

폴리그래프 검사의 정확도 추정*

엄진섭[†]

충북대학교 심리학과

지형기

대검찰청

박광배

충북대학교 심리학과

본 연구에서는 실제적 진실을 알지 못하는 상황에서 두 가지의 통계적인 방법을 이용하여 폴리그래프 검사의 정확도를 추정하였다. 한 가지는 대검찰청과 지방검찰청에서 2000년부터 2004년까지 범죄피의자를 대상으로 수행한 폴리그래프 검사와 검찰처분(기소여부)의 일치율을 이용하여 폴리그래프 검사의 정확도를 추정하였으며, 다른 한 가지는 대검찰청과 지방검찰청에서 2006년에 수행한 폴리그래프 검사결과와 검찰처분의 빈도분포를 가지고 잠재계층 분석을 이용하여 정확도를 추정하였다. 판단불능으로 결정된 사례를 제외하고 산출한 일치율로 폴리그래프 검사의 정확도를 추정한 결과, 2000-2004년 자료의 경우는 정확도가 .914 (SE=.004)로 산출되었으며, 2006년 자료의 경우는 정확도가 .885(SE=.021)인 것으로 산출되었다. 2006년 자료에 판단불능 사례를 포함한 후 잠재계층 분석을 이용하여 폴리그래프 검사의 정확도를 추정한 결과, 정확도가 .707~.734(SE=.027~.031)인 것으로 산출되었으며, 오류긍정율과 오류부정율은 각각 .078~.087(SE=.019~.023)과 .029~.078(SE=.010~.023)으로 추정되었다. 거짓을 말하는 사람을 정확하게 판별할 확률은 .912~.925(SE=.013~.016)로 높게 나타났고, 진실을 말하는 사람을 정확하게 판별할 확률은 .867~.955(SE=.011~.040)이었다.

주요어 : 폴리그래프, 거짓말 탐지, 정확도, 타당도, 잠재계층분석

* 이 논문은 2007년도 정부(과학기술부)의 재원으로 한국과학재단의 지원을 받아 수행된 연구임
(M10740030003-07N4003-00310)

† 교신저자 : 엄진섭, 충북대학교 심리학과, 충북 청주시 흥덕구 성봉로 410
E-mail : jseom2003@hanmail.net

한국의 검찰과 경찰, 군수사기관 등에서 거짓말을 탐지하기 위하여 매년 수천 건의 폴리그래프 검사를 실시하고 있지만, 폴리그래프 검사의 타당성은 논란의 대상이 되고 있어서 폴리그래프 검사가 법적인 효력을 인정받지 못하고 있으며(정도연, 2006), 미국의 경우에도 28개 주에서 폴리그래프 검사결과가 법적 효력을 인정받지 못하고 있다(Morris, 1989). 폴리그래프 검사의 타당도는 주로 준거타당도에 의해 평가되는데, 준거타당도는 폴리그래프의 검사결과가 거짓을 말하는 피검사자와 진실을 말하는 피검사자를 정확하게 구분해내는 정도인 정확도를 의미한다.

폴리그래프 검사의 정확도는 일반적으로 두 가지 상황에서 얻어진 자료에 근거하여 판단된다(Raskin & Honts, 2002). 한 가지는 실험실 연구를 통하여 자료를 수집하는 방법으로, 다수의 실험참여자들을 두 집단으로 나눈 후 한 집단의 실험참여자들에게만 가상의 범죄를 수행하도록 한다. 그런 다음, 모든 실험참여자들을 대상으로 폴리그래프 검사를 실시하여 검사결과와 실제상태가 얼마나 일치하는지를 판단한다. 다른 한 가지는 현장 연구로 검찰이나 경찰, 군수사기관 등에서 실제 피검사자에게 폴리그래프 검사를 실시한 자료를 수집하는 방법이다.

두 가지 자료를 이용한 폴리그래프 검사의 정확도 평가는 각각 장점과 단점을 가지고 있다. 모의범죄를 이용한 실험실 연구는 어떤 피검사자가 진실을 말하고 있고 어떤 피검사자가 거짓을 말하고 있는지에 대한 증거가 뚜렷한 반면, 현실성이 부족하여 일반화가능성이 낮다는 비평을 받고 있다. 폴리그래프 검사는 거짓말이 '탄로될 것에 대한 두려움'에 의해 야기된 반응을 측정하는데, 모의실험상

황에서의 두려움이 실제상황에서의 두려움과 동일하리라고 기대하기 어렵기 때문에 실제 폴리그래프의 정확도를 과소평가한다고 주장하는 연구자들이 있는 반면(Kircher, Horowitz, & Raskin, 1988), 실제 현장에서의 폴리그래프 검사는 실험실 실험에서와 같이 모든 것들을 완벽하게 통제된 상태에서 실시되기 어렵기 때문에 실험실 연구는 실제 폴리그래프 검사의 정확도를 과대평가한다고 주장하는 연구자도 있다(Lykken, 1998).

현장연구는 실제 피검사자에게 실시한 자료를 이용하므로 일반화 가능성의 문제는 발생하지 않지만, 검사결과와 정확도를 평가하기 위한 준거인 실제적 진실(ground truth)을 명확하게 설정하기 어렵다는 단점이 있다. 현장연구에서 피검사자가 진실을 말했는지 혹은 거짓을 말했는지에 대한 증거를 설정하는 방법은 두 가지가 주로 이용된다(Raskin, 1987). 한 가지 방법은 폴리그래프 검사 후에 피검사자가 자백을 했는지의 여부를 활용하는 것이다. 그러나, 자백은 폴리그래프 검사결과와 독립적이지 않으므로, 즉 폴리그래프 검사결과에 따라서 자백을 하는지의 여부가 좌우될 수 있고(박광배, 2002), 허위자백의 가능성이 있으며, 끝까지 자백을 하지 않는 것이 폴리그래프 검사에서 진실을 말한 것을 의미하지 않기 때문에(Kassin, 1997), 폴리그래프 검사의 정확도를 부정확하게 평가하는 경향이 있다. 다른 한 가지 방법은 전문가 패널이 범죄 사건을 검토한 후 범죄피의자가 유죄인지 혹은 무죄인지를 판단하는 것인데, 이 방법은 자백에 의존하는 방법보다도 더 큰 문제점을 가진다(Raskin, 1987). 미국에서 유죄에 대한 기준은 합리적인 의심을 넘어선(beyond reasonable doubt) 납득이 요구되기 때문에, 전문가 패널

은 실제로 무죄인 피의자를 유죄라고 잘 못 판단할 가능성보다 실제로 유죄인 피의자를 무죄라고 잘 못 판단할 가능성이 더 높은 것으로 알려져 있다(Raskin 등, 2002).

현장연구에서 폴리그래프의 정확도를 추정하는 새로운 방법으로 통계적인 방법을 이용할 수 있다. 심리검사의 분야에서는 피검사자의 진점수(true score)을 알지 못하는 상황에서도 관찰된 검사점수를 이용하여 피검사자의 진점수를 추정하며, 각 문항이 진점수에 반영되는 정도도 함께 추정된다(Crocker & Algina, 1986). 이 과정의 핵심은 진점수를 추정하기 위하여 다수의 문항을 이용하는 것이다. 실제적 진실(ground truth) 혹은 황금기준(gold standard)이 존재하지 않는 상황에서 다수의 검사결과(또는 판단)가 일치하는 정도를 이용하여 특정 검사(또는 판단)가 얼마나 정확한지를 추정할 수 있다. 두 판단의 일치율을 이용하여 정확도를 곧바로 추정할 수도 있고(Spencer, 2007), 두 판단의 교차빈도분포에 대하여 잠재계층분석을 실시하여 정확도를 추정할 수도 있다(Rindskopf & Rindskope, 1986; Pepe & Janes, 2007). 폴리그래프의 피검사자가 진실을 말하고 있는지 혹은 거짓을 말하고 있는지에 대한 또 다른 평가결과가 존재한다면, 폴리그래프 검사결과와 또 다른 평가결과간의 일치율(교차빈도분포)을 이용하여 폴리그래프의 정확도를 추정할 수 있을 것이다. 폴리그래프 검사의 정확도를 추정하는데 이용하는 표본을 검찰에서 수행된 검사로 한정한다면, 폴리그래프 피검사자에 대한 제 2의 평가결과로 검찰의 기소여부를 이용할 수 있다. 검찰은 모든 증거들을 활용하여 피검사자를 기소할 것인지(유죄의 취지), 불기소할 것인지(무죄의 취지)를 결정하므로, 검찰의 기소여부는 피검사자

가 거짓을 말하고 있는지 혹은 진실을 말하고 있는지에 대한 또 다른 평가라고 할 수 있다. 따라서 폴리그래프 검사 결과와 검찰의 기소 여부를 이용하여 폴리그래프 검사의 정확도를 추정할 수 있을 것이다.

일치율과 정확도

Spencer(2007)가 배심원 평결(유죄와 무죄)과 판사의 판단(유죄와 무죄)간 일치율로 배심원 평결의 정확도를 추정한 것과 같이, 검찰처분과 폴리그래프 검사결과 간 일치율로 폴리그래프 검사결과의 정확도를 추정할 수 있다. 어떤 사례 i 에 대하여 폴리그래프 검사결과가 옳을 확률을 p_i^B 로 표시하고, 검찰처분이 옳을 확률을 p_i^A 로 표시하고, 두 판단이 서로 일치할 확률을 ρ_i 로 표시하자. 두 판단이 서로 독립적으로 내려질 때, 두 판단이 일치할 확률은 $p_i^A p_i^B + (1-p_i^A)(1-p_i^B)$ 이 된다. 그러나 현실에서는 어떤 사례 i 에 대한 두 판단이 독립적으로 이루어지지 않는다. 두 판단이 서로 일치할 확률에서 두 판단이 서로 독립적으로 내려질 때의 일치확률을 뺀 값을 ω_i 라 하면, ρ_i 는 아래와 같이 정의된다.

$$\begin{aligned} \rho_i &= \omega_i + p_i^A p_i^B + (1-p_i^A)(1-p_i^B) \\ &= \omega_i + 2p_i^A p_i^B + 1 - p_i^A - p_i^B \end{aligned}$$

개별 사례에 대한 위의 방정식을 N 개의 사례에 대한 평균값으로 나타내면 방정식 (1)과 같다.

$$\rho = \omega + 2 \frac{\sum p_i^A p_i^B}{N} + 1 - p^A - p^B \quad (1)$$

p^A 와 p^B 의 공변량은 아래와 같으므로,

$$\begin{aligned}\sigma_{AB} &= \frac{\sum(\rho_i^A - \rho^A)(\rho_i^B - \rho^B)}{N} \\ &= \frac{\sum \rho_i^A \rho_i^B}{N} - \frac{\sum \rho_i^A \rho^B}{N} - \frac{\sum \rho^A \rho_i^B}{N} + \frac{\sum \rho^A \rho^B}{N} \\ &= \frac{\sum \rho_i^A \rho_i^B}{N} - \rho^A \rho^B - \rho^A \rho^B + \rho^A \rho^B\end{aligned}$$

방정식 (1)의 $\frac{\sum \rho_i^A \rho_i^B}{N}$ 을 $\sigma_{AB} + \rho^A \rho^B$ 로 대치하면 아래와 같고,

$$\rho = \omega + 2\sigma_{AB} + 2p^A p^B + 1 - p^A - p^B$$

$\omega + 2\sigma_{AB}$ 를 γ 로, $p^A - p^B$ 를 δ 로 바꾸어 표현하면 아래와 같이 정리된다.

$$\begin{aligned}\rho &= \gamma + 2p^A p^B + 1 - p^A - p^B \\ &= \gamma + 2(p^B + \delta)p^B + 1 - (p^B + \delta) - p^B \\ &= 2(p^B)^2 + 2(\delta - 1)p^B + 1 + \gamma - \delta\end{aligned}$$

따라서 일치율과 폴리그래프의 정확도간 관계는 아래와 같은 방정식으로 표시된다.

$$\rho = 2(p^B)^2 + 2(\delta - 1)p^B + 1 + \gamma - \delta \quad (2)$$

이 방정식에서 ρ 는 검찰처분과 폴리그래프 검사결과가 일치할 확률의 평균값이고, p^B 는 폴리그래프 검사결과가 옳을 확률의 평균값이며, δ 는 검찰처분이 옳을 확률의 평균값 p^A 에서 폴리그래프 검사결과가 옳을 확률의 평균값 p^B 를 뺀 값이다. γ 는 $\omega + 2\sigma_{AB}$ 로 정의되는 값으로, ω 는 실제로 두 판단이 일치할 확률과 두 판단이 서로 독립적일 때의 일치 확률(즉, 우연에 의한 일치확률) 간 차이값의 평균이며, σ_{AB} 는 검찰처분과 폴리그래프 검사결

과가 옳을 확률간 공변량을 나타낸다. 이 방정식은 한 가지 제약을 가지는데, 검찰처분과 폴리그래프 검사결과와의 일치율은 다음과 같은 하한값을 가진다.

$$\rho \geq \gamma + (1 - \delta^2)/2$$

폴리그래프의 정확도 p^B 를 계산하기 위하여 방정식 (2)를 풀 때에는 파악의 문제(identification problem)가 따른다(Spencer, 2007). 이 문제는 폴리그래프 검사결과와 검찰처분의 일치율이 두 판단이 모두 옳았을 경우와 더불어 두 판단이 모두 틀렸을 경우를 함께 반영하기 때문에 발생하는데, 일치율 자체만 가지고는 두 판단이 모두 옳은 경우가 두 판단이 모두 틀린 경우에 비하여 얼마나 많은지를 파악할 수 없다. 이러한 파악오류는 다음과 같은 수학적 형태로 나타난다.

방정식 (2)는 p^B 가 이차항으로 되어 있기 때문에 p^B 에 대하여 방정식을 정리하면,

$$p_{low}^B = .5(1 - \delta - \sqrt{2(\rho - \gamma) - 1 + \delta^2})$$

$$p_{high}^B = .5(1 - \delta + \sqrt{2(\rho - \gamma) - 1 + \delta^2})$$

으로 두 개의 근을 가진다. 두 개의 값 중 어떤 값이 p^B 와 일치하는 값인지를 알 수 없으므로 1을 초과하지 않는 값 중 더 큰 값을 선택하는 것이 일반적이다. 이 선택은 p^B 를 과대추정하게 되고, 파악오류(identification error)를 유발 하며, 0과 같거나 0보다 큰 오류를 가지게 된다.

폴리그래프 검사결과와의 정확도 추정치를 산출하기 위하여, p_{high}^B 를 산출하는 공식에 포함된 ρ 의 값으로 관찰된 일치율 $\hat{\rho}$ 을 사용할 수 있으며, p^B 가 1을 초과해서는 안 된다는 제약과 $\rho \geq \gamma + (1 - \delta^2)/2$ 의 제약을 이용할 수 있다. 결과적으로 정확도 추정치는 함수 (3)과

같이 표현될 수 있다.

$$\begin{aligned}
 &\text{if } 1+\delta < \hat{\rho}-\gamma, & (3) \\
 &\quad f(\hat{\rho}, \gamma, \delta) = 1 \\
 &\text{if } (1-\delta^2)/2 \leq \hat{\rho}-\gamma \leq 1+\delta, \\
 &\quad f(\hat{\rho}, \gamma, \delta) = 0.5(1-\delta + \sqrt{2(\hat{\rho}-\gamma)-1+\delta^2}) \\
 &\text{if } \hat{\rho}-\gamma \leq (1+\delta^2)/2, \\
 &\quad f(\hat{\rho}, \gamma, \delta) = 0.5(1-\delta)
 \end{aligned}$$

Spencer(2007)에 의하면 정확도 추정치인 p^B 는 세 가지 종류의 오류를 가질 수 있다. 첫 번째 오류는 앞서 언급한 파악오류로 $f(\rho, \gamma, \delta) - p^B$ 와 같다. 두 번째 오류는 두 판단의 모집단 일치율 ρ 와 추정된 일치율 $\hat{\rho}$ 을 다르게 만드는 모든 오류들(표본오차, 반응성 오차, 자료처리오차 등)을 포함하는 조사오류(survey error)로 $f(\hat{\rho}, \gamma, \delta) - f(\rho, \gamma, \delta)$ 와 같다. 세 번째는 모집단의 γ 와 δ 값을 대신하여 추정치($\hat{\gamma}$ 와 $\hat{\delta}$)을 사용하기 때문에 발생하는 오류로 명세오류(specification error)라고 하며, $f(\hat{\rho}, \hat{\gamma}, \hat{\delta}) - f(\hat{\rho}, \gamma, \delta)$ 와 같다.

잠재계층분석

잠재계층분석은 요인분석과 같이 관찰변인들 사이의 관계를 유발하는 잠재적인 공통요인을 규명하기 위하여 개발된 방법으로, 관찰변인과 잠재요인이 모두 범주변인일 때 사용할 수 있는 방법이며(박광배, 2006), 준거가 존재하지 않을 때 다수의 검사 결과들 간의 관련성을 분석하여 특정 검사의 정확도를 추정하는데 사용할 수 있다(Rindskopf & Rindskope, 1986; Faraone & Tsuang, 1994; Albert, McShane & Shih, 2001; Pepe & Janes, 2007). 즉, 잠재계

층분석을 통하여 폴리그래프 검사결과와 검찰처분간 관련성을 유발하는 잠재요인을 파악하고, 잠재요인을 준거로 하여 폴리그래프 검사의 정확도를 추정할 수 있다.

폴리그래프 검사를 A라고 하고, 검찰처분을 B라고 하자. A의 결과를 $i=1, 2, 3, \dots, I$ 로 표시하고, B의 결과를 $j=1, 2, 3, \dots, J$ 로 표시하자. 잠재적 범주요인을 U로 표시하고 이 요인의 계층들을 $t=1, 2, 3, \dots, T$ 로 표시하자. 그러면, 모집단 사례들이 잠재요인 U의 계층 t 에 속할 비조건확률은 p_t^U 로 정의되고, 잠재요인 U의 계층 t 에 속한 사례들이 폴리그래프 검사에서 i 로 결정될 조건확률은 p^{AU} 로, 잠재요인 U의 계층 t 에 속한 사례들이 폴리그래프 검사에서 i 로 결정되고 검찰처분에서 j 로 결정될 조건확률은 p_{ij}^{ABU} 로, 모집단의 전체 사례가 폴리그래프 검사에서는 i 로 결정되고, 검찰처분에서는 j 로 결정될 결합확률은 p_{ij}^{AB} 로, 모집단의 전체사례가 잠재요인 U의 계층 t 에 속하고 폴리그래프 검사에서는 i 로 결정되고 검찰처분에서는 j 로 결정될 결합확률은 p_{ijt}^{ABU} 로, 폴리그래프 검사에서는 i 로 결정되고 검찰처분에서는 j 로 결정된 사례들이 잠재요인 U의 계층 t 에 속할 사후확률은 p_{ijt}^{UAB} 로 정의된다.

어떤 사례가 특정한 잠재계층 t 에 속하고 폴리그래프 검사와 검찰처분의 결과가 ij 일 결합확률은 다음과 같이 조건확률과 외곽확률의 곱으로 표현될 수 있다.

$$p_{ijt}^{ABU} = p_{ijt}^{ABU} p_t^U$$

잠재계층분석은 각 잠재계층 내에서 검사결과들(즉, 폴리그래프 검사결과 검찰처분)이 서

로 독립적이 되도록 잠재계층을 형성한다. 즉, 전체 사례에 대해서는 검사 결과들 간에 상관성이 존재해야 하지만, 잠재계층분석을 통하여 파악한 각 잠재계층 내에서는 검사 결과들 간에 상관성이 없어야 한다. 이것을 지역독립성(local independence)이라고 하며, 폴리그래프 검사결과와 검찰처분 간 지역독립성이 잠재요인 U에 의해 성립된다면 다음의 수식이 성립한다.

$$p_{ijt}^{ABU} = p^{AU} p_{jt}^{BU}$$

위의 두 수식을 결합하면, 어떤 사례가 특정한 잠재계층 t에 속하고 폴리그래프 검사결과와 검찰처분이 ij일 결합확률은 다음과 같이 재정의 된다.

$$p_{ijt}^{ABU} = p_t^U p^{AU} p_{jt}^{BU}$$

특정한 사례가 폴리그래프 검사와 검찰처분에서 ij로 결정될 결합확률은 다음과 같이 p_{ijt}^{ABU} 를 모든 잠재계층 t에 걸쳐서 더하여 주면 된다.

$$p_{ij}^{AB} = \sum_{t=1}^T p_{ijt}^{ABU} = \sum_{t=1}^T p_t^U p^{AU} p_{jt}^{BU} \quad (4)$$

잠재계층분석은 표본자료에 계산된 두 검사 결과의 결합확률(p_{ij})을 방정식 (4)의 등호 왼쪽에 있는 p_{ijt}^{ABU} 에 대입한 후, 이 방정식을 최대한 충족하는 모수(p_t^U , p^{AU} , p_{jt}^{BU})의 값을 추정한다.

폴리그래프 검사결과는 0(진실을 말함), 1(거짓을 말함), 2(판단불능)의 값을 가질 수

있고, 검찰처분은 0(기소), 1(불기소), 2(기소 중지)의 값을 가질 수 있으며, 잠재계층 U는 0(진실을 말하는 집단)과 1(거짓을 말하는 집단)을 가질 수 있으므로, 검찰처분이 옳을 확률 $p^A = p_{00}^{AU} p_0^A + p_{11}^{AU} p_1^A$ 로 산출될 수 있으며, 폴리그래프 검사결과가 옳을 확률 $p^B = p_{00}^{BU} p_0^B + p_{11}^{BU} p_1^B$ 로 산출될 수 있다.

관찰된 결합확률을 이용하여 모수를 추정하기 위해서는 서로 독립적인 관찰된 결합확률의 개수(결합확률의 개수 - 1)가 추정하여 할 독립적인 모수의 개수보다 적어야 하며, 이러한 상태를 과파악(over-identification)이라고 한다. 만약, 추정하여야 할 독립적인 모수의 개수가 서로 독립적인 관찰된 결합확률의 개수보다 많으면 미파악(under-identification)되고, 추정할 독립적 모수의 개수가 서로 독립적인 관찰된 결합확률의 개수와 같으면 정파악(just-identification)되어 모형검증의 의미가 사라진다.

잠재계층분석을 하기 위해서는 모형이 과파악되어야 하므로, 정파악되거나 미파악되는 모형은 추정할 모수의 수를 줄여야 한다. 추정할 모수의 수를 줄이는 방법은 모수 추정치들에 대하여 연구자가 인위적인 제약(constraint)을 가하는 것이다. 모수제약에는 두 가지 방법이 있는데, 한 가지는 모수추정치 값의 연구자가 지정하는 것이고, 다른 한 가지는 특정한 모수들의 값이 서로 동일하다는 조건을 부여하는 것이다.

연구목적

한국에서 폴리그래프 검사의 정확도를 추정하기 위한 실험실 연구는 최근에 활발히 진행되고 있는 반면(조은경, 2006; 지형기, 정재영,

강민국, 김재홍, 김미영, 정규희, 이장한, 2008), 폴리그래프 검사의 정확도를 추정하기 위해 수행된 현장연구는 찾아볼 수 없다. 본 연구에서는 한국에서 처음으로 현장자료를 이용하여 폴리그래프 검사의 정확도를 추정하고자 하며, 기존 현장연구들에서 폴리그래프 검사의 정확도를 추정하는데 사용한 방법과는 다르게 통계적인 방법을 활용하고자 한다. 이를 위하여 검찰에서 수행된 폴리그래프 검사결과와 검찰처분간 일치도를 이용하여 폴리그래프 검사 결과의 정확도를 추정하고자 한다.

방 법

분석자료

두 개의 현장자료 세트를 폴리그래프 검사의 정확도 추정에 이용하였다. 한 세트의 자료(이하 '2000-2004 자료'라고 부르기로 함)는 2000년부터 2004년까지 대검찰청과 12개 지방검찰청에서 실시된 4472 건의 형사사건에 대한 폴리그래프 검사 중 검사자의 판정(진실 또는 거짓)과 검찰처분(혐의 없음 또는 혐의 있음)이 알려져 있는 2923건의 사례로 구성되

어 있다. 이 사례들은 흥부호흡, 복부호흡, 혈압 및 맥박, 피부전도수준을 측정하는 폴리그래프 장비를 이용하여 검사되었으며, 폴리그래프 검사는 거짓말탐지기 운영규정(대검찰청 예규 제284호)에 근거한 자격을 갖춘 검사관에 의해 실시되었다. 폴리그래프 검사기법은 형사사건에 가장 많이 이용하는 비교질문기법(comparison question technique)인 Beckster zone comparison test, Utah probable-lie test, 및 modified general question test 등을 사용하였다. 다른 한 세트의 자료(이하 '2006 자료'라고 부르기로 함)는 2006년에 대검찰청과 4개의 지방검찰청에서 형사사건에 대해 실시된 폴리그래프 검사 중 검사자의 판정(진실, 거짓 또는 판단불가)과 검찰처분(불기소, 기소 또는 기소중지)이 알려져 있는 208건의 사례로 구성되어 있다. 이 사례들도 흥부호흡, 복부호흡, 혈압맥박, 피부전도수준을 측정하는 폴리그래프 장비를 이용하여, 자격을 갖춘 폴리그래프 검사관에 의해 검사되었다. 폴리그래프 검사기법은 모두 Beckster zone comparison test를 사용하였다.

2000-2004 자료와 2006 자료를 구분한 이유는 두 자료 세트가 제공하는 정보가 다르기 때문이다. 표 1에 제시되어 있는 것처럼, 2000

표 1. 폴리그래프 검사 결과와 검찰처분의 교차빈도 분포

	검찰처분	폴리그래프 검사 결과		
		진실	거짓	판단불능
2000-2004 자료	불기소	1019 (34.9)	297 (10.2)	-
	기소	166 (5.7)	1441 (49.3)	-
2006 자료	불기소	54 (26.0)	19 (9.1)	24 (11.5)
	기소	12 (5.8)	68 (32.7)	18 (8.7)
	기소중지	3 (1.4)	8 (3.8)	2 (1.0)

-2004 자료는 폴리그래프 검사 결과가 ‘판단불능’인 사례들과 검찰처분이 ‘기소중지’인 사례들이 포함되지 않은 반면, 2006 자료는 이 자료들이 포함되어 있다. 따라서, 2006 자료는 판단불능 범주를 포함하였을 때의 폴리그래프 검사의 정확도를 산출해 볼 수 있을 것이다.

분석방법

일치율 이용한 정확도 추정

2000-2004 자료와 2006 자료에서 각각 산출한 검찰처분과 폴리그래프 검사 결과의 일치율을 함수 (3)에 대입하여, 폴리그래프 검사의 정확도를 추정하였다. 함수 (3)에서 검찰처분이 옳을 확률과 폴리그래프 검사결과가 옳을 확률의 차이값인 δ 와 실제 두 판단이 일치할 확률과 두 판단이 독립적일 때 일치할 확률간 차이값인 γ 를 정할 수 있는 방법이 없으므로, 일단 0으로 고정시키고 폴리그래프의 정확도를 추정하였다.

폴리그래프 정확도 추정치 p^B 는 세 가지 종류의 오류를 가질 수 있다고 하였는데, 첫 번째 오류는 파악오류로 $f(\rho, \gamma, \delta) - p^B$ 와 같다. 두 번째 오류는 조사오류로 $f(\hat{\rho}, \gamma, \delta) - f(\rho, \gamma, \delta)$ 와 같고, 세 번째는 명세오류로 $f(\hat{\rho}, 0, 0) - f(\hat{\rho}, \gamma, \delta)$ 와 같다.

명세오류는 잘못된 γ 와 δ 를 사용하기 때문에 발생하는데, 실제로 $\gamma = \omega + 2\sigma_{AB}$ 에서 $\omega \geq 0$ 일 가능성이 높다. 검찰의 처분은 폴리그래프 검사결과가 나온 이후에 내려지기 때문에, 검찰처분에 폴리그래프의 검사결과가 활용되었을 가능성이 높다. 따라서 두 판단이 독립적으로 내려졌을 때의 일치율 보다 현실에서 더 높은 일치율을 보일 가능성이 높다.

용의주도한 피검사자는 폴리그래프 검사에서 거짓이 탄로나지 않을 가능성이 높으며, 유죄의 증거도 남기지 않을 가능성이 높으므로 폴리그래프 검사결과와 검찰처분이 옳을 확률간 공변량인 σ_{AB} 도 0 보다 클 가능성이 높다. 수학적으로 γ 가 증가함에 따라 $f(\hat{\rho}, \gamma, \delta)$ 의 값은 감소하기 때문에, γ 의 추정치로 0을 사용하면 p^B 를 과대추정하게 된다. δ 도 0보다 클 가능성이 있는데, 검찰은 폴리그래프 검사자료와 다른 모든 증거들을 통합하여 기소여부를 판단하므로 폴리그래프 검사결과보다 검찰의 판단이 더 정확할 가능성이 있기 때문이다. δ 가 증가함에 따라서 $f(\hat{\rho}, \gamma, \delta)$ 의 값은 감소하기 때문에, δ 의 추정치로 0을 사용하면 p^B 를 과대추정하게 된다. 결국 $\gamma \geq 0$ 고 $\delta \geq 0$ 이므로, 명세오류는 0과 같거나 0보다 큰 값을 가지게 된다. 관찰된 일치율 $\hat{\rho}$ 의 기대값이 실제 일치율 ρ 라고 가정한다면, 파악오류와 명세오류는 p^B 를 다소 과대추정하도록 만들 수 있다.

잠재계층분석을 이용한 정확도 추정

판단불능 범주를 포함한 폴리그래프 검사의 정확도를 추정하기 위하여 Clogg(1977)이 만든 MLLSA 프로그램을 이용하여 잠재계층분석을 실시하였다. 잠재계층분석을 이용하여 모수를 추정하기 위해서는 관찰변인에 많은 양의 정보가 있는 것이 유용하다. 즉 관찰변인의 수가 많거나 관찰변인의 범주수가 많은 것이 유용하므로, 폴리그래프 검사결과가 판단불능인 사례와 검찰처분이 기소중지인 사례를 포함하고 있는 2006 자료를 이용하여 잠재계층분석을 통한 정확도를 추정하였다. 형사사건에서 폴리그래프의 피검사자들은 거짓을 말하거나 진실을 말하거나 둘 중 하나이므로, 2계층 모

형에 대하여 분석을 하였으며, 폴리그래프 검사결과(거짓, 진실, 판단불능)와 검찰처분(기소, 불기소, 기소중지)를 관찰변인으로 사용하였다.

3개씩의 범주를 가지는 두 개의 관찰변인으로 2계층 모형의 모수를 추정하기 위해서는 두 개 이상의 제약을 부과해야 모형 파악이 가능하기 때문에, 기존 연구로부터 4개의 동일성 제약 가능성을 고려하였다. 첫째로, 피검사자가 거짓 잠재계층에 속할 때 폴리그래프 검사결과가 판단불능일 조건확률과 피검사자가 진실 잠재계층에 속할 때 폴리그래프 검사결과가 판단불능일 조건확률이 동일하다는 제약을 가할 수 있다. 둘째로, 잠재계층이 진실일 때 폴리그래프 검사결과가 거짓일 조건확률과 잠재계층이 진실일 때 검찰처분이 기소일 조건확률이 동일하다는 제약을 할 수 있다. 셋째로, 잠재계층이 진실일 때 검찰처분이 기소중지일 조건확률과 잠재계층이 거짓일 때 검찰처분이 기소중지일 조건확률을 동일하게 제약할 수 있다. 넷째로, 잠재계층이 거짓일 때 폴리그래프 검사결과가 진실일 조건확률과 잠재계층이 진실일 때 검찰처분이 기소일 조건확률이 동일하다는 제약을 할 수 있다.

첫 번째 가정은 미국 의회의 기술평가국(1983)이 다수의 현장연구를 메타분석한 결과에 근거한 것으로(p.52, 표 4), 이 연구에서 진실조건의 판단불능 확률과 거짓조건의 판단불능 확률이 비슷한 것으로 나타난 것에 근거한 것이며, 두 번째 가정은 앞의 문헌에서 폴리그래프 검사결과가 오류공정을 보일 확률이 희박하다는 것과 한국의 형사재판에서 무죄평결이 드물다는 일반적인 경향성에 근거한 것이다(대검찰청, 2008). 세 번째 가정은 한국 검찰에 송치된 사건중 기소중지의 비율이 약 7%로 비교적 드물기 때문에 진실 잠재계층과

거짓 잠재계층에서 기소중지의 조건확률이 크게 다르지 않을 것이므로 두 가지의 조건확률이 동일하다고 제약할 수 있다(대검찰청, 2008). 네 번째 가정은 미국 의회의 기술평가국(1983)에서 수행한 메타분석결과와 Raskin과 Honts(2002)의 현장연구에 대한 비교질문검사의 타당도분석에 인용된 결과(p.32, 표 1.9)에서 피검사자가 거짓 잠재계층일 때 폴리그래프 검사 결과가 진실일 조건확률(1~10%)과 피검사자가 진실 잠재계층일 때 검찰처분이 기소일 조건확률이 유사한 것에 근거한 것이다(대검찰청, 2008).

정확도 추정치의 표준오차

폴리그래프 검사의 정확도 추정치가 표본에 따라서 어느 정도나 달라질 수 있는지를 파악하기 위하여 정확도 추정치의 표준오차(standard error)를 산출하였다. 표준오차는 표본에서 산출한 통계치(statistics)로 모집단의 모수치(parameters)를 추정할 때 발생하는 표준적인 오차의 양을 의미하며, 통계치의 표집분포가 정규분포한다면 모집단의 모수치가 표본통계치 - 1.96×표준오차 ~ 표본통계치 + 1.96×표준오차 내에 포함되어 있을 확률이 95%가 된다.

무선표본으로부터 산출된 통계치의 표준오차를 추정하는 유용한 방법 중 한 가지는 부트스트랩(bootstrap) 표집분포를 이용하는 것이다(Wasserman & Bockenholt, 1989; Efron & Tibshirani, 1993). 사례수가 N인 표본이 있을 때 이 표본으로부터 몇 개의 사례들을 다시 무선적으로 추출할 수 있다. 이 때, 원래의 표본으로부터 사례들을 추출할 때마다 그 사례의 값을 기록한 후 다시 원래의 표본에 넣는다면(즉, 복원추출을 의미함), 원래의 표본으로

부터 무수히 많은 값들을 재추출할 수 있다. 사례수가 N 인 원래의 표본으로부터 N 번의 복원추출과정을 거치면 N 개의 값을 얻을 수 있다. 원래의 표본에서 사례를 복원추출할 때 어떤 사례들은 여러 번 뽑힐 수도 있고, 어떤 사례들은 한 번도 안 뽑힐 수 있기 때문에, 이 N 개의 값들은 원래의 표본이 가진 N 개의 값들과 약간의 차이를 가지게 되는데, 이 값들을 부트스트랩 표본이라고 한다.

이와 같은 부트스트랩 표본을 1000개 뽑은 후(2006 자료의 예를 들면, 208개의 사례를 가지는 표본에서 208개의 값들을 복원추출하는 과정을 1000번 수행한 후), 1000개의 부트스트랩 표본 각각에서 폴리그래프 검사의 정확도 추정치를 산출하고, 1000개의 정확도 추정치로 부트스트랩 표집분포를 만들 수 있다. 이 분포의 표준편차가 폴리그래프 검사의 정확도 추정치의 표준오차가 된다.

결 과

폴리그래프 검사결과와 검찰처분의 일치율

2000-2004 자료에서 폴리그래프 검사결과와 검찰처분의 일치율은 .842였으며, 2006 자료에서 일치율은 .797(판단불능 사례와 기소중지 사례를 제외하고 산출함)로 두 자료의 일치율 차이는 .045로 큰 차이가 없었다. 2000-2004 자료에서 우연에 의한 일치율은 .510 ($.450 \times .405 + .550 \times .595$)이고 2006 자료에서 우연에 의한 일치율은 .503 ($.477 \times .431 + .523 \times .569$)으로, 폴리그래프 검사결과와 검찰처분의 일치율이 우연에 의한 일치율보다 약 .3 이상 높았다. 우연에 의한 일치율을 고려한 일치도

지수인 Cohen's Kappa는 .677과 .592로 적절한 수준(moderate to good)인 것으로 해석할 수 있다(Altman, 1991).

일치율을 이용한 폴리그래프 검사의 정확도 추정치

2000-2004 자료에 계산한 폴리그래프 검사와 검찰처분간 일치율인 .842를 함수 (3)에 대입하여 산출한 폴리그래프 검사의 정확도 추정치는 .914였으며, 2006 자료의 일치율인 .797을 적용한 정확도 추정치는 .885였다. 그러나, 이 추정치는 파악오류와 명세오류로 인하여 다소 과대추정되었을 수 있다는 사실을 인식하고 있어야 한다. 폴리그래프 검사가 검찰처분보다 더 정확하다면, 즉 δ 가 음수라면 폴리그래프 검사의 정확도 추정치는 더 높아지고, 검찰처분이 폴리그래프 검사보다 더 정확하다면, 즉 δ 가 양수라면 폴리그래프 검사의 정확도 추정치는 더 낮아진다. 폴리그래프 검사결과와 검찰처분이 서로 독립적으로 판단을 내릴 때의 일치율보다 실제의 일치율이 더 높다면, 즉 γ 가 양수라면 폴리그래프 검사의 정확도 추정치는 더 낮아진다. 표 2에 다양한 δ 와 γ 를 함수 (3)에 넣어 산출한 폴리그래프 검사의 정확도 추정치가 제시되어 있다.

정확도 추정치의 표준오차

일치율을 이용하여 산출한 폴리그래프 검사의 정확도 추정치가 표본에 따라서 어느 정도나 달라질 수 있는지를 파악하기 위하여 정확도 추정치의 표준오차를 산출하였다. 부트스트랩 표집분포를 이용하여 표준오차를 산출하였으며, 1000개의 부트스트랩 표본을 추출하여 표집분포를 만들었다.

표 2. 여러 가지 δ 값과 γ 값을 사용한 폴리그래프 검사의 정확도 추정치

δ ($\gamma=0$)	2000-2004 자료 $\hat{\rho}=.842$	2006 자료 $\hat{\rho}=.797$	γ ($\delta=0$)	2000-2004 자료 $\hat{\rho}=.842$	2006 자료 $\hat{\rho}=.797$
-0.25	1.000	1.000	-0.25	1.000	1.000
-0.20	1.000	0.998	-0.20	1.000	0.998
-0.15	0.995	0.968	-0.15	0.996	0.973
-0.10	0.967	0.939	-0.10	0.97	0.946
-0.05	0.939	0.911	-0.05	0.943	0.917
0.00	0.914	0.885	0.00	0.914	0.885
0.05	0.889	0.861	0.05	0.882	0.851
0.10	0.867	0.839	0.10	0.848	0.814
0.15	0.845	0.818	0.15	0.810	0.771
0.20	0.825	0.798	0.20	0.766	0.720
0.25	0.807	0.780	0.25	0.714	0.653

부트스트랩 표집분포를 이용하여 산출한 2000-2004 자료의 표준오차 추정치는 .004였으며, 2006 자료의 표준오차 추정치는 .021였다.

잠재계층분석을 이용한 폴리그래프 검사의 정확도 추정

분석방법에서 언급한 네 가지 제약들을 두 가지씩 조합하여 전체 6개의 잠재계층모형을 만든 후, 표 3과 같이 6개 모형 각각에 대하여 잠재계층확률과 조건확률들을 산출하였으며, 부트스트랩 표집분포를 이용하여 각 확률들의 표준오차를 추정하였다.

잠재계층분석결과 6개의 모형은 모두 자료와 잘 합치하는 것으로 나타났다($L^2 = .153 \sim 2.753$, $df=1$, $p>.05$). 그러나 6가지 모형들 중에서 제약 2를 포함한 모형들의 표준오차 추정치들이 다른 모형들에 비해서 큰 것으로 나

타나, 제약 2를 포함하지 않는 모형 2, 3, 6이 좋은 것으로 판단하였다. 따라서 이 모형들을 중심으로 결과를 해석하였다. 폴리그래프 검사결과가 옳을 확률의 추정치(\hat{p}^B)는 .707~.734의 범위를 보였으며 추정된 표준오차(SE)는 .027~.031였다. 검찰처분이 옳을 확률의 추정치(\hat{p}^A)는 .790~.818 (SE=.027~.032)로 폴리그래프 검사결과보다 약 .08 높은 것으로 나타났다. 피검사자가 무죄일 때 폴리그래프 검사결과가 거짓으로 산출될 확률(\hat{p}_{10}^{BU} , 오류긍정의 확률)은 .078~.087(SE=.019~.023)으로 산출되었으며, 피검사자가 거짓을 말하는 사람일 때 폴리그래프 검사결과가 진실로 산출될 확률(\hat{p}_{01}^{BU} , 오류부정의 확률)은 .029~.078(SE=.010~.023)으로 산출되어 오류긍정확률이 오류부정확률보다 근소하게 높은 것으로 추정되었다. 폴리그래프 검사결과가 거짓일 때 피검사자가 거짓을 말하는 사람일 사후확률(\hat{p}_{11}^{UB})은

표 3. 다양한 제약을 부과한 잠재계층모형의 확률추정치

	모형 1 제약 1 ¹⁾ +2 ²⁾	모형 2 제약 1+3 ³⁾	모형 3 제약 1+4 ⁴⁾	모형 4 제약 2+3	모형 5 제약 2+4	모형 6 제약 3+4
\hat{p}_{00}^U	.335 (.080)5)	.463 (.028)	.424 (.031)	.573 (.078)	.435 (.050)	.437 (.033)
\hat{p}_{01}^U	.665 (.080)	.537 (.028)	.576 (.031)	.427 (.078)	.565 (.050)	.563 (.033)
\hat{p}_{00}^{AU}	.971 (.059)	.808 (.040)	.890 (.032)	.765 (.059)	.886 (.033)	.859 (.027)
\hat{p}_{01}^{AU}	.213 (.066)	.163 (.039)	.153 (.035)	.051 (.068)	.143 (.049)	.152 (.036)
\hat{p}_{10}^{AU}	.002 (.055)	.130 (.037)	.077 (.022)	.172 (.057)	.081 (.023)	.078 (.023)
\hat{p}_{11}^{AU}	.707 (.063)	.775 (.041)	.762 (.042)	.887 (.070)	.772 (.053)	.786 (.038)
\hat{p}_{20}^{AU}	.027 (.026)	.063 (.016)	.033 (.025)	.063 (.016)	.033 (.026)	.063 (.017)
\hat{p}_{21}^{AU}	.080 (.027)	.063 (.016)	.085 (.026)	.063 (.016)	.085 (.029)	.063 (.017)
\hat{p}_{00}^{BU}	.786 (.061)	.704 (.031)	.701 (.034)	.578 (.069)	.658 (.059)	.659 (.051)
\hat{p}_{01}^{BU}	.117 (.048)	.029 (.010)	.077 (.022)	.002 (.039)	.081 (.023)	.078 (.023)
\hat{p}_{10}^{BU}	.002 (.055)	.085 (.019)	.087 (.023)	.172 (.057)	.081 (.023)	.078 (.022)
\hat{p}_{11}^{BU}	.671 (.055)	.760 (.027)	.711 (.035)	.839 (.064)	.746 (.043)	.750 (.043)
\hat{p}_{20}^{BU}	.212 (.029)	.212 (.028)	.212 (.028)	.250 (.049)	.261 (.053)	.263 (.055)
\hat{p}_{21}^{BU}	.212 (.029)	.212 (.028)	.212 (.028)	.159 (.051)	.173 (.042)	.172 (.042)
\hat{p}^A	.795 (.033)	.790 (.032)	.817 (.027)	.817 (.032)	.822 (.031)	.818 (.028)
\hat{p}^B	.710 (.035)	.734 (.027)	.707 (.031)	.689 (.039)	.708 (.035)	.710 (.030)
\hat{p}_{11}^{UA}	.998 (.063)	.874 (.035)	.931 (.021)	.793 (.076)	.926 (.028)	.928 (.023)
\hat{p}_{01}^{UA}	.002 (.063)	.126 (.035)	.069 (.021)	.207 (.076)	.074 (.028)	.072 (.023)
\hat{p}_{10}^{UA}	.303 (.103)	.190 (.039)	.189 (.038)	.047 (.088)	.173 (.066)	.185 (.039)
\hat{p}_{00}^{UA}	.697 (.103)	.810 (.039)	.811 (.038)	.953 (.088)	.827 (.066)	.815 (.039)
\hat{p}_{11}^{UB}	.998 (.067)	.912 (.013)	.917 (.015)	.784 (.081)	.923 (.031)	.925 (.016)
\hat{p}_{01}^{UB}	.002 (.067)	.088 (.013)	.083 (.015)	.216 (.081)	.077 (.031)	.075 (.016)
\hat{p}_{10}^{UB}	.229 (.096)	.045 (.011)	.130 (.037)	.002 (.071)	.137 (.038)	.133 (.040)
\hat{p}_{00}^{UB}	.771 (.096)	.955 (.011)	.870 (.037)	.998 (.071)	.863 (.038)	.867 (.040)
L^2	1.472	2.753	1.478	1.412	.153	1.409
df	1	1	1	1	1	1
% correctly allocated	93.96	90.89	93.90	92.50	90.50	90.04

1) 제약 1: $p_{20}^{BU} = p_{21}^{BU}$

2) 제약 2: $p_{10}^{AU} = p_{10}^{BU}$

3) 제약 3: $p_{20}^{AU} = p_{21}^{AU}$

4) 제약 4: $p_{01}^{BU} = p_{10}^{AU}$

5) 괄호안의 수치는 표준오차임

.912~.915(SE=.013~.016)로 상당히 높았으며, 폴리그래프 검사결과가 진실일 때 피검사자가 진실을 말하는 사람일 사후확률(\hat{p}_{00}^{UB})은 .867~.955 (SE=.011~.040)로 추정되었다.

논 의

본 연구에서는 한국에서 처음으로 현장자료를 이용하여 폴리그래프 검사의 정확도를 추정하였다. 기존에 국외에서 폴리그래프 검사의 정확도를 추정하기 위하여 수행된 현장연구에서는 실제적 진실을 결정하기 위하여 피검사자의 자백을 이용하거나 전문가 패널의 결정을 이용하였지만, 피검사자의 자백이나 전문가 패널의 결정이 부정확한 것이 현실이다. 따라서 본 연구에서는 한국 검찰청에서 수행한 폴리그래프 검사의 실제적 진실로도 사용될 수 있는 검찰의 처분을 폴리그래프 검사와 동일한 수준의 또 다른 평가로 간주하여, 폴리그래프 검사결과와 검찰의 처분 간 일치도와 교차빈도분포를 이용하여 통계적인 방법으로 폴리그래프 검사의 정확도를 추정하였다. 기존 현장연구들과 본 연구 사이의 가장 큰 차이점은 기존 연구들은 실제적 진실이 가지는 부정확성을 정확도를 추정하기 위한 모형에 포함시키지 않은 반면, 본 연구에서는 검찰처분이 실제적 진실을 추정하는데 부정확성을 가지고 있다는 것을 정확도 추정 모형에 명확히 하였다는 점이다. 이로 인하여, 기존 연구들의 정확도 추정치에는 부정확한 실제적 진실을 준거로 사용하기 때문에 발생하는 오류가 포함된 반면, 본 연구의 정확도 추정치에는 이러한 오류가 포함되지 않는다는 장점을 가진다.

폴리그래프 검사결과와 검찰처분간 일치율을 이용하여 추정한 폴리그래프의 정확도는 2000-2004 자료에서는 .914(SE=.004)였으며, 2006 자료에서는 .885(SE=.021)로 추정되었다. 따라서, 판단불능의 사례를 포함하지 않은 상태에서 폴리그래프 검사결과와 검찰처분간 일치율을 이용하여 추정한 폴리그래프의 정확도는 약 90%인 것으로 생각할 수 있다. 그러나 이 결과는 파악오류와 명세오류로 인하여 다소 과대 추정되었을 가능성이 있다.

판단불능 사례들을 포함한 후 잠재계층분석을 이용하여 추정한 폴리그래프 검사의 정확도는 .707~.734(SE=.027~.031) 범위인 것으로 나타났다. 이 값은 다소 과소추정되었을 가능성이 높다. 그 이유는 2000년에서 2004년까지 대검찰청과 지방검찰청에서 실시된 모든 폴리그래프 검사결과에서 ‘판단불능’인 사례의 비율은 15.2%였던 반면, 208개의 표본으로 구성된 2006 자료에서 폴리그래프 검사결과가 ‘판단불능’인 사례의 비율은 21.2%로 6%가 더 높았기 때문이다. 2006 자료에서 판단불능의 비율을 15% 정도로 낮추면, 폴리그래프의 정확도는 약 .055 정도가 높아져 .762~.789 정도일 것으로 추정된다. 만약 잠재계층분석의 결과에서 판단불능의 비율을 0으로 설정하고 정확도를 추정하면 약 .914로, 폴리그래프 검사결과와 검찰처분의 일치도를 이용하여 추정한 정확도 .885~.914와 비슷하여, 일치도를 추정하는 두 가지 방법이 일관성 있는 결과를 산출한 것으로 나타났다.

Ben-Shakhar와 Furedy(1990)는 자백을 준거로 사용하고 비교질문기법을 이용한 9개의 현장연구를 개관하였는데, 유죄의 84%, 무죄의 72%가 옳게 분류되었다고 보고하였으며, 비교질문기법을 사용한 4개의 현장연구를 개관한

Raskin과 Honts(2002)는 유죄의 89%, 무죄의 59%가 옳게 분류되었다고 하였다. 이 연구들은 모두 판정불능 범주를 포함하고 있었으며, 두 연구의 평균정확도는 76%로 본 연구에서 잠재계층분석을 이용하여 추정된 정확도 .762~.789와 유사하였다.

Ben-Shakhar 등(1990)의 연구에서 오류부정율은 13%, 오류긍정율은 23%였으며, Raskin 등(2002)의 연구에서는 오류부정율 1%, 오류긍정율 12%로 오류부정율이 오류긍정율보다 작은 것으로 나타났다. 본 연구에서도 오류부정율이 약 5%, 오류긍정율이 약 8%로 산출되어 기존연구들처럼 오류부정율이 오류긍정율보다 낮은 경향은 있었지만 그 차이는 크지 않은 것으로 나타났다.

최근에 국내에서 조은경(2006)이 비교질문기법으로 수행한 실험실 연구에서는 폴리그래프의 정확도가 판정불능 범주를 포함하였을 경우에는 71.4%로 산출되었고, 판정불능 범주를 제외하였을 경우에는 82.3%로 산출되어, 본 연구에서 현장자료를 이용하여 추정된 정확도가 5 ~ 8% 더 높았다. 52개의 실험실연구와 7개의 현장연구를 개관하여 정확도를 추정한 National Research Council(2003)의 연구에서도 현장 연구의 정확도 추정치($a=.89$)가 실험실 연구의 정확도 추정치($a=.86$)보다 약간 높은 것으로 나타난 것과 맥을 같이하고 있다. 특히, 조은경(2006)의 연구에 참여한 두 명의 폴리그래프 검사관은 본 연구에서 사용한 자료를 산출한 검사관들 중 일부이므로, 직접적인 비교 해석이 어느 정도 가능할 것이다.

폴리그래프 검사의 정확도가 약 90% 정도로 추정되므로 폴리그래프 검사의 정확도를 높이기 위한 연구가 필요하며, 폴리그래프 검사와 더불어 최근 한국에서 연구가 시작되고

있는 비언어적 행동분석 및 진술분석을 이용한 거짓말 탐지기법(김시업, 전우병, 김경하, 김미영, 전충현, 2005; 전우병, 김시업, 2005; 김시업, 전우병, 전충현, 2006; 노진아, 현명호, 2008)과 뇌파를 이용한 거짓말 탐지기법(함지선, 이장한, 2007; 김혁, 박판규, 이강희, 김현택, 2008)을 함께 사용하여 폴리그래프 검사의 단점을 보완할 필요성이 있을 것이다.

본 연구는 폴리그래프 검사의 타당도를 검증하기 위한 현장연구에서 실제적 진실(즉, 증거)을 얻기가 매우 어렵다는 일반적 상황에서, 검찰에서 수행한 폴리그래프 검사결과와 검찰처분을 이용하여 통계적인 방법으로 폴리그래프 검사의 정확도를 추정할 유일한 연구라는 큰 의미를 가진다. 그러나 이와 같은 방법을 이용하여 폴리그래프 검사의 정확도를 추정하기 위해서는 폴리그래프 검사와 검찰의 결정이 독립적으로 수행되었다는 가정이 필요한데, 검찰처분에 폴리그래프 검사결과가 활용되었을 가능성이 높으므로 이 가정이 충족되지 못하였을 가능성이 높다는 비판을 면하기 어려울 수 있다.

그러나 검찰처분에 폴리그래프 검사결과가 영향을 미쳤을 가능성은 있지만, 그 영향의 정도가 그리 크지는 않을 것으로 판단된다. 우선, 검찰에서 검사가 피의자의 진범여부가 불확실하여 그것을 파악하기 위해 폴리그래프 검사를 실시하는 경우는 거의 없다. 검사가 이미 피의자의 진범여부에 관한 판단을 하고 있는 상황에서 피의자의 자백을 유도하거나 자신의 판단이 옳다는 것을 보다 확실히 노출하기 위한 목적으로 실시하는 경우가 대부분이므로, 검사의 처분(혹은 판단)은 폴리그래프 검사를 실시하기 전에 이미 예비적으로 결정되는 경우가 많다. 따라서 실무감각에 의하면

검사의 처분이 폴리그래프 검사결과에 따라 좌우될 것으로 단정하기 어렵다. 둘째로, 결과들 간의 의존성(dependence)이 존재할 경우, 지역독립성을 가정하는 잠재계층모형의 적합도는 좋지 않게 나올 것으로 예상할 수 있는데 (Albert & Dodd, 2004), 본 연구에서 지역독립성을 가정한 잠재계층모형의 적합도는 매우 우수한 것으로 나타났다. 따라서 폴리그래프 검사결과에 대한 검찰처분의 의존성이 심각하지 않거나, 의존성이 존재하더라도 모형의 모수추정에 큰 영향을 미칠 정도는 아니라고 판단된다.

폴리그래프 검사결과와 검찰처분 외에 제3의 결과를 이용할 수 있다면, 폴리그래프 검사결과와 검찰처분 간의 의존성을 잠재계층모형에 포함시켜 편중되지 않은(unbiased) 결과를 얻을 수 있다(Albert 등 2004). 최근 대검찰청에서 범죄피의자를 대상으로 비언어적 행동분석 및 진술분석을 이용한 거짓말 탐지와 뇌파를 이용한 거짓말 탐지가 시도되고 있으므로, 폴리그래프 검사와 행동 및 진술분석, 뇌파를 이용한 검사결과들이 누적되면, 폴리그래프 검사의 정확도를 보다 정확하게 추정할 수 있을 것으로 기대한다.

본 연구는 표본과 관련된 몇 가지 한계점을 가진다. 첫째로, 본 연구에 사용한 표본은 검찰에서 수행한 폴리그래프 검사만 포함하므로, 본 연구결과를 경찰, 군수사기관, 정보기관 등에서 수행한 폴리그래프 검사의 정확도로 일반화하기가 어려울 수 있다. 둘째로, 2000-2004 자료는 대규모 자료로서 표본오차의 문제가 크게 발생하지 않지만, 2006 자료는 표본오차의 문제를 가질 수 있다. 특히, 2006 자료의 판단불능 비율이 2000 ~ 2004년의 판단불능 비율보다 높았던 점이 그렇다.

참고문헌

- 김 혁, 박판규, 이강희, 김현택 (2008). 관련자극과 무관련자극에서 사건관련전위 비교분석연구, 2008 한국심리학회 연차학술대회 논문집, 122-123.
- 김시업, 전우병, 김경하, 김미영, 전충현 (2005). 용의자의 거짓말 탐지를 위한 비언어적 단서탐색, 한국심리학회지: 사회 및 성격, 19(1), 151-161.
- 김시업, 전우병, 전충현 (2006). 군 수사현장에서 용의자의 비언어적 행동을 이용한 거짓말 탐지, 한국심리학회지: 사회문제, 12(2), 101-114.
- 노진아, 현명호 (2008). 거짓말의 음성 및 발화 행동 특징 연구, 한국심리학회지: 일반, 27(1), 119-137.
- 대검찰청 (2008). 2007년 종합심사분석. 서울:대검찰청.
- 박광배 (2002). 법심리학. 서울:학지사
- 박광배 (2006). 범주변인분석. 서울:학지사
- 전우병, 김시업 (2005). 이해득실 상황에 따른 거짓말 탐지에 대한 주관적 지표 - 대학생, 교도관, 재소자들을 대상으로 -, 한국심리학회지: 사회문제, 11(4), 1-22.
- 정도연 (2006). 거짓말 탐지기 검사의 신뢰성 확보와 적절한 운영 방안. 경북대학교. 석사학위논문.
- 조은경 (2006). 폴리그래프 검사의 타당성 및 행동분석과의 상관성 연구. 2006 대검찰청 정책연구용역과제 최종보고서.
- 지형기, 정재영, 강민국, 김재홍, 김미영, 정규희, 이장한 (2008). 폴리그래프 검사의 타당성에 관한 연구, 제4회 심리생리검사관세미나자료집, 57-81.

- 함지선, 이장한 (2007). 가상의 범죄환경과 뇌파를 이용한 거짓말 탐지 연구, 2007 한국심리학회 연차학술대회 논문집, 524-525.
- Abrams, S. (1973). Polygraph Validity and Reliability-a Review. *Journal of Forensic Sciences*, 18, 313-326.
- Albert, P. S., McShane, L. M., & Shih, J. H. (2001). Latent class modeling approaches for assessing diagnostic error without a gold standard: With applications to p.53 immuno- histochemical assays in bladder tumors. *Biometrics*, 57, 610-619.
- Albert, P. S., & Dodd, L. E. (2004). A Cautionary Note on the Robustness of Latent Class Models for Estimating Diagnostic Error without a Gold Standard. *Biometrics*, 60, 427-435.
- Altman, D. G. (1991). *Practical statistics for medical research*. London: Chapman and Hall.
- Ansley, N. (1990). Validity and Reliability of Polygraph decisions in Real Cases. *Polygraph*, 19(3), 169-181.
- Ben-Shakhar, G., & Furedy, J. (1990). *Theories and applications in the detection of deception*. New York: Springer-Verlag.
- Clark, J. P., & Tiffit, L. L. (1966). Polygraph and interview validation of self reported deviant behavior, *American Sociological Review*, 31, 516-523.
- Clogg, C. C. (1977). *Unrestricted and restricted maximum likelihood latent structure analysis: a manual for users*. Working paper 1977-09, Pennsylvania State University, Population Issues Research Center.
- Crocker, L. M., & Algina, James. (1986). *Introduction to Classical and Modern Test Theory*. New York: Holt, Rinehart and Winston, Inc.
- Efron, B. Tibshirani, R. (1993). *An Introduction to the Bootstrap*. New York: Chapman and Hall.
- Faraone, S. V., & Tsuang, M. T. (1994). Measuring diagnostic accuracy in the absence of a "gold standard". *American Journal of Psychiatry*, 151, 650-657.
- Horvath, F. S., & Reid, J. E. (1971). The reliability of polygraph examiner diagnosis of truth and deception. *The journal of Criminal Law, Criminology and Police Science*, 62, 276-281.
- Kassin, S. M. (1997). The psychology of confession evidence. *American Psychologist*, 52, 221-233.
- Kircher, J. C., Horowitz, D. W., & Raskin, D. C. (1988). Meta-analysis of mock crime studies of the control question polygraph technique. *Law and Human Behavior*, 12, 79-90.
- Lykken, D. T. (1998). *A Tremor in the Blood: Uses and Abuse of Lie Detection*. Now York: Plenum Trade.
- Morris, R. A. (1989). The admissibility of evidence derived from hypnosis and polygraph. In D. C. Raskin (Eds.), *Psychological Methods in Criminal Investigation and Evidence*. New York: Springer.
- National Research Council. (2003). *The Polygraph and Lie Detection*. Washington, D.C.: National Academies Press.
- Pepe, M. S., & Janes, H. (2007). Insight into latent class analysis of diagnostic test performance. *Biostatistics*, 8, 474-484.

- Raskin, D. C. (1987). Methodological issues in estimating polygraph accuracy in field applications. *Canadian Journal of Behavioural Science*, 19, 389-404.
- Raskin, D. C., & Honts, C. R. (2002). The Comparison Question Test. In Murray Kleiner (Ed.), *Handbook of Polygraph Testing*. San Diego: Academic Press.
- Rindskope, D., & Rindskopf, W. (1986). The Value of Latent Class Analysis in Medical Diagnosis. *Statistics in Medicine*, 5, 21-27.
- Spencer, B. D. (2007). Estimating the Accuracy of Jury Verdicts. *Journal of Empirical legal Studies*, 4, 305-329.
- U.S. Congress, Office of Technology Assessment. *Scientific Validity of Polygraph Testing: A Research Review and Evaluation - A Technical Memorandum*. OTA-TM-H-15, November, Washington D.C.: U.S. Government Printing Office. 1983, p.52.
- Wasserman, S., & Bockenholt, U. (1989). Bootstrapping: Applications to Psychophysiology. *Psychophysiology*, 26, 208-221.
- 논문투고일 : 2008. 08. 05
1차 심사일 : 2008. 08. 09
2차 심사일 : 2008. 09. 03
게재확정일 : 2008. 10. 10

Estimating the Accuracy of Polygraph Test

Jin-Sup Eom

Department of Psychology
Chungbuk National University

Hyung-Ki Ji

Forensic section
S.P.P.O. in Korea

Kwangbai Park

Department of Psychology
Chungbuk National University

The present study examined the accuracy of polygraph tests through two types of statistical methods with unknown ground truth. One method evaluated the accuracy based on the rates of agreements between polygraph test results of crime suspects and prosecutors' indictment decisions for them. Those crime suspects were tested with polygraph by the Prosecutors' Office of the Republic of Korea between 2000 and 2004. The other method estimated the accuracy by using the latent class analysis based on the frequency distributions of the polygraph results and indictments during 2006. Excluding cases that were 'inconclusive' on the polygraph test, the study showed that the accuracy of the polygraph tests is .914 (SE=.004) for the 2000-2004 data, and .885 (SE=.021) for the 2006 data. With the inclusion of 'inconclusive' cases in the 2006 data, the results from the latent class analysis showed the accuracy in the range between .707 and .734 (SE=.027~.031), with false positives between .078 and .087 (SE=.019~.023), and false negatives between .029 and .078 (SE=.010~.023). The probability that the polygraph test correctly classifies subjects appeared to be in the range between .912 and .925 (SE=.013-.016) for those who lie, and in the range between .867 to .955 (SE=.011-.040) for those who tell the truth.

Key words : polygraph, lie detection, accuracy, validity, latent class analysis