



Breve Historia de las Revisiones Científicas de la Investigación sobre la Precisión del Polígrafo

Raymond Nelson y Mark Handler

Qué tan preciso es el polígrafo? Una pregunta engañosamente simple. Es simple hasta el punto de ser simplista. Demasiado simple. Lo

que quiere decir que la respuesta no es tan simple. Parte de la complejidad es que no existe un índice numérico único que pueda representar adecuadamente el resultado de una prueba para el cual la respuesta es una probabilidad con cierto grado de incertidumbre o rango de error, y para el cual las discusiones sobre la precisión de la prueba pueden tomar vistas dimensionales diferentes, incluyendo sensibilidad, especificidad, tasas de error y otras formas descriptivas. Y aunque la cuestión de

la precisión no es tan simple, tampoco es tan compleja como para que no pueda estudiarse y responderse, aunque no de manera simplista.

Para reducir la complejidad de las discusiones científicas, las discusiones sobre la precisión en el contexto de la investigación y las pruebas a menudo tomarán la forma de discusiones sobre la tasa de error o la probabilidad de error (también conocido como valor-p y comúnmente expresado como decimal). La expectativa de cuantificar la posibilidad o probabilidad de un resultado erróneo es uno de los sellos distintivos o estándares de las actividades científicas.

Una Cronología de las Revisiones Científicas de la Precisión del Polígrafo

Las probabilidades de error bien pueden satisfacer las necesidades de conveniencia de los pensadores científicos, pero a las personas no capacitadas en metodologías científicas les puede resultar más fácil trabajar con un concepto diferente: el inverso de la tasa de error, denominado convenientemente como un nivel de confianza (NC) con el que nuestra estimación de la probabilidad de un resultado correcto se expresa comúnmente en términos de un porcentaje. Debido a que la investigación del polígrafo implica la precisión de los resultados verdaderos y engañosos, el promedio simple no ponderado de estos se encuentra entre los números más útiles que podemos usar para intentar describir el complejo fenómeno de la precisión del polígrafo con un solo número. Esto se puede considerar como una precisión no ponderada, para la cual promediamos la precisión de los resultados veraces con los resultados engañosos, asumiendo tasas base uniformes de personas culpables e inocentes.

De la misma manera que no esperamos lograr una estimación generalizable de la precisión del polígrafo a partir de un solo caso, tampoco esperamos lograr estimaciones generalizables de la precisión del polígrafo a partir de un solo estudio. En cambio, es más probable que se logren estimaciones generalizables de la precisión del polígrafo mediante la revisión sistemática de los resultados de múltiples estudios. Se puede obtener una comprensión aún más amplia de la precisión del polígrafo a partir de una evaluación histórica de los resultados de las revisiones científicas publicadas.

Entre las primeras revisiones científicas publicadas sobre la precisión del polígrafo se encuentra la de Abrams (1973), quien analizó la literatura publicada que data de la primera parte del siglo XX e informó una tasa de precisión agregada no poco impresionante de .980.

Unos años más tarde, Abrams (1977) publicó los resultados de otra revisión de la precisión del polígrafo y luego informó una estimación de .910 utilizando estudios más recientes y criterios de selección de estudios más cuidadosamente estructurados.

Podlesny y Raskin (1978) también examinaron la literatura disponible en un intento de cuantificar una respuesta a las preguntas sobre la precisión del polígrafo y reportaron una tasa de precisión estimada de .890 en ese momento.

Ansley (1983) también intentó cuantificar la literatura disponible sobre la precisión del polígrafo e informó los resultados de 1964 casos de laboratorio y 1113 casos de campo. Describió un nivel de precisión de decisión de .968, excluyendo los resultados no concluyentes. Los resultados que utilizaron la técnica relevante-irrelevante se informaron como más precisos (.960) que aquellos que utilizaron la técnica de preguntas de comparación (.952). Se informó que los resultados de las pruebas de información oculta tenían una tasa de precisión de .912.

Aproximadamente al mismo tiempo, la Oficina de Evaluación de Tecnología (OTA, 1983) informó que los resultados de los estudios de laboratorio tenían una tasa de precisión agregada de .832, mientras que la precisión de los estudios de campo se informó como .847.

Abrams (1989) informó los resultados de una tercera revisión sobre la precisión del polígrafo y describió una tasa de precisión agregada de .880, algo más conservadora que sus cálculos anteriores.

Ansley (1990) publicó otra revisión científica de la literatura sobre la precisión del polígrafo y resumió los resultados de 10 estudios que incluyeron 2042 polígrafos de investigación criminal desde 1980. Informó una tasa de precisión de .980 para casos de culpabilidad y .970 para casos de inocencia. También se informaron los resultados de una revisión de 11 estudios de evaluaciones ciegas de 922 polígrafos de investigación criminal, con una tasa de precisión general de .900, que incluye .940 para casos de culpabilidad y .890 para casos de inocencia.

Honts y Peterson (1997) resumieron los resultados de estudios de polígrafo de alta calidad en un informe “Amicus” presentado en U.S. Vs Scheffer, e informaron que la tasa de precisión excedía .900. Estos resultados también fueron informados por Raskin y Honts (2002).

En un estudio independiente financiado por el Departamento de Defensa de los EE. UU., Crewson (2002) analizó los resultados de los polígrafos diagnósticos y exploratorios, y comparó los resultados con las pruebas de exploración y diagnóstico utilizadas en medicina, y psicología. Se informó que las pruebas de polígrafo de diagnóstico tenían una tasa de precisión agregada de .880, mientras que las pruebas exploratorias tenían una tasa de precisión de .740 en ese momento.¹

El Consejo Nacional de Investigación (NRC, 2003) completó una revisión más reciente sobre la precisión del polígrafo, que describió la precisión de la prueba en términos del área bajo la curva (AUC) para la característica operativa del receptor (ROC). La media del AUC se informó como .860 para estudios de laboratorio y .890 para estudios de campo.²

¹Todos los estudios de pruebas exploratorias en Crewson (2002) se realizaron utilizando la técnica relevante-irrelevante. Esta cifra no se puede generalizar a los métodos de selección que utilizan técnicas con preguntas de comparación.

²La precisión ligeramente superior de los estudios de campo en comparación con los estudios de laboratorio se ha observado en varias ocasiones. Las diferencias no son estadísticamente significativas, pero no concuerdan con la tendencia de las observaciones en otros campos de la ciencia, en los que los estudios de campo con condiciones menos controladas suelen producir una precisión ligeramente menor que los estudios de laboratorio. Una interpretación del significado de esto podría ser la superioridad de los estudios de campo sobre la investigación de laboratorio. Una interpretación más conservadora sería que los estudios de campo están sujetos a sesgos inherentes y procesos de selección no aleatorios.

Usando un enfoque diferente para estimar la precisión de la prueba, Kokish, Levenson y Blasingame (2005) informaron los resultados de una encuesta de opinión de delincuentes sexuales condenados que estaban sujetos a requisitos de libertad condicional con prueba del polígrafo. Informaron que los encuestados describieron una alta tasa de acuerdo con los resultados de la prueba del polígrafo, más de .900.

En una revisión metaanalítica de validez de criterio, la American Polygraph Association (APA, 2011) estimó la precisión de los polígrafos de diagnóstico de eventos específicos en .890 con un rango de confianza del 95 % de .829 a .951. Se informó que las técnicas de polígrafo interpretadas con la suposición de que los estados de criterio de las preguntas objetivo de la prueba varían de forma independiente tienen una tasa de precisión agregada de .850 con un rango de confianza del 95% de .773 a .926. La tasa de

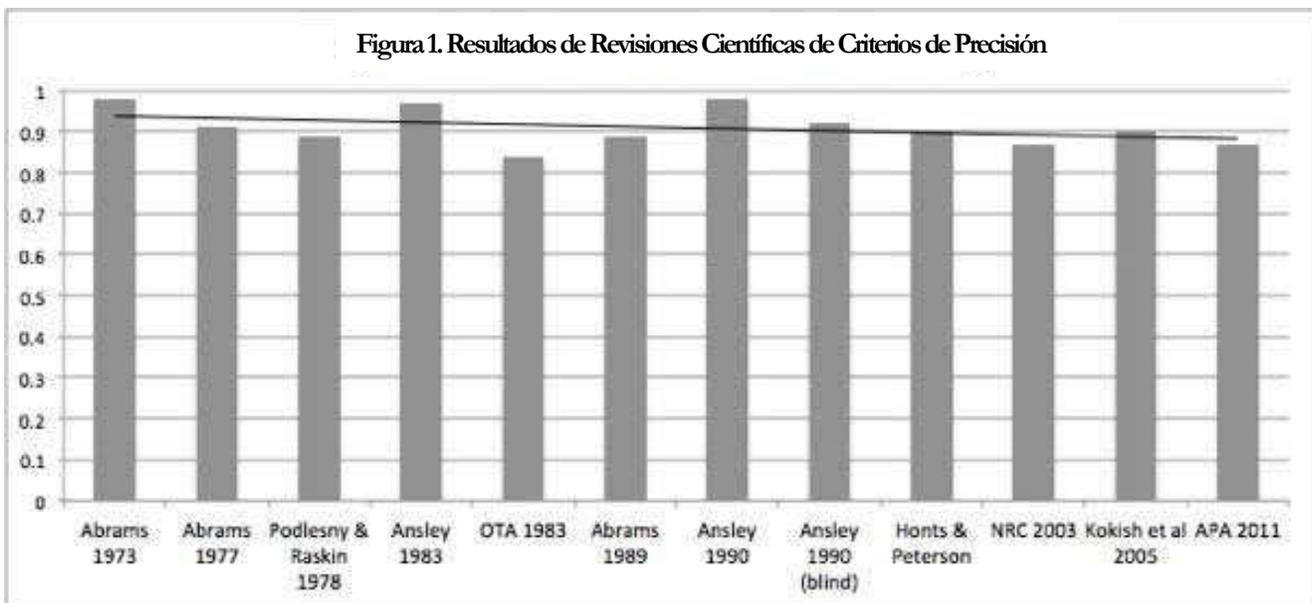
precisión agregada para todas las técnicas se informó como .869 con un rango de confianza del 95 % de .798 a .940.

La Figura 1 muestra los resultados de la serie de revisiones científicas de estudios publicados sobre la precisión de criterio de las pruebas de polígrafo, junto con la línea de tendencia.

Discusión

La precisión media de estas revisiones científicas fue de 0,900 y la mediana fue de 0,890. Excluyendo los tres estudios con una precisión extremadamente alta, tanto la precisión media como la mediana de las revisiones científicas de la investigación sobre la precisión del polígrafo fueron .890.

Si bien las revisiones científicas incluidas en este documento quizás no sean exhaustivas o inclusivas de todos los estudios de polígrafo,



incluyen razonablemente los intentos más útiles y visibles de explicar sistemáticamente la evidencia y la variación que está inevitablemente presente en los resultados de décadas de estudios de calidad variable. La mayoría de los resultados de estos estudios parecen estar dentro de tolerancias de error comprensibles entre sí. Aunque tres revisiones (Abrams, 1973; Ansley, 1983, 1990) son más optimistas que otras, con resultados en torno a los 0,90, los resultados de la mayoría de estos estudios (Abrams, 1977, 1989; Raskin y Podlesney, 1979; la OTA, 1983; Honts y Peterson, 1997); Raskin y Honts, 2002; la NRC, 2002; Kokish, Levenson y Blasingame, 2005; y la APA, 2011) sugieren que la precisión del polígrafo existe dentro de un rango predecible en los .80s medios a altos que se extienden hasta abajo de los .90s.

Una de las tendencias más obvias en estos datos es que los estudios realizados por académicos sin un interés particular en el polígrafo han tendido a producir resultados que son ligeramente inferiores a las estimaciones promedio históricas producidas por investigadores dentro de la profesión del polígrafo. Esta tendencia también parece corregirse con el tiempo, ya que los investigadores dentro de la profesión han comenzado a esforzarse por aplicar una metodología y un escrutinio más rigurosos para cuestionar la precisión de los criterios.

Una segunda tendencia obvia es que los resultados de los estudios de campo y de laboratorio han tendido a concordar estrechamente. En este momento no hay evidencia disponible que sugiera que una forma de investigación sea superior a otra. De hecho, los diferentes tipos de diseños de investigación tienen ventajas y desventajas. Es inherentemente beneficioso para la profesión utilizar una variedad de tipos de metodologías de investigación.

Otra tendencia es que los resultados producidos por investigadores dentro de la profesión del polígrafo y los proporcionados por académicos e investigadores fuera de la profesión han tendido a estar más de acuerdo que en desacuerdo, y a comenzar a converger más estrechamente con el tiempo.

Una cuarta tendencia obvia es que las estimaciones de la precisión del polígrafo tienden a disminuir con el tiempo. Si bien es posible que esto sea un reflejo de una disminución en la precisión de la prueba del polígrafo, posiblemente no sea probable que el polígrafo se haya vuelto menos preciso con el tiempo. Una explicación más probable es que las estimaciones de la precisión del polígrafo simplemente se han vuelto más realistas o precisas. Una mirada a los detalles del estudio mostrará que los primeros estudios se realizaron con gran énfasis en el uso de casos de campo confirmados, lo que puede presentar una vulnerabilidad a la sobreestimación de la precisión de la prueba como resultado de la

exclusión sistemática de los errores falsos negativos y falsos positivos para los cual no se pudo obtener ninguna confesión o evidencia de confirmación.³ No es probable que los investigadores estuvieran explotando intencionalmente este potencial; en cambio, es más probable que algunas de las primeras investigaciones y revisiones de investigación sobrestimaran la precisión del polígrafo como resultado de una vulnerabilidad de muestreo no reconocida. Independientemente, aunque las afirmaciones de precisión extremadamente alta o casi perfecta no están bien respaldadas, los datos están disponibles en este momento para respaldar la hipótesis de que las pruebas de polígrafo pueden proporcionar niveles de precisión de prueba utilizables. Una observación final también es interesante. Aunque se cree que los enfoques más rigurosos y científicos para las pruebas de polígrafo aumentan la precisión del polígrafo, la variación del grado en la evidencia publicada ha dificultado determinar si las mejoras previstas realmente se han manifestado en aumentos observables en la precisión de la prueba. Se puede ver que los resultados informados por Honts y Raskin (1997) y Raskin y Honts (2002) han excedido marginalmente el nivel de .900, pero la mayoría de las otras revisiones científicas de la literatura no superan este nivel.

Esto puede ser un reflejo de la práctica de los métodos poligráficos basados más en la tradición y menos en modelos científicos rigurosos. Pero también puede comenzar a ser un reflejo de que la precisión de la prueba del polígrafo se está acercando a una forma de techo dentro del cual los aumentos observables y medibles en la precisión pueden ser difíciles de obtener sin un cumplimiento riguroso de los protocolos científicos y las prácticas basadas en la evidencia.

La investigación del polígrafo ha estado en curso durante décadas y hay una gran cantidad de información en la literatura publicada. Es casi inevitable que los datos contengan información valiosa y al mismo tiempo sean algo ruidosos. De la misma manera que los resultados de las pruebas individuales tienen cierto potencial de error, los resultados de la investigación también tienen cierto potencial de error. Dada la duración del tiempo y la evolución de los métodos estadísticos y de investigación, no sorprende que también pueda haber alguna variabilidad observable en los resultados de las revisiones científicas de la evidencia publicada. La tendencia habitual de los datos agregados en la mayoría de los campos de la ciencia, la medición y la predicción estadística es que los resultados de las pruebas agregadas tienden a superar los

³ Las confesiones en sí mismas pueden ser el resultado de una decisión inherentemente no aleatoria de buscar más información, lo que genera una advertencia importante sobre la necesidad de replicar los resultados observados a través del estudio de los datos de muestreo adquiridos a través del muestreo de confesiones. Además, si la decisión de buscar confesiones se basa en los resultados de una prueba de polígrafo, entonces puede haber una falta de independencia inherente del criterio externo que conduce a la posibilidad de resultados de investigación autocumplidos.

resultados de las pruebas individuales y los resultados de las investigaciones agregadas tienden a superar los resultados de los estudios

individuales. Por esta razón, se espera que se pueda encontrar algún valor en esta breve descripción de los datos históricos.

Referencias

- 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.