Considering Base Rates in Testing Populations

Overview

The Base Rate of Guilt (BRG) is the percentage of people in a group to be tested that are likely to be guilty of the target behaviors covered by the test. That rate is never truly known, unless you are conducting a lab study and have that element controlled.

The Base Rate of Failure (BRF) is the percentage of people who fail a given test. Because of False Positive (FP) results, the group of people that fail a test can include innocent people. It will also include people that are guilty.

Most organizations can make good estimates of the BRF if they consider the number of previous failed background checks, interview data, etc. In addition, there are organizations that publish average failure rates or recidivism rates for specific roles and crimes, and in specific industries.

If a test is accurate at classifying innocent and guilty people, then the BRG and BRF will be similar numbers. However, because all testing methods render false positive (FP) and false negative (FN) results, the two rates are likely not to be equal.

For example, law enforcement applicants in the U.S. (consisting of recent college graduates) will generally fail a Law Enforcement Pre-Employment Test (LEPET) at about 35%. The typical LEPET would address drug use, commission of a serious crime, and work-related discipline. If 35% of the applicants are likely to fail the test, the BRF is 35%.

It is possible to estimate the BRG using historical BRF plus additional estimates of test accuracy and error rates with innocent and guilty people.

In a police investigation, the BRG might be higher because the police try to find and test those that are highly likely to be guilty. When testing average office workers for stealing at a previous employer, the BRG might be lower, say 25%.

Making a good estimate of the BRG is important when conducting credibility assessment testing. This document will explain that concept.

Why BRG Matters

The following explains what can happen in test results when a testing method does not properly consider in its decision model (algorithm) the BRG of the group being tested.

All tests, such as EyeDetect or polygraph, have published rates of accuracy and error rates. EyeDetect was originally researched and designed in a lab study. In a lab study, the base rate of guilt is approximately 50% (one-half of the participants are guilty and one-half are innocent).

When a testing solution is established to create test scores, the examiner or solution can either make an adjustment for the BRG of the group to be tested or do nothing.

The following shows what happens to test results when no adjustment is made in the testing solution's accuracy and error rate to physiological or cognitive reactions.

<u>Assumptions</u>
EyeDetect accuracy:

True Negative (TN) 89%	True Positive (TP) 83%
False Negative (FN) 17%	False Positive (FP) 11%

Sample size (N) = 100 people tested

Base Rate of Guilt (BRG) + Base Rate of Innocent (BRI) = 1. Thus, BRI = 1 - BRG.

(Some rounding has been done to make numbers whole.)

Formulas used:

True Negative (TN) = 100 x 0.89 x (1- BRG)	True Positive (TP) = 100 x 0.83 x BRG
False Negative (FN) = 100 x 0.17 x (BRG)	False Positive (FP) = 100 x 0.11 x (1- BRG)

Note: Most groups commonly tested have BRG below 50%, but the first example assumes a 50/50 split.

<u>Example: 50 examinees are innocent and 50 are guilty</u> (BRG 50%)

- Pass
    - TN: 44.5 of 50 - innocent and pass (89%)
    - FN: 8.5 of 50 - guilty but pass (17%)
    - 53.0 total
- Fail
    - TP: 41.5 of 50 - guilty and fail (83%)
    - FP: 5.5 of 50 - innocent but fail (11%)
    - 47.0 total

If a person passes the test, the probability he is innocent is 44.5/53.0 = 84%, and if a person fails the test, the probability he is guilty is 41.5/47.0 = 88.3%.

However, the following examples with differing BRGs reveal what happens where there is no accommodation for BRG in the testing tool.

<u>Example: 75 examinees are innocent and 25 are guilty</u> (BRG 25%)

- Pass
    - TN: 66.75 of 75 - innocent and pass (89%)
    - FN: 4.25 of 25 - guilty but pass (17%)
    - 71.0 total
- Fail
    - TP: 20.75 of 25 - guilty and fail (83%)
    - FP: 8.25 of 75 - innocent but fail (11%)
    - 29.0 total

In this case, 4.25 guilty people pass the test (FN); that's a 50% reduction in FN, which is good. And, 8.25 innocent people fail the test; that's a 33% increase in FP, which is not good.

Without adjusting for BRG, the ability of the test to predict who is guilty is now 20.75 / 29.0 = 71.6%. This is worse than the expected "rate of accuracy" published for the testing solution. However, the ability to predict the innocent participants is now 94%.

Now, for a more extreme, but plausible example.

<u>Example: 90 examinees are innocent and 10 are guilty</u> (BRG 10%)

- Pass
    - TN: 80.1 of 90 - innocent and pass (89%)
    - FN: 1.7 of 10 - guilty but pass (17%)
    - 81.8 total
- Fail
    - TP: 8.3 of 10 - guilty and fail (83%)
    - FP: 9.9 of 90 - innocent but fail (11%)
    - 18.2 total

In this case, now 1.7 guilty people pass the test (FN); that's another 60% reduction in FN, which is good. But, 9.9 innocent people fail the test; that's another 17% increase in FP, which is not good.

Without adjusting for BRG, the ability of the test to predict which person is guilty is now 45.6%. This is calculated as follows: 8.3 / 18.2. This is worse than a coin toss. However, the ability to predict who is innocent is now 98%. This is calculated as 80.1 / 81.8.

Achieving Better Results

As BRG decreases, there are more innocent people in the testing pool. If there is no accommodation in the decision model for this change, more innocent people will fail the test.

The good news is that more guilty people will fail the test. However, from our experience, the BRG for most testing groups is less than 50%. Therefore, the issue of failing innocent people applies to most testing groups.

The challenges inherent in BRG can be reduced if the decision model of the testing solution can be adjusted. By making a good estimate of the BRG prior to testing and by accommodating or adjusting the testing solution's decision model, better results will be achieved.

Solution: EyeDetect

The EyeDetect algorithm can be adjusted to accommodate for differing BRGs in its calculation to improve its ability to predict which examinees are truthful and which are deceptive.

With EyeDetect, using estimates for credibility scores belonging to the innocent and guilty distribution of scores, the Converus Science Team adjusts the algorithm cut score to balance false positive (FP) and false negative (FN) errors. In layman's terms, the algorithm adjusts to treat testing groups as if it was comprised of a 50/50 split of innocent and guilty to balance the errors.

When a customer makes an estimate of the BRG and Converus adjusts the algorithm for that estimate, the algorithm uses Bayes Rule to calculate credibility scores. Bayes Rule weighs the test score's p-value proportionally to the BRG or BRI. The "p-value" is the probability value. So, if the p-value is .25, there's a 25% chance of occurrence rather than a 50% chance when conducting a typical lab study where the BRG is 50%.

Polygraph examiners can also make adjustments for BRG in testing.

In polygraph, to make the same adjustment, the examiner can perform a manual calculation. It requires taking the examinee's score and subtracting the mean. That value is divided by the standard deviation (SD) to find a z-score. With a z-score, the examiner can calculate a probability density (PD). The examiner needs the PD for the innocent and guilty distributions. Then, use the BRI x PD (truthful) x BRI / PD (truthful) x BRI + PD of guilty x BRG.