseased, the share the test flags. Threshold-dependent. $1 - \mathrm{Se}$ = miss rate.

Sp $\mathrm{Sp} = \Pr(T^- \mid D^-)$
Of the healthy, the share correctly cleared. $1 - \mathrm{Sp}$ = false alarm rate — the number that does the damage.

PPV $\mathrm{PPV} = \Pr(D^+ \mid T^+)$
Given a positive, the chance of disease. **Not a property of the test** — moves with π.

NPV $\mathrm{NPV} = \Pr(D^- \mid T^-)$
Given a negative, the chance of health. The number you act on is the residual risk $1 - \mathrm{NPV}$.

LR $\mathrm{LR}^+ = \frac{\mathrm{Se}}{1-\mathrm{Sp}}$ $\mathrm{LR}^- = \frac{1-\mathrm{Se}}{\mathrm{Sp}}$
Prevalence-free, and the only one usable on an individual. $\mathrm{LR}^+ > 10$ or $\mathrm{LR}^- < 0.1$ is decisive.

BLOCK 02 One table, four quantities — and the numbers that make the point

	D^+	D^-	TOTAL
T^+	**TP**	**FP**	TP+FP
T^-	**FN**	**TN**	FN+TN
total	TP+FN	FP+TN	N

- **Down the columns** — the test: $\mathrm{Se} = \frac{\mathrm{TP}}{\mathrm{TP+FN}}$, $\mathrm{Sp} = \frac{\mathrm{TN}}{\mathrm{TN+FP}}$
- **Across the rows** — the result: $\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP+FP}}$, $\mathrm{NPV} = \frac{\mathrm{TN}}{\mathrm{TN+FN}}$
- Naming rule: the *second* word is what the test said, the *first* is whether it was right.

	ILL	HEALTHY	TOTAL
test +	**99**	**999**	1,098
test −	**1**	**98,901**	98,902
total	100	99,900	100,000

N = 100,000 π = 0.1% Se = 99% Sp = 99%
Se reads down: 99/100 = **99%**
PPV reads across: 99/1,098 = **9.0%**
Both true. Same table. A factor of eleven apart, because 1% of 99,900 healthy beats 99% of 100 ill.

BLOCK 03 Bayes, four ways — same statement, different instrument

COUNTS

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$

Reach for this first. Take a round cohort, split by prevalence, apply Se and $1 - \mathrm{Sp}$. Self-checking: the cells must sum to N.

EXPANDED

$$\mathrm{PPV} = \frac{\mathrm{Se} \cdot \pi}{\mathrm{Se} \cdot \pi + (1 - \mathrm{Sp}) \cdot (1 - \pi)} \qquad \mathrm{NPV} = \frac{\mathrm{Sp} \cdot (1 - \pi)}{\mathrm{Sp} \cdot (1 - \pi) + (1 - \mathrm{Se}) \cdot \pi}$$

When a paper gives Se, Sp and π but no counts. The denominator is just $\Pr(T^+)$, split into its two sources.

ODDS

$$\text{post-test odds} = \text{pre-test odds} \times \mathrm{LR} \qquad \mathrm{PPV} = \frac{\text{odds}}{1 + \text{odds}}$$

The form that generalises: separates prior from evidence, extends to multi-level results, and chains across tests — but only under conditional independence.

WORK-UPS

$$\frac{\mathrm{FP}}{\mathrm{TP}} = \frac{1}{\text{post-test odds}} \qquad \frac{1}{\mathrm{PPV}} = 1 + \frac{\mathrm{FP}}{\mathrm{TP}}$$

$1/\mathrm{PPV}$ = people worked up per true case found. PPV 9% means eleven biopsies per cancer.

BLOCK 04 Limits, approximations, and the specificity lever

- $\pi \to 0$: $\mathrm{PPV} \to 0$, $\mathrm{NPV} \to 1$
- Rare disease: $\mathrm{PPV} \approx \pi \cdot \mathrm{LR}^+$ (upper bound)
- Always-negative rule scores $\mathrm{NPV} = 1 - \pi$ — **a high NPV in screening means nothing.** Judge rule-out by LR^-.
- Accuracy $= \mathrm{Se} \cdot \pi + \mathrm{Sp} \cdot (1 - \pi) \to \mathrm{Sp}$. **Useless for screening.**

AT Π = 0.1%	SE	SP	PPV
baseline	90%	99%	8.3%
better sensitivity	99%	99%	9.0%
better specificity	90%	99.9%	**47%**

When disease is rare, **PPV is a specificity problem**. Improving sensitivity buys almost nothing. This is why screening runs in two stages: a sensitive first pass, then a highly specific confirmatory test whose whole job is to kill the false positives the first stage produced.

BLOCK 05 Where the three inputs actually come from

Se and Sp — a diagnostic accuracy study (STARD; bias graded by QUADAS-2). Best to worst: Cochrane DTA meta-analysis (pooled *jointly*, bivariate/HSROC — separate pooling is wrong) › primary study with a stated sampling frame › regulatory summary (IVDR SSP / EUDAMED, FDA 510(k)) › the manufacturer's insert (a ceiling).

- **Spectrum effect.** Sicker sample → higher Se. Tertiary figures do not transport to primary care.
- **Two-gate (case–control).** Cases vs healthy volunteers. Inflates Se and Sp together; its PPV is meaningless.
- **Partial verification.** Only positives get the gold standard → Se inflated, Sp deflated.
- **Imperfect standard.** Usually deflates the index test — unless the errors correlate, then it flatters it. Never accept discrepant analysis.
- **Unstated threshold.** Se/Sp are one point on an ROC curve. A cutoff picked by Youden on the same data is optimistic.

Prevalence — the input people replace with a guess, and the one that moves the answer most. Worst to best: population registry › setting-specific rate (your own lab: confirmed cases ÷ people tested, over a full year) › a validated clinical prediction rule (Wells, Centor, HEART) › never gestalt as the number itself.

- **Referral filter.** Every stage of triage enriches. The prior depends on where you are standing.
- **Time and place.** Season, outbreak phase, region. December's policy can be wrong in June.
- **Screening vs diagnosis.** The distinction that dominates this whole sheet: same assay, π one or two orders of magnitude apart, completely different meaning of a positive.

Ask the right question. Not "how common is this disease" but "of the people who present here, in this way, what fraction turn out to have it".

BLOCK 06 Typical values — orders of magnitude, not constants

TEST	SETTING	SE %	SP %	Π %	PPV	NPV %
HIV, 4th-gen immunoassay	screening, low-risk	99.9	99.8	0.10	33%	100.0
HIV, 4th-gen immunoassay	high-risk group	99.9	99.8	5.0	96%	100.0
Mammography, 50–69	population screening	85.0	90.0	0.80	6%	99.9
Faecal immunochemical	colorectal screening	79.0	94.0	0.40	5%	99.9
SARS-CoV-2 antigen	symptomatic	72.0	99.5	15.0	96%	95.3
SARS-CoV-2 antigen	asymptomatic screen	58.0	99.5	0.30	26%	99.9
D-dimer	suspected PE	97.0	40.0	15.0	22%	98.7
PSA > 4 ng/mL	prostate cancer	21.0	91.0	3.0	7%	97.4
Rapid strep antigen	adult sore throat	86.0	96.0	10.0	70%	98.4

- **Same assay, two settings, two worlds.** HIV immunoassay: PPV 33% screening the low-risk, 96% in a high-risk group. Nothing about the test changed.
- **A low PPV does not condemn a test.** D-dimer: Sp 40%, PPV 22% — indefensible as a rule-in test. But $\mathrm{LR}^- = 0.075$ drops a 15% pre-test probability of PE to 1.3%, below the treatment threshold. It does its job, which is rule-out. **Judge a test against the decision it is meant to support.**

BLOCK 07 Errors that survive into practice

"The test is 99% accurate, so a positive means you almost certainly have it."	Confuses $\Pr(T^+ \mid D^+)$ with $\Pr(D^+ \mid T^+)$ — the inverse fallacy. The answer needs prevalence, which the sentence never mentions.
Reading PPV off a case–control table.	The row totals encode the investigator's sampling ratio, not a prevalence. Extract Se and Sp; recompute PPV with *your* prevalence.
Applying published Se and Sp unchanged to a new setting.	Only if the case mix matches. Spectrum shifts with care level.
"NPV is 99.8% — excellent at ruling out."	At low prevalence a rock that always says "negative" scores nearly as well. Use LR^-.
Quoting accuracy for a screening test.	Accuracy ≈ specificity when disease is rare. It is blind to the errors that matter.
Chaining two tests by multiplying their LRs.	Valid only under conditional independence given D. Related biology ⇒ correlated errors ⇒ the second test adds less than advertised.
Reporting Se and Sp without the cutoff.	They are one point on a curve. Give the threshold, or the curve and its AUC.

BLOCK 08 The method that survives exam pressure

```
π = 1 in 1000 · Se = 99% · Sp = 99%
Cohort 100,000 → ill 100, healthy 99,900
Apply Se to the ill → TP 99, FN 1
Apply 1−Sp to the healthy → FP 999, TN 98,901
Check 99 + 1 + 999 + 98,901 = 100,000 ✓
PPV = 99 / 1,098 = 9.0%   NPV = 99.999%
```

Do not substitute into Bayes under time pressure. Pick a round cohort, push it through the two splits, read the answer off the table. Faster, self-checking, far harder to get wrong.

Cross-check with the odds form: pre-test odds ≈ 0.001, $\mathrm{LR}^+ = 99$, post-test odds ≈ 0.099, and 0.099 / 1.099 = 9.0%. The two routes must agree; if they do not, the arithmetic is wrong, not the theory.