EUROPEAN
POLYGRAPH
PUBLISHED SEMI-ANNUALLY
2020 VOLUME 14 NUMBER 2 (52)

DOI: 10.2478/EP-2020-0015

Electrodermal Response Ratios: Scoring Against the Stronger of Two Comparison Questions in Search of an Optimal Minimum Threshold*

Donald J. Krapohl

Past President of the American Polygraph Association
Questions and comments can be sent to APAkrapohl@gmail.com

Key words: polygraph examination, electrodermal activity, galvanic skin response (GSR), electrodermal response (EDR), Bigger-Is-Better Rule (BIBR)

Abstract

Previously, Krapohl (2020) evaluated the Bigger-Is-Better Rule (BIBR) on the polygraph electrodermal channel to assess whether there was a best minimum ratio between response sizes for assigning a score. Performance peaked at a minimum ratio between 10% and 20%. The ratios had been calculated by comparing the electrodermal responses for each relevant question

* Acknowledgments. I am very grateful to dr. John Kircher for providing the electrodermal data used in this study and to Mr. Brett Stern for his thoughtful comments and suggestions to an earlier draft. The views expressed in this article are solely those of the author and do not necessarily represent those of the Capital Center for Credibility Assessment.

against those of the immediately preceding comparison question. The analysis did not consider whether the same optimal ratio would be found if the relevant question electrodermal responses are compared to those of the stronger of two adjacent comparison questions. To investigate we analyzed responses from an independent sample of 255 laboratory cases. The data from those cases found the highest correlation between scores and ground truth occurred when the minimum difference between two electrodermal responses was 30%.

Introduction

Many or most polygraph schools teach the Bigger-Is-Better Rule (BIBR) in scoring. In simplest of terms, the BIBR states that a polygraph score can be assigned if the scorer perceives a stronger physiological response to one question than to another. By convention, when the stronger reaction is associated with a comparison question over a relevant question a positive score is assigned. Conversely, the stronger reaction to the relevant question warrants a negative score. There is no widely accepted minimum difference before a score can be given, however. The difference needs only be observed, and consequently, the decision to score is a subjective one. As with all subjective assessments, perceived differences in reaction intensity may be influenced by training, experience, preference, and how the data are displayed. These influences may be especially pertinent when differences are subtle.

The more frequently a subjective interpretation is called upon the greater is the opportunity for individual differences among scorers to be made manifest. Factors systematically affecting scoring will aggregate as more scores are assigned. Channels that tend to receive scores more often can be expected to have a disproportional influence on the final score, and hence the polygraph results. We drew a convenience sample of polygraph scores to get an impression of the frequency of scores assigned by polygraph examiners. In a large unpublished US Government laboratory study of polygraph screening methods there were 102 examinees who underwent polygraph examinations with the Test for Espionage and Sabotage (Dollins, Senter & Pollina, 2001). Across 612 opportunities to assign scores in that sample, non-zero scores were given in the pneumograph 52% of the time, 79% in the cardiograph and 91% in the electrodermal data. In a different analysis of field criminal cases, Ansley and Krapohl (2000) found 55% of the reactions in polygraph charts came from the electrodermal channel, followed by 26% in the cardiograph and 19% in the pneumograph. In a third approach, Bell et al. (1999) concluded "...the Utah scoring rules give greater weight to electrodermal reactions than to cardiovascular, respiration, or plethysmograph reactions". Virtually all carefully conducted analyses of polygraph data report the electrodermal channel tends to be dominant in polygraph scoring. As such the electrodermal channel has potentially more influence over the final polygraph decision than other traditional channels. If the goal is to increase polygraph decision accuracy, improving how electrodermal responses are scored offers one of the more impactful opportunities.

In a previous report Krapohl (2020) examined archival electrodermal measurements to determine whether simply being bigger is enough, or whether accuracy could be improved by requiring a minimum difference between two electrodermal responses (EDRs) to assign a score. Briefly, in that study the measurements of EDR amplitudes for 300 confirmed field Federal Zone Comparison Question Tests were systematically compared at minimum ratios between >1.0:1 to >1.8:1 in 0.1 increments. Correlation tests were conducted between ground truth and test results based exclusively on EDR scores. Electrodermal performance peaked when a minimum ratio difference between 10% and 20% was imposed. The findings could generalize to polygraph techniques in which the reaction of each relevant question is scored against a single designated comparison question. It was not established, however, whether they would generalize to the more common practice of scoring each relevant question against the stronger of two nearby comparison questions.

The present effort was designed to investigate this possibility. We combined the data from three laboratory studies to determine whether there was a best minimum to impose on EDR differences when scoring against the stronger response from two comparison questions.

Method

Data

Only the electrodermal data were used for this project. The data were collected during three separate doctoral research projects at the University of Utah (Bernhardt, 2005; Kircher, 1983; Podlesny & Raskin, 1978). The total sample size was 255 cases (128 deceptive, 127 non-deceptive). In this data set there were three relevant and three probable-lie comparison questions presented on three charts for a total of 2295 EDRs from relevant questions and the same number from comparison questions across the 255 cases.

In the three-question Utah Probable-Lie Technique (Handler, 2006) each relevant question is immediately preceded by a comparison question, but not followed immediately by one. Therefore, relevant questions are not directly bracketed by comparison questions as they are in many other techniques. For this effort the first two relevant questions were compared to the two comparison questions that were presented closest before and after each relevant question. The third relevant question in this technique is the final test question, and therefore has no comparison question following it. The third relevant question was scored against the comparison question immediately preceding it, and to the first comparison question in the sequence. In this way the EDR of each relevant question was gauged against two comparison questions and each comparison question was used for scoring exactly two relevant questions. This approach simulated a testing technique in which each relevant question is bracketed by two adjacent comparison questions.

Procedure

The EDR amplitude of each relevant question was compared to one of two probable-lie comparison question evoking the stronger response. This created three ratios per test chart, and nine ratios total for the three test charts per examinee. In the first assessment, any ratio greater than 1:1 was cause for assigning a score. It did not matter how much larger the EDR was. If the EDR to the relevant question was greater, a score of – 1 was assigned. If the EDR to one of the comparison questions was larger, a +1 was given. All equal amplitudes were assigned a 0. The scores were then tallied. With nine presentations of relevant questions a total score between – 9 and +9 per case was possible. This regimen was repeated for all 255 examinations. With ground truth coded as – 1 for deceptive and +1 for truthful, a point bi-serial correlation coefficient was calculated for the total score and the ground truth code. The point bi-serial correlation coefficient has a range of 0.0 to 1.0. The higher the coefficient becomes, the closer the relationship is between ground truth and the test score.

These steps were then repeated for all EDR amplitude ratios between 1.1:1 and 1.8:1 in 0.1 increments. Said another way, scores were assigned to minimum differences in EDR amplitudes beginning with any difference and progressing stepwise in 10% increments to ratios up to an 80% difference. When the individual scores were summed in each case, the coefficients were calculated in the manner described in the previous paragraph for each of these nine minimum ratios.

Results

Using any difference between EDR amplitudes greater than zero to assign a score produced a relatively strong correlation coefficient, in this case $r_{pb} = 0.649$. The coefficient did not become maximal until the minimum difference between two EDRs reached 30%, where $r_{pb} = 0.680$. Both before and after the 30% minimum difference in EDR amplitudes the coefficient falls. See Figure 1.

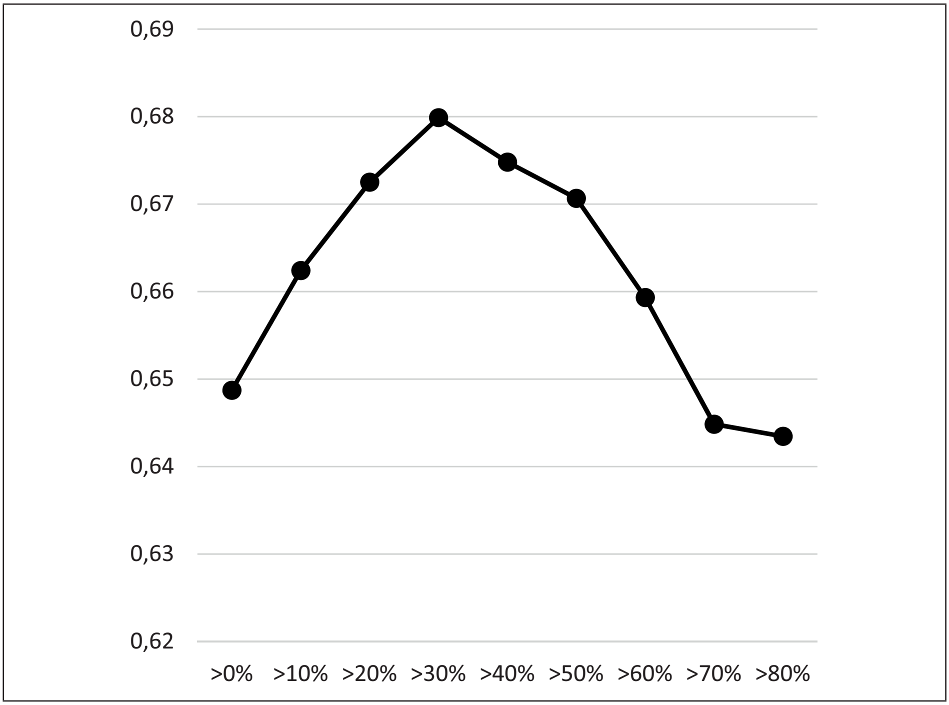

Figure 1. Point bi-serial coefficients between ground truth and EDR scores at escalating minimum differences between >0% and >80% in 10% increments for 255 laboratory cases.

As the minimum EDR differences for score assignment increases there is also a corresponding general increase in the proportion of cases in which EDR scores sum to zero. See Figure 2. As was observed with field cases in Krapohl (2020) there are virtually no cases in which EDR scores sum to zero when scores can be assigned to any difference in EDR amplitudes. There is an initial spike between >0% and >10% minimum difference in EDR amplitudes as there was in the previous Krapohl study. When requiring an 80% difference in EDR amplitudes for score assignment the proportion of cases with sums of zero is 0.094.

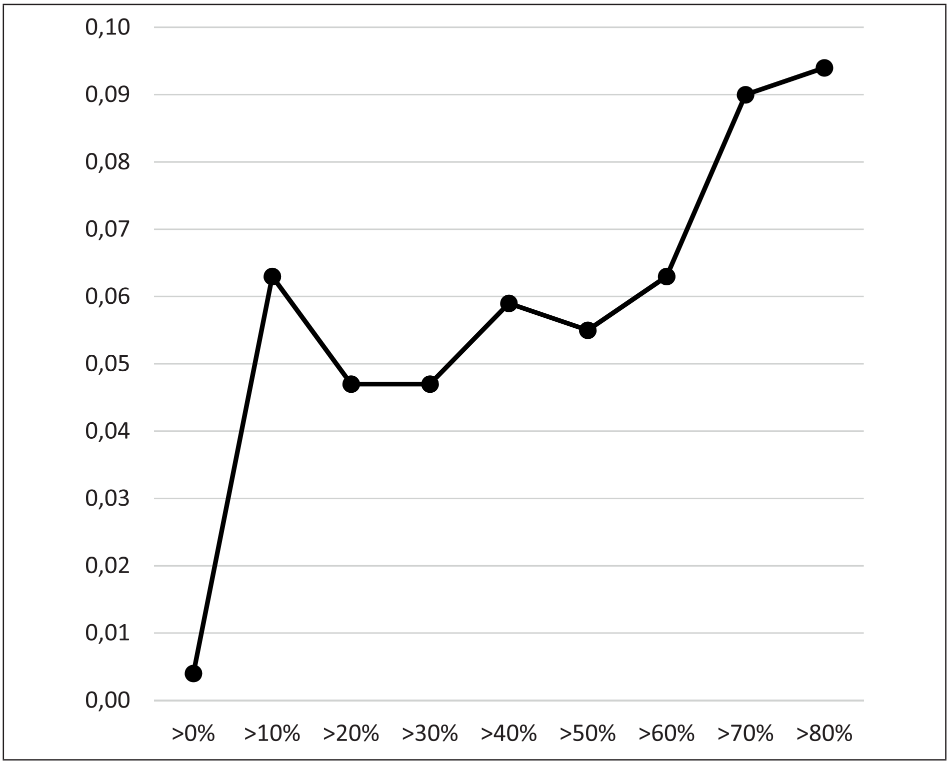

Figure 2. Proportion of cases with EDR scores summing to 0 at escalating minimum differences between >0% and >80% in 10% increments for 255 laboratory cases.

Discussion

The present data and those of Krapohl (2020) support the hypothesis that the BIBR is an effective heuristic. Both data sets, one lab and one field, found good detection efficiency at any minimum difference in EDRs. Bigger does seem to be better.

The two data sets also suggest the best performance does not occur when there is merely any difference between one EDR and another, but rather when there were specific minimum differences. In the Krapohl (2020) study the best performance was seen when the minimum EDR difference was set at 10%–20%. The current data set points to best performance when the minimum difference is 30%. A common finding from both studies is that scoring just any difference in EDR amplitudes, as is permitted with the BIBR, is acceptable but not necessarily optimal. Establishing a minimum difference seems to improve the contribution of EDA scores up to a point. The data from the present and previous study point to a minimum EDR difference between 10% and 30%.

Limitations

As in the earlier Krapohl (2020) study, generalizations of the present results are restricted to 3-position scoring systems, including the Empirical Scoring System. No evaluation was made for 7-position or rank order scoring.

The study also used single-issue examinations. Examinations where the examinee could be truthful to some questions while deceptive to others may produce a different outcome from what we found. Because mixed-issue examinations typically have fewer presentations of each issue than do single-issue examinations, variability would be expected to be greater as it typically is in smaller samples. This feature of mixed-issue examinations may affect where the best minimum difference in EDR amplitudes will be. More work is needed before generalizing the current findings to mixed-issue examinations.

Our study also used laboratory cases. A chief criticism of laboratory polygraph data is that the experience of lab examinees is quite unlike that of examinees in the field who face significant consequences for adverse test results (Cacioppo, Tassinary & Bernston, 2000). Consistent with this assertion, Pollina et al. (2004) did find differences in the response profiles in the physiological data between lab and field cases, though not in polygraph decision accuracy.

References

Ansley, N., and Krapohl, D.J. (2000), The frequency of appearance of evaluative criteria in field polygraph charts. *Polygraph*, 29 (2), 169–176.

Bell, B.G., Raskin, D.C., Honts, C.R., and Kircher, J.C. (1999), The Utah numerical scoring system. *Polygraph*, 28 (1), 1–9.

Bernhardt, P.C. (2005), *Effects of prior demonstrations of polygraph accuracy on outcomes of probable-lie and directed-lie polygraph tests*. The University of Utah, ProQuest Dissertations Publishing.

Cacioppo, J.T., Tassinary, L.G., and Berntson, G.G. (2000), *Handbook of Psychophysiology*, 2nd edition. New York, NY: Cambridge.

Dollins, A.B., Senter, S.M., and Pollina, D.A. (2001), *A Test of the Counterintelligence Screening Polygraph Process*. Report No. DoDPI01-R-0002. Department of Defense Polygraph Institute, Ft. Jackson, SC. Unpublished.

Handler, M.D. (2006), Utah Probable Lie Comparison Test. *Polygraph*, 35 (3), 139–148.

Kircher, J.C. (1983), Computerized decision-making and patterns of activation in the detection of deception. *Dissertation Abstracts International*, 44 (1-B).

Krapohl, D.J. (2020), *Electrodermal responses: When is bigger really better?* (in press).

Podlesny, J.A., and Raskin, D.C. (1978), Effectiveness of techniques and physiological measures in the detection of deception. *Psychophysiology*, 15 (4), 344–359.

Pollina, D.A. Dollins, A.B., Senter, S.M., Krapohl, D.J., & Ryan, A.H. (2004), Comparison of polygraph data obtained from individuals involved in mock crimes and actual criminal investigations. *Journal of Applied Psychology*, 89, 1099–1105.