한국심리학회지 : 임상
The Korean Journal of Clinical Psychology
2010, Vol. 29, No. 2, 505-523

한국판 치매 평가 검사(K-DRS)를 이용한 노인인지기능 평가에서 변화유의성 탐지: 50세 이상 장노년 검사-재검사 규준 연구*

김 호 영　　최 진 영†　　석 정 서
서울대학교 심리학과

시간 경과에 따른 기능의 변화를 측정하는 것은 신경심리검사의 중요한 용도지만, 검사 수행의 변화가 개인의 실제 변화를 반영하는 지 아니면 단지 정상적인 변산성을 반영하는 지 판단하는 것은 쉽지 않다. 본 연구에서는 신경학적으로 건강한 50세 이상의 장노년 291명을 대상으로 실시된 K-DRS의 검사-재검사 자료를 이용하여 반복측정에 따른 점수 변화에 영향을 미치는 변인을 탐색하고, 관찰된 수행변화가 정상 범위를 벗어나는 지 판단할 수 있는 검사-재검사 규준을 마련하였다. 분석 결과, K-DRS의 총점과 모든 소검사 점수의 변화에 대해 기저 검사 수행이 가장 강력한 예측변수로 밝혀졌다. 즉, 기저 수행이 낮을수록 점수가 더 많이 향상되는 평균으로의 회귀 현상이 일관되게 관찰되었다. 인구 통계학적 변수도 일부 소검사의 수행 변화를 유의미하게 예측하는 변수로 나타났는데, 교육 수준이 높을수록, 연령이 적을수록 재검사에서 더 많은 점수가 향상되는 것으로 밝혀졌다. 또한 평균적인 연습효과를 보정한 변화유의지수(Reliable Change Index adjusted for practice; RCIp, Chelune et al., 1993)와 표준회귀모형(Standardized Regression Based; SRB)을 이용하여 변화 유의성을 판단할 수 있는 예측구간을 계산하였다. 이 예측구간을 적용하여 유의미한 향상 혹은 감퇴를 보인 것으로 분류된 비율을 살펴본 결과, SRB가 RCIp보다 실제 재검사 자료의 분포를 합리적으로 반영하는 것으로 나타났다.

주요어 : 연습효과, 변화 유의지수(RCI), 표준회귀모델(SRB), 예측구간, 반복측정

---

* 이 논문은 2009년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No.2009-0068291).

† 교신저자(Corresponding Author) : 최진영 / 서울대학교 심리학과 / 서울시 관악구 관악로 599
Fax : 02-877-6428 / E-mail : jychey@snu.ac.kr

www.kci.go.kr

신경심리검사는 종종 현재 기능 수준을 평가하여 진단의 정확성을 높이는데 쓰이지만, 진행성 질환의 경과를 추적하고 뇌 손상으로부터 회복되는 정도와 치료 효과를 평가하는 등 시간 경과에 따른 기능 변화를 측정하는 것도 신경심리검사의 중요한 용도 중 하나이다. 신경심리검사에서 재평가는 진단적인 측면에서도 중요한 역할을 하는데, 진행성 치매를 가장 잘 진단할 수 있는 방법으로 반복적인 신경심리평가가 권장되기도 한다(Cummings & Benson, 1983; Winblad, Palmer, Kivipelto, Jelic, Fratiglioni, Wahlund, Nordberg, Backman, Albert, Almkvist, Arai, Basun, Blennow, DeLeon, Decarli, Erkinjuntti, Giacobini, Graff, Hardy, Jack, Jorm, Ritchie, Vanduijn, Visser & Petersen, 2004).

신경심리검사의 반복 측정에서 점수의 변화는 (1) 개인 인지 기능의 실제적인 변화, (2) 같은 검사를 반복 실시함으로써 나타나는 연습효과, (3) 검사의 측정 오차로 인해 발생하는 정상적인 변산성, (4) 평균으로의 회귀 현상 등이 복합적으로 영향을 주어 나타난 결과이다. 따라서 반복측정 결과 얻어진 점수의 차이를 실제적인 기능의 변화만으로 단순화하여 해석하기는 어렵다. 더욱이 노인이나 아동과 같이 발달적으로 인지 기능의 정상적인 변화가 예상되는 집단에서는 실제 인지 기능의 변화라도 그 연령에서 나타날 수 있는 정상 범위 내의 변화인지를 따져볼 필요가 있다. 즉, 개인의 수행 변화가 반복측정에 따른 정상 범위의 변화인지, 그 범위를 넘어선 유의미한 변화인지를 판단하기 위해서는 적절한 해석 기준이 필요하다.

변화 유의지수(Reliable Change Index; RCI, Jacobson & Truax, 1991)는 임상가가 심리평가에서 관찰된 수행 변화의 통계적 유의성을 검사의 변산성과 검사-재검사 신뢰도를 고려하여 판단할 수 있도록 돕기 위해 개발되었다. 이 지수는 참고 집단의 자료를 기초로 해당 영역의 실제적인 변화가 없을 때 관찰된 변화가 발생할 확률을 추정하여 변화의 유의성을 판단할 수 있게 해준다. 즉, RCI는 관찰된 변화가 임상적으로 중요한 사건이 발생하지 않은 상태에서도 정상적으로 일어날 수 있는 정도를 보여주기 위한 것으로, 정상 규준 집단의 검사-재검사 자료에 근거하여 산출된다. RCI를 처음 고안한 Jacobson과 Truax (1991)는 재검사 점수가 초기 점수와 동일할 것으로 간주하고 차이의 표준 오차만을 고려하여 점수 변화의 유의성으로 판단하였다. 즉, 재검사 점수를 기저 점수와 같은 점수로 예측하고 측정의 표준 오차를 고려하여 실제 관찰된 재검사 점수가 예측된 점수에서 유의미하게 벗어나면 유의미한 향상 혹은 감퇴로 해석하였다. 한편 Chelune, Naugle, Luders, Sedlak과 Awad(1993)는 연습효과로 인한 전반적인 수행 향상을 설명하기 위해 RCI를 수정하여 연습효과를 보정한 변화 유의지수(Reliable Change Index with Practice; RCIp)를 개발했다. 이 방법에서는 초기 점수에다 규준 집단에서 계산된 점수 변화의 평균값을 더하여 재검사 점수를 예측하고, 차이의 표준오차를 고려할 때 실제 관찰된 재검사 점수와 예측된 값의 차이가 유의미한 수준인지를 판단한다. 평가자가 설정한 유의수준에 따라 표준 오차에 z 분포상의 특정 값을 곱하여 검사-재검사 점수 차이에 대한 예측구간을 계산하고, 관찰된 재검사 점수가 예측된 점수에서 이 구간을 벗어나는 정도의 차이를 보이면 유의미하게 변한 것으로 해석한다. 표준 오차 값을 계산하는 방법에도 여러 가지가 제안되고 있고(Jacobson & Truax, 1991; Temkin

www.kci.go.kr

Heaton, Grant, & Dikmen, 1999; Massen, 2004) 표준 오차를 계산하는 방식에 따라 예측구간의 범위도 달라진다. 하지만 초기 검사 점수와 재검사 점수의 표준편차의 차이가 작을 때에는 각 계산 방식에 따라 산출되는 표준 오차들의 차이는 미미하다. 재검사 점수의 분포에 대한 사전 정보가 있다면, 기저 검사와 재검사의 표준편차를 모두 고려하여 계산하는 Massen(2004)의 방법이나 Temkin 등(1999)의 방법을 사용할 수 있다.

검사-재검사 신뢰도가 1이 아닌 이상 연습효과와 함께 평균으로의 회귀라는 통계적 현상이 정상적인 변산성(normal variability)의 일부로 나타난다. 평균으로의 회귀는 기저 검사에서 평균으로부터 거리가 먼 점수들이 재검사에서는 평균에 보다 가깝게 이동하는 경향으로 정의되며, 이 현상이 두드러지면 기저 수행 수준에 따라 연습효과가 차별적으로 나타날 것으로 예상해 볼 수 있다. Chelune 등(1993)의 RCIp에서는 초기 검사의 수행에 상관없이 재검사에서 평균적인 연습효과가 나타날 것으로 가정하고 초기 수행 수준에 따른 평균으로의 회귀 현상은 고려하지 않았다. 이를 보완하기 위해 McSweeny, Naugle, Chelune과 Luders (1993)는 회귀방정식을 통해 재검사 점수를 예측하는 방법을 개발하였다. 이 방법은 표준화된 회귀 기반 모델(Standardized Regression Based Model; SRB)로 불리는데, 재검사 점수를 예측하는 예측 변수로 초기 검사 점수와 연령, 교육 등 다양한 변수들을 포함시킬 수 있어 점수 변화에 영향을 미치는 요인들을 통합적으로 고려하여 변화의 유의성을 판단할 수 있다는 장점을 가진다.

RCIp 혹은 SRB와 같이 변화의 유의성을 판단하는 일련의 방법들에서 점수 변화의 통계적 유의성은 정상적인 변산성으로 설명되기 어려운 변화를 의미한다. 따라서 이러한 방법들로 정의된 '유의미한 변화'가 반드시 '병리적 변화'를 의미한다고는 볼 수 없으며, 이 같은 방법들의 임상적 타당성에 대해서는 경험적 연구가 필요하다. Heaton과 동료들(Heaton, Temkin, Dikmen, Avitable, Taylor, Marcotte, & Grant, 2001)은 정상 집단을 대상으로 수립된 변화 유의성 규준이 임상적으로 유용한지를 경험적으로 탐색한 바 있다. 이들은 신경학적으로 정상인 통제 집단 384명으로 대상으로 Wechsler 성인 지능 검사와 Halstead-Reitan Battery의 검사-재검사 규준을 수립하였다. 또한 이 규준을 신경학적으로 안정적이라고 간주할 수 있는 다른 정상 집단과 임상집단을 대상으로 교차 타당화하고, 새롭게 뇌 손상을 입은 환자 집단과 외상성 뇌 손상에서 회복된 환자 집단을 대상으로 일련의 변화 유의지수들이 이들 환자 집단에서 나타나는 실제 임상적 변화를 잘 변별하는지 확인함으로써 임상적 타당성을 살펴보았다. 그 결과 RCI와 SRB로 계산된 90% 예측구간은 교차 타당화를 위한 정상 집단의 85% 이상, 신경학적으로 안정적인 정신분열증 환자 78% 이상을 유의미한 변화가 없는 집단으로 정확히 분류하여 비교적 높은 특이도(specificity)를 보였다. 민감도(sensitivity) 측면에서는 모든 검사에서 90% 예측구간을 적용할 때 회복 집단과 손상 집단에서 각각 상승자와 감퇴자로 분류된 비율이 정상 집단에서 예상되는 Type I 오류율 5%보다 유의미하게 더 큰 것으로 관찰되었다. 구체적으로 뇌 손상에서 회복된 임상집단에서는 20-60%가 유의미하게 수행이 상승된 것으로 분류되었고, 새롭게 손상을 입은 환자의 26-80%가 수행이 유의미하게 감퇴된 것으로 분

www.kci.go.kr

류되었다. 또한 여러 검사에서 나타난 결과를 통합한 기준을 적용하면 특이도가 66-86%로 나타났고 민감도는 65-90%까지 상승하였다.

치매의 진단과 예언에서 신경심리검사의 변화 유의지수의 유용성을 살펴본 연구들도 있었다. Frerichs와 Tuokko(2005)는 캐나다 노인을 대상으로 한 5년 종단 연구에서 초기 검사와 재검사에서 모두 정상 노인으로 분류되었던 166명의 기억 검사 자료에 입각하여 변화 유의지수를 구하였다. 이 지수가 초기에는 정상이었으나 5년 후에는 치매가 발병했던 노인들(20명)과 5년 후에도 정상적인 인지 기능을 유지한 노인들(166명)을 얼마나 정확하게 판별하는지 살펴본 결과, 기억 검사의 종류에 따라 다소 차이는 있지만 분류 정확율이 62-94%로 진단적 변별에 비교적 유용한 것으로 밝혀졌다. 특히 변화 유의성을 판단하는 여러 가지 방법 중 RCIp와 SRB가 여러 검사들에 걸쳐 결과를 정확하게 분류하는 것으로 확인되었다. 75세 이상 고령자를 대상으로 한 독일의 한 종단 연구(Hensel, Angermeyer, & Riedel-Heller, 2007)도 RCI가 치매의 조기 진단에 도움이 될 가능성을 보여주었다. Hensel 등(2007)은 치매 및 인지 기능 저하 선별 검사를 반복 측정한 후 변화 유의성 판단으로 초기에는 치매가 아니었던 노인의 치매 발병을 예측할 수 있는지 살펴보았다. 초기에 치매가 아니었던 노인들 554명을 약 1.5년 후 재검사하여 치매가 발병했는지를 확인하고 두 번째 검사에서의 치매진단 여부를 결과변수로 하여 Receiver Operating Characteristic 곡선 분석을 한 결과, RCIp와 SRB 모두 치매 발병을 유의미하게 예측하는 것으로 나타났다.

본 연구의 목적은 노인의 전반적인 인지 기능을 측정하는 치매평가검사(Dementia Rating Scale; DRS, Mattis, 1988)의 한국판 검사(Korean-Dementia Rating Scale; K-DRS, 최진영, 1998)에서 검사–재검사에 따른 변화의 규준 자료를 제공하는 것이다. DRS는 치매의 조기 진단, 변별 진단, 단계 평가를 위해 임상 장면과 연구에서 널리 사용되며(Jurica, Leitten & Mattis, 2001), 최근 출시된 DRS-2(Jurica, P.J., Leitten, C.L., & Mattis, 2001)에서는 미국 노인들을 대상으로 표준화된 연령과 교육을 교정한 규준 자료가 제공되고 있다(Rilling, Lucas, Ivnik, Smith, Willis, Ferman, Petersen, Graff-Radford, 2005). 또한 Pedraza, Smith, Ivnik, Willis, Ferman, Petersen, Graff-Radford와 Lucas(2007)는 RCIp 방법을 적용하여 DRS의 반복측정에 따른 점수 변화 규준을 발표했다. 이들의 연구에서 DRS 총점의 변화는 성별, 교육, 연령, 그리고 치매 가족력 유무와 상관이 없는 것으로 나타났다. 또한 RCIp 경계점수를 적용하면 9-24개월 간격 재평가에서 초기에는 정상이었다가 재검사 시 임상 진단을 받은 이들의 72.7%가 유의미하게 감퇴한 것으로 나타났고, 특히 경도인지장애나 알츠하이머성 치매 진단을 받은 환자들의 66.7%가 유의미한 감퇴를 보이는 것으로 분류되었다. 국내에서는 최진영(1998)이 DRS를 우리나라 사정에 맞게 수정하여 한국판 치매평가 검사(K-DRS)를 표준화했다. 본 연구에서는 정상 노인을 대상으로 K-DRS의 검사-재검사에 따른 수행 변화에 영향을 미치는 변인을 탐색하고 재평가에 따른 수행 변화의 규준을 수립함으로써 변화의 유의성을 판단할 수 있는 근거를 제공하고자 한다. 또한 변화의 유의성을 계산하는 방법 중 RCIp과 SRB를 비교하여 어떤 모델이 실제 규준 자료의 분포를 보다 잘 반영하는 지 탐색하고자 한다.

www.kci.go.kr

방 법

연구대상

본 연구는 서울대학교 임상신경과학연구실에서 1998년에서 2008년까지 수행했던 지역사회 노인의 인지 노화 연구에 2회 이상 참여하였고, 다음의 배제 조건에 해당하지 않는 50세 이상의 정상 장노년을 대상으로 수행되었다. 초기 검사에서 배제 조건은 다음과 같다: (1) 검사 당시 신경과적 병력이나 주요 정신장애의 병력이 있는 자, (2) 두부 외상이나 뇌수술과 같은 신경외과적 병력이 있는 자, (3) 약물로도 조절이 안 되는 심한 당뇨와 고혈압 병력이 있는 자, (4) 안경이나 보청기와 같은 보조기 사용으로도 검사 실시가 불가능할 정도의 심한 감각 기능 손상이 있는 자. 두 번째 검사 시에 파킨슨씨 병, 뇌혈관 질환, 뇌졸증, 뇌종양 등의 신경과적 문제나 정신과적 문제가 추가로 확인된 자, 재검사 후 1년 이내에 신경학적 문제가 발생한 자, 최종 추적 기간 내(7년 175명, 4년 116명)에 치매 발명이 확인된 자도 배제되었다. 초기 검사에서 K-DRS 총점에서 5 백분위 이하의 낮은 수행을 보였더라도 당시 독립적인 생활에 문제가 없었고 보호자의 보고에 입각하여 이전에 비해 기능이 감퇴했다는 객관적인 증거가 없으면, 정상으로 간주하고 표본의 대표성을 위해 포함시켰다. 최종 분석대상은 총 291명이었으며, 연구 참여자의 인구 통계학적 특징과 검사 재검사 간격은 표 1과 같다. 종단 연구의 검사 간격은 약 1년 단위로 이루어졌으나 참여자에 사정에 따라 실시 간격이 년 단위로 통일되지 않아 검사 간격은 개월 단위로 계산하였다. 모든 참여자는 연구 참여에 앞서 연구에 대한 연구자의 설명을 들었으며 서면 동의서를 통해 참여를 허락하였다.

표 1. 전체 연구 참여자의 인구 통계학적 정보 (N=291)

변인	명 (%)	M (SD)
초기 검사 연령 (세)		66.31 (8.01)
50 - 59세	63 (21.7)	
60 - 69세	131 (45.0)	
70 - 79세	80 (27.5)	
80세 이상	17 (5.8)	
교육 (년)		8.11 (4.92)
무학	35 (12.0)	
1-6년	92 (31.6)	
7-12년	119 (40.9)	
13년 이상	45 (15.5)	
성별		
남	89 (30.6)	
여	202 (69.4)	
검사-재검사 간격(개월)		14.33 (6.05)
7-17개월	247 (84.9)	
21-25개월	31 (10.7)	
34-42개월	13 (4.4)	

측정도구

한국판 치매평가 검사(Korean-Dementia Rating Scale; K-DRS, 최진영, 1998). K-DRS는 한국 노인을 위해 DRS(Mattis, 1988)를 국내 노인 인구에 적합하게 수정하여 표준화한 치매 평가도구이다. 이 검사는 노년기 인지기능을 평가하는데 타당하고 신뢰로운 검사로서 주의, 관리기능, 개념화, 구성, 기억 소검사로 구성되어

www.kci.go.kr

있어 비교적 다양한 인지기능을 평가하고 있다(최진영, 1998). K-DRS는 주의 37점, 관리기능 37점, 구성 6점, 개념화 39점, 기억 25점을 만점으로 하며 총점은 144점 만점이다. DRS는 치매를 선별하고 진행 정도를 지수화하는 용도로 널리 사용되고 있다(Freedman, Kaplan, Delis & Morris, 1994; Shay, Duke, Conboy, Harrell, Callaway, & Folks, 1991). K-DRS는 한국판 간이정신상태검사(Mini-Mental State Examination-Korean; 권용철, 박종한, 1989)와 유의미한 상관을 보이고(r= 82, p<.01), 정상 집단과 치매 환자의 수행 비교에서도 유의미한 차이를 보여 타당도가 확인되었다(최진영, 나덕렬, 박선희, 박은희, 1998). 또한 구조적 뇌자기공명영상(magnetic resonance imaging, MRI) 연구를 통해 신경과학적 타당도도 지지되었다(Chey et al., 2006). 신뢰도의 경우에는 2주 간격으로 실시한 검사-재검사 신뢰도가 .96(p<.001)이었으며, 3명의 검사자를 통해 확인한 채점자간 신뢰도는 .99(p<.001)로 상당히 높게 보고되었다.

자료분석

본 연구에서는 전체 종단 자료 중 검사-재검사 간격에 상관없이 처음 2회의 K-DRS 점수가 분석되었다. 우선 전체 291명의 자료를 이용하여 검사-재검사 수행 변화에 영향을 미치는 요인들을 탐색하고, 이 중 검사-재검사 간격이 1년 내외(12±6개월)인 247명의 자료만 이용하여 변화 유의성에 대한 규준을 수립하였다. 모든 통계적 분석은 윈도우용 SPSS 12.0을 통해 실행되었다.

먼저 정상 장노년 291명을 대상으로 K-DRS 총점과 각 소검사에서 기저 검사 점수(T_1)와 재검사 점수(T_2) 차이의 유의성을 대응 표본 t 검정을 통해 살펴보았다. 또한 두 검사 점수의 차이값(T_2-T_1)을 종속변수로 할 때, 이에 영향을 미치는 변수를 단계적 회귀 절차를 통해 확인하였다. 단계적 회귀 분석에서 예측 변수로 투입된 변수는 각 검사의 기저 점수, 검사-재검사 간격, 그리고 연령, 교육 연한, 성별의 인구 통계학적 변수들이다. 각 단계에서 변수를 선택할 때 유의수준은 .05, 제거할 때 유의수준은 .10으로 지정하였다.

변화 유의지수

검사-재검사 점수의 차이는 측정 오차 이외에도 연습효과와 검사 간격 동안 일어났을 인지 노화의 영향을 받을 수 있다. 연습효과와 인지 노화 정도는 다시 검사 간격과 여러 인구 통계학적 변수들의 영향을 받을 수 있는데, 전체 자료의 85% 가량(n=247)이 1년 내외(12±6개월)의 검사 간격으로 수집되어 이들 자료만 이용하여 검사-재검사 규준을 계산하였다.

RCIp 계산 방정식은 Temkin 등(1999)이 사용한 방식을 따랐다. 이들은 기저 점수(T_1)에 두 검사점수의 평균적인 차이값(T_2-T_1의 평균값 =D)을 더하여 재검사 점수의 예측값을 구하고, 이 예측값과 실제 재검사 점수(T2) 간의 차이를 차이값의 표준편차(SD)로 나누었다(공식 [가]). 이때 사용된 분모는 변화 유의지수의 원 저자인 Jacobson과 Truax(1991)가 사용한 값과는 다르지만, 이에 대해 논의한 기존 연구들(Hinton-Byare, 2000; Maassen, 2004)에서 표준오차 계산 방식에 따른 차이가 크지 않았고, 본 연구에서와 같이 재검사 점수의 기초 통계량에 대한 정보를 가지고 있는 경우, 이를 반영하는 것이 보다 적절할 것으로 판단되어

www.kci.go.kr

Temkin 등(1999)의 방법을 채택하였다. 이를 공식화 하면 다음과 같고 z 점수와 같이 해석될 수 있다. 즉, RCIp>1.645이면 정상 집단에서 점수 변화가 상위 5% 이내에 속하며, RCIp<-1.645이면 점수 변화가 하위 5% 이내 범위에 속한다고 볼 수 있다.

$$\text{RCIp}= \frac{T_2-(T_1+D)}{SD} \qquad [가]$$

정규 분포를 가정할 때, 추정되는 재검사 점수의 90% 예측구간은 재검사 점수의 기대값에서 거리가 <1.645×차이값의 표준편차(SD)> 이내인 구간으로 다음과 같이 계산할 수 있다(공식 [나]). 만약 개인의 실제 재검사 점수(T_2)가 이 예측구간에 포함되지 않는다면 그 개인은 재검사에서 유의미하게 수행 변화를 보였다고 해석할 수 있다. 즉, 재검사 점수가 예측구간의 상한선보다 더 높게 나타났다면 유의미하게 향상된 것으로 해석할 수 있으며, 하한선보다 더 낮게 나타났다면 유의미하게 감퇴한 것으로 해석할 수 있다.

$$\text{RCIp의 90\% 예측구간}=(T_1+D) \pm (1.645 \cdot SD) \qquad [나]$$

표준화된 회귀기반 모델(Standardized Regression-Based Model)은 McSweeny 등(1993)의 방법을 따랐다. SRB 모델에서는 실제 재검사 점수(T_2)와 회귀 방정식에 의해 예측된 재검사 점수(T'_2)의 차이값을 회귀 모형의 추정의 표준오차(SE*est*)로 나눔으로써 표준화하고 이 값을 통해 변화의 유의성을 해석한다. 이를 공식화 하면 공식 [다]와 같으며, RCIp와 마찬가지로 z 점수처럼 해석할 수 있다.

$$\text{SRB}= \frac{T'_2-T_2}{SEest} \qquad [다]$$

단계적 회귀 절차를 이용하여 재검사 점수를 예측하는 회귀 방정식을 도출하였으며, 이때 연령, 교육 연한, 성별의 인구통계학적 변수들과 각 검사의 기저 점수, 검사-재검사 간격이 후보 예측 변수로 포함되었다. 기존 연구들에서 재검사의 강력한 예측 변수로 보고되어 온 기저 점수는 기본적으로 포함시켰고(Temkin et al., 1999; McSweeny et al., 1993; Salinsky et al., 2001; Raymond et al., 2006), 나머지 변수는 단계적 회귀 절차를 통해 선택되었다. 각 단계에서 변수를 선택할 때 유의수준은 .05, 제거할 때 유의수준은 .10으로 지정하였다. 기저 점수를 예측 변수로 한 회귀 방정식으로 재검사 점수(T'_2)를 추정하면 기저 점수(T_1)의 수준에 따라 점수 변화량이 차별적으로 계산되어 평균으로의 회귀 현상을 반영할 수 있다. 이는 연습효과의 개인차를 고려하지 않고 일괄적으로 검사-재검사 점수 차의 평균값을 연습효과로 간주하는 RCIp 모델과 대비된다. 이 모델에서의 예측구간은 다음 공식 [라]와 같이 계산될 수 있다.

$$\text{SRB의 90\% 예측구간}=T'_2 \pm (1.645 \cdot SEest) \qquad [라]$$

RCIp와 SRB 결과 비교

RCIp와 SRB 규준을 이용하여 유의미하게 상승 혹은 감퇴한 것으로 분류되는 사람들의 비율을 계산하고 각각의 비율이 통계적인 예상 비율인 5%와 차이가 있는지 근사적 z 검정을 하였다. 또한 각각의 방법에 의해 K-DRS 총점에서 상승자로 분류된 집단, 유의미한 변화가

www.kci.go.kr

없는 집단, 감퇴자로 분류된 집단 간의 연령, 교육 수준, 검사-재검사 간격, 기저 수행 수준에 차이가 있는지 일원분산분석을 통해 비교하고, Scheffe test로 사후 분석하였다.

결 과

검사-재검사에 따른 변화량에 영향을 미치는 요인들

검사-재검사간 점수 변화의 유의성을 확인하기 위해 대응표본 t 검정을 한 결과, 기억 소검사에서만 유의미한 상승을 보였다($t(290)=4.15$, $p<.001$, 표 2 참고). 하지만 K-DRS의 총점과 각 소검사의 변화량에 영향을 미치는 요인을 단계적 회귀 분석을 통해 탐색한 결과, 검사-재검사 간의 점수 변화량은 고정된 값이라기보다는 개인의 특징에 따라 차별적으로 나타나는 것으로 확인되었다. 회귀 분석의 구체적인 결과는 다음과 같다. K-DRS 총점의 변화량(T_2-T_1)은 K-DRS 총점의 기저 점수($t(288)=-4.34$, $p<.001$)와 학력($t(288)=2.07$, $p<.05$)이 유의미한 예측 변수로 확인되었는데($F(2, 288)=9.54$, $p<.001$), 기저 수행은 낮을수록, 학력은 높을수록 점수 증가량이 큰 것으로 나타났다. 주의와 개념화 점수의 변화도 기저 점수와 학력이 유의미한 예측변수로 확인되었으며($Fs(2, 288)>33.95$, $ps<.001$), 방향성은 총점과 동일하였다. 구성 점수의 변화에 대해서는 기저 점수($t(287)=-10.40$, $p<.001$), 학력($t(287)=4.62$, $p<.01$), 연령($t(287)=-2.53$, $p<.05$)이 유의미한 예측 변수로 확인되었으며($F(3, 287)=36.49$, $p<.001$), 기저 수행이 낮을수록, 교육 수준이 높을수록, 연령이 적을수록 증가폭이 더 큰 것으로 나타났다. 관리기능 점수의 변화량은 기저 점수($t(286)=-8.26$, $p<.001$), 연령($t(286)=-3.44$, $p<.01$), 성별($t(286)=-3.01$, $p<.01$)과 학력($t(286)=2.89$, $p< .01$)이 유의미한 예측 변수로 확인되었으며($F(4, 286)=17.74$, $p< .001$), 기저 수행이 낮을수록, 연령이 적을수록, 교육 수준이 높을수록, 남성보다 여성이 더 많은 증가를 보이는 것으로 나타났다. 기억 점수의 변화량은 기저 점수($t(288)=-10.81$, $p<.001$)와 연령($t(288)=-2.90$, $p<.01$)이 유의미한 예측 변수로 확인되었고($F(2, 288)=58.45$, $p<.001$), 기저

표 2. 기저 점수와 재검사 점수, 두 검사의 차이 점수의 기술 통계량(N=291)

검사	기저 검사(T_1)		재검사(T_2)		재검사-기저검사 (T_2 - T_1)	
	평균	표준편차	평균	표준편차	평균	표준편차
K-DRS 총점	129.77	10.15	130.23	10.69	.46	6.26
주 의	35.18	1.90	35.05	1.88	-.13	1.70
관 리	33.29	4.10	33.32	4.33	.02	3.54
구 성	5.41	1.09	5.46	1.10	.05	.99
개념화	34.77	3.60	34.67	4.17	-.10	3.64
기 억	21.11	2.80	21.73	2.64	.62**	2.56

** p< .001

www.kci.go.kr

수행이 낮을수록, 연령이 적을수록 증가량이 더 큰 것으로 나타났다. 모든 검사에서 기저 수행이 가장 강력한 예측변수로 확인되었으며, 회귀 계수가 모두 음수로 나타나 기저 수행이 높을수록 점수 변화량이 작은 것으로 나타났다. 더욱이 연령과 교육 수준이 평균 수준이고 여성인 피험자를 가정할 때, 모든 검사에서 X 절편이 평균 주변으로 나타나, 대체로 기저 수행이 평균보다 작을수록 재검사 시 수행이 증가하고, 평균보다 큰 경우 감소하는 평균으로의 회귀 현상이 뚜렷하게 나타나고 있음이 확인되었다.

검사-재검사에 따른 변화 유의성 규준

검사-재검사에 따른 K-DRS의 변화의 유의성 규준은 전체 연구 참여자 291명 중 검사-재검사 간격이 12±6개월(범위 7-17개월)인 247명(평균 연령=66.27세[SD=8.31], 평균 교육 수준=7.91년[SD=4.88])의 자료를 분석하였다. 검사-재검사 간격은 변화량에 유의미한 영향을 미치지 않는 것으로 나타났지만, 전체 참여자의 85%가 이 기간에 집중되어 있어 나머지 기간에 해당하는 참여자 수가 적다는 점과 노인기는 인지기능이 감퇴하는 시기라는 점을 감안하면 검사 간격이 2년 이상인 경우와 1년 내외인 경우의 변화량에 차이가 없다고 결론짓기 어려워 검사 간격인 1년 내외인 노인들로 규준집단을 제한하였다. 변화 유의지수는 앞서 기술한 RCIp 모델과 SRB 모델에 따라 계산되었으며, 각각의 기준에 따라 전체 규준집단에서 유의미하게 상승한 사람 수와 유의미하게 감퇴한 사람 수의 비율을 계산하여 비교하였다.

RCIp 모델에 따른 규준

K-DRS 총점과 각 소검사에서 검사-재검사 간 신뢰도, 차이 점수의 평균과 표준 편차, RCIp 모델에 따른 90% 예측구간은 표 3에 제시되었다. 검사-재검사 점수간의 차이는 전체 참여자를 대상으로 한 분석에서와 마찬가지로 기억 소검사에만 유의미했다($p < .001$). 검사-재검사 신뢰도는 .55에서 .83까지로 나타났으며, 각 소검사 점수의 총합인 K-DRS 총점의

표 3. 검사 재검사 간격이 12±6개월일 때, 차이값의 기초통계량과 RCIp 모델에 따른 90% 예측구간(N=247)

검사	검사-재검사 신뢰도	차이값의 평균	차이값의 SD	90% 예측구간	RC-	RC+
K-DRS 총점	.83	.48	6.46	±10.63	-10.15	11.10
주 의	.61	-.13	1.75	± 2.87	- 3.00	2.74
관 리	.65	.06	3.63	± 5.97	- 5.91	6.03
구 성	.61	.05	.99	± 1.63	- 1.57	1.68
개념화	.55	-.20	3.86	± 6.36	- 6.55	6.16
기 억	.57	.69**	2.62	± 4.32	- 3.62	5.01

주. RC- = 90% 예측구간의 하한선, RC+ = 90% 예측구간의 상한선

** $p < .001$

www.kci.go.kr

신뢰도가 가장 높았다. 개인의 점수 변화량 (T_2-T_1)이 평균적인 차이값에서 1.645×차이값의 SD 이내(90% 예측구간)에 포함된다면 그 개인은 두 검사 사이의 기간 동안 유의미한 변화가 없었던 것으로 해석할 수 있다.

SRB 모델에 따른 규준

K-DRS 총점과 각 소검사의 재검사 점수를 예측하는 회귀 방정식을 도출한 결과, 각 검사의 기저 점수는 모든 측정치에서 유의미할 뿐만 아니라 재검사 점수의 분산을 가장 많이 설명하는 변수로 나타났다. 또한 각 검사의 점수 변화량에 대한 회귀 분석에서 유의미한 예측 변수로 밝혀졌던 인구 통계학적 변수들이 재검사 점수도 유의미하게 예측하는 것으로 나타났다(표 4). 잔차들의 표준 편차를 이용하여 계산한 90% 예측구간은 표 4의 가장 우측 열에 제시된 바와 같으며, 개인의 관찰된 재검사 점수가 회귀방정식에 의해 추정된 점수에서 위, 아래로 이 구간 이상 벗어난다면 유의미하게 향상된 혹은 감퇴한 것으로 해석할 수 있다.

RCIp와 SRB의 비교

각 모델에 근거한 변화의 유의성 규준을 비교하기 위해, 전체 집단에서 각 규준에 의해 상승자 혹은 감퇴자로 분류되는 비율을 계산하였다(표 5). RCIp 모델에서 검사-재검사 점수 차이의 분포와 SRB 모델에서 잔차들의 분포가 정규 분포한다고 가정할 때, 90% 예측구간에 근거하여 각각의 모델에서 예상되는 상승자와 감퇴자의 비율은 이론적으로 모두 5%이다. 그러나 표 5에서 확인할 수 있듯이 각각의 모델에서 산출된 상승자와 감퇴자 비율은 이론적인 비율과 차이가 있었다. 상승자의 비율은 K-DRS 총점을 제외하면 SRB 모델에 따른 규준을 이용할 때 더 낮게 나타났으며, 감퇴자의 비율은 관리기능 소검사를 제외하면 RCIp 모델에 따른 규준을 이용할 때 더 낮게 나타났다.

표 3과 4의 결과에 의해 70세, 6년의 정규 교육을 받은 노인을 가정할 때 K-DRS 총점과

표 4. 검사 간격이 12±6개월일 때, SRB 모델에 따른 회귀방정식과 90% 예측구간(N=247)

검사	*F*	R^2	절편	회귀계수				SE	90% PI
				검사1	교육	연령	성별[a]		
K-DRS 총점	266.91	.69	21.63	.82**	.20*			6.30	±10.36
주 의	83.05	.41	16.97	.49**	.09**			1.51	±2.48
관 리	54.34	.46	19.79	.55**	.15**	-.09**	-1.48**	3.24	±5.33
구 성	61.69	.43	3.36	.48**	.05**	-.01*		.84	±1.38
개념화	70.39	.37	16.07	.47**	.27**			3.47	±5.71
기 억	61.46	.34	14.09	.50**		-.04*		2.23	±3.67

주. *SE* = Standard Error, PI= prediction interval

a. 남성=1, 여성=0

* $p < .05$, ** $p < .001$, 모든 *F*값의 *p*s $< .0001$

www.kci.go.kr

표 5. RCIp와 SRB에 규준에 따른 상승자 및 감퇴자의 비율

검사	RCIp		SRB	
	상승자 비율	감퇴자 비율	상승자 비율	감퇴자 비율
K-DRS 총점	3.2	4.0	3.6	6.5
주 의	5.7	2.8*	1.6*	6.5
관 리	4.5	6.5	3.2	4.9
구 성	4.5	4.0	3.6	5.3
개념화	4.0	7.7	2.8*	9.3*
기 억	4.9	5.7	2.8*	6.5

* 5%보다 유의미하게 크거나 작음 (유의수준 .05)

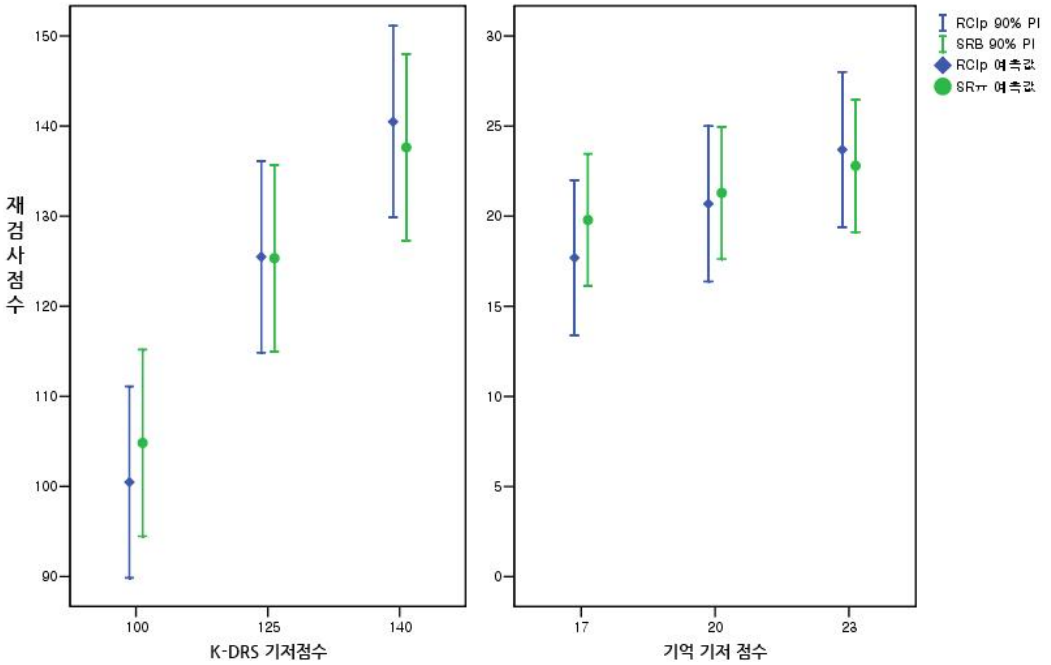

그림 1. 70세, 교육 연한 6년인 가설적인 사례의 기저 수행 수준에 따른 90% 예측구간의 예

기억 소검사의 기저 수행 수준에 따른 예측구간을 구해보면 그림 1과 같다. 그림 1에서 볼 수 있듯이, 총점과 기억 소검사에서 모두 기저 수행 수준이 낮은 경우 RCIp에 의한 예측구간보다 SRB에 의해 예측되는 구간이 보다 높은 반면, 기저 수행 수준이 높은 경우 SRB에 의해 예측되는 구간이 RCIp에 의한 예측구간보다 낮다. 이와 같이 기저 수행 수준에 따

www.kci.go.kr

른 두 모델의 예측구간의 차이로 상승자와 감퇴자를 분류할 때 불일치하는 사례가 발생하는 것으로 보인다. SRB에서 전체 상승자의 비율은 주의, 개념화, 기억 소검사에서 이론적인 예상치인 5%보다 유의미하게 낮았고, 전반적으로 3% 내외의 낮은 비율을 보였다. 이에 비해 RCIp에 따르면 전체 상승자 비율이 5% 내외로 나타났다. 계산된 예측구간에 따르면 기저 검사에서 평균 이상의 수행을 보인 경우 RCIp와 SRB 모두에서 예측구간의 상한선이 만점을 넘는 경우가 많아 기저 수행이 평균 이상인 사람들이 유의미하게 향상된 것으로 분류될 가능성은 낮다. 따라서 본 규준 집단에서는 수행이 유의미하게 상승한 사람들의 비율이 이론적인 기대치인 5%보다 작을 것으로 기대된다. 실제로 모든 측정치에서 기저검사에서 평균 이상의 수행을 보인 사람들 중 재검사에서 유의미한 상승을 보인 것으로 분류된 사람은 총점에 대해 SRB 기준을 적용했을 때 2명(0.8%)만 있었을 뿐이며, 전체 집단에서 상승자는 거의 대부분 기저 수행이 평균 이하였던 사례들에서 발생했다. 한편, 유의미하게 감퇴한 사람의 비율은 SRB에서 더 많게 나타났으며, RCIp가 통계적 기대치인 5%에 더 근접하게 나타났다.

RCIp와 SRB가 편향되지 않은 기준을 제시하는 지를 확인하기 위해, K-DRS 총점의 변화에 따라 RCIp, SRB 각각의 기준으로 상승집단, 무변화 집단, 감퇴 집단을 나눈 후 이 집단들 간의 교육 수준, 연령, 검사-재검사 간격, 기저 K-DRS 총점을 비교하였다(표 6). 그 결과 상승자로 분류된 집단은 RCIp와 SRB에서 모두 교육 수준이 낮고, 검사-재검사 간격이 긴 편이었다. K-DRS 총점의 기저 수행에 대해서는 SRB집단과 RCIp 집단 간에 차이를 보였는데, SRB에서는 상승집단과 감퇴집단의 기저 수행에 차이가 없는 데 비해, RCIp에서는 감퇴집단과 무변화 집단과 비교하여 상승집단만 기

표 6. K-DRS 총점 변화 집단간의 교육, 연령, 검사 간격 및 기저 K-DRS 총점 비교(M±SD)

		상승집단[a]	무변화집단[b]	감퇴집단[c]	*F* *df*=(2,244)	사후분석
RCIp	n (명)	8	229	10		
	교육 연한	2.75±4.13	8.13±4.75	7.10±6.33	5.00**	a < b
	연령	71.63±7.39	65.99±8.27	68.40±8.76	2.14	
	검사 간격	13.25±2.61	11.98±1.26	12.10±.57	3.69*	a > b
	K-DRS총점	113.13±10.29	129.79±10.20	131.40±8.68	10.63***	a< b, c
SRB	n (명)	9	222	16		
	교육 연한	2.00±2.24	8.23±4.65	6.88±6.80	7.84***	a < b, c
	연령	71.00±7.14	65.80±8.26	70.13±8.16	3.60*	
	검사 간격	13.33±2.35	11.96±1.28	12.19±.54	4.98**	a > b, c
	K-DRS 총점	114.67±10.85	130.53±8.55	120.69±21.31	17.62***	a, c < b

* P < .05, ** P < .01, *** P< .001

www.kci.go.kr

저 수행이 현저히 낮았다. 유의미한 차이는 아니지만 RCIp에서는 감퇴집단의 기저 수행의 평균이 무변화 집단보다도 높게 나타났으며, RCIp의 감퇴집단에서 기저 수행이 평균 이상이었던 사람은 10명 중 7명인데 비해 SRB의 감퇴집단에서 기저 수행이 평균 이상이었던 사람은 16명 중 7명으로 나타났다.

요약하자면, RCIp는 예측구간의 상한선을 감안할 때 상승자를 과잉 추정하고, 이 방법에 의해 상승자로 분류된 사람들은 기저 수행 수준이 낮고 감퇴자로 분류된 사람들은 기저 수행이 높은 경향이 있는 것으로 나타났다. 이에 비해, SRB는 예측구간의 상한선을 감안할 때 상승자로 분류되는 비율이 적절하고 기저 수행 수준에 상관없이 감퇴자나 상승자로 분류되는 것으로 나타났다.

논 의

본 연구는 50세 이상의 한국 장노년 인구를 대상으로 K-DRS의 검사-재검사 규준을 제공함으로써 노인의 인지 기능 변화를 평가할 때 해석의 정확성을 높이고자 하였다. 우선 검사-재검사 간격의 분산이 큰 전체 자료를 대상으로 반복 측정에 따른 점수 변화에 영향을 미치는 요인을 탐색하였고, 이 중 검사-재검사 간격이 1년 내외인 자료들을 따로 분석하여 연습효과를 보정한 변화 유의지수(RCIp)와 표준화된 회귀 기반 모델(SRB)에 따른 재검사 점수의 예측구간을 제시하였다. 또한 두 가지 방식으로 제시한 예측구간 중 어느 것이 실제 재검사 자료의 분포를 잘 반영하는 지 살펴보았다.

K-DRS의 점수 변화에 영향을 미치는 요인을 탐색한 결과, 총점과 각 소검사 점수에서 점수 변화량을 가장 잘 예측하는 변수는 각 측정치의 기저 점수인 것으로 나타났다. 각 회귀방정식에서 기저 점수에 대한 회귀 계수는 모두 음수로 기저 수행이 낮을수록 더 많은 점수 향상을 보인다는 것은 재검사 시 평균으로의 회귀 현상이 강력하게 나타나고 있음을 시사한다. 교육 수준도 기억 소검사를 제외한 모든 측정치의 점수 변화에 영향을 미치는 중요한 변수로 밝혀졌는데 교육 수준이 높을수록 점수 변화가 더 큰 것으로 나타나, 기존에 새로운 것을 학습해봤던 교육 경험이 낯선 심리검사 경험에서 검사 상황과 과제 유형에 대해 학습하는 데 도움이 되는 것으로 보인다. 이러한 결과는 DRS의 점수 변화와 교육 연한이 유의미한 상관을 보이지 않았던 Pedraza 등(2007)의 결과와 대조된다. 이들의 연구는 국내 노인들에 비해 전반적으로 교육 수준이 높고 교육 수준의 변산이 작은 미국 노인들을 대상으로 했기 때문에 본 연구에서 나타났던 교육의 영향이 축소되어 나타났던 것으로 보인다. 신경심리 검사 수행에 대한 교육의 영향은 횡단적인 연구들에서도 서구 선진국과 국내의 경우가 다르게 보고되고 있다. 교육의 변산이 작고 교육 수준이 비교적 높은 집단을 대상으로 한 서구 연구들에서는 신경심리검사 수행과 교육 수준의 상관을 크게 보고하지 않는 데 비해, 저교육자가 많고 교육 수준의 편차가 큰 국내 노인 연구에서는 신경심리 수행과 교육 간의 강한 관련성이 종종 보고되고 있다(강연욱, 박재설, 유경호, 이병철, 2009; 문혜성, 최진영, 2004; 안효정, 최진영, 2004; Chey, Na, Park, Park & Lee, 1999). 본 연구에서 연령의 효과는 기억과 관리기능 소검사의 변화량에 대해서만 유의미한 것으로

www.kci.go.kr

나타났는데, 연령이 적을수록 재검사 시 점수가 더 향상되는 경향을 보여 젊을수록 사전 경험으로부터 더 많은 이득을 얻는 것으로 나타났다. Pedraza 등의 연구(2007)에서는 총점만을 다루고 있어 각 소검사별 결과를 비교하기는 어려우나, 본 연구에서와 마찬가지로 총점의 변화에는 연령이 영향을 미치지 않는 것으로 나타났다. 전체 집단에서 유의미하게 상승한 점수는 기억 소검사에 국한되었지만, 위 결과를 종합해 볼 때 각 개인의 점수 변화량은 기저 수행수준, 연령, 교육에서 그 개인의 수준에 따라 차별적으로 관찰되었다. 즉, Chelune 등(1993)의 RCIp에서 가정하고 있는 것처럼 전체 집단에 고정된 크기의 연습효과를 기대하기는 어려운 것으로 시사된다.

검사-재검사의 시간 간격은 각 측정치의 점수 변화에 큰 영향을 미치지 않는 것으로 나타났다. 하지만 본 연구에서는 표본 수가 1년 전후 시간 간격에 밀집되어 있어, 본 연구의 결과만으로 2년, 3년, 4년 등 검사 간격이 보다 길어질 때 이것이 점수 변화에 영향을 미치지 않는다고 단정하기는 어렵다. Pedraza 등의 DRS 연구(2007)에서는 검사 간격이 9-15개월 집단과 16-24개월인 집단의 연습효과 크기를 비교하였는데, 본 연구 결과와 유사하게 큰 차이를 보이지 않았다. 하지만 Tempkin 등의 연구(1999)에서는 지능 검사에서 검사-재검사 간격이 길어질수록 재검사 점수가 낮아지는, 즉 연습효과가 상쇄되는 것으로 나타났다. MicgroCog를 이용한 Raymond 등(2006)의 노인 연구에서는 평가 영역에 따라 정도의 차이는 있지만, 2주 간격 재검사에 비해 3개월 간격 재검사에서 연습효과를 더 작게 보고하였다. 평가 영역에 따라 정도의 차이는 있겠으나 인지노화가 진행되는 노인들은 검사 간격이 길어질수록 인지노화가 연습효과를 상쇄시키는 방향으로 영향을 미칠 수 있다. 본 연구에서는 검사-재검사 간격이 1년 내외(7-17개월)인 사람들로 규준집단을 제한하여 규준을 제시하였는데, 검사 간격이 이 범위를 벗어나는 경우에도 이 규준을 적용할 수 있을 지는 추후 연구를 통해 밝혀야 할 것이다.

검사-재검사에 따른 변화의 유의성을 보다 잘 탐지해낼 수 있는 기준점을 마련하기 위해서 본 연구에서는 RCIp와 SRB를 이용하여 예측구간을 구하고, 각각의 기준에서 상승집단, 감퇴집단으로 분류된 비율을 제시하였다. 90% 예측구간을 적용했을 때 상승집단과 감퇴집단으로 분류될 비율은 이론적으로는 각각 5%로 기대된다. 본 연구의 규준 자료에서 실제 관찰된 비율은 상승집단, 감퇴집단 모두 RCIp를 적용했을 때 5%에 더 가깝게 나타났다. 하지만 RCIp에서는 상승자나 감퇴자로 분류되는 사람들이 기저수행 수준에 따라 특정 집단에 편향되는 경향이 관찰되었다. 구체적으로 기저 수행이 낮은 사람들을 상승자로 과잉추정하고 기저 수행이 높은 사람들을 감퇴자로 과잉 추정하는 편향성을 보였다. 이에 비해, SRB에서는 상승자 비율이 통계적 기대치인 5%에 유의미하게 못 미쳤지만 기저 수행이 평균 이상이면 예측구간의 상한선이 만점을 넘어 유의미하게 상승하는 사례가 발생할 수 없어 상승자가 대부분 기저 수행이 평균 이하인 사람들 중에서 발생한다는 점을 감안하면 납득할 만한 수준의 상승자 비율을 보였다. 또한 개념화 소검사에서는 감퇴자의 비율을 높게 추정하는 경향이 있었지만, RCIp에서처럼 기저 수행 수준이 높은 사람들에게 편향되어 감퇴자로 분류하는 경향은 보이지 않았다. Temkin 등(1999)은 감퇴자와 상승자의 비율을 분석하

www.kci.go.kr

여 변화 유의 지수를 계산하는 여러 모델을 비교하고 다중 회귀 모델과 함께 RCIp가 바람직한 모델이라는 결론을 내린 바 있다. 하지만 본 연구 결과에서 알 수 있듯이, 상승자와 감퇴자의 비율을 통계적 기대치인 5%와 단순히 비교하는 것은 때로는 잘못된 결론에 이를 수 있다. 이 연구의 RCIp 결과에서처럼 관찰된 감퇴자나 상승자의 비율이 통계적 기대치인 5%에 근접한다 해도 감퇴자나 상승자로 분류된 사람들이 기저 수행 수준에 따라 체계적으로 편향되어 나타난다면 변화 유의성을 평가하는 바람직한 방법이라고 보기 어렵다.

RCIp와 SRB에서 예측구간은 자료의 정규분포를 가정하여 계산된 구간이기 때문에 RCIp에서 검사-재검사 점수의 차이가 정규 분포하지 않거나 SRB에서 잔차들의 분포가 정규분포하지 않으면, 각 방법에 의해 계산된 90% 예측 구간으로 분류한 상승자와 감퇴자 각각의 비율은 5%와 다를 수 있다. 일반적으로 신경심리 검사는 임상적 진단을 목적으로 비정상적으로 저하된 수행을 변별해내기 위해 평균 수준 이하의 인지능력으로도 해결할 수 있는 쉬운 과제들이 많이 할당되어 있다(Crane, Narasimhalu, Gibbons, Mungas, Haneuse, Larson, Kuller, Hall, & van Belle, 2008). 이 때문에 신경심리 검사는 점수 분포가 정규 분포를 이루지 않고 부적으로 편향되는 경우가 흔하며, K-DRS도 예외는 아니다(Chey et al., 1999). 기저 수행의 분포가 지나치게 부적으로 편향된 검사에서는 천정효과(ceiling effect) 때문에 점수가 상승될 수 있는 사람이 제한되어서 검사-재검사 점수 차이의 분포가 정적으로 편향될 가능성이 있다. 이러한 자료에 변화량의 정규분포를 가정한 예측 구간을 그대로 적용하면 상승자의 비율이 5%보다 낮게 관찰될 것이다. 특히 본 연구에서처럼 어느 수준 이상의 기저 점수에 대한 예측구간 상한선이 해당 검사의 만점을 넘어선다면, 기저 수행이 높았던 사람들은 유의미한 상승자로 분류될 가능성이 전혀 없게 된다. 그럼에도 불구하고 본 연구에서 RCIp로 분류된 상승자의 비율이 5%에 근접한 것은 기저 수행이 낮았던 사람들 중에 상승자가 과잉 추정되었기 때문으로 보인다.

변화 유의성을 판단하기 위한 근거로 제시하는 예측구간은 통계적인 의미에서 정상적인 변화의 범위를 정의한다. 이 구간을 벗어날 만큼 극단적인 향상이나 감퇴를 보이는 사람들을 유의미한 상승이나 유의미한 감퇴를 보이는 것으로 간주한다. 하지만 이러한 예측구간이 실제 관심 임상 집단과 정상 노인 집단을 진단적으로 정확히 변별해주는 적정 경계 점수라고 단언하기는 어렵다. 정상 규준 집단만을 대상으로 설정한 이와 같은 경계 점수는 높은 특이도는 보장할 수 있지만, 실제 유의미한 변화가 발생한 임상 집단을 정확히 변별하는 민감도는 보장할 수는 없기 때문이다. 따라서 본 연구에서 제시한 변화 유의지수에 대한 예측구간이 노인기에 발병할 수 있는 여러 인지 장애의 발병을 예측하고 진단하는 데 실제로 도움이 되는 지에 대한 후속 연구가 필요하다. 본 연구에서 제시한 예측구간은 일차적으로 정상적인 범위의 점수 기복을 배제하는 수준으로 적용할 수 있을 것이다. 평가 목적에 따라서는 규준 자료에만 근거한 예측구간을 이용하기 보다는 관심 임상 집단과 정상 집단을 경험적으로 비교하여 경계점수를 탐색하는 집단 비교법(contrasting-group method; Koffler, 1980)으로 적정 경계 점수를 설정하여 적용하는 것이 더 적절할 수도 있다.

본 연구는 국내 최초로 신경심리검사의 재

www.kci.go.kr

검사 시 연습효과에 영향을 미치는 변수들을 탐색하고 변화의 유의성을 판단하기 위한 규준 자료를 제시하였다. 이 연구 결과는 추후 K-DRS를 이용하여 치매를 조기 진단하고, 치매의 진행 경과를 탐색하고, 치료적 개입의 효과를 논의하는 데 유용한 기준점이 될 수 있다. 국내에서 신경심리학은 최근 15 년간 급속도로 성장해왔으며, 임상장면과 연구에서 많은 신경심리 검사가 사용되고 있다. 또한 최근에는 신경심리검사 결과의 정확한 해석을 위해 연구와 임상 장면에서 모두 종단적인 추적이 강조되고 있는 추세이다. 그럼에도 불구하고 검사-재검사에 따른 변화에 대한 규준을 제공하고 있는 신경심리검사는 전무한 상태로 본 연구가 추후 신경심리검사의 반복측정과 관련된 연구들의 시발점이 될 수 있을 것으로 기대된다.

참고문헌

강연욱, 박재설, 유경호, 이병철 (2009). 혈관성 인지 장애 선별검사로서 Korean-Montreal Cognitive Assessment(K-MoCA)의 신뢰도, 타당도 및 규준 연구. 한국심리학회지: 임상, 28(2), 549-562.

권용철, 박종한 (1989). 노인용 한국판 Mini-Mental State Examination(MMSE-K)의 표준화 연구 - 제 1편: MMSE-K의 개발. 신경정신의학. 28(1), 125-135.

문혜성, 최진영 (2004). 한국 장노년 성인의 문식성과 신경심리기능. 노인정신의학, 8(2), 113-120.

안효정, 최진영 (2004). 노인용 이야기 회상 검사의 표준화 연구. 한국심리학회지: 임상, 23(2), 435-454.

최진영 (1998). 한국판 치매 평가 검사: Korean-Dementia Rating Scale. 서울: 학지사.

최진영, 나덕렬, 박선희, 박은희 (1998). 한국판 치매 평가 검사의 타당도와 신뢰도 연구. 한국심리학회지: 임상, 17(1). 247-258.

Chelune, G. J., Naugle,R. I., Luders, H., Sedlak, J.,&Awad, I. A. (1993). Individual change after epilepsy surgery: Practice effects and base-rate information. *Neuropsychology*, 7, 41-52.

Chey, J., Na, D. R., Park, S., Park, E., & Lee, S. (1999). Effects of education in dementia assessment: Evidence from standardizing the Korean-Dementia Rating Scale. *The Clinical Neuropsychologist,* 13(3), 293-302.

Chey, J., Na, D. G., Tae, W. S., Ryoo, J. W., Hong, S. B. (2006). Medial temporal lobe volume of nondemented elderly individuals with poor cognitive functions. *Neurobiology of Aging, 27(9)*, 1269-1279.

Crane, P. K., Narasimhalu, K., Gibbons, L. E., Mungas, D. M., Haneuse, S., Larson, E. B., Kuller, L., Hall, K., & van Belle, G.(2008). Item response theory facilitated cocalibrating cognitive tests and reduced bias in estimated rates of decline. *Journal of Clinical Epidemiology, 61(10)*, 1018-1027.

Cummings, J. L., & Benson, D. F. (1983). *Dementia: A clinical approach.* Boston: Butterworth.

Freedman M. Kaplan. E., Delis, D., & Morris, R. (1994). *Clock Drawing: A Neuropsychological Analysis.* New York: Oxford University Press.

Frerichs, R.J., Tuokko, H.A. (2005). A comparison of methods for measuring cognitive change in

www.kci.go.kr

older adults. *Archives of Clinical Neuropsychology,* 20, 321-333.

Heaton, R. K., Temkin, N., Dikmen, S., Avitable, N., Taylor, M. J., Marcotte, T. D., & Grant I. (2001). Detecting change: A comparison of three neuropsychological methods, using normal and clinical samples. *Archives of Clinical Neuropsychology,* 16, 75-91.

Hensel, A., Angermeyer, M. C., & Riedel-Heller, S.G. (2007). Measuring cognitive change in older adults: reliable change indices for the Mini-Mental State Examination. *Journal of Neurology, Neurosurgery and Psychiatry,* 78, 1298-1303.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology,* 59, 12-19.

Jurica, P. J., Leitten, C. L., & Mattis, S. (2001). *Dementia Rating Scale-2: Professional Manual.* Lutz, FL: Psychological Assessment Resources.

Koffler, S. L. (1980). A comparison of approaches for setting proficiency standards. *Journal of Educational measurement, 17*, 167-178.

Lucas, J. A., Ivnik, R. J., Smith, G. E., Bohac, D. L., Tangalos, E. G., Kokmen, E., Graff-Radford, N. R., & Petersen, R. C. (1998). Normative data for the Mattis Dementia Rating Scale. *Journal of Clinical and Experimental Neuropsychology, 20(4)*, 536-547.

Maassen, G. (2000). Principles of defining reliable change indices. *Journal of Clinical and Experimental Neuropsychology,* 22, 622-632.

Mattis, S. (1988). *Dementia Rating Scale (DRS): Professional Manual*. Odessa, FL: Psychological Assemssment Resources.

McCaffrey, R. J., Ortega, A., & Haase, R. F. (1993). Effects of repeated Neuropsychological assessment. *Archives in Clinical Neuropsychology*, 8, 519-524.

McSweeny, A. J., Naugle, R. I., Chelune, G. J., & Luders, H. (1993). "T scores for change":An illustration of a regression approach to depicting change in clinical neuropsychology. *The Clinical Neuropsychologist,* 7, 300-312.

Pedraza O, Smith GE, Ivnik R J, Willis F B, Ferman T J, Petersen R C, Graff-Radford N R, Lucas J A. (2007). Reliable change on the Dementia Rating Scale. *Journal of International Neuropsychological Society,* 13, 716-720.

Raymond, P. D., Hinton-Bayre, A. D., Radel, M., Ray, M., & Marsh, N. A. (2006). Test-retest norms and reliable change indices for the MicroCog battery in a healthy community population over 50 years of age. *The Clinical Neuropsychologist, 20*, 261-270.

Rilling, L. M., Lucas, J. A., Ivnik, R. J., Smith, G. E., Willis, F.B., Ferman, T.J., Petersen, R. C., & Graff-Radford, N. R. (2005). Mayo's Older African American Normative Studies: Norms for the Mattis Dementia Rating Scale. *The Clinical Neuropsychologist,* 19, 229-242.

Salinsky, M. C., Storzbach, D., Dodrill, C. B., & Binder, L. M. (2001). Test-retest bias, reliability, and regression equations for neuropsychological measures repeated over a 6-12 week period. *Journal of International Neuropsychological Society,* 7, 597-605.

Shay, K. A., Duke, L. W., Conboy, T., Harrell, L.

www.kci.go.kr

E., Callaway, R, & Folks, D. G. (1991). The clinical validity of the Mattis Dementia Rating Scale in staging Alzheimer's dementia. *Journal of Geriatric, Psychiatric, and Neurology, 4,* 18-25.

Temkin, N. R., Heaton, R. K., Grant, I., & Dikmen, S. S. (1999). Detecting significant change in neuropscyhological test performance: A comparison of four models. *Journal of the International Neuropsychological Society,* 5, 357-369.

Winblad, B., Palmer, K., Kivipelto, M., Jelic, V., Fratiglioni, L., Wahlund, L.-O., Nordberg, A., Bäckman, L., Albert, M., Almkvist, O., Arai, H., Basun, H., Blennow, K., de Leon, M., DeCarli, C., Erkinjuntti, T., Giacobini, E., Graff, C., Hardy, J., Jack, C., Jorm, A., Ritchie, K., van Duijn, C., Visser, P., & Petersen, R. C. (2004). Mild cognitive Impairment: beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. *Journal of Internal Medicine, 256, 240-246.*

원고접수일 : 2010. 2. 2.
게재결정일 : 2010. 3. 3.

www.kci.go.kr

한국심리학회지 : 임상
The Korean Journal of Clinical Psychology
2010, Vol. 29, No. 2, 505-523

Detecting Significant Cognitive Changes in Adults Aged Fifty Years and Older Using The Korean-Dementia Rating Scale

Hoyoung Kim **Jeanyung Chey** **Jungsuh Suk**

Department of Psychology, Seoul National University

A major use of neuropsychological evaluations is to measure changes in cognitive functioning over time. However, it is difficult to determine whether a difference in test performance indicates a real change in the individual or simply a product of normal variation. The current study established normative rates of change in the Korean Dementia Rating Scale (K-DRS) from baseline to follow-up testing among 247 neurologically normal adults aged 50 years and older. Normative retest data were presented for two analytic techniques: the Reliable Change Index adjusted for practice (RCIp) and the Standardized Regression-Based Technique (SRB). With respect to K-DRS total scores and all subtest scores, the most powerful predictor of changes in performance was initial performance. This results suggested strong regression to the mean. Education and age were also significant predictors for some subtests. Further, participants who were more highly educated as well as younger participants demonstrated more improvement with regard to test performance. Considering the prediction intervals and the base rates of decliner and improver, the SRB seemed to provide more unbiased, reasonable prediction intervals than the RCIp.

Key words : practice effect, reliable change index, standardized regression based, prediction interval, retest

www.kci.go.kr