

베일리영유아발달검사 제3판(Bayley-III)의 미국 기준 적용의 문제: 미숙아 집단을 대상으로*

임 유 진 방 희 정* 이 순 행†

이화여자대학교 심리학과

베일리영유아발달검사 제3판(Bayley-III)의 미국 기준이 영유아의 발달수행을 과대추정하고 발달지연율을 과소추정 한다는 논란이 있어서, 본 연구는 발달지연의 고위험군인 미숙아를 대상으로 미국 기준 적용시 Bayley-III의 임상적 유용성을 살펴보고자 하였다. 본 연구는 생후 1개월부터 24개월(교정원령)의 미숙아(재태기간 37주 미만)와 만삭아 93명을 대상으로 Bayley-III와 한국형베일리영유아발달검사 2판(K-BSID-II)을 실시하였다. 연구결과는 다음과 같다. 첫째, 미숙아/만삭아 각 93명을 대상으로 혼합설계 분산분석을 실시한 결과 재태기간(경증미숙아, 중증미숙아, 극중증미숙아, 만삭아)에 따른 Bayley-III의 발달지수 및 척도점수 간에 유의미한 차이가 있었다. 특히 경증미숙아 집단 내에서는 발달영역(인지, 언어, 운동)에 따라 유의미한 차이가 있는 것으로 나타났다. 둘째, Bayley-III와 K-BSID-II을 모두 실시한 미숙아(n=79)를 대상으로 환산점수 평균 간의 차이를 살펴본 결과, Bayley-III의 발달지수는 K-BSID-II의 발달지수에 비해 유의미하게 높은 것으로 나타났으며 발달지연률은 크게 감소하는 것으로 나타났다. 마지막으로 ROC 분석을 통해 발달지연 변별에 대한 Bayley-III의 민감도와 특이도를 분석하고 Bayley-III의 임상적 절단점을 탐색한 결과, Bayley-III의 하위 척도 모두 우수한 수준의 변별정확도를 나타냈으나, 미국 기준 적용시 절단점은 더 높게 나타났다. 본 연구는 미국 기준을 사용하여 Bayley-III를 임상 현장에 적용할 때 점수 해석의 주의를 알리는 실증적인 자료로서, 한국 기준 마련시 표집 문제와 관련된 논의를 제공하였다.

주요어 : 베일리 영유아 발달검사 3판(Bayley-III), 미숙아, 기준, 임상적 유용성

* 본 연구는 2012년 보건복지부 사회서비스 R&D ‘한국형 영유아 발달검사 서비스 개발’ 연구의 일부임.

본 연구는 임유진의 석사학위 논문 자료의 일부를 발췌, 수정한 것임.

† 교신저자 1: 방희정, 이화여자대학교, 서울특별시 서대문구 이화여대길 52, E-mail: hjbang@ewha.ac.kr

‡ 교신저자 2: 이순행, 이화여자대학교, 서울특별시 서대문구 이화여대길 22, E-mail: dynamic9@hanmail.net

소아과 의사이자 심리학자인 Gesell은 40여 년 전에 예일아동연구소(Yale Child Study Center)의 발달검사를 제작하였다. Gesell에 따르면, 어린 아동의 발달을 검사할 때는 의사소통 및 인지, 그리고 언어능력 등을 자연스럽게 이끌어내기 위해서 적절한 장난감을 사용하는 것이 중요하며, 구조화된 검사 상황을 통해서 이러한 행동 반응을 관찰하는 것이 필요하다고 강조하였다(Gesell, Harverson, Castner, Ames, & Amatruda, 1940). Bayley는 이러한 전통을 계승하고 체계적으로 정리하여 표준화된 발달 검사인 베일리영유아발달검사(Bayley Scales of Infant Development, 이후 BSID)를 개발하였다(Bayley, 1993, 2006). BSID는 현재 전 세계적으로 가장 널리 사용되는 영유아발달검사로서(Anderson, Cinzia, De Luca, Hutchinson, Roberts, & Doyle, 2010; Greene, Patra, Nelson, & Silvestri, 2012; Johnson, Moore, & Marlow, 2014; Lowe, Erickson, Schrader, & Duncan, 2011; Luttkhuizen dos Santos, de Kieviet, Konigs, van Elburg, & Oosterlaan, 2013; Moore, Johnson, Haider, Hennessy, & Marlow, 2012), 발달지연이 있는 영유아를 진단하고 조기개입이 필요한지 여부를 결정하며 개입에 따른 효과를 측정하기 위해 임상장면 및 연구용으로 광범위하게 활용되고 있다(Aylward, 2009; Dezoete, MacArthur, & Tuck, 2003; Greene et al., 2012; Ross, Lipper, & Auld, 1985; Skranes, Vik, Nilsen, Smevik, Anderson, & Brubakk, 1998; Siegel, 1983; William, Lewandowski, Coplan, & D'Eugenio, 1987).

BSID는 1969년에 1판이 발표되었고, 1993년에 베일리영유아발달검사 2판(Bayley Scales of Infant Development-2nd Edition, 이후 BSID-II)으로 개정되었다. BSID와 BSID-II는 인지(mental) 척도와 동작(motor) 척도로 구성되어 있어서, 인

지척도에서 인지와 언어발달을 함께 측정하였으며, 동작척도에서는 소근육운동과 대근육운동을 함께 측정하였다. BSID-II는 발달지연 판별을 위한 진단검사(Gold standard)로서 널리 사용됨에도 불구하고, 다음과 같은 측면에서 비판을 받아왔다. 첫째, BSID-II는 인지척도와 동작척도로만 구성되어 있어서 인지와 언어, 대근육운동, 소근육운동에 관한 개별적인 표준 점수를 제공하지 못한다는 것이다. 이로 인해 인지척도의 낮은 점수가 인지발달에서의 지연인지, 언어발달에서의 지연인지, 아니면 두 영역 모두의 지연인지에 대한 세부적인 정보를 제공해주지 못한다는 한계점을 지니고 있다(Siegel, Cooper, Fitzhardinge, & Ash, 1995). 두 번째, BSID-II는 검사실시방식에 있어서 연령별 문항세트(예를 들어, 32~34개월 영아에게는 138번~166번의 인지척도 문항 실시)를 두고 있어서 아동 능력의 기저선(basal limit)과 천정선(ceiling limit)을 제대로 반영하지 못한다는 것이다(Johnson & Marlow, 2006). BSID-II는 첫 번째 제한점으로 지적된 것처럼 인지척도와 동작척도로만 구성되어 있는데, 문항세트 내의 통과한 문항과 실패한 문항의 수에 따라서 기저선(예를 들어, 인지척도의 경우, 문항세트 내에서 5개 이상의 문항에서 통과)과 천정선(문항세트 내에서 3개 미만의 문항에서 실패)을 결정하므로, 발달영역에 따라 불균형한 발달을 보이는 영유아의 능력을 제대로 파악하지 못한다는 것이다. 예를 들어 뛰어난 언어능력을 가진 유아라도 문항세트 내의 문항만을 실시하므로 비언어적 능력에서의 한계로 3개 이상 실패하면(인지척도 천정선) 더 이상 문항을 실시할 수 없게 되며 이로 인해 아동의 뛰어난 언어적 능력에서의 최대 수행은 파악하기 힘들다는 한계점을 지니고 있었다. 세 번째는 교

정연령(corrected age) 사용과 관련하여 문항세트 및 규준표 결정의 문제이다. 미숙아(재태기간 37주 미만)의 경우, 생활연령에 따라서 문항세트를 실시하느냐 아니면 교정연령에 따라서 하느냐에 따라서 문항의 시작점이 달라지고 원점수 비교를 위한 규준표가 달라지므로 발달지수 역시 큰 차이가 있게 된다(Ross & Lawson, 1997). 또한 발달지연이 있는 영유아의 경우 생활연령에서 검사를 시작하게 되므로 기저선과 천정선을 충족시키는 문항세트를 찾기 위해서 상당히 많은 문항을 실시해야 하며, 혹시나 적절한 발달월령을 추정하여 검사를 시작한다고 해도 문항세트만을 실시하므로 생활연령에서 시작하는 것에 비해서 능력이 과소평가될 수 있다(Washington, Scott, Johnson, Wendel, & Hay, 1998).

베일리영유아발달검사 제 3판(Bayley Scales of Infant and Toddler Development-Third Edition, 이하 Bayley-III)은 BSID-II의 제한점을 보완하여 검사의 구성 및 실시방식을 다음과 같이 개선하였다. 먼저 검사의 구성에 있어서 Bayley-III는 BSID-II의 인지척도를 인지, 수용언어, 표현언어 척도로 세분화하였고, BSID-II의 동작척

도를 대근육운동과 소근육운동 척도로 세분화하였다(표 1 참조). 또한 기존의 BSID-II에서는 측정하지 않았던 사회정서영역과 적응행동영역을 추가하였다. 이로 인해, BSID-II의 문제점으로 지적되었던 발달 영역별 세분화된 표준점수를 제공하지 못한다는 한계점을 보완하였다. 즉, Bayley-III는 인지발달과 구분되는 언어발달의 표준점수인 언어발달지수를 제공할 뿐 아니라 수용언어척도점수와 표현언어척도점수로 세분화된 표준점수를 제공하고 있다. 두 번째, 검사의 실시방식에서도 기존의 연령별 문항세트 실시방식을 중단하고, Bayley-III는 기저선의 기준은 낮추고(3개 이상의 문항을 연속적으로 통과), 천정선의 기준은 높임으로써(5개 이상의 문항을 연속적으로 실패) 최대 수행과 최저 수행의 평가를 확장하였다. 또한 Bayley-III는 임상적 유용성을 확장하고자 표준화 과정에서도 이전의 버전들과 차이를 두어 전집에 고위험영아 집단들을 포함하였다(Bayley, 2006). 발달지연을 초래할 수 있는 고위험 집단인 다운증후군, 뇌성마비, 언어문제, 자폐스펙트럼장애, 미숙아 등을 포함하여 규준표집 인원의 10%를 구성하였다. 이는 “정상

표 1. BSID-II와 Bayley-III 검사의 구조

BSID-II ^a	Bayley-III ^b	
측정영역(발달지수)	측정영역(발달지수)	하위척도(척도점수)
인지척도 (인지발달지수, MDI ^c)	인지발달 (인지발달지수)	인지(인지척도)
	언어발달 (언어발달지수)	수용언어(수용언어척도) 표현언어(표현언어척도)
동작척도 (동작발달지수, PDI ^d)	운동발달 (운동발달지수)	대근육운동(대근육운동척도) 소근육운동(소근육운동척도)

a. BSID-II 베일리영유아발달검사 2판, b. Bayley-III 베일리영유아발달검사 3판,

c. MDI(Mental Developmental Index), d. PDI(Psychomotor Developmental Index)

(normal)” 모집단에는 고위험 영유아들이 포함되어 있는데, 표준 표집에서 이러한 영유아들을 배제한다면 그 검사의 점수는 정상 모집단의 특성을 반영하는 것이 아니라는 논의를 반영한 것이다(McFadden, 1996). 이상과 같이 Bayley-III는 검사 구조의 세분화 및 “정상” 모집단의 특성을 반영하는 검사 도구를 제작함으로써 이전 버전들과 차별을 두고자 하였으나, 오히려 점수의 해석과 관련하여 다음과 같은 2가지 논쟁을 지속적으로 불러일으키고 있다.

첫 번째는 검사 구조의 변화로 인해 이전 버전에서 얻은 표준점수와 Bayley-III 표준점수 간의 비교가 어렵다는 것이다(Anderson et al., 2010; Lowe et al., 2011; Moore et al., 2012). 특히, 동시대 또래 아이들과 비교하여 개입에 따른 발달적 진보를 중단적으로 추적하고자 할 때, 발달점수에서의 상승이 개입에 따른 상승인지 아니면 검사도구의 변화에 따른 것인지가 불명확하게 되었다. 이 문제는 두 번째 문제와 관련되어 더욱 큰 이슈가 되고 있다. 점수 해석과 관련된 두 번째 논쟁은 Bayley-III 표준화 및 미국 규준에 관한 것이다. 지능검사나 발달검사는 시간 경과에 따라 평균 점수가 전반적으로 상승하는 Flynn효과가 나타나곤 한다. 이에 이전 규준을 사용하면 Flynn효과에 의해서 검사 점수가 상승하나, 최근에 개정된 규준을 사용하면 동시대 또래들과 비교하므로 평균점수가 하락하는 결과가 나타난다(Herrnstein & Murray, 1994). 예를 들어, BSID-II가 개발되어 새로운 규준을 적용하였을 때 정상 영유아와 고위험군 영유아 모두 1판에 비해 낮은 표준점수를 받았다(Gagnon & Nagle, 2000; Glenn, Cunningham, & Dayus, 2001). 그런데 Bayley-III의 미국표준화 자료에서는 대조적

인 결과가 보고되었다. Bayley-III 매뉴얼에 따르면(Bayley, 2006), Bayley-III의 표준점수는 전반적으로 BSID-II 표준점수에 비해 7점정도 높아졌다고 보고하였다. 즉, 102명의 영유아에게 BSID-II와 Bayley-III를 실시하였을 때, BSID-II의 인지발달지수(Mental Developmental Index, 이후 MDI)의 평균점수는 93.9점이었으나 Bayley-III의 인지발달지수(Cognitive Composite Scores)와 언어발달지수(Language Composite Scores) 평균점수는 각각 102점과 99.7점으로 MDI에 비해 8.1점과 5.8점 상승하였으며, BSID-II의 동작발달지수(Psychomotor Developmental Index, 이후 PDI)의 평균점수는 94.8점이었으나 Bayley-III의 운동발달지수(Motor Composite Scores) 평균점수는 103.2점으로 PDI에 비해 8.4점 상승하였다. 이러한 결과는 Flynn효과와 매우 반대되는 것으로, 현재의 미국 규준을 사용하여 Bayley-III의 표준점수를 낼 경우, 영유아의 발달능력이 과대추정 될 수 있으며 발달지연으로 분류되는 영유아를 과소추정 할 수 있는 위험을 내포하고 있다. 규준이란 특정한 집단 구성원들이 해당 검사에서 얻은 점수 분포로서, 어느 한 검사에게 얻은 원점수를 의미 있게 비교하고 해석할 수 있도록 통계적으로 처리한 전환척도이다. 대부분의 표준화 심리검사는 규준지향적인 검사로서, 표준화 과정에서 규준집단에 실시하여 얻은 검사결과를 기초로 규준을 작성한다(이종승, 2005). 대부분의 나라들이 Bayley-III 검사 점수를 해석할 때 자국의 규준을 가지고 있지 않으므로 미국 규준을 대신 사용하고 있다. 그런데 미국 규준을 사용한 Bayley-III 점수 해석에서의 문제점이 계속해서 보고되고 있다.

Bayley-III의 발달능력의 과대추정에 관한 논쟁은 Anderson과 동료들(2010)의 연구에서 본격

화되었다. Anderson과 동료들은 재태연령 27주 이하의 호주의 미숙아를 대상으로 Bayley-III를 실시하고 미국 기준을 이용하여 결과를 분석하였다. 그 결과 미숙아 집단의 평균점수는 호주의 대조군에 비해 유의미하게 낮게 나오기는 하였으나 모두 정상범위(94.2~100.4점)에 속하였고, 대조군의 평균점수는 미국 표준화 연구에서 제시한 것보다 다소 더 높은 것(108.2~118.4점)으로 나타났다. 또한 Bayley-III에서 제시하는 미국 기준을 사용하여 미숙아 집단의 발달지연률을 조사했을 때, 인지, 언어, 운동발달에서 발달지연률은 각각 13%, 21%, 16%이었으며, 통제 집단에서의 발달지연률은 경도(mild), 중도(moderate), 최중도(severe)로 분류할 때 각각 13.6%, 2.0%, 0.3%였다. 반면, 미국 기준 대신에 연구에 참여했던 호주 자국의 영유아 발달수행 분포를 기준 자료로 이용하여 발달지연률을 조사했을 때, 미숙아 및 저체중아 집단의 3분의 1이 인지영역에서의 지연을 보였으며, 언어 및 운동영역에서의 지연은 그보다 더 높았다. 이는 미숙아를 대상으로 지속적으로 연구해 온 Anderson과 동료들의 임상적 인상 및 이전의 연구결과와 일치하는 결과라고 논의하면서, Bayley-III의 미국 기준을 적용했을 때 발달지연이 과소추정 된다고 보고하였다(Anderson et al., 2010).

Bayley-III는 표준화 과정에서 임상적 유용성을 증가하는데 초점을 두었음에도 불구하고(Bayley, 2006), 미숙아를 대상으로 Bayley-III를 실시한 독일(Reuner, Fields, Wittke, Löprrich, & Pietz, 2013), 영국(Moore et al., 2012; Johnson et al., Moore, & Marlow, 2014), 캐나다(Acton, Biggs, Creighton, Penner, Switzer, Thomas, Joffe, & Robertson, 2011) 연구에서도 미국 기준을 사용했을 때 발달능력의 과대추정 및 발달지연

률의 과소추정에 대한 유사한 결과들이 보고되고 있다. 또한 우리나라와 함께 동양권에 속하는 타이완연구(Yu, Chen, Wu, Jeng, Hsieh, Lee, Hsu, & Chiu, 2013)에서도 비슷한 결과가 보고되고 있어서 서양 문화권에 국한된 문제가 아닌 것으로 나타났다. 이에 국내 임상 현장 및 연구 분야에서 BSID의 활용도가 증가하고 있는 현 시점에서(김문선, 김진경, 2010; 신희선, 김정미, 임성오, 2007; 박인환, 석혜은, 김이영, 신석호, 방희정, 2015; 박주영, 2010; 신희선, 권범선, 임성오, 2005), 미국 기준을 이용한 Bayley-III의 임상적 변별력에 대한 연구가 시급한 것으로 보인다. 즉, 국내에서도 발달지연의 고위험군인 미숙아 집단을 대상으로 미국 기준을 적용하여 Bayley-III 점수를 해석할 때 발달 능력의 과대평가 및 발달지연률의 과소평가가 나타나는지를 확인하고 관련된 대안 논의가 필요할 것으로 보인다.

이를 위해 본 연구에서 구체적으로 다루고자 하는 연구 문제는 다음과 같다.

첫째, 미숙아 집단(경증, 중등도, 중증)과 만삭아 집단은 Bayley-III의 3가지 발달지수(인지, 언어, 운동)와 5가지 하위척도점수(인지, 수용언어, 표현언어, 소근육운동, 대근육운동)에서 유의미한 차이가 나타나는가?

둘째, Bayley-III와 K-BSID-II의 관련성은 어떠한가?

2-1. Bayley-III의 하위 검사(인지, 수용언어, 표현언어, 대근육운동, 소근육운동)와 K-BSID-II의 하위 검사(인지척도, 동작척도)의 관련성은 어떠한가?

2-2. 미숙아 집단의 Bayley-III 발달지수의 평균과 K-BSID-II 발달지수 평균 간에 차이가 나는가?

2-3. 1 표준편차와 2 표준편차를 적용한

Bayley- III와 K-BSID-II의 발달지연율은 어떠한가?

셋째, 미숙아 집단에서 Bayley-III의 임상적 유용성은 어떠한가?

3-1. 발달지연을 변별하기 위한 Bayley-III의 민감도와 특이도는 어떠한가?

3-2. 발달지연을 변별할 수 있는 Bayley-III의 적절한 임상적 절단점(cut-off)은 무엇인가?

연구 대상

본 연구는 재태기간 37주 미만의 미숙아들을 대상으로 이루어졌다. 미숙아는 그동안 발달검사의 임상적 유용성 연구의 주요 대상이 되어왔다(Aylward, 2009; Dezoete et al., 2003; Greene et al., 2012; Ross et al., 1985; Skranes et al., 1998; Siegel, 1983; William et al., 1987). 왜냐하면 미숙아는 인지, 언어, 운동 등의 영역에서 발달지연의 가능성이 높아 표준화된 발달검사를 통한 조기 선별 및 적절한 개입의 중요성이 강조되어 왔기 때문이다(Xu & Filler, 2005). 발달력 검토 결과 뇌실 내 출혈을 포함하여 심각한 신경학적 문제가 있거나 뇌성마비 진단을 받은 것으로 밝혀진 영유아의 검사 자료는 분석에서 제외하였다. 또한 본 연구에서는 생후 24개월 이상의 미숙아는 연구 대상에서 제외하였는데, 왜냐하면 생후 24개월까지 미숙아 출생으로 인한 영향을 가장 크게 받기 때문이다(이에, Bayley-III 미숙아 교정연령 적용 대상을 출생 후 24개월로 한정하고 있다). 총 111명(남: 70명, 여: 41명)이 모집되었으나, 제외 기준에 해당하는 아동을 제외하고 최종적으로 93명(남아 56명, 여아 37명)의 자료를 분석하였다. 미숙아의 발달 수행 결과에 영향을 미치는 요인으로 보고되고 있는 요

인(Wang, Wang, Huang, & Lin, 1998)인 출생시 체중, 재태기간, 부모의 학력에 대한 정보는 표 2에 제시하였다. 미숙아 집단의 수행을

표 2. 인구통계학적 변인(단위: 명, %)

		미숙아(93)	만삭아(93)
성 별	남	56(60.2%)	57(61.3%)
	여	37(39.8%)	36(38.7%)
교 정 연 령	1개월~5개월	22(22.4%)	26(28.0%)
	6개월~11개월	46(47.0%)	28(30.1%)
	12개월~17개월	21(21.4%)	30(32.3%)
	18개월~24개월	9(9.2%)	9(9.7%)
재 태 기 간	경증 미숙아 32주≤GA ^a <37주	39(41.9%)	
	중증 미숙아 28주≤GA<32주	30(32.3%)	해당사항 없음
	극중증 미숙아 GA<28주	24(25.8%)	
출 생 체 중	저체중 1500g~2499g	24(25.8%)	
	극소 저체중 1000g~1499g	25(26.9%)	해당사항 없음
	초극소 저체중 999g 이하	36(38.7%)	
	무응답	8(8.6%)	
모 교 육 수 준	중졸이하	2(2.2%)	0(0.0%)
	고졸이하	15(16.1%)	14(15.05%)
	대졸이하	64(68.8%)	66(70.97%)
	대학원재 이상	7(7.5%)	7(7.53%)
	무응답	5(5.4%)	6(6.45%)
합계		93(100%)	93(100%)

a. GA: Gestational Age 재태기간

만삭아 집단과 비교하고자 표준화 표본에서 성별 및 연령을 대응시켜 무작위로 대조군 93명을 추출하였다.

도구

베일리 영유아 발달검사 제 3판(Bayley Scales of Infant Development-III)

본 검사에서는 Bayley Scales of Infant and Toddler Development - Third Edition(Bayley, 2006)의 한국형 연구본을 사용하였다. Bayley-III 연구본은 2012년 실시된 예비 연구를 토대로 우리나라 영유아 특성과 언어, 문화에 맞도록 인지척도(오수경, 방희정, 이순행, 2014), 언어척도(이지연, 김영태, 방희정, 이순행, 2014), 운동척도(이순행, 안소현, 이은지, 방희정, 2014) 문항을 수정하여 제작된 것이다. 인지척도 91문항, 수용언어척도 49문항, 표현언어척도 48문항, 소근육운동척도 66문항, 대근육운동척도 72문항, 사회정서척도 35문항, 적응행동척도 241문항으로 구성되어 있다. 인지척도, 언어척도, 운동척도는 검사자가 직접 실시하며 시작 규칙과 중지규칙에 따라 실제 실시하는 문항의 수는 다를 수 있다.

Bayley-III 연구본의 신뢰도 및 타당도를 검증하고 문항분석을 실시한 결과, 한국 영유아에게 적용하는데 적합한 것으로 나타났다(오수경 외, 2014; 이순행 외, 2014; 이지연 외, 2014). Bayley-III 연구본의 Cronbach α 계수는 .71~.98로 적절한 수준이었으며 검사-재검사 신뢰도는 .98, 검사자간 신뢰도는 .98~.99로 신뢰할만한 도구임이 입증되었다. Bayley-III 연구본의 인지척도와 K-Leiter-R과의 상관은 중간 정도로 유의미한 관계를 나타냈으며($r=.503$, 오수경 외, 2014), 언어척도(이지연 외, 2014)도

PRES와 K-M-B-CDI와 각각 .847과 .921의 높은 상관을 보이며 수렴 타당도를 입증하였다. 또한 Bayley-III 연구본의 소근육운동 척도도 Bayley-III의 다른 하위척도 및 발달 영역과 수렴 및 변별타당도가 높은 것으로 입증되었다(이순행 외, 2014).

Bayley-III는 인지, 수용언어, 표현언어, 대근육운동, 소근육운동의 5가지 하위 척도에 대한 원점수(Raw Scores)를 제공한다. 원점수는 준거점이 없고 단위의 동간성이 없기 때문에 의미 있게 비교하고 해석할 수 없다(이종승, 2005). 그러므로 기준을 이용하여 점수를 변환시킨 환산점수를 사용한다. Bayley-III에서는 3가지 유형의 환산점수를 제공하는데, 척도점수(Scaled Scores), 발달지수(Composite Scores), 백분위(Percentile Ranks)가 그것이다. 척도점수는 5가지 하위 척도에서 얻어진 원점수를 평균 10점, 표준편차 3점인 정상분포의 기준을 토대로 변환한 점수로, 점수의 범위는 1점~19점이다. 발달지수는 각 발달 영역의 하위 척도들의 척도점수의 합을 원점수로 하여(예를 들어, 언어 발달 원점수는 수용언어 척도점수와 표현언어 척도점수의 합) 평균 100점, 표준편차 15점인 정상분포의 기준을 토대로 변환한 점수로, 범위는 40점~160점이다. 본 연구에서는 한국 기준이 아직 보고된 것이 없으므로 각 영역간 또는 집단 간 비교가 가능하도록 미국 기준을 이용한 척도점수 및 발달지수를 사용하였다. 한편, 발달지수 해석과 관련하여, 미국 Bayley-III 기준에 따르면 1 표준편차와 2 표준편차 사이에 해당하는 발달지수인 70점~84점 사이를 '경도(mild)' 발달지연으로 분류하고, 2 표준편차 미만에 해당하는 발달지수인 69점 이하를 '중도(moderate)' 발달지연으로 분류한다(Bayley, 2006).

한국 베이리 영유아 발달검사 2판(Korean Bayley Scales of Infant Development-II).

본 연구에서는 Bayley-III의 임상적 유용성을 확인하기 위한 준거 도구로써 한국 베이리 영유아 발달검사 2판의 인지척도와 동작척도를 사용하였다. 행동평정척도는 연구의 목적에 맞지 않아 제외하였고 인지발달지수(MDI)와 동작발달지수(PDI)만 사용하였다. 도구의 신뢰도는 표준화 연구에서 인지척도 $r=.91$, 동작척도 $r=.99$ 인 것으로 보고되었으며(조복희, 박혜원, 2004), 한국 웨슬러 유아지능검사(K-WPPSI)와의 상관성이 인지척도와 동작척도에서 각각

$r=.65$, $r=.52$ 로 나타나 공준타당도 역시 검증된 도구이다. K-BSID-II도 1 표준편차와 2 표준편차 사이에 해당하는 발달지수인 70점~84점 사이를 '경도' 발달지연으로 분류하고, 2 표준편차 미만에 해당하는 발달지수인 69점 이하를 '중도' 발달지연으로 분류한다(박혜원, 조복희, 2006).

본 연구의 대상인 1~24개월 연령집단에서 시행되는 K-BSID-II과 Bayley-III 문항의 변화를 요약하면 표 3과 같다.

절차

참여자 모집 및 검사 실시

본 연구를 위한 자료 수집은 2014년 3월부터 10월에 걸쳐 진행되었다. 온라인 육아 커뮤니티, 어린이집, 병원 등에 참여 홍보문을 게시하였고, 참여를 희망하는 영유아의 보호자가 직접 신청하도록 하였다. 검사는 E대학교 및 S센터, A병원에 마련된 검사실에서 이루어졌다. 보호자는 인구통계학적 정보와 기본적인 발달력에 관한 질문지를 작성하도록 요청받았고 사회정서척도와 적응행동척도에 대해 설명을 듣고 평정 방법을 이해한 것이 확인되면 질문지를 작성하도록 요청받았다. 검사자가 영유아에게 직접 실시하는 척도인 인지, 언어, 운동척도의 경우 먼저 Bayley-III를 실시한 후, K-BSID-II의 인지척도와 동작척도 각각에서 겹치는 문항을 채점한 뒤 남아있는 문항을 추가적으로 실시하는 방식으로 진행하였다. 같은 문항은 한번만 실행하도록 하여 연습효과를 방지하고 검사시간을 단축시켜 영유아의 집중력 저하 및 피로효과를 방지하고자 하였다. 만약 K-BSID-II에서 Bayley-III와 문항 내용이나 실시 방법은 같으나 구체적인 채점기준이 다른 문항

표 3. K-BSID-II와 Bayley-III의 문항 변화
(1-24개월 연령 문항세트)

K-BSID-II 문항세트	Bayley-III 문항			
	인지	언어	운동	
1-5 개월	유지	5(15%)	1(7%)	17(47%)
	수정	15(46%)	6(40%)	14(39%)
	삭제	11(33%)	2(13%)	3(8%)
	추가	2(6%)	6(40%)	2(6%)
6-11 개월	유지	6(38%)	0(0%)	8(19%)
	수정	4(25%)	3(27%)	22(54%)
	삭제	5(31%)	2(18%)	7(17%)
	추가	1(6%)	6(55%)	4(10%)
12-17 개월	유지	7(31%)	2(25%)	0(0%)
	수정	9(41%)	3(38%)	6(75%)
	삭제	3(14%)	0(0%)	2(25%)
	추가	3(14%)	3(37%)	0(0%)
18-24 개월	유지	2(40%)	3(18%)	0(0%)
	수정	3(60%)	5(29%)	11(69%)
	삭제	0(0%)	1(6%)	1(6%)
	추가	0(0%)	8(47%)	4(25%)

의 경우, Bayley-III 시행단계에서 실시하고 채점은 K-BSID-II의 기준에 따르는 방식을 적용하였다. 두 검사를 모두 끝내는데 소요되는 시간은 평균 1시간~1시간 40분 정도로 영유아의 연령 단계나 검사 수행 태도의 영향을 받아 개인차가 있었다. 검사 시간이 길어지거나 집중력 저하 등의 문제로 추가적으로 K-BSID-II를 실시하는 것이 어려웠던 경우는 Bayley-III를 완료한 후 검사를 종료하였다. 참여자에게는 개별적으로 발달보고서를 작성하여 메일로 받아볼 수 있도록 하였다. 본 연구는 E대학교 생명윤리위원회(Institutional Review Board, IRB)의 심의를 통과하였다.

자료 분석

자료 분석은 IBM SPSS for Windows Version 22.0와 Medcalc Version 14.0을 사용하였으며 구체적인 내용은 다음과 같다. Bayley-III 3가지 발달지수(인지, 언어, 운동)와 5가지 척도점수(인지, 수용언어, 표현언어, 소근육운동, 대근육운동)의 차이를 분석하고자, 미숙아 집단(경증, 중등도, 중증)과 만삭아 집단을 재태기간에 따라서 4집단으로 분류하여 집단 간 요인과 집단 내 요인을 통합하여 알아볼 수 있는 혼합설계 분산분석(between- and within-subjects design ANOVA)을 실시하였다. 이 때 반복측정에 따른 구형성검증을 위하여 Mauchly's 구형성 검증을 사용하였고, 구형성 가정을 충족시키지 못하는 경우 Greenhouse-Geisser 교정방법을 이용하여 분산분석 결과를 조정하여 보고하였다. 분산분석 결과가 유의미한 경우 사후분석을 실시하였는데, 각 요인의 수준이 세 개인 경우에는 Fisher's LSD 방법을 이용하였으며, 각 집단의 크기가 동일하지 않고 4개 이상의 수준

인 경우에는 Schaffé검증을 사용하였다. 다음으로, Bayley-III와 K-BSID-II 하위 척도들의 관련성을 알아보기 위해 상관분석을 실시하였으며 미숙아 집단에서 Bayley-III와 K-BSID-II 수행에 차이가 있는지 알아보기 위해 paired t-test를 실시하고, 각각의 발달지연률을 계산하였다. 마지막으로 K-BSID-II 실시 결과 중도발달지연으로 분류된 아동(MDI<70, PDI<70)에 대한 Bayley-III의 민감도와 특이도를 확인하고 미국 기준 점수를 이용했을 때의 적절한 절단점을 탐색하고자 ROC 분석을 실시하였다.

결 과

미숙아 집단의 Bayley-III 수행

재태기간에 따라서 미숙아의 발달상태가 상당히 차이가 있다는 선행연구(Hack, Friedman, & Fanaroff, 1996; Msall, 2009)에 따라 미숙아 집단을 극중증(GA<28주 미만), 중증(28주≤GA<32), 경증(32주≤GA<37주)으로 구획을 나누어서 Bayley-III의 발달지수와 척도점수를 살펴보았다. 재태기간에 따른 Bayley-III 발달지수 및 척도점수의 평균과 표준편차는 표 4, 그림 1, 그림 2와 같다. 그림 1과 그림 2에 제시된 것처럼, 재태기간이 길어질수록 Bayley-III의 발달지수와 척도점수의 평균이 모두 상승하는 양상을 보였다. 이에 재태기간을 구획변수로 하여 발달지수 및 척도점수가 다르게 나타나는지 알아보기 위하여 혼합설계에 의한 분산분석을 실시하였다. 먼저, 미숙아의 발달지수의 차이가 있는지 알아보기 위하여, 구획변수인 재태기간(극중증, 중증, 경증, 만삭아)을 집단간 변인으로 하고 발달지수(인지발달지수,

표 4. 재태기간에 따른 K-Bayley-III 척도점수와 발달지수의 평균과 표준편차(N=186)

척도	미숙아(n=93) ^a						만삭아(n=93)	
	극중증 (n=24)		중증(n=30)		경증 (n=39)			
	GA ^b <28		28≤GA<32		32≤GA<37		37≤GA<40	
	평균	(표준편차)	평균	(표준편차)	평균	(표준편차)	평균	(표준편차)
인지 ^c	8.08	(2.97)	9.37	(2.37)	9.70	(3.09)	10.72	(2.91)
수용언어 ^c	8.21	(3.36)	8.27	(2.39)	8.98	(2.58)	9.99	(3.20)
표현언어 ^c	8.00	(2.26)	8.73	(2.30)	8.23	(1.77)	10.16	(2.88)
소근육운동 ^c	8.83	(2.83)	10.17	(2.58)	10.55	(2.68)	10.62	(3.02)
대근육운동 ^c	8.67	(3.37)	9.20	(2.56)	10.34	(3.31)	10.42	(2.75)
인지발달 ^d	90.42	(14.88)	96.83	(11.85)	98.52	(15.19)	103.62	(14.88)
언어발달 ^d	89.04	(12.50)	91.33	(11.18)	92.00	(9.88)	103.61	(14.65)
운동발달 ^d	93.38	(17.42)	97.63	(13.87)	99.66	(14.95)	103.39	(14.25)

- a. 괄호 안은 인원(단위: 명) b. GA: 재태기간(단위: 주),
 c. 척도점수의 평균은 10점, 표준편차는 3점이며, 점수 범위는 1-19점이다.
 d. 발달지수의 평균은 100점, 표준편차는 15점, 점수범위는 40-140점이다.

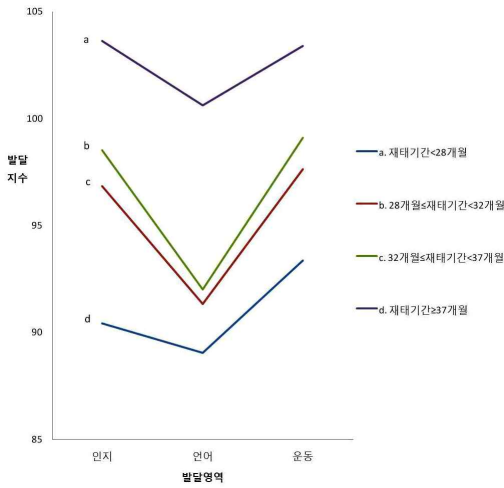


그림 1. 재태기간에 따른 발달지수 평균

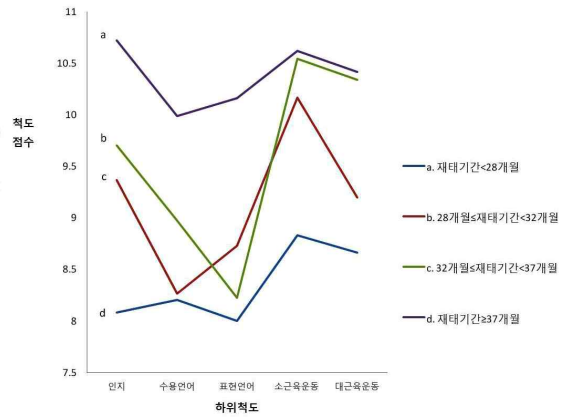


그림 2. 재태기간에 따른 척도점수 평균

언어발달지수, 운동발달지수)를 집단 내 변인으로 하여 분산분석을 실시하였다. 먼저, 각 집단의 차이점수들의 분산이 동일한지 알아보

기 위한 구형성 검정한 결과, 통계적으로 유의미하지 않아 공분산의 동질성 가정이 충족되는 것으로 나타났다. 재태기간에 따른 발달지

수의 차이를 알아본 결과는 표 5와 같다.

표 5에 제시된 것처럼, 재태기간에 따라서 집단 간 발달지수의 유의미한 차이가 있는 것으로 나타났으며[F(3, 192)=8.42, $p < .001$] 집단 내에서도 발달지수 영역 간의 유의미한 차이가 있는 것으로 나타났다[F(2, 384)=11.73, $p < .001$]. 그러나 재태기간과 발달지수의 상호작용 효과는 없는 것으로 나타났다.

발달영역에서 집단 간의 차이에 대한 사후검정을 실시한 결과, 인지발달영역과 운동발달영역에서 극중증 미숙아 집단(GA<28)의 발달지수 평균은 각각 90.42점과 93.38점으로 만삭아 집단(GA≥37)의 인지발달지수와 운동발달지수의 평균인 103.62점과 103.39점에 비해 유의미하게 낮게 나타났다($p < .05$). 또한 언어발달영역에서 극중증(GA<28), 중증(28≤GA <32), 경증(32≤GA<27) 미숙아 집단의 발달지수 평균은 각각 89.04점, 91.33점, 92.00점으로 만삭아 집단(GA≥37)의 언어발달지수 평균인 103.61점에 비해 유의미하게 낮게 나타났다($p < .05$).

또한 집단 내의 발달영역 간의 차이에 대한 집단 사후검정을 실시한 결과, 극중증 미

숙아(GA<28)의 경우 인지(M=90.42), 언어(M=89.04), 운동(M=93.38) 발달지수가 평균 하단 지점에 속하며 발달 영역 간의 유의미한 차이를 보이지 않았으며, 중증 미숙아(28≤GA<32) 집단의 경우 언어발달지수(M=91.33)가 인지발달지수(M=96.83) 및 운동발달지수(M=97.63)에 비해 낮게 나타났지만 그 차이가 유의미한 정도는 아니었다. 반면 경증 미숙아(32≤GA<37) 집단의 경우, 언어발달지수(M=92.00)가 인지발달지수(M=98.52, $p < .01$) 및 운동발달지수(M=99.66, $p < .001$)에 비해 유의미하게 낮은 것으로 나타났다.

다음으로, 미숙아의 하위척도점수의 차이가 있는지 알아보기 위하여, 구획변인인 재태기간(극중증, 중증, 경증, 만삭아)을 집단간 변인으로 하고, 하위척도점수(인지척도, 수용언어척도, 표현언어척도, 대근육운동척도, 소근육운동척도)를 집단 내 변인으로 하여 분산분석을 실시하였다. 먼저 구형성 검정한 결과, 구형성 가정이 충족되지 않아 Greenhouse-Geisser에 의해 조정된 값으로 결과를 해석하였다. 재태기간에 따른 척도점수를 알아본 결과는 표 6과 같다. 표 6에 제시된 것처럼 재태기간에 따라

표 5. 재태기간에 따른 K-Bayley-III 발달지수의 비교(N=186)

분산원	df	F	η^2	p
집단간	195			
재태기간	3	8.42	.12	< .001
오차	192			
집단내	392			
발달영역	2	11.73	.06	< .001
발달영역×재태기간	6	0.89	.01	0.50
오차	384			
합계	587			

표 6. 재태기간에 따른 K-Bayley-III 척도점수의 비교(N=186)

분산원	df	F	η^2	p
집단간	195.00			
재태기간	3.00	7.82	.11	0.00
오차	192.00			
집단내	732.12			
하위척도 ⁺	3.74	8.84	.04	0.00
하위척도×재태기간 ⁺	11.21	1.69	.03	0.07
오차	717.17			
합계	927.12			

+는 Greenhouse-Geisser에 기초해 조정된 교정값

서 하위척도점수 간에 유의미한 차이가 있는 것으로 나타났으며[F(3, 192)=7.82, $p < .001$] 각 집단 내에서도 하위척도점수 간에 유의미한 차이가 있는 것으로 나타났다[F(3.74, 717.17)=8.84, $p < .001$]. 그러나 재태기간과 하위척도점수 간의 상호작용 효과는 없는 것으로 나타났다.

재태기간에 따른 집단 간 하위척도점수 차이에 대한 사후검정을 실시한 결과, 인지척도에서 극중증 미숙아(GA<28) 집단의 평균점수는 8.08점으로 만삭아 집단(GA≥37)의 인지척도 평균점수인 10.72점에 비해 유의미하게 낮게 나타났으며($p < .05$), 표현언어척도에서 극중증 미숙아(GA<28) 집단과 경중 미숙아(32≤GA<37) 집단의 평균점수는 각각 8.00점, 8.23점으로 만삭아 집단의 표현언어척도 평균점수인 10.16점에 비해 유의미하게 낮게 나타났다($p < .05$).

또한 각 재태기간 내 Bayley-III 5가지 하위척도점수에 대한 사후검정을 실시한 결과, 극중증 미숙아(GA<28) 집단의 경우 모든 척도점수가 평균 하단 지점에 속하며 하위 척도 간

의 유의미한 차이를 보이지 않았다. 그러나 중증 미숙아(28≤GA<32) 집단의 경우 소근육운동의 수행(M=10.17)은 미국 표준평균에 해당할 만큼 잘 발달되어 있는 반면 수용언어척도의 수행(M=8.27)은 상대적으로 낮아서 두 척도 간의 유의미한 차이를 나타내었다($p < .05$). 경중 미숙아(32≤GA<37) 집단의 경우는 척도 간의 불균형적인 발달이 더욱 두드러지게 나타나서 소근육운동척도 점수(M=10.55)는 수용언어척도 점수(M=8.98, $p < .05$) 및 표현언어척도 점수(M=8.23, $p < .001$)에 비해 유의미하게 높은 점수를 나타내었으며, 대근육운동척도 점수(M=10.34)도 표현언어척도 점수(M=8.23)에 비해 유의미하게 높은 것으로 나타났다($p < .001$). 인지척도 점수(M=9.70)도 표현언어척도 점수(M=8.23)에 비해 유의미하게 높은 것으로 나타났다($p < .05$).

한 가지 주목할 점은 집단 간 평균차이에도 불구하고, 미국 표준을 이용한 미숙아 집단의 척도점수(8~10점)와 발달지수(89~99점)는 대체로 ‘평균(average)’ 범위 안에 있는 것으로 나타났다는 점이다.

Bayley-III와 K-BSID-II의 관련성 분석

Bayley-III와 K-BSID-II를 동시에 실시한 미숙아 79명을 대상으로 두 검사 점수 간의 관련성을 알아보기 위해 상관분석을 실시하였고, 그 결과는 표 7 상단에 제시하였다. 표 7에서 보는 것처럼, K-BSID-II의 MDI는 이론적으로 관련성이 높은 Bayley-III의 인지척도와 상호상관이 더 높은 것으로 나타났으며($r=.69, p < .001$), 수용언어척도, 표현언어척도, 소근육운동척도, 대근육운동척도, 언어발달지수, 운동발달지수와는 중간 정도의 정적 상관이 있는 것으

로 나타났다($r=.43\sim.58, p < .001$). K-BSID-II의 PDI도 이론적으로 관련성이 높은 Bayley-III의 대근육운동척도($r=.74, p < .001$) 및 운동발달지수($r=.62, p < .001$)와 상호상관이 더 높은 것으로 나타났으며, 인지척도, 수용언어척도, 소근육운동척도, 대근육운동척도, 언어발달지수와는 중간 정도의 유의한 정적 상관을 나타냈다($r=.43\sim.51, p < .001$).

미숙아 집단의 Bayley-III 발달지수 평균과 K-BSID-II의 발달지수 평균 간의 차이를 알아보기 위해 paired t-test를 실시한 결과는 표 7

표 7. 미숙아 집단의 BSID-II와 Bayley-III의 상관관계 및 평균의 차이($n=79$)

Bayley-III ^b	BSID-II ^a		Bayley-III ^b	
	MDI ^c	PDI ^d	평균	표준편차
인지척도	0.69	0.49	9.14	2.76
수용언어	0.48	0.43	8.47	2.76
표현언어	0.43	0.20	8.19	1.98
소근육운동	0.54	0.51	9.97	2.78
대근육운동	0.52	0.74	9.22	3.04
인지발달	0.69	0.49	95.70	13.81
언어발달	0.58	0.43	90.38	10.88
운동발달	0.55	0.62	96.27	15.52
BSID-II	MDI	PDI		
평균	83.97	79.22		
표준편차	20.76	19.48		
차이점수	M	SD	t	p
COG ^e -MDI	11.72	14.67	6.965	<.001
LAN ^f -MDI	6.40	16.93	3.36	.001
MOT ^g -PDI	17.05	15.50	9.687	<.001

a. BSID-II: 베일리영유아발달검사2판, b. Bayley-III: 베일리영유아발달검사3판, c. MDI: BSID-II 인지발달지수, d. PDI: BSID-II 동작발달지수, e. COG: Bayley-III 인지발달지수, f. LAN: Bayley-III 언어발달지수, g. MOT: Bayley-III 운동발달지수

하단에 제시하였다. 표 7에 나타난 것처럼, Bayley-III 발달지수가 K-BSID-II 발달지수에 비해 대략적으로 6점~17점정도 유의미하게 더 높은 것으로 나타났다($p < .001$). 구체적으로 살펴보면, 미숙아 집단의 Bayley-III의 인지 발달지수의 평균은 95.70점($SD=13.81$), 언어발달지수의 평균은 90.38점($SD=10.88$), 운동발달지수의 평균은 96.27점($SD=15.52$)으로 모두 '평균(average)' 범주에 속한 반면, K-BSID-II의 MDI 평균은 83.97점($SD=20.76$), PDI 평균은 79.22점($SD=19.48$)으로 '경계선~평균 하' 범주에 속하였다.

또한 Bayley-III와 K-BSID-II의 1 표준편차와 2 표준편차를 기준으로 발달지연률을 탐색해 본 결과는 표 8과 같다.

1 표준편차(85점 미만)를 기준으로 한 BSID-II의 MDI와 PDI의 발달지연률은 각각 51.9%(41명)와 58.2%(46명)이었으며, 2 표준편차(69점 이하)를 기준으로 한 BSID-II의 MDI와 PDI의 발달지연률은 각각 30.4%(24명)와 35.4%(28명)이었다. 반면에, 1 표준편차(85점 미만)를 기준으로 한 Bayley-III의 인지발달, 언어발달, 운동발달의 발달지연률은 각각 12.7%(10명), 22.8%(18명), 16.5%(13명)로 크게 감소하였으며, 2 표준편차(69점 이하)를 기준으로 한 Bayley-III의 인지발달, 언어발달, 운동발달의 발달지연률도 각각 2.5%(2명), 3.8%(3명), 3.8%(3명)으로 크게 감소하였다. 그러나 하위척도별로 Bayley-III의 발달지연률을 탐색하였을 때는 어느 정도 상승하였는데, 1 표준편차(7점 미만)를 기준으로 한 Bayley-III의 인지척도, 수용언어척도, 표현언어척도, 소근육운동척도, 대근육운동척도의 발달지연률은 각각 30.4%(24명), 36.7%(29명), 34.2%(27명), 13.9%(11명), 26.6%(21명)이었다. 또한 2 표준편차(4점 미만)를 기준으로 한

표 8. BSID-II와 Bayley-III의 발달지연률($n=79$)

BSID-II ^a	N(<85)	N(<70)
MDI ^b	41(51.9%)	24(30.4%)
PDI ^c	46(58.2%)	28(35.4%)
Bayley-III ^d	N(<85)	N(<70)
인지발달지수 ^e	10(12.7%)	2(2.5%)
언어발달지수 ^e	18(22.8%)	3(3.8%)
운동발달지수 ^e	13(16.5%)	3(3.8%)
Bayley-III	N(<7)	N(<4)
인지척도 ^f	24(30.4%)	4(5.1%)
수용언어 ^f	29(36.7%)	6(7.6%)
표현언어 ^f	27(34.2%)	4(5.1%)
소근육운동 ^f	11(13.9%)	4(5.1%)
대근육운동 ^f	21(26.6%)	4(5.1%)

a. BSID-II: 베일리영유아발달검사2판, b. MDI: BSID-II 인지발달지수, c. PDI: BSID-II 동작발달지수, d. Bayley-III: 베일리영유아발달검사3판, e. 발달지수의 평균은 100점, 표준편차는 15점, 점수범위는 40-140점이다. f. 척도점수의 평균은 10점, 표준편차는 3점이며, 점수 범위는 1-19점이다.

Bayley-III의 인지척도, 수용언어척도, 표현언어척도, 소근육운동척도, 대근육운동척도의 발달지연률도 5.1%(4명), 7.6%(6명), 5.1%(4명), 5.1%(4명), 5.1%(4명)로 Bayley-III 발달지수로 살펴본 발달지연률보다 약간 상승하였다.

Bayley-III의 임상적 변별력 확인

Bayley-III의 3가지 발달지수(인지발달지수, 언어발달지수, 운동발달지수)와 5가지 하위척도점수(인지척도, 수용언어척도, 표현언어척도, 소근육운동척도, 대근육운동척도)가 미숙아의 발달지연을 얼마나 정확하게 변별하는지 임상

적 유용성을 확인하고자 ROC(Receiver Operation Characteristic curve, 이하 ROC) 분석을 실시하였다. ROC는 특정 검사의 민감도와 특이도를 표현한 그래프로써, 진단의 정확도는 그래프 아래의 면적(Area under the curve, 이하 AUC)에 의해 측정될 수 있다. AUC의 면적이 1에 가까울수록 완벽한 검사를 의미한다. 구체적인 기준은 Miska와 Jan(2005)의 연구와 이진, 김영아와 오경자(2009)의 연구에 적용되었던 것을 참조하였으며 .90~1.00은 “판별의 정확도가 매우 우수한” 수준, .80~.90은 “우수한” 수준, .70~.80은 “적합한” 수준, .60~.70은 “판별력은 지지될 수 있지만 우수하지는 않은 수준”, .60~.50은 “덜 적합한” 수준으로 보았다.

먼저, K-BSID-II의 MDI점수 2 표준편차(69점 이하)를 발달지연의 기준점으로 하고 인지척도, 수용언어척도, 표현언어척도, 인지발달지수, 언어발달지수를 예측변인으로 한 ROC 분석결과는 그림 3, 표 9와 같다. 그림 3과 표 9에서 보듯이, 인지척도와 인지발달지수의 AUC는 .90으로 매우 우수한 수준이었고($p < .001$),

표 9. Bayley-III 소근육운동, 대근육운동, 운동발달지수에 대한 ROC분석($n=79$)

영역	표준 오차 ^a	근사 95% 신뢰구간	
		하한	상한
FM ^d	.725**	.614	.837
GM ^b	.916***	.855	.976
MOT ^c	0.809***	0.711	0.907

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

a. FM: 소근육운동척도, b. GM: 대근육운동척도, c. MOT: 운동발달지수

언어발달지수의 AUC 역시 .81로 우수한 수준인 것으로 나타났다($p < .001$). 반면, 수용언어척도의 AUC는 .75로 적합한 수준이지만, 표현언어척도의 AUC는 .69로서 판별력이 지지되지만 우수하지는 않은 수준으로 나타났다.

또한, K-BSID-II의 PDI점수 2 표준편차(69점 이하)를 발달지연의 기준점으로 하고 소근육운동척도, 대근육운동척도, 운동발달지수를 예측변인으로 한 ROC 분석결과는 그림 4와 표 10

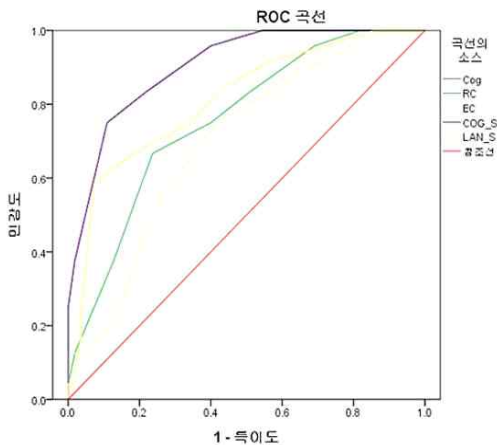


그림 3. Bayley-III 인지척도, 수용언어척도, 표현언어척도, 인지발달, 언어발달지수에 대한 AUC곡선

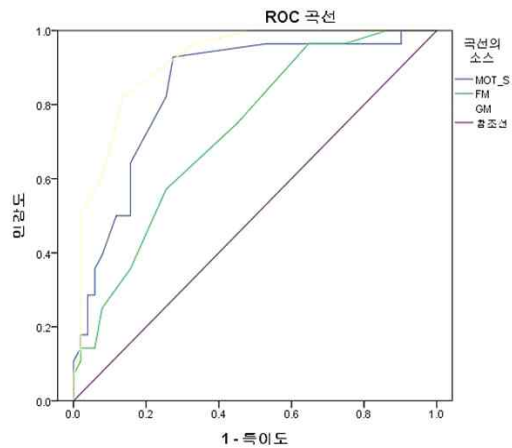


그림 4. Bayley-III 소근육운동척도, 대근육운동척도, 운동발달지수에 대한 AUC곡선

표 10. Bayley-III 소근육운동, 대근육운동, 운동발달지수에 대한 ROC분석($n=79$)

영역	표준 오차 ^a	근사 95% 신뢰구간	
		하한	상한
FM ^a	.725**	.057	.614 .837
GM ^b	.916***	.031	.855 .976
MOT ^c	0.809***	0.05	0.711 0.907

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

a. FM: 소근육운동척도, b. GM: 대근육운동척도, c. MOT: 운동발달지수

과 같다. 그림 4와 표 10에서 보듯이, 대근육운동의 AUC는 .91로 매우 우수한 수준이었고 ($p < .001$), 운동발달지수의 AUC 역시 .80으로 우수한 수준인 반면($p < .001$), 소근육운동척도의 AUC는 .72로 적합한 수준으로 나타났다.

한편, ROC분석을 통하여 발달지연을 가장 잘 변별할 수 있는 절단점은 민감도와 특이도가 동시에 높은 지점이라고 할 수 있다. K-BSID-II에서의 발달지연(≤ 69)을 가장 잘 예측하는 Bayley-III 발달지수를 확인하고자 각 발달영역별로 민감도와 특이도를 산출하였고 그 결과를 표 11에 제시하였다. 본 연구에서는 민감도와 특이도의 합이 최대가 되는 지점을 찾되, 결정이 불분명할 경우 발달지연인 아동을 지연으로 판별해내는 것이 발달지연이 아닌 아동을 아니라고 판별해내는 것보다 더 중요하다고 보고 민감도가 높은 점수를 선택하였다. 인지발달지수는 90점 이하일 때 민감도와 특이도가 최대가 되었고 민감도는 84.6%, 특이도는 75.4%로 나타났다. 언어발달지수는 89점 이하일 때 민감도와 특이도가 최대가 되었고 민감도는 80.8%, 특이도는 66.2%로 나타났다. 운동발달지수는 97점 이하일 때 민감도와

표 11. Bayley-III 인지, 언어, 운동발달지수의 민감도, 특이도에 따른 절단점

	절단점(점)	민감도	특이도	민감도+특이도
인지 척도	≤ 80	.500	0.969	1.469
	≤ 85	.769	0.862	1.631
	≤ 90	.846	0.754	1.600
	≤ 95	.962	0.569	1.531
	≤ 100	1.000	0.446	1.446
언어 척도	≤ 79	.500	0.938	1.438
	≤ 83	.654	0.908	1.562
	≤ 86	.731	0.815	1.546
	≤ 89	.808	0.662	1.469
	≤ 91	.846	0.554	1.400
운동 척도	≤ 91	.727	0.914	1.641
	≤ 94	.818	0.845	1.663
	≤ 97	.879	0.793	1.672
	≤ 100	.970	0.586	1.556
	≤ 103	.970	0.534	1.504

와 특이도가 최대가 되었고 민감도는 87.9%, 특이도는 79.3%였다.

논 의

BSID는 전 세계적으로 가장 널리 사용되는 영유아발달검사로서, 2006년에 3판(Bayley-III)이 출판되었다. 발달 지연의 위험이 있는 영유아를 조기에 규명하고 개입하는 것을 주요 목적으로 하는 발달검사에서 임상적 유용성은 매우 중대한 문제인데, 현재 Bayley-III 미국 규준은 영유아의 발달 수행을 과대추정하고 발달지연률을 과소추정한다는 논의가 계속되고 있다. 이에 본 연구는 발달지연의 고위험군

중 하나인 미숙아를 대상으로 미국 기준을 적용하였을 때의 Bayley-III의 임상적 유용성을 살펴보고자 하였다. 본 연구의 결과 및 그에 따른 논의는 다음과 같다.

첫째, 미숙아 집단(경중, 중등도, 중증)과 만삭아 집단은 Bayley-III의 3가지 발달지수(인지, 언어, 운동)와 5가지 하위 척도점수(인지, 수용언어, 표현언어, 소근육운동, 대근육운동)에서 유의미한 차이가 나타나는지 알아본 결과, 선행연구(Greene et al., 2012; Msall, 2009; Sansavini, Savini, Guarini, Broccoli, Alessandroni, & Faldella, 2010)와 일치하게 재태기간이 길어짐에 따라 척도점수 및 발달지수가 상승하는 양상을 보였을 뿐만 아니라 극중증 미숙아 집단은 만삭아 집단에 비해 인지발달, 언어발달, 운동발달 모두에서 유의미하게 낮은 발달지수를 보였다. 이는 24개월의 호주 미숙아 집단을 연구한 Anderson과 동료들(2010) 연구와 대만의 6~24개월 미숙아를 연구한 Yu와 동료들(2013) 연구, 2.5세의 스웨덴 미숙아 집단을 연구한 Månsson과 Stjernqvist의 연구(2014)와 일치하는 결과이다. 한편, 미숙아 집단은 Bayley-III 하위 척도 및 발달 영역 간의 유의미한 수행 차이를 보였다. 이러한 발달 영역 간의 불균형은 점수가 전반적으로 평균 하단에 분포되어 있는 극중증 미숙아 집단 보다는 중증 및 경중 미숙아 집단에서 두드러지게 나타났다. 즉, 이들 집단들은 인지 및 운동 발달에 비해 언어 발달의 상대적인 지연을 나타내었다. 미국의 24개월의 미숙아 집단을 연구한 Greene와 동료들(2012) 연구에서도 언어영역, 특히 표현언어에서의 상대적인 지연을 보고하였으며, 우리와 같이 동양권 문화에 있는 Yu와 동료들(2013) 연구에서도 언어영역에서의 상대적인 지연을 보고하였다. 이는 BSID-II의 MDI에서 언어영역

을 세분화시킨 Bayley-III의 임상적 유용성을 드러내는 결과라고 할 수 있다. 이와 같이 Bayley-III는 만삭아와 변별되는 미숙아 집단의 발달적 특성을 반영하고 있을 뿐 아니라 미숙아 집단 내에서 발생할 수 있는 발달 영역 간의 미묘한 차이를 파악할 수 있는 장점을 보여주고 있다.

하지만, 미국 기준을 적용했을 때 미숙아 집단의 척도점수(8~11점)와 발달지수(90~109점) 모두 대체로 '평균' 범위 안에 있었다. 이는 미국의 Bayley-III의 표준화 연구 결과와 유사한데, 1~42개월의 미숙아를 대상으로 한 실시하였을 미국 표준화연구에서도 평균 범위의 척도점수(9점대)와 발달지수(90점대)를 보고하였다. 교정 연령 1세 미만의 조산 위험아 및 만삭 위험아를 대상으로 한 국내의 박주영(2010)의 연구에서도 평균 범위의 발달지수(90점대)를 보고한 것과 일치하는 결과이다. 이외에도 호주의 Anderson과 동료들(2010)의 연구, 캐나다의 Acton과 동료들(2011)의 연구, 미국의 Greene와 동료들(2012)의 연구, 대만의 Yu와 동료들(2013)의 연구, 스웨덴의 Månsson과 Stjernqvist(2014)의 연구에서도 유사한 결과가 보고되었는데, 이는 미국 기준을 적용하여 Bayley-III의 발달지수 해석하면 아동 능력이 과대평가된다는 논쟁을 불러일으켜 왔다.

두 번째, 미국 기준을 적용한 Bayley-III의 과대평가가 우리나라 미숙아 연구에서도 나타나는지 확인하기 위하여 K-BSID-II 관계를 살펴보았다. 먼저, K-BSID-II 발달지수와 Bayley-III 하위 척도 및 발달지수와의 상관관계를 알아본 결과, K-BSID-II의 MDI는 Bayley-III 인지척도 및 인지발달지수와 상관이 높은 것으로 나타났으며, K-BSID-II의 PDI는 Bayley-III의 대근육운동척도 및 운동발달지수와 상관이 높은

것으로 나타났다. 다만 Bayley-III 언어발달지수 및 하위척도(수용언어척도, 표현언어척도)를 K-BSID-II의 MDI와의 관계에서 해석할 때는 주의가 요구되는 것으로 보인다. 표 3에서 제시된 것처럼 BSID-II에서 Bayley-III로 개정되는 과정에서 인지발달영역과 운동발달영역의 문항은 BSID-II의 문항을 그대로 유지하거나 약간의 수정만을 하고(문항의 61~100%), 문항의 추가는 매우 적은 비율을 차지하였으나(0~25%), Bayley-III에서 새롭게 형성된 구인인 언어발달영역의 문항은 대부분의 월령대에서 40% 이상의 문항이 새로 추가되었기 때문에(12~17개월 문항세트만 65% 정도가 그대로 유지되거나 약간만 수정됨), Bayley-III 인지발달지수만큼 언어발달지수는 K-BSID-II의 MDI와 높은 상관을 보이지는 않았다. 본 연구에서 Bayley-III 언어발달영역이 PDI보다 MDI와 더 높은 상관을 보였을 뿐 아니라 미국 표준화 연구에서 MDI와 더 높은 구인적 특성(Bayley, 2006)을 제안하였기 때문에, Bayley-III 언어발달영역을 K-BSID-II의 MDI 관계에서 탐색하였다.

다음으로, Bayley-III와 K-BSID-II를 모두 실시한 미숙아 79명을 대상으로 두 버전 간의 평균 점수 차이를 살펴보았다. Bayley-III의 인지척도와 언어척도는 K-BSID-II의 인지발달지수(MDI)와 비교하였고, Bayley-III의 운동척도는 K-BSID-II의 동작발달지수(PDI)와 비교하였다. 그 결과, 미숙아 집단의 K-BSID-II의 MDI와 PDI의 평균은 각각 83.97점($SD=20.79$)과 79.22점($SD=19.48$)이었으나 Bayley-III의 인지발달, 언어발달, 운동발달지수의 평균은 각각 95.70점, 90.38점, 96.27점으로 나타났다. 즉, Bayley-III의 인지 및 언어발달지수의 평균은 K-BSID-II의 MDI 보다 각각 11.73점, 6.41점이 더 높

았으며, Bayley-III의 운동발달지수 평균은 K-BSID-II의 PDI 보다 17.05점 더 높게 나타났다. 이는 미국 Bayley-III 표준화 연구(Bayley, 2006)와 유사한 연구 결과로서, Bayley-III가 BSID-II보다 대략 7점정도 더 높은 것으로 보고되었다. 다만, 미국 규준을 적용한 본 연구에서는 Bayley-III 언어발달지수만 BSID-II의 MDI에 비해 7점정도 높은 것으로 나타났으며, Bayley-III의 인지발달지수와 운동발달지수는 BSID-II에 비해 10점 이상 높은 것으로 나타났다. 이는 미국 Bayley-III 표준화 연구 결과보다는 더 높은 수치이나, 미숙아를 대상으로 한 다른 연구들과는 일치하는 결과이다. 22~24개월의 초극소저체중 미숙아를 대상으로 한 Silveira, Filipouski, Gostein, O'Shea와 Procianoy (2012)의 연구에서도 Bayley-III의 인지발달지수와 언어발달지수는 BSID-II의 MDI에 비해 각각 9점과 8점 정도 더 높으며, Bayley-III 운동발달지수는 BSID-II의 PDI에 비해 14점 더 높은 것으로 보고되었다. 또한 초극소저체중의 미숙아를 대상으로 한 Vohr와 동료들(2012)의 연구에서도 Bayley-III 인지발달지수가 BSID-II의 MDI에 비해 11점 높게 나타났으며 Bayley-III 운동발달지수는 K-BSID-II의 PDI에 비해 18점정도 높은 것으로 보고하였다. 이렇게 미국 Bayley-III 표준화에서 발표된 점수 차이와 이후 연구에서 발표된 점수 차이가 다른 것은 연구 대상의 차이에서 기인하는 것으로 생각된다. 미국 Bayley-III 표준화 연구의 대상은 1~42개월의 '정상'아동을 대상으로 한 반면, 이후의 연구들은 24개월을 전후한 '미숙아' 집단을 대상으로 임상적 유용성을 확인하기 위해 실시된 연구들이다. 이와 관련하여 Lowe와 동료들(2011), Moore와 동료들(2012)에 따르면, Bayley-III와 BSID-II 발달지수 간의 차이는 모

든 점수대에서 유사하게 점수가 상승하는 것이 아니라 정상분포의 '하단에 속할수록 점수 차이가 더욱 심해진다고 보고하였다. 예를 들어 Lowe와 동료들(2011)의 연구에서는 BSID-II의 MDI에서는 60점을 얻었던 아동이 Bayley-III의 인지발달지수에서는 87점을 얻었다고 보고하면서, 정상분포 하단에 있는 아동의 Bayley-III와 BSID-II 점수를 비교할 때는 더욱 큰 주의를 요한다고 경고하였다. 이와 같이 정상분포 하단의 발달지수의 상승은 발달지연률과 관련하여 임상적 유용성 측면에서 심각한 문제를 야기할 수 있다.

한편, Bayley-III가 BSID-II에 비해 점수가 상승하는 것은 선행 연구들(Anderson et al., 2010; Moore et al., 2012; Vohr, Stephens, Higgins, Bann, Newman, Hintz, Das, Peralta-Carcelen, Yolton, Dusick, Evans, Goldstein, Ehrenkranz, Pappas, Adams-Chapman, Wilson-Costello, Bauer, Bodnar, Heyne, Vaucher, Dillard, Acarregui, McGowan, Myers, & Fuller, 2012)과 동일하게 Flynn효과에 상반되는 결과이다. 이렇게 Flynn효과와 상충되는 결과는 Bayley-III의 구조적 측면과 미국 표준화 기준 집단의 문제로 해석될 수 있다. 먼저, Bayley-III와 K-BSID-II의 구조적 차이에 기인하는 것으로 추정될 수 있다. K-BSID-II에서는 각 연령마다 시작과 중지 문항이 정해진 문항세트를 제시하고 있어 같은 연령일 경우 획득할 수 있는 점수의 최대치가 동일하게 되지만, Bayley-III에서는 문항세트 대신 시작규칙과 중지규칙을 도입하여 같은 연령일지라도 최대치로 받을 수 있는 점수는 동일하지 않게 된다. 따라서 똑같은 능력을 가진 아동일지라도 Bayley-III에서는 중지규칙을 충족할 때까지 더 많은 문항을 수행할 수 있게 되므로 더 높은 범위의 점수를 획득하는 것이

가능해지게 되는 것이다. 두 번째는 Bayley-III 미국 표준화 기준 집단의 문제로 인해 이러한 차이가 발생한다고 추정할 수 있다. 많은 선행 연구들(Anderson et al., 2010; Acton et al., 2011; Johnson et al., 2014; Moore et al., 2012; Reuner et al., 2013; Yu et al., 2013)이 Bayley-III의 기준 제작과 관련하여 미국 표준화 과정에서 임상 집단을 10% 포함한 것에 대해서 비판을 하고 있다. 이러한 비율은 정상분포에서의 비율(2.3%)을 훨씬 초과하는 것으로, 발달 문제가 있는 아동들을 과잉포집하게 되면 검사를 더 '쉽게' 하는 효과를 가져오며 발달지연률을 과소추정하게 할 수 있다는 것이다(Acton et al., 2011; Johnson et al., 2014).

이에, Bayley-III의 발달지연률을 탐색하기 위해 1 표준편차와 2 표준편차를 기준으로 K-BSID-II의 발달지연률과 비교하였을 때, Bayley-III의 발달지연률은 K-BSID-II의 발달지연률에 비해 크게 감소하였다. 1 표준편차(85점 미만)를 기준으로 하였을 때, K-BSID-II의 MDI와 PDI에 의한 발달지연률은 각각 51.9%와 58.2%였으나 Bayley-III의 인지, 언어, 발달 지수에 따른 발달지연률은 각각 12.7%, 22.8%, 16.5%로 크게 감소하였다. 또한 2 표준편차(69점 이하)를 기준으로 하였을 때, K-BSID-II의 MDI와 PDI에 따른 발달지연률은 각각 30.4%와 35.4%였으나 Bayley-III의 인지, 언어, 발달 지수에 따른 발달지연률은 각각 2.5%, 3.8%, 3.8%로 대폭 감소하였다. 이는 미숙아를 대상으로 Bayley-III 미국 표준을 적용한 호주(Anderson et al., 2010), 캐나다(Acton et al., 2011), 영국(Moore et al., 2012; Johnson et al., 2014), 독일(Reuner et al., 2013), 타이완 연구(Yu et al., 2013)에서 제기되어 온 핵심적인 논란으로서 우리나라에서도 미숙아 집단에서의 발달

지연률의 과소추정이 동일하게 발생하였다. 그러나, 발달지연률에 있어서 BSID-II와 Bayley-III 간의 차이와 관련하여 또 다른 해석이 이루어지기도 한다. 일부 학자들은 Bayley-III가 발달지연률을 과소추정 하는 것이 아니라 BSID-II가 발달지연률을 과대추정 해왔다고 주장하고 있다. 다른 말로 하면 BSID-II가 아동의 능력을 과소평가 해왔다는 것이다. Hack, Taylor, Drotar, Schluchter, Cartar, Wilson-Costello, Klein, Friedman, mercuri-Minich와 Morrow(2005)에 의해 이루어진 종단연구에 따르면, 교정연령 20개월의 미숙아를 대상으로 MDI를 측정하고 8세에 K-ABC(Kaufman Assessment Battery for Children)로 지적능력을 평가했을 때 평균적으로 12.2점이 상승하였다고 보고하면서 BSID-II가 아동의 능력을 과소평가했다고 보고하였다. 또한 BSID 원판 및 Bayley-III와 비교할 때 유독 BSID-II의 MDI와 PDI만 더 낮아지는 특성을 보였으며(Synnes, Rogers, Thomas, Wright, & Butt, 2010), Glenn과 동료들(2001)은 BSID 원판과 BSID-II를 반복 측정한 결과 정상집단과 다운증후군 집단 모두 BSID-II에서 8~9점 정도 낮은 점수를 얻었다고 보고하였다. 이러한 논의들은 Bayley-III가 발달지연률을 과소추정하는 것일수도 있지만 BSID-II가 발달지연률을 과잉추정하는 것일수도 있다는 가능성을 모두 열어두고 두 버전 간의 논쟁을 더욱 뜨겁게 하고 있다. 그리고 Bayley-III의 발달능력의 과대평가 및 발달지연률의 과소추정 논란은 발달지연 진단과 관련하여 절단점 논의로 이어지고 있다.

세 번째, Bayley-III의 미숙아 집단에 대한 임상적 변별력을 확인하고 절단점을 탐색해보고자 ROC분석을 실시하였다. AUC곡선을 통해 각 발달영역별 변별력을 확인한 결과, 인지발

달, 언어발달, 운동발달의 변별 정확도는 우수한 수준으로 나타났다. 이는 Yu와 동료들(2013)의 연구에서 BSID-II에 대한 Bayley-III의 AUC곡선을 산출한 결과 대부분의 연령에서 80%이상 나타냈다는 것과 일치하는 결과이며 Bayley-III가 미숙아에게서 발달지연을 변별하는 목적으로 매우 유용하게 사용될 수 있다는 것을 의미하는 결과이다.

더 나아가 발달지연을 가장 잘 변별할 수 있는 절단점은 민감도와 특이도가 동시에 높은 지점으로 탐색해 보았다. 그 결과, 인지발달지수는 90점일 때, 언어발달지수는 86~89점 사이일 때, 운동발달지수는 97점일 때 민감도와 특이도가 모두 높은 절단점이 되었다. 본 연구에서 제안하는 인지발달지수의 절단점인 90점은 Lowe와 동료들(2011)이 발달지연 진단을 위한 인지발달지수의 절단점으로 93점을 제안한 것과 Yu와 동료들(2013)이 12~24개월 발달지연을 위한 절단점으로 90점을 제안한 것과 유사한 점수라고 할 수 있다. 한편 본 연구의 운동발달지수의 절단점인 97점은 다른 연구의 절단점들보다 상당히 높은 점수이다. 이와 관련하여 운동영역의 과대평가적 속성과 문화적 차이로 해석해 볼 수 있다. 먼저, 만 2세의 미숙아를 대상으로 한 Anderson과 동료들(2010)의 연구에서도 운동발달지수는 규준평균에 해당하는 100점을 얻었으며, 비교정상집단은 무려 118점이나 얻을 만큼 높은 점수의 상승을 보였다. 또한 22~24개월 미숙아를 대상으로 연구한 Silveira와 동료들(2012)의 연구에서는 운동발달지수가 BSID-II에 비해 각각 무려 14점이나 상승한 점을 볼 수 있다. 또 다른 가능성은 문화적인 차이로 해석할 수 있다. 우리와 같이 아시아 문화에 속해 있는 Yu와 동료들(2013)의 연구에서도 24개월 아동의 운동발

달지연의 절단점으로 94점을 제안하고 있고, 한국의 고위험 영아들을 대상으로 Bayley-III를 연구한 박주영(2010)의 연구에서도 조산 위험아와 건강 만삭아의 운동발달지수 평균이 98점과 107점이라고 제시하고 있어 아시아 문화권에서 운동발달 점수의 상승을 고려해볼 수 있다. 현재 만 2세 미만의 미숙아 대상의 Bayley-III 연구의 대부분이 인지영역이나 언어영역에 집중된 논의가 이루어지고 있고 (Johnson et al, 2013; Lowe et al, 2011; Moore et al, 2012), 운동영역에 대한 연구가 부족한 상황이라서 이에 대한 추후 탐색이 필요한 것으로 보인다.

이상과 같이 Bayley-III는 미숙아를 대상으로 한 연구에서 다양한 발달영역에 대한 정보를 제공하여 발달 영역 간의 미묘한 불균형을 파악하도록 도움을 줄 것으로 보인다. 특히 언어 발달의 독자적인 측정은 미숙아 집단에서 흔하게 나타나는 언어문제를 민감하게 파악할 수 있을 뿐만 아니라 표현언어와 수용언어를 세분화하여 측정하고 있어서 각 발달영역에서 개입을 계획하거나 평가할 때 유용한 자료로 활용될 수 있을 것으로 보인다. 또한 발달 진단과 관련하여 우수한 임상적 변별력을 보이고 있다. 그러나 선행 연구들에서처럼 한국 미숙아 연구에서도 미국 기준을 사용할 때 발달 능력의 과대평가 및 발달지연률의 과소평가라는 한계점을 여전히 보이고 있었다. 이에 편파된 미국 기준을 보완할 수 있는 대안으로 자국의 비교집단을 '기준'으로 이용하는 연구결과들이 호주(Anderson et al., 2010)와 캐나다(Acton et al., 2011), 스웨덴(Månsson & Strjernqvist, 2014), 이탈리아(Tallandini, Franco, & Morsan, 2013), 네덜란드(Steenis, Verhoeven, Hessen, & Baar, 2014)에서 보고되었다. 아울러

동일하게 영국에 거주하더라도 민족에 따라서 운동 발달의 경로가 다르다는 것 또한 보고되었으며(Kelly, Sacker, Schoon, & Nazroo, 2006), 언어발달에 있어서도 민족에 따라서 서로 다른 발달경로가 보고되었다(Duncan, Watterberg, Nolen, Vohr, Adams-Chapman, Das, & Lowe, 2012). 이러한 결과들은 자국의 모집단을 대표하는 표본에 기초한 기준의 중요성을 강조하고 있다. 최근 Spencer-Smith, Spittle, Lee, Doyle 와 Anderson(2015)의 연구에서 미숙아를 대상으로 Bayley-III에서 제안한 미국 표준화 연구에서 제안한 기준을 사용하여 인지척도 및 언어척도의 민감도와 특이도를 측정한 결과, 민감도는 각각 29.4%~38.5%, 40.0%~46.7%로 낮았으나 자국의 지역 기준을 사용했을 때 민감도가 다소 상승하는 결과를 얻었다. 이에 우리나라 영유아의 인지, 언어, 운동 발달을 반영할 수 있는 우리나라의 기준 마련이 대안일 수 있으며, 이때 미국 기준제작 과정과 동일하게 임상집단을 과도하게 포함(10%)하기보다는, 이후의 연구에서 제안되는 것처럼 한국 영유아의 모집단의 특성을 잘 반영하는 표집을 통한 기준제작이 요구된다.

본 연구의 의의와 제한점은 다음과 같다. 첫째, Bayley-III가 개발된 이래로 세계 여러 나라에서 임상적 유용성에 대한 논의가 이루어져오고 있으며 아직 일치된 연구 결과들이 보고되고 있지는 않다. 그러나 미국 기준을 사용한 Bayley-III의 발달지연률 과소평가 문제는 초기 발달지연 확인 및 개입이 어려워지도록 만들 위험이 있다는 점에서 반드시 고려해야 하는 문제이다. 이에 아직 한국 기준이 제공되지 않는 시점에서 Bayley-III를 사용하는 국내의 연구자 및 임상가들에게 점수 해석의 주의

를 알리는 실증적인 자료가 될 것이다.

둘째, 아직 한국 규준이 마련되지 않는 시점에서 미숙아를 포함한 고위험군 영유아에게 Bayley-III가 유용하게 사용될 수 있기 위한 각 척도점수의 적절한 절단점을 제안했다는 데 의의가 있다. 본 연구를 통해 Bayley-III의 임상집단에 대한 예측력과 변별 정확도가 우수한 수준인 것으로 나타나 임상 현장에서 발달 진단 검사 도구로써 유용하게 사용될 수 있음을 확인하였으며, K-BSID-II에서의 발달지연에 대한 Bayley-III의 미숙아 집단에서의 적절한 절단점을 탐색하였다는 점에서 임상적 활용에 대한 의의가 있다.

셋째, Bayley-III 미국 표준화 연구의 규준을 이용한 임상적 유용성의 한계는 한국의 규준 제작 과정에서 반면교사의 자료가 될 수 있을 것이다. 한국 규준 마련 시에는 미국 표준화 연구에서와 같이 임상집단을 과도하게 포함시키기 보다는 한국 영유아의 모집단을 잘 반영할 수 있는 표집이 요구되는 것으로 보인다. 몇몇 외국의 선행 연구들에서도 규준과 다른 절단점을 제시하고 있지만, 우리나라 영유아와 문화나 발달, 발육에서의 차이 등으로 수행에서 차이를 보일 수 있어 그것을 국내에 직접적으로 적용하는 데에는 한계가 있다. 이에 K-Bayley-III 예비연구에서도 제안되었듯이(김지혜 외, 2014), 한국 진단기준점 산출의 필요성이 제기되고 있다.

본 연구의 제한점은 다음과 같다. 첫째, 첫째, 재태기간에 따라 Bayley-III 발달지수와 척도점수의 차이는 유의미했지만, 설명량인 r^2 는 작은 것으로 나타나서 이에 대한 추후 검토가 요구된다. 둘째, 본 연구에서는 발달지연의 고위험군 중에서도 미숙아로 대상을 제한시켰으나, 임상적 유용성을 검증하기 위해서는 자폐

스펙트럼 장애, 언어지연, 뇌성마비 등 좀 더 다양한 임상집단에 대한 연구가 이루어질 필요가 있다. 국내에서 뇌성마비 영유아의 발달 평가에 관한 연구들은 이루어지고 있으나 다른 임상집단에 대한 연구는 부족한 실정이다. 후속 연구에서는 다양한 임상집단에 대한 임상적 유용성을 살펴볼 것을 제안한다.

참고문헌

- 김문선, 김진경 (2010). 발달지연아에서 K-ASQ (Korean-Ages & Stages Questionnaires) 발달 선별 검사와 베일리 영유아 발달 검사의 비교. *대한소아신경학회지*, 18(1), 49-57.
- 김지혜, 방희정, 이순행, 김현미, 남민 (2014). 한국형 베일리 영유아 발달검사 3판 사회-정서척도의 표준화 예비 연구. *한국인간발달학회지*, 21(2), 201-225.
- 신희선, 권범선, 임성오 (2005). 발달문제 위험 아동 선별을 위한 한국형 Denver II 검사의 타당성에 관한 연구. *아동간호학회지*, 11(3), 316-321.
- 신희선, 김정미, 임성오 (2007). 0-5세를 위한 한국형 영·유아 발달 선별검사(K-CDR)의 타당성에 대한 연구. *아동간호학회지*, 13(2), 222-229.
- 박인환, 석혜은, 김아영, 신석호, 방희정 (2015). 한국형 베일리 영유아 발달검사 제 3판 사회-정서 척도의 타당도 연구. *한국심리학회지: 발달*, 28(3), 29-50.
- 박주영 (2010). Bayley 영유아 발달검사(III)를 이용한 고위험군 영아의 발달검사. *인제대학교 대학원 석사학위 논문*.
- 오수경, 방희정, 이순행 (2014). 한국형 베일리

- 영유아 발달검사 3판 인지척도 예비 연구. 한국심리학회지: 발달, 27(1), 117-140.
- 이순행, 안소현, 이은지, 방희정 (2014). 한국형 베일리 영유아 발달검사 3판 소근육 운동 척도 예비 연구. 한국아동학회, 35(5), 37-61.
- 이종승 (2005). 표준화 심리검사. 서울: 교육과학사.
- 이지연, 김영태, 방희정, 이순행 (2014). 한국형 베일리 영유아 발달검사 제3판 표현언어 척도 예비연구. 한국언어정각임상학회, 19(3), 320-330.
- 이진, 김영아, 오경자 (2009). 한국판 CBCL 1.5-5의 변별타당도 및 임상적 유용성. 한국임상심리학회지: 임상, 28(1), 171-186.
- 조복희, 박혜원 (2004). 한국 Bayley 영유아 발달검사(K-BSID-2) 표준화 연구(1). 한국심리학회지: 발달, 17(1), 191-206.
- Acton, B. V., Biggs, W. S. G., Creighton, D. E., Penner, K. A. H., Switzer, H. N., Thomas, J. H. P., Joffe, A. R., & Robertson, C. M. T. (2011). Overestimating neurodevelopment using the Bayley-III after early complex cardiac surgery. *Pediatrics*, 128(4), 794-800.
- Anderson, P. J., Doyle, L. W., De Luca, C. R., Hutchinson, E., Roberts, G., Callanan, C., Davis, N., Duff, J., Kelly, E., McDonald, M., Stewart, M., Ung, L., Hunt, R., Opie, G., Watkins, A., Williamson, A., Woods, H., Carse, E., Charlton, M. P., & Hayes, M. (2010). Underestimation of developmental delay by the new bayley-III scale. *Archives of Pediatrics and Adolescent Medicine*, 164(4), 352-356.
- Aylward, G. P. (2009). Developmental screening and assessment: What are we thinking? *Journal of Developmental & Behavioral Pediatrics*, 30(2), 169-173.
- Bayley, N. (1993). *Bayley scales of infant development: manual*. Psychological Corporation.
- Bayley, N. (2006b). *Bayley Scales of Infant Development and Toddler development: Technical manual*. San Antonio, TX: The Psychological Corporation.
- Dezoete, J. A., MacArthur, B. A., & Tuck, B. (2003). Prediction of Bayley and Stanford-Binet scores with a group of very low birthweight children. *Child: Care, Health and Development*, 29(5), 367-372.
- Duncan, A. F., Watterberg, K. L., Nolen, T. L., Vohr, B. R., Adams-Chapman, I., Das, A., & Lowe, J. (2012). Effect of ethnicity and race on cognitive and language testing at age 18-22 months in extremely preterm infants. *The Journal of Pediatrics*, 160(6), 966-971.
- Gagnon, S. G., & Nagle, R. J. (2000). Comparison of the revised and original versions of the Bayley Scales of Infant Development. *School Psychology International*, 21(3), 293-305.
- Gesell, A., Halverson, H. M., Thomson, H., Ilg, F. L., Castner, B. M., Ames, L. B., & Amatruda, C. S. (1940). *The first five years of life, the preschool years*. Oxford, England.
- Glenn, S. M., Cunningham, C. C., & Dayus, B. (2001). Comparison of the 1969 and 1993 standardizations of the Bayley Mental Scales of Infant Development for infants with down syndrome. *Journal of Intellectual Disability Research*, 45(1), 56-62.
- Greene, M. M., Patra, K., Nelson, M. N., &

- Silvestri, J. M. (2012). Evaluating preterm infants with the Bayley-III: patterns and correlates of development. *Research in Developmental Disability, 33*(6), 1948-1956.
- Hack, M., Friedman, H., & Fanaroff, A. A. (1996). Outcomes of extremely low birth weight infants. *Pediatrics, 98*(5), 931-937.
- Hack, M., Taylor, H. G., Drotar, D., Schluchter, M., Cartar, L., Wilson-Costello, D., Klein, N., Friedman, H., Mercuri-Minich, N., & Morrow, M. (2005). Poor predictive validity of the Bayley Scales of Infant Development for cognitive function of extremely low birth weight children at school age. *Pediatrics, 116*(2), 333-341.
- Herrnstein, R. J., & Murray, C. (1994). *The bell curve: The reshaping of American life by differences in intelligence*. New York: Free.
- Johnson, S., & Marlow, N. (2006). Developmental screen or developmental testing?. *Early Human Development, 82*(3), 173-183.
- Johnson, S., Moore, T., & Marlow, N. (2014). Using the Bayley-III to assess neurodevelopmental delay: which cut-off should be used? *Pediatric Research, 75*(5), 670-674.
- Kelly, Y., Sacker, A., Schoon, I., & Nazroo, J. (2006). Ethnic differences in achievement of developmental milestones by 9 months of age: the Millennium Cohort Study. *Developmental Medicine & Child Neurology, 48*(10), 825-830.
- Lowe, J. R., Erickson, S. J., Schrader, R., Duncan, A. F. (2011). Comparison of the Bayley II Mental Developmental Index and the Bayley III Cognitive Scale: are we measuring the same thing? *Acta Paediatrica, 101*(2), e55-e58.
- Luttikhuisen dos Santos, E. S. L., de Kieviet, J. F., Königs, M., van Elburg, R. M., & Oosterlaan, J. (2013). Predictive value of the Bayley scales of infant development on development of very preterm/very low birth weight children: a meta-analysis. *Early human development, 89*(7), 487-496.
- McFadden, T. U. (1996). Creating language impairments in typically achieving children: the pitfalls of "normal" normative sampling. *Language, Speech & Hearing Services in Schools, 27*, 3-9.
- Månsson, J., & Stjernqvist, K. (2014). Children born extremely preterm show significant lower cognitive, language and motor function levels compared with children born at term, as measured by the Bayley III at 2.5 years. *Acta Paediatrica, 103*(5), 504-511.
- Miska, L., Jan, H. (2005). Evaluation of current statistical approaches for predictive geomorphological mapping. *Geomorphology, 67* (3-4), 299-315.
- Moore, T., Johnson, S., Marlow, N., Haider, S., & Hennessy, E. (2012). Relationship between test scores using the second and third editions of the bayley scales in extremely preterm children. *Journal of Pediatrics, 160*(4), 553-558.
- Msall, M. E. (2009). Optimizing neuromotor outcomes among very preterm, very low-birth-weight infants. *JAMA, 302*(20), 2257-2258.
- Reuner, G., Fields, A. C., Wittke, A., Pietz, J., & Löfflich, M. (2013). Comparison of the developmental tests Bayley-III and Bayley-II in 7-month-old infants born preterm. *European*

- Journal of Pediatrics*, 172(3), 393-400.
- Ross, G., & Lawson, K. (1997). Using the Bayley-II: Unresolved Issues in Assessing the Development of Prematurely Born Children. *Journal of Developmental & Behavioral Pediatrics*, 18(2), 109-111.
- Ross, G., Lipper, E. G., & Auld, P. A. (1985). Consistency and change in the development of premature infants weighing less than 1,501 grams at birth. *Pediatrics*, 76(6), 885-891.
- Sansavini, A., Guarini, A., Justice, L. M., Savini, S., Broccoli, S., Alessandrini, R., & Faldella, G. (2010). Does preterm birth increase a child's risk for language impairment?. *Early human development*, 86(12), 765-772.
- Siegel, L. S. (1983). Correction for Prematurity and Its Consequences for the Assessment of the Very Low Birth Weight Infant. *Child Development*, 54(5), 1176-1188.
- Siegel, L. S., Cooper, D. C., Fitzhardinge, P. M., & Ash, A. J. (1995). The use of the mental development index of the Bayley Scale to diagnose language delay in 2-year-old high-risk infants. *Infant Behavior and Development*, 18(4), 483-486.
- Silveira, R. C., Filipouski, G. R., Goldstein, D. J., O'Shea, T. M., & Procianny, R. S. (2012). Agreement between Bayley scales second and third edition assessments of very low-birth-weight infants. *Archives of Pediatrics & Adolescent Medicine*, 166(11), 1075-1076.
- Skranes, J., Vik, T., Nilsen, G., Smevik, O., Andersson, H. W., & Brubakk, A. M. (1998). Can cerebral MRI at age 1 year predict motor and intellectual outcomes in very low birth weight children?. *Developmental Medicine & Child Neurology*, 40(4), 256-262.
- Spencer-Smith, M. M., Spittle, A. J., Lee, K. J., Doyle, L. W., & Anderson, P. J. (2015). Bayley-III cognitive and language scales in preterm children. *Pediatrics*, 135(5), e1258-e1265.
- Steenis, L. J., Verhoeven, M., Hessen, D. J., & van Baar, A. L. (2014). First steps in developing the Dutch version of the Bayley III: Is the original Bayley III and its item sequence also adequate for Dutch children?. *European Journal of Developmental Psychology*, 11(4), 494-511.
- Synnes, A. R., Rogers, M., Petrie Thomas, J., Wright, S., & Butt, A. (2010). The Bayley Scales of Infant and Toddler Development in extremely low birth weight survivors at 18 months corrected age. *Pediatr Child Health*, 15(suppl A), 41A.
- Tallandini, M. A., Franco, A., & Morsan, V. (2013). Bayley scales of infant and toddler development (bsitd-iii): un contributo alla taratura italiana su un campione di bambini prematuri. *Ricerche di Psicologia*.
- Vohr, B. R., Stephens, B. E., Higgins, R. D., Bann, C. M., Newman, J. E., Hintz, S. R., Das, A., Peralta-Carcelen, M., Yolton, K., Dusick, A. M., Evans, P. W., Goldstein, R. F., Ehrenkranz, R. A., Pappas, A., Adams-Chapman, I., Wilson-Costello, D. E., Bauer, C. R., Bodnar, A., Heyne, R. J., Vaucher, Y. E., Dillard, R. G., Acarregui, M. J., McGowan, E. C., Myers, G. J., & Fuller, J. (2012). Are outcomes of extremely preterm

- infants improving? Impact of Bayley assessment on outcomes. *Journal of Pediatrics*, 161(2), 222-228.
- Washington, K., Scott, D. T., Johnson, K. A., Wendel, S., & Hay, A. E. (1998). The Bayley Scales of Infant Development-II and children with developmental delays: a clinical perspective. *Journal of Developmental & Behavioral Pediatrics*, 19(5), 346-349.
- Williams, M. L., Lewandowski, L. J., Coplan, J., & D'Eugenio, D. B. (1987). Neurodevelopmental outcome of preschool children born preterm with and without intracranial hemorrhage. *Developmental Medicine & Child Neurology*, 29(2), 243-249.
- Xu, Y., & Filler, J. W. (2005). Linking assessment and intervention for developmental/functional outcomes of premature, low-birth-weight children. *Early Childhood Education Journal*, 32(6), 383-389.
- Yu, Y. T., Chen, L. C., Wu, Y. C., Jeng, S. F., Hsieh, W. S., Lee, W. T., Hsu, C. H., & Chiu, N. C. (2013). A psychometric study of the Bayley scales of infant and toddler development - 3rd edition for term and preterm Taiwanese infants. *Research in Developmental Disabilities*, 34(11), 3875-3883.
- Wang, S. T., Wang, C. J., Huang, C. C., & Lin, C. H. (1998). Neurodevelopment of surviving infants at age two years, with a birthweight less than 2000g and cared for in neonatal intensive care units(NICU)-results from a population based longitudinal study in Taiwan. *Public Health*, 112, 331-336.

1차원고접수 : 2016. 04. 22.

수정원고접수 : 2017. 03. 08.

최종게재결정 : 2017. 03. 16.

The Clinical Utility of Korean Bayley Scales of Infant and Toddler Development-III - Focusing on using of the US norm -

Yoo Jin Lim

Hee Jeong Bang

Soon Hang Lee

Ewha Womans University

The study aims to investigate the clinical utility of Bayley-III using US norm in Korea. A total of 98 preterm infants and 93 term infants were assessed with the K-Bayley-III. The performance pattern of preterm infants was analyzed with mixed design ANOVA which examined the differences of scaled scores and composite scores of Bayley-III between full term- and preterm- infant group and within preterm infants group. Then, We have investigated agreement between classifications of delay made using the BSID-II and Bayley-III. In addition, ROC plots were constructed to identify a Bayley-III cut-off score with optimum diagnostic utility in this sample. The results were as follows. (1) Preterm infants have significantly lower function levels in areas of 5 scaled scores and 3 developmental indexes compared with infants born at term. Significant differences among scores within preterm infant group were also found. (2) Bayley-III had the higher scores of the Mental Development Index and Psychomotor Developmental Index comparing to the scores of K-BSID- II, and had the lower rates of developmental delay. (3) All scales of Bayley-III, Cognitive, Language and Motor scale had the appropriate level of discrimination, but the cut-off composite scores of Bayley-III were adjusted 13~28 points higher than 69 for prediction of delay, as defined by the K-BSID- II. It explains the lower rates of developmental delay using the standard of two standard deviation. This study has provided empirical data to inform that we must careful when interpreting the score for clinical applications, identified the discriminating power, and proposed more appropriate cut-off scores. In addition, discussion about the sampling for making the Korean norm of Bayley-III was provided. It is preferable that infants in Korea should use our own validated norms. The standardization process to get Korean normative data must be performed carefully.

Keywords : Bayley scales of Infant and Toddler Development-Third Edition, clinical utility, preterm infants