EyeDetect Hybrid Directed-lie Comparison Test (HDLC)

Development and Validation Summary

May 16, 2021

Converus recently introduced the EyeDetect Hybrid Directed-Lie Comparison Test (HDLC). The HDLC is a specific incident test that covers a single relevant issue, such as a murder, robbery, or parole violations. The comparison issue for this relevant issue (R1) is a directed lie question. Directed lie questions are concerned with inappropriate acts, and the computer directs the test subject to lie to those questions. Examples are, “In your entire life, have your ever lied to get out of trouble?” or ”In your entire life, have you always told the truth?” Test subjects are instructed to lie to those questions – to claim they have never lied and say they have always told the truth. The computer informs the subject that it will use the physiological and ocular-motor data to try to distinguish their deceptive answers to directed lie questions from presumed truthful answers to relevant questions. The subjects are told if the computer can’t distinguish between the directed lie and relevant questions, the subjects are told they could fail the test. The HDLC assumes that truthful subjects will be concerned that they might fail the test because they don’t react as they should to the directed lie questions. Consequently, the test predicts that truthful test subjects will be more concerned about directed lie questions than the relevant questions. On the other hand, the test predicts that guilty subjects will react most strongly to the relevant questions because they are lying to those questions that pertain to the matter under investigation. For test subjects who answer relevant questions truthfully, directed lie questions should pose the greater threat to passing the test and capture more attention. Conversely, for test subjects who lie to relevant questions, relevant questions are more salient and should pose the greater threat.

The HDLC combines traditional polygraph measures and newer ocular-motor measures to distinguish between truthful and deceptive people. In addition to ocular-motor signals from the eye tracker, EyeDetect+ 2.0 records electrodermal activity from disposable electrodes attached to the palmar surface of fingers of the non-preferred hand. It also records respiration from a strain gauge around the chest or abdomen, the electrocardiogram (ECG) from electrodes attached to each arm, and the photoplethysmograph (PPG) from a finger on the non-preferred hand.

From the ECG and PPG signals, the computer derives an indirect measure of blood pressure known as Pulse Transit Time (PTT). PTT varies inversely with changes in blood pressure (Geddes et al., 1981; Obrist et al., 1978; Obrist et al., 1979). As blood pressure increases, the time it takes the pulse to travel from the heart to the periphery decreases. Research has shown that PTT is at least as diagnostic as the traditional cardiograph for probable-lie polygraph tests (Webb & Kircher, 2005). EyeDetect+ 2.0 replaces the cardiograph with PTT. PTT is less invasive than the cardiograph, alleviates test subject discomfort, and offers more flexibility in test construction.

We developed and cross-validated two decisions models. The first model assumed availability of polygraph and ocular-motor measures. Outcomes were inconclusive when the credibility score was between 45 and 55. On cross-validation, HDLC decisions were 88.6% correct, excluding 8.1% inconclusive outcomes. Accuracy was slightly higher for innocent (89.3%) than for guilty subjects

(87.9%). Without an inconclusive region, mean accuracy was 86.2%. Accuracy was slightly higher for truthful (87.1%) than for guilty subjects (85.4%).

The second model assumed availability of only ocular-motor measures. On cross-validation, 89.5% of innocent and 84.5% of guilty subjects were classified correctly. Mean accuracy was 87.0%, excluding 7.3% inconclusive outcomes. With no inconclusive region, HDLC decisions were 82.3% and 87.1% correct for innocent and guilty subjects, respectively. Mean accuracy was 84.7%.

The remainder of the present paper describes the procedures used to develop the HDLC test and estimate its accuracy.

Experiment

We conducted a mock crime experiment modeled after Cook et al. (2012) to collect the physiological data needed to develop and cross-validate a decision model composed of polygraph and ocular-motor measures. We recruited 124 test subjects from the local community. Subjects were told that some participants would commit a mock theft of $20 from a secretary's wallet, whereas others would be innocent and would not commit the crime. Subjects were randomly assigned to guilty and innocent groups. Guilty subjects (n=60) took $20 from a secretary's office and lied about it on the test. Innocent subjects (n=60) answered truthfully to questions about taking the $20.

Subjects completed their assigned task and were given the HDLC test. The test proctor asked the subject to sit at a desk in front of a computer screen. The proctor attached sensors for skin conductance, ECG, photoplethysmograph (PPG), and respiration to the subject and then calibrated the eye tracker. We used a multi-channel physiology monitor to record skin conductance and respiration at 340 Hz and ECG and PPG at 1000 Hz. A Tobii 4C remote eye tracker attached to the bottom of the computer monitor recorded bilateral gaze position and pupil diameter at 60 Hz. Prior to the HDLC test, subjects were required to complete a practice test with DLC questions. Before they could take the HDLC test, subjects were required to demonstrate that they understood the instruction to lie to directed-lie questions. A text-to-speech digital voice presented test instructions and questions over headphones. Subjects were instructed to avoid moving, to keep their feet flat on the floor, and to look at the computer screen during the test.

The HDLC test consisted of a protocol optimized for traditional polygraph measures followed by a protocol optimized for ocular-motor measures. The polygraph-optimized phase consisted of a set of eight Yes/No questions that was repeated three times with 22-second inter-question intervals. That was followed by three repetitions of a set of 36 True/False statements optimized for the ocular-motor measures. Subjects were instructed to answer the True/False statements as quickly and accurately as possible or they might fail the test. Analysis conducted midway through data collection indicated that the data from the ocular-motor optimized phase did not increase the accuracy achieved in the initial polygraph-optimized phase. At that point, we dropped the ocular optimized phase and increased the number of polygraph sessions from three to four. At the conclusion of data collection, we compared the diagnostic accuracy achieved with three or four polygraph sessions. The accuracy achieved with four polygraph sessions was no higher than that obtained with only three polygraph sessions. Consequently, the final HDLC test contained a set of eight Yes/No questions that was repeated three times.

Table 1 contains the questions presented in one of the three HDLC sessions. The computer presented the same directed lie and relevant questions in the three sessions, but the order of question presentation and the arithmetic questions varied over sessions.

Table 1. Hybrid directed-lie comparison test questions

Question Number	Question Type	Question
1	Directed lie	In your entire life, have you ever broken a rule?
2	Relevant	Did you take that $20?
3	Arithmetic	Does 6 + 2 equal 8?
4	Relevant	Do you have that $20 with you now?
5	Directed lie	In your entire life, have you ever made a mistake?
6	Arithmetic	Does 5 + 1 equal 6?
7	Directed lie	In your entire life, have you always told the truth?
8	Relevant	Did you take the $20 from the backpack?

Subjects were paid $40 for participating in the study and had been offered an additional $30 bonus if they passed the test.

Analysis

Ocular-motor data were analyzed to identify features that discriminated between questions and statements answered truthfully and deceptively. We created a logistic regression equation from a subset of diagnostic polygraph and ocular-motor features. The logistic regression equation provided the probability that the test subject was truthful to the relevant questions on the test (credibility score). The credibility score was used to classify the test subject as truthful or deceptive. In some cases, the credibility score was near 50% (chance) and was classified as inconclusive. Accuracy results are reported with, and without, an inconclusive region.

K-Fold Validation

A statistical model that is optimal for classifying the cases in a particular experiment is rarely optimal for the population from which the subjects were sampled. The model is not optimal because the sample does not perfectly represent the general population from which it was drawn. Consequently, we obtain biased estimates of accuracy if we test the model on the cases that were used to create the model.

Better estimates of accuracy may be obtained with k-fold validation. A k-fold validation divides the data set into *k* folds (subsets). The first subset comprises a hold-out subsample and is removed from the dataset. The remaining subsets are combined to create a training set. A logistic regression model is developed using the cases in the training set. That regression model is then used to classify the cases in the hold-out subsample. The accuracy observed in the hold-out sample provides a less biased estimate of accuracy because the holdout cases were not used to optimize feature coefficients in the regression equation. The accuracy achieved in the hold-out sample is recorded.

This process continues for each partition of the data set. The first subset is returned to the training set, and the second subset is removed to serve as a new holdout sample. A new logistic regression equation is created with all but the second subset of cases. That model is used to classify cases in the holdout sample, and its accuracy is recorded. This process is repeated for each of the remaining subsets. The best estimate of accuracy for the model is mean accuracy for the *k* holdout samples.

Validation of the HDLC with Polygraph and Ocular Motor Measures

A 4-fold validation was performed on the decision model. The sample of 124 subjects was split into 4 subsamples of 30 subjects each. The fourth subsample contained 34 subjects. Half of the subjects in each subsample were innocent and half were guilty. Table 2 shows percent correct for truthful and deceptive questions for each fold, as well as the mean accuracy across the four folds.

Table 2. Percent correct decisions for innocent and guilty subjects in 4-fold validation

	Fold 1	Fold 2	Fold 3	Fold 4	Mean
n	30	30	30	34	
Innocent	94.1	86.7	80.0	86.7	87.1
Guilty	76.5	73.3	93.3	100.0	85.5
			Mean Accuracy		**86.3**

Mean accuracy was slightly higher for innocent (87.1%) than for guilty subjects (85.5%). Based on these results, *we would expect the HDLC to produce 86% correct decisions when the model is used in a new sample*.

Because the HDLC test is more likely than the RCT or MCT to be used in criminal investigations, it may be difficult to justify making a definite decision when the credibility score is near 50 (chance). We used the unbiased credibility scores for each holdout sample of 30 cases to compute accuracy rates when tests that produced credibility scores between 45 and 55 were considered inconclusive. The results are presented in Table 3. Of the 124 HDLC tests, 10 (8.1%) were inconclusive. Percent correct decisions in Table 3 exclude the inconclusive outcomes.

Table 3. Frequency of correct, wrong, and inconclusive outcomes for innocent and guilty subjects

	Sample size	Correct	Wrong	Inconclusive	% Correct Decisions
Innocent	62	50	6	6	89.3
Guilty	62	51	7	4	87.9
	124	101	13	**Mean Accuracy**	**88.6**

When tested on a new sample of cases, *we would expect the HDLC test to produce about 89% correct decisions excluding 8% inconclusive test results.*

Validation of the HDLC with Only Ocular Motor Measures

To obtain traditional polygraph measures, we must attach multiple sensors to the test subject and ensure that the various sensor produce high quality signals. This direct contact with the subject and additional training of the test proctor that is not needed with an ocular-motor deception test that uses a remote eye tracker.

A new HDLC model was developed and cross validated that used only ocular-motor measures. 4-fold validation results are presented in Table 4.

Table 4. Percent correct decisions for innocent and guilty subjects in 4-fold validation using only ocular-motor measures

	Fold 1	Fold 2	Fold 3	Fold 4	Mean
n	30	30	30	34	
Innocent	86.7	66.7	86.7	88.2	82.3
Guilty	80.0	100.0	86.7	82.4	87.1
				Mean Accuracy	**84.7**

The mean accuracy of HDLC with only ocular-motor measures was 84.7% on cross-validation. We obtained the results in Table 5 with an inconclusive region for unbiased credibility scores that ranged from 45 to 55. Of the 124 tests, 9 were inconclusive (7.3%). Excluding inconclusive outcomes, mean accuracy was 87%. The addition of measures from traditional polygraph signals had increased accuracy from 87% to almost 89%.

Table 5. Frequency of correct, wrong, and inconclusive outcomes for innocent and guilty subjects using only ocular-motor measures

	Sample size	Correct	Wrong	Inconclusive	% Correct Decisions
Innocent	62	51	6	5	89.5
Guilty	62	49	9	4	84.5
	124	100	15	**Mean Accuracy**	**87.0**

References

Cook, A. E., Hacker, D. J., Webb, A. K., Osher, D., Kristjansson, S., Woltz, D. J., & Kircher, J. C. (2012). Lyin' Eyes: Ocular-motor Measures of Reading Reveal Deception. Journal of *Experimental Psychology: Applied, 18(3)*, 301-313.

Geddes L.A., Voelz, M.H., Babbs C.F., Bourland J.D., & Tacker W.A. (1981). Pulse transit time as an indicator of arterial blood pressure. *Psychophysiology, 18,* 71–74.

Obrist, P. A., Light, K. C., McCubbin, J. A., Hutcheson, J. S., & Hoffer, J. L. (1978). Pulse transit time: Relationship to blood pressure. *Behavior Research Methods & Instrumentation, 10,* 623-626.

Obrist, P. A., Light, K. C., McCubbin, J. A., Hutcheson, J. S., & Hoffer, J. L. (1979). Pulse transit time: Relationship to blood pressure and myocardial performance. *Psychophysiology, 16,* 292-301.

Webb, A. K. & Kircher, J. C. (2005). *Use of Pulse Transit Time for the Psychophysiological Detection of Deception.* Final report to the U.S. Department of Defense. Salt Lake City: University of Utah, Department of Educational Psychology.