

A Test That Is Ninety-Nine Per Cent Accurate Can Be Wrong Nine Times in Ten

— With sensitivity and specificity both **0.99**, the predictive value is **0.990000** at a prevalence of **0.5** but **0.090164** at **0.001** — it falls to a half at a prevalence of **0.010000000**, matching the closed form — raising specificity to **0.999** gives **0.497738** while raising sensitivity gives only **0.090909** — [Paper 470]

Yuuki YAMAGISHI (山岸 勇輝)
Yamagishi Complexity Torus Lab, Japan
September 2026

— *From thinking to syncing.* —

Abstract

A test of high accuracy is thought to make a positive result trustworthy. This paper shows that the predictive value of one test runs from **0.09** to **0.99** with prevalence—and that what to improve is specificity, not sensitivity. No new theorem or law is claimed. First, at a prevalence of 0.5 the predictive value is exactly **0.990000000000**, equal to the accuracy (Section 2). Second, at a prevalence of 0.001 it falls to **0.090164** (Section 2). Third, this is the core of the paper. It reaches a half at a prevalence of **0.010000000**, matching the closed form $(1 - sp) / ((1 - sp) + se)$ (Section 3). Fourth, raising specificity to 0.999 gives **0.497738**, a factor of **5.4750** (Section 3). Fifth, raising sensitivity to 0.999 gives only **0.090909** (Section 3). Sixth, the separator is whether the quantity belongs to the test or to the test paired with a population (Section 4). **Keywords:** Bayes' theorem, positive predictive value, sensitivity, specificity, prevalence.

Scope of this paper (scope note): No new theorem or law is claimed—Bayes’ theorem, sensitivity and specificity, and positive predictive value are all standard. No medical advice is given—which test to take and how to read a result are not the subject; clinical judgement belongs to professionals. No particular test is discussed—sensitivity and specificity of 0.99 are illustrative, not the performance of any real test. Test sequences are excluded—two-stage testing, confirmatory tests and independence assumptions are not the subject. Cost and harm are excluded—the weight of follow-up and anxiety from a false positive is not the subject. Relation to earlier papers: Paper 469 showed that the same ten per cent differs ninefold with the starting point—there too, the effect of an operation was set by the state. Paper 300 showed that common root and distinct root can be decided—“the accuracy of a test” and “the reliability of a positive” are distinct roots by that criterion. Paper 460 showed that the limit of heat is set by two quantities together—there too, one number alone did not settle an ordering. What is added is tabulating the predictive value at four prevalences, solving for the half-way prevalence as 0.010000000 by bisection and closed form, setting the effect of specificity against sensitivity as 0.497738 to 0.090909, giving the ratio as 5.4750, and placing the separator at test versus test-and-population.

1. Notation and the Question

Setting: For a test of sensitivity se and specificity sp applied at prevalence p , the probability that a positive is genuine follows from Bayes’ theorem as

$$PPV = \frac{p se}{p se + (1 - p)(1 - sp)}$$

Below, $se = sp = 0.99$ is the baseline.

Observation 1.1

A test of high accuracy is thought to make a positive trustworthy. A test that is right 99 per cent of the time—should surely be right 99 per cent of the time.

Observation 1.2

But the prevalence p appears in the expression. The test alone does not settle the predictive value—who it is used on holds half the answer.

Question: By how much does it move.

2. From 0.09 to 0.99 with Prevalence

Computation 2.1: Take the predictive value at four prevalences.

Prevalence	Positive predictive value
0.5000	0.990000
0.1000	0.916667
0.0100	0.500000
0.0010	0.090164

Observation 2.2

At a prevalence of 0.5 it is exactly **0.990000000000**.Equal to the sensitivity and specificity—look only here and “accuracy equals predictive value” seems true.

Observation 2.3

At a prevalence of 0.001 it is **0.090164**.About nine positives in ten are wrong—with the same test and the same accuracy.

Observation 2.4

The span is 0.09 to 0.99, a factor of 11.Without changing the test at all—“ninety-nine per cent accurate” does not carry that span.

Observation 2.5

Because the healthy vastly outnumber the sick.At a prevalence of 0.001, one in a thousand is ill and 999 are well—and one per cent of false positives makes about ten of those 999 test positive.

3. What to Improve Is Specificity, Not Sensitivity

Computation 3.1: Solve for the prevalence at which the predictive value is a half, by bisection and by closed form.

0.010000000 (agreeing with $(1 - sp)/((1 - sp) + se)$ to within 10^{-9} relative).

Observation 3.2

A prevalence of one per cent is the break-even line. Below it, a positive is more likely wrong than right—and many screened conditions lie below it.

Computation 3.3: At a prevalence of 0.001, raise each of sensitivity and specificity to 0.999.

Condition	Predictive value
se = 0.99, sp = 0.99	0.090164
se = 0.999, sp = 0.99	0.090909
se = 0.99, sp = 0.999	0.497738

Observation 3.4

Raising sensitivity moves it from 0.090164 to 0.090909. An improvement of 0.8 per cent—catching more of the sick barely moves the predictive value.

Observation 3.5

Raising specificity jumps it to **0.497738**. A factor of **5.4750**—the same operation, “take 0.99 to 0.999”, with wholly different effect.

Observation 3.6

Because false positives scale with the number of healthy people. They outnumber the sick 999 to one, so their errors count 999 times—cut the larger denominator.

Observation 3.7

So “reduce the misses” does not raise the predictive value for a rare condition. It points the wrong way—what must fall is the wrongful catching of the healthy.

4. The Separator

The separator: does the quantity belong to the test, or to the test paired with a population.

Observation 4.1

To the test, and one number suffices. Sensitivity and specificity are set by the test alone—quantities that go on a datasheet.

Observation 4.2

To the pair, and it changes with who is tested. From 0.09 to 0.99—the same test becoming a different thing by changing the population.

Observation 4.3

These are distinct roots by the criterion of Paper 300. Both are called “how often the test is right”—one a property of the test, one of the test paired with a population.

Observation 4.4

The same shape as Paper 460. There, heat was set by temperature and humidity together and one number gave no ordering—here the predictive value is set by test and prevalence together.

Observation 4.5

It overlaps Paper 469 as well. There, the same ten per cent differed ninefold with the starting point—here the same “0.99 to 0.999” differs by 5.47. In both, the state sets the effect of the operation.

Observation 4.6

So state the population whenever test performance is quoted. “Ninety-nine per cent accurate” does not say how to read a positive—the same number matching both 0.09 and 0.99.

Observation 4.7

To be honest—only a single test has been treated. In practice a confirmatory second stage is used, and the predictive value rises greatly there—this paper does not design that, but this calculation is why confirmation is needed.

5. Roll Call and Closing

1. **0.990000000000 at prevalence 0.5 (Observation 2.2): equal to the accuracy.**
2. **0.090164 at 0.001 (Computation 2.1): nine in ten wrong.**
3. **A span of 11 (Observation 2.4): the test unchanged.**
4. **The healthy outnumber 999 to one (Observation 2.5): false positives pile up.**
5. **Break-even at prevalence 0.010000000 (Computation 3.1): the core.**
6. **Matching the closed form (Computation 3.1): checked by bisection.**
7. **Sensitivity gives 0.090909 (Computation 3.3): a 0.8 % gain.**
8. **Specificity gives 0.497738 (Computation 3.3): a factor of 5.4750.**
9. **The separator is test or test-and-population (Observation 4.3): distinct roots by Paper 300.**
10. **Why confirmation is needed (Observation 4.7): this calculation explains it.**

Closing: A test of high accuracy is thought to make a positive trustworthy—a test right 99 per cent of the time should surely be right 99 per cent of the time. But the prevalence appears in the expression: the test alone does not settle the predictive value, and who it is used on holds half the answer. The predictive value of one test moves elevenfold—at a prevalence of 0.5 it is exactly **0.990000000000**, equal to the accuracy, and looking only there

makes “accuracy equals predictive value” seem true. At a prevalence of 0.001 it is **0.090164**: about nine positives in ten are wrong, with the same test and the same accuracy. Because the healthy vastly outnumber the sick—one in a thousand ill against 999 well, so one per cent of false positives makes about ten of those 999 test positive. And what to improve is specificity, not sensitivity—the predictive value reaches a half at a prevalence of **0.010000000**, matching $(1 - sp)/((1 - sp) + se)$ to within 10^{-9} relative. One per cent is the break-even line, and many screened conditions lie below it. At a prevalence of 0.001, raising sensitivity to 0.999 moves the value only from 0.090164 to **0.090909**, a gain of 0.8 per cent. Raising specificity to 0.999 jumps it to **0.497738**, a factor of **5.4750**—the same operation with wholly different effect. Because false positives scale with the number of healthy people, who outnumber the sick 999 to one, so their errors count 999 times: cut the larger denominator. So “reduce the misses” points the wrong way for a rare condition; what must fall is the wrongful catching of the healthy. One thing separates them—does the quantity belong to the test, or to the test paired with a population. To the test, and one number suffices: sensitivity and specificity go on a datasheet. To the pair, and it runs from 0.09 to 0.99, the same test becoming a different thing by changing the population. They are distinct roots by the criterion of Paper 300, one a property of the test and one of the test paired with a population. It is the same shape as Paper 460, where heat was set by temperature and humidity together. It overlaps Paper 469 too, where the same ten per cent differed ninefold with the starting point: in both, the state sets the effect of the operation. So state the population whenever test performance is quoted—“ninety-nine per cent accurate” matches both 0.09 and 0.99. To be honest—only a single test has been treated. In practice a confirmatory second stage is used and the predictive value rises greatly there. This paper does not design that, but this calculation is why confirmation is needed.

References

1. Bayes, T. (1763). An Essay towards Solving a Problem in the Doctrine of Chances. *Philosophical Transactions of the Royal Society*.
2. Yerushalmy, J. (1947). Statistical Problems in Assessing Methods of Medical Diagnosis. *Public Health Reports*.
3. Gigerenzer, G., & Hoffrage, U. (1995). How to Improve Bayesian Reasoning Without Instruction. *Psychological Review*.
4. Yamagishi, Y. (2026). *The Three Hundredth Census: Whether Two Things Share a Root Is Decidable* [Paper 300]. Zenodo. 10.5281/zenodo.22171914
5. Yamagishi, Y. (2026). *The Limit of Heat Is Set by Temperature and Humidity Together* [Paper 460]. Zenodo. 10.5281/zenodo.22733136
6. Yamagishi, Y. (2026). *What Ten Per Cent More Traffic Costs Depends on Where You Already Are* [Paper 469]. Zenodo. 10.5281/zenodo.22733401

— *From thinking to syncing.* —

Yuuki YAMAGISHI (山岸 勇輝)