

전문가판단모델(SPJ) 기반 폭력위험성평가(HCR-20) 버전 3 타당화 연구: 재범예측에서 예방으로의 패러다임 전환

서종한* · 조병철** · 김희송*** · 김경일****

국 | 문 | 요 | 약

본 연구의 목적은 국내에 HCR-20 버전 3을 소개하고 그 활용가능성을 탐색하고자 하였다. 이를 위해 국내 수용기관에 수용중인 80명의 재소자를 대상으로 표집이 이루어졌다. 이를 대상으로 HCR-20 버전 3의 채점자간 신뢰도와 공인타당도를 알아보았다. 표본의 대표성을 감안해 각 지역별 수용인원수에 따라 차등적으로 전국 수용시설의 참가자를 모집하였고 그 결과 총 85명의 샘플이 이루어졌지만 최종적으로 수집 자료의 신뢰도를 따져 5건을 제외한 80건을 최종 분석 대상으로 정하였다. 주요결과, 전체적으로 HCR-20의 3개 위험요인과 전문가 판단 항목이 CAPP, KORAS-G, LSI-R, PCL-R 등과 유의미한 수준에서 강한 상관을 보였다. 채점자간 신뢰도 또한 높은 일치성(.58-.91)을 보였는데 상위 3개 위험요인 뿐 아니라 최종 전문가 판단 내 5개 하위영역에서도 비교적 높은 일치성(.60-.99)을 보였다. 이런 결과는 향후 HCR-20 버전 3의 국내 활용과 추가적인 연구에 중요한 디딤돌이 될 것으로 보인다. HCR-20 버전 3에 대한 시사점 등을 추가적으로 논의하였다.

❖ 주제어 : 근거기반, 폭력, 전문가판단모델, HCR-20 버전 3

한국판 HCR-20 Version 3에 대한 국내 저작권[©]은 원저자들과의 계약을 통해 서종한, 김희송(국과수대표)에게 있음. 본 연구에 도움을 주신 Stephen D. Hart, Kevin S. Douglas, 전국의 경찰청 프러파일러와 법무부 교정심리학자, 그리고 연구원에 감사를 전함.

* 사이먼프레이저대학교 심리학과 정신건강법정책연구소 수석연구원

** 아주대학교 일반대학원 심리학과 연구원

*** 국립과학수사연구원 법심리과 과장

**** 교신저자: 아주대학교 심리학과 교수

I. 서론

폭력위험성을 평가하는 HCR-20(Historical-Clinical-Risk Management-20; Douglas et al., 2013)은 아직 한국에서는 타당화가 이루어지지 않은 대표적인 위험성 평가 중 하나이다. Webster 등(1993)이 주도하여 캐나다 밴쿠버에서 개발된 4세대 HCR-20 위험성 평가는 국외에서 가석방 심사, 교정 내 처우 분류, 치료감호소, 보호관찰, 우범자관리, 보호관찰, 사법기관의 폭력위험성 평가 등에서 폭넓게 사용하고 있는 도구이며 이와 직·간접적으로 관련된 저서와 문헌만 해도 500편을 넘고 있다(Douglas et al., 2013; 최이문 등, 2014; 서종한 등, 2017). 주로 북미의 캐나다와 미국, 북유럽 대다수의 국가, 아시아 중 싱가포르, 홍콩, 중국, 일본 등의 법무부, 경찰국, 교정국과 법무치료감호소 등 국가 공공기관에서 폭넓게 활용 중에 있다. 국내의 경우 HCR-20의 사용은 거의 이루어지지 않고 있으며 일부 연구자가 연구 목적으로 사용하고 있는 추세이다(최금화, 2010; 정지숙 등, 2012; 이수정, 2010)¹⁾. HCR-20는 소위 [보험] 계리적 도구(actuarial instrument; 자세한 내용은 서종한 등, 2017참고)가 아닌 구조화된 전문가 판단 모델(structured professional judgement model; 이하 SPJ; 자세한 내용은 서종한 등, 2017 참고)을 바탕으로 만들어졌다. 따

- 1) Douglas, K. S., & Belfrage, H. (2014). Interrater reliability and concurrent validity of the HCR-20 Version 3. *International Journal of Forensic Mental Health*, 13(2), 130-139.; 최이문, 강태경, & 조은경. (2014). 미국 법원에서 심리학 도구 (사이코패시 체크리스트, PCL-R) 의 역할에 대한 연구 (2005~ 2012). *형사정책연구*, 100(단일호), 375-414.; Douglas, K. S., & Webster, C. D. (1999). The HCR-20 violence risk assessment scheme: Concurrent validity in a sample of incarcerated offenders. *Criminal Justice and Behavior*, 26, 3-19.; Douglas, K. S., Webster, C. D., Eaves, D., Wintrup, A., & Hart, S. D. (1996, March). *A new scheme for the assessment of dangerousness and the prediction of violence*. Paper presented at the Biennial Meeting of the American Psychology-Law Society (APA Div. 41), Hilton Head, SC.; Douglas, K. S., Webster, C. D., Hart, S. D., Eaves, D., & Ogloff, J. R. P. (Eds.) (2001). *HCR-20: Violence risk management companion guide*. Burnaby, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University.; Douglas, K. S., Webster, C. D., & Wintrup, A. (1996, August). *The HCR-20 risk assessment scheme: Psychometric properties in two samples*. Poster session presented at the Annual Meeting of the American Psychological Association, Toronto, ON.; 정지숙·신정·이장규. (2012). 재범위험성 평가도구 타당도 검증. *한국심리학회지: 일반*, 31(3), 643-665.; 최금화 (2009). 폭력 위험성 측정 도구 (Violence Risk Assessment: HCR-20)의 타당도 연구. *경기대학교 대학원 석사 학위 논문*).

라서 이 도구의 근본 취지는 규범정립적(nomothetic) 접근에서 벗어나 임상가가 개별기술적(idiographic)인 방식에서 폭력 위험성을 평가 하고 이를 바탕으로 사례를 개념화하여 향후 펼쳐질 수 있는 위험성 시나리오(scenario)를 발견하는데 도움을 주는데 있다. 더 나아가 가능한 위험성 시나리오 탐색을 통해 향후 관리계획(management plan)과 전문가의 개입(expert opinion)에 중요한 단서를 제공하고자 하는데 주요 목적이 있다.

HCR-20은 국내에서 개발되어 활용중인 HAGSOR(조은경, 2012), KORAS-G(이수정, 고려진, 2011), KSORAS(이수정, 고려진, 최혜림, 2010), PCL-R(Hare, 2003; 한국판, 조은경, 이수정, 2008) 등과 같은 도구와는 달리 위험성 등급을 구분 지을 수 있는 산술적 알고리즘(algorithm)은 전혀 존재하지 않는다. 따라서 원저자들은 HCR-20을 계리적 도구처럼 각 위험요인을 점수를 합산하여 등급을 분류하고자하는 것은 본 도구의 근본적인 목적을 훼손하는 것이라 지적하였다.²⁾ 본 연구의 일차적 목적은 아직 국내에 소개되지 않은 HCR-20 버전 3의 핵심적인 전문가판단모델과 그 주변 개념을 상세히 개관하고자 하였고 국내에서 활용할 수 있는 발판을 마련하고자 신뢰도와 타당도, 형사정책적 시사점을 찾는데 두었다.

1.1 HCR-20 버전 3의 핵심적 개념

폭력. Hart(2004/2011)가 지정한 바와 같이, 만약 폭력(violence)을 정확하게 정의할 수 없다면 무엇이 폭력을 어떻게 유발하는지 정확하게 이해를 할 수 없을 뿐 아니라 이를 예측하거나 예방할 수도 없을 것이다. 여기서는 폭력을 다른 사람에게 실제로 행사하였거나 미수에 그친 신체적 피해 또는 신체적 피해 위협이라고 정의한다. 이 정의는 HCR-20 Version 1과 Version 2에서의 정의 및 세계보건기구

2) 윤정숙, 이수정 (2012). 성범죄자 치료 프로그램의 주요 원칙과 방향 정립을 위한 이론적 고찰. 한국심리학회지: 일반, 31(3), 847-875.; 이수정 (2010). 최신 범죄 심리학. 서울: 북카페.; 이수정·고려진 (2011), 한국 성범죄자 위험성 평가척도(KSORAS) 실시요강. 서울: 맥스메디카; 이수정·윤옥경. (2003). 범죄위험성의 평가와 활용방안. 한국심리학회지: 일반, 22(2), 99-126.; 이춘화 (2001). 소년에 대한 조사제도의 문제점과 개선방안. 소년보호연구, (3), 26-85.; 조은경 (2010). 성폭력사범 재범위험성평가척도 개발(HAGSOR). 과천: 법무부[법무부용역연구보고서].

(WHO)의 “대인폭력”의 정의와 맥락을 같이한다(Krug, Dahlberg, Mercy, Zwi, & Lozano, 2002, p.6). 이 정의에서 폭력은 여러 가지 측면을 담고 있다. 첫째, 신체 혹은 심각한 심리피해를 유발하거나 가능성이 있는 행동에 초점을 둔다. 심각한 심리피해는 신체부상에 대한 불안뿐만 아니라 건강이나 삶의 질을 상당히 해치는 정서·정신·인지적 결과를 포괄한다. 둘째, 신체 혹은 심각한 심리피해는 행위자가 피해자의 신원을 아는지 여부와 상관없이 한 명 이상의 사람에게 영향을 미쳐야 한다. 자해(즉, 자살 혹은 자기-손상) 행동은 행동의 결과로 인해 타인이 신체피해를 입을 수 있다면 폭력 범위에 들어간다(예: 살해 후 자살, 경찰관을 위협하여 발포하게 만드는 자살 방법 등). 셋째, 완결된 행위, 미완성 혹은 시작 단계인 행위(예: 시도 혹은 계획), 소극적 행위(예: 태만 혹은 회피, 적어도 행위자가 돌봄의 의무가 있는 경우), 의사소통 행위(예: 구두 혹은 서면 발언), 집단 행위(예: 공모 혹은 음모)를 포함한 다양한 행동을 폭력이라고 볼 수 있다. 넷째, 우연 혹은 반사적인 것보다는 분명한 목적성이 있어야 한다. 적어도 부분적으로 신체피해 혹은 심각한 심리피해를 입히는 것(즉, 의도적으로 혹은 계획적으로) 또는 자신의 행동이 타인에게 신체피해나 심각한 심리피해를 유발할 만 한 가능성이 있다는 것을 알고(즉, 무모하게) 특정 행동을 선택한 것이어야 한다. 마지막으로, 허가되지 않는 행동이다. 이는 자신의 행동으로 인해 영향을 받을 수 있는 타인에게 동의를 받지 않았고 어떤 법적 허가도 받지 않았음을 의미한다. 따라서 의도적으로 타인에게 신체적 피해를 입히지만 적절한(즉, 완전한, 통보된, 자발적인, 법적인) 동의와 함께 행해진 행동(예: 수술, 종합 이종 격투기)은 폭력이 아님을 의미한다.³⁾

위험성. 위험성(risk)은 완전히 이해하기 어려운 위협 혹은 위험이기 때문에 이를 예측하는 데 불확실성이 동반될 수밖에 없다(Bernstein, 1996). 위험성은 일종의 다면적 개념으로, 위험유형, 위험 발생 가능성, 위험 빈도나 지속기간, 위험 결과의 심각성, 위험 임박성과 관련이 있다. 또한 위험이 특정 상황에서 발생하고 존재하기

3) Hart, S. D. (2011). Complexity, uncertainty, and the reconceptualization of violence risk assessment. In R. Abrunhosa (Ed.), *Victims and offenders: Chapters on psychology and law* (pp. 57-69). Brussels: Politeia. (Original work published in 2004.); Krug, E. G., Dahlberg, L. L., Mercy, J. A., Zwi, A. B., & Lozano, R. (Eds.). (2002). *World report on violence and health*. Geneva: World Health Organization.

때문에 위험성의 개념은 본질적으로 역동적이고 맥락과 관련이 있을 수밖에 없다 (Douglas & Belfrage, 2014).

평가. 평가(assessment)는 의사 결정을 위해 필요한 정보를 수집하는 과정이다 (American Educational Research Association/American Psychological Association /National Council on Measurement in Education, 1999; Segal & Coolidge, 2000). 수집정보의 형태와 해당 정보수집에 적용되는 방식은 무엇을 평가할지, 향후 어떤 결정을 내릴지 그리고 어떤 사례인지에 따라 결정될 수 있는데, 평가는 특정한 개인에 대한 고려를 넘어 상황 전체를 고려해야 된다. 따라서 폭력 위험성 평가의 경우, 폭력 예방을 위해 개인의 과거나 현재의 기능뿐 아니라 향후 목표와 계획을 평가하는 것도 필요하다(예: Hart, 2001; Heilbrun, 1997; Monahan, 1981/1995; Monahan & Steadman, 1994).⁴⁾

관리. 관리(management)는 상황을 통제하고 유지하기 위해 조치를 취하는 과정이다. HCR-20에서 폭력 위험성과 관련 있는 관리 전략은 네 가지 범주로 나뉜다(Hart, Douglas, & Webster, 2001; Kropp, Hart, Lyon, & LePard, 2002). 모니터링 (monitoring)은 감시 혹은 반복적인 평가를 포함한다. 예를 들어, 보호관찰관의 가정방문, 간호사의 정신상태 평가, 사법당국 부서의 밀착 감시, 우편물 검사와 같은 특정 전략을 나타낸다. 관리감독(supervision)은 감금, 강제치료감호, 무기압수 그리고 이동, 단체, 통신의 법적 제한 등과 같이 개인의 자유를 통제하거나 제한하는 것을 말한다. 치료(treatment)는 정신건강 혹은 약물사용 문제에 대한 약물치료와 심리치료, 고용문제에 대한 교육 및 고용 프로그램, 법적 문제에 대한 분쟁 해결 서비스와 같은 구체적인 전략 등을 포함한다. 피해자 안전 계획(victim safety planning)은 향후 폭력의 잠재적인

4) Hart, S. D. (2001). Assessing and managing violence risk. In K. S. Douglas, C. D. Webster, S. D. Hart, D. Eaves, & J. R. P. Ogloff (Eds.), *HCR-20 violence risk management companion guide* (pp. 13-25). Burnaby, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University.; Heilbrun, K. S. (1997). Prediction versus management models relevant to risk assessment: The importance of legal decision-making context. *Law and Human Behavior*, 21, 380-359.; Monahan, J. (1995). *The clinical prediction of violent behavior*. Northvale, NJ: Jason Aronson. (Original work published in 1981.); Monahan, J., & Steadman, H. J. (Eds.). (1994). *Violence and mental disorder: Developments in risk assessment*. Chicago: University of Chicago.; Segal, D. L., & Coolidge, F. L. (2000). Assessment. In A. E. Kazdin (Ed.), *Encyclopedia of psychology*, Vol. 1 (pp. 264-272). Washington, DC: American Psychological Association.

희생자라고 추정되는 사람을 보호하는 것을 말한다. 이를테면 피해자 교육 및 상담, 잠재적 피해자의 거주지 혹은 직장에서의 물리적 보안 강화와 같은 구체적인 전략이 예가 될 수 있다.⁵⁾

근거기반(Evidence-based). 근거기반(evidence-based)이란 “관찰이나 연구, 통계 또는 잘 타당화 된 이론 형태의 관련 정보를 지침 혹은 바탕으로 하여 검토해 도출한 행동이나 의사 결정”을 말한다(예: 서중한 등, 2017, p. 219; Mears, 2007; Pfeffer & Sutton, 2006; Sackett & Rosenberg, 1995). 현재 많은 분야에서 널리 사용되는 이 개념은 의학에서 대중화되었고, 고전적 정의는 “개별 환자 치료에 관한 의사결정을 할 때 현재 최적의 증거를 양심적이고 분명하게 그리고 현명하게 사용하는 것”이다(Sackett, Rosenberg, & Muir Gray, Haynes, & Richardson, 1996, pp. 71-72).⁶⁾ 근거기반 의사결정의 관습적 혹은 표준화된 개념은 자유주의파(latitudinarian)로 분류될 수 있는데(Hart, 2003a; Hart & Logan, 2011) 근거 기반 의사 결정을 지도 철학, 핵심 가치, 기대 기준으로 고려하고 있다. 이와는 반대로, Hart(2003a)가 언급한 것처럼, 일부는 훨씬 더 편협하거나 엄격한 관점의 근거 기반을 고수하고 있기도 한다(예, 정통파(orthodox)]. 예를 들어 특정 절차가 경험적 연구에서 직접적으로 생산되거나 사실임이 확인되고 입증된 경우에만 근거기반 의사 결정이라고 여긴다. 현재, 폭력 위험성의 근거기반 평가에 대해 광범위한 또는 편협한 관점을 가진 사람들 간의 논쟁은 현장에서 활발하고 치열하게 이루어지고 있는데 자유주의파적 관점과 정통파적 관점 간의 차이를 제대로 알고 있을 때 연구자간

- 5) Hart, S. D., Douglas, K. S., & Webster, C. D. (2001). Risk management using the HCR-20: A general overview focusing on historical factors. In K. S. Douglas, C. D. Webster, S. D. Hart, D. Eaves, & J. R. P. Ogloff (Eds.), *HCR-20 violence risk management companion guide* (pp. 27-40). Burnaby, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University.; Kropp, P. R., Hart, S. D., Lyon, D., & LePard, D. (2002). Managing stalkers: Coordinating treatment and supervision. In L. Sheridan & J. Boon (Eds.), *Stalking and psychosexual obsession: Psychological perspectives for prevention, policing and treatment* (pp. 138-160). Chichester, UK: Wiley.
- 6) 추상 혹은 일반 용어에서, 근거기반 의사결정 개념은 널리 알려져 있고 수용된다. 하지만 근거 기반 의사 결정의 기본적 정의에 관해서 상당한 의견 차이가 있고(예: Justice, 2008; Tanenbaum, 2005; Timmermans & Mauck, 2005), 그것의 원칙과 관행에 관한 의견 차이는 훨씬 더 확연하다(예: Atkins, Slegel, & Slutsky, 2005; Borry, Schotsmans, & Dierickx, 2006; Kemm, 2006; Miles, Polychronis, & Grey, 2006; Steinberg & Luce, 2005; Tonelli, 2006).

근거기반 위험성 평가의 방향성이 다를 수 있음을 이해할 수 있고 접근 방식이 다소 다름을 알 수 있게 된다.

이를 모두 통합해 보자면, ‘폭력 위험성의 근거 기반 평가 및 관리’는 ‘사람에 대한 정보를 수집하는 절차로, 수집 방식은 가용한 최신의 과학적·전문적 지식에 따르거나 혹은 이와 일치하는 방식으로 이루어지며 최종적으로는 향후 잠재적 폭력 가능성을 이해하고 이를 방지하기 위해 필요한 조치를 결정하는 것’⁷⁾으로 정의할 수 있다(Hart, 2009; Hart & Logan, 2011).

이 정의에 따르면, 폭력 위험성 평가와 관리는 단계에 따라 구분 지을 수 있다. 첫 번째 단계는 평가 대상자의 폭력 가능성을 이해하는 것이다. 이 부분은 대상자의 폭력 이력(예: 폭력 행위, 시도, 위협, 사고, 의도), 전반적인 심리사회적 적응(예: 문제, 취약성, 강점) 그리고 주거환경(예: 사회적 및 물리적 환경)의 체계적인 분석이 뒤따른다. 이 단계의 목표는 대상자가 어떤 종류의 폭력을, 누구를 상대로, 무슨 이유로, 어떤 상태에서 저지를 가능성이 있는지 이해하는 것이다. 다른 말로 예측(prediction)의 과정으로 설명될 수 있지만 예측은 폭력의 정확한 양적 확률 추정치를 내리는 것보다는 ‘대략적’으로 미래에 무슨 일이 일어날지 이해하는 의미로 사용되어야 한다(가능성을 이해하는 것과 확률적 예측은 완전히 다른 것이다). 두 번째 단계는 어떤 사건이 대상자의 향후 폭력 가능성을 증가 혹은 감소시킬 것인지 알아보는 것으로 폭력을 통제하는 요인을 밝혀내기 위해 주거환경을 체계적으로 분석해 볼 수도 있다. 이 단계의 목표는 개입에 대한 계획을 세움으로써 미래의 폭력을 예방(즉, 위험성 완화)하는데 있는데 소위 관리(management) 과정으로 일컬어질 수 있다.

1.2 HCR-20의 접근: 자유재량 과정

폭력 위험성 평가와 관리에 관한 의사결정 접근은 검토된 정보나 수집방식과는

7) the process of gathering information about people in a way that is consistent with and guided by *the best available scientific and professional knowledge* to understand their potential for engaging in violence in the future and to determine what should be done to prevent them from doing so (Hart, 2009; Hart & Logan, 2011)

무관하게 최종 결정에 도달하기 위해서 정보가 어떻게 가중되고 결합되는지에 기초해서 크게 두 가지 형태로 분류될 수 있다(예: Meehl, 1954/1996; Monahan, 1981/1995). 첫 번째, **자유재량 의사결정**으로, “임상적, 직관적인, 비공식적인, 주관적인, 인상주의적”이라고도 말한다(Grove & Meehl, 1996, p. 293). 자유재량 의사결정의 특징은 어떤 정보를 고려하고 수집할지 그리고 그 정보를 어떻게 가중하고 결합시킬 것인지를 포함한 의사결정 상당부분을 전문가 판단에 맡기는 것이다. 두 번째 접근은 **비-자유재량 의사결정**으로, “계리적인, 통계적인, 기계적인, 알고리즘적인”이라고도 한다(Grove & Meehl, 1996, p. 293). 이 접근의 특징은 평가자가 선형적으로 개발된 고정적이고 분명한 규칙에 따라서 최종 결정을 내리는 것이다. **비-자유재량 접근**은 어떤 정보를 고려하고, 어떻게 정보를 수집하며, 수집한 정보를 어떻게 가중하고 결합시키는지를 결정하기 위해서 수식을 사용하는 경우로 수식은 경험적 연구를 통해 얻은 것이고, 특정 환경, 인구, 결과를 위해 최적화되어 있다고 볼 수 있다. 본 연구에서는 우리가 타당화하려는 도구와 관련성이 있는 자유재량 과정에 국한해서만 논하도록 하겠다.⁸⁾

자유재량 과정(discretionary procedures). 자유재량 접근에는 최소한 세 개의 다른 과정으로 구분 지을 수 있다. 첫 번째는 비구조화 된 전문가 판단(unstructured professional judgement)(Hart, 1998)으로, Hanson(1998) 등은 독자적 임상 판단(unaided clinical judgement)이라고 부르기도 하였다. 이것은 구조가 완전히 없는 상황에서의 의사결정으로, 직관·경험적인 판단에 의존하는 접근이다. 역사적으로 볼 때, 과거 가장 흔하게 사용하는 방법이기 때문에 법원이나 법정에서도 매우 익숙하게 사용되어져 오고 있다. 하지만 가장 오래 적용되어 왔던 방식임에도 불구하고(예: Hart, 2001), 혐의나 광의의 정의로든 근거기반으로는 여겨질 수 없다. 왜냐면 첫째, 비구조화 된 전문가 판단이 과학적이고 전문적인 문헌의 정보를 포함하였거나 구조화된 정도를 알 수 있는 방법은 없다. 둘째, 폭력 위험성의 비구조화 된 전문

8) Meehl, P. E. (1996). *Clinical versus statistical prediction: A theoretical analysis and a review of the literature*. Northvale, NJ: Jason Aronson. (Original work published in 1954.); Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law*, 2, 293-323.;

적 판단이 전문가들 사이에 일치한다거나(“신뢰로운”) 미래의 폭력을 정확하게 예측한다는(“타당한”) 과학적 증거는 없기 때문이다.

두 번째 자유재량 방법은 기왕병력 위험성 평가(anamnestic risk assessment)라고 불린다(예: Otto, 2000). 그리스어 “기억” 혹은 “추억”에서 유래된 병력(anamnesis)은 의학에서 과거력을 알아보는 과정을 일컫는다. 병력 위험성 평가는 대상자의 과거 폭력과 관련된 개인 및 상황요인을 확인하는 등 평가과정에 제한적인 구조를 설정한 것이 특징이다. 다시 말해 과거 일련의 행동 형태 이를테면 연쇄 사건과 상황이 향후 폭력 행위로 이어진다는 가정이 있다. 따라서 평가자의 과제는 이 일련의 연쇄 관계를 이해하고 어떻게 끊을 수 있는지를 제안하는데 있다. 이런 제한적인 구조에도 불구하고, 병력 위험성 평가는 비구조화된 전문가 판단과 동일한 이유로 근거 기반이라고 할 수 없다. 마지막으로 구조화된 전문가 판단(structured professional judgement; Kropp & Hart, 2000)으로, 구조화된 임상 판단(예: Hart, 1998) 혹은 가이드 된 임상 접근(guided clinical approach; Hanson, 1998)이라고도 불린다. 이 접근에서의 의사결정은 과학적 지식과 전문적 실무에 대한 “운용규정”을 반영하기 위해 개발된 지침서의 도움을 받는다. 이 지침서는 임상적 지침서, 합의 지침서, 임상 치료 권고안으로도 불리는 데 의학에서 꽤 흔하게 사용되지만 정신의학, 심리학 혹은 교정 평가에서는 빈번하게 사용되지 않는다(Kapp & Mossman, 1996). 구조화된 전문가 판단 절차는 협의와 광의의 정의에 따라 근거를 기반으로 한다. 왜냐면, 첫째, 지침서는 과학적이고 전문적인 문헌에 기초해 직접적으로 정보를 얻고 구조화되었기 때문이다. 둘째로, 구조화된 전문가 판단 지침서를 사용하여 내린 폭력 위험성에 관한 결정이 신뢰로우면서 타당하다는 의견을 지지하는 과학적 증거가 이미 상당하고 증가 추세에 있다.⁹⁾

9) Otto, R. K. (2000). Assessing and managing violence risk in outpatient settings. *Journal of Clinical Psychology*, 56, 1239-1262.; Hanson, R. K. (1998). What do we know about sex offender risk assessment? *Psychology, Public Policy, and Law*, 4, 50-72.; Kropp, P. R., & Hart, S. D. (2000). The Spousal Assault Risk Assessment (SARA) guide: Reliability and validity in adult male offenders. *Law and Human Behavior*, 24, 101-118.; Kapp, M. B., & Mossman, D. (1996). Measuring decisional competency: Cautions on the construction of a “capacimeter.” *Psychology, Public Policy, and Law*, 2, 73-95.

1.3 폭력위험성평가의 진화: 계리적 도구와 구조화된 전문가 판단 모델의 ‘역사적 분열’

폭력 위험성의 체계성 있는 평가와 관리는 1800년대 중반에 시작된 것으로 볼 수 있다. 당시 세계의 국가 기관은 민간 및 법정 정신 건강, 사법집행, 교정 그리고 아동 보호의 책임이 있는 중앙집권 협회 및 기관을 설립하기 시작했다. 폭력 위험성 평가 및 관리의 책임이 있었던 협회 및 기관에 고용된 사람들은 당시 비구조화 된 임상 판단 외에는 어떤 접근도 사용할 선택권이 없었고 순전히 직관과 경험에 의존하였다(Gottfredson, 1987; Gottfredson & Moriarty, 2006; Harcourt, 2007). 다른 전략은 평가의 초점을 경험적 연구에 의해 밝혀진 소수의 폭력 재범 위험 요인에 맞추는 것이었다. 이것이 계리적 심리 검사 사용의 시작이었다. 미네소타 다면성격 검사(Minnesota Multiphasic Personality Inventory; Megargee, Cook, & Mendelsohn, 1967)의 적대감 과잉통제 척도 혹은 PCL, PCL-R을 사용한 폭력 예측 연구가 전형적인 예이다.¹⁰⁾

1990년대 초반, 의사결정 구조화에 대한 두 가지 다른 전략이 뚜렷이 갈라졌는데 당시 이 갈라짐을 분열이라고 표현할 수도 있다. 앞서 언급된 병력접근은 실무에 대한 권고사항을 더 상세하게 그리고 구체적으로 발전시켜나갔다. 이를테면, 정보수집 방법과 위험 요인의 존재 및 관련성에 초점을 맞추므로써 의사결정을 위해 정보를 사용하는 방법을 제안하였다. 이러한 발전은 지속적으로 누적되는 증거에 의해 지지되었고, 구체적인 위험 요인과 폭력의 이론적 모델에 관한 경험적 연구를 통합했다. 이것이 구조화된 전문가 판단 접근의 시작이었다, 전형적인 예로는 배우자 폭력 위험성 평가(Spousal Assault Risk Assessment; 이하 SARA)와 HCR-20, 뿐만 아니라 성폭력위험성-20(Sexual Violence Risk-20; 이하 SVR-20; Boer, Hart,

10) Gottfredson, S. D. (1987). Prediction: An overview of selected methodology issues. In D. M. Gottfredson & M. Tonry (Eds.), *Prediction and classification: Criminal justice decision making, crime and justice: A review of research*, Vol. 9 (pp. 21-53), Chicago: University of Chicago Press.; Gottfredson, S. D., & Moriarty, L. J. (2006). Statistical risk assessment: Old problems and new applications. *Crime & Delinquency*, 52, 178-200.; Megargee, E. I., Cook, P. E., & Mendelsohn, G. A. (1967). Development and evaluation of an MMPI scale of assaultiveness in overcontrolled individuals. *Journal of Abnormal Psychology*, 72, 519-5

Kropp, & Webster, 1997), 성폭력위험성프로토콜(Risk for Sexual Violence Protocol; 이하 RSVP; Hart et al., 2003), 스토킹 평가 및 관리 가이드(Guideline for Stalking Assessment and management; 이하 SAM; Kropp, Hart, & Lyon, 2008) 등 기타 다수가 있다.¹¹⁾ 국내에서는 서종한 등(2017)이 처음으로 성폭력위험성프로토콜(RSVP)을 타당화하였다.

그에 반해서, 심리 검사의 계리적 사용은 다문항 계리적 위험성 평가도구의 개발로 발전되었고 범죄자, 민간 및 법정 환자에 대한 대규모 후향(추적 혹은 준예측) 연구에 의해 지지되고 있다. 이것이 계리적 접근의 시작이었고, 국외에서는 VRAG(Harris & Rice, 1993), Rapid Risk Assessment of Sex Offender Recidivism(RRASOR; Hanson, 1997), Sex Offender Risk Appraisal Guide (SORAG; Quinsey et al., 1996)에 이어서 STATIC-99(Hanson & Thorn, 1999), Classification of Violence Risk(COVR; Monahan et al., 2005) 등이 전형적인 예이다.¹²⁾ 재미있는 점은 학자들은 이러한 점진적인 발달을 분열이라고 묘사하였는데 (Buchanan, 2008; Heilbrun & Dvoskin, 2001; Webster, Hucker, & Bloom, 2002) 이는 구조화된 전문가 판단과 계리적 접근 양 주창자들이 근본적으로 서로 양립할 수 없다는 우려를 공식적으로 표명했기 때문이다.

구조화된 전문가 판단과 계리적 접근의 시작을 더 정확하게는 1993년이라고 할

- 11) 이는 구체적으로 폭력 위험성 평가와 관련해 구성되었다. 구조화된 전문가 판단의 발전에 앞서 혹은 이와 동시에 다른 영역에서의 근거 기반 지침서 혹은 임상치료권고안을 만들기 위한 강한 움직임이 있었다. 예를 들어, *Level of Supervision Inventory*(1982년에 처음 개발, 이후 *Level of Service Inventory*, *Level of Service Inventory-Revised*, *Level of Service/Case Management Inventory* (이 접근의 발달 검토를 위해서는 Andrews, Bonta, & Wormith, 2010을 참고)로 개칭)와 *Dangerous Behaviour Rating Scale*(Menzies, Webster, & Sepejak, 1985)은 모두 구조화된 전문가 판단 지침서와 유사하다. 명백하게 구조화된 전문가 판단 접근의 발달에 강한 영향을 미쳤다. 하지만 *Level of Supervision Inventory*는 폭력의 위험성이 아니라 전반적인 범죄의 위험성을 평가했고, *Dangerous Behaviour Rating Scale*은 구조화된 전문가 판단 접근의 특징인 실무에 대한 상세한 권고가 부족했다. 그리고 둘 다 구조화된 전문가 판단과 일치하는 방법으로 숫자를 이용한 항목평가와 절단점에 의존했고 이것은 SPJ 접근과 일관적이지 않다.
- 12) 전반적인 범죄 위험성의 계리적 평가는 약 100년 전으로 거슬러 올라간다(Gottfredson & Moriarty, 2006; Harcourt, 2006). 초기에 Rand Corporation 관련 연구자들이 판결 시 사용하기 위해 폭력 위험성과 관련된 연구기반의 ARAIs를 개발하기 위한 여러 시도를 하였다(예: Chaiken & Chaiken, 1982; Greenwood, 1982). 하지만 이러한 시도는 제한적인 성공만 이루어졌을 뿐, 널리 받아들여지지 않았다.

수 있다. Christopher Webster 등으로 구성된 전문가 집단이 폭력 위험성 평가 실무에 대한 권고를 작성하기 위해서 캐나다 온타리오에서 함께 작업을 했다. 그들은 Violence Prediction Scheme이라고 하는 시스템을 개발한 후 매뉴얼을 출판했다(Webster, Harris, Rice, Cormier, & Quinsey, 1994).¹³⁾ 이 시스템에는 크게 두 파트로 나누어져 있다. 첫째, 계리적 위험성 평가 도구는 처음에 Risk Assessment Guide(RAG)라고 불렀다가 Violence Risk Appraisal Guide(VRAG)라고 개칭되었고, 통계에 기초한 폭력 재범의 위험성 추정치를 산출해 주었다. 둘째, 초기 수준의 구조화된 전문가 판단 지침서는 ASSESS-LIST라고 불리는데 이는 임상가가 평가해야 하는 요인 목록의 앞 글자를 따서 만든 것으로 폭력 위험성의 통계적 추정치를 조정하기 위해 사용하였다. 하지만 Violence Prediction Scheme을 작성하는 동안, 저자들은 계리적 접근과 구조화된 전문가 판단 접근을 사용하여 얻은 결과를 적절히 통합하는 데 있어 조정하기 어려운 차이가 있음을 알게 되었다.

애초에 Harris, Rice, Quinsey(1993)와 Webster와 동료들(1994)은 **계리적 추정치가 임상적 조정을 통해서 10%까지 향상될 수 있다는 것에 동의했지만 얼마 지나지 않아** 사람들의 의견이 바뀌자 두 집단으로 뚜렷이 나뉘어졌고 두 집단 모두 계리적 추정치의 임상적 조정 접근을 폐기하였다. Quinsey, Harris, Rice와 Cormier(1996/2008)는 자신들의 입장을 관철했고 계속해서 순수 계리적 접근만을 추구했다. 그들에 따르면, “우리가 권고하는 것은 현존하는 관습에 계리적 방법을 추가하는 것이 아니라, 현존하는 관습을 계리적 방법으로 완전히 대체하는 것”이며 “계리적 방법과 임상적 판단 모두 너무 우수하므로 전자를 후자로 오염시키는 위험을 치를 수 없다”(Quinsey et al., 1998, p. 171; Quinsey et al., 2006, p. 182)고 하였다.¹⁴⁾ 여기서 떨어져 나온

13) Webster, C. D., Martin, M. L., Brink, J., Nicholls, T. L., & Desmarais, S. L. (2009). *Manual for the Short-Term Assessment of Risk and Treatability (START)*, Version 1.1. Coquitlam, Canada: British Columbia Mental Health & Addiction Services.

14) Meehl은 계리적 추정치를 조정하기 위해 전문적 판단을 사용하는 것이 “우호적이고 관용적이며 공평한 것처럼 들리지만, 실제로는 어리석은 일이다”(p. 435)라고 말했다.; Harris, G. T., Rice, M. E., & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior*, 20, 315-335.; Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. (1998). *Violent offenders: Appraising and managing risk*. Washington, DC: American Psychological Association.; Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. (2006). *Violent offenders: Appraising and managing*

Webster(SPJ 1세대 연구자)는 Henrik Belfrage, Douglas Boer, Johann Brink, Kevin Douglas, Derek Eaves, Stephen Hart, P. Randall Kropp, Tonia Nicholls(SPJ 2세대 연구자) 등의 사이먼 프레이저 대학 동료와 함께 계속해서 구조화된 전문가 판단 접근을 발전시켰다(Bloom, Webster, Hucker, & De Freitas, 2004). 국내에서는 2017년부터 아주대학교 심리학과 서종한이 슈퍼바이저였던 Stephen Hart, Kevin Douglas, Randall Kropp 등과 연구를 함께 진행하고 있다.¹⁵⁾

1.4 HCR-20 버전 3의 특징

〈표 1〉 HCR-20^{v3} 20 위험요인과 세부 하위항목

이력 척도 (문제이력)	하위항목
H1. 폭력	a. 아동기 (12세 이하) b. 청소년기 (13-17세) c. 성인기 (18세 이상)
H2. 기타 반사회적 행동	a. 아동기 (12세 이하) b. 청소년기 (13-17세) c. 성인기 (18세 이상)
H3. 대인관계	a. 이성적 b. 비이성적
H4. 고용	
H5. 물질 사용	
H6. 주요 정신 장애	a. 정신병적 장애 b. 주요 기분장애 c. 기타 주요 정신장애
H7. 성격 장애	a. 반사회적, 사이코패스, 비사회적 b. 기타
H8. 외상 경험	a. 피해경험/외상 b. 역기능적 양육경험
H9. 폭력적 태도	
H10. 치료 혹은 관리감독 반응	

risk, 2nd ed. Washington, DC: American Psychological Association.

- 15) Bloom, H., Webster, C. D., Hucker, S., & De Freitas, K. (2005). The Canadian contribution to violence risk assessment: History and implications for current psychiatric practice. *Canadian Journal of Psychiatry*, 50, 3-11.

임상 척도 (최근문제)	
C1. 통찰력	a. 정신장애 b. 폭력 위험성 c. 치료 필요성
C2. 폭력 사고 혹은 의도	
C3. 주요 정신장애 증상	a. 정신병적 장애 b. 주요 기분장애 c. 기타 주요 정신장애
C4. 불안정성	a. 정서적 기능 b. 행동적 기능 c. 인지적 기능
C5. 치료 혹은 관리감독 반응	a. 순응도 b. 반응도
위험 관리 척도 (미래문제)	
R1. 전문서비스와 계획	
R2. 주거 환경	
R3. 개인적 지지	
R4. 치료 혹은 관리감독 반응	a. 순응도 b. 반응도
R5. 스트레스 혹은 대처	

Note. © [서종한]. 재사용은 저작권자 허락이 필요하며 개별 위험요인에 대한 한국번역은 저작권자의 한국판을 사용해야 함.

<표 1>에 제시된 바와 같이 HCR-20은 이력, 임상, 위험관리 척도 등 3개척도 20문항으로 이루어져 있다. 이력(H) 척도는 10개의 문제이력 문항으로 구성, 임상(C) 척도는 5개의 최근문제 문항으로 구성, 위험관리(R) 척도는 5개의 미래문항으로 구성되어 있다. 코딩은 각 위험요인에 정보가 존재하는 경우 “있음”, 해당요인이 잠재적으로 혹은 부분적으로 존재하는 경우 “가능”, 마지막으로 해당요인이 존재하지 않거나 적용되지 않은 경우는 “없음”으로 평가한다. 하지만 해당 위험요인의 존재를 판단할 수 있는 신뢰로운 정보가 없는 경우 “생략”으로 코딩할 수 있다. HCR-20은 크게 7단계에 따라 평가가 이루어지고 있으며 1단계 관련정보탐색, 2단계 위험요인 존재유무 코딩, 3단계 위험요인의 관련성 코딩, 4단계 위험성 사례개념화, 5단계 위험성 시나리오구성, 6단계 위험성 시나리오에 따른 관리계획 수립, 7단계 최종의견이다. 아래는 각 단계에 대한 세부 특징을 논하였다.

단계1: 관련정보탐색 혹은 사례정보. 해당 사례와 관련된 정보를 탐색하여 필요

한 자료를 수집하는 절차를 말한다. 이 과정에서는 면담 및 관련 자료 검토가 포함되어 있다.

단계2: 위험요인 존재유무 코딩. 이력(H) 척도는 전생애를 기반으로 평가 대상자의 심리사회적 적응 이력의 특징을 반영한다. 이력 척도는 한 개인의 과거 몇몇 지점에서 발생했을 지도 모르는 핵심적이고 중요한 폭력 위험 요인들을 나타내기 위한 것이다. 10개의 위험 요인들을 여러 개의 대략적인 범주에 넣어 보면 적응이나 삶에서의 문제는 H3(대인관계), H4(고용), H8(외상경험)에서 기록한다. 정신건강과 관련된 문제는 H5(물질사용), H6(주요 정신장애), H7(성격장애)을 포함한다. 과거 반사회성은 H1(폭력), H2(다른 반사회성 행동), H9(폭력적 태도), H10(치로나 관리 감독에 대한 반응)에서 기록한다. 임상(C) 척도는 평가대상자의 최근 심리 사회적 적응의 특정 특징을 반영한다.

현재 평가대상자가 통제된 환경에 있고 그런 환경에서 지내왔다 할지라도, 심리 사회적 적응의 특정 특징이 존재한다면, 항상 중요하게 고려할 필요가 있다. 이 요인을 코딩하기 위한 기간범위는 평가 전 몇 주에서 몇 달과 같은 최근 과거이다. 임상 척도 위험요인들을 코딩하기 위해 평가자들은 평가의 목적과 특성, 평가 대상자의 삶의 상황과 과거 몇 주 혹은 몇 달에 걸친 적응에서의 주요한 변화들, 정보의 가용성 등을 고려하여 평가할 특정 기간을 정해야 한다. 그러나 여러 장면에 걸쳐 적용할 수 있는 균일한 평가 틀이 없다. 몇몇 내담자들은 이전에 평가받은 적이 없을 수도 있고 혹은 만약 그들이 평가 받았다고 해도 현재 평가자가 평가 결과를 알지 못하거나 가용하지 않았을 수 있다.

하위 항목 위험요인 코딩. HCR-20^{v3}의 일부 위험 요인은 내용 면에서 다면적 혹은 광범위한 반면에 다른 것들은 내용이 상대적으로 단순하거나 좁은 의미를 가진다. 몇몇 위험 요인에 대해서는, 두 개 이상의 내용 영역 혹은 하위 문항이 확인되었다. 하위문항은 평가자가 복잡한 위험 요인에 대한 존재 유무 평가를 더 확실하게 이해하고 실시할 수 있게끔 도움을 줄 수 있다. 또한 평가자는 각 하위문항의 존재 유무를 개별적으로 평가하도록 선택할 수 있다. 이런 방식은 HCR-20^{v3}의 임상적 혹은 통계적 분석을 더 정교하게 한다. 하위 문항의 평가도 중요 위험 요인의 경우처럼 일반 평가 방식을 따르도록 한다(즉, 부재, 잠재적/부분적, 존재).

지표코딩. 각각의 HCR-20^{v3}의 위험요인에 대해 세부 지표 목록을 제공하고 있다. 하위문항이 있는 위험 요인에 대해서도 각 하위 문항에 대한 세부지표를 제공하였다. 지표는 평가자가 위험 요인의 존재 유무 판단을 내릴 때 찾아야 하는 대표적인 정보의 예라고 할 수 있다. 이것을 통해 평가자가 위험 요인에 대한 존재 유무 평가를 더 확실하게 이해하고 실시하는데 도움을 준다. 또한 개별위험 요인의 징후에 대한 이해를 용이하게 하는데 그 목적을 두었다. 하지만 지표가 주로 위험 요인에 필요한 안내지침을 제공하는 것일지라도, 평가자 혹은 연구자는 각 지표의 존재 유무를 개별적으로 평가할 수도 있다. 이런 식으로 HCR-20^{v3}의 임상적 혹은 통계적 분석을 정교화 하는 데 도움을 줄 수 있다. 지표의 존재 유무를 평가할 때는 “Y”(Yes)는 존재, “N”(No)은 부재라는 단순한 2점 척도를 사용하여 평가한다. 연구 목적을 위해서는 위험요인의 존재 유무 평가와 관련성 평가 시 사용하는 것과 같은 3 수준 체계를 사용하여 지표를 평가할 수도 있다.

단계3: 위험요인 관련성. 평가자는 미래 위험의 관리전략을 염두에 두고 위험요인의 관련성을 판단해야 한다. 관련성이란, 평가 대상자가 폭력을 저지른 원인과 추후 폭력 예방을 위한 최선책에 대해 사례개념화 할 때 어떤 요인이 얼마나 중요하지 그 정도를 의미한다고 볼 수 있다. 관련성이 확인된 위험요인은 모니터링, 치료, 관리감독, 피해자 안전에 관한 계획을 세울 때 중요하게 고려해야 한다. 관련성은 3가지 수준으로 코딩한다. 낮음(Low)은 수집한 사례 정보에 근거하여 평가자가 그 요인이 개인의 폭력 위험성 혹은 위험성 관리 전략의 개발과 관련성이 낮다고 결론 내린 것을 의미한다. 이는 요인이 부재하거나, 존재하지만 미래 폭력에 인과적 역할을 할 가능성이 낮거나, 존재하지만 위험성 관리 전략의 효과성을 실질적으로 손상시킬 가능성이 낮기 때문일 수 있다. 중간(Moderate)은 평가자가 요인이 어느 정도 관련이 있지만 폭력을 유발하거나 위험성 관리 전략의 효과성을 손상시키는 데 기여하는 역할이 명확하지 않고 제한적일 가능성이 높다고 보는 것이다. 마지막으로, 높음(High)은 평가자가 보기에 요인이 존재하며, 그 요인이 폭력을 유발하거나 위험성 관리 전략의 효과성을 손상시키는 데 기여한 역할이 명확하고 실질적임(즉, 높은 관련성)을 의미한다.

단계 4: 위험성 사례개념화. 사례개념화는 일반 심리치료 문헌에 근거하고 있다.

목적은 임상가가 통찰력을 가지고 개인이 지닌 문제의 원인을 쉽게 개념화 하고자 하는 것이다. 사례개념화는 3단계에서 위험요인의 개별적인 관련성을 결정하면서 시작되고, 이때 평가자는 개인에게 직접적인 영향을 끼치는 위험요인을 고려해야 한다. 평가자는 개별 위험요인을 개념상 중요한 틀로 통합시킴으로써 위험성 사례개념화 과정을 발전시킨다. 이상적으로는, 가용한 단편 정보를 통합하여 개인에 대한 이야기를 구성해야 한다. 이 과정은 개인별 위험성 이론을 도출하고 위험성 이론을 이해하기 위해 필요하며, 이를 통해 위험성에 대한 최상의 개입 및 관리가 가능해진다. Hart와 Logan(2011)은 사례개념화의 필수적인 특징을 논의하였는데, 사례개념화는 다음 특징들을 포함하고 있어야 한다: ① 추론적(사례개념화는 행동의 이유 혹은 원인을 설명한다), ② 행동 지향적(사례개념화는 임상가 혹은 형사사법 전문가의 위험성 감소 조치를 안내한다), ③ 이론에 근거한(사례개념화는 원리에 입각한 중요한 지식 구조의 안내를 추구한다), ④ 개별화된(사례개념화는 반드시 한 개인의 행동을 설명한다), ⑤ 서술적인(사례개념화는 중요한 “기준점(anchor)”이 있는 일시적인 이야기를 만들어내는 질적인 과정이다), ⑥ 통시적인(사례개념화는 다양한 시기를 다룬다—과거, 현재, 미래), ⑦ 검증 가능한(사례개념화는 평가될 수 있어야 한다), ⑧ 확충적인(사례개념화는 단지 알려진 사실을 요약하기 보다는 개인에 관한 새로운 정보 혹은 지식을 창조한다). 더 구체적인 수준에서, 개인수준에서 폭력 관련 가설을 검증할 수 있는 여러 개념 모델이 있다. 대표적으로, 평가자는 **결정이론 프레임워크(decision theory framework; Hart & Logan, 2011; 서종한 등, 2017 참고)**를 사용하여 특정 위험요인의 수준이나 특정 요인들 간의 조합이 **동기, 탈억제, 불안정성 요인**으로 어느 정도 작용할지 고려해볼 수 있다.¹⁶⁾

단계 5: 위험성 시나리오 4단계에서 논의한 위험성 사례개념화는 사람들이 폭력적인 이유를 밝히는데 도움을 준다. 5단계에서는 위험성 관리에 대한 최종 가고를 만들기 위해 위험성 시나리오를 사용한다. 어떤 면에서 그것은 한 개인이 미래에 무엇을 할 것인가, 그리고 그것을 왜 할 것인지와 같이 미래지향적인 사례개념화이

16) Hart, S. D., & Logan, C. (2011). Formulation of violence risk using evidence-based assessments: The structured professional judgment approach. In P. Sturmey & M. McMurran (Eds.), *Forensic case formulation* (pp. 83-106). Chichester, UK: Wiley-Blackwell; 서종한, 김희송, & 김정일. (2017). 성폭력위험성평가의 진화. 형사정책연구, 28(2), 215-256.

다. 위험성 시나리오의 구성은 이전 단계들을 많이 참고한다.

Hart 등(2003; Hart et al., 2003; Hart & Logan, 2011)에 따르면, 위험성 평가 기반 시나리오에는 경영, 보건, 군대 영역과 같은 곳에서 50년 이상 사용해 온 시나리오 계획, 관리 전략 방식이다(Ringland, 1998; Schwartz, 1990; van der Heijden, 1997). Chermack과 Lynham(2002, p. 366)에서 “시나리오 계획은 몇 가지 알려진, 가능한 그리고 상상했던 대안적 환경을 상정하는 과정인데, 이 과정을 통해 미래 결정이 최근의 생각을 바꾸고, 의사결정을 개선하며, 인간 및 조직 학습을 향상시키고, 수행 향상을 목적으로 이루어질 수 있는 것”이라고 규정했다. 이는 불완전한 지식에 근거하여 결정을 내려야 하는 상황, 즉 빈도주의 학파의 확률에 대한 견해에 기반을 둔 접근법으로 결과를 예측하는 게 불가능한 불확실한 자원들이 다양한 상황에서 가장 적절하다(예: Epstein, 1998; van der Heijden, 1994). 이 방법이 폭력 위험성 평가에 관해 현황을 정확하게 규정하는 것처럼 보일 수 있다(Hart, 2003; Hart et al., 2003). HCR-20^{v3}에서 시나리오는 개인이 저지를 수 있는 여러 종류의 폭력에 대한 이야기이다. 무엇이 일어날 것인지가 아닌, 무엇이 일어날 수 있을지에 관한 예상내지는 추정이라 볼 수 있다. 구성 가능한 시나리오 개수가 무궁무진할지라도, 알려진 사실과 이론에 근거하면 어떤 사례에 해당하는 몇 가지 두드러지는 시나리오들만 평가자에게 타당하고, 믿을 만하며, 일관성 있어 보일 것이다(예: Chermack & van der Merwe, 2003). 여러 가지 다른 시나리오를 작성하는 것이 이론적으로 가능할 수 있지만 현실성이 떨어질 수 있으므로 최종적으로 전문가가 기각하거나 “제거”할 수 있을 것이다(예: Pomerol, 2001).¹⁷⁾

6단계: 관리전략 이 단계에서 평가자는 위험 관리 전략을 권고한다. 전략은 2, 3단계에서 위험 요인의 존재 유무와 관련성, 4단계에서 만든 위험성 사례개념화 뿐

17) Chermack, T. J., & Van Der Merwe, L. (2003). The role of constructivist learning in scenario planning. *Futures*, 35(5), 445-460.; Ringland, G. (1998). *Scenario planning: Managing for the future*. Chichester, UK: Wiley.; Schwartz, P. (1990). *The art of the long view*. New York: Doubleday.; Chermack, T. J., & Lynham, S. A. (2002). Definitions and outcome variables of resource planning. *Human Resource Development Review*, 1, 366-383.; van der Heijden, K. (1997). *Scenarios: The art of strategic conversation*. New York: Wiley.; Pomerol, J.-C. (2001). Scenario development and practical decision making under uncertainty. *Decision Support Systems*, 31, 197-204.

만 아니라 5단계에서 확인된 폭력성에 대한 다양한 시나리오에 기초한다. 위험성 관리에서는 4가지 기본 활동 즉, 모니터링, 치료, 관리감독, 피해자 안전 계획에 대해 생각해 보는 것이 유용하다(Hart et al., 2001; Kropp et al., 2002). 평가자는 개인의 위험 수준, 위험요인의 유형과 관련성, 사례개념화와 시나리오 계획을 통해 확인한 가능한 이유들 중 폭력을 가장 잘 설명하는 이유를 선택해야 한다. **모니터링/감시.** 모니터링의 목표는 관리 전략을 적절하게 수정할 수 있도록 시간에 따른 위험성의 변화를 평가하는 것이다. 모니터링은 관리감독과는 달리, 개인의 자유를 통제하거나 제한하기 보다는 감시하는 것에 초점을 맞춘다. 따라서 침해를 최소화한다. **치료/평가.** 치료는 훈련 혹은 사회 복귀 서비스를 제공하는 것을 포함한다. 치료 목표는 심리사회적 적응 혹은 기능에서의 결함을 개선시키는 것이다. **관리감독/통제.** 관리감독은 개인의 권리 혹은 자유에 대한 규제를 포함하고 있다. 관리감독의 목표는 미래에 개인이 폭력에 관여하게 되는 것을 더 어려워지게 만드는 것이다. 일반적으로, 관리감독은 개인이 가진 위험성과 비례하는 강도로 시행해야 한다. 또한, 미래 폭력 예방률을 가장 최적화 하려는 원리는 개인의 시민권을 보호하며, 관리감독 서비스 제공에 포함된 법적 책임을 줄여준다. **피해자 안전 계획.** 피해자 안전 계획은 잠재적 피해자의 역동적 및 정적 안전 자원을 개선시키는 것, 즉 때로 “범지표적 강화(target hardening)”라고 일컫는 과정을 포함한다. 만약 모든 모니터링, 치료, 관리감독 노력에도 불구하고 폭력이 재발한다면, 피해자의 심리적 및 신체적 안녕에 미치는 모든 부정적인 영향을 최소화시키는 것이 목표이다.¹⁸⁾

7단계: 잠정적 의견. 마지막 단계에서는 적절한 조치를 쉽게 할 수 있도록 판단을 분명하고 간단하게 요약한다. 이 때 미래 폭력 위험성/사례 우선순위 매기기, 심각한 신체적 피해의 위험성, 임박한 폭력의 위험성, 사례 검토, 폭력 외에 다른 위험성이 보이는지 여부와 같은 주제를 다뤄야 한다.

미래 폭력의 위험성/사례 우선순위 매기기. 사례 우선순위 매기기는 폭력을 저지르는 것을 막기 위해 요구되는 노력 혹은 개입의 정도에 따라, 3가지 수준의 반응

18) Kropp, P. R., Hart, S. D., Lyon, D., & LePard, D. (2002). Managing stalkers: Coordinating treatment and supervision. In L. Sheridan & J. Boon (Eds.), *Stalking and psychosexual obsession: Psychological perspectives for prevention, policing and treatment* (pp. 138-160). Chichester, UK: Wiley.

양식을 사용하여 입력해야 한다. 반응 양식은 다음과 같다. “낮음” 혹은 “보통”은 그 사람에게 폭력 위험성 관리를 위한 어떤 특별한 개입 혹은 관리감독 전략이 필요하지 않으며, 위험성이 변화하는지 그 사람을 유심히 모니터링 할 필요가 없음을 의미한다. “중간의” 혹은 “증가한”은 최소한 모니터링 하는 빈도가 증가하는 것을 포함하여, 약간의 특별한 관리전략이 필요함을 의미한다. “높은” 혹은 “긴급한”은 위험 관리 계획을 긴급히 개발할 필요가 있음을 의미한다. 주로 (최소한)직원에게 알리기, 관리감독 수준 높이기, 그 사람에게 가용한 치료자원의 우선순위를 매기기, 정기적인 재평가 일정 정하기를 포함한다. 일부 위험성이 높은 사례들에서는 비상 대응(예: 입원, 가석방 유예)이 필요할 것이다.

심각한 신체적 위해 위험성. 심각한 신체적 피해의 위험성은 그 사람이 저지를 수 있는 폭력의 심각도를 반영한다. 위험성의 이러한 측면은 법정 의사결정에서 특히 더 중요하다. 예를 들어, 강제 치료감호, 신상정보공개, 경고 혹은 보호해야 하는 의무이행에 관한 결정을 내릴 때 “낮은”, “중간의”, “높은” 위험성이라는 3 수준의 체계를 사용하여 피해 위험성을 코딩해야 한다. **임박한 폭력 위험성.** 이 평가는 평가 대상자가 저지를 수 있는 폭력의 임박성을 반영한다. 심각한 신체적 피해의 위험성과 같이 이 위험성의 측면은 강제 치료감호, 신상정보공개, 경고 혹은 보호할 의무 이행에 대한 의사결정시 특히 더 중요할 수 있다. “낮은”, “중간의”, “높은” 위험성이라는 3 수준의 체계를 사용하여 임박한 폭력의 위험성을 코딩하도록 권한다. 또한, 평가자는 법적 책임 관리의 목적으로 과거에 행해졌고, 혹은 미래에 행해질 수 있는 모든 즉각적인 조치에 대해 분명하게 기록해야 한다. 사례 검토. 평가자는 정기 사례 검토(즉, 위험성 재평가) 날짜뿐만 아니라 빠른 시일 내에 재평가를 해야 한다는 경고 신호를 남겨야 한다. 이렇게 하는 이유는 위험성이 시간에 따라 변하기 때문이다. 개인력의 실제 변화(예: 새로운 유죄 선고) 혹은 사례 정보의 변화(예: 범 죄경력에 관한 새로운 정보 등장) 때문에 폭력 이력이나 사이코패스 성격 장애와 같은 정적 요인조차도 시간에 따라 변화한다. 지역사회에 살고 있는 경우, 이런 변동이 매우 빠르게 일어날 수 있다. 예를 들어, 단일 폭력 행위가 여러 요인에 대한 판단을 바꿀 수 있고 단일 위험요인의 악화로 인해 사례 우선순위 매기기에 대한 평가자의 판단이 극단적으로 변할 수도 있다. 재평가에서는 위험관리 계획의 재검

토를 권장하며, 이를 바탕으로 평가자는 평가 대상자의 긍정적인 혹은 부정적인 진전의 특징을 설명할 수 있다. **기타 위험성.** 마지막으로, 평가자는 다른 위험성의 존재 여부를 파악하고 이후 평가에 초점을 맞춰야 한다. 이는 특수한 형태의 폭력(즉, 성적으로 친밀한 파트너) 혹은 다른 해로운 활동(즉, 자살 관련 행동: 비자살적 자해)의 형태로 나타날 수 있다.

1.5 HCR-20의 계량적 특징

채점자 간 신뢰도

de Vries Robbé와 de Vogel(2010)은 HCR-20^{V3} 원고 1을 86명의 범법 정신질환자에게 시행하였다. 25개의 사례에 대해 각기 다른 평가자가 개별적으로 평가하였다. 급내상관계수(ICC)에 나타나듯이 평가자 간 신뢰도는 위험성 유무에 있어 .84, SRR은 .72로 나타났다. HCR-20^{V2}의 위험요인 존재 유무의 평가자 간 신뢰도는 .83으로 두 버전 사이의 평가자 간 신뢰도는 거의 동일하다. Belfrage와 Douglas(2012)는 HCR-20^{V3}의 원고 3을 스웨덴의 Sundsvall 범법병원 범법 정신질환자 35명에게 평가하였다. 위험요인 존재 유무의 평가자 간 신뢰도는 적절한 정도에서 우수한 정도였다. 이력, ICC₁ = .94, ICC₂ = .98; 임상적, ICC₁ = .86, ICC₂ = .95; 위험 관리(기관 내 평가), ICC₁ = .69, ICC₂ = .87; 위험 관리(지역사회 평가), ICC₁ = .75, ICC₂ = .90; 총합(기관 내), ICC₁ = .94, ICC₂ = .98; 총합(지역사회), ICC₁ = .94, ICC₂ = .98. SRR의 평가자 간 신뢰도 또한 우수했다. 내, ICC₁ = .81, ICC₂ = .93; 외, ICC₁ = .75, ICC₂ = .90. Doyle, Shaw와 Coid(2013)는 HCR-20^{V3} 원고 2를 네 명의 평가자로 하여금 영국과 웨일즈의 범법 정신질환자 20명을 대상으로 평가를 진행하였다. 이력, ICC₁ = .72, ICC₂ = .91; 임상적, ICC₁ = .69, ICC₂ = .90; 위험 관리(지역사회), ICC₁ = .76, ICC₂ = .93; 총합(지역사회), ICC₁ = .73, ICC₂ = .92의 결과를 보였다. HCR-20^{V3} 원고 1의 이력(H) 척도에 관한 파일 코딩 연구는 80명의 캐나다 범법 정신질환자를 대상으로 Strub과 Douglas(2009)가 세 명의 평가자에게 여섯 개 파일 평가를 진행시켰다. HCR-20^{V3}의 이력 위험요인에 대한 존재 유무 평가는 수용가능 했다, ICC₁ = .75, ICC₂ = .85이었고, HCR-20^{V2}를 사용했을 때 보다

약간 높았다($ICC_1 = .69$, $ICC_2 = .82$).¹⁹⁾

공인 타당도

일부 연구에서 HCR-20 버전 2와 3 사이의 연계성을 평가하였다.²⁰⁾ 우선 위험요인 존재 유무에 대한 종합평가를 보면 버전 2와 버전 3 간의 상관관계 계수는 노르웨이 범법환자 20명: $r=.84$ (Eidhammer et al., 2013), 스웨덴 범법 환자 35명: 기관 내 평가=.85와 지역사회 평가=.90(Belfrage & Douglas, 2012), 캐나다 민간 정신질환자 및 범죄자 43명: .86(Blanchard & Douglas, 2011), 네덜란드 범법 환자 86명: .93(de Vries Robbé & de Vogel, 2010)으로 높았다. 다음은 위험요인 이력에 대한 위험성 존재 유무 평가의 상관관계수 보면 두 버전 사이의 상관관계는 노르웨이 범법 환자 20명: $r=.85$ (Eidhammer et al., 2013), 스웨덴 범법 환자 35명: .87(Belfrage & Douglas, 2012), 캐나다 범법 환자 80명: .60(Strub & Douglas, 2009). 캐나다 민간 정신질환자 및 범죄자 43명 .86(Blanchard & Douglas, 2011)으로 보편적으로 높았다.²¹⁾

19) de Vries Robbé, M., & de Vogel, V. (2010). *Pilot study of HCR-20 (Version 3)*. Unpublished data analyses. Van der Hoeven Kliniek, Utrecht, The Netherlands.; Belfrage, H. & Douglas, K. S. (2012). *Interrater reliability and concurrent validity of HCR-20 (Version 3)*. Unpublished data analyses. Mid-Sweden University, Sundsvall, Sweden.; Strub, D. S., & Douglas, K. S. (2009, March). *New version of the HCR-20 violence risk assessment scheme: Evaluation of a draft of the revised Historical scale*. Poster presented at the American-Psychology Law Society conference, San Antonio, Texas.

20) 독립적인 평가자가 HCR Version 2와 Version 3을 평가하는 것은 불가능 했다. 원칙적으로 오차 분산이 제거되기 때문에 버전사이의 관계가 부풀려 질 수 있다. 그러나 이런 코딩 방법은 임상적인 장면에서 같은 평가자 혹은 평가자가 수검자를 평가할 때 Version 2에서 Version 3로 바꾸는 것과 유사하여 현실적이라고 볼 수 있다. 더욱이 표면적으로는 동일한 평가자가 두 가지 버전을 사용함으로써 의도적이지는 않겠지만 두 버전 사이의 차이점을 “찾으려 할 수도”있기 때문에 두 버전사이의 연계성이 줄어들 수 있다. 요약하자면, 이들 연구가 합당한 방법으로 두 버전 사이의 연계성을 평가하였다고 믿는다.

21) Eidhammer, G., Selmer, L. E., & Bjørkly, S., and the SAFE Pilot Project Group (2013). *A brief report of a comparison of the HCR-20 (Version 2) and HCR-20 (Version 3) concerning internal consistency and clinical utility*. Centre for Research and Education in Forensic Psychiatry, Oslo University Hospital, Norway.; Blanchard, A., & Douglas, K. S. (2011, March). *APLS Undergraduate Paper Award (First Place): The Historical-Clinical-Risk Management-Version 3: The inclusion of idiographic relevance ratings in violence risk assessment*. Invited poster presented

임상 척도에 대한 위험성 존재 유무 평가의 숫자 평점을 보면 두 버전 사이의 상관관계는 노르웨이 범법 환자 20명: $r=.59$ (Eidhammer et al., 2013), 스웨덴 범법 환자 35명: $.76$ (Belfrage & Douglas, 2012), 캐나다 민간 정신질환자 및 범죄자 43명: $.78$ (Blanchard & Douglas, 2011)로 중간정도에서 높은 수준이다. 위험 관리 척도에 대한 위험성 존재 유무 평가의 상관계수를 보면 스웨덴 범법 환자 35명: 지역사회 평가= $.82$ (Belfrage & Douglas, 2012), 캐나다 민간 정신질환자 및 범죄자 43명: 지역사회 평가= $.67$ (Blanchard & Douglas, 2011), 노르웨이 범법 환자 20명: 지역사회 평가= $.81$ (Eidhammer et al., 2013)로 두 버전 사이에 중간정도에서 높은 수준의 상관관계를 보였다.

한 연구는 HCR-20 각 버전에 걸쳐 나타나는 위험요인평가(SRRs) 사이의 상관관계에 대해 조사하였다. Blanchard와 Douglas(2011)는 캐나다 민간 정신질환자 및 범죄자 43명의 SRRs를 비교 후 $r=.94$ 의 상관관계를 보고하였다. 이 집단($N=142$)의 후속연구에서 보고된 상관관계는 $r=.98$ 이다(Douglas & Strub, 2013). Blanchard와 Douglas(2011)의 연구에서 HCR-20^{V3}의 위험성 존재 유무, 관련성, 그리고 SRR은 Psychopathy Checklist: Screening Version(PCL: SV; Hart, Cox, & Hare, 1995)와 $r=.71$ (관련성), $r=.79$ (위험성 존재 유무), $r=.82$ (SRRs)로 강력한 상관관계를 보였다. Smith와 Edens(2013)는 범죄자 32명을 대상으로 HCR-20^{V3}의 원고 3과 PCL-R 점수의 동시 타당도 분석을 실시하였다. HCR-20^{V3}을 사용한 두 가지 도구의 총점 상관관계는 $.66(p<.001)$, 위험 관리(기관 내 평가) $.70(p<.001)$. 위험관리(지역사회 평가) $r=.80$ 로 나타났다. PCL-R과 HCR-20^{V3}의 이력, 임상, 위험관리(지역사회 평가) 척도는 순차적으로 유의미한 수준에서 $r=.66$, $.62$, $.63$ 로 유사하였다. PCL-R의 요인 1(정서/대인관계)은 요인 2(충동/반사회성; $rs = .64-.83$, all $ps<.001$)에 비해 HCR-20^{V3}의 점수($rs = .12-.32$, all ns)와 상대적으로 낮은 연관성을 보였다. 이력 척도를 분석하기 위해, H7(성격장애)을 제외하였다. 단면(Facet) 수준에서 PCL-R의 단면 2(정서성)는 HCR-20^{V3}의 점수와 다소 강한 연관을 보였으나($rs=.16$ 에서 $.39$),

at the annual convention of the American-Psychology Law Society, Miami, FL.; Smith, S., & Edens, J. F. (2013). *Concurrent validity between the HCR-20 (Version 3) and the PCL-R among male pre-trial offenders in Texas*. Unpublished data analyses. Texas A&M University, College Station, Texas.

단면 1(대인관계)과의 연관성은 약하고 유의미하지 않은 것으로 나타났다. 임상 척도와 상관계수는 유의미 했지만($r=.39, p<.05$) 이력척도와 상관계수는 상대적으로 미미했다($r=.34, p=.058$). PCL-R의 단면3(생활방식)은 HCR-20^{V3} 점수와 관련성이 높았다($r=.80$ 에서 $.66, p<.01$ 에서 $.001$). 단면4(반사회성)는 HCR-20^{V3} 점수와 가장 강력하게 관련되어 있다($r=.62$ 에서 $.81, p<.001$).

예측 타당도

Doyle, Shaw와 Coid(2013)는 387명의 범법 정신질환자들을 대상으로 전향적 연구를 실시하였다. 영국과 웨일즈의 중간단계의 보안 감옥에서 출소한 지 12개월 된 집단을 대표한 샘플이었다. 이력, 임상, 위험 관리 척도 총점의 평균을 보면 각각 6개월과 12개월에, 폭력적인 사람이 그렇지 않은 사람에 비하여 유의미 하게 높은 점수를 보였다. 6개월 차의 AUC 값은 통계적으로 유의미하였다. 총점 척도에 대한 로지스틱 회귀분석은 연령과 성별을 통제 한 후 6개월과 12개월의 폭력성을 유의미 하게 예측하는 것으로 나타났다($ps<.001$).²²⁾

Douglas와 Strub(2013)은 범죄자 106명, 정신질환자(55% 남성)를 대상으로 HCR-20^{V3}의 세 번째 원고의 SRRs을 평가하였다. 혼합 집단, 각각의 하위 집단의 결과를 따로 산출하였다. 범죄자 집단을 모집 했을 당시 범죄자들은 2년 이하의 선고를 받아 교정 기관에서 나가는 과도기이거나 혹은 보호관찰 중이었다. 지역사회에서의 폭력성은 4-6주 후 그리고 6-8개월 후 평가되었다. 6-8개월의 추적조사 기간 동안($N=106$)을 보면 SRRs의 폭력성 예측은 유의미하였다($r=.41, p<.001$). 정신과 집단($r=.48, p<.001, n=50$)과 교정집단($r=.33, p<.01, n=56$) 등의 하위집단도 마찬가지였다. AUC값은 혼합 $.73, p<.001$, 정신질환 $.74, p=.01$, 교정집단 $.68, p<.05$ 으로 상관 값과 비례했다.

Douglas와 Strub(2013)은 HCR-20^{V3}의 SRRs이 표본 집단(정신질환 집단 vs. 범죄자 집단)에 의해 조절 되는지 알아보기 위해 로지스틱 회귀분석을 실시하였다. 그

22) Doyle, M., Shaw, J., & Coid, J. (2013). *Validation of risk assessments in medium secure services (VORAMSS): Preliminary HCR-20 (Version 3) analyses*. Unpublished data analyses. University of Manchester, UK, and Queen Mary University of London, UK.;

결과 표본 자체 그리고 표본 상호작용(각각 $B=.84$ 그리고 $-.91$, 모두 $p>.19$) 모두 폭력성을 예측하지 못했다. 마지막 모델에서 SRRs만(B = 2.67; $e^b = 14.41$; Wald = 4.61, $p<.05$) 폭력성을 예측가능 하였다(Model $\chi^2 = 20.25$; -2LL = 115.60; Neglkerke $R^2 = .24$). 이는 앞서 언급한 것처럼 집단 간 이변량 효과크기에 있어 약간의 변량이 존재하나 SRRs이 6-8개월 동안 환자와 범죄자에 대해 다르게 예측하지 않았고, HCR-20^{V3}의 SRR은 집단을 통제된 후에 예측력이 있음을 의미한다. 또한 Douglas와 Strub(2013)은 HCR-20^{V3}의 SRRs의 예측 타당도가 성별에 의해 조절 되는지 알아보았다. 결과를 보면 성별, 성별 상호작용(각각 $B=-.55$ 그리고 $-.20$, 모두 $p>.46$) 모두 폭력성을 예측하지 못했다. 마지막 모델에서 SRRs만(B=1.16; $eb=3.19$; Wald=9.13, $p=.003$) 폭력성을 예측할 수 있었다(Model $\chi^2=20.68$; -2LL=115.17; Neglkerke $R^2=.25$). 이는 SRRs이 6-8개월 동안 남녀를 다르게 예측하지 않음을 의미한다.

4-6주 후속연구도(N=101) 유사하게 SRRs이 폭력성을 유의미하게 예측한다는 것을($r=.43$, $p<.001$) 밝혔다. 6-8개월 후속 연구와 마찬가지로 Douglas와 Strub(2013)은 HCR-20^{V3}의 SRRs이 표본 집단(정신질환 집단 vs. 범죄자 집단)에 의해 조절되는지 알아보기 위해 로지스틱 회귀분석을 실시하였다. 결과를 보면 표본 자체 그리고 표본 상호작용(각각 $B=2.39$ 그리고 -1.88 , 모두 $p>.15$) 모두 폭력성을 예측하지 못했다. 마지막 모델에서 SRRs만(B=4.92; $e^b = 137.43$; Wald=3.99, $p<.05$) 폭력성을 예측할 수 있었다(Model $\chi^2=21.84$; -2LL= 66.44; Neglkerke $R^2 = .33$). 이는 SRRs이 4-6주 동안 환자와 범죄자를 다르게 예측하지 않는다는 것과 집단을 통제된 후 HCR-20^{V3}의 SRRs이 예측 가능함을 의미한다.

6-8개월 후속 연구와 마찬가지로 4-6주 후속 평가에서 HCR-20^{V3} SRRs 또한 성별에 의해 조절되지 않았다. 결과를 보면 성별, 성별 상호작용(각각 $B=.74$, $.53$, 모두 $p>.57$) 모두 폭력성을 예측하지 못했다. 마지막 모델에서 SRRs만(B=1.44; $eb=4.24$; Wald=6.83, $p=.009$) 폭력성을 예측 가능하였다(Model $\chi^2=19.52$; -2LL = 68.76; Neglkerke $R^2=.30$). 이는 SRRs이 4-6주 동안 남녀를 다르게 예측하지 않음을 의미한다.

주로 SRRs에 초점을 둔 분석이었지만, 몇몇 분석은 HCR-20^{V2}와 HCR-20^{V3}의

존재 유무 총점을 비교한 것이었다. 결합된 표본을 이용하여 6-8개월간 후속연구 끝에 버전 2는 폭력성과 유의미하게 관련이 되어있고($r=.46, p<.001$), 버전 3 또한 동일했다($r=.45, p<.001$). 4-6주 간 추적연구 결과 버전 2는 폭력성과 유의미하게 관련이 되어있고($r=.36, p<.001$), 버전 3 또한 동일했다($r=.38, p<.001$). 이 표본 집단에서는 버전 2와 3의 존재 유무 점수의 상관관계가 높았다는 것을($r=.91, p<.001$) 기억할 필요가 있다.

또한 본 연구에서 버전 2와 3의 SRRs을 비교한 결과 둘 다 동등하게 폭력성을 예측하는 것으로 나타났다. 6-8개월간 후속연구에서 두 도구의 폭력성과의 상관관계는 .41($p<.001$)이었고 4-6주 후 후속연구에서는 .42($p<.001$)로 나타났다. 이는 본 연구에서 버전 2와 버전 3의 SRRs이 근본적으로 동일하다는 것을 고려해보면 놀라운 결과는 아니다($r=.98, p<.001$).

Blanchard와 Douglas(2011)는 위험성 존재 유무, 관련성 평가, 그리고 신체적 폭력의 연관성에 대해 사후예측분석을 실시하였다. 관련성 평가의 AUC 값은 .81이었던 반면, 위험성 존재 유무는 .72였다. 이렇듯 관련성 평가가 신체적 폭력과 연관성이 좀 더 높은 패턴이 각각의 척도에서 나타난다. 이러한 분석이 사후 예측적이지만, 위험성 존재 유무와 관련성 평가가 폭력성과 관련되어있다는 초기 지표를 제공한다.

II. 연구방법

2.1 평가자 구성 및 훈련절차

본 연구를 위해 평가자 35명으로 구성된 연구원을 집체 훈련하였다. 평가자들은 크게 교정심리학자와 범죄심리학자, 석사과정 이상의 대학원생 연구원으로 구성되었으며 본 연구를 위해 HCR-20 버전 3.0 원저자(SDH, KJD)를 초빙하여 이론, 실습, 슈퍼비전이 차례대로 이루어졌다. 구체적으로 교정심리학자는 10명, 범죄심리학자는 15명, 대학원 석사과정 이상 연구원은 10명으로 이루어졌다. 교정심리학자의

경우 임상현장에서 심리치료와 위험성 평가를 담당하고 있으며 근무기간은 1년에서 10년 사이로 다양했다. 범죄심리학자로는 최소 2년에서 10년 사이 경찰청에서 범죄 분석 업무와 피의자 면담, 위험성평가를 담당하였고 모두 심리학이나 사회학을 전공하여 석사학위 이상을 소지하고 있었다.

구체적으로 원저자가 진행하는 집중 워크숍 참여, 슈퍼비전을 받은 사례 실습 등을 통한 다양한 방법으로 훈련을 받도록 하였다. 본 연구에서 원저자(SDH, KJD) 두 명을 국내에 초청하여 5일간 평가자 훈련을 실시하였고 크게 다음의 필수 요소를 포함하였다. 첫 째, HCR-20의 기초 정보와 시행 주제에 강조점을 두며 HCR-20 매뉴얼 및 지침을 검토하였다. 둘째, HCR-20(버전 3.0) 매뉴얼 출간 후, 폭력위험성 평가와 관련해 문헌상 중요한 발전이 있었는지 검토하여 소개하였다. 세 번째, 문서 검토에 기초한 사례 연습(3건)을 실시하였다. 네 번째, 경험 있는 동료의 슈퍼비전 혹은 조언을 구하여 실제 사례평가가 이루어졌다. 마지막으로 추후 보수교육을 실시하였다. 크게 구조화된 전문가 판단(SPJ) 모델과 이론, 위험성평가 절차와 최근 문헌 검토에 8-16시간의 학습이 필요하였고 마지막 사례실습 등에 8-16시간이 필요하였다. 요약하면, HCR-20를 이용한 폭력 위험성 평가의 기본적인 훈련은 대략 16-32시간 (2일에서 4일 정도)의 학습시간을 가졌다.

2.2. 한국판 HCR-20 3판 번역절차

20개 위험요인과 하위위험요인 번역을 위한 절차는 다음과 같다. 먼저 번역을 위한 위원회를 구성하였다. 위원회 리더는 1저자(SJH)였고 관련 위원으로는 임상심리학, 사회심리학, 범죄심리학, 계량심리 전공 자 등 7명으로 구성하였다. 1차적으로 용어에 대한 정확한 개념 파악을 위해 1저자가 원저자 2명(SDH, KSD)과 함께 용어 정의와 확인 절차를 위해 10여 차례 밴쿠버에서 회의를 진행하였다. 위험요인의 개념을 정확하게 반영하는 수준에서 번역 리더인 1저자가 이를 1차적으로 번역하였다. 이 후 7명의 번역위원회에서 번역된 용어에 대한 토의 과정을 7차례 거쳐 최종적으로 1차 번역본을 완성하였다. 이 후 해당분야와 상관없는 언어심리학자, 사회심리학자, 그리고 대학원생에게 1차 번역본에 대한 이해와 개념, 한국과의 언어학적

차이 등에 대한 피드백을 받았다. 이후 최종적으로 이를 반영하여 2차 번역본을 완성하였다. 마지막으로 2명의 전문번역가에게 2차 번역본에 대한 역본역을 의뢰하였고 역본역을 두고 다시 1저자와 원자 간에 언어학적 그리고 의미론적 차이가 나는 용어를 중심으로 합일점을 찾고자 하였다. 이후 이를 반영한 수정본에서 재차 역본역을 [1차 역본역과 상관없는] 2명의 번역 전문가에게 의뢰하였다. 이런 과정이 원저자와 1저자 간에 문제가 없다고 판단할 때까지 계속 이어졌다. 최종적으로 본 연구에서 사용된 최종번역본을 완성을 하였다.

2.3 연구절차

수용시설에 수용중인 수용자를 대상으로 이들의 자발적 지원을 받아 최종 면담 대상자로 선정하였다. 이 과정에서 정신병적 증세가 심하거나 중증 정신지체, 급격한 기분장애를 보이는 경우는 해당 교도관과 담당자의 도움을 받아 연구 참여에서 제외시켰다. 크게 연구 절차는 두 기관인 교정청과 경찰청으로 나누어 진행되었다. 지방 교정청은 담당 교정심리학자가, 경찰청은 해당 범죄심리학자가 주도적으로 면담을 진행하였고 필요한 경우 보조 면담자로 대학원 석사과정 중에 있던 연구원 1명을 추가하여 총 2명이 공동으로 면담을 진행하도록 하였다. 참가자에게 면담 전에 연구 목적에 대해 설명하였고 개인정보의 활용방식과 향후 권리 제한에 대해 공지한 이후 수용자 연구 참여 동의서에 서명하도록 하였다. 면담 전·후 그리고 종결 후라도 언제든지 연구 참여 철회와 거부 권리를 고지하였다. 주 면담자 1명과 보조 면담원 1명 등 총 2명이 진행하였고 이 중 한명은 HCR-20와 관련된 질문을 하였고 나머지 한명은 추가적으로 LSI-R, KORAS-G, PCL-R, 그리고 CAPP과 관련된 세부 질문을 하여 평가를 완료하였다. 이 후 면담에 참여하지 않은 대학원 석사과정 연구원 2명이 같은 사례에 대해 같은 정보(주면담자에게 전달받은)를 바탕으로 HCR-20 평가지를 동일하게 블라인드 상태에서 독립적으로 평가하도록 하였다. 전체적으로 요약하면, 각 사례에 총 4명의 HCR-20 평가가 이루어졌지만 2명은 면담에 직접 참여하였고 2명은 전달 받은 자료를 바탕으로 최종 평가가 이루어졌다. 그리고 면담에 참여 했던 1명은 동시에 동시타당도를 알아보기 위해 사용된 4개의 척

도(CAPP, IRS-R, KORAS-G, PCL-R)에 대해서도 평가를 추가적으로 완료하였다. 면담을 종결하면서 연구 참여자에 대해서는 3만원이 보상으로 주어졌다. 본 연구 과정에 대한 전반적인 절차와 진행과정은 연구책임자가 최종 관리 통제하였다.

2.4 참가자

전국 18개 교도소와 구치소, 유치장 등 수용시설(예, 안양(3), 원주(5), 대구(6), 대전(6), 의정부(2), 서울(22), 수원(5), 순천(3), 여주(1), 청주(11), 포항(5), 해남(1), 밀양(1), 공주(1), 광주(5), 경북(5), 제주(1), 인천(2)) 중심으로 총 85명의 남성 수용자를 대상으로 타당화 연구가 이루어졌다. 이 중 연구과정 중 면담을 포기한 경우, 자료가 부족한 경우, 정신이상 등으로 인해 진술내용의 신빙성이 떨어지는 경우, 그리고 종료 이후 철회한 경우 등 총 5건을 본 연구에서 제외하여 최종적으로 80건이 분석에 사용되었다. 조사 대상자들의 평균 나이는 38.78세($SD=9.86$; 범위, 18-78) 이었고 모두 남성 범죄자였다. 본 범죄 유형으로는 살인(미수, 영아살해 등 포함) 18건(22.5%), 상해/폭력 18건(22.5%), 강도/절도 6건(7.5%), 성폭력(미성년자성범죄도 포함) 23건(28.7%), 사기 4건(5%), 마약관리법위반 1건(1.2%), 방화 5건(6.2%), 기타(공무집행방해 등) 5건(6.2%)이었다. 과거범죄 평균은 5.5회($SD=2.6$)였다. 범죄자의 41%가 미혼상태였고 38.2%가 본 건으로 수용되기 전에 무직상태였다. 77%가 과거 범죄로 인해 교도소에 구속수감(징역형)된 적이 있었고 평균 89%가량이 범죄경력 1회 이상 존재했다.

2.5 평가도구

2.5.1 한국판 사이코패스종합평가(Comprehensive Assessment of Psychopathy Personality Disorder; 이하 CAPP)

사이코패스종합평가(이하 CAPP; Cooke et al., 2004; 한국판, 서종한, 정혜정, 김경일, 2018, 출판 중)는 사이코패스와 관련된 모든 영역의 증상을 포괄적으로 잡아내기 위해 고안된 사이코패스 종합평가도구이다. CAPP은 비교적 근래 개발된 도구

로 현재 계량적 평가와 국제적인 연구가 함께 축적되고 있는 상태이다. 지금까지 영국, 캐나다 그리고 미국 등 서구 국가를 중심으로 실증적 연구들이 축적되어 오고 있다. CAPP은 사이코패스 특징을 종합적이면서 역동적인 수준에서 그 증상을 잡아내고자 하는 의도를 갖고 만들어졌다. 이런 특징은 특정 영역에 한정된 활용보다는 다양한 임상 치료현장에서 유용하게 활용할 수 있도록 하는데 그 목적이 있다. 뿐만 아니라 시간에 따른 사이코패스 증상과 기능상의 변화를 민감하게 잡아내도록 하였다. CAPP은 6가지 영역별 사이코패스 개념 모델을 가지고 있다. 6개의 CAPP 영역인 애착, 행동, 인지, 지배성, 정서성, 자기 영역은 크게 33개의 사이코패스 증상을 포함하고 있다. 각 증상은 7점 척도로 이루어져 있고 최고점수는 198점이다. 본 연구에서는 각 증상은 전생애를 기반으로 채점되었고 따라서 일정 기간에 따른 증상의 변화는 조사하지 않았다. 본 연구에서는 한국판 CAPP에 따라 평가가 이루어졌다(자세한 내용은 서종한, 정혜정, 김경일, 2018, 출판 중).²³⁾

2.5.2 한국범죄자위험성평가-일반(Korean Offender Risk Assessment General; 이하 KORAS-G)

정진수, 이수정, 황의갑과 박선영(2010)은 죄명에 관계없이 비교적 포괄적으로 사용할 수 있는 한국판 위험성 평가도구를 개발하였다. KORAS-G라고 불리는 이 도구는 205명의 강력범죄자를 대상으로 하여 위험요인을 추출하였고 신뢰도와 타당도 검증을 통하여 최종 17개의 위험요인을 선택하였다. KORAS의 17개 문항은 다음과 같다. 피의자의 나이, 교육 수준, 혼인 상태, 최초 경찰 입건 나이, 청소년 시설 수용 경험, 과거 범죄 점수, 강력범죄 횟수, 동종 전과 여부, 처분기간 동안의 문제행동, 학령기 문제행동, 알코올 혹은 약물 사용 등이다. 이후 국내에서 타당도와 신뢰도 연구가 추가적으로 이루어졌다(이수정, 고려진, 2011; 최혜림, 2011). KORAS-G의 각 문항에 대한 평가는 수사자료, 범죄경력조회, 공소장 판결문 등 공식 기록 문서를 검토하여 정보를 획득하고 평가대상자 등과의 면담 자료 등 모든 정보 출처를 통해 평가하는 것을 원칙으로 한다. 피의자 나이, 본 범행점수, 반사회

23) 서종한, 정혜정, 김경일 (2018). 한국판 사이코패스종합성격장애모델 연구와 타당화. 출판 중.

적 성향은 4점 척도(0-3)로 최초 경찰입건 나이, 강력범죄 횟수, 처분기간 동안의 문제행동, 기타 취약요인 등은 3점 척도(0-2)로 그 이외의 요인들은 모두 2점 척도(예-1, 아니오-0)로 이루어져 있다.²⁴⁾

2.5.3 Level of Service Inventory-Revised(이하 LSI-R)

LSI-R(Andrew & Bonta, 2000)은 분류심사관이 수용자의 구금 수준 및 교도소의 적정 보안수준을 결정하는 분류과정에서 면담방식에 의해 적용하도록 되어 있다. LSI-R은 10개의 하위범주로 구성되어 있으며, 총 54개의 문항으로 이루어져 있다. 각 질문은 ‘예’ 또는 ‘아니오’로 대답하는 문항(2점 척도)과 0에서 3의 4점 척도로 구성되어 있다. 구체적으로 ‘0’은 반드시 개선되기를 바라는 매우 불만족스러운 상황, ‘1’은 개선이 되어야 한다고 느끼는 상대적으로 불만족스러운 상황, ‘2’는 개선의 여지가 있지만 상대적으로 만족스러운 상황, ‘3’은 개선의 여지가 전혀 없는 만족스러운 상황을 의미한다. 각 문항에 부정적으로 대답한 점수를 합하면 위험성 점수가 산출되도록 구성되었고 0점에서 54점의 위험성 점수가 산출된다. 전체 점수가 높을수록 위험성이 높은 수용자로 분류된다.

국내연구에서는 이민식과 김혜선(2009)등의 연구에 따르면 폭력범의 경우 비폭력범보다 유의미하게 높은 LSI-R 점수를 보였다. 정지숙 등(2011)의 연구에서도 높은 폭력성을 보이는 고위험군의 경우 그렇지 않은 집단보다 유의미하게 높은 LSI-R 점수를 보였다. 특히 범죄력, 교육 및 고용, 알코올 및 약물 문제 등에서 높은 점수를 보였다. 또한 재범위험성을 살펴본 결과, AUC=.769를 보였다. 국외에서는 Furguson 등(2009)은 호주 멜버른 정신병원을 퇴소한 집단을 대상으로 타당도를 검증한 결과 전체 집단의 AUC가 .67, 폭력 범죄 .60으로 나타나 국내외가 유사한 수준의 예측력을 보였다고 할 수 있다.²⁵⁾

24) 최혜림 (2011) 초기 한국판 재범위험성 평가도구 : 일반위험요인(KORAS-G)의 신뢰도 분석. 경기대학교 석사논문.

25) Andrews, D. A., & Bonta, J. (2000). The level of service inventory-revised. Toronto, Canada: Multi-Health Systems.; Andrews, D. A., Bonta, J., & Wormith, S. J. (2000). Level of service/case management inventory: LS/CMI. Toronto, Canada: Multi-Health Systems.

2.5.4 한국판 Psychopathy Checklist-Revised (이하 PCL-R)

PCL-R(Hare, 2003; 한국판, 조은경 & 이수정, 2008)은 20개 항목으로 구성된 척도이며 반구조화된 질문지를 이용하여, 전문가 훈련을 받은 평가자가 실시하도록 되어 있다. 크게 2요인 4단면으로 구성되어 있다. 1요인은 대인관계와 정서성, 2요인은 생활양식과 반사회성으로 이루어져 있다. 단면 1 대인관계 문항은 피상적 매력, 과도한 자존감, 병적인 거짓말, 남을 잘 속임/ 조종함이며, 단면 2 정서성은 후회 혹은 죄책감 결여, 얕은 감정, 냉담/공감 능력 결여, 자신의 행동에 대한 책임감 못느낌, 단면 3 생활양식은 자극욕구/쉽게 지루해함, 기생적인 생활방식, 현실적이고 장기적인 목표부재, 충동성, 무책임성, 마지막으로 단면 4 반사회성은 행동통제력 부족, 어릴 때 문제 행동, 청소년 비행, 조건부 석방 혹은 유예의 취소, 다양한 범죄력 등으로 구성되어 있다. 평가자들은 각 문항을 3점 척도를 기준으로 평가하며 ‘0’점은 해당되지 않음, ‘1’점은 어떤 상황에서도 아마도 해당됨, ‘2’점은 해당됨으로 채점하며, 점수 분포는 0점에서 40점의 범위가 된다.²⁶⁾

2.6 자료분석

개별 HCR-20 항목에 대한 채점자간 신뢰도는 ICC_c(개별 점수 간 상관관계)와 ICC_s(평가자 평균 간 상관관계)를 통해 평가하였다. 백분율(중간값과 평균) 일치성도 함께 조사했다. 2, 3 그리고 6 단계에서 표준화된 HCR-20 위험성 요인에 대해 개별적으로 분석이 이루어졌고 추가적으로 4와 5단계에서 부분적인 항목도 함께 분석했다.

ICC 모델은 평균 변동성 추정치에 근거하고 있고 집단 간 분산과 전체 분산 간 비율로 개념화 할 수 있다. ICCs는 모든 판단과 항목에 걸쳐 나타나는 전체 분산에 똑같은 테스트 항목에 대한 다른 판단의 변산도(variability)를 비교하여 정정기회 동의 척도(*chance-corrected agreement*)를 제공해 준다. 특히 2명 이상의 채점자가 있거나 데이터가 범주화 속성을 띠는 경우 신뢰도를 평가하는데 추천되는 통계적 방

26) 조은경, & 이수정. (2008). PCL-R 전문가 지침서: 한국판 표준화. 서울: 학지사.

식이다(Uebersax, 2009). ICC가 가중 Kappa와 산술적으로 동일한 값을 산출 (Cicchetti & Sparrow, 1981; Landis & Koch, 1977)할 뿐 아니라 임상-법 연구 분야에서 더 일반적으로 사용하고 있기에 선택되었고 이후 다른 연구 결과와도 비교할 수 있을 것으로 보인다. ICC 통계치에 대한 몇 가지 형태가 존재하고(McGraw & Wong, 1996) 각각이 다른 연구 설계에 적합할 수 있다. 모든 채점자가 각 사례를 평가했고 채점자와 사례 모두 더 큰 전집에서 표집 될 수 있는 것으로 보이기 때문에 ICC_s(이원무선효과: two-way random effect)가 여기서는 더 적합할 것으로 보인다. 각 사례와 채점자에 대한 무선효가 순수하게 이루어지지 않았지만, 본 연구가 유목 내 평균효과 혹은 고정효과(population average effect or fixed effect)와 반대로 대상자-특정 효과(subject-specific effect)와 관련 있기 때문에 무선효과를 가정했다. 또한 가정과 해석이 다르기는 하지만 다른 이원 ICC 모델에 대한 수치가 동일하다(Nichols, 1998). ICCs는 절대 동의서(absolute agreement)와 단일 측정(single measures)으로 나누어져 계산된다.²⁷⁾

HCR-20의 채점자 간 신뢰도를 평가했던 이전 연구를 따라 ICC 해석 가이드라인은 Fleiss(1981)을 따랐다. 기준은 다음과 같다. ICC<.39인 경우 나쁨, .40 < ICC < .59인 경우 보통, .60 < ICC < .74인 경우 좋음, .75 < ICC인 경우 아주 좋음으로 평가했다. Landis와 Koch(1977)는 0이하 이면 나쁨, 0-.20은 약간, .21-.40은 적정한, .41-.60은 보통, .61-.80 좋음, 그리고 .81-1.00은 거의 완벽함으로 평가할 수 있다. 백분율 일치성도 실용적 수준에서 대략적인 일치성에 대한 정보를 제공해 준다 (Uebersax, 2009). 본 연구에서는 두 가지의 백분율 즉 평균과 중앙값(가까운 정수로 반올림 한 값)에 대한 일치성을 살펴보았다. 각 일치성 측정은 각각 평균과 중앙값과 일치하는 관찰비율을 나타낸다고 볼 수 있다. 백분율 일치성은 각 항목과 전체 사례 및 채점자를 대상으로 계산되어졌다. 결측치는 모두 계산에 조정하여 처리했다.

27) Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *American journal of mental deficiency*; Uebersax, J. (2009). Raw agreement indices. URL: <http://ourworld.compuserve.com/homepages/jsuebersax/raw.htm> Accessed, 10.

Ⅲ. 연구결과

3.1 HCR-20채점자 간 신뢰도

3.1.1 HCR-20 단계 2와 3의 채점자간 신뢰도

〈표 2〉 HCR-20 위험요인 존재(유무)와 관련성 채점 ICCs

HCR-20 척도	존재		관련성	
	ICC ₁	ICC ₂	ICC ₁	ICC ₂
단계 2 & 3: 이력척도	.732(.650-.804)	.916(.881-.943)	.738(.657-.809)	.919(.885-.944)
단계 2 & 3: 임상척도	.898(.859-.928)	.972(.961-.981)	.873(.827-.910)	.965(.950-.976)
단계 2 & 3: 위험관리	.861(.811-.902)	.961(.945-.973)	.830(.772-.879)	.951(.931-.967)
단계 2 & 3: 전체척도점수	.790(.721-.849)	.938(.912-.957)	.859(.809-.901)	.961(.944-.973)

Note. 95% confidence intervals in parentheses.

<표2>은 HCR-20 2, 3단계에 대한 전체 ICC를 나타내고 있다. Fleiss(1981)의 기준에 따르면 이력척도(ICC₁=.732, ICC₂=.916)는 “좋은” 신뢰도(.60<ICC<.74)를 보여주고 있고 그 이외 척도에서는 존재(유무)와 관련성에 상관없이 “아주 좋은” 신뢰도(ICC>.75; Fleiss, 1981)를 보이고 있다(존재-임상척도; ICC₁=.898, ICC₂=.972: 위험관리척도; ICC₁=.861, ICC₂=.961).

〈표 3〉 HCR-20 이력척도에서의 ICCs

HCR-20 항목		ICCs			
		사례 수	평정자수	ICC ₁	ICC ₂
영역 A: 이력척도					
1. 폭력	존재	80	4	.777(.673-.861)	.903(.845-.964)
	관련성	80	4	.335(.221-.461)	.669(.531-.774)
2. 기타 반사회적 행동	존재	80	4	.785(.715-.845)	.936(.909-.956)
	관련성	80	4	.782(.711-.843)	.935(.908-.956)
3. 대인관계	존재	80	4	.921(.890-.945)	.979(.970-.986)
	관련성	80	4	.761(.685-.827)	.927(.897-.950)
4. 고용	존재	80	4	.491(.377-.605)	.794(.708-.860)
	관련성	80	4	.599(.495-.697)	.857(.797-.902)

HCR-20 항목		ICCs			
		사례 수	평정자수	ICC ₁	ICC ₂
5. 물질사용	존재	80	4	.780(.709-.842)	.934(.709-.842)
	관련성	80	4	.813(.749-.866)	.946(.923-.963)
6. 주요 정신장애	존재	80	4	.887(.845-.921)	.969(.956-.979)
	관련성	80	4	.842(.787-.889)	.955(.936-.970)
7. 성격장애	존재	80	4	.727(.643-.800)	.914(.878-.941)
	관련성	80	4	.786(.715-.846)	.936(.910-.956)
8. 외상경험	존재	80	4	.730(.648-.803)	.915(.880-.942)
	관련성	80	4	.776(.704-.839)	.933(.905-.954)
9. 폭력적 태도	존재	80	4	.538(.428-.646)	.823(.750-.879)
	관련성	80	4	.366(.251-.491)	.698(.573-.794)
10. 치료 혹은 관리감독 반응	존재	80	4	.727(.644-.800)	.914(.878-.941)
	관련성	80	4	.837(.767-.877)	.950(.930-.966)

Note. 95% confidence intervals in parentheses.

<표 3>은 각 영역 내 세부항목에 대한 채점자간 신뢰도를 구한 것이다. Landis와 Koch(1977)의 기준에 따르면 3개 항목을 제외 하고는 존재와 관련성에 상관없이 최소한 “좋은”(.60<ICCs<.74) 정도의 신뢰도를 보여주고 있다.

하지만 고용(ICC1=.491, ICC2=.764)과 폭력적 태도(ICC1=.538, ICC2=.823) 항목에서 ‘존재’의 경우 “중간”(.60<ICCs<.80) 정도의 일치성을 보여 주고 있고 폭력적 태도(ICC1=.366, ICC2=.698)의 관련성의 경우 “보통”(.21<ICCs<.40) 정도의 일치성을 보여 주어 상대적으로 낮은 일치성을 나타냈다.

<표 4> 임상척도와 위험관리척도 ICCs

HCR-20 항목		ICCs			
		사례 수	평정자수	ICC ₁	ICC ₂
영역 B: 임상척도					
11. 통찰력	존재	80	4	.708(.621-.786)	.907(.868-.936)
	관련성	80	4	.729(.646-.802)	.915(.880-.942)
12. 폭력사고와 의도	존재	80	4	.764(.689-.829)	.928(.899-.951)
	관련성	80	4	.769(.695-.833)	.930(.901-.952)
13. 주요 정신장애증상	존재	80	4	.872(.826-.910)	.965(.950-.976)
	관련성	80	4	.821(.759-.872)	.948(.927-.965)

HCR-20 항목		ICCs			
		사례 수	평정자수	ICC ₁	ICC ₂
14. 불안정성	존재	80	4	.812(.748-.865)	.945(.922-.963)
	관련성	80	4	.814(.751-.867)	.946(.923-.963)
15. 치료 혹은 관리감독 반응	존재	80	4	.756(.679-.823)	.925(.894-.949)
	관련성	80	4	.734(.653-.806)	.917(.883-.943)
영역 C: 위험관리척도					
16. 전문서비스와 계획	존재	80	4	.700(.611-.779)	.903(.863-.934)
	관련성	80	4	.786(.716-.846)	.936(.910-.957)
17. 주거환경	존재	80	4	.746(.666-.899)	.921(.889-.946)
	관련성	80	4	.718(.633-.793)	.911(.873-.939)
18. 개인적지지	존재	80	4	.843(.787-.888)	.955(.937-.970)
	관련성	80	4	.855(.804-.898)	.959(.943-.972)
19. 치료 혹은 관리감독 반응	존재	80	4	.782(.711-.843)	.935(.908-.955)
	관련성	80	4	.809(.745-.864)	.944(.921-.962)
20. 스트레스 혹은 대처 반응	존재	80	4	.640(.541-.731)	.877(.825-.916)
	관련성	80	4	.581(.475-.682)	.847(.784-.896)

Note. 95% confidence intervals in parentheses.

<표 4>는 임상척도와 위험관리 척도 세부항목에 대한 ICC를 보여주고 있다. ‘존재’의 경우, 폭력사고와 의도, 주요 정신장애증상, 불안정성, 치료 혹은 관리감독 반응(임상척도), 주거환경, 개인적지지, 그리고 치료 혹은 관리감독 반응(위험관리척도) 등 7개 항목에서 “아주 좋은”(ICCs>.75) 정도의 일치성을 보여 주고 있고 그 이외의 나머지 3개 항목의 경우 “좋은”(.60<ICCs<.74) 범위의 일치성을 보여주어 전반적으로 만족할 만한 신뢰도였다(Fleiss, 1981).²⁸⁾

‘관련성’의 경우, 폭력사고와 의도, 주요 정신장애증상, 불안정성, 전문서비스와 계획, 개인적지지, 그리고 치료 혹은 관리감독 반응(위험관리척도) 등 6개 항목에서 “아주 좋은”(ICCs>.75)범위의 일치성을 보였고 3개 항목에서는 “좋은”(.60<ICCs<.74) 범위의 일치성을 보인 반면 1개의 항목에서 “중간”(.40<ICCs<.59) 정도의 일치성을 보여 ‘존재’와 유사한 일치성을 나타냈다(Fleiss, 1981).

28) Fleiss, J. L. (1981). Balanced incomplete block designs for inter-rater reliability studies. *Applied Psychological Measurement*, 5(1), 105-112.

3.1.2 HCR-20 단계 6의 채점자간 신뢰도

〈표5〉 전문가 요약 판단에서의 ICCs

	ICCs			
	사례수	평정자수	ICC ₁	ICC ₂
단계 6: 전문가 요약 판단				
사례 우선순위	80	4	.725(.641-.800)	.913(.877-.941)
심각한 신체 위해 위험성	80	4	.612(.492-.696)	.855(.795-.902)
긴급조치의 필요성	80	4	.726(.643-.801)	.914(.878-.941)
제기된 다른 위험성	80	4	.783(.713-.844)	.935(.908-.956)
사례검토	80	4	.712(.623-.792)	.903(.825-.912)

Note. 95% confidence intervals in parentheses.

<표 5>는 5가지 전문가 요약 판단 세부 항목에 대한 신뢰도를 보여주고 있다. 제기된 다른 위험성 항목은 “아주 좋은” 수준의 일치성(ICC₁=.783, ICC₂=.935)을 보였고 그 이외의 4개 항목에서는 “좋은” 수준의 일치성을 보여 전문가들 간의 요약 판단 신뢰도가 높았다(Fleiss, 1981). Landis와 Koch (1977)의 기준을 적용해 보아도 비슷한데 심각한 신체 위해 위험성(“중간”; ICC₁=.612, ICC₂=.855)을 제외한 나머지 영역에서 모두 “좋은” 일치성을 보여 주고 있음을 알 수 있었다.

3.2 HCR-20 공인타당도

3.2.1 HCR-20 요인과 척도 간 전체 상관관계

〈표 6〉 HCR-20와 척도들 간 상관관계 (r)

HCR-20	CAPP 전체점수	PCL-R 전체점수	KORAS-G 전체점수	LSI-R 전체점수
단계 2 & 3: 이력척도(존재)	0.41	.458**	.415**	.577**
단계 2 & 3: 이력척도(관련성)	-.177	.436**	.339**	.675**
단계 2 & 3: 임상척도(존재)	.344*	.183	-.085	.176
단계 2 & 3: 임상척도(관련성)	.405**	.239*	-.099	.182
단계 2 & 3: 위험관리(존재)	.268*	.259*	-.003	.417**
단계 2 & 3: 위험관리(관련성)	.445**	.262*	-.004	.361**
단계 2 & 3: 전체점수(존재)	.266*	.414**	.147	.477**

HCR-20	CAPP 전체점수	PCL-R 전체점수	KORAS-G 전체점수	LSI-R 전체점수
단계 2 & 3: 전체점수(관련성)	.326 [*]	.389 ^{***}	.116	.515 ^{***}

Note. ^{*} $p < .05$, ^{**} $p < .01$, ^{***} $p < .001$

<표 6>는 다른 척도간에 상관관계를 나타내어주고 있는 표이다. CAPP, PCL-R, LSI-R 전체점수 간에 유의미한 수준에서 상관관계를 보여주고 있다. 하지만 KORAS-G의 경우 이력척도(존재와 관련성 모두)를 제외하고는 다른 HCR-20 하위 척도 간에는 유의미한 관련성을 보이지 않았다. KORAS-G는 정적 위험요인만을 포함하고 있는 점을 감안한다면 충분히 예측할 수 있는 점이다.

3.2.2 HCR-20 척도와 CAPP 하위영역 간 상관관계

<표 7> HCR-20 척도와 CAPP 하위 6개 영역 간 상관관계 (n)

HCR-20 척도		CAPP 하위 영역					
		애착	행동	인지	지배성	정서성	자기 전체
이력 (문제이력)	존재	-.177	.258 [*]	.468 ^{***}	.129	-.455^{***}	-.163
	관련성	-.214	.006	.296 [*]	-.132	-.359^{**}	.340 [*]
임상 (최근문제)	존재	.245 [*]	.355 ^{***}	.673 ^{***}	.071	.059	.103
	관련성	.216	.308 [*]	.644 ^{***}	.183	.206 [*]	.199
위험관리 (미래문제)	존재	.491 ^{***}	.231	.401 ^{***}	-.019	.180	.028
	관련성	.612 ^{***}	.296 [*]	.541 ^{***}	.063	.423 ^{***}	.167
전체	존재	.188	.358 [*]	.659 ^{***}	.087	-.188	-.016
	관련성	.264	.288 [*]	.687 ^{***}	.076	.131	.014

Note. ^{*} $p < .05$, ^{**} $p < .01$, ^{***} $p < .001$

<표 7>은 HCR-20 하위척도와 CAPP 하위영역 간 상관관계 정도를 보여주고 있다. 이력 척도의 경우 ‘존재’와 ‘관련성’에 상관없이 인지영역과 정적 상관을 보여주는 반면 정서성과는 오히려 부적상관을 보여주고 있다. 임상 척도의 경우 행동, 인지영역과 유의미한 수준에서 정적 상관을 보여주었다. 위험관리 영역에서는 ‘존재’

와 ‘관련성’에 관계없이 인지와 애착 영역 간에 유의미한 상관을 보였고 정서성과는 ‘관련성’에서만 높은 관련성을 보였다. 전체적으로 HCR-20는 CAPP 하위 영역중 인지, 행동, 애착 순으로 유의미한 관련성을 보이고 있다.

3.2.3 HCR-20 척도와 PCL-R 간 상관관계

〈표 8〉 HCR-20 항목과 PCL-R 하위요인 간 상관관계 (r)

HCR-20 척도		PCL-R					
		요인1 정서/대인	요인2 충동 /반사회성	단면1 대인관계	단면2 정서성	단면3 생활양식	단면4 반사회성
이력 (문제이력)	존재	.186	.533**	.109	.189	.519**	.318**
	관련성	.235*	.459**	.059	.314**	.506**	.217
임상 (최근문제)	존재	.170	.124	.088	.182	.274*	-.075
	관련성	.271*	.114	.120	.310**	.300**	-.116
위험관리 (미래문제)	존재	.325**	.125	.105	.411**	.221	-.022
	관련성	.389**	.077	.084	.532**	.155	-.031
전체점수	존재	.311**	.362**	.161	.334**	.452**	.120
	관련성	.353**	.285**	.102	.457**	.408**	.045

Note. * $p < .05$, ** $p < .01$, *** $p < .001$

〈표 8〉은 HCR-20 하위 척도와 PCL-R 하위 요인 간 상관관계를 보여주고 있다. 이력 척도의 경우 생활양식과 유의미한 상관관계를 보여주었고 정서성은 관련성과 반사회성은 존재 부분에서 관련성을 보였다. 임상영역의 경우 생활양식과 정적 관련성을 보여주었고 정서성과는 관련성 부분에서 유의미한 상관관계를 보였다. 위험 관리 척도의 경우 정서성 요인에서만 유의미한 수준에서 높은 상관을 보였다 전체 적으로 단면3 생활양식과 단면2 정서성 요인에서 HCR-20 척도와 높은 관련성을 보여준 반면 단면1 대인관계와는 관련성이 전혀 보이지 않았다.

3.2.4 HCR-20 척도와 LSI-R 하위요인 간 상관관계

〈표 9〉 HCR-20 척도와 LSI-R 하위요인 간 상관관계

HCR-20 척도		LSI-R										
		범죄 기록	교육 직업	재정 상태	가족 결혼	주거 환경	레저 레크.	동료 친구	음주 약물	정서 인성	태도 지향	전체
이력	존재	.175	.395*	.221	.248	.025	.357**	.230	.326*	.522**	.452**	.577**
	관련성	.139	.393*	.264	.374**	.143	.476**	.275*	.468**	.380*	.531**	.675**
임상	존재	-.104	.074	.075	.023	-.130	.230	.133	-.072	.591**	.348*	.176
	관련성	-.162	.091	.117	.054	-.115	.175	.167	-.063	.448**	.475**	.182
위험 관리	존재	.087	.341*	.244	.436**	.344**	.304*	.279*	.046	.049	.318*	.417**
	관련성	-.059	.234	.210	.516**	.342**	.297*	.275*	.033	.061	.286*	.361*
전체	존재	.104	.357**	.209	.328*	.122	.295*	.279*	.101	.379**	.478**	.477**
	관련성	-.024	.297*	.233	.383**	.130	.396**	.292*	.209	.381**	.550**	.515**

Note. * $p < .05$, ** $p < .01$, *** $p < .001$; 레크: 레크리에이션

<표 9>는 HCR-20 하위 척도와 LSI-R 하위 요인 간에 상관관계를 보여주고 있다. 이력척도의 경우 교육직업, 가족/결혼(관련성만), 레저/레크리에이션, 음주/약물 문제, 정서/인성, 태도/지향 등 6개 요인과 유의미한 관련성을 보였다. 임상척도의 경우 정서/인성과 태도지향 두 요인에서만 높은 관련성을 보였다. 위험관리 척도에서는 교육/직업(존재), 가족/결혼, 주거/환경, 레저/레크리에이션, 동료/친구, 태도/지향 등 6개 요인과 유의미한 관련성을 보였다. 관련성을 보이는 요인들 중 많은 부분이 위험관리 척도가 다루는 하위항목과 유사한 개념을 나타내기 때문이다.

3.2.5 HCR-20 전문가 요약판단 항목과 척도 간 상관관계

〈표 10〉 HCR-20 전문가 요약판단 항목과 다른 척도 간 상관관계 (r)

요약판단 항목	CAPP 총점	PLC-R 총점	KORAS-G 총점	LSI-R 총점
단계 6: 전문가 요약 판단				
사례 우선순위	.465**	.383*	.101	.418**
심각한 신체적 위해 위험성	-.299*	.174	-.051	.269*
긴급조치의 필요성	.223*	.374*	.169	.420**
제기된 다른 위험성	.070	.088	.054	.128

요약판단 항목	CAPP 총점	PLC-R 총점	KORAS-G 총점	LSI-R 총점
사례검토	.547**	.478**	.298*	.612**

Note. * $p < .05$, ** $p < .01$, *** $p < .001$

<표 10>은 전문가의 최종 요약 판단과 척도 간에 상관관계를 보여주고 있다. 사례우선순위는 KORAS-G을 제외한 모든 척도들에서 높은 상관관계를 보였다. 심각한 신체적 위해 위험성은 CAPP과는 부적관련성을 LSI-R과는 정적 관련성을 보였다. 긴급조치의 필요성에서는 KORAS-G을 제외한 모든 척도들 간에 유의미한 상관을 보였다. 제기된 다른 위험성은 상관을 보이는 척도가 없었다. 사례검토의 경우 모든 척도와 상관관계를 보였다. 전체적으로 사례 우선순위와 사례검토 주기가 다른 척도들 간 높은 관련성을 보인 반면 제기된 다른 위험성은 전혀 존재하지 않았다. KORAS-G의 경우 사례검토를 제외한 모든 전문가 요약 판단 항목과 관련성을 보여주지 않았다.

IV. 논의

HCR-20(Douglas et al., 2013) 버전 3은 SPJ 모델을 기반으로 버전 2.0에서 한층 더 이를 구현하였고 버전 2.0(Webster et al., 1997)의 고유한 특성들을 충실히 반영하여 업그레이드 된 것이다. 앞 서론에서 이미 버전 2.0과의 차이점에 대하여 설명한 바 있다. 본 연구는 버전 3.0에 대한 국내 타당화를 목적으로 하였고 국외에서는 활발하게 이용하고 있을 뿐 아니라 사법 정책적 관점에서 놀라운 도움을 주고 있는 도구를 소개하고자 하였다. 버전 3.0의 채점자간 신뢰도와 공인타당도가 많지 않은 시점에서 의미 있는 시사점을 줄 수 있으리라 본다.

4.1 채점자간 신뢰도

채점자간 신뢰도는 결과에 제시되어 있는 것처럼, 척도 수준의 경우 전체적으로

위험요인 존재(유무)와 관련성 여부에 상관없이 최소한 “좋은” 수준의(채점자간) 일치성을 보여 주었다. 다시 말해 해당 위험요인이 존재할 경우 그리고 미래와의 폭력 위험성을 연관 지어 볼 경우 모두 채점자들이 강한 일치성을 본질적으로 보여 주고 있음을 나타낸다고 볼 수 있다.

Landis와 Koch(1977)의 수준에 따르면 척도 간 하위항목 수준 중 ‘존재’의 경우 20개 항목 중 5개 항목이 “거의 완벽한” 수준의 일치성을 보였고 12개 항목이 “좋은” 수준의 일치성을 보였다. ‘관련성’의 경우에서도 ‘존재’와 비슷한 일치성을 보여 3개의 항목을 제외하고는 최소 ‘좋은’ 수준 이상의 일치성을 보였다.²⁹⁾

본 연구의 결과는 Douglas와 Belfrage(2014)의 HCR-20 버전 3.0 연구에서 제시한 채점자 간 신뢰도와는 비슷하거나 다소 낮은 결과를 보여 주었다. 또한 Douglas와 Reeve(2010)는 HCR-20 버전 2.0의 채점자간 신뢰도를 36개 연구를 기반으로 보고 하였는데 본 연구의 결과와 비교해 볼 때 유의미한 차이를 보이고 있지 않았다. 이 연구에서의 중앙 ICC 값은 전체점수 .85, 이력 .86, 임상 .72 그리고 위험관리 .68 수준을 보여 이 전 연구와 별반 차이가 없었으나 Douglas와 Belfrage(2014)의 결과(전체 .94, 이력 .94, 임상 .94, 위험관리 .75 versus 우리연구 결과; 전체 .79, 이력 .73, 임상 .89, 위험관리 .86)와 비교했을 때는 다소 낮은 점수를 보여 주었다. 그럼에도 불구하고 버전 2.0 연구에서 보여주었던 이전 연구(Robbe & Vogel, 2010; Belfrage & Douglas, 2012)와 버전 3.0 연구(Doyle et al., 2013; Strub & Douglas, 2009)와 대등한 수준의 일치성을 보이는 것으로 볼 때 국내에서는 본 연구의 결과만으로 일반화하기는 이르지만 신뢰도의 일관성이 수용할만한 것으로 받아들여질 수 있다.

위험성 요약 판단에서 전문가가 내린 향후 폭력성에 대한 결정의 신뢰도도 존중할(“좋은” 혹은 “아주 좋음”)만한 것으로 나타났다. Douglas와 Reeves(2010)의 연구에서 다루었던 9개의 샘플에서 보인 중앙값은 .65 수준이었고 가장 최근의 논문인 Douglas와 Belfrage(2014)의 연구 결과에서는 .75이거나 최소한 이를 초과하는 값을 보였다. 본 연구에서는 사례우선순위 .75, 심각한 신체 위험성 .61, 긴급조치

29) Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159-174.

.72, 제기된 다른 위험성 .78 등을 보면 이전 연구와 비슷한 수준을 보이고 있음을 알 수 있다.³⁰⁾

20개 항목에 대한 개별 하위항목에 대한 구체적인 채점자 간 신뢰도는 생략하였다. 하위항목에 대한 평정은 연구자의 환경과 조건에 따라 할 수도 있는 선택적 항목이기 때문이다. 각 항목에 대한 정확한 평정을 위해 원저자가 하위 항목을 두었기 때문에 평정을 하고자 한다면 전체 20개 항목에 대한 평정 후에 정보가 명확하고 충분한 상태에서 평정할 수 있을 것이다. 본 연구의 결과 또한 이전 연구의 결과(Douglas & Belfrage, 2014) 비슷하게 각 하위항목의 신뢰도는 개별 항목보다 일관되게 낮은 수준을 보일 것이다. 이는 정의상 하위 항목이 상위 항목보다 더 협소한 개념을 갖고 있고 각 세부 하위항목에 대한 구체적인 정보를 찾아내는데 어려움이 있을 수 있기 때문이다. 또한 평가자가 면담과정에서 세부 하위항목에 대한 구체적인 정보를 확인하는 과정을 놓칠 수도 있기 때문이다.

이전 연구와는 달리 본 연구에서는 비교적 ‘관련성’ 신뢰도도 ‘존재’의 신뢰도와 큰 차이 없이 수용할만한 일치성을 보여주었다. Douglas와 Belfrage(2014)의 연구에서 20개 항목 중 6개 항목에서 부적합한 신뢰도($ICC > .20$; Landis & Koch, 1977)를 보여 주었고 이 항목을 제외한 나머지 항목에서의 평균치도 “좋은”, “중간”, 그리고 “아주 좋음”에 고루하게 분포하고 있었다(Fleiss, 1981). 하지만 본 연구에서도 폭력(H1) ‘관련성’, 고용(H4), 폭력적 태도(H9), ‘관련성’ 그리고 스트레스 혹은 대처 반응(H20) ‘관련성’에서 모두 “중간” 점수를 보여주고 있어 향후 연구에서는 ‘존재’에 비해 상대적으로 낮은 일관성을 보이는 이유에 대해서 연구할 필요가 있다고 본다.

Douglas와 Belfrage(2014)는 폭력(H1) ‘관련성’의 경우 이전 폭력 존재가 어떻게 향후 폭력에 영향을 줄 수 있는지 그리고 실제 그 영향정도는 얼마인지에 대한 명확한 설명을 할 수는 없다고 지적한바 있다. 평가자도 과거의 폭력 존재 여부가 향후 폭력에 이르는 경로에 대해 불명확한 태도를 가지고 있을 뿐만 아니라 평가자가 과거 폭력성을 중요한 예측 인자로 고려하느냐 아니냐에 따라 그 정도가 달라질 수

30) Douglas, K. S., & Reeves, K. (2010). The HCR-20 violence risk assessment scheme: Overview and review of the research. In R. K. Otto & K. S. Douglas (Eds.), *Handbook of violence risk assessment* (pp. 180-185). New York: Routledge/Taylor & Francis Group.

있었다는 점이다. 폭력적 태도(H9)의 경우도 마찬가지다. 과거에 폭력적 행동이 있었다고 해서 현재 평가 대상자의 의도와 태도를 명확하게 견지하고 있느냐는 의문스러울 수밖에 없다. 폭력적 태도와 의도(H9)가 있다고 평가자가 판단하는 과정에서 그 정도와 범위가 어느 정도까지 항목에 반영해야 될지도 의문스러울 수밖에 없다. 특히 순응도와 반응도 항목(H10)에서도 낮은 일치성을 보이는데 평가자의 수련과 전문가적 역량에 따라 대상자 치료 순응도와 반응도를 다르게 해석하거나 이해할 가능성이 있다. 하지만 이 항목이 낮은 수치를 보이는 이유에 대해 과거 연구에서 명확하게 설명하지 못하고 있기 때문에 향후 국내연구에서 좀 더 면밀한 연구가 뒤따라야한다고 본다.

하지만 본 연구에서의 ‘관련성’에서의 채점이 이전 연구보다 신뢰도가 높게 나타나는 이유를 살펴보면, 전체 채점자들에게 집체교육과 보수 교육, 이 후 사례 슈퍼비전이 동반된 것이 이유일 수 있다. 또한 대상자에 대한 직접 조사 과정이 가장 큰 원인 중 하나라고 보인다. 직접 조사 과정에서 각 항목 뿐 아니라 하위항목에 대한 직/간접적인 질문과 [재]확인 과정을 통해서 함께 참여한 나머지 평가자들이 쉽게 향후 폭력성과 연관성을 일관되게 평가한 것으로 보인다.

전체적으로, 관련성 항목은 국내에서 실시된 성폭력위험성프로토콜(서종한 등, 2017) 연구에서도 알 수 있듯이 높은 신뢰도를 보이고 있다(Hart et al., 2003; Hart & Boer, 2010). 관련성 항목은 전문가 지침서에 잘 설명되어 있지만, 관련성은 본질적으로 향후 위험성 개념화, 시나리오와 관리계획 등을 위해 평가되고 있다. 이런 과정은 다양한 이론과 개념적 모델에 근거하여 이루어지고 있는데 모델에 따라서 각 위험요인이 다른 관련성을 갖고 나타날 수도 있기 때문에 낮은 신뢰도를 보일 수도 있다. 혹은 일부 평가자들이 위험요인들 간 원인분석 혹은 인과적 연결을 잇는 과정에서 특정 위험요인의 관련성을 위계적으로 접근하여 하위 위험요인으로 통합하거나 구분하고자 하는 과정에서 발생할 수도 있다.

4.2 공인타당도

HCR-20 버전 3.0에 대한 공인타당도 연구는 거의 이루어지지 않았다. 다만

Douglas와 Belfrage(2014)의 연구에서는 이전 버전인 2.0과의 관련성을 조사하였다. 그 결과 전체적으로 높은 상관을 보였다. 다른 연구에서도 비슷한 결과를 보였는데 노르웨이 범법 환자 20명: $r=.84$ (Eidhammer et al., 2013), 스웨덴 범법 환자 35명: 기관 내 평가=.85와 지역사회 평가=.90(Belfrage & Douglas, 2012), 캐나다 민간 정신질환자 및 범죄자 43명: .86(Blanchard & Douglas, 2011), 네덜란드 범법 환자 86명: .93(de Vries Robbé & de Vogel, 2010)으로 높은 상관관계를 보였다.

PCL-R 혹은 PCL: SV과의 동시타당도 연구를 살펴보면, Blanchard와 Douglas(2011)의 연구에서 HCR-20^{V3}의 위험성 존재(유무), 관련성, 그리고 SRR은 Psychopathy Checklist: Screening Version (PCL: SV; Hart, Cox, & Hare, 1995)와 $r=.71$ (관련성), $r=.79$ (위험성 존재), $r=.82$ (SRRs)로 높은 상관관계를 보였다. Smith와 Edens(2013)는 텍사스 교도소에서 판결을 기다리고 있는 범죄자 32명을 대상으로 HCR-20^{V3} 원고 3과 PCL-R 점수의 동시 타당도 분석을 실시하였다. HCR-20^{V3}을 사용한 두 가지 도구의 총점 상관관계는 .66($p<.001$), 위험관리(기관 내 평가) .70($p<.001$). 위험관리(지역사회 평가)로 나타났다. PCL-R과 이력(H), 임상(C), 지역사회 평가(R-Out)는 순차적으로 $r=.66$, .62, .63(모두 $p<.001$)로 유사하였다. 본 연구에서 PCL-R과의 상관관계는 선행 연구보다 다소 낮았지만($r=.47$ (위험성 존재), $r=.51$ (관련성), 모두 $p<.01$) 전체적으로 유의미한 수준의 상관을 보여 타당도가 있음을 보였다.³¹⁾

PCL-R 하위요인과 단면측면에서 선행연구를 살펴보면, Smith와 Edens(2013)에서 PCL-R의 요인1(정서/대인관계)은 요인2(충동/반사회성; $r_s = .64-.83$, all $p<.001$)에 비해 HCR-20^{V3}의 점수($r_s = .12-.32$, all ns)와 상대적으로 낮은 관련성을 보였다. PCL-R의 단면2(정서성)는 HCR-20^{V3}의 점수와 다소 강한 연관을 보였으나($r_s=.16-.39$) 단면1(대인관계)은 약하고 유의미하지 않은 것으로 나타났다($r_s=.05-.21$, all ns). 임상 척도와의 상관관계는 유의미 했고($r=.39$, $p<.05$), 이력 척도와의 상관관계는 미미했다($r=.34$, $p=.058$). 단면3(생활방식)은 HCR-20^{V3} 점수와 관련성이 높았다($r_s=.80$ 에서 .66, $p<.01$ 에서 .001). 단면 4(반사회성)는 HCR-20^{V3}

31) Hart, S. D., Cox, D. N., & Hare, R. D. (1995). Hare psychopathy checklist: Screening version (PCL: SV). Multi-Heath Systems.

점수와 가장 강력하게 관련되어 있다($r=.62$ 에서 $.81$, $p<.001$). 본 연구에서는 선행 연구에 비해 ‘관련성’이 비교적 낮게 나타났지만 유의미한 상관관계를 보였다(요인 1, $r=.17-.38$; 요인2, $r=.12-.53$). 또한 전체적으로 이력 척도($r=.109-.553$)가 임상 척도($r=.088-.310$)보다 유의미한 수준에서 고루한 상관관계를 보였다. 특히 이력 척도는 요인2와 단면3(생활양식)과 높은 관련성을 보인 반면 임상 척도는 상대적으로 미미한 상관관계를 보였다. 이는 단면4가 가장 관련성이 있다는 기존 연구와는 차이가 나는 점이다. 위험관리 척도는 단면2(정서성)와 가장 높은 관련성을 보였다.

CAPP, LSI-R과 KORAS-R 척도들과 HCR-20 간에 동시 타당도를 살펴본 연구(버전 3.0인 경우)는 국내외 문헌을 살펴볼 때 본 연구가 처음이다. CAPP과의 상관관계에서는 집중력, 융통성, 편협성 등을 나타내는 인지영역이 이력 척도, 임상 척도 그리고 위험관리 척도와 전반적으로 유의미한 관련성을 보였다. 이력척도는 정서성 영역과는 부적상관을 보인 반면 위험관리와는 높은 상관을 보였다. 위험관리 척도는 대인관계에서 나타나는 개별 특질인 무심한, 공감력 결핍, 무정한 등을 포함하고 있는 애착영역과 유의미한 수준의 상관을 보였다. 하지만 자기영역은 HCR-20 하위 척도와 전반적으로 관련성이 나타나지 않았다. 이는 기존 연구에서도 지적하는 바와 같이 자기영역이 다른 영역에 비해 사이코패스를 나타내는 원형성이 떨어진다고 보고한 바와 일치한다고 볼 수 있겠다(Hart, Logan, 2013). LSI-R과는 전체적으로 유의미한 상관관계를 보여 주고 있다. 특히 위험관리 척도는 LSI-R 하위요인 중 개념적으로 함께 공유될 수 있는 척도에서 유의미한 상관을 보였는데, 이를테면 가족/결혼, 주거환경, 레저/레크레이션, 동료/친구 등과 유의미한 관련성을 보였다. 이력 척도의 경우도 교육/직업, 가족/결혼, 음주/약물, 정서/인성, 태도/지향 등과 유의미한 관련성을 보였다. 특히 임상 척도의 경우 정서/인성과 태도/지향과 관련성을 보였다. 하지만 KORAS-G는 이력 척도 이외의 다른 두 개의 척도에서는 유의미한 상관을 보이지 않았다. 이는 KORAS-G는 순전히 정적요인에 근거를 두고 있기 때문에 동적요인으로 구성되고 있는 HCR-20 하위 척도들과는 관련성을 보이지 않는 것으로 생각된다.

마지막으로 전문가 최종 요약판단 항목과 다른 척도와의 관련성인데 사례우선순위와 긴급조치 필요성, 사례[재]검토 주기가 KORAS-G를 제외한 대부분의 척도와

유의미한 관련성을 보였다. 사례[재]검토만이 KORAS-G와 유의미한 상관을 보였다. 반면 심각한 신체적 위해 위험성은 LSI-R과만 유의미한 정적상관을 보였으며, 제기된 다른 위험성은 모든 척도들과 유의미한 상관을 보이지 않았다. 따라서 전반적인 사례 개입 우선순위 매김, 긴급조치의 필요성 그리고 사례 검토의 주기를 고려할 경우 전문가 판단 항목으로 활용될 수 있음을 나타낸다. 하지만 CAPP과 유의미한 부적상관 상관을 보인 심각한 신체적 위해 위험성이나 아예 유의미한 상관을 보이지 않은 제기된 다른 위험성은 사례관리 적용에서 추가적인 연구나 연구자의 주의가 요구될 수 있다.

4.3 장점 및 제한점

본 연구의 가장 큰 장점은 면담과 파일검토가 동시에 이루어졌다는 점이다. 그리고 외적타당도정도가 높을 것으로 보이는데 이는 평가자간의 대부분이 임상장면에서 심리평가를 직접 실시하는 전문가이기 때문이다. 또 다른 강점은 본 연구에서는 2-3명이 아닌 4명의 평가자가 동시에 채점이 이루어졌다는 점이다. 많은 수의 평가자는 결국 필요한 사례수를 줄여 줄 뿐만 아니라 안정적인 신뢰도를 얻는데 도움을 줄 수 있다. 특히 사례수의 경우 이전의 연구(Douglas와 Belfrage, 2014; Belfrage와 Douglas, 2012; Doyle, Shaw와 Coid, 2013)에 비해 본 연구에서는 최소 3배 이상 더 많은 사례수가 포함되었다. 무엇보다도 대상자 모두 평가자와는 전혀 면식이 없었던 상태였다. 따라서 평가자가 대상자의 중요한 정보에 대해 더 민감하게 포착할 수 있을 뿐만 아니라 대상자에 대해 가질 수 있는 편견과 오염을 사전에 막을 수 있었다.

동시타당도의 경우, 같은 평가자가 동시에 다른 척도들을 함께 평가했다는 점이 제한점으로 따를 수 있다. 원칙적으로는 한 척도에 대한 채점자가 다른 척도를 평가할 경우 분명 영향을 받을 수밖에 없다는 점이다. 따라서 향후에는 다른 평가자가 척도를 달리해서 채점할 수 있도록 보다 정교화 한 실험 설계가 필요할 것으로 보인다. 하지만 다른 채점자에 의한 평가도 일정 부분 보조 정보의 종류와 추가 출처라는 오류에 의해 영향을 받을 수밖에 없다. 본 연구에서는 채점자가 다른 척도를

평가하는 중에 척도 간 차이점이 나타날 경우 과보상하는 형태로 채점할 수도 있기 때문에 평가 자체가 오염될 수 있다. 즉 타당화하고자 하는 척도 평가 자체가 다른 척도의 평가에 작간접적으로 영향을 줄 수도 있다는 점이다. 다른 채점자가 평가한 것과 동일한 채점자가 평가한 것 사이의 상관관계 일치정도를 비교해보아야 진실 된 상관관계 정도를 파악할 수 있을 것으로 보인다.

4.4 형사정책적 시사점

HCR-20은 한국과 일부 국가를 제외한 많은 국가의 교정국, 보호관찰, 군과 법무 감호치료소 등에서 공식적으로 활용하고 있는 추세에 있으나 국내에서는 본 도구의 소개와 연구가 미진했던 것이 사실이다. 따라서 본 연구를 통해 얻을 수 있었던 가장 큰 시사점은 서론에서 서술하였던 바, HCR-20의 핵심개념과 평가체계에 대해 국내에 구체적으로 언급할 수 있었다는 점이다. 본 연구 결과가 국외 연구 결과와 크게 다르지 않았을 뿐 아니라 그 신뢰도와 타당도가 일관성 있는 수준을 보였다. 평가자가 평가도구를 활용함에 있어 평가자간의 신뢰도와 핵심 특질을 잘 포착했느냐가 중요한 쟁점일 수밖에 없다. 결국 본 도구의 안정성을 보여주는 첫걸음을 통해 국내 연구자와 실무자에게 문을 열어 주었다는 점에서 그 의의가 있다고 본다.

본 HCR-20은 서종한 등(2017)이 언급한 바와 같이 4세대 모델로 현장에서 실무자가 미래 예측이라는 단순한 환상적 패러다임에서 벗어나 통합적 수준에서 위험요인을 찾고 향후 폭력성과의 관련성을 찾는 데 있다. 한걸음 더 나아가 합리적인 모델에 근거하여 사례개념화(forensic case formulation)를 실시하며 향후 관리계획(management plan) 측면까지 평가자가 연속적으로 판단할 수 있다. 하지만 그 중에서도 가장 중요한 포인트는, 단순 폭력 관련 위험요인의 존재 유무만을 평가하는 것이 아니라 전문가가 해당 위험요인이 향후 폭력 위험성과 관련성이 있는지 진지한 수준의 의사결정 과정을 열어주었다는 점에서 가장 큰 의미가 있다고 본다. Judge 등(2014)과 Prescott(2014)의 연구에서 잘 지적한 바와 같이 국외에서 본 도구가 활발히 사용되고 있는 이유 중에 하나도 평가자가 기계적으로 도구 자체에 모든 의사결정을 ‘무책임’하게 맡기는 것이 아니라 평가자 스스로가 향후 폭력성 간의 관련성을

찾고 이에 대한 개입과 관리계획을 주체적으로 고민하는 계기를 마련해 준다는 점이다. 아쉽게도, 국내에서는 전문가 판단모델(SPJ)이 소개된 바가 없고 전문가들이 이를 적용할 수 있는 기회조차 없었다.

결언으로 형사정책 측면에서 사회적 비용과 개인희생을 최종적으로 줄여 나가기 위해서는 결국 범법자에 대한 처우를 어떻게 결정하고 적용할 것인지에 대해 근본적인 변화가 필요하다. 단순 재범위험성에 따라 가둘지 말지 아니면 통제할지의 과거의 패러다임에서 전환하여 이들 폭력성의 근본적인 원인이 무엇인지 그리고 이들 위험요인 간 어떤 위계관계가 형성되어 있고 고리가 맺어 있는지, 폭력에 이르는 개별 매커니즘에 대한 이해와 치료적 접근이 개별기술적(idiographic) 수준에서 절실하다(서종한 등, 2017).

참고문헌

- 서종한·김희송·김경일. (2017). 성폭력위험성평가의 진화. 형사정책연구, 28(2), 215-256.
- 윤정숙·이수정 (2012). 성범죄자 치료 프로그램의 주요 원칙과 방향 정립을 위한 이론적 고찰. 한국심리학회지: 일반, 31(3), 847-875.
- 이수정 (2010). 최신 범죄 심리학. 서울: 북카페.
- 이수정·고려진 (2011), 한국 성범죄자 위험성 평가척도(KSORAS) 실시요강. 서울: 맥스메디카
- 이수정·윤옥경. (2003). 범죄위험성의 평가와 활용방안. 한국심리학회지: 일반, 22(2), 99-126.
- 이춘화 (2001). 소년에 대한 조사제도의 문제점과 개선방안. 소년보호연구, (3), 26-85.
- 조은경 (2010). 성폭력사범재범위험성평가척도 개발(HAGSOR). 과천: 법무부[법무부용역연구보고서].
- 조은경·이수정. (2008). PCL-R 전문가 지침서: 한국판 표준화. 서울: 학지사.
- 최이문·강태경·조은경. (2014). 미국 법원에서 심리학 도구 (사이코패시 체크리스트, PCL-R) 의 역할에 대한 연구 (2005~ 2012). 형사정책연구, 25(4), 375-414.
- 최혜림 (2011) 초기 한국판 재범위험성 평가도구 : 일반위험요인(KORAS-G)의 신뢰도 분석. 경기대학교 석사논문.
- American Psychological Association, and National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Andrews, D. A., Bonta, J., & Wormith, S. J. (2010). The Level of Service (LS) assessment of adults and older adolescents. In R. K. Otto & K. S. Douglas (Eds.), *Handbook of violence risk assessment* (pp. 199-225). New York: Routledge.

- Andrews, D. A., & Bonta, J. (2000). The level of service inventory-revised. Toronto, Canada: Multi-Health Systems.; Andrews, D. A., Bonta, J., & Wormith, S. J. (2000). Level of service/case management inventory: LS/CMI. Toronto, Canada: Multi-Health Systems.
- Atkins, D., Slegel, J., & Slutsky, J. (2005). Making policy when the evidence is in dispute. *Health Affairs*, 24, 102-113.
- Belfrage, H. & Douglas, K. S. (2012). *Interrater reliability and concurrent validity of HCR-20 (Version 3)*. Unpublished data analyses. Mid-Sweden University, Sundsvall, Sweden.
- Blanchard, A., & Douglas, K. S. (2011, March). *APLS Undergraduate Paper Award (First Place): The Historical-Clinical-Risk Management-Version 3: The inclusion of idiographic relevance ratings in violence risk assessment*. Invited poster presented at the annual convention of the American-Psychology Law Society, Miami, FL.
- Bloom, H., Webster, C. D., Hucker, S., & De Freitas, K. (2005). The Canadian contribution to violence risk assessment: History and implications for current psychiatric practice. *Canadian Journal of Psychiatry*, 50, 3 - 11.
- Boer, D. P., Hart, S. D., Kropp, P. R., & Webster, C. D. (1997). *Manual for the Sexual Violence Risk - 20: Professional guidelines for assessing risk of sexual violence*. Burnaby, Canada: Mental Health, Law, & Policy Institute, Simon Fraser University.
- Buchanan, A. (2008). Risk of violence by psychiatric patients: beyond the actuarial versus clinical assessment debate. *Psychiatric Services*, 59, 184-190.
- Chermack, T. J., & Van Der Merwe, L. (2003). The role of constructivist learning in scenario planning. *Futures*, 35(5), 445-460.
- Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: applications to assessment of

- adaptive behavior. American journal of mental deficiency.
- de Vries Robbé, M., & de Vogel, V. (2010). *Pilot study of HCR-20 (Version 3)*. Unpublished data analyses. Van der Hoeven Kliniek, Utrecht, The Netherlands.
- Douglas, K. S., Blanchard, A., & Hendry, M. (2013). Violence risk assessment and management: Putting structured professional judgment into practice. In C. Logan & L. Johnstone (Eds.), *Risk assessment and management: Clinical guidelines for effective practice* (pp. 29-55). New York: Routledge/Taylor and Francis Group.
- Douglas, K. S., & Reeves, K. (2010). The HCR-20 violence risk assessment scheme: Overview and review of the research. In R. K. Otto & K. S. Douglas (Eds.), *Handbook of violence risk assessment* (pp. 180-185). New York: Routledge/Taylor & Francis Group.
- Doyle, M., Shaw, J., & Coid, J. (2013). *Validation of risk assessments in medium secure services (VORAMSS): Preliminary HCR-20 (Version 3) analyses*. Unpublished data analyses. University of Manchester, UK, and Queen Mary University of London, UK.
- Eidhammer, G., Selmer, L. E., & Bjørkly, S., and the SAFE Pilot Project Group (2013). *A brief report of a comparison of the HCR-20 (Version 2) and HCR-20 (Version 3) concerning internal consistency and clinical utility*. Centre for Research and Education in Forensic Psychiatry, Oslo University Hospital, Norway.
- Gottfredson, S. D. (1987). Prediction: An overview of selected methodology issues. In D. M. Gottfredson & M. Tonry (Eds.), *Prediction and classification: Criminal justice decision making, crime and justice: A review of research*, Vol. 9 (pp. 21-53), Chicago: University of Chicago Press.
- Gottfredson, S. D., & Moriarty, L. J. (2006). Statistical risk assessment: Old

- problems and new applications. *Crime & Delinquency*, 52, 178-200.
- Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law*, 2, 293-323.
- Hanson, R. K. (1997). *The development of a brief actuarial risk scale for sexual offense recidivism*. Ottawa: Solicitor General of Canada.
- Hanson, R. K. (1998). What do we know about sex offender risk assessment? *Psychology, Public Policy, and Law*, 4, 50 - 72.
- Hanson, R. K., & Thornton, D. (1999). *Static-99: Improving actuarial risk assessments for sex offenders*. Ottawa: Solicitor General of Canada.
- Harcourt, B. E. (2007). *Against prediction: Profiling, policing and punishing in an actuarial age*. Chicago: University of Chicago Press.
- Harris, G. T., Rice, M. E., & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior*, 20, 315-335.
- Hart, S. D. (1998). The role of psychopathy in assessing risk for violence: Conceptual and methodological issues. *Legal and Criminological Psychology*, 3, 121 - 137.
- Hart, S. D. (2003a). Actuarial risk assessment: Commentary on Berlin et al. *Sexual Abuse: A Journal of Research and Treatment*, 15, 383-388.
- Hart, S. D. (2003b). Violence risk assessment: An anchored narrative approach. In M. Vanderhallen, G. Vervaeke, P. J. Van Koppen, & J. Goethals (Eds.), *Much ado about crime: Chapters on psychology and law* (pp. 209-230). Brussels: Uitgeverij Politeia NV.
- Hart, S. D. (2011). Complexity, uncertainty, and the reconceptualization of violence risk assessment. In R. Abrunhosa (Ed.), *Victims and offenders: Chapters on psychology and law* (pp. 57-69). Brussels: Politeia. (Original

work published in 2004.)

- Hart, S. D. (2009). Evidence-based assessment of risk for sexual violence. *Chapman Journal of Criminal Justice*, 1, 143-165.
- Hart, S. D., Cox, D. N., & Hare, R. D. (1995). Hare psychopathy checklist: Screening version (PCL: SV). Multi-Heath Systems.
- Hart, S. D., Douglas, K. S., & Webster, C. D. (2001). Risk management using the HCR-20: A general overview focusing on historical factors. In K. S. Douglas, C. D. Webster, S. D. Hart, D. Eaves, & J. R. P. Ogloff (Eds.), *HCR-20 violence risk management companion guide* (pp. 27-40). Burnaby, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University.
- Hart, S. D., Kropp, P. R., Laws, D. R., Klaver, J., Logan, C., & Watt, K. A. (2003). *The Risk for Sexual Violence Protocol (HCR-20): Structured professional guidelines for assessing risk of sexual violence*. Burnaby, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University.
- Hart, S. D., & Logan, C. (2011). Formulation of violence risk using evidence-based assessments: The structured professional judgment approach. In P. Sturmey & M. McMurrin (Eds.), *Forensic case formulation* (pp. 83-106). Chichester, UK: Wiley-Blackwell.
- Heilbrun, K. S., & Dvoskin, J. (2001). Risk assessment and release decision-making: Toward resolving the great debate. *Journal of the American Academy of Psychiatry and the Law*, 29, 6-10.
- Kapp, M. B., & Mossman, D. (1996). Measuring decisional competency: Cautions on the construction of a “capacimeter.” *Psychology, Public Policy, and Law*, 2, 73-95.
- Kropp, P. R., Hart, S. D., & Lyon, D. (2008). *Guidelines for Stalking Assessment and Management (SAM): User manual*. Vancouver, Canada: ProActive

ReSolutions Inc.

- Kropp, P. R., Hart, S. D., Lyon, D., & LePard, D. (2002). Managing stalkers: Coordinating treatment and supervision. In L. Sheridan & J. Boon (Eds.), *Stalking and psychosexual obsession: Psychological perspectives for prevention, policing and treatment* (pp. 138-160). Chichester, UK: Wiley.
- Krug, E. G., Dahlberg, L. L., Mercy, J. A., Zwi, A. B., & Lozano, R. (Eds.). (2002). *World report on violence and health*. Geneva: World Health Organization.
- Mears, D. P. (2007). Towards rational and evidence-based crime policy. *Journal of Criminal Justice*, 35, 667-682.
- Meehl, P. E. (1996). *Clinical versus statistical prediction: A theoretical analysis and a review of the literature*. Northvale, NJ: Jason Aronson. (Original work published in 1954.)
- Megargee, E. I., Cook, P. E., & Mendelsohn, G. A. (1967). Development and evaluation of an MMPI scale of assaultiveness in overcontrolled individuals. *Journal of Abnormal Psychology*, 72, 519-528.
- Monahan, J. (1995). *The clinical prediction of violent behavior*. Northvale, NJ: Jason Aronson. (Original work published in 1981.)
- Monahan, J., & Steadman, H. J. (Eds.). (1994). *Violence and mental disorder: Developments in risk assessment*. Chicago: University of Chicago.
- Monahan, J., Steadman, H. J., Appelbaum, P. S., Grisso, T., Mulvey, E. P., Roth, L. H., Robbins, P. C., Banks, S., & Silver, E. (2005). *Classification of Violence Risk (COVR)*. Lutz, FL: Psychological Assessment Resources, Inc.
- Otto, R. K. (2000). Assessing and managing violence risk in outpatient settings. *Journal of Clinical Psychology*, 56, 1239-1262.
- Pfeffer, J., & Sutton, R. I. (2006). Evidence-based management. *Harvard Business Review*, January, 63-74.

- Pomerol, J.-C. (2001). Scenario development and practical decision making under uncertainty. *Decision Support Systems*, 31, 197-204.
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. (1998). *Violent offenders: Appraising and managing risk*. Washington, DC: American Psychological Association.
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. (2006). *Violent offenders: Appraising and managing risk*, 2nd ed. Washington, DC: American Psychological Association.
- Ringland, G. (1998). *Scenario planning: Managing for the future*. Chichester, UK: Wiley.
- Sackett, D. L., & Rosenberg, W. M. (1995). On the need for evidence-based medicine. *Journal of Public Health Medicine*, 17, 330-334.
- Sackett, D. L., Rosenberg, W. M., Muir Gray, J. A., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn't—it's about integrating individual clinical expertise and the best external evidence. *British Medical Journal*, 312, 71-72.
- Schwartz, P. (1990). *The art of the long view*. New York: Doubleday.
- Segal, D. L., & Coolidge, F. L. (2000). Assessment. In A. E. Kazdin (Ed.), *Encyclopedia of psychology*, Vol. 1 (pp. 264-272). Washington, DC: American Psychological Association.
- Smith, S., & Edens, J. F. (2013). *Concurrent validity between the HCR-20 (Version 3) and the PCL-R among male pre-trial offenders in Texas*. Unpublished data analyses. Texas A&M University, College Station, Texas.
- Strub, D. S., & Douglas, K. S. (2009, March). *New version of the HCR-20 violence risk assessment scheme: Evaluation of a draft of the revised Historical scale*. Poster presented at the American-Psychology Law Society conference, San Antonio, Texas.

- Uebersax, J. (2009). Raw agreement indices. URL: <http://ourworld.compuserve.com/homepages/jsuebersax/raw.htm> Accessed, 10.
- van der Heijden, K. (1997). *Scenarios: The art of strategic conversation*. New York: Wiley.
- Webster, C. D., Hucker, S. J., & Bloom, H. (2002). Transcending the actuarial versus clinical polemic in assessing risk for violence. *Criminal Justice and Behavior*, 29, 659 - 665.
- Webster, C. D., Martin, M. L., Brink, J., Nicholls, T. L., & Desmarais, S. L. (2009). *Manual for the Short-Term Assessment of Risk and Treatability (START)*, Version 1.1. Coquitlam, Canada: British Columbia Mental Health & Addiction Services.

The Paradigm Shift from prediction to prevention: the Validity Study of HCR-20 Version 3

Sea, Jong-han* · Joe, Byung-chul** · Kim, Hee-song*** · Kim, Kyung-il****

We evaluated the interrater reliability and concurrent validity of the HCR-20 Version 3(HCR-20^{V3}) for a total of 80 inmates across 18 prisons. Four sets of final ratings were blindly conduct by correctional clinicians, criminal analyzers and graduate-level students for 20 risk factors and five final opinions. Reliability analyses focused on ratings of the presence of Version 3 risk factors in step 2 & 3 and five opinions in step 7. Concurrent validity analyses focused on the correlational associations with between Version 3 and four well-validated measurements(CAPP, LSI-R, PCL-R, KORAS-G). Findings indicated that Version 3 and other measurements were strongly correlated. Interrater reliability was consistently excellent for the presence of risk factors and for final opinions. The majority of relevance ratings in the good to excellent range, although there was a minority of such ratings in the fair or poor categories. Findings support the concurrent validity and interrater reliability of HCR-20^{V3}. Implication for the use of HCR-20^{V3} in Korean settings by professionals and agencies are further discussed.

❖ Keyword: evidence-base, violence, HCR-20 Version 3, violence risk assessment

투고일 : 2월 28일 / 심사일 : 3월 23일 / 게재확정일: 3월 23일

* Simon fraser University, Dept. of Psychology

** Ajou University, Dept. of Psychology

*** National Forensic Service, Dpet. of Forensic Psychology

**** Corresponding Author: Ajou University, Dept. of Psychology