



A Brief History of Scientific Reviews of Polygraph Accuracy Research

Raymond Nelson and Mark Handler

How accurate is the polygraph? A deceptively simple question. It is simple to the point of being simplistic. Oversimplified. Which is to say that the answer is not so simple. Part of the complexity is that there is no single numerical index that can adequately represent a test result for which the answer is a probability with some degree of uncertainty or range of error – and for which discussions of test accuracy can take several different dimensional views, including sensitivity, specificity, error rates and other descriptive forms. And while

the question of accuracy is not so simple, the question is also not so complex that it cannot be studied and answered, albeit not in a simplistic manner.

To reduce the complexity of scientific discussions, accuracy discussions in the context of research and testing will often take the form of discussions of the error rate or probability of error (also referred to as a p-value, and commonly expressed as a decimal). The expectation to quantify the likelihood or probability of an erroneous result is among the hallmarks or standards of scientific pursuits.

Probabilities of error may well serve the expediency needs of scientific thinkers, but persons untrained in scientific methodologies may find it easier to work with a different concept – the inverse of the error rate, referred to conveniently as a confidence level (CL) with which our estimate of the probability of a correct result is commonly expressed in terms of a percentage. Because polygraph research involves the accuracy of both truthful and deceptive results, the simple unweighted average of these is among the more useful numbers we can use in attempt to describe the complex phenomena of polygraph accuracy with a single number. This can be thought of as unweighted accuracy, for which we average the accuracy of truthful results with deceptive results, while assuming uniform base-rates of guilty and innocent persons.

In the same way that we do not expect to achieve a generalizable estimate of polygraph accuracy from a single case, we also do not expect to achieve generalizable estimates of polygraph accuracy from a single study. Instead generalizable estimates of polygraph accuracy are more likely to be achieved by systematically reviewing the results from multiple studies. An even broader understanding of polygraph accuracy might be gained from a historical evaluation of the results of published scientific reviews.

A Chronology of Scientific Reviews of Polygraph Accuracy

Among the earliest published scientific reviews of polygraph accuracy was that of Abrams (1973), who surveyed the published literature dating to the earlier part of the 20th century, and reported a not unimpressive aggregated accuracy rate of .980.

A few years later, Abrams (1977) published the results of another review of polygraph accuracy, and then reported an estimate of .910 using more recent studies and more carefully structured study selection criteria.

Podlesny and Raskin (1978) also surveyed the available literature in attempt to quantify an answer to the questions of polygraph accuracy and reported an estimated accurate rate of .890 at that time.

Ansley (1983) also attempted to quantify the available literature on polygraph accuracy and reported the results of 1,964 laboratory cases and 1,113 field cases. He described a decision accuracy level of .968, excluding inconclusive results. Results using the relevant-irrelevant technique were reported as more accurate (.960) than those using comparison question technique (.952). Results from concealed information tests were reported as having an accuracy rate of .912.

At about that same time the Office of Technology Assessment (OTA, 1983) reported the results of laboratory studies as having an aggregated accuracy rate of .832, while the accuracy of field studies were reported as .847.

Abrams (1989) reported the results of a third survey of polygraph accuracy, and described an aggregated accuracy rate of .880 – somewhat more conservative than his earlier calculations.

Ansley (1990) published the another scientific review of the literature on polygraph accuracy, and summarized the results of 10 studies including 2,042 criminal investigation polygraphs since 1980. He reported an accuracy rate of .980 for guilty cases and .970 for innocent cases. Also reported were the results of a survey of 11 studies of blind evaluations of 922 criminal investigation polygraphs, with an overall accuracy rate of .900, including .940 for guilty cases and .890 for innocent cases.

Honts and Peterson (1997) summarized the results of high quality polygraph studies in an Amicus brief filed in *U.S. Vs Scheffer*, and reported the accurate rate as exceeding .900. These results were also reported by Raskin and Honts (2002).

In a independent study funded by the U.S. Department of Defense, Crewson (2002) surveyed the results of diagnostic and screening polygraphs and compared the results to screening and diagnostic tests used in medicine and psychology. Diagnostic polygraph tests were reported to have an aggregated accuracy rate of .880 while screening tests were reported to have an accuracy rate of .740 at that time.¹

A more recent survey of polygraph accuracy was completed by the National Research Council (NRC, 2003) who described test accuracy in terms of the area under the curve (AUC) for the receiver operating characteristic (ROC). Median AUC was reported as .860 for laboratory studies and .890 for field studies.²

¹All of the screening studies in Crewson (2002) were conducted using the relevant-irrelevant technique. This figure cannot be generalized to screening methods using comparison question techniques.

²The slightly higher accuracy of field studies compared to laboratory studies has been observed on various occasions. Differences are not statistically significant but is inconsistent with the trend of observations in other fields of science for which field studies with more less tightly controlled conditions most often produce slightly lower accuracy than laboratory studies. One interpretation of the meaning of this might be superiority of field studies over laboratory research. A more conservative interpretation would be that field studies are subject to inherent bias and non-random selection processes.

Using a different approach to estimate test accuracy, Kokish, Levenson and Blasingame (2005) reported the results of an opinion survey of convicted sex offenders who were subject to probation requirements for polygraph testing. They reported that respondents described a high rate of agreement with the polygraph test results, over .900.

In a meta-analytic survey of criterion validity, the American Polygraph Association (APA, 2011) estimated the accuracy of event-specific diagnostic polygraphs as .890 with a 95% confidence range of .829 to .951. Polygraph techniques interpreted with the assumption that the criterion states of the test target questions vary independently were reported as having an aggregated accurate rate of .850 with a 95% confidence range of .773 to .926. The aggregated accuracy rate for all

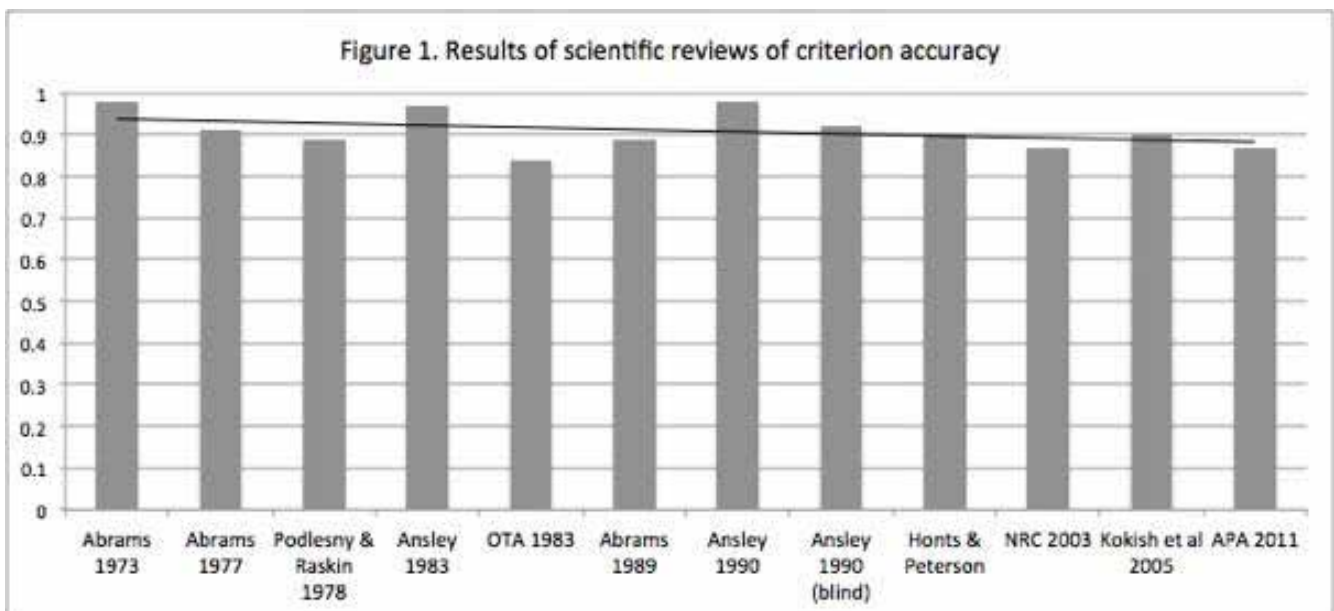
techniques was reported as .869 with a 95% confidence range of .798 to .940.

Figure 1 shows the results of the series of scientific reviews of published studies on criterion accuracy of polygraph tests, along with the trend-line.

Discussion

The mean accuracy for these scientific reviews was .900, and the median was .890. Excluding the three studies with extremely high accuracy, both the mean and median accuracy of the scientific reviews of polygraph accuracy research were .890.

While the scientific reviews included herein are perhaps not comprehensive or inclusive of all polygraph studies, they



are reasonably inclusive of most useful and visible attempts to systematically account for the evidence and variance that is inevitably present in the results of decades of studies of varying quality. The majority of the results of these studies appear to be within understandable error tolerances of each other. Although three reviews (Abrams, 1973; Ansley, 1983, 1990) are more optimistic than others, with results in the high .90s, the results of the majority of these studies (Abrams; 1977, 1989; Raskin and Podlesney, 1979; the OTA, 1983; Honts and Peterson, 1997); Raskin and Honts, 2002; the NRC, 2002; Kokish, Levenson and Blasingame, 2005; and the APA, 2011) suggest that polygraph accuracy exists within a predictable range in the mid to high .80s extending into the low .90s.

One of the most obvious trends in this data is that studies conducted by academics with no particular interest in the polygraph have tended to produce results that are slightly less than the historical average estimates produce by researchers within the polygraph profession. This trend also appears to be correcting itself over time as researchers within the profession have begun to endeavor to apply more rigorous methodology and scrutiny to question about criterion accuracy.

A second obvious trend is that the results of field and laboratory studies have tended to agree closely. At this time there is no available evidence to suggest one form of research as superior to another. Indeed different types of research designs all have advantages and disadvantages. It is inherently beneficial to the profession to use a variety of types of research methodologies.

Another trend is that results produced by researchers within the polygraph profession and those provided by academics and researchers outside the profession have tended to both agree more than disagree, and to begin to converge more closely over time.

A fourth obvious trend is that estimates of polygraph accuracy have tended to decrease over time. While it is possible that this is a reflection of a decrease in polygraph test accuracy, it is probably not likely that the polygraph has become less accurate over time. A more likely explanation is that estimates of polygraph accuracy have simply become more realistic or more accurate. A look at the study details will show that early studies were conducted with heavy emphasis on the use of confirmed field cases – which may introduce a vulnerability to overestimation of test accuracy as a result

of the systematic exclusion of both false-negative and false-positive error case for which no confession or confirmation evidence could be obtained.³ It is not likely that researchers were intentionally exploiting this potential; instead it is more likely that some early research and research reviews overestimated polygraph accuracy as a result of an unrecognized sampling vulnerability. Regardless, although claims of extremely high or near-perfect accurate are not well supported, data are available at this time to support the hypothesis that polygraph testing can provide useably high levels of test accuracy. A final observation is also interesting. Although more rigorous and scientific approaches to polygraph testing is thought to increase polygraph accuracy, the degree variance in the published evidence has made it difficult to ascertain whether intended improvements have actually manifested themselves in observable increases in test accuracy. It can be seen that results reported by Honts and Raskin (1997) and Raskin and Honts (2002) have marginally exceeded the .900 level, but most other scientific reviews of the literature do not exceed this level.

This may be a reflection of the practice of polygraph methods based more on tradition and less on rigorous scientific models. But it may also begin to be a reflection that polygraph test accuracy is nearing a form of ceiling within which observable and measurable increases in accuracy may be difficult to obtain without rigorous adherence to scientific protocols and evidence-based practices.

Polygraph research has been ongoing for decades, and there is a great deal of information in the published literature. It is almost inevitable that the data contain some valuable information while also being somewhat noisy. In the same way that individual test results have some potential for error, research results also have some potential for error. Given the length of time and evolution of statistical and research methods, it is not surprising that there may also be some observable variability in the results of scientific reviews of the published evidence. The usual trend of aggregated data in most fields of science, measurement and statistical prediction is that aggregated testing results tend to outperform

³ Confessions themselves may be the result of an inherently non-random decision to pursue more information, raising an important caution about the need for replication of results observed through the study of sampling data acquired through confession sampling. Moreover, if the decision to pursue confessions is based on the results of a polygraph test, then there may be inherent non-independence of the external criterion leading to the potential for self-fulfilling research results.

individual testing results and aggregated research results have tended to outperform the results of individual studies. For this

reason, it is hoped that some value can be found in this brief overview of the historical data.

References

- Abrams, S. (1973). Polygraph validity and reliability: A review. *Journal of Forensic Sciences*, 18, 313-326.
- Abrams, S. (1977). *A polygraph handbook for attorneys*. Lexington, MA: Lexington Books.
- Abrams, S. (1989). *The complete polygraph handbook*. Lexington, MA: Lexington Books.
- American Polygraph Association (2011). Report of the Ad Hoc Committee on validated techniques. *Polygraph*, 40(4), 196-305.
- Ansley, N. (1983). A compendium on polygraph validity. *Polygraph*, 12, 53-61.
- Ansley, N. (1990). The validity and reliability of polygraph decisions in real cases. *Polygraph*, 19, 169-181.
- Crewson, P. E. (2001). A comparative analysis of polygraph with other screening and diagnostic tools. Research Support Service. Report No. DoDPI01-R-0003. Reprinted in *Polygraph* 32, (57-85).
- Honts, C. R. & Peterson, C.F. (1997). Brief of the Committee of Concerned Social Scientists as Amicus Curiae *United States v Scheffer*. Available from the author.
- National Research Council (2003). *The Polygraph and Lie Detection*. Washington, D.C.: National Academy of Sciences.
- Office of Technology Assessment (1983). *The validity of polygraph testing: A research review and evaluation*. Washington, D.C.: U.S. Congress, Office of Technology Assessment.
- Kokish, R., Levenson, J. S., & Blasingame, G. D. (2005). Post-conviction sex offender polygraph examination: client-reported perceptions of utility and accuracy. *Sexual Abuse: A Journal of Research and Treatment*, 17, 211-21.
- Podlesny, J. A. & Raskin, D. C. (1978). Effectiveness of techniques and physiological measures in the detection of deception. *Psychophysiology*, 15, 344-359.
- Raskin, D. C. & Honts, C. R. (2002). Handbook of polygraph testing. In M. Kleiner (Ed.), *Handbook of Polygraph Testing*. San Diego: Academic Press.