The Police Polygraph Digest

JANUARY 2016

TABLE OF CONTENTS

Deadline for the March 2015 issue is February 1, 2016

Screening Survival Analysis - Understanding Base Rates, Accuracies, & Successive Hurdles in Credibility Assessment Screening with Complementary Technologies

Mark Handler and Charles Honts

At the recent 50th Anniversary APA seminar in Chicago, we were introduced to a new commercial alternative credibility assessment technology (Raskin, 2015). The new technology, EyeDetect, was originally known as Ocular Motor Deception Testing. Within a reading exercise, it uses a number of physiological, reading behavior, and answering behavior metrics to categorize subjects along a truthful or deceptive continuum. EyeDetect was developed by a team of researchers at the University of Utah. It has been tested in a number of Latin American countries where the results are encouraging.

The idea of the technique is not to replace polygraph (PDD), but rather to improve the overall accuracy of the credibility assessment screening process by using both EyeDetect and PDD in a "successive hurdles" configuration. Successive Hurdles (Meehl & Rosen, 1955) is the sequential application of tests with the goal of maximizing information gained in the context of varying base rates and technology accuracies. Hopefully this short article will help the reader better understand how base rates of deception and test accuracies affect testing outcomes. We also introduce the concept of "screening survival analysis" posited by Charles Honts. We find it an intuitive way to understand the outcomes of a successive hurdles approach when applying multiple technologies. We also believe it opens a dialogue about considering the goals of testing in the screening process and developing strategies towards achieving those goals.

We will illustrate the interaction of test accuracy, base rates, confidence in test outcomes and information gained during the implementation of a successive hurdles testing configuration. One of our goals is to show how using one technology to adjust the base rate of deception can improve the outcome of the second stage of testing. Depending on your testing goals, you can choose how to follow-up first stage testing. Some examples will hopefully help.

Equal Base Rates and Equal Accuracies

First consider a hypothetical credibility assessment test (Stage 1) in Table 1.

Under the Ground Truth column, we list the actual state of the subject. For consistency sake with published scientific studies, we refer to these as Innocent and Guilty. We realize in a criminal justice setting these are determined by triers of fact but here we use the terms to refer to status in the real world. TN is True Negatives or correct hits with Innocent cases. FN is False Negatives or misses with Guilty cases. TP is True Positives or correct hits with Guilty cases. FP is False Positives or misses with Innocent cases.

We apply that test to 1000 individuals, 500 of whom are Innocent and 500 of whom are Guilty, so the base rate of Guilt is 50% or .50. Assume the test is 90% accurate with both Innocent and Guilty subjects. For simplicity, we did not include the possibility of an inconclusive outcome.

Table 1. Contingency Table with equal accuracy (90%) and equal base rates (50%)

Ground Truth	Pass Test	Fail Test	Totals
Innocent	450 (TN)	50 (FP)	500
Guilty	50 (FN)	450 (TP)	500
Totals	500	500	1000
Outcome Confidence	**0.9 (NPV)**	**0.9 (PPV)**	

The numbers in the bottom row can be thought of as your confidence in the accuracy of the various outcomes. These proportions also have statistical names. The proportion of correct truthful outcomes to total truthful outcomes (here, 450/500) is known as the Negative Predictive Value (NPV). The proportion of correct deceptive outcomes to the total number of deceptive outcomes is known as the Positive Predictive Value (PPV), here, 450/500). Notably in this example with equal base rates your confidence in the test outcomes directly mirrors the accuracy of the test.
Equal Base Rates and Different Accuracies

Let's complicate this just a bit. Table 2 illustrates a second case with equal base rates, but where the test has an accuracy of 95% with Guilty, but is only 85% accurate with Innocent. The notable thing from Table 2 is that although the test is more accurate with Guilty, you have more confidence in Pass Test outcomes (NPV = .94) than you do in Fail Test outcomes (PPV = .86). That is because the number of Guilty and Innocent subjects who pass or fail the test changes disproportionally due to the imbalanced accuracies. This imbalance is reflected in the trade-off in Outcome Confidence.

Table 2. Contingency Table with equal base rates (50%) but different accuracies (Guilty = 95%, Innocent = 85%)			
Ground Truth	Pass Test	Fail Test	Totals
Innocent	425 (TN)	75 (FP)	500
Guilty	25 (FN)	475 (TP)	500
Totals	450	550	1000
Outcome Confidence	**0.94 (NPV)**	**0.86 (PPV)**	

Different Base Rates and Equal Accuracies

Finally, a third example that shows the impact of base rates on confidence in outcomes. Table 3 illustrates an example where the target of the screening test is a relatively rare event and occurs in only 10% of the people who are tested. For simplicity, Table 3 goes back to a test with equal accuracy of 90%. In this situation where the target occurs in only 10% of your subjects, your confidence in a Pass Test outcome is extremely high (NPV = .99) but your confidence in a Fail Test outcome is poor (PPV = .50). Half of the subjects who failed the test are actually Innocent. This is because the opportunity to make an error with an Innocent subject occurs nine times for every one chance to make an error with a Guilty subject. If an agency selects a very low base rate target, they can expect similar results – even with a highly accurate test.

Table 3. Contingency Table with equal accuracy (90%) but a rare target (BR of Guilt = 0.1)			
Ground Truth	Pass Test	Fail test	Totals
Innocent	810 (TN)	90 (FP)	900
Guilty	10 (FN)	90 (TP)	100
Totals	820	180	1000
Outcome Confidence	**0.99 (NPV)**	**0.50 (PPV)**	

Increasing information gained by considering accuracy and base rates

To maximize the information gained from a testing situation we have to consider accuracies and base rates. We may not be able to do much in the short run to improve on testing accuracies of PDD. Test sensitivities and specificities seem fairly stable, though we can possibly make some marginal improvements. But base rates are something within our potential control. We argue one can make a reasonable initial estimate of target base rates by reviewing past PDD test results and admissions. Many agencies keep excellent records of test results and admissions and reviewing this could provide reasonable estimates of base rates for various targets. Such estimates need not be perfect as even a rough approximation would be informative.

Agencies should be cautious about testing for extreme base rate targets. It is a truism in the science of diagnostic testing (psychometrics) with tests that have error (essentially all tests), at some point the base rate will become so extreme that simply predicting the base rate extracts all of the possible information from the situation. In such a situation conducting an error prone test will actually decrease the amount of useful information you have after conducting the test. In other words, under extreme base rate situations conducting an error prone test will make you know less after running the test than what you knew before (Meehl & Rosen, 1955).

But what if we could use one technology to adjust the base rate of Guilt up or down before applying the second technology. Let's suppose you had a lot of applicants for a job where the hiring authority wanted to make very sure the applicant was being truthful. For example, a technician who will have access to highly classified communications and encryption keys. If you were not concerned with false positive rates ("throwing a lot of babies out with the bath water") you would want a very conservative testing scenario.

To consider EyeDetect and PDD for possible use in such a successive hurdle configuration we began with the accuracy estimates from the multiple EyeDetect studies conducted on the EyeDetect as described by Dr. David Raskin at the 2015 APA seminar in Chicago, Illinois (Raskin, 2015). Those rates are shown in Table 4. (To make the data comparable we are ignoring inconclusive outcomes. The numbers were taken from Dr. Raskin's Table 2, excluding Outlier Results, D Correct and T Correct.)

Table 4. Accuracy Rates for EyeDetect (Raskin, 2015) and PDD

Ground Truth	Pass Test	Fail Test
EyeDetect		
Innocent	0.88 (TN)	0.12 (FP)
Guilt	0.17 (FN)	0.83 (TP)
PDD		
Innocent	0.83 (TN)	0.27 (FP)
Guilt	0.09 (FN)	0.91 (TP)

Successive Hurdles with Equal Base Rates

Table 5 illustrates the first hurdle (Stage 1) outcome matrix for EyeDetect with equal base rate and 1000 examinations.

Table 5. 1000 EyeDetect Outcomes with equal base rates			
Ground Truth	Pass Test	Fail test	Totals
Innocent	440 (TN)	60 (FP)	500
Guilt	85 (FN)	415 (TP)	500
Totals	525	475	100
Outcome Confidence	**0.84 (NPV)**	**0.87 (PPV)**	

In our *conservative* successive hurdles model the 475 individuals who failed the EyeDetect would be eliminated at this point. The 525 individuals who passed the EyeDetect would then be tested with a PDD test to further reduce false negative results. This second hurdle (Stage 2) PDD test would produce Table 6.

Table 6. Outcomes for the 525 PDD tests run as the second hurdle			
Ground Truth	Pass Test	Fail Test	Totals
Innocent	365 (TN)	75 (FP)	440
Guilty	8 (FN)	77 (TP)	85
Totals	373	152	525
Outcome Confidence	**0.98 (NPV)**	**0.51 (PPV)**	

Recall we began the process with 500 Guilty and 500 Innocent applicants. After the two stages of different screening tests we are left with 365 Innocent and 8 Guilty applicants. We reduced the Guilty population by approximately 98%. However, the reduction in Guilty was achieved at the cost of 135 (27%) of the 500 Innocent who were eliminated by the process.

By using EyeDetect first we adjusted the base rate of Innocent applicants up and the base rate of Guilty applicants down. Another way to look at this is if you began this process as an Innocent person your likelihood of surviving the screening is about 73%, while if you started as a Guilty applicant your likelihood of surviving the process is approximately 2%.

Successive Hurdles with a Low Base Rate Target

Table 7 illustrates the outcomes with a first hurdle EyeDetect test and a target with a 10% base rate of Guilt. This may be comparable to a screening test looking for a person involved in organized crime trying to join a police force or a person trying to be hired by a company to do industrial espionage. As you can see the confidence in passing the test is quite high but the confidence in failing the test is less than chance.

Table 7. One Thousand EyeDetect Outcomes with a 10% Base Rate for Guilt.			
Ground Truth	Pass Test	Fail test	Totals
Innocent	792 (TN)	108 (FP)	900
Guilty	17 (FN)	83 (TP)	100
Totals	809	191	1000
Outcome Confidence	**0.98**	**0.43**	

Table 8 Illustrates applying the PDD as a second hurdle outcomes with the 809 people who passed the first hurdle, our *conservative* approach. Note that we have reduced the base rate of Guilt to about 2% by using EyeDetect first. We can see that our confidence in outcome for a person passing the test is extremely high, even with the low base rate target. The confidence in a failed test in the second stage is low but recall we stated a priori we were willing to lose some applicants to make very sure we hired as few Guilty as possible.

Table 8. Outcomes for the 809 PDD tests run as the second hurdle			
Ground Truth	Pass Test	Fail Test	Totals
Innocent	657	135	792
Guilty	2	15	17
Totals	659	150	809
Outcome Confidence	**0.997**	**0.10**	

In this low base rate example, the screener began with 100 Guilty applicants and ended the two hurdle process with only two. This was done at a cost of eliminating from consideration 243 of 900 Innocent applicants. Another way of looking at this is to look at survival across the two tests. With the Innocent 900 who began the screening and 657 survived. Thus, as an Innocent applicant your likelihood of surviving this screening process would be 73%. However, for the Guilty 100 who began the process and only two survived, a likelihood of survival of 2%. A review of tables 6 and 8 shows the survival rates are quite similar across the two base rate conditions, thanks to the successive hurdles process.

Using EyeDetect to Adjust Base Rates prior to PDD Testing

The examples given above prioritized minimizing the number of Guilty who survive the screening process. Our examples show that eliminating Guilty applicants can be done quite effectively in a two hurdle process. However, eliminating Guilty necessarily involves eliminating some number of Innocent applicants who inevitably fail the fallible tests.

Different end users may place different priorities on who they want to eliminate from consideration. We could have easily reversed the process and placed a priority on reducing false positive outcomes. This *liberal* approach would of course come at the cost of increased false negative rates. The methodology provided here shows a process for formally assessing the costs and benefits of different screening processes. We encourage end users to consider their testing goals and adjust targets and base rates to achieve them. One of the greatest values we see in EyeDetect is providing an empirically proven manner of adjusting base rates to improve PDD screening outcomes. This adjunct technology allows screening programs to engage in a true successive hurdles model as recommended by Meehl & Rosen (1955) when faced with imperfect tests and varying base rates.

Multiple and Successive Hurdles approach to hiring.

Texas DPS lieutenant Dr. Adam Park's Master's thesis (Park & Herndon, 2015) is reprinted in the journal *Polygraph* as "Police Cadet Attrition and Training Performance Outcomes." In it he discusses "multiple hurdles" approach which requires applicants to pass each stage of the hiring process to be considered for employment.

If, for example, during the initial recruiting interview the applicant discloses he or she smoked marijuana within the past month, they would likely be disqualified. They don't meet the minimum requirements to clear that hurdle and they don't move forward in the process. Likewise, an applicant who can't successfully complete the physical agility course will probably not move on in the process. Subsequent hurdles become more invasive and more expensive. It seems best to weed out unqualified applicants as early as possible to be fiscally responsible and to improve the odds that the qualified candidates get a better chance at being hired.

"Successive hurdles" is an approach used to compensate for imperfect testing techniques. A successive hurdles approach would start with a technology having either good sensitivity or specificity to the issue of concern, depending on testing goals. If there is a positive or negative result (in this case a passed or failed aspect of an early hurdle) a more specific testing tool is used to try to confirm or disconfirm the area of concern. Historically PDD examiners use a single-issue screening PDD test to follow-up on positive multiple-issue screenings tests. Research has not supported the single-issue screening test is better able to identify areas of concern. It would seem that using alternative technologies to follow up failed hurdles might be a better approach in a successive hurdles model. Also, as we have tried to point out, adjusting base rates with alternative technologies as a first stage is more efficacious than trying to improve PDD accuracy.

Assuming it has accuracy levels stated above, EyeDetect would be a good addition to any multiple hurdles and/or successive hurdles hiring program that uses PDD. Polygraph testing is more intrusive and more time consuming, but it is the most powerful tool for finding out the disqualifying information. EyeDetect seems best suited as an early hurdle for those applicants who claim to have no disqualifying acts in their history. The reported sensitivity and specificity is as high, or higher, than PDD.

The EyeDetect testing process and milieu is not designed to get disqualifying information from the applicant. It is an early interview hurdle - much better than a human to human interview where ability to detect deception hovers near chance (54%, Bond & DePaulo, 2006). EyeDetect can be used to more quickly move forward those applicants or test subjects who have not (or are likely to have not) committed disqualifying acts and thus alter the base rate of guilt subjected to the more expensive PDD testing. EyeDetect would also seem to have the potential to be used as an early risk assessment tool in Post-Conviction Sex Offender testing as well. Using EyeDetect could help quickly identify those test subjects who may need to be further investigated with the PDD, or other means, by the treatment team. By altering the base rate of guilt subjects moving to PDD we can make our confidence in a test result higher.

Finally, having alternative technologies can help defend against claims of unfair testing processes in hiring. Should an applicant or test subject fail PDD after failing EyeDetect, a hiring authority, probation officer, or treatment provider would seem to be in a more defensible position to explain to the subject that they employed multiple technologies to try to clear the disqualifying behavior hurdle. As we have shown, the use multiple testing technologies can reduce the final overall error rates that are most important to the screening agency. While no test is perfect, having multiple technologies is safer for the decision maker and the applicant than a single technology approach.

References:

Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, *10*, 214-234.

Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin*, *52*, 194-216.

Park, A., & Herndon, J. (2015). Police Cadet Attrition and Training Performance Outcomes. Polygraph 44(2), 143-161.

Raskin, D. C. (2015, September). The Utah Technique, Presentation at the Presented at the American Polygraph Association 50th Annual Seminar, Chicago, IL.