

Diferencias en el diagnóstico del desarrollo, comparación de las escalas de Bayley II y III

Differences in the diagnosis of development, comparison of the Bayley II and III scales.

Ismene Corral-Guillé, Rolando Rivera-González

Resumen

ANTECEDENTES: Las escalas Bayley son una herramienta ampliamente reportada para la evaluación del desarrollo en los primeros años de vida. Estudios que comparan la segunda edición (BSID-II) y la tercera (B-III), informan puntuaciones más altas de lo esperado usando B-III, otros asumen que el BSID-II presenta mayores exigencias a las de varias poblaciones. No existen reportes en población hispanohablante.

OBJETIVO: Comparar puntuaciones y clasificación de ambas versiones en niños con antecedentes de riesgo perinatal.

MATERIALES Y MÉTODOS: Estudio transversal descriptivo en 671 niños de 0 a 42 meses, con antecedentes de riesgo perinatal. Se utilizaron las escalas BSID-II y B-III. Cada participante fue evaluado en una sola sesión. Para la comparación entre ambas pruebas, se calculó la media de las puntuaciones del B-III de las escalas cognitiva y de lenguaje, denominado Puntaje Compuesto Lenguaje-Cognitivo (PC-LC)

RESULTADOS: La media del PC-LC del B-III fue 14.04 puntos mayor a la del Índice de Desarrollo (ID) Mental BSID-II, y el Puntaje Compuesto Motor (PC-Motora del B-III) 17.87 puntos más alta que el ID Motor BSID-II. Las regresiones lineales para estimar los ID Mental BSID-II a partir del B-III mostraron valores de $r^2 = 0.55$ y por rango de edad con r^2 de 0.52 a 0.84 y para el ID Motor BSID-II con r^2 de 0.58 en general, y para los 7 rangos con r^2 entre 0.51 a 0.75

CONCLUSIONES. Con el BSID-II se observa mayor presencia de alteraciones en ambas subescalas. El BSID-II subestima el desarrollo del niño respecto al B-III.

PALABRAS CLAVE: Escalas Bayley de desarrollo infantil; desarrollo del niño; psicométrica; infantes, discapacidades del desarrollo.

Abstract

BACKGROUND: The Bayley scales are a widely reported tool for the evaluation of development in the first years of life. Studies comparing the second edition (BSID-II) and the third (B-III) report higher than expected scores using B-III, others assume that the BSID-II places higher demands on various populations. There are no reports in the Spanish-speaking population.

OBJECTIVE: Compare scores and classification of both versions in children with a history of perinatal risk.

Centro de Investigación del Neurodesarrollo, Instituto Nacional de Pediatría.

ORDICD

<https://orcid.org/0000-0002-7993-4687>
<https://orcid.org/0000-0002-8312-0248>

Recibido: 28 de abril de 2022

Aceptado: 24 de mayo de 2023

Correspondencia

Rolando Rivera González
rolandorivera66@gmail.com

Este artículo debe citarse como: Corral-Guillé I, Rivera-González R. Diferencias en el diagnóstico del desarrollo, comparación de las escalas de Bayley II y III. Acta Pediatr Mex 2023; 44 (3): 187-197.

MATERIALS AND METHODS: Descriptive cross-sectional study in 671 children from 0 to 42 months, with a history of perinatal risk. The BSID-II and B-III scales were used. Each participant was evaluated in a single session. Average scores of the B-III Cognitive Language Composite (CLC) for comparison with Mental Development Index (MDI).

RESULTS: The mean CLC (B-III) was 14.04 points higher than the MDI, and Motor Composite (MC) (B-III) 17.87 points higher than the Psychomotor Developmental Index (PDI). Linear regressions to estimate MDI from B-III showed values of $r^2 = 0.55$ and by age range with r^2 from 0.52 to 0.84 and for PDI with r^2 of 0.58 overall, and for the 7 ranges with r^2 between 0.51 to 0.75

CONCLUSIONS: With the BSID-II, a greater presence of alterations is observed in both subscales. The BSID-II underestimates the child's development compared to the B-III.

KEYWORDS: Bayley scales of infant development; child development; psychometrics; child, developmental disabilities.

ANTECEDENTES

Bayley Scales of Infant and Toddler Development (BSID) es una de las herramientas más utilizadas en el campo clínico y de investigación para medir y diagnosticar el desarrollo.¹ Cuenta con tres versiones publicadas en 1969, 1993 y 2006.^{2,3} Se ha comentado en la literatura que los resultados con las dos últimas dan resultados distintos al valorar a un mismo niño. La segunda versión conocida como *Bayley's Scales of Infant Development* (BSID-II) y la tercera como Bayley-III (B-III).

La estrategia fundamental de ambas versiones consiste en ofrecer a los niños un conjunto de ítems con creciente complejidad para su ejecución; del total que el niño acredita se obtienen puntuaciones directas que se transforman a *puntuaciones estandarizadas*. Una de las primeras reglas que puede llevar a resultados distintos es determinar a partir de cuántos ítems consecutivos resueltos de forma positiva se asume que todos los previos son positivos, lo que se

se denomina línea inicio o piso; y con cuántos indicadores negativos consecutivos se debe asumir que el niño ya no es capaz de realizar más conductas dejándose de explorar y calificar los indicadores subsecuentes, lo que se denomina línea techo o de suspensión.^{3,4} Esto se basa en el supuesto de que el desarrollo procede en una secuencia ordinal; sin embargo, cuando se evalúan niños que presentan alteraciones, pueden observarse secuencias heterogéneas según los componentes específicos o subdominios sobre los que se enfatiza en cada edad o momento del desarrollo en cada área o escala particular.^{3,5,6}

La BSID-II cuenta con dos escalas, la *Mental* y la *Motora* y la B-III con 3 escalas: *Cognitiva*, *Lenguaje* y *Motora*, y cuatro subescalas: *Lenguaje expresivo* (LE), *Lenguaje receptivo* (LR), *Motor fino* (MF) y *Motor grueso* (MG). Las escalas expresan sus *puntuaciones estandarizadas* con medias (\bar{X}) de 100 y desviaciones estándar (SD) de 15, denominadas *Índices de Desarrollo* en la BSID-II (ID-Mental e ID-Motor) y *Puntuaciones Compuestas* en la B-III (PC-Cog, PC-Leng y PC-

Motora), las subescalas del B-III se expresan con *Puntuaciones Escalares con (\bar{X}) de 10 y SD de 3: LR, LE, MF y MG.*

Esta limitación de carecer de puntuaciones específicas en el BSID-II trató de ser resuelta en la B-III, pasando de dos a siete parámetros de medición permitiendo identificar retrasos o cambios en cada uno de ellos. Con la separación de las puntuaciones específicas se provee una mayor tolerancia de avance, ya que el punto de suspensión se considera subescala por subescala, para luego ser transformados a puntuaciones escalares, compuestas y percentiles.^{3,7} La clasificación de los puntajes finales del B-III se modificaron respecto a su versión previa preservándose el criterio de retraso severo en el desarrollo cuando la *puntuación estandarizada* es <70 (2 SD por debajo de la media).^{2,3} Algunos autores han propuesto otros puntos de corte para clasificar e interpretar los resultados a parámetros que se ajusten a prevalencias poblacionales conocidas.⁸

Se informan puntuaciones más altas de lo esperado usando B-III;^{1,6,8,9} lo que lleva a suponer que sobreestima el desarrollo y, como tal, subestima el retraso en comparación con BSID-II. Los editores de B-III revelaron que la muestra de estandarización era siete puntos más alta que las puntuaciones de los ID del BSID-II. La inclusión de niños con deterioro leve (10 %) en la estandarización de B-III, junto con niños hispanoamericanos (16 %), se ha formulado como una probable causa de puntuaciones mayores.³ Algunos autores han señalado que, en los primeros meses de vida, el B-III sobreestima el desarrollo.¹⁰

OBJETIVO

Comparar los resultados de las evaluaciones (puntuaciones y clasificación) BSID-II y B-III en niños con antecedentes de riesgo perinatal y predecir los resultados del BSID-II a partir del B-III.

MATERIALES Y MÉTODOS

Participantes

Se realizó un estudio transversal y prospectivo en el Instituto Nacional de Pediatría de la Ciudad de México. Los participantes fueron lactantes y preescolares de 0 a 42 meses de edad, con antecedentes de riesgo perinatal, nacidos entre abril 2013 y noviembre del 2019. En los casos de niños prematuros se utilizó la corrección de la edad indicada en los manuales, similar en ambas versiones.

Procedimiento

Cada participante fue evaluado en una sola sesión por un experto estandarizado en la administración de ambas versiones. Los ítems fueron aplicados y calificados de acuerdo con las instrucciones de los manuales de forma independiente. Se cumplieron los criterios de puntuación de líneas de inicio y suspensión. Los ítems comunes a ambas escalas se puntuaron simultáneamente a partir de un ítem, evitando así los efectos de la práctica o entrenamiento y reduciendo la duración de la prueba. Para efectos de comparación, se generó el Puntaje Compuesto Lenguaje-Cognitivo del B-III; PC-LC (B-III), definido como el puntaje promedio de las PC-Cog y PC-Leng, para ser comparado con el ID Mental BSID-II. El retraso severo en el BSID-II equivale al resultado extremadamente bajo del B-III (69 o menos puntos).

Instrumentos

BSID-II. Consta de una Escala Mental con 178 elementos que evalúan las capacidades de memoria, habituación, solución de problemas, concepto de número, capacidad de generalización, clasificación, vocalizaciones, lenguaje y habilidades sociales. Una Escala Motora con 111 elementos que valora el control de los grupos musculares gruesos y finos; y manipulaciones

motrices finas implicadas en la prensión, uso adaptativo de los útiles de escritura e imitaciones de los movimientos de la mano.²

B-III. Consta de tres escalas: Escala cognitiva, basada en respuestas no verbales del niño y mide los procesos de aprendizaje, la capacidad de resolver problemas, la atención, la habilidad para contar objetos y clasificarlos, y las habilidades para jugar. Escala de lenguaje; dividida en las subescalas de Lenguaje Receptivo (LR) mide la capacidad del niño para comprender distintos estímulos, palabras o instrucciones en el entorno y Lenguaje Expresivo (LE) evalúa vocalizaciones, uso de palabras y construcción de oraciones. Escala motora incluye la subescala de MF que mide la coordinación manos-dedos y manos-ojos, y la subescala Motor Grueso (MG) que evalúa el control del niño sobre su cuerpo y las habilidades para mover torso y extremidades.³

Análisis estadístico

Las variables generales se analizaron a través de estadística descriptiva. Se compararon las medias (\bar{X}) y desviaciones estándar (SD) de los ID del BSID-II (Mental y Motor) con las PC del B-III (Cognitiva-Lenguaje y Motora) a través de análisis univariado, con prueba t de *Student* estimándose la magnitud de las diferencias entre el ID Mental BSID-II y el PC-LC (B-III) y entre el ID Motor BSID-II y la PC-Motora (B-III), verificando la presencia de los supuestos de uniformidad de varianza y distribución normal para aplicar pruebas paramétricas. Se establecieron las diferencias en la proporción de alteraciones entre las escalas correspondientes (tomando como punto de corte para alteración ID o PC < 70) mediante prueba chi² para tablas de contingencia, calculándose además la concordancia interpruebas mediante coeficiente *Kappa*.

Para determinar la capacidad de predicción de una versión de la escala respecto a su correspondiente, se realizó análisis de regresión

lineal múltiple, verificando la baja colinealidad, solo se incluyeron los predictores significativos mediante *stepwise forward*; utilizando como variables independientes las puntuaciones escalares Cog, LE, LR, MF y MG del B-III para la predicción de los ID Mental y Motor del BSID-II. La estimación de los coeficientes en la regresión se realizó sobre las puntuaciones escalares que expresan valores estandarizados para cada edad transformados a \bar{X} de 10 y SD de 3.

El estudio y la carta de consentimiento informado fueron aprobados por los comités de ética e investigación del Instituto Nacional de Pediatría (registro 086/2014) apegándose a los Principios éticos para las investigaciones médicas en seres humanos de la Declaración de Helsinki y al Art. 17 del Reglamento de la Ley General de Salud. Todos los padres o tutores de los pacientes firmaron el consentimiento informado.

RESULTADOS

Se realizaron 671 valoraciones de desarrollo clasificadas en 7 rangos de edad; 50.90% pertenecían al género femenino. Presentaron como antecedentes: peso al nacimiento 2237 ± 1171 g; 46.3 % con peso inferior a 2500 g, 49 % fueron prematuros, 48 % de término y 3% postérmino. Por su condición de riesgo tuvieron hospitalización al nacimiento entre 4 a 99 días (22.6 ± 8). La edad materna fue de 26.8 ± 6.5 años; 45% tuvo escolaridad materna de 3 años de secundaria o menor y, sólo 1.2 %, tuvo licenciatura o mayor. El 81 % de las madres o cuidadoras se dedicaba al hogar, 10 % con empleo informal y 9 % empleo formal.

El análisis de las diferencias de medias por prueba t de *Student* entre los resultados de las escalas correspondientes a los ID del BSID-II fueron significativamente menores que los PC del B-III ($p<0.0001$). El promedio de la diferencia entre el PC-LC (B-III) y el ID Mental BSID-II fue 14.04 puntos ($SD = 10.09$); la PC-Motora (B-III)

fue en promedio 17.87 puntos ($SD = 10.32$) más alta que los puntajes del ID Motor del BSID-II ($p < 0.0001$) (**Cuadro 1**). Correspondientemente, se detectó mayor proporción de alteraciones en ambas subescalas del BSID-II respecto a las del B-III, siendo mayor esta proporción entre las motoras (**Cuadro 1**).

La distribución de medias del ID Mental BSID-II y del PC-LC (B-III) por rangos de edad se muestra en la Figura 1. Podemos observar que las mayores diferencias se presentaron en el rango 5 (25 - 30 meses) con 19.3 puntos, seguido del rango 4 (19 - 24 meses) con 18.1 puntos. Las puntuaciones más altas en todos los rangos de edad fueron con el PC-LC (B-III) (**Figura 1**).

Respecto a las diferencias entre las medias del ID Motor BSID-II y PC-Motora (B-III), observamos la misma tendencia, fueron mayores en los últimos tres rangos: Rango 5 (25-30 meses), 6 (31-36 meses) y 7 (37-42 meses); con 27.8, 25.4 y 25.3 puntos de diferencia respectivamente, a favor del B-III (**Figura 2**).

Se analizó la frecuencia de resultados normales o alterados según los PC-LC (B-III) y el ID Mental BSID-II. De las 671 valoraciones realizadas,

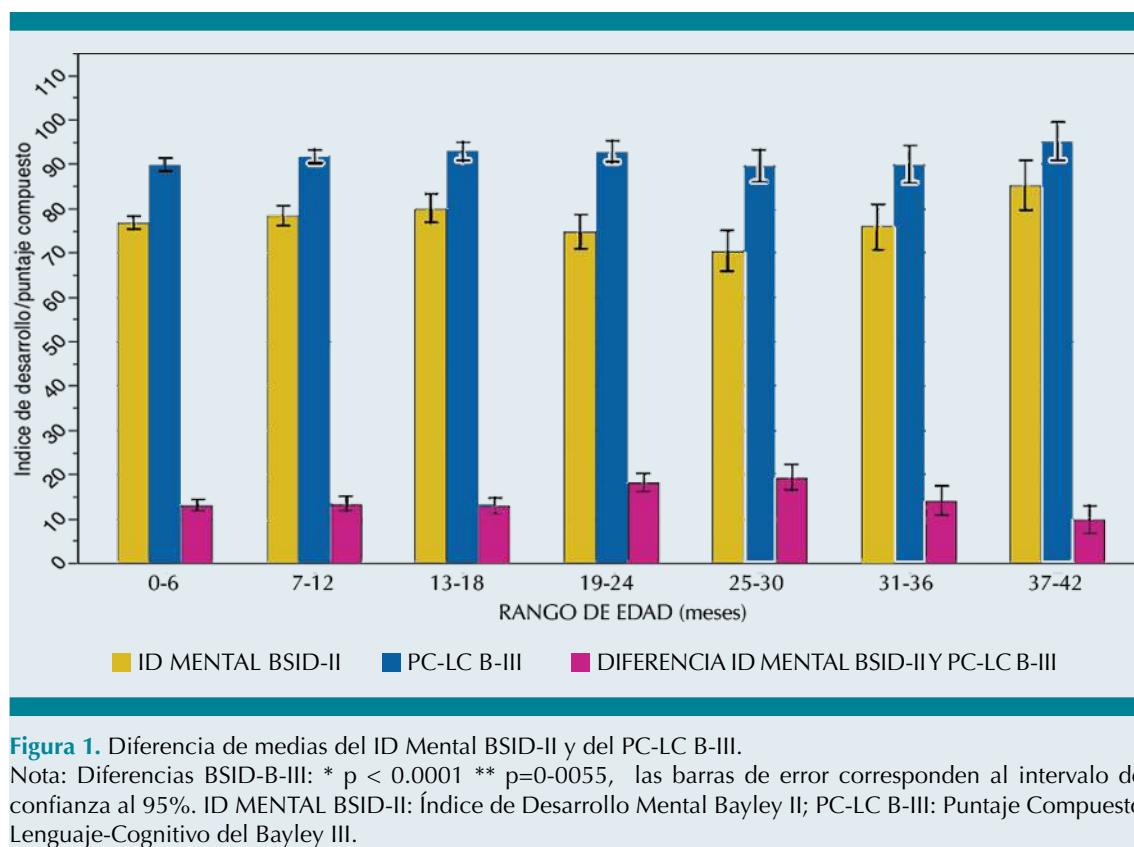
491 (73.2 %) fueron clasificadas como normales según el ID Mental BSID-II, pero 24 de éstas (3.6 %) resultaron alteradas con el PC-LC (B-III). El B-III caracterizó a 598 (89.1 %) con resultado normal; de las cuales 131 (19.5 %) fueron consideradas alteradas con el BSID-II ($\kappa = 0.2751$). A partir de estos resultados podemos observar que, en 155 valoraciones, las calificaciones no coincidieron entre ambas versiones; 131 fueron clasificadas por el BSID-II como alteradas y normales por el B-III; mientras que 24 valoraciones consideradas como alteradas por el B-III, fueron normales por el BSID-II (**Cuadro 2**). Respecto al componente motor de ambas escalas, el coeficiente $Kappa$ fue de 0.1373; de las 283 (42.2 %) evaluaciones con alteraciones en el ID Motor BSID-II, sólo 35 (5.21 %) también resultaron alteradas en la PC-Motora (B-III). Sólo una valoración (0.15 %) resultó alterada en la PC-Motora (B-III), pero normal en el ID Motor BSID-II (**Cuadro 2**).

Se realizaron análisis predictivos con regresión múltiple a partir de las puntuaciones escalares de las subescalas B-III. Para la predicción del ID Mental BSID-II de manera global, las subescalas cognitiva y de lenguaje tuvieron mayor peso en la predicción ($p < 0.001$) y la mejor determina-

Cuadro 1. Distribución de la presencia de alteración en las puntuaciones del BSID-II y B-III

| Escala | Subescala | \bar{X} | SD | Presencia de alteraciones | N | % |
|---------|-----------------------------|-----------|------|---------------------------|-----|-------------------|
| BSID-II | Índice de desarrollo mental | 76.9 | 13.9 | Normal | 491 | 73.2 |
| | | | | Alterado | 180 | 26.8 ^a |
| | Índice de desarrollo motor | 71.6 | 13.9 | Normal | 388 | 57.8 |
| | | | | Alterado | 283 | 42.2 ^b |
| B-III | Puntaje compuesto cognitivo | 95.3 | 13.1 | Normal | 645 | 96.1 |
| | | | | Alterado | 26 | 3.9 ^a |
| | Puntaje compuesto lenguaje | 86.9 | 12.4 | Normal | 614 | 91.5 |
| | | | | Alterado | 57 | 8.5 ^a |
| | Puntaje compuesto motor | 89.5 | 12.8 | Normal | 635 | 94.6 |
| | | | | Alterado | 36 | 5.4 ^b |

^a prueba de chi cuadrada diferencia de proporciones entre escala BSID-II mental y B-III cognitiva-lenguaje $p < 0.0001$; ^b diferencia entre BSID-II motor y B-III motor $p < 0.0001$.



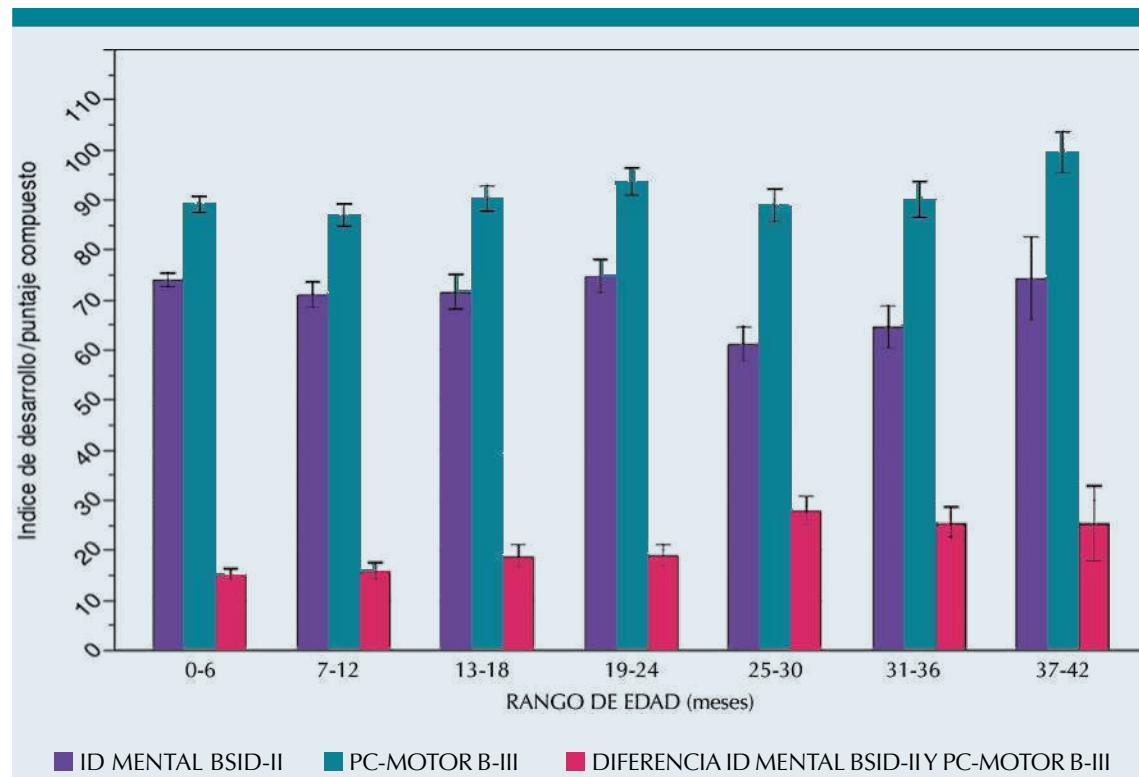
ción ($r^2 \geq 0.67$) sucedió del rango 3 en adelante (12 a 42 meses); mientras que para el área Motora fueron las Puntuaciones MG y MF en los rangos 1 al 4 (0 a 24 meses) y en el rango 6 (30 a 36 meses) con $r^2 \geq 0.63$ (**Cuadro 3**).

DISCUSIÓN

Las diferencias encontradas entre las dos versiones del Bayley^{2,3} muestran una diferencia de 10 y 18.4 puntos para la subescala mental y de 17.9 puntos para el motor, al grado de suponer que alguna de las dos versiones no hace una estimación adecuada del desarrollo o ambas presentan problemas de sobreestimación o subestimación del desarrollo. Nuestros resultados muestran valores bajos en las estimaciones del desarrollo con el BSID-II, considerando que la población que se tomó como base fueron niños de condi-

ción socioeconómica baja con algún grado de riesgo perinatal, especialmente asfixia perinatal y prematuridad. Estudios anteriores han informado una media de las puntuaciones cognitivas y del lenguaje del B-III mayores en seis a 10 puntos y en la motora en tres a 18 puntos más que sus correspondientes del BSID-II. Estas diferencias se han atribuido a aspectos relacionados con las características metodológicas y de cohortes entre los estudios.⁶⁻¹¹

Se han reportado diferencias con igual magnitud a la descrita por nosotros en el área motora; pero con menor discrepancia en la mental. Jari *et al.* en el 2013 encontraron que la mediana de la PC-Cog fue 7 puntos más alta que el ID Mental BSID-II; y la PC-Motora (B-III) fue 18 puntos mayor que el ID Motor BSID-II.¹¹ Nuestros resultados muestran que las diferencias

**Figura 2.** Diferencia de medias del ID Motor BSID-II y PC-Motora (B-III) por rangos de edad*.

Nota: Diferencia, $p < 0.0001$ en todos los pares de barras, las barras de error corresponden al intervalo de confianza al 95%, ID MOTOR BSID-II: Índice de Desarrollo Motor BSID-II; PC MOTOR B-III: Puntaje Compuesto Motor del B-III

Cuadro 2. Presencia de alteraciones en el BSID-II respecto a alteraciones en las subescalas correspondientes al B-III

| B-III Cognitivas y/o Lenguaje | | | |
|---|--------------------------|------------|--------------|
| BSID-II Mental | Normal | Alterado | Total |
| | Normal 467 (69.6%) | 24 (3.6%) | 491 (73.2%) |
| | Alterado 131 (19.5%) | 49 (7.3%) | 180 (26.8%) |
| 598 (89.1%) | | | |
| 73 (10.9%) | | | |
| 671 (100%) | | | |
| Kappa 0.2751 (0.1964 - 0.3539) $p < 0.0001$ | | | |
| B-III Motora | | | |
| BSID-II Motora | Normal | Alterado | Total |
| | Normal 387 (57.68%) | 1 (0.15%) | 388 (57.83%) |
| | Alterado 248 (36.96%) | 35 (5.21%) | 283 (42.17%) |
| 635 (94.64%) | | | |
| 36 (5.36%) | | | |
| 671 (100%) | | | |
| Kappa 0.1373 (0.0936 - 0.1811) $p < 0.0001$ | | | |

Cuadro 3. Regresión lineal entre los índices de desarrollo y puntuaciones motoras para la predicción del BSID-II a partir de las puntuaciones del B-III

| RANGO EDAD | ID MENTAL BSID-II | | ID MOTOR BSID-II | |
|---------------|---|----------------|---|----------------|
| | FÓRMULA DE PREDICCIÓN | r ² | FÓRMULA DE PREDICCIÓN | r ² |
| 1 | 45.58+2.06(Cog)+1.13 (MF)+0.52 (MG) ** | .54 | 40.06+1.03(MF)+3.15(MG) ** | .61 |
| 2 | 34.50+3.29 (Cog)+1.64 (LE) ** | .54 | 38.71+0.73 (MF)+3.68 (MG) ** | .69 |
| 3 | 15.07+2.54 (Cog)+1.13 (LE)+1.60 (LR)+1.19 (MF)+1.04 (MG) ** | .75 | 26.82+0.69(MF)+5.42(MG) ** | .75 |
| 4 | 3.13+2.67 (Cog)+3.49 (LE)+1.18 (LR)+0.99 (MF) ** | .84 | 26.41+1.25(MF)+5.04(MG) ** | .65 |
| 5 | 12.76+2.27 (Cog)+3.29 (LE)+1.45 (MF) ** | .70 | 30.90+4.19 (MG) ** | .51 |
| 6 | 21.73+4.79 (LE)+1.99 (MF) ** | .68 | 25.90+4.71 (MG) ** | .63 |
| 7 | 37.07+5.17 (Cog) ** | .81 | 38.70+3.81 (MG) * | .27 |
| GLOBAL | 35.43+2.48 (Cog)+1.00 (LE)+0.97 (MF)+0.63 (MG) -0.14* Meses** | .56 | 37.38+1.11 (MF)+3.62 (MG) -0.27*Meses** | .58 |

Nota. Cog: Puntuación Escalar Cognitiva; LR: Puntuación Escalar Lenguaje Receptivo; LE: Puntuación Escalar Lenguaje Expresivo; MF: Puntuación Escalar Motor Fino; MG: Puntuación Escalar Motor Grueso. **p < 0.0001; *p < 0.05

entre las áreas motoras evaluadas en diferentes grupos de edad, fueron muy parecidas (17.87 puntos), pero aproximadamente el doble (si se considera el promedio de la diferencia del área lenguaje y cognitiva = 14.04) para el caso de las áreas mentales. Por su parte, Yi *et al.* en 2018 informaron que la media de la PC-Cog fue 5.8 puntos superior a los reportados en el ID Mental BSID-II, y la PC-Motora (B-III) fue 7.9 puntos más alta que los puntajes del ID Motor BSID-II.⁷ Ambos estudios concluyen que el B-III identifica significativamente menos niños con retrasos en el desarrollo.

Sin embargo, la magnitud de las diferencias entre las áreas cognitivas y de lenguaje del B-III respecto al BSID-II, encontradas en este trabajo también se han descrito en otras investigaciones. Recientemente (en 2020), Çelik *et al.* reportaron diferencias de 13.1 ± 9.1 puntos más altos en el PC-Cog y 8.6 ± 8 en el PC-LC (B-III) que el ID Mental BSID-II; y 14.4 ± 10.5 puntos más en la PC-Motora (B-III) respecto al ID Motor BSID-II.¹² Resultados similares ya habían sido descritos por Vohr en el 2012, al describir puntuaciones

más altas con el B-III que las obtenidas en el BSID-II en ambos ID. También plantearon que el manual de B-III establece que la PC-Cog es siete puntos más alta que el ID Mental BSID-II; y que, en su cohorte, fueron 11 puntos más altos.⁹ Otro estudio, observaron diferencias de hasta 1 SD más alta en el PC-Cog que el ID Mental BSID-II; las medias del PC-Motora (B-III) fueron 9.0 ± 11.9 puntos superiores al ID Motor BSID-II.¹³ Los mismos autores al utilizar los puntos de corte propuestos por las pruebas, reportaron que el 40 % de los niños de su muestra, fueron clasificados con menor retraso con el B-III en la parte cognitiva, mientras que sólo 1 (< 2%) como más severo. En nuestros datos 19.5 % fueron calificados más altos con el B-III y sólo 3.6 % más bajo (Tabla 2) considerando que sólo diferenciamos normal contra alterado (PC o ID < 70: < 2 SD). De lo anterior se desprende como una vía de salida, ajustar los puntos de corte siguiendo criterios de nuevas normalizaciones o predicciones.^{11,12,14,15} Aunque ello limita las posibilidades de establecer comparaciones entre distintas versiones e incluso pruebas a partir de las puntuaciones estandarizadas.

El análisis que se presenta toma como base a las *puntuaciones estandarizadas*, las cuales representan la posición de esa puntuación respecto a su distribución alrededor de la media de una muestra de referencia o normativa. Con relación a esas normas, los resultados pueden interpretarse en ambas escalas con relación a la media de 100 y su magnitud en términos de la desviación estándar (DS) igual a 15 para los Índices de Desarrollo del BSID-II y Puntuaciones compuestas para el B-III. Por lo mismo las puntuaciones de la muestra inferiores a la norma en ambas versiones, pero más cercanas al B-III (0.29 a 0.87 DS) que en el BSID-II (1.54 a 1.89 DS). Otros estudios han reportado desempeños más bajos en población de niños mexicanos aplicando el Bayley u otras pruebas de desarrollo, por lo mismo los resultados mostrados por nuestros niños nos parecen adecuados, considerando que se trata de una población de alto riesgo.¹⁶⁻¹⁹

No existe un supuesto por el cual se deba asumir que las diferencias entre una y otra versión de las escalas sean constantes con relación a la edad; los valores pueden variar por razones intrínsecas a la construcción de las mismas escalas ofreciendo mayor o menor dificultad para transformar las puntuaciones en las poblaciones normativas y en las poblaciones clínicas; y por razones extrínsecas como la mayor o menor afectación del desarrollo en los grupos de edad.

Los estudios deben asumir que las diferencias pueden cambiar con la edad o momento del desarrollo, en nuestro estudio, por ejemplo, tuvieron mayor variación en los componentes mentales/cognitivo-Lenguaje y para el caso del BSID-II, las medias fueron muy cercanas al valor 70 lo cual origina una mayor proporción de casos clasificados como alterados, incrementando la diferencia de proporciones de alteración de una prueba respecto a otra. Sin embargo, es en estos tipos de población de riesgo en las que debe compararse y evaluarse las estimaciones que realizan las pruebas pues son los usuarios de mayor interés y

en quienes las decisiones se toman a partir de los parámetros numéricos y categóricos.

Otro aspecto que puede generar sesgos en la puntuación alcanzada corresponde a los criterios de inicio y suspensión, el inicio cuando el niño es capaz de realizar un número de ítems de forma consecutiva asumiendo que puede realizar todos los previos y la detención, cuando el niño ya no puede realizar de forma continua un número determinado de ítems, deteniéndose la evaluación de los subsecuentes. El que los formatos se dividan en cinco para el caso del B-III, mientras que en el BSID-II sólo en dos, influye en que las posibilidades de detención sean mayores en el BSID-II ya que el rezago de uno de sus componentes puede frenar el desarrollo de toda la escala. Mientras que el B-III permite el avance de las otras escalas si la dificultad se expresa solo en una de ellas. Ambos aspectos determinan que la variación entre las escalas dependa de la distribución de ítems que representan las subescalas que la componen y, en ese sentido, al existir una mayor probabilidad de detener la aplicación con el BSID-II, se estaría subestimando el potencial de desarrollo con que cuenta el niño.^{8,20}

Se ha señalado que las diferencias en los puntajes en pruebas similares, pero con inicio o suspensión inadecuados no permiten una distinción precisa de los niveles de habilidad.²⁰ Los puntos de inicio y suspensión se han ampliado y mejorado en el B-III buscando impactar en las discrepancias en los puntajes; se estableció que los niños deben acreditar los primeros tres ítems consecutivos de su rango de edad para avanzar en la evaluación y se detiene cuando el niño deja de acreditar cinco ítems consecutivos, esta indicación supone una mayor probabilidad de que el niño logre un puntaje directo mayor considerando que cada punto puede equivaler a unos dos a cuatro puntos de la *puntuación estandarizada*, según la edad y la escala. También debe considerarse cambios seculares en las muestras y un probable efecto Flynn.²¹

El análisis de regresión está orientado a encontrar una fórmula que permita transformar los datos de una valoración hacia la otra y comparar resultados, así como integrar datos en seguimientos a largo plazo en determinados grupos poblacionales o en riesgo. Cuando se intenta hacer la predicción de los valores del BSID-II Mental se encontró que la incorporación de la Escala Cognitiva (Cog) tiene la mayor relevancia, pero incorporar además los valores de la subescala de LE y MF mejora la r^2 de la regresión. La predicción del BSID-II Motora dependió predominantemente de las puntuaciones MG del B-III y en algunas edades aportaba, en menor importancia, la puntuación MF (**Cuadro 3**), esto se explica porque indicadores de la función motriz fina que implican procesos visoperceptuales y praxias tempranas tienen también un componente cognitivo y de esa forma estaban ubicados en la BSID-II. Una limitación del estudio es no poder predecir escalas específicas del B-III a partir del BSID-II, pues al ser éstas más globales, el error de la predicción es alto. La validez externa de los hallazgos es relativa a la prueba, aunque las características del desarrollo de los niños pueden implicar algunas variaciones, al trabajarse sobre valores normalizados y ser los valores de la varianza similares a los previstos en los manuales y otros reportes, debe asumirse como susceptibles de generalizarse en otros grupos de riesgo.

CONCLUSIÓN

Existen diferencias importantes en las estimaciones del desarrollo hechas con las dos versiones del Bayley, El BSID-II califica más bajo el desarrollo que el B-III en una población de riesgo perinatal. Dichas diferencias en la estimación se pueden explicar por las muestras de referencia utilizadas en la normalización, los criterios de inicio y suspensión durante la aplicación de las escalas y los métodos de determinación de las puntuaciones.

Agradecimientos

Agradecemos a los expertos en diagnóstico del desarrollo infantil del Centro de Investigación del Neurodesarrollo del Instituto Nacional de Pediatría: Jasbet Zapoteco Nava; Aline Morales Ramírez; Esperanza Ontiveros Mendoza; Beatriz Menes Arzate y Olivia Arzate Rivera.

REFERENCIAS

1. Lowe JR, Erickson SJ, Schrader, R., Duncan AF. Comparison of the Bayley II Mental Developmental Index and the Bayley III Cognitive Scale: are we measuring the same thing? *Acta Paediatr.* 2012;101(2):e55-e58. <https://doi.org/10.1111/j.1651-2227.2011.02517.x>
2. Bayley, N. *Bayley Scales of Infant Development*. San Antonio, TX: The Psychological Corporation. 1993.
3. Bayley, N. *Bayley Scales of Infant and Toddler Development*. San Antonio, TX: The Psychological Corporation. 2006.
4. Johnson S, Marlow N. Developmental screen or developmental testing? *Early Hum Dev.* 2006;82(3):173-183. <https://doi.org/10.1016/j.earlhumdev.2006.01.008>
5. Pérez-López J, de la Nuez AGB, Martínez-Fuentes MT, Díaz-Herrero Á, Sánchez-Caravaca J, Fernández-Rego FJ, et al. Las escalas Bayley BSID-I frente a BSID-II como instrumento de evaluación en Atención Temprana. *Anales de Psicología/Annals of Psychology.* 2012;28(2):484-489. <https://doi.org/10.6018/analeps>
6. Anderson PJ, De Luca CR., Hutchinson, E., Roberts, G., Doyle, L. W. Underestimation of developmental delay by the new Bayley-III Scale. *Arch Pediatr Adolesc Med.* 2010;164(4):352-356. <https://doi.org/10.1016/j.siny.2013.11.012>
7. Yi YG, Sung IY, Yuk JS. Comparison of second and third editions of the Bayley Scales in children with suspected developmental delay. *Ann Phys Rehabil Med.* 2018;42(2):313. <https://doi.org/10.5535/arm.2018.42.2.313>
8. Moore T, Johnson S, Haider S, Hennessy E, Marlow N. Relationship between test scores using the second and third editions of the Bayley Scales in extremely preterm children. *J Pediatr.* 2012;160(4):553-558. <https://doi.org/10.1016/j.jpeds.2011.09.047>
9. Vohr BR, Stephens BE, Higgins RD, Bann CM, Hintz SR, Das A, et al. Are outcomes of extremely preterm infants improving? Impact of Bayley assessment on outcomes. *The Journal of Pediatrics.* 2012;161(2):222-228. <https://doi.org/10.1016/j.jpeds.2012.01.057>
10. Reuner G, Fields AC, Wittke A, Löpprich M, Pietz J. Comparison of the developmental tests Bayley-III and Bayley-II in 7-month-old infants born preterm. *Eur J Pediatr.* 2013; 172:393–400. <https://doi.org/10.1007/s00431-012-1902-6>
11. Jary S, Whitelaw A, Walløe L, Thoresen M. Comparison of Bayley-2 and Bayley-3 scores at 18 months in term infants

- following neonatal encephalopathy and therapeutic hypothermia. *Dev Med Child Neurol.* 2013;55(11):1053-1059. <https://doi.org/10.1111/dmcn.12208>
- 12. Çelik P, Ayrancı Sucaklı I, Yakut HI. Which Bayley-III cut-off values should be used in different developmental levels? *Turk J Med Sci.* 2020; 50:764-770. <https://doi.org/10.3906/sag-1910-69>
 - 13. Sharp M, DeMauro SB. Counterbalanced comparison of the BSID-II and Bayley-III at eighteen to twenty-two months corrected age. *J Dev Behav Pediatr.* 2017;38(5):322-329. <https://doi.org/10.1097/DBP.0000000000000441>
 - 14. Kaya-Kara O, Kerem-Gunel M, Yigit S. Correlation of the Bayley scales of infant-toddler development-and neurosensory motor assessment in preterm infants during the first year of life. *Turk J Pediatr.* 2019;61(3): 399-406. <https://doi.org/10.24953/turkjped.2019.03.012>
 - 15. Duncan AF, Bann C, Boatman C, Hintz SR, Vaucher YE, Vohr BR, et al. Do currently recommended Bayley-III cutoffs overestimate motor impairment in infants born < 27 weeks gestation? *J Perinatol.* 2015;35(7):516-521. <https://doi.org/10.1038/jp.2014.243>
 - 16. Mendoza-Sánchez V, Hernández-Negrete LE, Cázares-Ortiz M, et al. Neurodesarrollo en niños con cardiopatía congénita a los 30 meses de edad. *Rev Mex Pediatr.* 2019;86(4):143-146. <https://doi:10.35366/SP194C>
 - 17. Rivera-González R, Villanueva-Romero Y, Amaro-López L, Sánchez-Pérez C, Figueroa-Olea M, Soler-Limón KM. Validez concurrente de las cartillas de vigilancia para identificar alteraciones en el desarrollo del lactante. *Ciencias Clínicas,* 2014;15: 22-29. <https://doi:10.1016/j.cc.2014.08.001>
 - 18. Rivera-González R, Sánchez C, Corral-Guille I, Figueroa-Olea M, Soler- Limón K, Martínez-Vázquez I, et al. Edad de presentación de los reactivos del Test de Denver II en Niños de 0 a 4 años de edad del Estado de Morelos. *Salud mental,* 2013; 36(6): 459-470. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0185-33252013000600003&lng=es
 - 19. Sánchez C, Rivera-González R, Martínez-Vázquez I, Corral-Guille I, Figueroa-Olea M, Cázares A, Soler-Limón KM. Indicadores de desarrollo del CAT/CLAMS en lactantes de una comunidad urbana de México. Reporte preliminar. *Revista Mexicana de Pediatría,* 2008;75(5); 217-227.
 - 20. Gagnon SG, Nagle RJ. Comparison of the revised and original versions of the Bayley Scales of Infant Development. *Sch Psychol Int.* 2000;21(3),293-305. <https://doi.org/10.1177/0143034300213006>
 - 21. Flynn JR. Searching for justice: The discovery of IQ gains over time. *American Psychologist,* 1999;54(1):5-20. <https://doi.org/10.1037/0003-066X.54.1.5>