

## 소년범죄자 재범예측 평가 시 가중치 계산 방법에 관한 실증 연구\*

권 해 수

윤 일 홍<sup>†</sup>

김 현 정

조선대학교 상담심리학과    조선대학교 경찰행정학과    법무부 소년과

소년범죄자의 재범 위험성을 예측할 수 있는 계리적 평가 도구의 개발에 대한 실무자와 학계의 관심이 높다. 그러나 이러한 평가 도구를 개발함에 있어 가중치를 사용해야 하는지의 여부와 가중치를 사용할 경우 구체적으로 어떤 방법을 활용하는 것이 타당한가에 대해서는 그동안 충분한 논의가 이루어지지 않았다. 본 연구는 이러한 연구 공백을 메꾸고자 요인분석, 원점수, 너필드, 단순로지스틱, 다중로지스틱 등의 다섯 가지 방법으로 도출된 가중치를 활용하여 평가도구를 개발한 후 ROC 분석과 부트스트랩 기법을 활용하여 이들의 예측정확도를 비교·분석하였다. 그 결과 다중로지스틱 회귀분석법과 너필드법의 효용 가치가 가장 높은 것으로 판명되었다. 이러한 분석 결과를 바탕으로 정책적 시사점 및 향후 연구에 대한 함의를 제시하였다. 특히 예측력이 높은 평가 도구를 개발하기 위해서는 소년원 출원 후의 일상 생활과 관련된 변인들 및 정성적 변인들의 산입이 필요함을 강조하였다.

주요어 : 재범 위험성, 계리적 평가도구, 소년원, 부트스트랩 기법, ROC 분석

\* 이 논문은 2015년도 조선대학교 학술연구비의 지원을 받아 연구되었으며, 한국소년정책학회 추계학술대회(2015년 12월 4일, 안양소년원)에서 발표되었음.

<sup>†</sup> 교신저자: 윤일홍, 조선대학교 경찰행정학과, 광주광역시 동구 필문대로 309

Tel: 062-230-6782, E-mail: yun.ilhong@gmail.com

일반 시민들은 사회가 발달할수록 범죄의 양이 증가한다고 믿는 반면, 범죄학자들은 사회의 발달 수준과 범죄의 양이 사실상 반비례 관계에 놓여 있다는 점을 명확하게 인식하고 있다. 범죄의 양에 대한 역사적 연구(Spitzer, 1975) 및 최근 서구 선진국 연구(Van Dijk, 2006)를 감안컨대 우리나라의 경우도 사회 발전에 따라 절대적 범죄의 양은 계속해서 줄어들 것이라고 예상할 수 있다(황지태, 2009 참조).

범죄의 양을 표시하는 범죄 통계는 학자들에게는 감정이 개입할 여지가 없는 객관적 숫자들의 조합에 불과할 수 있다. 그러나 이러한 숫자들이 범죄 피해자와 그 가족들에게는 눈물과 고통을 의미하며 빼앗겨버린 재산, 심지어는 강탈당한 소중한 생명을 뜻하기도 한다. 국가 차원에서는 피해자 구조, 가해자의 처벌, 수용, 처우, 재활 등에 소요되는 막대한 경제적 비용 부담을 의미한다. 따라서 비록 범죄의 양이 줄어들고 있다고 하더라도 형사 정책 관계자와 연구자들은 긴장의 끈을 놓을 수 없다.

범죄량의 감소에도 불구하고 성인범죄와 구분되는 소년범죄는 최근 들어 전체적으로 증가추세를 보이고 있다. 특히 관련 실무자와 학자들은 소년범죄자 재범 비율이 지속적으로 증가하고 있는 현실을 주목해야 한다(강동욱, 2012; 윤옥경, 2011). 코호트(cohort) 중 10% 미만의 소수가 전체 코호트가 일생동안 저지르는 폭력 범죄의 50% 이상을 저지른다고 한다(Wolfgang, Figlio, & Sellin, 1972). 따라서 이러한 소년범죄자들에 대해 시의적절한 개입이 이루어져 이들이 ‘생애지속형범죄자(life-course persistent offender)’로 진행되는 것을 미리 방지할 수 있다면 장래 우리 사회 전체의 범죄량

은 눈에 띄게 줄어들 수 있을 것이다.

소년원은 우리나라에서 생애지속형범죄자로 성장할 가능성이 많은 다수의 소년범들이 동일한 시간대에 동일한 장소에 수용되어 있는 소년보호기관이다. 소년원의 구조적 특성은 소년범들에 대해 체계적이고 조직적인 개입과 교정을 가능하게 해준다. 즉, 소년원 내에서 소년범죄자들이 받는 처우, 훈련, 서비스의 종류, 강도 등은 이들이 출원 후 범죄자로서의 길을 걷게 될지, 선량한 시민으로서의 삶을 살게 될지의 여부에 영향을 미칠 수 있다. 한발 더 나아가 소년원의 교정정책은 소년범죄자들이 생애지속형범죄자로 진행될 것인지 아닌지의 여부에 영향을 미쳐 장래 우리 사회 전체 범죄량을 결정짓는 유의미한 역할을 담당할 수도 있을 것이다.

이와 같은 소년원의 중요한 역할을 염두에 두고 그동안 법무부는 소년원에서의 교정은 감시와 처벌보다 교육과 보호가 재범 방지에 효과적이라는 입장을 유지하며 교과교육, 인성교육, 그리고 직업능력개발훈련에 중점을 두어 왔다(권해수, 이연상, 2013; 박영규, 2011). 그리고 과거 소년원의 부정적인 이미지를 없애기 위해 소년원이라는 명칭 대신 OO학교로 그 명칭을 변경하고 소년범죄자의 교정에 힘쓰고 있다. 그러나 최근 들어 소년범 재범률이 지속적으로 증가하자, 그동안 소년사법기관이 소년범들의 위험성 - 욕구 - 반응성(risk - need - responsivity)을 충분히 파악하지 못하였으며, 차별적이고 선택적인 교정 처우를 행하지 못했다는 비판이 점차 확산되었다(윤옥경, 2011). 이에 따라 소년범의 재범위험성의 판단을 실무자의 직관적 또는 자의적인 판단에 맡기기보다는 통계분석기법을 활용하여 보다 객관적인 계리적<sup>1)</sup>(actuarial) 평가도구를 개발하여

이에 입각한 차별적, 선택적 교정처우를 해야 한다는 목소리가 높아졌다(서석교, 2011). 우리나라 성인 보호관찰 대상자에 대해서는 이미 그러한 도구인 K-PRAI(Korean Probationers Risk Assessment Inventory)가 개발, 활용되고 있으며 소년보호관찰 영역에서도 재범예측 평가도구에 대한 다양한 연구가 행해졌다(김양곤, 이수정, 이민식, 2005; 노일석, 2007, 2009, 2010; 노일석, 정진경, 2009). 그러나 소년원생의 재범위험성 예측에 대한 경험적 연구는 상대적으로 미흡하다. 2008년 범죄예방정책국 소년과에서는 비행위험성 진단체계 개발을 위해 소년보호 교육종합관리시스템(Total Education and Management System: 이하 TEAMS)을 구축하였으나, TEAMS를 활용한 재범위험성 판단을 위한 계리적 평가도구의 개발은 상당히 미흡했다. 단, 서석교(2011)가 2008년 8월부터 2년에 걸쳐 전국 8개 소년원에 입원한 소년들에 대한 TEAMS 자료를 바탕으로 재범 위험성을 조사한 연구가 유일하다.

일반적으로 계리적 평가도구는 기존 연구결과들을 반영하여 재범위험성을 예측할 수 있을 것으로 판단되는 다수의 위험요인(risk factor) 또는 문항들로 구성되어 있다. 특정 위험요인이 한 개인에게 존재하지 않을 경우 그 개인에게 0점이 부여되고 위험요인이 존재하는 경우에는 일정한 양의 정수가 부여된다. 따라서 점수가 높을수록 해당 위험요인이 더 강하게 존재함을 의미한다. 각 항목에 부여된 점수들을 전부 합한 점수가 최종 진단점수가 되고 이 점수의 고저에 따라 개인의 재범위험성의 정도가 예측된다. 결국 이와 같은 계

리적 평가도구는 재범 위험성 판단의 결정적 자료로 활용되는 최종 진단점수를 산출하기 위한 것이다.

그러나 최종 진단점수를 산출하는 최적의 방법에 관해서는 아직까지 학자들 간에 의견의 일치가 이루어지지 않고 있다. 특히 최종 점수를 산출하기 위해 가중치를 사용해야 하는지의 여부와 가중치를 사용할 경우 구체적으로 어떤 방법을 사용하는 것이 타당한가에 대해서는 전혀 논의가 이루어지지 않은 채 연구자들의 개별적 선호에 따라 다양한 방식이 활용되어 왔다. 예를 들어 서석교(2011)는 요인분석 결과 도출된 요인부하량(factor loading)을 이용하여 가중치로 사용하였다. 반면, 노일석(2007; 2009; 2010) 및 노일석과 정진경(2009)은 다중 로지스틱 회귀분석 결과의 산출물인 승산비(odds ratio) 또는 회귀계수를 가중치로 활용하였다. 이에 비해 김양곤, 이수정, 이민식(2005)은 가중치를 아예 사용하지 않고 원점수(raw score)를 그대로 이용하여 최종점수를 계산해 냈다. 이와 같이 일관성 없이 다양한 방법이 사용되고 있기 때문에 연구자들과 실무자들 또한 적지 않은 혼란을 경험할 가능성이 높다. 따라서 재범위험성 평가를 위한 계리적 도구의 개발에 앞서, 최종점수 산출 시 과연 가중치를 사용해야 하는지, 또 그럴 경우 어떤 방법으로 가중치를 도출할 때 변별력이 가장 높은 최종점수를 계산해 낼 수 있는지를 검증해보는 것은 이론적으로 또한 실무적으로도 그 의의가 크다고 할 수 있다.

재범위험성 평가를 위해 계리적 도구를 활용하여 최종진단점수를 산출할 경우 가중치를 사용할 것인가의 여부에 관해 서구 학자들 사이에서 일부 논의가 있어왔다. 먼저 스웨덴의 Grann과 Langstrom(2007)은 가중치 없이 원점수

1) ‘계리적’ 평가라 함은 정해진 규칙과 체계적 방법에 따라 도출된 숫자들을 평가의 자료로 활용하는 평가방법을 의미한다.

(raw score)를 그대로 사용할 것을 권장하였다. 이들은 폭력사범의 재범위험성을 예측하는 연구에서 10개의 항목으로 구성된 평가도구를 이용하였다. 그들은 먼저 위와 같은 10개 항목에 대해 회귀분석을 포함한 네 가지 다른 방법을 이용해 각각의 가중치를 도출한 후 이에 따라 최종점수를 계산해 내고 마지막 방법으로는 가중치를 부여하지 않고 최종점수를 산출하여 보았다. 이와 같이 다섯 가지 방법으로 산출된 최종점수들의 예측정확도를 ROC (Receiver Operating Characteristics; 이하 ROC) 분석을 통하여 비교해 본 결과, 가중치를 부여하지 않는 방법이 예측 정확도가 가장 높은 것으로 판명되었다. 특히 이들은 회귀분석을 사용하여 가중치를 도출해 낸 방법을 원점수를 사용한 방법보다 재범 여부를 더 정확하게 예측하지 못하는 이유는 바로 범죄학계의 이론 및 경험적 연구 수준이 아직까지도 재범에 이르는 인과적 메커니즘을 명확하게 규명하지 못하고 있기 때문이라고 주장하였다. 따라서 이들은 재범 연구의 획기적인 진전이 있기 전까지는 회귀분석을 통한 가중치를 사용하기보다는 예측항목들의 원점수를 이용해야 한다고 주장하였다(Grann & Langstrom, 2007). 같은 맥락에서 Cohen과 Cohen(1983) 회귀분석법과 원점수법의 예측정확도를 비교하여 살펴본 바, 회귀계수를 가중치로 사용하는 방법이 원점수를 사용하는 방법에 비해 결코 우수하지 않다는 결론에 도달했다.

기존 서구 선진국 연구에서 가중치를 사용하는 방법들로는 계산의 복잡성 정도에 따라 너필드(Nuffield)법, 단순로지스틱 방법, 그리고 다중로지스틱 방법 등을 들 수 있다. 먼저 계산방법이 가장 단순한 너필드 방법은 캐나다에서 가석방 여부를 결정하기 위해 개발되었

다. 이 방법은 Nuffield(1982)에 의해 사용된 후, 주로 캐나다와 미국에서 활용되었고 그 효용성 또한 수차례 입증되었다(Bonta, Harman, Hann, & Cormier, 1996; Quinsey, Rice, & Harris, 1995). 가중치를 계산하는 방법은 특정 예측변인의 특정 점수가 표본의 평균 재범률을 5% 증가시킬 때마다 각 +1의 가중치를, 5% 감소시킬 때마다 각 -1의 가중치를 부여하는 방법이다.<sup>2)</sup>

회귀분석을 이용한 방법으로는 단순로지스틱 방법과 다중로지스틱 방법이 있다. 두 가지 방식 모두 로지스틱 회귀분석 결과 도출된 회귀계수 또는 승산비(odds ratio)를 가중치로 사용한다는 점에서는 동일하다. 그러나 다중로지스틱 방법은 예측변인들 사이의 공변량을 제외하고 종속변수와 특정 예측변인 사이의 고유한 공변량만을 기준으로 가중치를 설정하는 반면, 단순로지스틱 방법은 예측변인들 사이의 공변량을 제외하지 않은 상태에서 종속변수와 예측변인 사이의 공변량을 측정하여 이를 기준으로 가중치를 설정한다는 점에서 차이가 있다. 사실상 지난 20여 년 동안 다중로지스틱 방법을 사용하여 가중치를 설정하는 방법이 학자들 사이에서 가장 빈번하게 활용되어 왔다. 그 이유로는 이 방법이 다른 예측변인들이 재범 여부에 미치는 영향력을 통제

2) 예를 들어 전체 표본의 평균 재범율이 20%인 경우, 만약 아동학대를 받은 경험이 있는 자들의 재범율은 그보다 11% 증가한 31%였다면 아동학대라는 예측변인은 +2의 가중치를 부여받게 된다. 만약 고등학교를 졸업한 자들의 평균 재범율이 4%에 불과하다면 고등학교 졸업이라는 예측요인은 -3의 가중치를 부여받게 된다. 이와 같이 도출된 가중치와 각 해당 항목의 점수를 곱한 후 이 값을 전부 더한 점수가 최종 진단점수가 된다.

한 상태에서 특정 변인의 고유한 영향력을 측정할 수 있다는 점, 주효과(main effect) 외에도 두 개의 예측변인 사이의 상호작용효과(interaction effect)를 살펴볼 수 있다는 점 등을 들 수 있다(Duwe & Freske, 2012). 더불어 다중로지스틱 회귀분석 결과 생성되는 예측 확률(predicted probability)을 최종 진단점수로 대응할 경우 그 개념의 해석이 원점수의 해석에 비해 용이하다는 점도 들 수 있다(Grann & Langstrom, 2007)<sup>3)</sup>.

이러한 장점들 때문에 다중로지스틱법이 자주 활용되고 있으나 이에 대해서도 여러 비판이 제기된다. 가장 큰 문제점은 회귀분석이 모형의 과적합화(overfitting) 현상을 유발할 수 있다는 것이다(Duwe & Freske, 2012; Quinsey et al., 1995). 즉, 회귀모형은 사용된 표본자료와의 적합도(fit)가 가장 큰 상태에서의 회귀계수를 계산해내는 알고리즘을 사용하기 때문에 이러한 모형은 표본자료가 갖는 구조적 특성뿐만 아니라 우연적 성질 또는 잡음(noise)까지도 모두 반영하게 된다. 그리하여 결과적으로 이러한 회귀모형을 다른 표본에 적용했을 때는 그 적합도가 원래의 적합도보다 현저하게 떨어질 수밖에 없다는 것이다. 이러한 문제점을 해결하기 위해 일부 학자들은 애초에 표본을 둘로 나누고 나뉜 표본의 일부(subsample)는 회귀분석을 통한 계리적 도구를 개발하는데 사용하고 나머지 일부는 이 도구의 적합도를 검증하는 데 사용하는 내적 검증(internal validation)방법을 택하기도 하였다(Grann & Langstrom, 2007). 그러나 이러한 경우에도 오

3) 예를 들어 ‘2년 이내에 재범을 저지를 위험성이 10점이다’라는 표현보다는 ‘재범을 저지를 확률이 10%이다’라고 하는 것이 이해하기가 훨씬 수월하기 때문이다.

직 한 두 개의 추가표본에 대해서만 검증을 하는 경우 우연성을 배제하기가 어렵기 때문에 다수의 표본에 대해 검증을 하는 것이 바람직하다는 지적이 있다(Cohen, 1990).

마지막으로 요인분석을 활용하여 가중치를 계산하는 방법이 있다. 문헌분석 결과 요인분석법이 서구 선진국에서 재범 위험성을 예측하기 위한 연구에서 사용된 경우는 찾아볼 수 없었다. 원래 요인분석법은 도시·지역개발 분야에서 삶의 질이나 개발 격차를 측정하기 위해 개발되었고(Drewnowski, 1974) 우리나라에서도 역시 같은 분야의 학자들에 의해 비교적 빈번하게 활용되었다(김덕준, 2003; 김태일, 1999; 박성복, 1996). 요인분석법이 형사사법 분야에서 재범위험성 평가도구에 활용된 것은 서석교(2011)의 연구가 유일하다. 이 연구는 TEAMS에 있는 8,823명의 소년범죄자의 자료를 사용해 총 3단계의 요인분석을 실시하여 최종진단점수를 산출해냈다. 구체적으로 먼저 TEAMS에 내재해 있는 107개 세부 항목에 대해 탐색적 요인분석을 실시하여 총 27개 하위요인을 추출하고, 각 요인에 적재된 세부항목들과 그 항목들의 요인부하량을 곱한 후(즉, 요인부하량을 각 세부항목의 가중치로 상정함)<sup>4)</sup> 이들을 전부 합한 값을 해당 하위요인의 값으로 하였다. 다음 단계로 이와 같이 추출된 27개 하위요인들의 상위요인으로 8개의 진단 영역을 가정하여 확인적 요인분석을 실시

4) 서석교는 요인부하량(factor loading)을 가중치로 활용하였으나 도시·지역 개발 연구들에서는 주로 요인점수계수(factor scoring coefficient)를 가중치로 사용하였다. 요인점수계수가 요인부하량을 이용하여 산출되므로 둘 사이의 상관관계는 높다고 할 수 있으나 원칙적으로 두 값은 서로 상이하므로 최종 결과물인 최종점수 또한 어느 정도의 차이가 있게 된다.

하고 각 하위요인값과 하위요인의 요인부하량을 곱한 후 이들을 전부 합하여 각 진단 영역의 값으로 하였다. 마지막 단계로 위 8개의 진단 영역을 최상위 요인인 최종진단점수의 하위요인들로 간주하여 다시 한 번 확인적 요인분석을 실시하고 각 진단영역 값과 그에 해당하는 요인부하량을 곱한 후 이들을 전부 합하여 이 값을 최종진단점수로 사용하였다.

요인분석을 사용하여 가중치를 계산하는 방법은 예측변인의 수가 많을 때 변인 간 체계적 구조를 설립할 수 있다는 점에서 활용 가치가 있다(김태일, 1999). 그러나 이 방법에 대해서는 다음과 같은 비판들이 제기될 수 있다. 먼저, 재범위험성 평가 체계와 요인분석의 인과 구조가 그 방향성에 있어 서로 상반된 원리에 기초한다는 점이다. 전자의 경우에는 하위 예측요인들(개인적 요인, 환경적 요인, 범 죄관련 요인 등)이 상위목표인 재범위험성에 영향력을 미치는 하위 → 상위의 구조, 즉 하위 예측요인이 원인, 상위목표인 재범위험성이 결과가 되는 구조를 지닌다. 그러나 요인분석의 경우는 상위목표인 재범위험성이 원인, 하위지표인 예측요인들이 결과인 상위 → 하위의 구조를 지니게 된다.<sup>5)</sup> 즉, 요인분석상의 인과 구조는 재범위험성 예측모형의 인과 구조와는 원칙적으로 동일하지 않기 때문에 요인부하량을 가중치로 사용하는 경우 계산상 오류에 노출될 위험성이 있다(김태일, 1999 참조). 두 번째로는 요인분석을 사용해 가중치를 추출해내는 경우 가중치가 오로지 각 지표들 사이의 상관관계에 의해서만 결정될 수밖에 없다는 점이다(Cattell, 1958). 따라서 특정 지표

가 상위요인의 값을 결정하는데 이론적으로 아주 중요한 지표라 할지라도 이 지표와 동일 요인에 적재된 다른 지표들과의 상관관계가 비교적 낮은 경우는 낮은 가중치가 설정될 수밖에 없다. 반대로 특정 지표가 상위요인과의 상관관계는 비교적 낮지만 다른 지표들과의 상관관계가 높은 경우에는 높은 가중치가 부여된다는 취약점이 있다. 서석교(2011)의 연구에서는 최종점수를 산출해내기 위해 TEAMS에 내재한 107개 세부 항목들을 모두 사용하였다. 통상적으로 서구 선진국에서 사용된 계리적 도구들이 10개 이내의 항목들로 구성되어 있다(Duwe & Freske, 2012; Grann & Langstrom, 2007; Quinsey et al., 1995)는 점을 감안하면 서석교(2011)의 연구는 필요 이상의 많은 항목들을 사용하였다고 평가할 수 있다. 더구나 사용된 항목들의 일부는 최저 요인부하량의 기준인 .32(Tabachnick & Fidell, 2001)에 미치지 못하기도 하였고 또한 일부는 최종 결과변수인 재범 여부와의 상관관계가 유의미하지 않은 경우도 있었다. 이는 간명한 통계모형이 우수한 모형이라는 간명성의 원칙(Cohen, 1990)에 반한다고 할 수 있다. 마지막으로 회귀분석방법의 경우 과적합화가 문제될 수 있다는 지적은 요인분석법의 경우에도 마찬가지로 적용된다(김태일, 1999). 즉, 요인분석에 의한 가중치의 부여가 분석에 사용된 표본에 따라 그 값이 가변적일 수 있기 때문에 추출된 가중치를 다른 복수의 표본들을 상대로 검증해보지 않는다면 그 결과의 신뢰성을 담보하기가 쉽지 않을 것이다.

재범위험성을 평가하는 계리적 도구를 개발함에 있어 가중치 부여에 관한 문제는 최종진단점수 계산에 직접적으로 영향을 미치기 때문에 극히 중요한 쟁점이라고 할 수 있다.

5) 구조방정식모형 그림에서 화살표가 잠재요인에서 관찰지표들 쪽으로 향해져 있음을 떠올리면 쉽게 이해할 수 있다.

서구의 연구에서는 재범위험성 평가도구와 관련하여 가중치를 부여할 것인지 여부와 부여한다면 어떤 종류의 가중치를 활용하는 것이 가장 합당한지에 대해 학자들 간에 논의가 있어왔다(Cohen, 1990; Duwe & Freske, 2012; Grann & Langstrom, 2007, Quinsey et al., 1995). 국내에서는 도시·지역개발 관련 평가지표 체계와 관련된 일부 경험적 연구가 진행된 바 있다(김태일, 1999; 2000; 김덕준, 2003). 그러나 형사사법 분야의 재범위험성 평가와 관련된 선행연구는 찾아볼 수 없다. 기존 연구들은 이론적 근거를 명시하지 않은 채 연구자의 선호에 따라 가중치 없는 원점수를 사용하거나(김양곤, 이수정, 이민식, 2005) 회귀분석 후 회귀계수나 승산비를 활용하거나(노일석, 2009; 2010) 또는 요인분석 후 요인적재량을 가중치로 사용하였다(서석교, 2011).

통계적 방법을 사용해서 재범위험성 예측도구를 개발할 경우 통계모형이 그 모형을 도출하기 위해 사용된 표본에만 과적합하는 현상이 발생할 수 있으므로 그 모형의 적합성을 내·외적으로 검증해보는 절차가 필요하다(Cohen & Cohen, 1983). 재범위험성 평가와 관련한 기존 국내 연구의 경우 아직까지 위와 같은 검증을 행한 사례는 없다. 따라서 연구 당시 개발된 평가도구들이 추후 다른 표본에서도 동일한 정도의 예측력을 발휘하는지에 대해서도 확인된 바가 없다.

이에 본 연구에서는 재범위험성 평가도구의 개발에 있어서 기존 연구에서 다루지 않았던 가중치 부여 문제를 다루고자 한다. 특히 소년원에 입원한 소년범죄자들의 재범위험성 평가도구 개발과 관련된 연구가 미흡한 관계로 이들의 재범위험성 평가도구에 있어서의 가중치 부여 문제를 조사해 보고자 한다. 구체적

으로 TEAMS의 자료를 바탕으로 기존의 국내 또는 서구의 연구에서 활용되었던 요인분석법, 원점수법, 너필드법, 단순로지스틱 방법, 다중로지스틱 방법 등 총 다섯 가지의 방법을 이용하여 최종점수를 산출한 후, ROC 분석을 활용해 가중치를 사용한 경우와 하지 않은 경우, 또한 어떤 종류의 가중치를 사용한 경우에 예측정확도가 가장 높은지를 비교·검증해보고자 한다. 또한 기존 국내 연구들이 개발된 평가도구들의 추정치에 대해 그 외 다른 추가 표본을 대상으로 검정하는 절차를 거치지 않았다는 점을 고려하여 본 연구에서는 부트스트랩(bootstrapping) 기법을 통해 예측 정확도에 대한 유의성을 검증해 보고자 한다.

## 방 법

### 연구 대상

본 연구는 소년원에 재원했다가 출원한 사실이 있는 남자 소년범죄자들을 그 대상으로 하였다. 여자 재원생들은 남자에 비해 그 숫자가 훨씬 적고 또한 재범위험성 평가도구는 성별을 구분하여 개발하는 것이 원칙(노일석, 2000)이라는 점을 감안하여 여자 소년범죄자는 본 연구에서 제외하였다. 연구대상은 법원 소년부로부터 장기소년원 처분(10호)과 단기소년원 처분(9호)을 받고 전국 8개 소년원 중 한 곳에 입원했다가 2013년 6월 1일부터 2014년 12월 31일까지 사이에 출원한 남자 소년범죄자 1,394명이다. 재범 여부의 관찰기간은 대상자들의 각 출원 시점부터 2015년 5월 30일까지의 기간이다. 재범 여부는 소년원에서 출원한 소년이 다시 비행을 저지른 후, 소년원에

표 1. 연구대상의 인구통계학적 특성

구분		N	%	구분		N	%
연령 <sup>6)</sup>	14세 이하	112	8.1	비행명	강도 등	80	5.7
	15-16세	548	39.3		성폭력	89	6.4
	17-18세	671	48.1		폭력행위 등	128	9.2
	19세 이상	63	4.5		절도 등	482	34.6
보호관찰법위반					474	34.0	
학력	초등	11	.8	기타	141	1.1	
	중등	496	35.6	처분종류	9호 처분	804	57.7
	고등	877	62.9		10호 처분	590	42.3
	대학	10	.7	재범여부	재범 없음	1,144	82.1
가정환경	친부모가정	590	42.3		재범	250	17.9
	결손가정	804	57.7				

재입원하였는지 여부를 기준으로 판단하였다. 재입원 여부를 기준으로 재범 여부를 판단할 경우 경찰입건 여부를 기준으로 하는 것보다 재범의 범위를 다소 축소시킬 우려가 있다. 그러나 소년원에서 출원한 소년이 다시 재비행을 저지르는 경우 재입원될 가능성이 매우 높기 때문에 재입원 여부만으로도 진값에 가까운 재범률을 측정할 수 있을 것으로 판단되었다(서석교, 2011).

표 1은 본 연구에 사용된 표본의 인구통계학적 특성을 나타내고 있다. 연령별로는 약 절반 정도가 17~18세였고 15~16세는 약 39.3%, 나머지는 14세 이하 또는 19세 이상이였다. 고등학교를 중퇴, 재학, 또는 졸업한 소년들이 62.9%로 가장 많았고, 중학교를 유예, 재학, 또는 퇴학한 소년들이 35.6%를 차지하였다. 가정환경별로는, 친부모가 현재 모두 있는 정상적인 가정은 42.3%에 불과한 반면, 편

부모, 조손, 또는 이혼가정 등 결손가정 소년은 57.7%에 달하였다. 소년원 입원 당시의 비행별로는 절도(34.6%)와 보호관찰법위반(34.0%)이 대부분을 차지했고, 폭력사범의 비율은 그다지 높지 않은 편이었다. 장기처분인 10호처분(42.3%)보다 단기처분인 9호처분(57.7%)으로 입원한 소년의 비율이 약간 더 높았다. 마지막으로 재비행으로 인해 소년원에 재입원한 소년은 전체소년의 17.9%로서 이는 표본의 재범율이 17.9%였음을 의미한다. 본 연구에서 종속변수는 출원생이 재비행으로 인해 다시 소년원에 재입원을 했는지의 여부, 즉 재비행 여부이다.

#### 자료수집 방법 및 내용

본 연구에 사용된 자료는 법무부 소년보호 교육종합관리시스템(TEAMS)로부터 추출하였다. TEAMS에는 기본적으로 재원생의 신상정보, 교육정보, 원내 생활정보 등이 포함되어 있고 그 외에도 총 5개의 측정도구<sup>7)</sup>로 구성되

6) 연령은 재범관찰기간이 경과한 2015년 6월 1일을 기준으로 하여 소년원에 입원 또는 재입원해 있는 원생들의 나이를 측정한 것이다.



어 있다. 본 연구는 재범 여부와 직접적 연관성이 다소 떨어지는 노동부 직업흥미검사를 제외한 나머지 4개의 측정도구로부터 추출된 자료들을 사용하였다. 소년범의 부모 및 가정 환경, 친구관계, 흡연 및 약물복용, 비행경력, 학교생활 등에 관한 내용은 TEAMS 기본 정보와 분류심사설문지로부터 입수하였고, 소년원 내에서의 생활 성적, 상·벌점, 교육에 임하는 성실성 및 교육태도에 관한 자료들은 위탁생 행동평가척도에 있는 항목들을 사용하였다. 특수인성검사-R은 비행청소년의 재비행 가능성을 진단할 목적으로 법무부에서 자체 개발한 집단 성격검사이다. 연구대상자들의 일탈 가능성을 측정하기 위해서 이 측정도구의 반사회성, 공격성, 충동성, 주의력결핍 항목을 사용하였고, 성격 및 성향에 관해서는 불안, 우울, 성실성, 자아존중감, 자기효능감과 관련된 항목들을 사용하였다. 마지막으로 특수지능·적성검사는 비행청소년들의 지적 수준 및 적성을 파악할 목적으로 개발된 도구로서 본 연구에서는 대상자들의 종합지능지수를 추출하기 위해 이 도구를 사용하였다.

#### 자료분석 방법 및 절차

이 연구를 위한 자료분석은 Stata 13 (StataCorp, 2013)을 이용하였고 다음과 같은 분석단계를 거쳤다. 먼저 TEAMS 진단체계를 구성하는 위 네 가지의 측정도구로부터 범죄학 이론 및 경험적 연구 결과 재범에 영향을 미치는 요인들로 추정되어지는 변인 115개를 추출하였다. 이들 변인들은 크게 가정환경, 인성,

비행성, 약물남용, 여가활동, 친구관계, 소년원 내 생활태도의 진단영역으로 구별될 수 있었다. 각 진단영역에 속한 구체적 변인들은 서석교(2011)의 논문을 참조하였다.

## 결 과

### 요인분석법

먼저 요인분석을 이용한 가중치를 계산해내기 위해서 서석교(2011)가 사용하였던 분석방법을 되풀이하여 3단계의 요인분석을 실시하였다. 단, 위 115개 변수를 모두 사용하지는 않고 통계모형의 간명성(parsimony)을 추구하여 재범 여부와 상관관계가 .05 유의수준에 미치지 못하는 변수들은 제거하고 유의한 상관관계를 유지한 47개의 변인들로만 요인분석을 실시하였다. 그리고 요인분석에서 요인부하량이 .32에 이르지 못하는 변인들은 탈락시켰다 (Tabachnick & Fidell, 2001).

먼저 요인분석에 앞서 위 47개 변수들의 값을 모두 z 점수로 표준화하였다. 제1단계에서는 위 변수들을 대상으로 각 진단영역별로 탐색적 요인분석을 실시하였다. 고유값(eigenvalue) > 1과 스크리도표의 급경사를 판단 기준으로 삼아 요인을 추출하였다. 진단영역별로 1개부터 7개까지, 총 17개의 요인이 추출되었다. 각 요인별로 적재된 변인들의 값과 이 변인들의 요인부하량을 곱한 후 이를 전부 합한 값을 해당요인의 값으로 책정하였다. 제2단계에서는 위 17개 요인들의 상위요인으로 8개의 진단영역을 가정하여 확인적 요인분석을 실시하고 이때 발생한 요인부하량을 가중치로 활용하여 각 진단영역의 값을 계산

7) 여기에는 분류심사설문지, 위탁생 행동평가척도, 특수인성검사-R, 특수지능·적성검사, 노동부 직업흥미검사가 포함된다.

하였다. 마지막 단계에서는 위 8개 진단영역을 최상위 요인인 최종진단점수의 하위요인으로 간주하여 확인적 요인분석을 실시한 후 요인부하량을 가중치를 이용하여 최종진단점수를 산출하였다. 최종점수의 범위는 -.84에서 1.18에 이르렀고 평균은 .01( $SD = .31$ )이었다.

### 원점수법

다음 단계로 원점수법, 너필드법, 단순로지스틱 방법, 그리고 다중로지스틱 방법을 이용하여 각각의 최종진단점수를 구하였다. 이들 방법들은 변인 간 체계적 구조를 확립하고자 하는 요인분석과는 그 목적이 다르기 때문에 굳이 위 47개 변인들을 전부 사용할 필요는 없다. 서구 선진국의 재범위험성 평가도구가 통상 10개 미만의 항목들로 구성된 점을 감안한다면 본 연구에서도 위 47개 변인들 중 재범 여부에 가장 큰 영향력을 미칠 수 있는 변인들을 선별할 필요성이 있었고 이를 위하여 다중로지스틱 회귀분석을 실시하였다.

회귀분석에 앞서 위 47개 변인들 사이의 공선성(collinearity)을 공차한계(tolerance)와 분산팽창인자(VIF)값을 기준으로 조사하였다. 그 결과 나이와 출원 나이 사이에 공선성이 확인되어 나이 변인을 탈락시켰다. 재범 여부를 종속변인으로 한 후 나머지 46개 변인들에 대해 후진제거법(backward elimination method)을 활용하여 로지스틱 모형을 추정하였다. 유의수준 .05에서 재범 여부에 영향을 미치는 것으로 확인된 변인들은 총 8개였다. 이들은 구체적으로 ① 평소 외출 시 본인의 행선지를 부모(보호자)가 알고 있는 지 여부, ② 비행력 횟수<sup>8)</sup>, ③ 일주일 이상 무단결석을 처음 시작한

8) '비행력횟수'는 소년원 처우담당자가 범 죄 및 수

시기, ④ 가출횟수, ⑤ 소년원 재원 시 면회횟수, ⑥ 취업여부<sup>9)</sup>, ⑦ 출원 나이, ⑧ 소년범이 출원한 후 담임선생이 직접방문, 전화 또는 SNS 등을 통하여 소년범을 사후지도한 건수<sup>10)</sup>였다. 표 2는 8개 변인들의 평균값, 표준편차, 범위, 재범 여부와의 상관계수를 나타내고 있다. 재범 여부와 가장 높은 상관계수를 보이는 변인은 소년범의 출원 시 나이( $r = -.26$ )였고, 취업 여부( $r = -.13$ )와 가출 횟수( $r = .12$ )가 그 뒤를 이었다.

일반적으로 위험성 평가도구는 해석 및 사용상의 편의를 위해 연속형 변인들을 주로 이분변수 또는 삼분 범주형 변수로 전환, 표준화하여 사용한다(Duwe & Freske, 2012; Grann & Langstrom, 2007). 본 연구에서도 유의성이 입증된 위 8개 변인들을 '0'은 재범위험성이 낮고 '1'은 위험성이 높음을 의미하도록 이분변인으로 표준화하였다. 이때 각 변수의 이분점은 노일석(2010)이 제시한 대로 각 변인들의 분류정확도가 최대에 이르는 지점을 기준으로 하였다. 구체적으로는 ① 외출 시 부모 또는 보호자가 소년의 행선지를 잘 모르는 경우, ② 비행력 횟수가 3회 이상인 경우, ③ 재원하는 동안 면회횟수가 총 15회 미만인 경우,

---

사자료표, 분류심사서 등을 근거로 작성한 것으로 소년원 입원 전에 경찰에 인지되어 검찰청, 법원 등에서 조사 또는 처분을 받은 횟수를 TEAMS에 기록한 것이다.

9) 취업 여부는 재원 중 또는 출원 후 유선 또는 방문지도 등을 통해 취업사실을 확인하여 TEAMS에 기록한 것이다.

10) 법무부에서는 소년원생이 출원한 후 6개월에서 1년 동안 현장방문 또는 전화상담 등 사후지도를 실시하여 안정적으로 사회에 복귀할 수 있도록 지원하는 '희망 도우미 제도'를 2013년부터 도입하여 실시하고 있다.

표 2. 유의한 변인들의 평균, 표준편차, 범위 및 재범여부와 상관계수

변인	평균	표준 편차	범위(코딩)	재범여부와 상관계수
외출 시 행선지를 부모가 알고 있는지	2.84	.87	1(모름) ~ 4(항상 알고 있음)	-.12***
비행력 횟수	4.42	3.92	0 ~ 34회	.09**
일주일 이상 무단결석을 처음 시작한 시기	3.11	1.13	1(초등시절) ~ 5(고교시절)	-.18***
가출 횟수	2.73	1.11	1(없음) ~ 4(5회 이상)	.12***
재원 시 면회 횟수	18.59	25.88	0회 ~ 362회	-.09***
취업 여부	.17	.38	0(무직) ~ 1(취업)	-.13***
출원 나이	17.30	1.38	13세 ~ 21세	-.26***
사후지도건수	1.50	8.39	0 ~ 70	-.11***

\*\*  $p < .01$  \*\*\*  $p < .001$

④ 처음으로 일주일 이상 무단결석한 때가 중학교 2학년 이전인 경우, ⑤ 가출횟수가 3회 이상인 경우, ⑥ 출원 나이가 16세 이전인 경우, ⑦ 사후지도건수가 13회 이하인 경우는 재범위험성이 높은 것으로 간주하여 '1', 그 반대의 경우는 '0'으로 처리하였다. 별도의 가중치를 부여하지 않는 원점수법은 위 8개 변인들의 값을 전부 합한 값을 최종점수로 보는 것이다. 원점수법으로 최종점수를 산출한 경우 범위는 0에서 8이며 평균값은 4.91( $SD=1.54$ )이었다.

#### 너필드(Nuffiled)법

위와 같이 이분변수로 표준화된 8개의 변인들을 사용하여 가중치를 부여하는 방법들 중에서 가장 단순한 방법은 너필드법이다. 전술한 바와 같이 이 방법은 예측변인이 표본의 평균 재범률을 5% 증가시킬 때마다 각 +1의 가중치를 부여하는 방법이다. 본 연구 표본의 재범률은 17.9%이다. 이에 비해 외출 시 부모

가 행선지를 모르는 경우 소년들의 재범률은 20%였다. 비록 이들의 재범률이 표본 전체의 평균 재범률보다 높지만 그 차이가 5%를 초과하지 않았기 때문에 외출 시 행선지를 부모가 알고 있는지 여부 변인에는 별도의 가중치가 부여되지 않았다. 이와 반대로 출원 나이의 경우, 16세 이전에 출원한 소년들의 재범률이 33.8%로서 평균 재범률보다 15.9% 더 높았다. 따라서 출원 나이 변인에는 +3의 가중치가 부여되었다. 이와 같은 방법으로 부여된 가중치들을 활용하여 각 소년별로 최종진단점수를 계산하였다. 표 3은 너필드법에 의해 산출된 가중치를 나타내고 있다. 각 변인의 값에 이 가중치를 곱한 후 전부 합산한 것이 최종진단점수이며, 이 경우 범위는 0에서 12, 평균은 6.63( $SD=2.52$ )이었다.

#### 단순로지스틱법

다음으로 단순로지스틱법에 의해 산출된 승산비( $Exp(B)$ )를 사용하여 최종진단점수를 계산

표 3. 너필드법에 의한 가중치 부여

변인	재범률	가중치
외출 시 행선지를 부모가 알고 있는지 여부	20.0%	1
비행력 횡수	23.32%	2
일주일 이상 무단결석을 처음 시작한 시기	23.9%	2
가출한 횡수	21.7%	1
재원 시 면회 횡수	21.0%	1
취업 여부	21.0%	1
출원 나이	33.8%	3
사후지도건수	21.1%	1

※ 별도의 가중치가 부여되지 않는 경우 기저 가중치(base weight) '1'이 부여되었음.

표 4. 단순로지스틱법에 의한 가중치 부여

변인	B	S.E.	승산비 (Exp(B))
외출시 행선지를 부모가 알고 있는지 여부	.56***	.17	1.76
비행력 횡수	.92***	.16	2.52
일주일 이상 무단결석을 처음 시작한 시기	.93***	.15	2.55
가출한 횡수	.62***	.16	1.88
재원 시 면회 횡수	.67***	.16	1.96
취업 여부	1.16***	.25	3.19
출원 나이	1.31***	.14	3.73
사후지도건수	.73***	.17	2.08

\*\*\*  $p < .001$

하였다. 단순로지스틱법의 경우, 특정 예측변인과 다른 예측변인들 사이의 공변량을 제외하지 않은 상태에서 그 예측변인과 종속변수 사이의 공변량을 기준으로 가중치를 계산한다. 즉, 이 방법에 의한 가중치는 다른 예측변인들의 값을 통제하지 않는 상태에서 특정 예측변인의 값이 '0'에서 '1'로 변할 때 재범위험성의 변화를 나타낸다. 표 4는 가중치로 사용된 각 변인들의 단순로지틱 승산비를 나타내고 있다. 최종점수의 범위는 0에서 15.94에 이르렀고 평균은 1.67( $SD=$ .31)이었다.

#### 다중로지스틱법

다중로지스틱법은 특정 예측변인과 다른 예측변인들 사이의 공변량을 제외한 상태에서 그 예측변인과 종속변인 사이의 고유한 공변량만을 기준으로 가중치(승산비)를 설정한다. 즉, 이때의 가중치는 다른 예측변인들의 값을

모두 통제된 상태에서 특정 예측변인의 값이 '0'에서 '1'로 변할 때의 재범위험성 증가의 정도를 의미한다. 다중로지스틱 분석결과 도출된 모형은 표 5와 같다. 이 모형의 카이자 승값은 199.24( $df=8, p<.$ 001)이고 Pseudo- $R^2$  값은 .15였다. 각 변인들의 승산비를 가중치로 사용하여 최종진단점수를 계산한 결과 범위는 0에서 17.04에 이르고, 평균은 9.94( $SD=$  3.32)였다.

#### ROC 분석

자료 분석의 마지막 단계로 각기 다른 방법으로 도출된 최종진단점수들의 재범예측능력을 ROC분석을 통하여 비교·평가하였다. ROC 분석은 민감도(sensitivity)와 특이도(specificity)의 개념을 바탕으로 분류 또는 예측모형을 평가하기 위해 빈번하게 활용되는 통계기법이다 (Pundir & Amala, 2014). 본 연구의 경우 특정 최종진단점수가 실제 재범을 저지른 소년들을

표 5. 다중로지스틱법에 의한 가중치 부여

변인	B	S.E.	승산비 (Exp(B))
외출 시 행선지를 부모가 알고 있는지 여부	.48**	.18	1.62
비행력 횟수	1.03***	.18	2.81
일주일 이상 무단결석을 처음 시작한 시기	.48**	.16	1.62
가출한 횟수	.51**	.16	1.67
재원 시 면회 횟수	.61***	.17	1.85
취업 여부	.77**	.26	2.17
출원 나이	1.28***	.16	3.61
사후지도건수	.53**	.18	1.69

\*\* $p < .01$  \*\*\* $p < .001$

재범자로 예측해내는 정도를 민감도, 재범을 저지르지 않은 소년들을 비재범자로 예측해내는 정도를 특이도라고 할 수 있다. 즉 민감도는 참양성률(true positive rate)을 의미하고 특이도는 참음성률(true negative rate)을 의미한다. ROC 분석은 참양성률(민감도)을 Y축에, 허위음성률<sup>11)</sup>을 X축에 그래프로 표시한 ROC curve를 도출해낸다. 이 ROC curve 아래의 면적(area under the ROC curve; AUC)이 특정 최종진단점수의 예측정확도 또는 그 효과크기(effect size)를 의미하는 지표이다. AUC 값이 1에 가까울수록 참양성률이 높아지고 허위음성률은 낮아지기 때문에 해당 예측모형의 예측정확도가 높다고 할 수 있다. 명확한 기준은 없으나 노일석(2010)은 서구의 한 연구(Eher, Rettenberger, Schilling, & Pfafflin, 2008)를 인용하며 AUC 값이 .71 이상이면 예측정확도가 높고, .63 이하이면 낮으며, 그 사이의 값은 보통의 예측정

11) 허위음성률은 1-특이도를 의미한다.

확도를 갖는다고 하였다.

전술한 바와 같이 기존 국내 연구들은 평가 도구의 적합성을 최초 표본 외에 추가 표본을 대상으로 검증하는 절차를 거치지 않았다. 본 연구에서는 부트스트랩(bootstrapping) 기법을 통해 개발된 계리적 도구의 예측 정확도에 대한 내적 검증을 실시하였다. 부트스트랩 기법은 전통적 유의성 검정의 한계를 극복하기 위해 Efron(1979)에 의해 소개된 비모수적 통계 추론(nonparametric statistical inference) 방법이다. 기존의 유의성 검정 방법은 통상적으로 모수적 추론(parametric inference)으로서 특정 추정치가 일정한 확률분포를 따를 것이라고 가정하고 그 분포의 모수에 대한 검정을 실시하는 것이다. 따라서 추정치가 특정한 확률분포를 따른다고 가정할 수 없는 경우에는 모수적 추론방법이 적합하지 않다. 또한 모수적 추론의 경우, 표본의 크기가 특정 수준 이상 커지게 되면 거의 모든 통계적 검정이 유의하게 나타날 수 있다는 단점을 지니고 있다. 이러한 단점들을 극복할 수 있는 부트스트랩 기법은 표본 자료로부터 그 표본과 크기가 동일한 표본을 통상 1,000회 이상 복원추출(resampling with replacement)하는 방법을 사용하여 추정치를 반복 추정하게 된다. 이 과정을 통해 1,000개 이상의 추정치가 반복 생성되므로 이 추정치들 자체가 경험적 분포(empirical distribution)를 형성하게 되어 그 추정치가 특정한 확률분포를 따를 것이라고 가정할 필요가 없어진다. 따라서 부트스트랩 기법은 정상분포를 가정하여 표준오차를 도출할 필요 없이 추정치의 경험적 분포로부터 표준오차를 직접 도출해 낼 수 있다(Thompson, 1993). 이러한 부트스트랩 기법은 표본을 몇 개의 소표본으로 나누어 평가도구를 검증하는 방법보다 훨씬 우수한 내적 검

증(internal validation) 기법으로 평가받고 있다 (Steyerberg, 2009).

ROC 분석 결과 가중치 부여 방법에 따라 산출된 AUC값, 표준편차, 그리고 95% 신뢰구간이 표 6에 제시되었다. 표 6의 왼쪽 부분 값은 부트스트랩 기법을 사용하지 않고 얻은 값이며 오른쪽 부분은 부트스트랩 기법을 사용하여 얻은 값이다. 먼저 왼쪽 부분의 AUC 값을 보면 다중로지스틱법(.766)이 가장 높은 예측정확도를 보이고, 너필드법이 다음으로 높은 값(.759)을 보였으며, 그 뒤를 원점수법(.748)과 단순로지스틱법(.723) 따랐다. 결과적으로 요인분석법(.487)은 가장 낮은 예측정확도를 보였다. 요인분석법을 제외하고는 모두 .71을 상회하는 AUC 값을 보여 노일석(2010)의 기준에 따르면 모두 예측정확도가 높다고 할 수 있다.

그러나 표 6의 오른쪽에 부트스트랩 기법을 사용하여 얻은 AUC값을 살펴보면 전반적으로 그 값들이 눈에 띄게 낮아진 점을 확인할 수 있다. .71을 넘는 값은 오로지 너필드법(.713)과 다중로지스틱법(.757) 뿐이었다. 이러한 사실은 부트스트랩 기법이 평가도구를 검증하는 방법으로서 타당도가 우수한 검증기법이라는

점(Steyerberg, 2009)을 감안할 때, 기존의 국내 연구들이 제시한 재범위험성 평가도구모형들은 해당 모형을 산출하기 위해 사용된 연구표본에 과적합(overfit)되어 실제 예측정확도보다 더 높은 추정치를 제시하였을 가능성을 시사한다. 마지막으로 .71을 넘는 다중로지스틱법 AUC값(.757)과 너필드법 AUC값(.713)의 차이가 통계적으로 유의한지 여부를 조사한 바, 그 차이는 유의한 것으로 나타났다( $z=8.7, p < .001$ ). 각 기법에 따른 AUC 값의 차이에 대한 시각적 이해를 돕기 위해서 ROC 곡선과 AUC 값을 그림 1에 제시하였다.

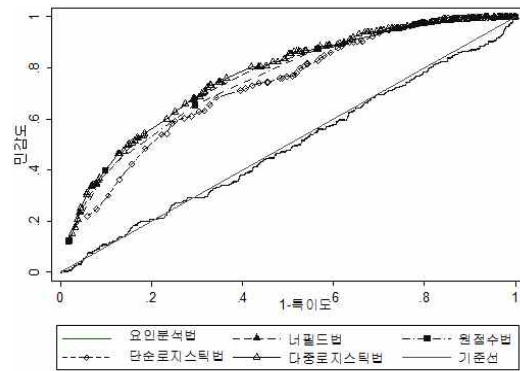


그림 1. 각 가중치 모델의 ROC곡선 및 AUC값

표 6. 다섯 가지 가중치부여 방법에 따른 재범예측 정확도 비교

구분	ROC분석			부트스트랩 ROC 분석		
	AUC	S.E.	95% CI	AUC	S.E.	95% CI
요인분석법	.487	.02	.44 ~ .53	.487	.02	.45 ~ .53
원점수법	.748	.02	.72 ~ .78	.668	.02	.63 ~ .70
너필드법	.759	.02	.73 ~ .79	.713	.02	.68 ~ .75
단순로지스틱법	.723	.02	.69 ~ .76	.708	.02	.68 ~ .74
다중로지스틱법	.766	.02	.73 ~ .80	.757	.02	.73 ~ .79

※ AUC값의 차이를 세밀하게 비교하기 위하여 소수점 이하 세 자리로 표시하였음.

## 논 의

최근 들어 우리 사회의 전체 범죄량은 감소하고 있는 추세이나 소년범죄의 감소율은 상대적으로 미흡하며 특히 소년범죄자의 재범률은 오히려 증가하는 경향을 보이고 있다. 이러한 현상에 대해 소년교정기관이 주먹구구식 교정 처우와 프로그램을 시행해오며 소년범의 위험성 - 욕구 - 반응성(risk-need-responsivity)을 충분히 고려하지 않았기 때문이라는 비판과 함께 관계자들의 자성의 소리 또한 높아가고 있다(윤옥경, 2011). 이런 배경 속에서 소년범의 재범위험성에 따른 차별적이고 선택적인 교정 처우를 목적으로 재범예측도구 개발요구가 증가하였고, 이에 부응하여 소년 보호관찰 대상자 및 소년원에 재원 중인 소년범들을 대상으로 재범위험성에 대한 계리적 평가도구를 개발하고자 하는 노력이 있어 왔다(김양곤, 이수정, 이민식, 2005; 노일석, 2007; 2009; 2010; 서석교, 2011). 본 연구는 이와 같은 계리적 도구를 개발함에 있어서 논의의 필요성이 절실함에도 불구하고 그동안 학계 내에서 적절한 관심을 받지 못했던 가중치 선정 방법에 대한 실증적 검증을 실시하고자 하였다.

본 연구에서는 소년원에서 출원한 남자 소년범 1,394명의 자료를 사용하여 다섯 가지 다른 방법으로 가중치를 산출한 후 ROC 분석을 통해 각각의 예측정확도를 비교하여 보았다. 그 결과 요인분석법을 활용한 계리적 예측도구의 개발은 추천할 바가 되지 못한 것으로 판단되었다. 이 방법은 그 절차가 상당히 복잡하여 실무자들이 활용하기가 곤란하며 또한 표 6에 나타난 바와 같이 부트스트랩 기법 여부에 상관없이 AUC값이 현저히 낮았기 때문이다. 원점수법과 단순로지스틱법의 경우,

연구표본만을 사용하여 최종진단점수를 산출한 경우에는 그 예측정확도가 적절한 것으로 나타났다. 그러나 부트스트랩 기법을 사용하여 내적 검증절차를 거친 결과 AUC값이 저하되어 역시 추천할만하지 않았다. 이런 점에서 보건대 본 연구의 결과는 가중치 없이 원점수만을 사용할 것을 권장한 Grann과 Langstrom (2007)의 스웨덴 연구 결과와는 부합하지 않았다. 물론 사회·문화별로 재범예측 변인들의 영향력 또한 다를 것이기 때문에 위와 같은 상이한 결론이 특이하다고 할 수는 없을 것이다.

부트스트랩 기법을 활용한 ROC 분석 결과, 재범 예측에 대해 적절한 예측력을 발휘할 수 있을 것으로 확인된 방법은 오직 너필드법과 다중로지스틱법 두 가지였다. 그 중에서도 다중로지스틱법의 AUC가 더 높아 예측력이 가장 높은 것으로 확인되었다. 그러나 계리적 도구의 효용가치는 예측정확도 외에 투명성(transparency)도 고려되어야 한다(Grann & Langstrom, 2007). 즉, 예측모델이 예측에 실패했을 때 그 예측과정을 역으로 되짚어가면서 과연 어느 부분에서 착오가 발생했는지를 확인할 수 있을 정도로 모형의 구성이 쉬워야 한다. 투명성의 관점에서 본다면 가중치 계산 방법이 가장 단순한 너필드법이 그 효용 가치가 높다고 할 것이다. 특히 회귀분석에 익숙하지 않은 교정기관 실무자들에게는 너필드법이 재범 예측에 있어 가장 적절한 방법이 될 수도 있을 것이다.

본 연구의 분석 결과를 종합해보면, 다중로지스틱 회귀분석 결과 산출된 승산비를 가중치로 활용한 경우 소년원 출원생의 재범위험성을 가장 정확하게 예측할 수 있었다. 너필드법에 의해 산출된 가중치 활용은 그 방법이

용이하고 예측력 또한 비교적 높은 편이어서 특히 소년교정 실무자들이 사용하기에 적절할 것으로 판단되었다. 그 외 요인분석법, 원점수법, 단순로지스틱법으로 산출된 가중치들은 예측력이 낮은 편이어서 재범위험성 평가를 위한 계리적 도구의 개발에 있어서 효용가치는 다소 떨어질 것이라고 예상할 수 있다.

본 연구는 그동안 연구자들 사이에 논의되지 않았던 가중치활용 문제를 다룸으로써 재범위험성 평가를 위한 계리적 도구의 개발을 시도하는 후속 연구자들에게 가중치 사용에 관한 이론적·경험적 함의를 제공하고자 하였다. 또한 본 연구는 부트스트랩 기법을 활용하여 계리적 도구의 예측정확도에 대해 내적검증을 시도하였고, 그 결과 연구표본으로부터 도출된 예측정확도(AUC)는 과적합화 현상으로 인해 과대추정되기 쉽다는 결론을 도출하였다는 점에서 적지 않은 의미를 부여할 수 있다. 따라서 후속 연구자들에게도 도출된 추정치에 대한 검증 절차를 거치는 것이 바람직하다는 점을 강조하고 싶다. 본 연구에서는 최초 사용된 표본자료를 복원추출(resampling)하는 내적 검증 절차를 거쳤으나 후속연구에서는 사용된 표본과는 별도의 모집단으로부터 추출된 독립된 표본을 사용하여 외적타당도(external validity) 또한 검증할 필요가 있다.<sup>12)</sup>

최근 소년교정분야에서 계리적 평가도구의 개발에 대한 노력은 그 어느 때보다 더 활발한 편이다. 왜냐하면, 첫째, 예측력이 우수한

재범위험성 평가도구를 개발하면, 소년범죄자의 위험성과 욕구에 상응하는 교정처우와 프로그램을 시행하여 궁극적으로 소년범죄자의 재범률을 낮출 수 있고, 둘째, 과거 소년법에 대한 처분 또는 처우 결정이 수리적·객관적 기준보다는 평가자의 자의적 판단에 의한 경우가 많아 그 공정성을 담보할 수 없었다는 비판에서 비교적 자유로워질 수 있기 때문이다(서석교, 2011; Krysik & LeCroy, 2002). 이러한 이유로 서구 선진국에서는 십 수 년 전부터 이미 계리적 도구의 개발에 앞장섰고, 미국의 경우 이러한 도구의 활용을 법으로 강제하기도 하였다(Grann & Langstrom, 2007). 이 외에도 계리적 도구의 개발과 활용은 소년교정기관에게 추가적인 이점을 제공할 수 있다. 수리적이고 객관적인 도구의 활용은 21세기의 첨단과학지향이라는 시대 사조에 부합할 뿐더러 만약 소년범죄자의 재범률이 증가하여 교정기관에 대해 비난이 쏟아질 경우 관계자들은 이와 같은 계리적·과학적 도구에 의한 교정 처우를 근거로 그 비난을 일부 회피할 수 있기 때문이다.

예측정확도가 높은 계리적 도구의 개발에 대한 노력은 물론 바람직하다. 그러나 평가도구의 효용성에 대한 지나친 낙관과 의존은 경계되어야만 한다. 그 이유로는 먼저 대부분의 평가도구가 교정기관들이 '얻기 쉬운' 자료들만을 위주로 하여 개발된다는 점이다. 본 연구를 포함한 대부분의 기존 연구들은 소년범의 분류심사 시 획득할 수 있는 인구나사회학적 특성, 인성, 비행력, 전과관련 변인들을 사용하며, 추가적으로는 소년원 내에서의 교정성적, 징계여부, 수용기간 등의 자료를 활용한다. 그러나 출원 후 재범여부는 이와 같은 출원 전 단계의 변인들에 의해 영향을 받기도 하지

12) 한편, 서구의 연구에 의하면 부트스트랩 기법에 의해 도출된 추정치와 독립된 표본으로부터 도출된 추정치는 아주 유사하며 신뢰도면에서는 부트스트랩 기법이 오히려 더 우수한 것으로 판명되기도 했다(Steyerberg, Bleeker, Moll, Grobbee, & Moons, 2003).



만 오히려 출원 후의 일상생활과 관련된 변인들에 의해 더 큰 영향을 받을 수 있다. 이러한 점을 강조하여 Laub과 Sampson (2001)은 아무리 개인의 과거가 비행성향과 비행행위로 점철되어 있다고 할지라도 미래에 발생할 수 있는 취업, 연인, 멘토, 심지어는 심리적으로 밑바닥을 경험하는 등의 다양한 전환점(turning point)들이 결국 탈비행(desistance)을 이끌어 낼 수 있다는 점을 강조한다. Maruna(2001) 역시 전과자들 중 과거 범죄경험이나 기존의 비행성향과는 독립적으로 어느 순간 다양한 통로를 통해 일종의 '정체성의 변화'를 경험하게 되고 이런 변화가 결국 탈비행을 이끌게 된다고 주장한다. 따라서 소년원을 출원한 후 개인들이 겪게 되는 외적, 내적 경험과 관련된 변인들이 고려되지 않은 위험성 평가도구는 결국 그 예측력에 한계가 있을 수밖에 없다. 본 연구에서 최종 다중로지스틱 모형의 Pseudo  $R^2$ 값이 .15에 불과했고 다른 국내 연구들이 검토한 모형들도 그 설명력이 결코 높지 않다는 사실은 이와 같은 한계점을 잘 나타내고 있다. 후속 연구에서는 소년원 출원 전 변인 외에도 출원 후 변인들까지 고려하는 노력이 필요할 것이다.

계리적 도구에 대한 지나친 의존을 경계하는 또 하나의 이유는 수리적 측정방법의 내재적 한계와 관련 있다. 계리적 도구는 본질적으로 수치화가 가능한 요소들만을 활용한다. 그러나 소년범죄자의 탈비행에 영향을 미칠 수 있는 요인들 중에는 계량화하기 곤란한 요소들이 반드시 존재한다. 일례로 소년원 담임 교사의 헌신, 열정, 관심 또는 특정 재원생과의 관계의 질이 재범 예방에 기여한 경우, 이러한 요소들을 정확하게 측정하기는 쉽지 않다. 2013년부터 소년원 담임교사들은 재범을

방지하고 출원생들의 사회정착을 돕고자 그들을 직접방문하거나 전화, SNS 등을 통한 사후 지도에 힘쓰고 있다.<sup>13)</sup> 그러나 이러한 희망도우미 활동은 TEAMS내에 그 빈도만이 기록되고 있을 뿐이다. 그러나 만약 희망도우미 활동이 재범 방지에 기여하였다면 이는 만남의 횟수가 많아서라기보다는 그 관계의 질이 높았기 때문이라고 쉽게 예상할 수 있다.<sup>14)</sup> 계리적 도구는 이러한 정성적인 요소들을 정확하게 측정·활용함에 있어서는 태생적인 한계를 지닐 수밖에 없다. 따라서 실무에서는 계리적 도구의 한계를 인식하고 소년범에 대한 교정 처우, 훈련, 프로그램의 제공에 있어 일종의 참고적 도구로 활용함이 바람직할 것으로 보인다.

## 참고문헌

- 강동욱 (2012). 소년사범과 소년범 처리실태에 관한 고찰. *비교법연구*, 12(2), 9-48.
- 권해수, 이연상 (2013). 탈비행에 성공원 소년원 출원생의 생활 실태 분석. *청소년시설환경*, 11(3), 125-136.
- 김덕준 (2003). 지역개발격차의 측정 및 원인 분석에 관한 연구. *행정논총*, 41(4), 279-308.
- 김양곤, 이수정, 이민식 (2005). 소년보호관찰대

13) 실무상 이러한 활동을 “희망도우미”제도로라고 한다.

14) 본 연구에서는 사후지도 빈도수만을 활용했음에도 불구하고 재범방지에 유의한 영향을 미치는 것으로 나타났다. 이는 희망도우미제도의 긍정적 효과를 암시하는 바, 이 제도의 효과성에 대한 후속 연구가 필요할 것이다.

- 상자에 대한 분류평가도구 개발에 관한 연구. 서울: 한국형사정책연구원.
- 김태일 (1999). 요인분석을 활용한 종합점수화 기법 II. 정책분석평가학회보, 9(2), 185-200.
- 김태일 (2000). 수리적 기법에 의한 평가모형 체계의 가중치 부여방식에 관한 논의. 한국행정학보, 33(4), 243-258.
- 노일석 (2007). 성폭력 보호관찰청소년 재범위험성 예측도구 개발: 이종, 동종을 포함한 일반재범위험성을 중심으로. 연세대학교 대학원 석사학위논문.
- 노일석 (2009). 청소년 절도사범 재범 예측 요인: 절도 소년보호관찰대상자 재범위험성 평가도구(LJP-RRAR) 개발 연구. 한국심리학회: 일반, 28(2), 449-470.
- 노일석 (2010). 남자 청소년 폭력사범 재범 예측 요인: 남자 폭력소년 보호관찰대상자 재범위험성 평가도구(MVJP-RRAR) 개발 연구. 형사정책연구, 21(3), 350-379.
- 노일석, 정진경 (2009). 여자 청소년 폭력사범 재범 예측 요인: 여자 폭력소년 보호관찰대상자 재범위험성 평가도구(FVJP-PRAR) 개발 연구. 한국심리학회: 여성, 14(3), 365-386.
- 박성복 (1996). 지역발전격차의 분석 시론. 한국행정논집, 8(2), 1-19.
- 박영규 (2011). 소년원 송치처분의 개선방안. 교정연구, 54, 53-84.
- 서석교 (2011). TEAMS 비행위험성 진단체계 평가기준 개발에 관한 연구. 소년보호연구, 16, 107-164.
- 윤옥경 (2011). 한국 소년범죄의 추세와 소년 보호관찰의 정책방향. 소년보호연구, 17, 35-57.
- 황지태 (2009). 공식통계와 범죄피해조사의 상반된 결과에 대한 설명. 형사정책연구, 20(1), 279-303.
- Bonta, J., Harman, W., Hann, R., & Cormier, R. (1996). The prediction of recidivism among federally sentenced offenders: A revalidation of the SIR scale. *Canadian Journal of Criminology*, 38, 61-79.
- Cattell, R. (1958). Extracting the correct number of factors in factor analysis. *Educational and Psychological Measurement*, 18, 791-837.
- Cohen, J. & Cohen, P. (1983). *Applied Multiple Regression/correlation Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum.
- Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, 45, 1304-1312.
- Duwe, G., & Freske, P. (2012). Using logistic regression modeling to predict sexual recidivism: The Minnesota sex offender screening tool-3 (MnSOST-3). *Sexual Abuse: A Journal of Research and Treatment*, 24(4), 350-377.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7, 1-26.
- Eher, R., Rettenberger, M., Schilling, F., & Pfafflin, F. (2008). Failure of Static-99 and SORAG to predict relevant re-offense categories in relevant sexual offender subtypes: A prospective study. *Sexual Offender Treatment*, 8(1), 1-20.
- Grann, M., & Langstrom, N. (2007). Actuarial assessment of violence risk: To weigh or not to weigh? *Criminal Justice and Behavior*, 34(1), 22-36.
- Krasyk, J., & LeCroy, C. (2002). The empirical validation of an instrument to predict risk of

- recidivism among juvenile offenders. *Research on Social Work Practice, 12*(1), 71-81.
- Laub, J., & Sampson, R. (2001). Understanding desistance from crime. *Crime and Justice, 28*, 1-69.
- Maruna, S., 2001. *Making Good: How Ex-offenders Reform and Reclaim Their lives*. Washington, DC.: American Psychological Association Books.
- Nuffield, J. (1982). *Parole Decision-making in Canada: Research Towards Decision Guidelines*. Ottawa: Solicitor General of Canada.
- Pundir, S., & Amala, R. (2014). Evaluation of area under the constant shape bi-Weibull ROC curve. *Journal of Modern Applied Statistical Methods, 13*(1), 20.
- Quinsey, V., Rice, M., & Harris, C. (1995). Actuarial prediction of sexual recidivism.. *Journal of Interpersonal Violence, 10*, 85-105.
- Spitzer, S. (1975). Punishment and social organization: A study of Durkheim's theory of penal evolution. *Law and Society Review, 9*, 613-637.
- StataCorp. (2013). *Stata Statistical Software: Release 13*. Stata Corporation: College Station, TX.
- Steyerberg, E. (2009). *Clinical prediction models: A Practical Approach to Development, Validation, and Updating*. New York, NY: Springer.
- Steyerberg, E., Bleeker, S., Moll, H., Grobbee, D., & Moons, K. (2003). Internal and external validation of predictive models: A simulation study of bias and precision in small samples. *Journal of Clinical Epidemiology, 56*(5), 441-447.
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using Multivariate Statistics*. Boston: Allyn and Bacon.
- Thompson, B. (1993). The use of statistical significance tests in research: Bootstrap and other alternatives. *Journal of Experimental Education, 61*, 361-377.
- Van Dijk, J. (2006). *International Comparisons of Crime: The ICVS and More*. Presented at the British Crime Survey 25th Anniversary Conference at Cumberland Lodge. England, 16-17 October, 2006.
- Wolfgang, M., Figlio, R., & Sellin, T. (1972). *Delinquency in a Birth Cohort*. Chicago, IL: The University of Chicago Press.
- 1차원고접수 : 2015. 09. 30.  
수정원고접수 : 2015. 11. 18.  
최종게재결정 : 2015. 12. 03.

## Assessment of Actuarial Tools for the Prediction of Recidivism among Incarcerated Juvenile Offenders

**Kwon Haesoo**

**Yun Ilhong**

**Kim Hyunjung**

Chosun University

Ministry of Justice

There exists a high degree of interest in actuarial tools that may help predict the recidivism of juvenile offenders. Yet, it has rarely been a focus of debate among scholars whether to use weights and what kinds of weights should be employed when developing such a tool. The current study fills this research gap. Specifically, the present study examined and compared the predictive accuracy of five different actuarial tools created by different methods of calculating weights. The results revealed that weights calculated by multivariate logistic regression and the Nuffield method proved to be of the highest predictive efficacy. Based on the results, we offer policy and research implications for both practitioners and future researchers. In particular, we underscore the need to incorporate variables pertaining to the routine activities of the offender after release as well as qualitative variables that are not easily quantified if a researcher wishes to develop an actuarial tool with a high predictive efficacy.

*Key words* : *recidivism risk, actuarial tool, juvenile detention center, bootstrapping, ROC analysis*