Bigger is Better for Automated Scoring:

Analysis of Minimum Constraints for RQ/CQ Ratios

Raymond Nelson

Abstract

An archival sample of n=300 confirmed field polygraph examinations was used to study the effects of minimum constraint ratios, from 1:1 to 2:1 in increments of .05, for automated feature extraction and automated score assignment. For respiration data 95% of the scores were zero (0) at a minimum constraint ratio of 1.6:1. In contrast, approximately 55% of the EDA scores were non-zero and 39% of the cardio scores were non-zero at the same (1.6:1) ratio. LogRC Ratios were optimal with no minimum constraint, indicating that automated scoring methods are reasonable when they attempt to make use of any measurable difference that can be extracted from relevant and comparison questions. For signed integer scores, the correlation coefficient (similar to DEC) for the 2700 numerical scores was largely unaffected by any constraint for cardio data. The correlation coefficient for numerical scores of EDA data was minimally affected by the series of constraints, beginning at .425 at 1.05:1, then rising slightly to .450 at ratio of 1.2:1, and ending at .385 at the maximum constraint ratio of 2:1. Score correlations for respiration scores, together with the aggregated score correlation (shown in orange), suggest that constraining the respiration score extraction to the range from 1.2:1 and 1.6:1 may be useful to optimize the contribution of respiration scores to correct vs incorrect conclusions. Data from this analysis indicate that no minimum constraint ratio is needed for automated analysis methods for EDA or cardio and provide general support for the validity of the bigger-is-better rule.

Introduction

All scientific conclusions, in both scientific research and scientific testing, are made with regard to other possible conclusions. The process of science is intended to evaluate the strength of available evidence to support each of the different possible conclusions that attempt to answer basic questions about the universe and reality. What is it? How does it work? Why? Regardless of whether science occurs at the level of theoretical physics or at the level of practical forensic and risk-management decisions about how best to proceed with a single individual, reproducibility of analytic results has become a de facto standard or expectation for all areas of scientific research and scientific testing (Peng, 2011). The need for reproducibility can be easily observed in credibility assessment testing – beginning with the fact that test data are permanently recorded. An ability to record data is foundational to an ability to study the signals in different ways so that analytic methods can be optimized.

Advancements in technology during the early history of the polygraph profession involved both the development of sensors that can provide access to physiological signals that are correlated with deception and truth-telling, and methods to record changes in physiology so that they may be studied more carefully and studied repeatedly. Today we know that although deception itself cannot be measured physically, we are not likely to ever find any physiological activity that is uniquely associated deception. Instead, all polygraph signals will most likely continue to involve the autonomic nervous system and multiple aspects of the cerebral cortex. And all physiological activity will likely continue to be associated with multiple types of human behavior. In short: there is no such thing as “Pinocchio’s nose.”

The strength of correlation of different physiological signals, along with the degree to which different signals may covary, will remain an underlying concern to anyone involved in polygraph validity research (or discussions about polygraph validity). To be use useful, physiological signals will ideally correlate with the criterion at a statistically significant level, but will not covary so strongly that they are redundant. Useful signals will contribute unique, non-redundant, information to a structural model (i.e., a mathematical/statistical representation of the phenomena of interest). Redundancy of physiological signal can be observed when adding additional data to the model does not increase the effectiveness of the model, even though the added information is known to be correlated with the phenomena of interest. In other words, the simplistic adage "more information is always better" is untrue: more information is better when it increases the effect size of interest. If the added information does not increase the effect size of interest the actual effect will be an increase in risk for confusion and unreliability. The result of all of this is that scientific tests are often constructed of signals of moderate correlation strength – because signals for which the criterion correlation is strong will tend to covary so much they can become redundant.

During the early part of the 20th century the kymograph was the best available technology to record polygraph data for subsequent analysis, and re-analysis. Analytic methods through the mid-century period relied almost uniformly on the un-quantified intuition and experience of the expert observer. Over time, toward the latter half of the 20th century, the need for improved consistency and skill development among a variety of experts led to an emphasis on numerical scoring systems such as the seven-position system and three-position scoring method. Towards the latter half of the 20th century we saw an exponential increase in the availability computing technology. Powerful and (relatively) inexpensive computing technology has influenced virtually every aspect of social and professional life – including recreation, entertainment, communication, transportation, education, administration, employment, news and information, publication, and even science and scientific testing.

Today – well into the 21st century – there is no area of society and no area of science that does not make use of computing technology to record and analyze data. The kymograph of the early 20th century is today virtually completely supplanted, in both polygraph field practice and polygraph research, with analog to digital converters and computerized encoding systems that record polygraph data not as a tracing on a cylinder or paper scroll but as a time-series of recorded numbers stored on an electronic media. Data are processed for display in the familiar form of time-series tracings on a computer screen. A convenient aspect of all of this is that older polygraph examiners can plot or print their "charts" onto paper and inspect them visually in ways similar to what they have done in decades past.

Whereas the factors that influence the plotted lines in the days of early polygraph instrumentation were entirely mechanical – involving the moving mass of carefully engineered hardware, including the friction coefficients of pivots and bearings along the myriad of adjustments and calibrations necessary to ensure that recorded data, encoded as ink on paper, would be useful – polygraph signal processing can today be more carefully and precisely designed through the careful efforts of electrical engineers who understand our hardware requirements and through software engineers and data scientists who can enable us to make use of digital signal processing methods and statistical methods with more power than those we used during the era when all computations were done manually.

Electronic engineering and digital signal processing methods can provide far greater precision and reliability, and with much greater convenience and economy, than mechanical solutions of the past. In contrast to mechanical polygraph systems, in which filtering and smoothing was sometimes an unintended or unanticipated byproduct of the friction of the weight of the capillary ink pen on the scrolled paper, computerized polygraph systems of today – with high sampling rates and high resolution analog-to-digital conversion – can provide data that is of higher fidelity, in terms of recording and representing physiological activity, than ever in the past.

Simultaneous with advances in polygraph testing methods, signal processing and data recording, data analysis methods have also advanced as a result of available computing technologies. Polygraph professionals now have access to both empirical reference distributions (Krapohl & McManus, 1999; Krapohl, 2002; Nelson, Krapohl & Handler, 2008; Nelson & Handler, 2015) and multinomial reference distributions (Nelson, 2017; 2018). The availability of computer-based statistical reference models has led to the potential for convenient application of both frequentist and Bayesian statistical methods in polygraph field practice.

Today, in the 21st century, we have the capability for both digital recording of polygraph signals and the potential convenient use of powerful mathematical and statistical methods that can go well beyond what polygraph professionals are willing to attempt with pencil and paper. We also have the capability for automated feature extraction – and this will be inherently more reliable than feature extraction through visual pattern recognition methods that may have been the best available solution for analog polygraph instruments. Deception and truth-telling are complex problems – beginning with the complex asymmetry of even achieving a completely satisfactory epistemological/philosophical definition of deception and truth. It is also not surprising that the analysis of credibility assessment test data is inherently complex – and therefore subject to a variety of forms of bias, subjectivity and inconsistency.

The magnitude of complexity surrounding polygraph feature extraction becomes quickly apparent when considering the combination or interaction of factors that can influence a numerical score, including feature extraction at both the relevant-question (RQ) and comparison-question (CQ) along with the comparison of these two values. The most realistic solution for the future of the polygraph profession will be to harness the power of digital computers to record process and analyze the variety of complexities and interactions, including the task of automated feature extraction and automated score assignment. To do otherwise – to limit polygraph methodology to mid-century methods from the pre-computer epoch – will be to invite eventual disruption. Fortunately, a great deal of knowledge and methodology exists for this purpose. Computing power and analysis tools today are abundant and inexpensive – quite often they are free and open source. This project is an optimization study of automated numerical score assignment as a function of the ratio of RQ and CQ and pairs. The question of interest is whether there exists a set of minimum RQ/CQ ratio constraints that will maximize the diagnostic information that is achieved in the numerical scores for each of the polygraph recording sensors.

Methods

Data

Data for this project were n=300 confirmed field exams that were conducted using the Federal Zone Comparison Test (FZCT, Department of Defense, 2006) format. Sample cases were conducted by a variety of federal, state, and municipal law enforcement agency and were subsequently included in the confirmed case archive at the Department of Defense Polygraph Institute (now the National Center for Credibility Assessment). All cases consisted of three iterations of a question sequence that included three relevant-questions (RQs) and three comparison-questions (CQs) in addition to other procedural questions that are not subject to numerical or statistical analysis. All exams consisted of three completed test charts. [Refer to Nelson (2015) and Department of Defense (2006) for general information on the comparison question test and how the sample cases were conducted.]

All of the sample cases included the standard array of sensors, including upper and lower respiration sensors, an electrodermal activity sensor, and cardiovascular activity sensor from which responses would be extracted and numerical scores assigned. This sample was previously used in the development of the OSS-2 scoring method (Krapohl, 2002), at which time response features were extracted from the recorded data using a computer software program (Extract.exe, Harris, in Krapohl & McManus, 1999) that was developed to objectively extract the Kircher feature measurements from respiration, EDA and cardio data of computerized polygraph data. A total of 21600 measurements were available for the n=300 field cases with three iterations of a question sequence that included three RQs and three CQs. Data were imported to the R Language for Statistical Computing (R Core Team, 2019) for analysis.

Analysis

All iterations of all relevant questions (RQs) were evaluated using the comparison question selected according to the standardized procedure for the FZCT format. The RQs are labeled R5, R7 and R10, while the comparison questions (CQs) are labelled C4, C6, and C9. For each sensor, the first RQ, R5, was evaluated with CQs that are immediately preceding and immediately following the RQ – either C4 or C6 on the first recorded chart, though the questions may be rotated for subsequent charts – depending on which CQ produced the greater change in physiological activity. The second and third RQs, R7, and R10 were evaluated with the preceding CQ – C6 for R7 and C9 for R10, though the order may be rotated for some recorded test charts. An RQ/CQ ratio., referred to as an RC Ratio, was calculated for each pair of questions. For EDA and cardio sensors greater extracted values indicate greater changes in physiology. In contrast, for the respiration sensor, smaller extracted values represent greater changes in physiological activity.

RC ratios will conform to an asymmetrical distribution, bounded by 0 and ∞ (infinity) with a mean of 1 and a potentially infinite range of values between 0 and 1, along with a potentially infinite range of values between 1 and infinity. When the RQ value was greater than the CQ value the RC Ratio was a value between 1 and infinity. When the CQ value was greater than the RQ value the RC Ration was a decimal value between 0 and 1. To avoid this asymmetry the natural logarithm was taken for each RC Ratio, referred to as a logRC Ratio, The resulting distribution of logRC Ratios was a symmetrical distribution with a mean of 0 and an infinite number of potential values between 0 and ∞ (infinity) along with an infinity number potential values between 0 and -∞ (negative infinity). RC Ratios between 0 and 1 produced negative logRC Ratio values between 1 and negative-infinity, while RC Ratios between 1 and infinity produced positive logRC Ratios between 0 and infinity. Table 1 shows an example of the use of the natural logarithm to produce ratios that are symmetrical around 0.

Table 1. Examples showing use of the natural logarithm to achieve a symmetrical distribution of logRC ratios.

	RQ Value	CQ Value	RC Ratio	logRC Ratio
Ex 1	300	200	1.5	0.4054651
Ex 2	200	300	.67	-0.4054651

Notice, in Table 1, how RC Ratios are not symmetrical around 1 while logRC Ratios are symmetrical around 0. This symmetry make it possible to make use of linear statistical calculations such as the correlation coefficient. Before proceeding further, it was necessary to adjust the sign values of the logRC Ratios for EDA and cardio data so that negative logRC Ratios correspond to deceptive scores while positive logRC Ratios correspond to truthful scores for all sensor, including the respiration, EDA and cardio.

Twenty-one thousand six-hundred (21600) measurements were taken from the n=300 cases, from which a total of 10800 logRC Ratios were calculated for the three iterations of three RQs, for the thoracic and abdominal respiration sensors, EDA sensor and cardio sensor for each of the n=300 sample cases. After combining the data for the thoracic and abdominal respiration sensors there were 8100 logRC Ratios, including 2700 values for each recording sensor: respiration, EDA, and cardio. To remain consistent with the familiar intuition for integer scores used in polygraph field practice, logRC Ratio of + sign value correspond to truth-telling while integer scores of – sign value correspond to deception. During the course of the analysis, automated signed integer scores were assigned to the 10800 logRC Ratios using the bigger-is-better-rule (BIBR: National Center for Credibility Assessment, 2017). Numerical scores of this type are similar to the scores that human experts would assign using manual scoring methods such as the Federal three-position scoring method (National Center for Credibility Assessment, 2017) or the Empirical Scoring System (ESS: Nelson, Krapohl & Handler, 2008).

Respiration data

Thoracic and abdominal logRC Ratios were combined to a single vector of 5400 values and the point-biserial correlation for respiration scores was r_{pb}=.184. For the thoracic respiration sensor alone, the value was r_{pb}=.209, and for the abdominal sensor alone it was r_{pb}=.161. For the combined respiration sensor data, the maximum logRC Ratio was 3.8 (a ratio of 45:1). A maximum constraint value was applied iteratively from +/-2 to +/0 3.5 and was found to optimize the point-biserial correlations at +/-2.7 (a ratio of 14.9). That is, logRC Ratios were coerced to zero (0) if they exceeded the values 2.7 or -2.7. LogRC Ratios for respiration data were more likely to contribute to incorrect scores than to correct scores when they exceeded this level. Twenty-four (24) of the 5400 respiration scores (<0.5%) exceeded the maximum constraint. With the maximum constraint value, the point-biserial correlations were r_{pb}=.216 for the thoracic sensor and r_{pb}=.182 for the abdominal sensor. An outer or maximum constraint can improve the correlation coefficients for aggregated respiration scores. However, because the goal of this project was to study optimal minimum constraints no further optimization was performed on the outer constraint for respiration scores. The maximum constraint ratio was retained for the remainder of the analysis.

Respiration scores for the thoracic and abdominal sensors were then combined to a single set of 2700 scores using the procedure described by Nelson and Krapohl (2017). Using that procedure. The combined logRC Ratio was coerced to zero (0) if the sign values are opposite, and was set to the value with the greater absolute value if not opposite. After combining the two logRC Ratios to a single respiration score for each iteration of each RQ for each case the point-biserial correlation was r_{pb}=.211. Respiration data were also combined by averaging the logRC Ratios for the two respiration sensors. The point-biserial correlation for the averaging method of combining the sensor data was r_{pb}=.215, and exceeded that of the procedural method. For the combined respiration vector, five (5) of the 2700 respiration logRC Ratios were zero (0). Separate vectors of logRC Ratios were retained for analysis, including 2700 values for the thoracic sensor and 2700 values for the abdominal sensors. Thoracic and abdominal information would be combined in a later step for each iteration of a series of minimum ratio constraints.

The logRC Ratios were then standardized to a mean of zero (0) and standard deviation of one (1). There was no difference in the correlation (r_{pb}=.215) for the standardized logRC Ratios for the averaged thoracic and abdominal respiration sensors. Because standardization offered no advantage, the remainder of the analysis was completed with the un-standardized respiration data. Standardized values will have a common metric, with mean=0 and sd=1, and can be calculated at a later time when combining data from different sensors.

LogRC Ratios for combined respiration data were then aggregated by averaging the three iterations of the three RQs for each case. The point-biserial correlation for the mean logRC Ratios for respiration data was r_{pb}=.408. For the individual sensors the point-biserial correlation after aggregating the scores for each case was r_{pb}=.420 for the thoracic sensor and r_{pb}=.359 for the abdominal sensor with the outer constraint. Without the maximum constraint the correlations were r_{pb}=.401 for the thoracic and r_{pb}b=.317 for the abdominal. It is not surprising that the aggregated logRC Ratios for each case produced a substantially stronger correlation coefficient than the logRC Ratios for each presentation of each RQ. Aggregating data has the effect of improving the signal-to-noise ratio within the information extracted from the recorded data.

A series of minimum constraints was evaluated, from 1.05:1 to 2:1 in increments of .05. Ratios from 1.05 to 2 were transformed to their natural logarithms so that they could be applied symmetrically to the logRC Ratios which have + and – sign values similar to the intuition that field polygraph examiners use for truthful and deceptive numerical scores. The series of possible constraints was applied iteratively to the sample of n=300 cases. For each iteration of the series of minimum constraints, logRC Ratios were coerced to zero if they did not exceed the constraint. Integer scores using the BIBR were also coerced to zero for values for which the logRC was coerced to zero. The constraint value was applied separately to the thoracic and abdominal sensor data before combining the sensor data using the method described by Nelson & Krapohl (2017). Log RC Ratios were aggregated by averaging all iterations of all RQs for each case, and the point-biserial correlation was then calculated for the case criterion states coded as [+1, -1].

LogRC Ratios for each presentation of each RQ were then transformed to signed integer scores for which the thoracic and abdominal scores were combined using the method described earlier. Sign values for both the logRC Ratios and the signed integer scores conform to the familiar intuition for sign scores used by polygraph field examiners. Scores of positive (+) value correspond to truth-telling, and scores of negative (-) sign value correspond to deception. For each iteration of the series of minimum constraint ratios the proportion of non-zero logRC Ratios was calculated – also the proportion of non-zero signed integer scores – for the 2700 scores after combining the thoracic and abdominal sensor data. The proportion of correct non-zero signed integer scores was calculated by comparing the integer scores with the case criterion state coded as [+, -]. Finally, the signed integer scores were summed for each case and the point-biserial correlation was calculated for the numerical scores and the criterion state, coded as [+1, -1] for each iteration of the minimum constraint ratio.

For each iteration of the minimum constraint ratio the correlation of the 2700 signed integer scores and the case criterion state was calculated using a procedure similar to the detection efficiency coefficient (DEC; Kircher, Horowitz, & Raskin, 1988). DEC is calculated as the Pearson correlation between integer score codes [+1, 0, -1] and the criterion state [+1, -1], and are informative because they represent strength of information about correct, incorrect, and inconclusive outcomes in a single correlation statistic. This application of the DEC differed from its normal use in that DEC was initially described for use with classifications made with numerical or statistical cutscores using aggregated scores for each case using a complete array of recording sensors; it is use here with the individual scores for a single senor and with no numerical cutscores. In this usage the DEC correlation can the thought of as a numerical score correlation; it provides a measurement of the strength of information from the numerical scores at each minimum constraint ratio.

EDA data

For EDA scores the point-biserial correlation for the 2700 logRC Ratios was r_{pb}=.433. The maximum logRC Ratio was 3.4, corresponding to a ratio of 30:1 where the CQ value exceeded the value of the RQ. The minimum logRC Ratio was -4.7, corresponding to a ratio of 110:1, where the RQ value exceeded the CQ value.

A maximum constraint ratio was applied iteratively from +/-2 to +/-20 and was found to maximize the point-biserial correlation with a maximum constraint ratio of 7:1 with r_{pb}=.439. EDA LogRC Ratios that exceeded this level were more likely to contribute to incorrect scores than to correct scores. The remainder of the analysis was completed with this maximum constraint ratio. Ninety (90) of the 2700 logRC Ratios (3.3%) exceeded this constraint value.

LogRC Ratios were then standardized to evaluate the effect on the correlation coefficients. Standardizing the logRC Ratios did not change the point-biserial correlations. Because this project did not involve the aggregation of data between sensors, the remainder of the analysis was completed with the un-standardized logRC Ratios. Standardization will be advantageous, and can be accomplished at a later

After aggregating the logRC Ratios for all iterations of all RQs for each of the n=300 sample cases the point-biserial correlation was r_{pb}=.751. Aggregating data has the effect of improving the signal to noise ratio, and it is therefore not surprising that the correlation for aggregated logRC Ratios exceeded the correlation for the individual logRC Ratios.

A series of minimum constraint ratios was evaluated from 1.05:1 to 2:1 in increments of .05. Results for the EDA data were re-calculated for each iteration of the constraint, including the point-biserial correlation of the aggregated logRC Ratios with the criterion state of each of the n=300 sample cases. For each iteration of the series of minimum constraints, logRC Ratios were coerced to zero if they did not exceed the constraint. The 2700 logRC Ratios were also transformed to signed integer scores [+1, 0, -1] and the criterion correlation with the sign scores was calculated for each iteration of the minimum constraint ratio. Results were also calculated for the proportion of correct sign scores and the proportion of non-zero scores. Finally, the signed integer scores were summed for each of the n=300 cases and the correlation was calculated of the aggregated integer scores with the case criterion state.

Cardio data

For cardio scores the point-biserial correlation for the 2700 logRC Ratios was r_{pb}=.179. The maximum logRC Ratio was 1.9 which corresponds to a ratio of 6.7:1. The minimum logRC Ratio was -3.9, corresponding to a ratio of 52:1.

A maximum constraint ratio was applied iteratively to the cardio data from +/-2 to +/-20 and was found to optimize the point-biserial correlation with a maximum constraint ratio of 12:1 with rpb=.180. LogRC Ratios that exceeded this the 12:1 level were more likely to contribute to incorrect cardio scores than to correct scores. The remainder of the analysis was completed with this maximum constraint ratio. Two (2) of the 2700 logRC Ratios exceeded the 12:1 constraint value.

LogRC Ratios were then standardized to evaluate the effect on the correlation coefficients. Standardizing the logRC Ratios did not change the point-biserial correlations. Because this project did not involve the aggregation of data between sensors, analysis further analysis of the cardio data was completed with the un-standardized logRC Ratios. Standardization of the cardio can be accomplished at a later time when data are combined for the array of recording sensors.

After aggregating the logRC Ratios for all iterations of all RQs for each of the n=300 sample cases the point-biserial correlation was r_{pb}=.460. Aggregating the cardio data has the effect of improving the signal to noise ratio, and it is therefore not surprising that the correlation for aggregated logRC Ratios exceeded the correlation for the individual logRC Ratios.

The same series of minimum constraint ratios, from 1.05:1 to 2:1 in increments of .05, was applied to the cardio data. Results were re-calculated for each iteration of the constraint, including the point-biserial correlation of the aggregated logRC Ratios with the criterion state of each of the n=300 sample cases. For each iteration of the series of minimum constraints, logRC Ratios were coerced to zero if they did not exceed the constraint. The 2700 logRC Ratios were also transformed to signed integer scores [+1, 0, -1] and the criterion correlation with the sign scores was calculated. Results were also calculated for the proportion of correct sign scores and the proportion of non-zero scores for the cardio data. Finally, the signed integer scores were summed for each of the n=300 cases and the correlation was calculated of the aggregated cardio integer scores with the case criterion state.

Results

Table 2 shows the mean point-biserial correlation coefficients for the 2700 logRC Ratios and the criterion state for each recording sensor, along with the mean point-biserial correlation coefficient for the aggregated logRC Ratios and summed integer scores. It can be seen in Table 2 that correlations for aggregated scores exceed those of the individual scores. This is an example of the value of using polygraph test formats with multiple RQs and multiple iterations of the question sequence.

Table 2. Point-biserial correlations for logRC Ratios, aggregated logRC Ratio, and summed integer scores.

	logRC Ratios (2700 scores)	Aggregated logRC Ratios (n=300)	Summed Integer Scores (n=300)
Respiration	UP=.216, LP=.182, comb.=.211	UP=.420, LP=.359, comb=.401	.228
EDA	.439	.751	.749
Cardio	.183	.460	.410

Figure 1 shows a plot of the series of minimum ratio constraints, from 1.05:1 to 2:1, with the respiration data, including the point-biserial correlation for the logRC Ratios and numerical scores after aggregating the data for each case. Also shown is the point-biserial correlation for the 2700 integer scores, along with proportion of non-zero scores at each minimum constraint level and the proportion of correct non-zero scores. Figure 2 shows a plot of the same information for the EDA data. Figure 3 shows the same information for the cardio data.

Figure 1. Pneumo Constraint Analysis

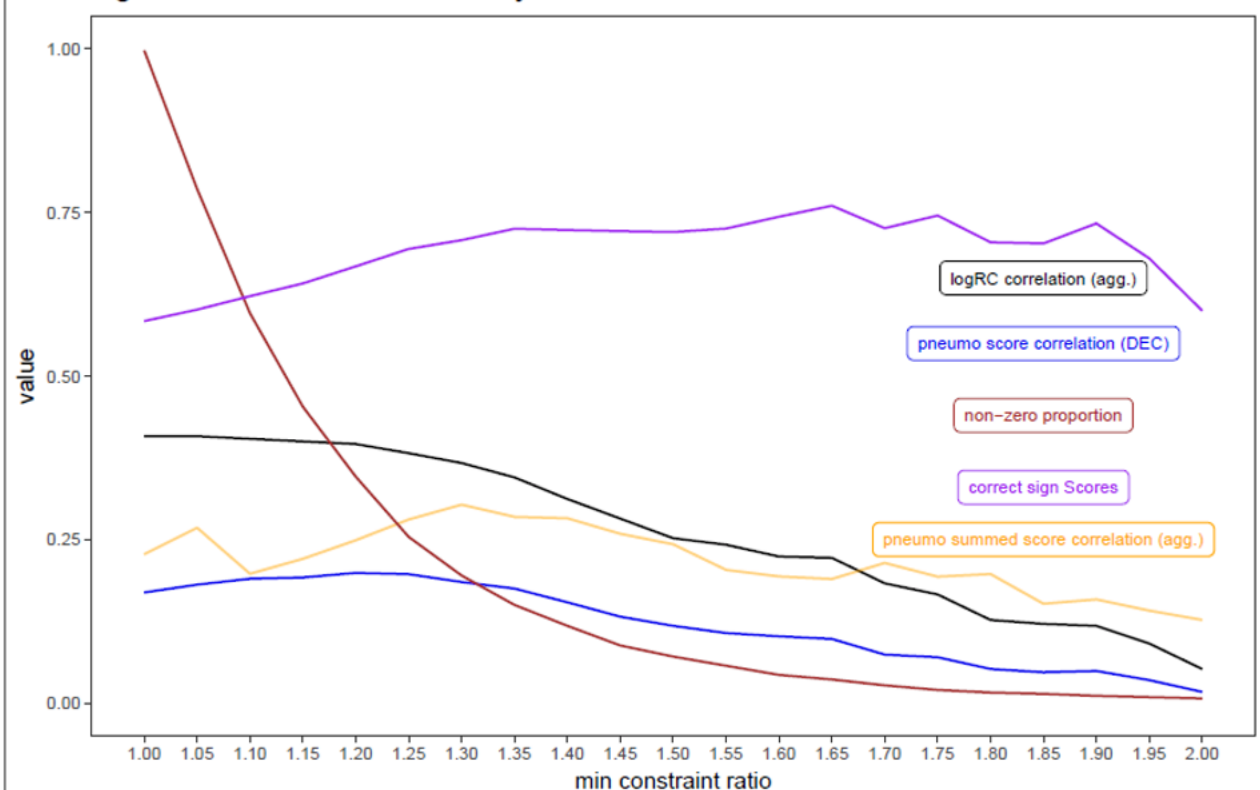


Figure 2. EDA Constraint Analysis

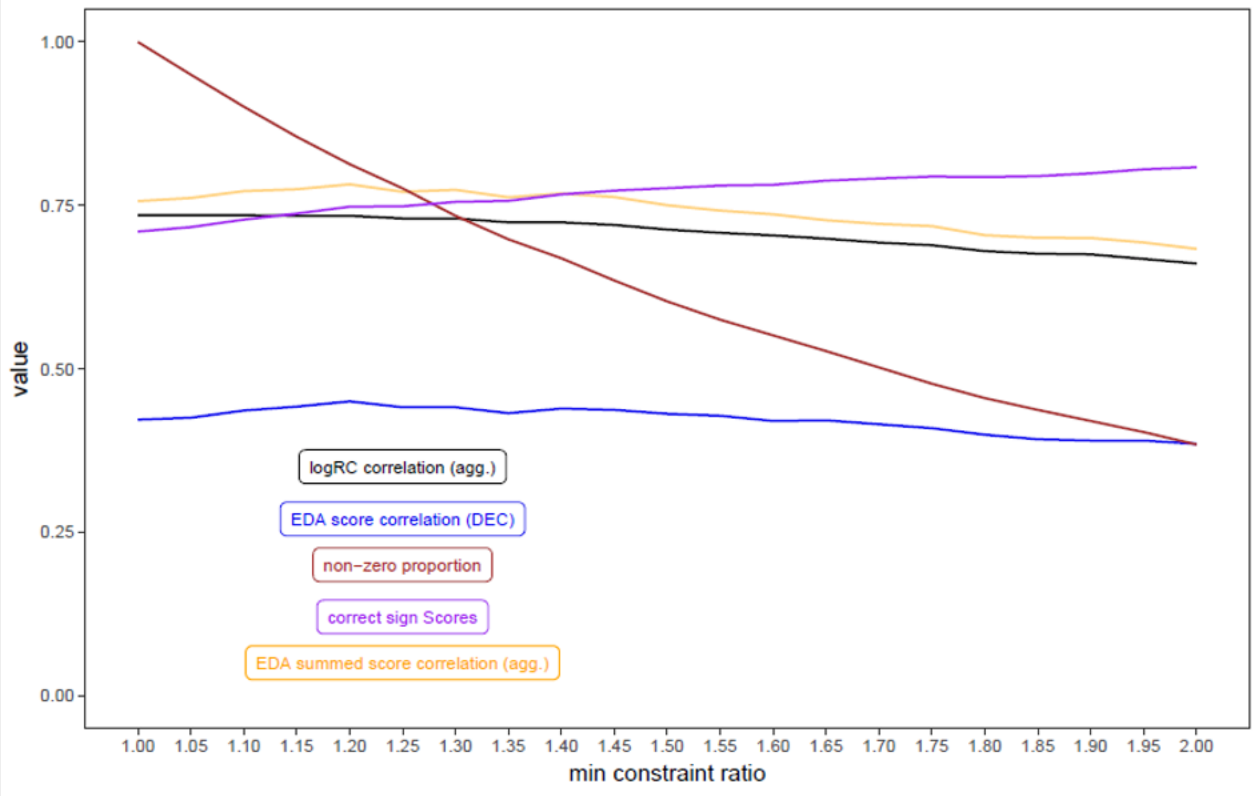
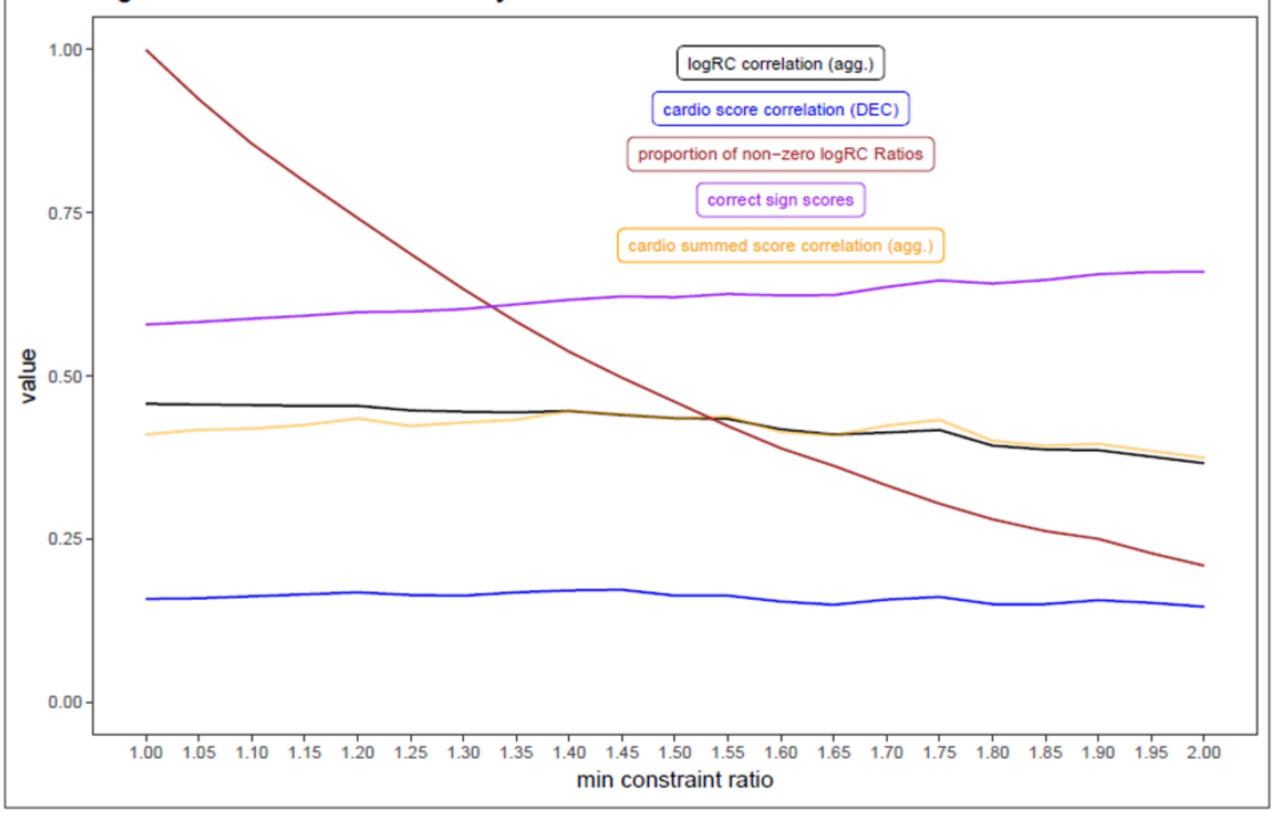


Figure 3. Cardio Constraint Analysis



Discussion

For each of the recording sensors, respiration, EDA, and cardio, the point-biserial correlation between the aggregated logRC Ratios was greatest with no minimum constraint. Application of the series of increasing minimum constraints resulted in continuous weakening of the point-biserial coefficient. Not surprisingly, for each of the recording sensors the proportion of non-zero values – both logRC Ratios and numerical scores – was greatest with no minimum constraint and became progressively smaller as the minimum constraint ratio increase. This effect was most pronounced for the respiration data, for which 95% of the scores were zero (0) at a minimum constraint ratio of 1.6:1. In contrast, approximately 55% of the EDA scores were non-zero and 39% of the cardio scores were non-zero at the same (1.6:1) ratio. The correlation coefficient (similar to DEC) for the 2700 numerical scores was largely unaffected for cardio data. The correlation coefficient for EDA scores was minimally affected by the series of constraints, beginning at .411 and rising slightly to .450 at ratio of 1.2:1, and ending at .385 at the maximum constraint ratio of 2:1.

The proportion of correct signed scores – both logRC Ratios and numerical scores – increased to small degree across the range of increasing minimum constraint ratios for the EDA and cardio data. However, the magnitude of this increase was substantially less than the increase in the number of scores of zero (0) for these sensors. For respiration data, the proportion of correct signed scores increased to a peak at a ratio of 1.35:1 and became unstable a ratios exceeding 1.6:1. This instability can be attributed to the small number of non-zero-scores that remained at constraint ratios of 1.6:1 and higher.

The point-biserial correlation for the logRC Ratios (shown in black in Figures 1, 2, and 3), together with the score correlation (shown in blue) – similar to a DEC correlation for result, but calculated in this analysis with no numerical cutscores – provides a convenient synthesis of the complex information contained in this analysis. This correlations captures information about correct, incorrect and null (0) scores in a single numerical index for which the familiar intuition for correlation coefficients can be applied. For cardio scores no minimum constraint can be identified that will increase the effectiveness of the scores that can be extracted from recorded data. For EDA scores the effect of a minimum constraint ratio to improve the correlation coefficient of numerical scores was minimal. What remains is whether any statistically significant advantage exists for the use of a minimum constraint ratio for numerical scores. However, these data suggest that there is no advantage to the use of a minimum constraint with the logRC Ratios used in automated scoring methods – and this same conclusion would be observed using ratios without a log transformation. Score correlations for respiration scores, together with the aggregated score correlation (shown in orange), suggest that constraining the score extraction to the range from 1.2:1 and 1.6:1 may be useful to optimize the contribution of respiration scores to correct vs incorrect conclusions.

One obvious limitation of this study is the lack of any test of statistical significance. Inclusion of such a test is possible, but would require complex methodology that would substantially increase the burden to readers, and might reduce the level of interest in this important topic. Statistical optimization of feature extraction and numerical scores is a non-trivial analytic challenge that deserves greater attention in publication. It was thought that limiting this project to a correlation study, and the presentation of high dimensional analytic results in the form of a three graphic plots, might serve to maintain the readability and clarity among interested readers. Data from this analysis support the validity of the BIBR as a reasonable solution, and suggest that no minimum constraint ratio is needed for automated analysis methods. (Computer scoring algorithms have already presumably made use of any measurable difference that could be extracted from relevant and comparison questions.) Another limitation of this analysis is the lack of a second sample with which to compare these results. Continued interest and research is recommended in the optimization of feature extraction and numerical transformation for both automated and manual test data analysis methods.

References

Kircher, J.C., Horowitz, S.W, and Raskin, D.C. (1988). Meta-analysis of mock crime studies of the control question polygraph technique. *Law and Human Behavior, 12*(1), 79 – 90.

Krapohl, D. & McManus, B. (1999). An objective method for manually scoring polygraph data. *Polygraph, 28*, 209-222.

Krapohl, D.J. (2002). Short report: Update for the Objective Scoring System. *Polygraph, 31(*4), 298 – 302.

National Center for Credibility Assessment (2017). *Test Data Analysis: Numerical Evaluation Scoring System Pamphlet*. Available from the author. (Retrieved from http://www.antipolygraph.org on 4-13-2019).

Nelson, R. & Krapohl, D. K. (2017) Practical polygraph: a recommendation for combining the upper and lower respiration data for a single respiration score. *APA Magazine. 50*(6), 31-41.

Nelson, R., Krapohl, D.J., and Handler, M. (2008). Brute force comparison: A Monte Carlo study of the Objective Scoring System version 3 (OSS-3) and human polygraph scorers. *Polygraph, 37*(3), 185-215.

Nelson, R. & Handler, M. (2015). Statistical reference distributions for comparison question polygraphs. *Polygraph, 44*(1), 91-114.

Nelson, R. & Handler, M. (2018). Practical polygraph: seven things to know about feature extraction with electrodermal and cardio data. *APA Magazine, 51*(1), 116-132.

Nelson, R. (2017). Multinomial reference distributions for the Empirical Scoring System. *Polygraph and Forensic Credibility Assessment, 46* (2), 81-115.

Nelson, R. (2018). Multinomial reference distributions for three-position scores of comparison question polygraph examinations. *Polygraph and Forensic Credibility Assessment, 47*(2), 158-175.

Peng, R. (2011). Reproducible research in computational science. *Science, 334*(6060), 1226–1227.

R Core Team (2019). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.