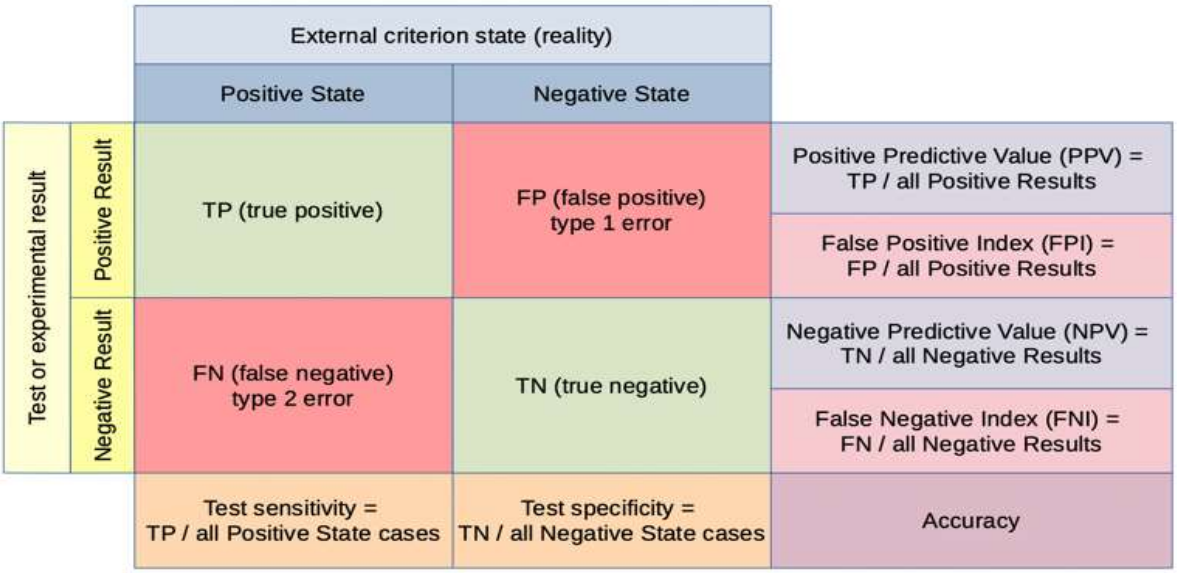
Five Minute Science Lesson: Test Accuracy Metrics
Raymond Nelson
APA MAGAZINE
THE MAGAZINE FOR THE POLYGRAPH PROFESSIONAL
MAY• JUN Vol. 53.3

Conceptual terms for understanding accuracy



Criterion state (external criterion): is any phenomena of interest that we wish to quantify, classify or predict. This is sometime referred to as an *unknown parameter* or *unknown phenomena* of interest.

Test result: often refers to the categorical test result. Scientific test results are not discussed in terms of pass or *fail* – though in some contexts there may be a tendency towards practical interpretations at this level – but instead use the terms *positive* and *negative* to signify whether the likelihood
is sufficient to support a categorical conclusion about the presence or absence of the unknown parameter.

Positive result: is a term that signifies when a probabilistic test result supports a categorical conclusion that the unknown parameter or unknown phenomena of interest is *present* in a case.

Negative result: signifies when a probabilistic test result supports a categorical conclusion that the unknown parameter or unknown phenomena of interest is *absent* in a case.

Positive state: refers to whether the unknown parameter or unknown phenomena of interest is actually *present* (in reality) for a case.

Negative state: refers to whether the unknown parameter or unknown phenomena of interest is actually *absent* (in reality) for a case.

True positive (TP): describes a *positive result* that concurs with a *positive state.* A test has *correctly* identified the *presence* of the unknown phenomena of interest for a case.

True negative (TN): describes a *negative result* that concurs with a *negative state.* A test has *correctly* identified the *absence* of the unknown phenomena of interest for a case.

False positive (FP): describes a p*ositive result* that concurs with a *negative* state. A test has *incorrectly* identified the *presence* of the unknown phenomena in a case for which the external criterion state (reality) is actually negative (the unknown phenomena of interest is actually *absent*). FP is sometimes calculated as 1-specificity. However, the 1-specificity calculation will be incorrect for tests that include the use of inconclusive classifications (when a probabilistic result is not statically significant for either positive or negative classification).

False negative (FN): describes a negative result that concurs with a *negative state*. A test has *incorrectly* identified the *presence* of the unknown phenomena in a case for which the external criterion state (reality) is actually positive (the unknown phenomena of interest is actually *present*). FP is sometimes calculated as 1-sensitivity. However, the 1-sensitivity calculation will be incorrect for tests that include the use of inconclusive classifications (when a test result is neither positive nor negative).

Positive predictive value (PPV): refers to the proportion of TP outcomes to (TP + FP) outcomes for a group of cases for which the actual positive state or negative state is known. (Data scientists sometimes refer to these as *labelled cases*.) Can be useful to estimate the likelihood that a positive result is *correct* for an unknown case. However, PPV is non-resistant to differences in the prior proportion of positive state and negative state cases (base rate or incidence rate). That is, PPV – the likelihood that a positive result is correct – will be a function of the proportion of positive state cases in the group of cases, in addition to being influenced by the test sensitivity and FP rates.

Negative predictive value (NPV): refers to the proportion of TN outcomes to (TN + FN) outcomes for a group of labelled cases (i.e., cases for which the actual positive state or negative state is known). Can be useful to estimate the likelihood that a negative result is *correct* for an unknown case. NPV is also non-resistant to differences in the prior proportion of positive state and negative state cases (base rate or incidence rate). NPV – the likelihood that a negative result is correct – will vary with both the test specificity rate and the proportion of positive state cases (i.e., prior probability, base rate or incidence rate) in a group of cases.

False positive index (FPI): refers to the proportion of FP outcomes to all positive (FP + TP) outcomes for a group of labelled cases. Can be useful to estimate the likelihood that a positive result is *incorrect* for an unknown case. FPI is non-resistant to group imbalance (i.e., differences in the proportion of positive state and negative state cases), and will vary with both the prior base rate or incidence rate and the test sensitivity rate.

False negative index (FNI): refers to the proportion of FN outcomes to all negative (FN + TN) outcomes for a group of labelled cases. Can be useful to estimate the likelihood that a negative result is *incorrect* for an unknown case. FNI is non-resistant to group imbalance. FNI will vary with both the test specificity rate and also will differences in the proportion of negative state and positive state cases (prior base rate or incidence rate).

Sensitivity: refers to the proportion of a group of positive state cases for which a test can *correctly* identify the *presence* of the unknown phenomena of interest. Test sensitivity, because it is calculated only within the subgroup of positive state cases, is resistant to differences in prior incidence rate. That is, test sensitivity will be invariant to group imbalance.

Specificity: refers to the proportion of a group of negative state cases for which a test can *correctly* identify the *absence* of the unknown phenomena of interest. Test specificity is calculated within the subgroup of negative state cases and is resistant to differences group imbalance (prior probability, base rate or incidence rate).