



UNIVERSIDAD  
CATÓLICA  
DE CUENCA

**UNIVERSIDAD CATÓLICA DE CUENCA**

*Comunidad Educativa al Servicio del Pueblo*

**FACULTAD DE CIENCIAS DE LA SALUD**

**CARRERA DE MEDICINA**

**EFFECTIVIDAD DEL TEST DE DENVER VERSUS EL TEST  
BAYLEY-III EN LA EVALUACIÓN DEL NEURODESARROLLO  
INFANTIL EN PACIENTES DE 1-42 MESES: REVISIÓN  
SISTEMÁTICA**

**PROYECTO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL  
TÍTULO DE MÉDICA**

**AUTOR: DANNA EMILIA BERMÚDEZ PADILLA**

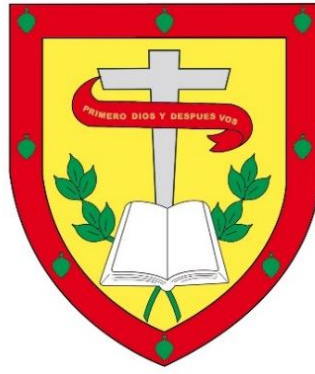
**JOSELYN LIZBETH GONZÁLEZ MINCHALA**

**DIRECTOR: DRA. JOHANA ALICIA PALACIOS SACOTO**

**CUENCA - ECUADOR**

**2026**

**DIOS, PATRIA, CULTURA Y DESARROLLO**



**UNIVERSIDAD CATÓLICA DE CUENCA**

*Comunidad Educativa al Servicio del Pueblo*

**FACULTAD DE CIENCIAS DE LA SALUD**

**CARRERA DE MEDICINA**

**EFFECTIVIDAD DEL TEST DE DENVER VERSUS EL TEST  
BAYLEY-III EN LA EVALUACIÓN DEL NEURODESARROLLO  
INFANTIL EN PACIENTES DE 1-42 MESES: REVISIÓN  
SISTEMÁTICA**

**PROYECTO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL  
TÍTULO DE MÉDICA**

**AUTOR: DANNA EMILIA BERMÚDEZ PADILLA**

**JOSELYN LIZBETH GONZÁLEZ MINCHALA**

**DIRECTOR: DRA. JOHANA ALICIA PALACIOS SACOTO**

**CUENCA - ECUADOR**

**2026**

**DIOS, PATRIA, CULTURA Y DESARROLLO**



**Declaratoria de Autoría y Responsabilidad**

**Danna Emilia Bermúdez Padilla** portador(a) de la cédula de ciudadanía N° **0106511140**. Declaro ser el autor de la obra: “**efectividad del test de denver versus el test bayley-iii en la evaluación del neurodesarrollo infantil en pacientes de 1-42 meses: revisión sistemática**”, sobre la cual me hago responsable sobre las opiniones, versiones e ideas expresadas. Declaro que la misma ha sido elaborada respetando los derechos de propiedad intelectual de terceros y eximo a la Universidad Católica de Cuenca sobre cualquier reclamación que pudiera existir al respecto. Declaro finalmente que mi obra ha sido realizada cumpliendo con todos los requisitos legales, éticos y bioéticos de investigación, que la misma no incumple con la normativa nacional e internacional en el área específica de investigación, sobre la que también me responsabilizo y eximo a la Universidad Católica de Cuenca de toda reclamación al respecto.

Azogues, **13 de mayo del 2026**

F: 

**Danna Emilia Bermúdez Padilla**

**C.I. 0106511140**



**Declaratoria de Autoría y Responsabilidad**

**Joselyn Lizbeth González Minchala** portador(a) de la cédula de ciudadanía N° **0302471420**. Declaro ser el autor de la obra: “**efectividad del test de denver versus el test bayley-iii en la evaluación del neurodesarrollo infantil en pacientes de 1-42 meses: revisión sistemática**”, sobre la cual me hago responsable sobre las opiniones, versiones e ideas expresadas. Declaro que la misma ha sido elaborada respetando los derechos de propiedad intelectual de terceros y eximo a la Universidad Católica de Cuenca sobre cualquier reclamación que pudiera existir al respecto. Declaro finalmente que mi obra ha sido realizada cumpliendo con todos los requisitos legales, éticos y bioéticos de investigación, que la misma no incumple con la normativa nacional e internacional en el área específica de investigación, sobre la que también me responsabilizo y eximo a la Universidad Católica de Cuenca de toda reclamación al respecto.

Azogues, 13 de mayo del 2026

F: 

**Joselyn Lizbeth González Minchala**

**C.I. 0302471420**

## CERTIFICACIÓN DEL DIRECTOR DE TESIS

Johana Alicia Palacios Sacoto

DOCENTE DE LA CARRERA DE MEDICINA

De mi consideración:

Certifico que el presente trabajo de titulación denominado: "**Efectividad del test de Denver versus el test Bayley-III en la evaluación del neurodesarrollo infantil en pacientes de 1-42 meses: revisión sistemática**", realizado por: **Danna Emilia Bermúdez Padilla y Joselyn Lizbeth González Minchala**, con documentos de identidad: **0106511140 y 0302471420**, previo a la obtención del título de **Médica** ha sido asesorado, orientado, revisado y supervisado durante su ejecución, bajo mi tutoría en todo el proceso, por lo que certifico que el presente documento, fue desarrollado siguiendo los parámetros del método científico, se sujeta a las normas éticas de investigación que exige la Universidad Católica de Cuenca, por lo que está expedito para su presentación y sustentación ante el respectivo tribunal.

Azogues, 14 de mayo del 2026

  
Dra. Johana Palacios S.  
MÉDICO PEDIATRA  
C.C. 0302023569

JOHANA ALICIA PALACIOS SACOTO

C.I. 0302023569

DIRECTOR

## **AGRADECIMIENTO**

La culminación de nuestro primer escalón de la vida profesional no habría sido posible sin el apoyo incondicional, guía y motivación de muchas personas, a quienes deseamos expresar nuestro más profundo y sincero agradecimiento.

En primer lugar, quiero manifestar mi gratitud a Dios, quien nos ha brindado la fortaleza para avanzar en los momentos más duros, la salud y la sabiduría para complementar esta grata profesión.

A nuestra familia, por ser un pilar fundamental en nuestras vidas además de ser un motor que día a día nos está impulsando para alcanzar nuestras metas. A nuestros padres, quienes, con su ejemplo de trabajo incansable, su perseverancia y valores nos han enseñado que la dedicación y el esfuerzo es capaz de superar cualquier obstáculo que nos pone la vida. A nuestras madres, quienes con mucha paciencia y el amor incondicional que solo ellas nos saben dar han sido nuestro más grande refugio y motivación en todo momento. Sus palabras de aliento, su capacidad para hacernos sentir en paz y que nunca nos diéramos por vencidas. A nuestros hermanos quienes nos han hecho compañía en los momentos llenos de estrés y han sido capaces de hacernos reír, quienes nos ven como su ejemplo a seguir por la perseverancia que hemos tenido. Cada uno de ustedes ha sido un modelo de fortaleza y resiliencia, y esta tesis es un reflejo del cariño y la unión que compartimos como familia.

Expresamos nuestro especial agradecimiento a nuestra tutora de tesis, Dra. Johana Palacios, por su invaluable orientación, paciencia y dedicación. Su experiencia y conocimiento han sido fundamentales. Gracias por cada consejo, por el tiempo que nos ha brindado a pesar de haber estado durante su momento de maternidad y por la paciencia que nos ha tenido para realizar cada una de las correcciones.

Por último, queremos agradecer a todas las personas que de cierta manera han contribuido a que este gran paso se hiciera posible. Sus palabras, su apoyo y sus gestos siempre estarán presente en nuestros corazones.

Esta tesis no nos representa un logro más, sino el fruto de nuestro esfuerzo en conjunto con nuestra familia y su amor incondicional. A todos ellos, con todo nuestro cariño y gratitud les agradecemos por ser parte de esta gran trayectoria y por siempre habernos tenido fe.

Con aprecio,

*Danna Emilia Bermúdez Padilla & Joselyn Lizbeth González Minchala*

## **DEDICATORIA**

A mis padres, Barny y Ana quienes me han acompañado a lo largo de estos años de mi carrera y han visto cada obstáculo que se ha atravesado durante el camino y, aun así, han buscado la manera de apoyarme. A mi enamorado Pedro quien en algunas ocasiones se ha convertido en mi hombro derecho y a pesar de las adversidades continúa siendo mi mejor amigo y cómplice. A mi abuelita María, que, a pesar de no vivir en la misma ciudad, ha estado en todo momento esperando con ansias que llegará el final de esta meta. Finalmente, esto va dedicado a mí, a esa chica que jamás se le cruzó por la mente que llegaría a convertirse en médico y que pese haber tenido duros momentos dentro de la carrera y haber atravesado por pérdidas de seres importantes y valiosas durante este trayecto, supo que luego de la tormenta sale el sol y que rendirse jamás fue una opción.

*(Danna Emilia Bermúdez Padilla)*

A mis padres, Byron y Angélica, cuya confianza inquebrantable, apoyo constante, sabios consejos y paciencia infinita han sido la piedra angular que ha sostenido cada paso en mi camino hacia este logro. Ustedes son y serán siempre mi mayor fuente de inspiración para alcanzar mis metas y sueños. A mis hermanos, Daniela y Lenin, quienes, con su apoyo incondicional y amor desbordante, me han brindado no solo su confianza en mis habilidades, sino también su tiempo y disposición para ser parte de mis prácticas y aprendizajes. A mis queridos tíos y abuelos, quienes, a pesar de las ocupaciones de sus vidas, siempre encontraban un momento para aconsejarme, escucharme y brindarme su respaldo. Y a mi preciada compañera, mi perrita Muñeca, cuya compañía iluminó innumerables noches de estudio y cuyos desvelos, caricias y muestras de cariño hicieron que los días más difíciles se sintieran más llevaderos. Aunque su partida dejó un vacío, su memoria y amor permanecen vivos en mi corazón, recordándome que su espíritu estará conmigo en cada paso que dé en el futuro. A todos ustedes, dedico este pequeño gran logro, con profunda gratitud y amor eterno.

*(Joselyn Lizbeth González Minchala)*

# Efectividad del test de Denver versus el test Bayley-III en la evaluación del neurodesarrollo infantil en pacientes de 1-42 meses: revisión sistemática

Danna Emilia Bermúdez Padilla, Joselyn Lizbeth González Minchala, Johana Alicia Palacios Sacoto

Universidad Católica de Cuenca, [danna.bermudez.40@est.ucacue.edu.ec](mailto:danna.bermudez.40@est.ucacue.edu.ec),

[joselyn.gonzalez.20@est.ucacue.edu.ec](mailto:joselyn.gonzalez.20@est.ucacue.edu.ec)

## RESUMEN

**Objetivos:** Evaluar la efectividad del Test de Denver en comparación con el Test Bayley-III para la evaluación del neurodesarrollo infantil en pacientes de 1-42 meses mediante una revisión sistemática. **Metodología:** Este estudio revisó cualitativamente estudios aleatorizados sobre la sensibilidad y especificidad del Test de Denver y Bayley-III. Incluyó revistas internacionales entre 2014 y 2024. La revisión sigue la declaración PRISMA-2020, utilizando términos MeSH y operadores lógicos. **Resultados:** El Test de Bayley-III presenta una sensibilidad en el índice motor entre 26.3%-76% y una especificidad de 27%-99.4%. Los VPP varían entre 38%-96.4%, con un LR+ de 1.04 hasta 103.2. Mientras que el Test de Denver, presenta una sensibilidad en la motricidad gruesa de 65.16%-86.6%. Los VPP y VPN son altos (72.5%-97.12%), con un LR+ entre 1.1 y 46.2. **Discusión:** Las dos pruebas presentan diferencias en sensibilidad, especificidad y valores predictivos. El test de Denver muestra alta sensibilidad en el área del lenguaje, aunque baja especificidad, mientras que el test de Bayley-III cuenta con una alta precisión en distintas áreas. **Limitaciones:** Dificultad para encontrar artículos en relación a la validez del test de Denver. Además, existen artículos de acceso restringido que no aportan información necesaria con alto riesgo en los sesgos afectando su validez. **Conclusiones:** Ambos test resultaron ser de gran utilidad, el test de Bayley-III presenta mayor precisión como sensibilidad y especificidad en áreas como cognitiva, motora y lenguaje, mientras que Denver es adecuado para un cribado inicial.

*Palabras clave:* desarrollo infantil, desempeño psicomotor, diagnóstico precoz, sensibilidad y especificidad, trastornos del neurodesarrollo

Effectiveness of the Denver Test versus the Bayley-III Test in the Assessment of Child Neurodevelopment in Patients Aged 1 to 42 Months: A Systematic Review

**ABSTRACT**

**Objectives:** To evaluate the effectiveness of the Denver Test compared to the Bayley-III Test for assessing child neurodevelopment in children from 1 to 42 months of age through a systematic review. **Methodology:** This study conducted a qualitative review of randomized trials examining the sensitivity and specificity of the Denver Test and the Bayley-III Test. It included international journal articles published between 2014 to 2024. The review follows the PRISMA-2020 statement, using MeSH terms and logical operators. **Results:** The Bayley-III Test showed a sensitivity in the motor index ranging from 26.3% to 76% and a specificity of 27% to 99.4%. The PPVs range from 38% to 96.4%, with an LR+ of 1.04 to 103.2. Meanwhile, the Denver Test showed a sensitivity for gross motor skills ranging from 65.16% to 86.6%. The PPV and NPV are high (72.5%–97.12%), with an LR+ between 1.1 and 46.2. **Discussion:** The two tests differ in terms of sensitivity, specificity, and predictive values. The Denver Test shows high sensitivity in the area of language, although it demonstrates low specificity, while the Bayley-III Test demonstrates high accuracy in various areas. **Limitations:** There was a difficulty in finding articles related to the validity of the Denver Test. Additionally, some restricted-access articles did not provide necessary information, posing a high risk of bias that affects their validity. **Conclusions:** Both tests proved to be highly useful; the Bayley-III Test demonstrates greater accuracy in terms of sensitivity and specificity in areas such as cognition, motor skills, and language, while the Denver Test is suitable for initial screening.

*Keywords:* child development, psychomotor performance, early diagnosis, sensitivity and specificity, neurodevelopmental disorders

## ÍNDICE

|   |     |
|---|-----|
| RESUMEN.....  | VI  |
| ABSTRACT.....   | VII |
| 1. Introducción .....                                       | 1   |
| 2. Objetivos .....  | 3   |
| 2.1. Objetivo general: .....                                | 3   |
| 2.2. Objetivos específicos:.....                            | 3   |
| 3. Métodos.....   | 3   |
| 3.1. Diseño del estudio .....                               | 3   |
| 3.1.1. Criterios de elegibilidad: .....                     | 4   |
| 3.2. Extracción de datos.....                               | 4   |
| 3.2.1. Fuentes de información: .....                        | 4   |
| 3.2.2. Estrategia de búsqueda: .....                        | 4   |
| 3.2.3. Selección de estudio:.....                           | 5   |
| 3.2.4. Lista de los datos: .....                            | 5   |
| 3.3. Evaluación de la calidad de los estudios .....         | 5   |
| 3.4. Evaluación de riesgo de sesgos basada en dominios..... | 5   |
| 3.5. Análisis de la información.....                        | 6   |
| 3.5.1. Medidas de resumen:.....                             | 6   |
| 3.5.2. Síntesis de Resultados: .....                        | 6   |
| 4. Resultados .....   | 6   |
| 4.1. Diagrama de flujo .....                                | 6   |
| 4.2. Calidad de estudios.....                               | 8   |
| 4.3. Riesgo de sesgos basada en dominios .....              | 10  |
| 4.4. Características de los estudios .....                  | 11  |
| 4.5. Sensibilidad y especificidad. ....                     | 14  |
| 4.6. Valor predictivo negativo y positivo .....             | 17  |
| 4.7. Coeficientes de razón de verosimilitud .....           | 20  |

|                                     |    |
|-------------------------------------|----|
| 5. Discusión.....                   | 23 |
| 6. Limitaciones .....               | 26 |
| 7. Conclusiones .....               | 27 |
| 8. Referencias Bibliográficas ..... | 28 |
| 9. Anexos.....                      | 34 |

## 1. Introducción

Los estudios realizados en base al neurodesarrollo empezaron en 1787 con investigaciones relacionadas al desarrollo en infantes. (1) En términos generales, se puede definir al neurodesarrollo como una serie de capacidades que se adquieren de forma natural y de manera secuencial durante los primeros meses de vida, convirtiéndose en un proceso continuo para los próximos años, llegando a alcanzar hasta en un 90% de madurez dentro de las capacidades cognitivas. (1-4)

El neurodesarrollo conlleva un proceso donde participan miles de genes encargados de dar origen al sistema nervioso periférico y central, influenciado a su vez por factores genéticos y modificaciones del ambiente que ejercen sobre la expresión de los mismos conocido como factores epigenéticos. Este proceso tiene lugar al momento de la concepción y continua durante toda la vida, siendo los primeros mil días luego de la concepción, el período de mayor aceleración en el desarrollo. (5)

Por su parte, el cerebro cumple un papel fundamental dentro de esta etapa, ya que es el encargado de atravesar por distintas fases desde el periodo gestacional hasta finalizar la edad preescolar. Durante los diferentes periodos de formación, las neuronas participan en un proceso donde se originan, forman conexiones y por medio del proceso de mielinización y apoptosis logran su sintonía fina. Posteriormente, ocurre el proceso de maduración para finalmente completar con el desarrollo de cada área específica del mismo cerebro como es el lenguaje, la función sensorial, y la parte cognitiva. (3,4,6,7)

Alrededor de 200 millones de niños que habitan en países en vía de desarrollo no logran un desarrollo adecuado. Entre el 6% y el 9% de los niños menores de seis años experimentan retrasos o trastornos del desarrollo requiriendo de una intervención temprana. Por ejemplo, en Estados Unidos, 1 de cada 6 niños presentan trastornos del desarrollo del habla y lenguaje, retraso mental, parálisis cerebral o autismo. (8) La Organización Mundial de la salud estima por su parte que el 5.1% de la población mundial infantil (0 hasta los 14 años) padece de alguna discapacidad. (9)

La manifestación clínica más conocida dentro de los trastornos en el desarrollo neurológico son las insuficientes habilidades de desarrollo acorde a la edad de cada niño, por esta razón, existen métodos de detección, que ofrecen un diagnóstico temprano para cada trastorno o dificultad, y de esta manera, brindan una intervención oportuna que corrija o atenúe cada alteración para el beneficio del niño, su familia y el entorno que lo rodea. (6,10,11)

Entre las herramientas usadas para la evaluación del neurodesarrollo se encuentra al test de Denver, mismo que fue elaborado por tres especialistas en el área de psicología, entre ellos Frankenburg, Doods y Fordal en 1967 en la Universidad de Colorado, Denver. Es considerado como un test evolutivo, con el principal objetivo de permitir la valoración de capacidades intelectuales, físicas y desarrollo psicomotriz en niños, permitiendo de esta manera la evaluación de la motricidad gruesa

y fina, lenguaje y desarrollo social de cada niño desde el primer mes de vida hasta cumplir los seis años. (12–16)

La prueba cuenta con 125 ítems, que se divide de la siguiente manera: 25 ítems abarcan habilidades sociales y personales, 29 ítem motricidad fina y adaptativa, 39 ítems lenguaje, 32 ítems motricidad gruesa. Es necesario tomar en consideración que solamente se aplicará aquellos ítems que se encuentren del lado izquierdo de la línea, misma que corresponde a la edad del niño. Los materiales a utilizarse suelen ser un bloque de madera de diferentes colores, cubos y botella de plástico, tomándose un tiempo aproximado entre 10 a 20 minutos en un ambiente que resulte agradable tanto para el niño como su familiar. (12,13,17,18)

Ante la posibilidad de que el niño no pueda completar los ítems correspondientes a la edad cronológica, en donde el 90% de los niños de la misma edad si lo pueden realizar, solo se conocerá como un “fallo absoluto”. Por otro lado, si el niño no rebasa la prueba en donde el 75% de niños de su edad si son capaces de realizarlo, se lo considerará como un “fallo relativo”. (12,17)

Para su respectiva interpretación, consideramos como “Normal” aquel niño que no presenta ningún tipo de retraso o máximo un fallo considerado como precaución, para ello se necesitará realizar otra visita para comprobar que se cumpla las respuestas anteriores. “Sospechoso” aquel que presenta entre dos o más errores o uno o más retardos, se necesitará reevaluar después de una o dos semanas para excluir distintos factores como fatiga, enfermedad o miedo. “Alterado” aquel niño que rechaza realizar una o varias actividades que están a la izquierda de línea correspondiente a la edad. La calificación se obtiene dependiendo si las respuestas están acertadas o descartadas del rango de su edad normal. (18,19)

El test de Bayley, por su parte fue publicado en el año de 1969, siendo en 1993 actualizado y adaptado con el nombre de BSID-II; el cual se compone de dos índices como lo es el desarrollo mental y psicomotor, considerada como una herramienta esencial para la evaluación del desarrollo neurológico en niños pequeños. Para el año 2006 se presentó a la tercera edición conocida como Bayley-III, indicado a partir del primer mes hasta los 42 meses de edad. (20–22)

El instrumento evalúa cinco apartados, cada uno de ellos presenta un número de ítems específicos. Empezando por el apartado cognitivo que posee 91 ítems; seguido del lenguaje que se lo categoriza como receptivo 49 ítems, expresivo 48 ítems; en tercer lugar, el motor en donde se encuentra la motricidad fina con 66 ítems y gruesa 72 ítems; en cuarto lugar, encontramos al apartado socioemocional con 35 ítems; y finalmente, el comportamiento adaptativo con 241 ítems. (23,24)

La puntuación se obtiene comparando al niño con un grupo de igual edad en el área cognitiva, lenguaje y motora. Es así como si se obtiene 100 (DE=15) se la considera una puntuación media estandarizada, lo que indica que el niño se encuentra en el percentil 50, es decir, su desarrollo es adecuado para su edad. Si la puntuación es <85 (1 DE inferior de la media) sugiere una deficiencia

leve o un riesgo en el desarrollo, por lo que se recomienda realizar un seguimiento. Una puntuación <70 (2 DE inferior de la media) indica una deficiencia moderada a grave en el desarrollo, lo que requiere de una intervención de equipo especializado. (25,26)

Dado el impacto del desarrollo infantil y la necesidad de disponer de herramientas que resulten precisas para el diagnóstico, la presente revisión sistemática tiene como objetivo evaluar cuál de las escalas resulta más efectiva en la detección temprana de trastornos del neurodesarrollo. La comparación entre ambos test permitirá identificar la herramienta que ofrezca un diagnóstico más preciso, facilitando intervenciones oportunas que mejoren el pronóstico y bienestar de los niños. Teniendo en el presente estudio como pregunta de investigación ¿Cuál es la sensibilidad y especificidad comparativa del Test de Denver y el Test Bayley-III en la evaluación del neurodesarrollo en niños en edades comprendidas entre 1 a 42 meses?

## **2. Objetivos**

### ***2.1. Objetivo general:***

- Evaluar la efectividad del Test de Denver en comparación con el Test Bayley-III para la evaluación del neurodesarrollo infantil en pacientes de 1-42 meses a través de una revisión sistemática.

### ***2.2. Objetivos específicos:***

- Analizar críticamente la calidad y la metodología de los estudios incluidos que investigan la eficacia del Test de Denver y la Escala Bayley-III en la evaluación del neurodesarrollo infantil.
- Comparar la sensibilidad y la especificidad de ambos tests para detectar retrasos en el desarrollo.
- Determinar el valor predictivo positivo y negativo, LR+ y LR- del Test de Denver y la Escala Bayley-III en la detección de trastornos del neurodesarrollo en niños de 1 a 42 meses explorando sus aplicaciones clínicas.

## **3. Métodos**

### ***3.1. Diseño del estudio***

En esta investigación sistemática se empleó un enfoque de revisión cualitativa, analizando diversos estudios aleatorizados que proporcionen datos sobre la sensibilidad y especificidad del Test de Denver y del Test de Bayley III. Se incluyeron únicamente estudios publicados en revistas de prestigio internacional, sin importar el cuartil de clasificación de la revista en las bases de datos científicas indexadas en el ámbito de la salud. El presente estudio permitirá

analizar y comparar los resultados de estudios observacionales, los cuales serán extraídos de bases de datos literarias de alto impacto.

Pregunta PICO:

P: Niños en edades comprendidas entre 1 mes y 42 meses.

I: Evaluación del neurodesarrollo utilizando el Test de Denver.

C: Comparar la eficacia entre el Test de Denver y Bayley-III.

O: Sensibilidad y especificidad de ambos tests.

**3.1.1. Criterios de elegibilidad:**

Entre los criterios de elegibilidad se considerará la estrategia PICO (Paciente, Intervención, Comparación, Resultados), en donde se incluyeron estudios observacionales retrospectivos y prospectivos, que detallaron la efectividad del Test de Denver y el Test de Bayley-III en la evaluación del neurodesarrollo infantil en niños de 1 a 42 meses. Se consideraron artículos originales publicados en inglés y español entre 2014 y 2024 que evaluaron la sensibilidad y especificidad de los tests, la concordancia de resultados y la identificación de retrasos en los dominios del desarrollo (motor, cognitivo, lenguaje, socioemocional). De igual forma publicaciones en su versión final y artículos de paga; se excluyeron aquellos que no se obtuvieron en su formato completo, así como literatura gris (tesis no publicadas, reportes de conferencias no revisados por pares, informes técnicos, etc.) y estudios no relacionados directamente con la efectividad de los tests mencionados.

**3.2. Extracción de datos**

**3.2.1. Fuentes de información:**

Se explorarán diversas bases de datos especializadas en el campo de la salud para realizar una búsqueda exhaustiva de recursos literarios disponibles de libre acceso como de pago, tales como Scielo, MPDI (Publisher of Open Access Journals), Elsevier, Taylor & Francis, NIH (National Institutes of Health), Springer link y Scopus.

La revisión de la literatura se llevó a cabo desde junio de 2024 hasta la fecha actual, con el propósito de identificar y analizar estudios observacionales relevantes, permitiendo una síntesis crítica de la evidencia disponible.

**3.2.2. Estrategia de búsqueda:**

Esta revisión sistemática sigue la Declaración PRISMA-2020, una guía para la presentación de investigaciones que abarca los métodos avanzados para identificar, seleccionar, evaluar y sintetizar estudios. Las bases de datos se utilizarán para mejorar

la búsqueda mediante palabras clave obtenidas de “DeCS/MeSH” y operadores booleanos como “AND, NOT y OR”, además de aplicar filtros para seleccionar los criterios de inclusión y acceder a los textos completos de los artículos.

Se llevó a cabo una búsqueda sistemática de estudios que evaluaron la efectividad del Test de Denver y del Bayley-III en la Evaluación del Neurodesarrollo Infantil. Se utilizaron los términos en inglés: “Denver Developmental Screening Test”, “Bayley III”, y términos relacionados. Con las formulas de búsqueda: “Bayley III and sensitive”, “DDST and II and sensitivity and development”.

### **3.2.3. Selección de estudio:**

Se analizaron los títulos y resúmenes obtenidos a partir de búsquedas electrónicas en las bases de datos para determinar su posible inclusión y proceder a la recuperación de los textos completos. La calidad de cada artículo completo fue evaluada para verificar su elegibilidad y confirmar que cumplía con los criterios de inclusión y exclusión establecidos. Se consideraron elegibles los estudios observacionales que abordaban la efectividad del Test de Bayley y del Test de Denver en la evaluación del neurodesarrollo pediátrico. Se excluyeron estudios de tipo reseña, editorial, carta al editor, ensayo, revisión bibliográfica, revisión sistemática y meta-análisis.

### **3.2.4. Lista de los datos:**

Se considerarán las siguientes variables: primer autor y año de publicación, país del estudio, población estudiada, calidad del estudio, resultados de efectividad, resultados en el desarrollo neuropsicológico infantil (PICO), y fuentes de financiamiento.

### **3.3. Evaluación de la calidad de los estudios**

Se utilizó la herramienta de evaluación de la calidad para estudios observacionales de cohorte y transversales de la "National Heart, Lung and Blood Institute", en donde se incluyen diferentes criterios que deben cumplir los estudios como la pregunta de investigación, la población de estudio, los grupos reclutados de la misma población y criterios de elegibilidad uniformes, justificación del tamaño de la muestra, exposición evaluada antes de la medición de resultados, plazo suficiente para ver un efecto, diferentes niveles de exposición de interés, medidas de resultados, cegamiento de los evaluadores de resultados, análisis estadísticos. (27)

### **3.4. Evaluación de riesgo de sesgos basada en dominios**

Para asegurar la validez de la revisión sistemática sobre la efectividad del Test de Denver y el Test Bayley III en la evaluación del neurodesarrollo infantil, se realizará una evaluación exhaustiva del riesgo de sesgo en los estudios primarios utilizando la herramienta "ROB - 1" para los sesgos de estudios.

### **3.5. Análisis de la información**

#### **3.5.1. Medidas de resumen:**

Las medidas de resumen principales empleadas en este estudio incluyen la efectividad de los instrumentos de evaluación, las limitaciones asociadas y su influencia en la detección precoz de problemas de desarrollo, junto con sus respectivos intervalos de confianza. Estas métricas facilitarán la comparación de la efectividad y la seguridad del Test de Denver y el Test de Bayley-III en la evaluación del neurodesarrollo infantil.

#### **3.5.2. Síntesis de Resultados:**

Los resultados se organizarán en una tabla resumen que condensa las características primordiales de cada artículo selecto, facilitando la comparación de los diversos resultados encontrados en estos estudios para identificar variaciones y similitudes. Estas tablas detallarán información como el autor, el título, el tipo de estudio, el país, la sensibilidad, la especificidad, el valor predictivo negativo, el valor predictivo positivo, sexo, edad y limitaciones.

## **4. Resultados**

### **4.1. Diagrama de flujo**

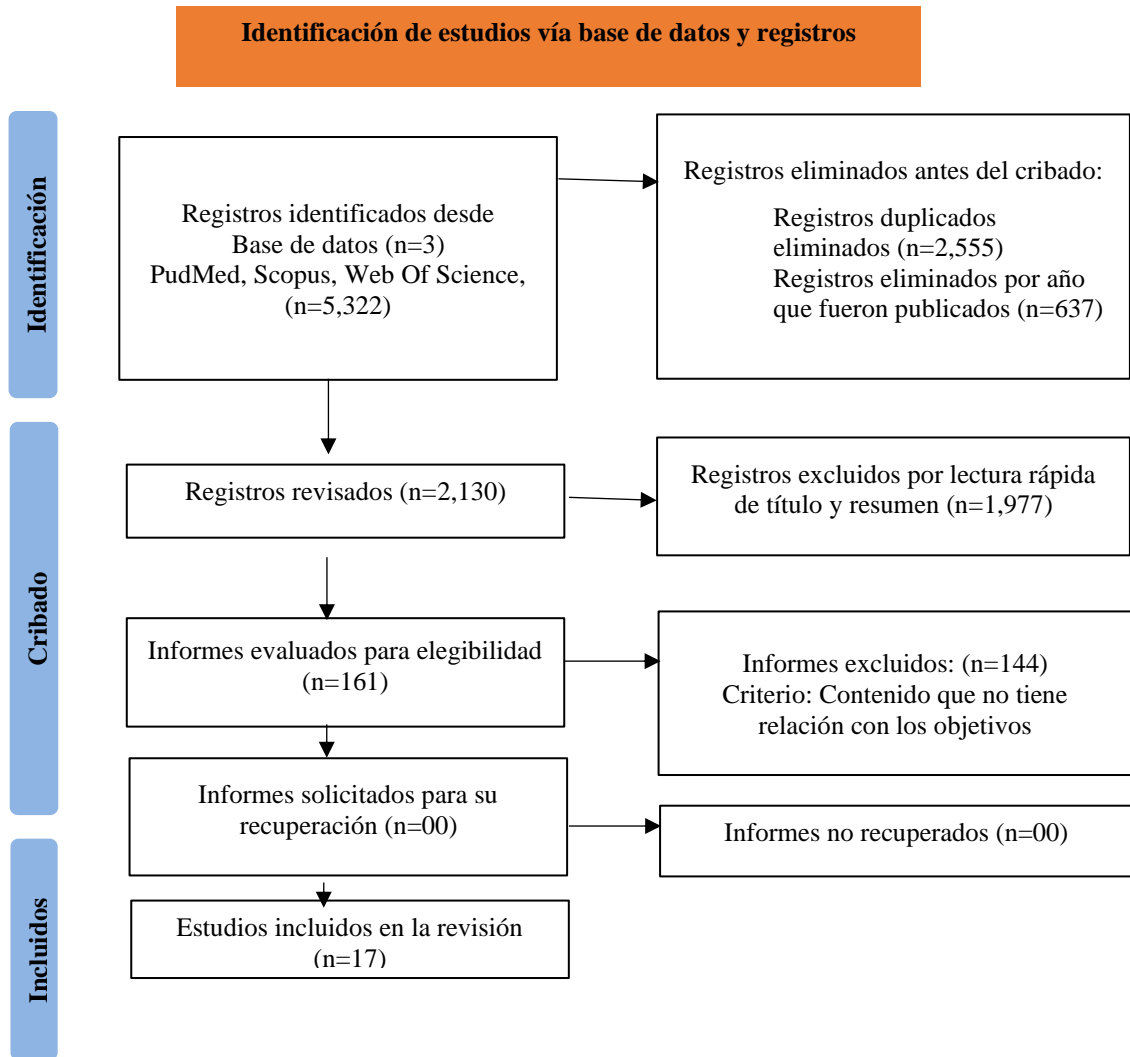
Tras llevar a cabo una búsqueda exhaustiva en las bases de datos literarias, centrada en artículos científicos experimentales de acceso abierto y ciertos artículos de paga, se presenta un resumen en el diagrama de flujo que ilustra las búsquedas realizadas y los resultados obtenidos en la fase inicial del cribado de la presente revisión sistemática.

El flujograma se basa en el modelo de PRISMA en donde se muestra un resumen de los artículos que han sido evaluados. Se empieza con la identificación de registros a partir de tres bases de datos, incluyendo PubMed, Scopus y Web Of Science (n=5,322). Durante esta fase, se eliminan los registros duplicados (n=2,555) y aquellos que no cumplen con los criterios de inclusión basados en el año de publicación (n=637). Posteriormente, en la fase de cribado, los registros revisados (n=2,130) son evaluados rápidamente a través de sus títulos y resúmenes (n=1,977). Aquellos que no están relacionados con los objetivos del estudio son excluidos en esta etapa.

En la fase de elegibilidad, los informes restantes se evalúan más detalladamente para determinar su pertinencia y calidad. Los informes que no cumplen con los criterios específicos son excluidos (n=144). Luego, se solicitan los informes completos (n=00) para una revisión exhaustiva. Si algunos informes no pueden ser recuperados (n=00). Finalmente, en la fase de

inclusión, los estudios que cumplen con todos los criterios de selección y son pertinentes para los objetivos de la revisión sistemática son incluidos en el análisis final (n=17).

**Gráfico 1. Diagrama de Flujo**



*Elaborado por:* Bermúdez D, González J.

## 4.2. Calidad de estudios

**Tabla 1.** Herramienta de evaluación de la calidad para estudios observacionales de cohorte y transversales

| Referencia            | A | B | C | D | E | F | G | H | I | J | K | L | M | N | PUNTOS | CALIDAD |
|-----------------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|--------|---------|
| Kharawala et al. 2023 | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | 10     | MEDIA   |
| Liu et al. 2021       | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | 11     | ALTA    |
| Ryu et al. 2019       | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | 10     | MEDIA   |
| Mannson et al. 2021   | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | 11     | ALTA    |
| Blazek et al. 2024    | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | 11     | ALTA    |
| Wong et al. 2018      | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | 10     | MEDIA   |
| Creighton et al. 2018 | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | 9      | MEDIA   |
| Çelik et al. 2020     | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | 11     | ALTA    |
| Milne et al. 2014     | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | 10     | MEDIA   |
| Fairbairn et al. 2017 | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | 10     | MEDIA   |
| Duncan et al. 2016    | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | 9      | MEDIA   |
| Bode et al. 2014      | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | 11     | ALTA    |
| Johnson, et al. 2021  | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | 12     | ALTA    |
| Kharlukhi et al. 2021 | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | 11     | ALTA    |
| Montedori et al. 2024 | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | 11     | ALTA    |
| Nair et al. 2014      | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | 8      | MEDIA   |
| Coelho et al. 2016    | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | 11     | ALTA    |

*Elaborado por:* Bermúdez D, González J.

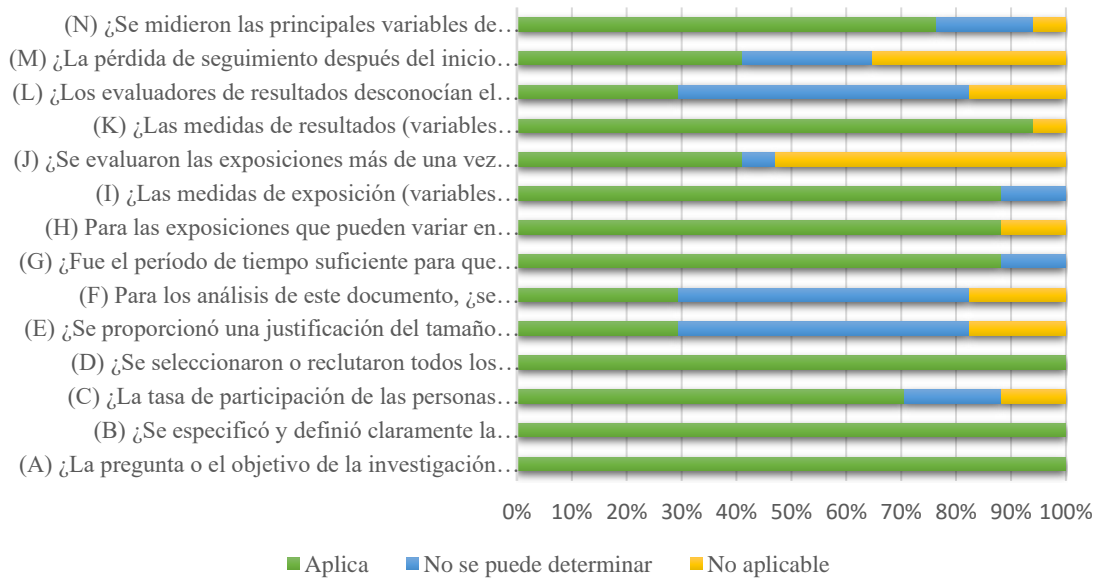
**Interpretación:** Al evaluar la calidad de estudios transversales, se observa que el 53% (n=9) son de alta calidad y el 47% (n=8) son de calidad media. No obstante, su media aritmética es de 10,35, su moda de 11 y la mediana es 11 en relación a la calidad de los estudios. Se observó que varios estudios no cumplieron con ciertos criterios específicos establecidos dentro del análisis. En concreto, 2 estudios no cumplieron la pregunta C, 3 estudios no cumplieron con la pregunta E, 2 estudios no cumplieron con la pregunta H, 9 estudios no cumplieron la pregunta J, 1 estudio no cumplió con la pregunta K, 3 estudios no cumplieron con la pregunta L, 6 estudios no cumplieron con la pregunta M y 1 estudio no cumplió con la pregunta N.

**Tabla 2.** Herramienta de evaluación de la calidad para estudios observacionales de cohorte y transversales

|                               |  | Datos por cantidad |    |    |    |   |   |    |    |    |   |    |   |   |    |
|-------------------------------|--|--------------------|----|----|----|---|---|----|----|----|---|----|---|---|----|
|                               |  | A                  | B  | C  | D  | E | F | G  | H  | I  | J | K  | L | M | N  |
| <b>Aplica</b>                 |  | 17                 | 17 | 12 | 17 | 5 | 5 | 15 | 15 | 15 | 7 | 16 | 5 | 7 | 13 |
| <b>No se puede determinar</b> |  | 0                  | 0  | 3  | 0  | 9 | 9 | 2  | 0  | 2  | 1 | 0  | 9 | 4 | 3  |
| <b>No aplicable</b>           |  | 0                  | 0  | 2  | 0  | 3 | 3 | 0  | 2  | 0  | 9 | 1  | 3 | 6 | 1  |

*Elaborado por:* Bermúdez D, González J.

**Gráfico 2.** Diagrama de flujo de evaluación de la calidad de estudios



*Elaborado por:* Bermúdez D, González J.

**Interpretación:** Al evaluar los datos por cantidad se puede evidenciar que la pregunta A, B y D son aplicables en todos los casos evaluados, la pregunta C es aplicable en 12 estudios, mientras que 3 estudios no se pueden determinar y en 2 estudios no es aplicable. La pregunta E son aplicables 5 estudios, 9 estudios no se pueden determinar y 3 estudios no es aplicable. La pregunta F, G e I son aplicables 15 estudios, 2 estudios no se pueden determinar. La pregunta H 15 estudios son aplicables y 2 estudios no es aplicable. La pregunta J son aplicables 7 estudios, 1 estudio no se puede determinar y 9 estudios no es aplicable. La pregunta K son aplicables 16 estudios y 1 estudio no es aplicable. La pregunta L son aplicables 5 estudios, 9 estudios no se pueden determinar y 3 estudios no es aplicable. La pregunta M son aplicables 7 estudios, 4

estudios no se pueden determinar y 6 estudios no es aplicable. La pregunta N son aplicables 13 estudios, 3 estudios no se pueden determinar y 1 estudio no es aplicable.

### 4.3. Riesgo de sesgos basada en dominios

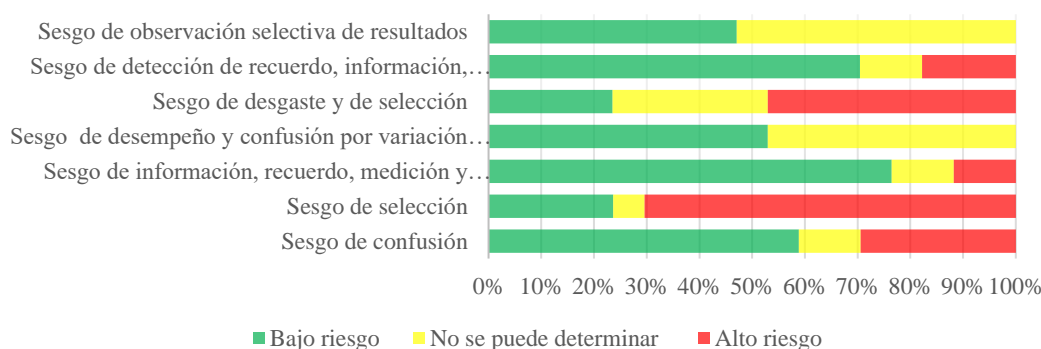
Tabla 3. Herramienta de evaluación de riesgo de sesgos

| Datos en porcentajes   |                    |                    |   |   |                                  |   |  |
|------------------------|--------------------|--------------------|---|---|----------------------------------|---|--|
|                        | Sesgo de confusión | Sesgo de selección | Sesgo de información, recuerdo, medición y del observador | Sesgo de desempeño y confusión por variación temporal | Sesgo de desgaste y de selección | Sesgo de detección de recuerdo, información, clasificación errónea, de observador y de medición | Sesgo de observación selectiva de resultados |
| Bajo riesgo            | 59                 | 24                 | 76  | 53  | 24                               | 70  | 47   |
| No se puede determinar | 12                 | 6                  | 12  | 47  | 29                               | 12  | 53   |
| Alto riesgo            | 29                 | 70                 | 12  | 0   | 47                               | 18  | 0  |

Elaborado por: Bermúdez D, González J.

**Interpretación:** Al evaluar el riesgo de sesgos, se observa de manera general que el sesgo de confusión presenta un predominio de bajo riesgo del 59%, mientras que el 29% restante es de alto riesgo. Por otro lado, el sesgo de selección muestra una preocupación significativa, ya que el 70% es de alto riesgo. El sesgo de desempeño y confusión por variación temporal por su parte no es de alto riesgo, sin embargo, el 47% no se puede determinar generando de esta manera cierta incertidumbre. De igual manera, el sesgo de desgaste y selección, el 53% es de bajo riesgo, mientras que el 47% restante es de alto riesgo, llegando a comprometer la validez. Asimismo, el sesgo de detección de recuerdo, información, clasificación errónea del observador y de medición es en su mayoría de bajo riesgo con un 70%. Finalmente, el sesgo de observación selectiva de resultados muestra un 47% de bajo riesgo, mientras que el 53% no se lo puede determinar, dejando dudas en dicho apartado.

Gráfico 3. Diagrama de flujo de evaluación de la calidad de estudios



Elaborado por: Bermúdez D, González J.

**Interpretación:** Se observa la mayor parte de sesgos están catalogadas como bajo riesgo, en especial el sesgo de información, recuerdo, medición y del observador con un 76% y el sesgo de detección de recuerdo, información, clasificación errónea, de observador y de medición con un 70%, lo que no sugiere una preocupación. No obstante, se observa que el sesgo de confusión 29%, el sesgo de selección 70% y el sesgo de desgaste y de selección 47% son de alto riesgo, sugiriendo que podrían comprometer la validez de los estudios. Por último, aquellos estudios que no se pudieron determinar pertenecen al sesgo de desempeño y confusión por variación temporal 47% y al sesgo de observación selectiva de resultados 53%, lo que puede reflejar cierta incertidumbre para la correcta evaluación del nivel de riesgo.

#### 4.4. Características de los estudios

Las características de estudios se presentan en la Tabla 4. Cada artículo ha sido revisado y seleccionado por dos revisores que han estado de acuerdo con cada uno de ellos.

**Tabla 4.** Características de los estudios incluidos.

| Autor                                 | Título   | Tipo de estudio                  | País          | Edad   | Sexo                  | Limitaciones  |
|---------------------------------------|--|----------------------------------|---------------|--|-----------------------|---|
| <b>Montgomery, et al (2023).</b> (28) | Valor predictivo del índice motor Bayley-III para posteriores dificultades motoras en niños nacidos extremadamente prematuros.                                       | Cohorte observacional            | Suecia        | Bebés prematuros (<27 semanas de gestación) a los 2,5 años de edad corregida | Masculino<br>Femenino | Muestra limitada, riesgo de sesgos, versión desactualizada de Bayley, y exclusión de niños con discapacidades neurosensoriales. |
| <b>Liu, et al (2021).</b> (29)        | Validez predictiva de las puntuaciones cognitivas del Bayley-III a los 6 meses para los resultados cognitivos a los 24 meses en lactantes de muy bajo peso al nacer. | Estudio retrospectivo            | Taiwán        | Bebés a los 6 y 24 meses de edad de concepción                               | Masculino<br>Femenino | Estudio retrospectivo, sesgo de seguimiento, variabilidad y falta de estandarización limitan la generalización                  |
| <b>Ryu, et al (2019).</b> (30)        | Validez y fiabilidad del DDST II y el Bayley III en niños con retraso en el desarrollo del lenguaje.   | Estudio retrospectivo            | Corea del Sur | Niños de 12 a los 36 meses de edad   | Masculino<br>Femenino | Muestra limitada, evaluación única, sesión de seguimiento y falta de estandarización limitan validez.                           |
| <b>Mansson, et al (2021).</b> (31)    | La capacidad de las puntuaciones Bayley-III para predecir la inteligencia posterior en niños nacidos extremadamente prematuros.                                      | Estudio prospectivo longitudinal | Suecia        | Niños <27 semanas de gestación a los 2,5 años de edad                        | Masculino<br>Femenino | Sesgo de abandono, falta de normas locales y subestimación limitan validez externa.   |

|                                      |   |  |            |   |                       |   |
|--------------------------------------|---|--|------------|---|-----------------------|---|
| <b>Blazek, et al (2024).</b> (32)    | Sensibilidad y especificidad de la evaluación visual neonatal para predecir los resultados motores y cognitivos en recién nacidos muy prematuros.   | Estudio de cohorte observacional prospectivo | Australia  | Bebés a los 12 meses de edad corregida                    | Masculino<br>Femenino | Falta de datos del NVA a 3 meses y evaluación única a los 12 meses.                                       |
| <b>Wong, et al (2018).</b> (33)      | Validez de los resultados del neurodesarrollo de niños nacidos muy prematuros evaluados durante el seguimiento clínico rutinario en Inglaterra.   | Estudio transversal                          | Inglaterra | Niños de 20 y 28 meses de edad postérmino.                | Masculino<br>Femenino | Sesgo de selección, falta de aleatorización, muestra insuficiente y exclusión por idioma limitan validez. |
| <b>Creighton, et al (2018).</b> (34) | Establecimiento de puntuaciones de corte Bayley-III a los 21 meses para predecir puntuaciones bajas de CI a los 3 años de edad en una cohorte de prematuros.  | Estudio de cohorte prospectivo               | Canadá     | Niños a los 21 meses de edad corregida.                   | Masculino<br>Femenino | Falta de grupo control, datos incompletos, variabilidad en evaluaciones y seguimiento limitado.           |
| <b>Celik, et al (2020).</b> (21)     | ¿Qué valores de corte del Bayley-III deben utilizarse en los distintos niveles de desarrollo?   | Estudio prospectivo                          | Turquía    | Niños en una edad comprendida entre 6 a 42 meses de edad. | Masculino<br>Femenino | Cortes variables, baja sensibilidad de Bayley -III limita detección temprana de retrasos.                 |
| <b>Milne, et al (2014).</b> (35)     | La puntuación alternativa del Bayley-III mejora la predicción del rendimiento en las Escalas de Desarrollo Mental de Griffiths antes del ingreso escolar en preescolares con problemas de desarrollo. | Estudio retrospectivo.                       | Australia  | Niños menores de 3.5 años de edad.                        | Masculino<br>Femenino | NR  |
| <b>Fairbairn, et al (2017).</b> (36) | Predicción de resultados a tres años utilizando el Bayley-III para lactantes australianos quirúrgicos, cardíacos y sanos al año de edad.  | Estudio de cohorte prospectivo.              | Australia  | Niños en edad comprendida de 1 a 3 años de edad.          | Masculino<br>Femenino | Tamaños desiguales, datos parentales retrospectivos y falta de detalles de intervenciones.                |
| <b>Duncan, et al (2016).</b> (37)    | ¿Sobreestiman los umbrales de Bayley III recomendados actualmente el  | Estudio retrospectivo                        | Australia  | Niños <27 semanas de gestación a los                      | Masculino<br>Femenino | Diseño retrospectivo, sin grupo control, opiniones no   |

|                                      |   |  |                 |  |                       |   |
|--------------------------------------|---|--|-----------------|--|-----------------------|---|
|                                      | deterioro motor en bebés nacidos con menos de 27 semanas de gestación?  |  |                 | 18 a 22 meses de edad.   |                       | validadas y métodos no estandarizados.  |
| <b>Bode, et al (2014).</b> (38)      | Validez predictiva del Bayley, tercera edición a los 2 años para el cociente de inteligencia a los 4 años en lactantes prematuros.  | Estudio prospectivo  | Estados Unidos. | Niños $\leq$ 30 semanas de gestación a los 2 años de edad.           | Masculino<br>Femenino | NR  |
| <b>Johnson, et al (2014).</b> (39)   | Uso del Bayley-III para evaluar el retraso del neurodesarrollo: ¿qué punto de corte debe utilizarse?  | Estudio retrospectivo.                                       | Inglaterra      | Niños $<$ 27 semanas de gestación a los 2-3 años de edad corregida.  | Masculino<br>Femenino | NR  |
| <b>Kharlukhi, et al (2021).</b> (40) | Eficacia de las herramientas de cribado clínico a pie de cama para predecir el retraso del desarrollo neurológico a corto plazo entre los prematuros de muy bajo peso al nacer. | Estudio observacional prospectivo                            | India           | Niños con una edad gestacional media de 30,26 ( $\pm$ 2,21) semanas. | Masculino<br>Femenino | Estudio unicéntrico, falta de resonancia magnética, pérdida de sujetos y evaluación limitada.       |
| <b>Montedori, et al (2024).</b> (41) | Vigilancia del desarrollo infantil durante los primeros mil días: Investigación de la relación entre el instrumento de vigilancia del desarrollo y las escalas estandarizadas.  | Estudio longitudinal prospectivo con un enfoque cuantitativo | Brasil          | Niños hasta 36 meses de edad   | Masculino<br>Femenino | Tamaño muestral reducido, sesgos por autoinformes y alteraciones por adaptaciones pandémicas.       |
| <b>Nair, et al (2014).</b> (42)      | CDC Kerala 2: Paquete de intervención del desarrollo para bebés $<$ 1800 g: resultado a los 6 meses con DASII   | Estudio de cohorte prospectivo                               | India           | Niños a los 4 meses de edad.   | Masculino<br>Femenino | NR. (No Reporta)  |
| <b>Coelho, et al (2016).</b> (43)    | Desarrollo infantil en atención primaria: una propuesta de vigilancia.  | Estudio transversal.   | Brasil          | Niños en la edad comprendida de 0 a 36 meses de edad.                | Masculino<br>Femenino | Muestra localizada, pérdidas en evaluación limitan la generalización de resultados epidemiológicos. |

**Elaborado por:** Bermúdez D, González J.

**Interpretación:** De acuerdo a los artículos que han sido revisados se puede observar que 1 estudio es observacional, 2 son estudios transversales, 5 estudios retrospectivos, y 9 estudios

prospectivos. Las edades en las que se aplicaron los dos test son entre 0 a 3.5 años de edad. Sin embargo, ambos instrumentos comparten limitaciones, como el sesgo de selección, la falta de estandarización en algunas aplicaciones y la variabilidad en las evaluaciones, lo que puede afectar la validez externa y la capacidad predictiva de los resultados. Además, muchos estudios presentan muestras pequeñas o limitadas y seguimiento insuficiente, lo que limita la generalización de los hallazgos.

#### 4.5. Sensibilidad y especificidad.

*Tabla 5. Sensibilidad y especificidad del test de Bayley-III*

| Referencia                            | Tipo                                    | Bayley-III (%) |
|---------------------------------------|---|----------------|
| <b>Montgomery, et al (2023).</b> (28) | Sensibilidad del índice motor           | 26,3%          |
|                                       | Especificidad del índice motor          | 92,9%          |
| <b>Liu, et al (2021).</b> (29)        | Sensibilidad cognitiva <70              | 28.7%.         |
|                                       | Especificidad cognitiva <70             | 97.1%          |
|                                       | Sensibilidad cognitiva <85              | 22,5%          |
|                                       | Especificidad cognitiva <85             | 92,5%          |
| <b>Ryu, et al (2019).</b> (30)        | Sensibilidad del lenguaje               | 93.1%          |
|                                       | Especificidad del lenguaje              | 83.3%          |
| <b>Mansson, et al (2021).</b> (31)    | Sensibilidad intelectual <70            | 18%            |
|                                       | Especificidad intelectual <70           | -              |
|                                       | Sensibilidad intelectual <85            | 44%            |
|                                       | Especificidad intelectual <85           | -              |
| <b>Blazek, et al (2024).</b> (32)     | Sensibilidad cognitiva                  | 78%            |
|                                       | Especificidad cognitiva                 | 27%            |
|                                       | Sensibilidad motora                     | 76%            |
|                                       | Especificidad motora                    | 27%            |
| <b>Wong, et al (2018).</b> (33)       | Sensibilidad cognitiva                  | 69.7%          |
|                                       | Especificidad cognitiva                 | 83.9%          |
|                                       | Sensibilidad comunicativa               | 53.2%          |
|                                       | Especificidad comunicativa              | 90.8%          |
|                                       | Sensibilidad motora                     | 61.9%          |
|                                       | Especificidad motora                    | 99.4%          |
| <b>Creighton, et al (2018).</b> (34)  | Sensibilidad cognitiva <80              | 77%            |
|                                       | Especificidad cognitiva <80             | 90%            |
|                                       | Sensibilidad del lenguaje <80           | 76%            |
|                                       | Especificidad del lenguaje <80          | 72%            |
|                                       | Sensibilidad cognitiva-lingüística <83  | 76%            |
|                                       | Especificidad cognitiva-lingüística <83 | 78%            |

|                                      |   |       |
|--------------------------------------|---|-------|
| <b>Celik, et al (2020).</b> (21)     | Sensibilidad del compuesto cognitivo      | 86.4% |
|                                      | Especificidad del compuesto cognitivo     | 94.1% |
|                                      | Sensibilidad del lenguaje cognitivo       | 86.4% |
|                                      | Especificidad del lenguaje cognitivo      | 91.1% |
| <b>Milne, et al (2014).</b> (35)     | Sensibilidad retraso leve DSM – IV        | 87%   |
|                                      | Especificidad retraso leve DSM – IV       | 49%   |
|                                      | Sensibilidad retraso moderado DSM – IV    | 76%   |
|                                      | Especificidad retraso moderado DSM – IV   | 100%  |
|                                      | Sensibilidad retraso severo DSM – IV      | 63%   |
|                                      | Especificidad retraso severo DSM – IV     | 96%   |
| <b>Fairbairn, et al (2017).</b> (36) | Sensibilidad del índice cognitivo         | 67%   |
|                                      | Especificidad del índice cognitivo        | 99%   |
|                                      | Sensibilidad del lenguaje receptivo       | 57%   |
|                                      | Especificidad del lenguaje receptivo      | 90%   |
|                                      | Sensibilidad del lenguaje expresivo       | 71%   |
|                                      | Especificidad del lenguaje expresivo      | 77%   |
|                                      | Sensibilidad del índice motor fino        | 63%   |
|                                      | Especificidad del índice motor fino       | 95%   |
|                                      | Sensibilidad del índice motor grueso      | 89%   |
|                                      | Especificidad del índice motor grueso     | 69%   |
| <b>Duncan, et al (2016).</b> (37)    | Sensibilidad índice motor grueso <70      | 88%   |
|                                      | Especificidad índice motor grueso <70     | 94%   |
|                                      | Sensibilidad índice motor fino <70        | 100%  |
|                                      | Especificidad índice motor fino <70       | 93%   |
| <b>Bode, et al (2014).</b> (38)      | Sensibilidad del índice cognitivo         | 85%   |
|                                      | Especificidad del índice cognitivo        | 87%   |
|                                      | Sensibilidad del índice del lenguaje      | 76%   |
|                                      | Especificidad del índice del lenguaje     | 84%   |
| <b>Johnson, et al (2014).</b> (39)   | Sensibilidad del índice cognitivo <85     | 89%   |
|                                      | Especificidad del índice cognitivo <85    | 98%   |
|                                      | Sensibilidad del índice del lenguaje <85  | 100%  |
|                                      | Especificidad del índice del lenguaje <85 | 96%   |
|                                      | Sensibilidad del índice cognitivo <70     | 42%   |
|                                      | Especificidad del índice cognitivo <70    | 100%  |
|                                      | Sensibilidad del índice del lenguaje <70  | 58%   |
|                                      | Especificidad del índice del lenguaje <70 | 99%   |

*Elaborado por:* Bermúdez D, González J.

**Interpretación:** Según los datos expuestos en la tabla 5 se puede determinar que la sensibilidad del índice motor parte desde 26.3% hasta el 76%, la especificidad de 27% hasta 99.4%; el índice motor fino presenta una sensibilidad del 63% al 100%, una especificidad del 93% al 95%; el índice motor grueso del 88% al 89%, una especificidad del 69% al 94%; el índice cognitivo con un coeficiente intelectual del <70 hasta el <85 presenta una sensibilidad del 22.5% hasta el 89%, de la misma forma la especificidad de 27% hasta 100%; el índice del lenguaje con un coeficiente intelectual del <70 hasta el <85 presenta una sensibilidad del 58% al 100%, una especificidad de 72% al 99%; el índice cognitivo lingüístico con un coeficiente intelectual <83 presenta una sensibilidad del 76% hasta el 86.4%, especificidad de 78% hasta el 91.1%; la sensibilidad del lenguaje receptivo es del 57%, la especificidad del 90%; la sensibilidad del lenguaje expresivo es de 71%, la especificidad del 77%; la sensibilidad retraso leve según el DSM – IV es de 87%, la especificidad de 49%; la sensibilidad retraso moderado según DSM – IV es de 76%, la especificidad del 100%; la sensibilidad retraso severo DSM – IV es de 63%, la especificidad retraso severo DSM – IV es de 96%; el índice intelectual con un coeficiente que varía del <70 al <85 la sensibilidad varía del 18% al 44%; la sensibilidad comunicativa es del 53.2% y la especificidad del 90.8%.

**Tabla 6.** Sensibilidad y especificidad del test de Denver.

| Referencia                           | Tipo                                  | Denver (%) |
|--------------------------------------|---------------------------------------|------------|
| <b>Ryu, et al (2019).</b> (30)       | Sensibilidad del lenguaje             | 93.3%      |
|                                      | Especificidad del lenguaje            | 60%        |
| <b>Kharlukhi, et al (2021).</b> (40) | Sensibilidad de la motricidad gruesa  | 86.6%      |
|                                      | Especificidad de la motricidad gruesa | 24.1%      |
| <b>Montedori, et al (2024).</b> (41) | Sensibilidad d+el retraso probable    | 70%        |
|                                      | Especificidad del retraso probable    | 56%        |
|                                      | Sensibilidad del estado de alerta     | 57%        |
|                                      | Especificidad del estado de alerta    | 70%        |
|                                      | Sensibilidad del desarrollo normal    | 21%        |
|                                      | Especificidad del desarrollo normal   | 74%        |
| <b>Nair, et al (2014).</b> (42)      | Sensibilidad de la motricidad gruesa  | 65.16%     |
|                                      | Especificidad de la motricidad gruesa | 98.59%     |
| <b>Coelho, et al (2016).</b> (43)    | Sensibilidad del retraso probable     | 70%        |
|                                      | Especificidad del retraso probable    | 56%        |
|                                      | Sensibilidad del estado de alerta     | 57%        |
|                                      | Especificidad del estado de alerta    | 70%        |

*Elaborado por:* Bermúdez D, González J

**Interpretación:** De acuerdo a la tabla 6 se puede establecer que la sensibilidad de lenguaje 93.3% y la especificidad de 60%; sensibilidad de la motricidad gruesa varía desde 65.16% hasta el 86,6%, mientras que su especificidad va desde 24.1% hasta 98.59%; sensibilidad del retraso probable 70% y su especificidad 56%; sensibilidad del estado de alerta 57% y especificidad 70%; sensibilidad del desarrollo normal 21% y especificidad 74%.

#### 4.6. Valor predictivo negativo y positivo

**Tabla 7.** Valor predictivo positivo y valor predictivo negativo del test de Bayley-III

| Referencia                            | Tipo                          | Bayley-III (%) |
|---------------------------------------|-------------------------------|----------------|
| <b>Montgomery, et al (2023).</b> (28) | VPP del índice motor          | 48.4%          |
|                                       | VPN del índice motor          | 83.3%          |
| <b>Liu, et al (2021).</b> (29)        | VPP cognitiva <70             | 27.4%          |
|                                       | VPN cognitiva <70             | 97.3%          |
|                                       | VPP cognitiva <85             | 33.6%          |
|                                       | VPN cognitiva <85             | 87.7%          |
| <b>Ryu, et al (2019).</b> (30)        | VPP del lenguaje              | 96.4%          |
|                                       | VPN del lenguaje              | 71.4%          |
| <b>Mansson, et al (2021).</b> (31)    | VPP intelectual <70           | -              |
|                                       | VPN intelectual <70           | -              |
|                                       | VPP intelectual <85           | -              |
|                                       | VPN intelectual <85           | -              |
| <b>Blazek, et al (2024).</b> (32)     | VPP cognitiva                 | -              |
|                                       | VPN cognitiva                 | -              |
|                                       | VPP motora                    | -              |
|                                       | VPN motora                    | -              |
| <b>Wong, et al (2018).</b> (33)       | VPP cognitiva                 | 47.9%          |
|                                       | VPN cognitiva                 | 92.9%          |
|                                       | VPP comunicativa              | 75%            |
|                                       | VPN comunicativa              | 79%            |
|                                       | VPP motora                    | 92.9%          |
|                                       | VPN motora                    | 95.3%          |
| <b>Creighton, et al (2018).</b> (34)  | VPP cognitiva <80             | 30%            |
|                                       | VPN cognitiva <80             | 90%            |
|                                       | VPP del lenguaje <80          | 12%            |
|                                       | VPN del lenguaje <80          | 98%            |
|                                       | VPP cognitiva-lingüística <83 | 15%            |
|                                       | VPN cognitiva-lingüística <83 | 99%            |

|                                      |                                 |      |
|--------------------------------------|---------------------------------|------|
| <b>Celik, et al (2020).</b> (21)     | VPP del compuesto cognitivo     | -    |
|                                      | VPN del compuesto cognitivo     | -    |
|                                      | VPP del lenguaje cognitivo      | -    |
|                                      | VPN del lenguaje cognitivo      | -    |
| <b>Milne, et al (2014).</b> (35)     | VPP retraso leve DSM – IV       | -    |
|                                      | VPN retraso leve DSM – IV       | -    |
|                                      | VPP retraso moderado DSM – IV   | -    |
|                                      | VPN retraso moderado DSM – IV   | -    |
|                                      | VPP retraso severo DSM – IV     | -    |
|                                      | VPN retraso severo DSM – IV     | -    |
| <b>Fairbairn, et al (2017).</b> (36) | VPP del índice cognitivo        | 86%  |
|                                      | VPN del índice cognitivo        | 97%  |
|                                      | VPP del lenguaje receptivo      | 47%  |
|                                      | VPN del lenguaje receptivo      | 93%  |
|                                      | VPP del lenguaje expresivo      | 38%  |
|                                      | VPN del lenguaje expresivo      | 93%  |
|                                      | VPP del índice motor fino       | 50%  |
|                                      | VPN del índice motor fino       | 97%  |
|                                      | VPP del índice motor grueso     | 38%  |
|                                      | VPN del índice motor grueso     | 97%  |
| <b>Duncan, et al (2016).</b> (37)    | VPP índice motor grueso <70     | 58%  |
|                                      | VPN índice motor grueso <70     | 99%  |
|                                      | VPP índice motor fino <70       | 38%  |
|                                      | VPN índice motor fino <70       | 100% |
| <b>Bode, et al (2014).</b> (38)      | VPP del índice cognitivo        | 64%  |
|                                      | VPN del índice cognitivo        | 95%  |
|                                      | VPP del índice del lenguaje     | 57%  |
|                                      | VPN del índice del lenguaje     | 93%  |
| <b>Johnson, et al (2014).</b> (39)   | VPP del índice cognitivo <85    | 85%  |
|                                      | VPN del índice cognitivo <85    | 99%  |
|                                      | VPP del índice del lenguaje <85 | 73%  |
|                                      | VPN del índice del lenguaje <85 | 100% |
|                                      | VPP del índice cognitivo <70    | 100% |
|                                      | VPN del índice cognitivo <70    | 94%  |
|                                      | VPP del índice del lenguaje <70 | 85%  |
|                                      | VPN del índice del lenguaje <70 | 95%  |

*Elaborado por:* Bermúdez D, González J.

**Interpretación:** Se según lo indicado en la tabla 7 es posible concluir que el VPP del índice motor parte desde 48.4% hasta el 92.9%, el VPN de 83.3% hasta 95.3%; el índice motor fino presenta un VPP del 38% al 50%, un VPN del 97% al 100%; el índice motor grueso del 38% al 58%, un VPN del 97% al 99%; el índice cognitivo con un coeficiente intelectual del <70 hasta el <85 presenta un VPP del 27.4% hasta el 100%, de la mismo forma un VPN de 87.7% hasta 99%; el índice del lenguaje con un coeficiente intelectual del <70 hasta el <85 presenta un VPP del 12% al 96.4%, un VPN de 71.4% al 100%; el índice cognitivo lingüístico con un coeficiente intelectual <83 presenta un VPP del 15%, VPN de 99%; el VPP del lenguaje receptivo es del 47%, la VPN del 93%; el VPP del lenguaje expresivo es de 38%, el VPN del 93%; el VPP comunicativa es del 75% y el VPN del 79%.

**Tabla 8.** Valor predictivo positivo y valor predictivo negativo del test de Denver.

| Referencia                           | Tipo                        | Denver (%) |
|--------------------------------------|-----------------------------|------------|
| <b>Ryu, et al (2019).</b> (30)       | VPP del lenguaje            | 93.3%      |
|                                      | VPN del lenguaje            | 60%        |
| <b>Kharlukhi, et al (2021).</b> (40) | VPP de la motricidad gruesa | 72.5%      |
|                                      | VPN de la motricidad gruesa | 43.7%      |
| <b>Montedori, et al (2024).</b> (41) | VPP del retraso probable    | -          |
|                                      | VPN del retraso probable    | -          |
|                                      | VPP del estado de alerta    | -          |
|                                      | VPN del estado de alerta    | -          |
|                                      | VPP del desarrollo normal   | -          |
|                                      | VPN del desarrollo normal   | -          |
| <b>Nair, et al, 2014.</b> (42)       | VPP de la motricidad gruesa | 97.12%     |
|                                      | VPN de la motricidad gruesa | 79.55%     |
| <b>Coelho, et al (2016).</b> (43)    | VPP del retraso probable    | 70%        |
|                                      | VPN del retraso probable    | 56%        |
|                                      | VPP del estado de alerta    | 57%        |
|                                      | VPN del estado de alerta    | 70%        |

**Elaborado por:** Bermúdez D, González J.

**Interpretación:** De acuerdo con lo indicado en la tabla 8 se puede establecer que el VPP de lenguaje 93.3% y un VPN del 60%; el VPP de la motricidad gruesa varía desde el 72.5% hasta el 97.12%, el VPN desde el 43.7% hasta el 79.55%; sobre el retraso probable el VPP del 70% y el VPN del 56%; del estado de alerta un VPP del 57% y un VPN del 70%.

#### 4.7. Coeficientes de razón de verosimilitud

Tabla 9. LR+ y LR- del test de Bayley-III

| Referencias                           | Tipo                          | Bayley-III | Test      |
|---------------------------------------|-------------------------------|------------|-----------|
| <b>Montgomery, et al (2023).</b> (28) | LR+ del índice motor          | 3.7        | Regular   |
|                                       | LR- del índice motor          | 0.79       | Malo      |
| <b>Liu, et al (2021).</b> (29)        | LR+ cognitiva <70             | 9.9        | Bueno     |
|                                       | LR- cognitiva <70             | 0.73       | Malo      |
|                                       | LR+ cognitiva <85             | 3          | Regular   |
|                                       | LR- cognitiva <85             | 0.84       | Malo      |
| <b>Ryu, et al (2019).</b> (30)        | LR+ del lenguaje              | 5.6        | Bueno     |
|                                       | LR- del lenguaje              | 0.08       | Excelente |
| <b>Mansson, et al (2021).</b> (31)    | LR+ intelectual <70           | -          | -         |
|                                       | LR- intelectual <70           | -          | -         |
|                                       | LR+ intelectual <85           | -          | -         |
|                                       | LR- intelectual <85           | -          | -         |
| <b>Blazek, et al (2024).</b> (32)     | LR+ cognitiva                 | 1.1        | Malo      |
|                                       | LR- cognitiva                 | 0.8        | Malo      |
|                                       | LR+ motora                    | 1.04       | Malo      |
|                                       | LR- motora                    | 0.89       | Malo      |
| <b>Wong, et al (2018).</b> (33)       | LR+ cognitiva                 | 4.3        | Regular   |
|                                       | LR- cognitiva                 | 0.36       | Regular   |
|                                       | LR+ comunicativa              | 5.8        | Bueno     |
|                                       | LR- comunicativa              | 0.51       | Malo      |
|                                       | LR+ motora                    | 103.2      | Excelente |
|                                       | LR- motora                    | 0.38       | Regular   |
| <b>Creighton, et al (2018).</b> (34)  | LR+ cognitiva <80             | 7.7        | Bueno     |
|                                       | LR- cognitiva <80             | 0.25       | Regular   |
|                                       | LR+ del lenguaje <80          | 2.7        | Regular   |
|                                       | LR- del lenguaje <80          | 0.33       | Regular   |
|                                       | LR+ cognitiva-lingüística <83 | 3.45       | Regular   |
|                                       | LR- cognitiva-lingüística <83 | 0.3        | Regular   |
| <b>Celik, et al (2020).</b> (21)      | LR+ del compuesto cognitivo   | 14.6       | Excelente |
|                                       | LR- del compuesto cognitivo   | 0.14       | Bueno     |
|                                       | LR+ del lenguaje cognitivo    | 9.7        | Bueno     |
|                                       | LR- del lenguaje cognitivo    | 0.15       | Bueno     |
|                                       | LR+ retraso leve DSM – IV     | 1.7        | Malo      |
|                                       | LR- retraso leve DSM – IV     | 0.26       | Regular   |

|                                      |                                 |       |           |
|--------------------------------------|---------------------------------|-------|-----------|
| <b>Milne, et al (2014).</b> (35)     | LR+ retraso moderado DSM – IV   | 0.84  | Malo      |
|                                      | LR- retraso moderado DSM – IV   | 2.4   | Malo      |
|                                      | LR+ retraso severo DSM – IV     | 15.75 | Excelente |
|                                      | LR- retraso severo DSM – IV     | 0.38  | Regular   |
| <b>Fairbairn, et al (2017).</b> (36) | LR+ del índice cognitivo        | 67    | Excelente |
|                                      | LR- del índice cognitivo        | 0.33  | Regular   |
|                                      | LR+ del lenguaje receptivo      | 5.7   | Bueno     |
|                                      | LR- del lenguaje receptivo      | 0.48  | Regular   |
|                                      | LR+ del lenguaje expresivo      | 3.1   | Regular   |
|                                      | LR- del lenguaje expresivo      | 0.38  | Regular   |
|                                      | LR+ del índice motor fino       | 12.6  | Excelente |
|                                      | LR- del índice motor fino       | 0.39  | Regular   |
|                                      | LR+ del índice motor grueso     | 2.9   | Regular   |
|                                      | LR- del índice motor grueso     | 0.16  | Bueno     |
| <b>Duncan, et al (2016).</b> (37)    | LR+ índice motor grueso <70     | 14.7  | Excelente |
|                                      | LR- índice motor grueso <70     | 0.13  | Bueno     |
|                                      | LR+ índice motor fino <70       | 1.4   | Malo      |
|                                      | LR- índice motor fino <70       | 0.97  | Malo      |
| <b>Bode, et al (2014).</b> (38)      | LR+ del índice cognitivo        | 6.5   | Bueno     |
|                                      | LR- del índice cognitivo        | 0.17  | Bueno     |
|                                      | LR+ del índice del lenguaje     | 4.7   | Regular   |
|                                      | LR- del índice del lenguaje     | 0.28  | Regular   |
| <b>Johnson, et al (2014).</b> (39)   | LR+ del índice cognitivo <85    | 44.5  | Excelente |
|                                      | LR- del índice cognitivo <85    | 0.11  | Bueno     |
|                                      | LR+ del índice del lenguaje <85 | 2.5   | Regular   |
|                                      | LR- del índice del lenguaje <85 | 0.94  | Malo      |
|                                      | LR+ del índice cognitivo <70    | 0.47  | Malo      |
|                                      | LR- del índice cognitivo <70    | 5.8   | Malo      |
|                                      | LR+ del índice del lenguaje <70 | 58    | Excelente |
|                                      | LR- del índice del lenguaje <70 | 0.42  | Regular   |

*Elaborado por:* Bermúdez D, González J.

**Interpretación:** En base a los datos mostrados en la tabla 9 se puede establecer que el LR+ del índice motor varía desde 1.04 hasta 103.2, LR- de 0.38 hasta 0.89; LR+ del índice motor fino es de 1.4 y del 12.6 con un LR- de 0.39 y 0.97; el LR+ del índice motor grueso 2.9 y de 14.7 con un LR- de 0.13 y de 0.16; el índice cognitivo con un coeficiente intelectual del <70 hasta el <85 presenta un LR+ del 0.47 al 67 con un LR- de 0.11 al 5.8; el índice del lenguaje con un coeficiente intelectual del <70 hasta el <85 presenta un LR+ del 2.7 al 58 con un LR- de 0.08 al 0.94; el índice cognitivo lingüístico con un coeficiente intelectual <83 tiene un LR+ de 9.7 y

3.45 con un LR- 0.15 y 0.3; el lenguaje receptivo tiene un LR+ de 5.7 y un LR- de 0.48; el lenguaje expresivo con un LR+ de 3.1 y un LR- de 0.38; el LR+ de retraso leve según el DSM – IV es de 1.7 con un LR- de 0.26; el retraso moderado según DSM – IV presenta un LR+ de 0.84 con un LR- 2.4; el LR+ del retraso severo DSM – IV es de 15.75 con un LR- del 0.38; el LR+ del índice comunicativo es del 5.8 con un LR- del 0.51. Por ende, según los datos obtenidos en todos los tipos de LR se determinó que el 27% de LR+ indican un test excelente, 23% un test bueno, 30% test regular y un 20% un test malo; en cambio, un 3% de LR- indican un test excelente, 20% test bueno, 43% test regular y un 33% un test malo.

**Tabla 10.** LR+ y LR- del test de Denver.

| Referencia                           | Tipo                        | Denver | Test      |
|--------------------------------------|-----------------------------|--------|-----------|
| <b>Ryu, et al (2019).</b> (30)       | LR+ del lenguaje            | 2.3    | Regular   |
|                                      | LR- del lenguaje            | 0.11   | Bueno     |
| <b>Kharlukhi, et al (2021).</b> (40) | LR+ de la motricidad gruesa | 1.1    | Malo      |
|                                      | LR- de la motricidad gruesa | 0.56   | Malo      |
| <b>Montedori, et al (2024).</b> (41) | LR+ del retraso probable    | 1.6    | Malo      |
|                                      | LR- del retraso probable    | 0.53   | Malo      |
|                                      | LR+ del estado de alerta    | 1.9    | Malo      |
|                                      | LR- del estado de alerta    | 0.61   | Malo      |
|                                      | LR+ del desarrollo normal   | 0.8    | Malo      |
|                                      | LR- del desarrollo normal   | 1.07   | Malo      |
| <b>Nair, et al (2014).</b> (42)      | LR+ de la motricidad gruesa | 46.2   | Excelente |
|                                      | LR- de la motricidad gruesa | 0.35   | Regular   |
| <b>Coelho, et al (2016).</b> (43)    | LR+ del retraso probable    | 1.6    | Malo      |
|                                      | LR- del retraso probable    | 0.53   | Malo      |
|                                      | LR+ del estado de alerta    | 1.9    | Malo      |
|                                      | LR- del estado de alerta    | 0.61   | Malo      |

*Elaborado por:* Bermúdez D, González J.

**Interpretación:** Basándose en lo escrito en la tabla 10 se puede determinar que el LR+ de lenguaje 2.3 y el LR- de 0.11; la motricidad gruesa presenta un LR+ de 1.1 y 46.2 con un LR- de 0.56 y 0.35 respectivamente; el retraso probable con un LR+ de 1.6 y un LR- de 0.53; el LR+ del estado de alerta 1.9 con un LR- de 0.61; LR+ del desarrollo normal 0.8 con un LR- de 1.07. Por lo tanto, a partir de los datos obtenidos en los diferentes tipos de LR, se determinó que el 12% de los LR+ corresponden a un test excelente, el 12% a un test regular y el 75% a un test deficiente. Por otro lado, el 12% de los LR- se clasifican como test bueno, el 12% como test regular y el 75% como test deficiente.

## 5. Discusión

La evaluación del neurodesarrollo infantil es fundamental para la detección temprana de alteraciones que llegan a impactar de manera negativa en el desarrollo cognitivo, motor y socioemocional del niño. Bajo los criterios previamente establecidos por la pregunta PICO, se permitió evaluar la sensibilidad, especificidad, valores predictivos y razones de verosimilitud LR (+/-) del Test de Denver y el Test de Bayley – III para la identificación de retrasos dentro del neurodesarrollo infantil.

La selección de estudios se centró en un grupo de población de niños de 1 a 42 meses, tomando en consideración investigaciones publicadas en un período de 10 años. Se excluyó la literatura gris con el fin de garantizar la inclusión de artículos completos y con alta calidad. Además, este enfoque permitió limitar posibles sesgos y llevar a cabo una selección rigurosa que permita abarcar estudios transversales obtenidos previamente.

En la evaluación de la calidad de estudios se analizaron únicamente de 17 estudios donde se pudo observar una distribución equitativa, ya que el 53% fue clasificado como alta calidad, mientras que el 47% restante fue considerado como calidad media. A pesar de que la mayoría de estudios cumplieron con los criterios metodológicos, aún persisten desafíos en la implementación de criterios específicos dentro de la evaluación.

Por otra parte, en la evaluación de sesgos se pudo evidenciar que el sesgo de confusión presentó un predominio de bajo riesgo del 59% indicando que los estudios parecen estar diseñados adecuadamente, mientras que el 29% que es catalogado como de alto riesgo, resaltando la necesidad de mejorar con el fin de evitar el entorpecimiento en los resultados. No obstante, el sesgo de selección presenta una mayor preocupación, ya que el 70% es de alto riesgo, pudiendo comprometer la representatividad y generalización de los hallazgos.

En el sesgo de información, recuerdo, medición y del observador predomina el de bajo riesgo con un 76%, lo que indica que dichos aspectos son manejados de manera correcta en la mayoría de los estudios. Esto contrasta con el sesgo de desempeño y confusión por variación temporal, ya que al no presentar datos que lo cataloguen como alto riesgo, si presenta un 47% de datos que no se lo pueden determinar, sugiriendo cierta incertidumbre en la metodología abordada en dichos estudios.

Además, se pudo evidenciar que el sesgo de desgaste y selección presenta un alto riesgo con un 47%, lo que pone de manifiesto un desafío muy significativo, ya que puede alterar la validez en los resultados, en especial si ocurre una pérdida de participantes. Por otro lado, el sesgo de detección de recuerdo, información, clasificación errónea, del observador y de medición, el 70% es de bajo riesgo, mientras que el 18% se lo clasificó como alto riesgo mostrando áreas de mejora, en especial en el uso de herramientas de medición y en la capacitación para cada uno de los observadores.

Por último, el sesgo de observación selectiva de resultados presenta una similitud en sus resultados, ya que el 47% pertenece a bajo riesgo, mientras que el 53% no se lo puede determinar, siendo este último un dato crítico, ya que la carencia de claridad acerca de la selectividad de los resultados puede influenciar de manera considerable en la interpretación y confianza de los hallazgos encontrados en los diferentes estudios.

Con respecto a los resultados de esta revisión, se evidenció que ambos test presentan una sensibilidad elevada en ciertos dominios. En lo que respecta al test de Denver en la publicación de Ryu, et al., presenta una sensibilidad en el lenguaje del 93.3%, lo que sugiere una alta capacidad para identificar niños con retrasos en este aspecto. Sin embargo, su especificidad en el lenguaje es relativamente baja (60%), lo que podría conducir a falsos positivos y, en consecuencia, a un sobrediagnóstico. En el caso de la publicación de Kharlukhi, et al., en la motricidad gruesa la sensibilidad es del 86.6%, mientras que la especificidad fluctúa en 24.1%. Estas diferencias reflejan una variabilidad que puede influir en la confiabilidad del test dependiendo del contexto de aplicación.

Por otro lado, el Test Bayley-III muestra ventajas significativas en términos de sensibilidad y especificidad en áreas como el desarrollo cognitivo, motor fino y lingüístico. Por ejemplo, en el artículo de Johnson, et al en el índice de lenguaje con un coeficiente intelectual (CI) del <85, esta herramienta alcanza una sensibilidad del 100% y una especificidad del 96%. Asimismo, en el artículo de Duncan, et al., el motor fino con un CI de <70 demuestra una sensibilidad del 100% y una especificidad del 93%. Estas características hacen de la Escala Bayley-III sea una herramienta especialmente adecuada para entornos especializados, donde es fundamental confirmar diagnósticos y planificar intervenciones basadas en evaluaciones altamente precisas.

De igual forma se evidenció que los test presentan sensibilidades bajas en algunos dominios. En relación al test de Denver en el escrito de Montedori, et al., manifiesta una sensibilidad del desarrollo normal del 21% con una especificidad del 74%; de manera similar en el escrito de Coelho, et al., manifiesta una sensibilidad del estado de alerta con un 57% y una especificidad del 70% lo que indica que esta prueba genera más falsos negativos que falsos positivos. En el test de Bayley-III en el escrito de Mansson, et al., presenta una sensibilidad intelectual con un CI <70 del 18% sin una especificidad demostrada; en Liu, et al., presenta una sensibilidad cognitiva con un CI <85 del 22.5 % con una sensibilidad del 92.5%. Para una mejora en la eficacia de estas pruebas es necesario implementar distintos tipos de enfoques como es el uso de múltiples herramientas evaluativas y la capacitación al personal para una adecuada interpretación de los resultados.

El análisis de los valores predictivos también destaca diferencias relevantes. El Test de Denver, Nair et al., tiene un valor predictivo positivo (VPP) alto en motricidad gruesa de 97.12%, lo que respalda su efectividad para confirmar casos positivos. Sin embargo, su valor predictivo negativo (VPN) es de 79.55%, lo que podría implicar que algunos casos negativos no sean correctamente identificados. En

Ryu et al., describe un VPP de lenguaje del 93.3%, reflejando una capacidad de identificación de casos positivos, pero con un VPN del 60% lo que indicaría un porcentaje significativo de casos negativos no identificados correctamente. Por el contrario, la Escala Bayley-III según Johnson, et al., muestra un VPP cognitivo con un CI del <70 del 100% y un VPN del 94%, mostrando una alta precisión en la confirmación y exclusión de los casos; en Ryu et al., presenta un VPP del lenguaje del 96.4% con un VPN del 71.4% llegando a indicar una capacidad de identificación en casos positivos como negativos en el dominio lingüístico.

De manera similar se pudo demostrar que hay una deficiencia en ciertos dominios de acuerdo a sus valores predictivos. En relación al test de Denver, Coelho et al., manifiesta un VPP del estado de alerta del 57% y un VPN del 70%, lo que indica una capacidad moderada para la identificación de casos positivos. En contraste, la escala de Bayley-III, Creighton et al., presenta un VPP del lenguaje con un CI <80 de un 12% y un VPN del 98%, en el dominio cognitivo lingüístico con un CI <83 de un 15% y un valor VPN del 99%, lo que refleja una excelente capacidad para la detección de falsos negativos, pero con una precisión baja de casos positivos; Liu et al., reporta un VPP cognitivo con un CI <70 del 27.4% y un VPN del 97.3%, sugiriendo una capacidad moderada en la identificación de casos positivos y una alta capacidad en identificar casos negativos.

Los coeficientes de razón de verosimilitud (LR+ y LR-) aportan información adicional sobre la utilidad diagnóstica de estas herramientas. En la Bayley-III, los valores de LR+ en dominios como el lenguaje y la motricidad gruesa son notablemente altos, lo que indica una excelente capacidad para aumentar la probabilidad post-prueba de un diagnóstico positivo. Por ejemplo, Wong et al., indica un LR+ en el índice motor del 103.2 indicando que es altamente efectiva para confirmar diagnósticos positivos, presenta un LR- del 0.38 presentando una buena capacidad para identificar correctamente la ausencia de una enfermedad. Fairbairn et al., manifiesta un LR+ del índice cognitivo del 67 con un LR- del 0.33; LR+ del índice motor fino del 12.6 con un LR- de 0.39. Duncan et al., indica un LR+ del índice motor grueso con un CI <70 del 14.7 y un LR- del 0.13. Por último, Milne et al., con un LR+ del retraso severo según la DSM-IV de 15.75 con un LR- del 0.38.

Por lo contrario, el test de Denver según Nair et al., indica un LR+ de la motricidad gruesa del 46.2, lo que sugiere una capacidad relativamente moderada para la confirmación de diagnósticos positivos, sin embargo, presenta un LR- de 0.35 lo que implica una capacidad deficiente para descartar diagnósticos negativos en comparación con el test de Bayley-III. Además, Ryu et al., con un LR+ en el índice lingüístico del 2.3 presenta una capacidad limitada de un diagnóstico positivo, mientras que su LR- de 0.11 nos indica una capacidad razonable para eliminar un diagnóstico negativo.

De manera análoga también presentan coeficientes de razón de verosimilitud relativamente bajos en ciertos dominios. En el test de Bayley-III, Blazek et al., manifiesta un LR+ en el índice cognitivo de 1.1 con un LR- del 0.8, lo que indica una capacidad limitada para confirmar diagnósticos positivos y una

capacidad moderada para excluir los negativos. Milne et al., por su parte indica un LR+ en el retraso moderado según la DSM-IV del 0.84 con un LR- del 2.4 demostrando una baja capacidad para reconocer casos positivos y una capacidad moderada para los diagnósticos negativos.

En comparación, los LR+ y LR- del Test de Denver son menos consistentes, especialmente en dominios como el desarrollo normal, el estado de alerta, motricidad gruesa y retraso probable, donde los valores no alcanzan niveles diagnósticamente significativos. Esto podría limitar su aplicabilidad en ciertos contextos clínicos donde se requiere una alta precisión diagnóstica. Por ejemplo, Kharlukhi et al., indicó un LR+ de la motricidad gruesa del 1.1 con un LR- del 0.56. Montedori et al., con un LR+ del desarrollo normal del 0.8 con un LR- del 1.07. Coelho et al., manifestó un LR+ del retraso probable del 1.6 con un LR- del 0.53; con un LR+ del estado de alerta del 1.9 con un LR- del 0.61, todos ellos reflejan una capacidad deficiente en las diferentes pruebas, ya sea para su confirmación o su exclusión en diagnósticos.

A pesar de sus fortalezas, ambas herramientas presentan limitaciones que deben considerarse al seleccionar la más adecuada para un contexto clínico específico. En el caso del Test de Denver, la variabilidad en la estandarización de las aplicaciones y su dependencia de la interpretación subjetiva del evaluador pueden introducir sesgos que afecten la validez de los resultados. Por su parte, la Escala Bayley-III, aunque es más detallada y precisa, requiere una capacitación específica para su aplicación, además de un mayor tiempo y recursos, lo que podría limitar su uso en contextos de atención primaria o en áreas con recursos limitados.

Los estudios revisados también muestran que estas herramientas tienen diferencias en términos de su utilidad para poblaciones específicas. Mientras que el Test de Denver es más adecuado para el cribado inicial en poblaciones amplias debido a su practicidad y bajo costo, la Escala Bayley-III es preferida en entornos especializados, donde la precisión diagnóstica es crucial. Por ejemplo, en niños prematuros o con antecedentes perinatales de riesgo, Bayley-III se ha utilizado ampliamente para evaluar el impacto de estos factores en el desarrollo neurocognitivo y motor, proporcionando datos que pueden guiar intervenciones personalizadas.

## **6. Limitaciones**

La presente revisión sistemática presenta ciertas limitaciones que deben considerarse al momento de interpretar sus hallazgos. En primer lugar, se dificultó la búsqueda de artículos que evaluaban específicamente la validez de Denver debido a la disponibilidad, razón por la cual nos vimos en la obligación de ampliar el rango de años de hasta diez años atrás, a pesar de que existía una gran cantidad de artículos actuales sobre su descripción.

En segundo lugar, al momento de realizar la búsqueda de artículos y seleccionarlos de acuerdo a los criterios de exclusión e inclusión, encontramos artículos de acceso restringido que en un inicio parecían ser los adecuados, sin embargo, la mayoría de ellos al momento de adquirirlos no brindaban la

información necesaria. Finalmente, la gran parte de los estudios seleccionados para los resultados presentaban un alto riesgo de sesgo, lo que generó incertidumbre acerca de su impacto en su validez y aplicabilidad en la práctica clínica.

## **7. Conclusiones**

A lo largo de la revisión sistemática, se pudo evidenciar que tanto el test de Bayley-III como el test de Denver son herramientas de gran utilidad para una correcta evaluación en el neurodesarrollo infantil en pacientes de grupo etario de 1 a 42 meses. No obstante, ambas tienen diferencias en relación a la sensibilidad, especificidad como también dentro de la práctica clínica. El test de Bayley-III mostró mayor precisión, debido a la alta sensibilidad y especificidad dentro de la parte cognitiva, motora y lingüística, por lo que es considerada como una herramienta adecuada al momento de diagnosticar.

La calidad de estudios que fueron revisados, la gran parte fueron catalogados como media o alta. A pesar de ello, se encontraron limitaciones metodológicas de gran relevancia, como sesgo de selección, una disminución en el tamaño de muestra e inconvenientes para la estandarización en sus aplicaciones. Por lo que dichas limitaciones resaltan la urgencia de ser mejoradas para futuras investigaciones aplicando diseños que fortalezcan la evidencia disponible.

En cuanto a la sensibilidad y especificidad, el test de Bayley-III se destacó por ser de gran ayuda en el diagnóstico de retrasos dentro del desarrollo, llegando incluso a superar al test de Denver. Para la interpretación de los valores predictivos como de las razones de verosimilitud, se pudo observar que el test de Bayley-III mostró datos más consistentes y una mayor capacidad para poder confirmar o descartar diagnósticos, a diferencia del test de Denver, ya que este presentó ciertas limitaciones, llegando a restringir su uso en aquellos aspectos donde se requiera de una alta precisión dentro de la práctica clínica.

Finalmente, se puede decir que el test de Denver es el que más se recomienda para aquellas situaciones donde se desee realizar un cribado inicial dentro de la atención primaria, y para confirmaciones diagnósticas donde sea necesario obtener una evaluación detallada a profundidad en aquellos niños que presenten factores de riesgo, el test de Bayley-III es la opción más adecuada dado que su sensibilidad, especificidad y valor predictivo es superior al resto, sin embargo, la implementación se encuentra limitada debido al tiempo y su capacitación.

## 8. Referencias Bibliográficas

1. Abreu L, Morilla A, Parada Y, Tamayo V, Cabrerías E, Rodríguez D. Factores de riesgo perinatales y evolución del neurodesarrollo hasta el primer año de edad. *Rev Cubana Pediatr* [Internet]. 2021 [citado el 17 de julio de 2024];93(4). Disponible en: <https://www.medigraphic.com/pdfs/revcubped/cup-2021/cup214g.pdf>
2. Yamaguchi B, Silva AZ, Araujo LB, Guimarães ATB, Israel VL. Psychomotor evaluation of children attending Child Education Centers in the south of Brazil. *Early Child Dev Care* [Internet]. 2021 [citado el 17 de julio de 2024];191(11). Disponible en: <https://doi.org/10.1080/03004430.2019.1672165>
3. Castro Y, García X, Bermúdez I. La estimulación del neurodesarrollo infantil como contenido de la formación inicial de los profesionales de la educación . *Conrado* [Internet]. 2021 [citado el 17 de julio de 2024];17(83). Disponible en: <http://scielo.sld.cu/pdf/rc/v17n83/1990-8644-rc-17-83-174.pdf>
4. Álvarez-Izazaga MA, Galindo-Gómez C, Roldán-Amaro JA, Saucedo-Arteaga G, Díaz-Martínez M, Chávez-Villasana A, et al. Neurodevelopment and timely early stimulation in children of indigenous migrant and non-migrant mothers in Chihuahua, Mexico. *Anales de Psicología* [Internet]. 2022 [citado el 17 de julio de 2024];38(2). Disponible en: [https://scielo.isciii.es/scielo.php?pid=S0212-97282022000200005&script=sci\\_arttext](https://scielo.isciii.es/scielo.php?pid=S0212-97282022000200005&script=sci_arttext)
5. Medina-Alva P. La atención del neurodesarrollo en los primeros años de vida. *Revista Peru Pediátrica* . 2023;75(1).
6. Martínez-Moreno A, Imbernón Giménez S, Díaz Suárez A. The Psychomotor Profile of Pupils in Early Childhood Education. *Sustainability* [Internet]. 2020 [citado el 17 de julio de 2024];12(6). Disponible en: <https://doi.org/10.3390/su12062564>
7. Förster J, López I. Neurodesarrollo humano: un proceso de cambio continuo de un sistema abierto y sensible al contexto. *Revista Médica Clínica Las Condes* [Internet]. 2022 [citado el 17 de julio de 2024];33(4). Disponible en: <https://www.sciencedirect.com/science/article/pii/S0716864022000724>
8. Soleimani F, Azari N, Vameghi R, Barekati SH, Lornejad H, Kraskian A. Standardization of the Bayley Scales of Infant and Toddler Development for Persian Children. *J Rehabil* [Internet]. 2022 [citado el 4 de agosto de 2024];23(1). Disponible en: [https://rehabilitationj.uswr.ac.ir/browse.php?a\\_id=2890&sid=1&slc\\_lang=en&html=1](https://rehabilitationj.uswr.ac.ir/browse.php?a_id=2890&sid=1&slc_lang=en&html=1)
9. Cruz-Lara A, Ángeles-Llerenas A, Katz-Guss G, Astudillo-García CI, Rangel-Eudave NG, Rivero-Rangel GM, et al. Conocimiento sobre trastornos del neurodesarrollo

asociado con la aceptación del modelo de educación inclusiva en docentes de educación básica. *Salud Publica Mex* [Internet]. 2020 [citado el 4 de agosto de 2024];62. Disponible en: <https://www.medigraphic.com/pdfs/salpubmex/sal-2020/sal205m.pdf>

10. Flores P, Teixeira JE, Leal AK, Ribeiro J, Monteiro AM, Fonseca RB, et al. The Necessity of a Reduced Version of the Psychomotor Battery to Screen for Learning Difficulties in Preschool Children. *Sustainability* [Internet]. 2022 [citado el 17 de julio de 2024];14(12). Disponible en: <https://www.mdpi.com/2071-1050/14/12/7263>
11. Chung HJ, Yang D, Kim GH, Kim SK, Kim SW, Kim YK, et al. Development of the Korean Developmental Screening Test for Infants and Children (K-DST). *Clin Exp Pediatr* [Internet]. 2020 [citado el 17 de julio de 2024];63(11). Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7642140/>
12. Arias V, Benavides E. Evaluación de la psicomotricidad en niños menores de 3 años durante la teleeducación en tiempos de confinamiento. *Ciencia Latina Revista Científica Multidisciplinar* [Internet]. 2021 [citado el 17 de julio de 2024];5(6). Disponible en: <https://ciencialatina.org/index.php/cienciala/article/view/1266>
13. Hubacova V. Assessment of Psychomotor Development of Preschool Children: A Review of Eight Psychomotor Developmental Tools. *World Academy of Science, Engineering and Technology International Journal of Sport and Health Sciences* [Internet]. 2020 [citado el 17 de julio de 2024];14(11). Disponible en: [https://d1wqtxts1xzle7.cloudfront.net/102357131/pdf-libre.pdf?1684401170=&response-content-disposition=inline%3B+filename%3DAssessment\\_of\\_Psychomotor\\_Development\\_of.pdf&Expires=1721049251&Signature=IyYvqvDSwjp4O1WNQYy9ELuyJelCaAVwB238~9YPvZR6MhVjAEXST1Oxn4lLAzbEoEtsOAcMC3RuyAgp-gLKKg9rNM4CQmZuxN1KEpuwJ7zGUphC1YO5PvYhKrm84VGG-t0dpgBvOnzea1Up2oBcle1s3VPnlCtUKrOft04RL-12McYyYm28OE9MbAS3KbmMYyfvaM66sXP-tIWBO3sezemV-WqN9RG28yHG7EhU8AkTw488YWBDd9EoafsHxbuY-0GAXJn3fV~DdmnZbR~5811CuRwA6p91SaEaRUX1lbVPX18bbjkMKXdWKAqPq737vzsH1zaYozLdOtPvb7RzQ\\_\\_&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA](https://d1wqtxts1xzle7.cloudfront.net/102357131/pdf-libre.pdf?1684401170=&response-content-disposition=inline%3B+filename%3DAssessment_of_Psychomotor_Development_of.pdf&Expires=1721049251&Signature=IyYvqvDSwjp4O1WNQYy9ELuyJelCaAVwB238~9YPvZR6MhVjAEXST1Oxn4lLAzbEoEtsOAcMC3RuyAgp-gLKKg9rNM4CQmZuxN1KEpuwJ7zGUphC1YO5PvYhKrm84VGG-t0dpgBvOnzea1Up2oBcle1s3VPnlCtUKrOft04RL-12McYyYm28OE9MbAS3KbmMYyfvaM66sXP-tIWBO3sezemV-WqN9RG28yHG7EhU8AkTw488YWBDd9EoafsHxbuY-0GAXJn3fV~DdmnZbR~5811CuRwA6p91SaEaRUX1lbVPX18bbjkMKXdWKAqPq737vzsH1zaYozLdOtPvb7RzQ__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA)
14. Mustakim M, Irawan R, Irmawati M, Setyoboedi B. Impact of Stunting on Development of Children between 1-3 Years of Age. *Ethiop J Health Sci*. 2022;32(3).
15. Kim KM, Choi JW. Associations between breastfeeding and cognitive function in children from early childhood to school age: a prospective birth cohort study. *Int*

- Breastfeed J [Internet]. 2020 [citado el 17 de julio de 2024];15(1). Disponible en: <https://internationalbreastfeedingjournal.biomedcentral.com/articles/10.1186/s13006-020-00326-4>
16. Cardoso TF, Santos RS dos, Corrêa RM, Campos JV, Silva R de B, Tobias CC, et al. Congenital Zika infection: neurology can occur without microcephaly. Arch Dis Child [Internet]. 2019 [citado el 17 de julio de 2024];104(2). Disponible en: <https://adc.bmj.com/content/104/2/199.abstract>
  17. Costa CC, Blascovi-Assis SM. Teste de triagem Denver II no monitoramento do desenvolvimento de crianças com transtorno do espectro autista: uma revisão de literatura . Centro de Pesquisas Avançadas em Qualidade de Vida [Internet]. 2022 [citado el 17 de julio de 2024];14. Disponible en: [https://www.researchgate.net/profile/Silvana-Blascovi-Assis/publication/364838212\\_TESTE\\_DE\\_TRIAGEM\\_DENVER\\_II\\_NO\\_MONITORAMENTO\\_DO\\_DESENVOLVIMENTO\\_DE\\_CRIANCAS\\_COM\\_TRANSTORNO\\_DO\\_ESPECTRO\\_AUTISTA\\_UMA\\_REVISAO\\_DE\\_LITERATURA/links/635d3a9296e83c26eb64d4ee/TESTE-DE-TRIAGEM-DENVER-II-NO-MONITORAMENTO-DO-DESENVOLVIMENTO-DE-CRIANCAS-COM-TRANSTORNO-DO-ESPECTRO-AUTISTA-UMA-REVISAO-DE-LITERATURA.pdf](https://www.researchgate.net/profile/Silvana-Blascovi-Assis/publication/364838212_TESTE_DE_TRIAGEM_DENVER_II_NO_MONITORAMENTO_DO_DESENVOLVIMENTO_DE_CRIANCAS_COM_TRANSTORNO_DO_ESPECTRO_AUTISTA_UMA_REVISAO_DE_LITERATURA/links/635d3a9296e83c26eb64d4ee/TESTE-DE-TRIAGEM-DENVER-II-NO-MONITORAMENTO-DO-DESENVOLVIMENTO-DE-CRIANCAS-COM-TRANSTORNO-DO-ESPECTRO-AUTISTA-UMA-REVISAO-DE-LITERATURA.pdf)
  18. Jumbo Salazar FF, Salazar Villacis MG, Acosta Gavilánez RI, Torres Constante DV. Test de Denver y el test Prunape, instrumentos para identificar alteraciones de desarrollo psicomotor. Revista Científica UISRAEL [Internet]. 2021 [citado el 17 de julio de 2024];8(1). Disponible en: [http://scielo.senescyt.gob.ec/scielo.php?pid=S2631-27862021000300123&script=sci\\_arttext](http://scielo.senescyt.gob.ec/scielo.php?pid=S2631-27862021000300123&script=sci_arttext)
  19. Herrera G, Gutiérrez Z. Resultados del tratamiento de rehabilitación física en niños con retardo en el desarrollo psicomotor. Interdisciplinary Rehabilitation / Rehabilitacion Interdisciplinaria [Internet]. 2023 [citado el 17 de julio de 2024];3(28). Disponible en: <https://ri.saludcyt.ar/index.php/ri/article/view/28/15>
  20. Goldstone AB, Baiocchi M, Wypij D, Stopp C, Andropoulos DB, Atallah J, et al. The Bayley-III scale may underestimate neurodevelopmental disability after cardiac surgery in infants. European Journal of Cardio-Thoracic Surgery [Internet]. 2020 [citado el 17 de julio de 2024];57(1). Disponible en: <https://academic.oup.com/ejcts/article/57/1/63/5475811>
  21. Celik P, Ayracin İ, Yakut H. Which Bayley-III cut-off values should be used in different developmental levels? Turk J Med Sci [Internet]. 2020 [citado el 17 de julio de 2024];50(4). Disponible en: <https://journals.tubitak.gov.tr/medical/vol50/iss4/15/>

22. Hwan S, Sim YJ. The validity and reliability of DDST II and Bayley III in children with language development delay . *Neurology Asia* [Internet]. 2019 [citado el 17 de julio de 2024];24(4). Disponible en: [https://www.neurologyasia.org/articles/neuroasia-2019-24\(4\)-355.pdf](https://www.neurologyasia.org/articles/neuroasia-2019-24(4)-355.pdf)
23. Balasundaram P, Avulakunta I. Bayley Scales Of Infant and Toddler Development. *Europe PMC* [Internet]. 2021 [citado el 17 de julio de 2024]; Disponible en: <https://europepmc.org/article/NBK/nbk567715>
24. Sun L, Sabanathan S, Thanh PN, Kim A, Doa TTM, Thwaites CL, et al. Bayley III in Vietnamese children: lessons for cross-cultural comparisons. *Wellcome Open Res* [Internet]. 2019 [citado el 17 de julio de 2024];4(98). Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6685399/>
25. Del Rosario C, Slevin M, Molloy EJ, Quigley J, Nixon E. How to use the Bayley Scales of Infant and Toddler Development. *Arch Dis Child Educ Pract Ed* [Internet]. 2021 [citado el 17 de julio de 2024];106(2). Disponible en: <https://ep.bmj.com/content/106/2/108.abstract>
26. Månsson J, Källén K, Eklöf E, Serenius F, Ådén U, Stjernqvist K. The ability of Bayley-III scores to predict later intelligence in children born extremely preterm. *Acta Paediatr* [Internet]. 2021 [citado el 17 de julio de 2024];110(11). Disponible en: <https://onlinelibrary.wiley.com/doi/epdf/10.1111/apa.16037>
27. Herramientas de evaluación de la calidad de los estudios . *NHLBI, NIH* [Internet]. [citado el 17 de julio de 2024]; Disponible en: <https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools>
28. Montgomery C, Setänen S, Kaul YF, Farooqi A, Broström L, Aden U, et al. Predictive value of Bayley-III Motor Index for later motor difficulties in children born extremely preterm. *Acta Paediatr* [Internet]. 2023 [citado el 5 de enero de 2025];112(4). Disponible en: <https://pubmed.ncbi.nlm.nih.gov/36723223/#:~:text=Conclusion%3A%20The%20Bayley%20III%20at, follow%20Dup%20beyond%20preschool%20age.>
29. Liu TY, Chang JH, Peng CC, Hsu CH, Jim WT, Lin JY, et al. Predictive Validity of the Bayley-III Cognitive Scores at 6 Months for Cognitive Outcomes at 24 Months in Very-Low-Birth-Weight Infants. *Front Pediatr* [Internet]. 2021 [citado el 5 de enero de 2025];9. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/34026684/>
30. Ryu S, Sim YJ. The validity and reliability of DDST II and Bayley III in children with language development delay . *Neurology Asia* [Internet]. 2019 [citado el 5 de enero de

- 2025];24(4). Disponible en: [https://www.neurology-asia.org/articles/neuroasia-2019-24\(4\)-355.pdf](https://www.neurology-asia.org/articles/neuroasia-2019-24(4)-355.pdf)
31. Månsson J, Källén K, Eklöf E, Serenius F, Ådén U, Stjernqvist K. The ability of Bayley-III scores to predict later intelligence in children born extremely preterm. *Acta Paediatr* [Internet]. 2021 [citado el 5 de enero de 2025];110(11). Disponible en: <https://pubmed.ncbi.nlm.nih.gov/34289173/>
  32. Blazek JW, Colditz PB, Guzzetta A, Ware RS, Chatfield MD, Hough JL, et al. Sensitivity and specificity of the Neonatal Visual Assessment to predict motor and cognitive outcomes in infants born very preterm. *Early Hum Dev* [Internet]. 2024 [citado el 5 de enero de 2025]; Disponible en: <https://pubmed.ncbi.nlm.nih.gov/38968818/>
  33. Wong HS, Cowan FM, Modi N. Validity of neurodevelopmental outcomes of children born very preterm assessed during routine clinical follow-up in England. *Arch Dis Child Fetal Neonatal Ed* [Internet]. 2018 [citado el 5 de enero de 2025];103(5). Disponible en: <https://pubmed.ncbi.nlm.nih.gov/29079650/>
  34. Creighton DE, Tang S, Newman J, Henderson L, Sauve R. Establishing Bayley-III cut-off scores at 21 months for predicting low IQ scores at 3 years of age in a preterm cohort. *Paediatrics Child Health* [Internet]. 2018 [citado el 5 de enero de 2025];23(8). Disponible en: <https://pubmed.ncbi.nlm.nih.gov/30842698/>
  35. Milne SL, McDonald JL, Comino EJ. Alternate scoring of the Bayley-III improves prediction of performance on Griffiths Mental Development Scales before school entry in preschoolers with developmental concerns. *Child Care Health Dev* [Internet]. 2014 [citado el 5 de enero de 2025];41(2). Disponible en: <https://pubmed.ncbi.nlm.nih.gov/25040260/>
  36. Fairbairn N, Galea C, Loughran-Fowlds A, Hodge A, Badawi N, Walker K. Prediction of three year outcomes using the Bayley-III for surgical, cardiac and healthy Australian infants at one year of age. *Early Hum Dev* [Internet]. 2018 [citado el 5 de enero de 2025];117. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/29288912/>
  37. 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? *Journal of Perinatology* [Internet]. 2016 [citado el 5 de enero de 2025];35(7). Disponible en: <https://pubmed.ncbi.nlm.nih.gov/25634519/>
  38. Bode MM, D'Eugenio DB, Mettelman BB, Gross SJ. Predictive Validity of the Bayley, Third Edition at 2 Years for Intelligence Quotient at 4 Years in Preterm Infants. *Journal*

- of Developmental Behavioral Pediatrics [Internet]. 2014 [citado el 5 de enero de 2025];35(9). Disponible en: <https://pubmed.ncbi.nlm.nih.gov/25370298/>
39. Johnson S, Moore T, Marlow N. Using the Bayley-III to assess neurodevelopmental delay: which cut-off should be used? *Pediatr Res* [Internet]. 2014 [citado el 5 de enero de 2025];75(5). Disponible en: <https://pubmed.ncbi.nlm.nih.gov/24492622/>
40. Kharlukhi J, Narasimhan U, James S, Anitha FS, Suresh S, Polina SI. Effectiveness of Bedside Clinical Screening Tools in Predicting Short-Term Neurodevelopmental Delay Among Very-Low-Birth-Weight Pre-terms: A Prospective Observational Study. *Cureus* [Internet]. 2021 [citado el 5 de enero de 2025];13(12). Disponible en: <https://pubmed.ncbi.nlm.nih.gov/35036195/>
41. Montedori K, Lima M. Early childhood development monitoring during the first thousand days: Investigating the relationship between the developmental surveillance instrument and standardized scales. *Early Hum Dev* [Internet]. 2024 [citado el 5 de enero de 2025];190. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0378378224000343>
42. Nair MKC, Sunitha RM, Leena ML, George B, Bhaskaran D, Russell PSS. CDC Kerala 2: Developmental Intervention Package for Babies <1,800 g – Outcome at 6 mo Using DASII. *The Indian Journal of Pediatrics* [Internet]. 2014 [citado el 5 de enero de 2025];81(2). Disponible en: <https://www.scopus.com/record/display.uri?eid=2-s2.0-84922105021&doi=10.1007%2fs12098-014-1624-z&origin=inward&txGid=770c2709498255decf3c9d742c70847c>
43. Coelho R, Ferreira JP, Sukiennik R, Halpern R. Child development in primary care: a surveillance proposal. *J Pediatr (Rio J)* [Internet]. 2016 [citado el 5 de enero de 2025];92(5). Disponible en: <https://www.sciencedirect.com/science/article/pii/S0021755716300419?via%3Dihub#sec0015>

## 9. Anexos

*Anexo 1. Definición de las letras utilizadas en la Herramienta de Evaluación de Calidad de los estudios. (27)*

- A. ¿La pregunta o el objetivo de la investigación en este documento se enunciaron claramente?
- B. ¿Se especificó y definió claramente la población del estudio?
- C. ¿La tasa de participación de las personas elegibles fue al menos del 50%?
- D. ¿Se seleccionaron o reclutaron todos los sujetos de la misma población o de poblaciones similares (incluido el mismo período de tiempo)? ¿Se especificaron previamente los criterios de inclusión y exclusión para participar en el estudio y se aplicaron de manera uniforme a todos los participantes?
- E. ¿Se proporcionó una justificación del tamaño de la muestra, una descripción de la potencia o estimaciones de varianza y efecto?
- F. Para los análisis de este documento, ¿se midieron las exposiciones de interés antes de medir los resultados?
- G. ¿Fue el período de tiempo suficiente para que uno pudiera esperar razonablemente ver una asociación entre la exposición y el resultado, si existiera?
- H. Para las exposiciones que pueden variar en cantidad o nivel, ¿el estudio examinó diferentes niveles de exposición en relación con el resultado (por ejemplo, categorías de exposición o exposición medida como variable continua)?
- I. ¿Las medidas de exposición (variables independientes) estaban claramente definidas, eran válidas, confiables y se implementaron de manera consistente entre todos los participantes del estudio?
- J. ¿Se evaluaron las exposiciones más de una vez a lo largo del tiempo?
- K. ¿Las medidas de resultados (variables dependientes) fueron claramente definidas, válidas, confiables y se implementaron de manera consistente entre todos los participantes del estudio?
- L. ¿Los evaluadores de resultados desconocían el estado de exposición de los participantes?
- M. ¿La pérdida de seguimiento después del inicio fue del 20% o menos?
- N. ¿Se midieron las principales variables de confusión potenciales y se ajustaron estadísticamente para determinar su impacto en la relación entre exposición(es) y resultado(s)?



## **AUTORIZACIÓN DE PUBLICACIÓN EN EL REPOSITORIO INSTITUCIONAL**

**Danna Emilia Bermúdez Padilla** portador(a) de la cédula de ciudadanía N° **0106511140**. En calidad de autor/a y titular de los derechos patrimoniales del proyecto de titulación **“Efectividad del test de denver versus el test bayley-iii en la evaluación del neurodesarrollo Infantil en pacientes de 1-42 meses: revisión sistemática”** de conformidad a lo establecido en el artículo 114 Código Orgánico de la Economía Social de los Conocimientos, Creatividad e Innovación, reconozco a favor de la Universidad Católica de Cuenca una licencia gratuita, intransferible y no exclusiva para el uso no comercial de la obra, con fines estrictamente académicos y no comerciales. Autorizo además a la Universidad Católica de Cuenca, para que realice la publicación de éste proyecto de titulación en el Repositorio Institucional de conformidad a lo dispuesto en el artículo 144 de la Ley Orgánica de Educación Superior.

Azogues, **13 de mayo de 2026**

F: 

**Danna Emilia Bermúdez Padilla**

**C.I. 0106511140**

## **AUTORIZACIÓN DE PUBLICACIÓN EN EL REPOSITORIO INSTITUCIONAL**

**Joselyn Lizbeth González Minchala** portador(a) de la cédula de ciudadanía N° **0302471420**. En calidad de autor/a y titular de los derechos patrimoniales del proyecto de titulación **“Efectividad del test de denver versus el test bayley-iii en la evaluación del neurodesarrollo infantil en pacientes de 1-42 meses: revisión sistemática”** de conformidad a lo establecido en el artículo 114 Código Orgánico de la Economía Social de los Conocimientos, Creatividad e Innovación, reconozco a favor de la Universidad Católica de Cuenca una licencia gratuita, intransferible y no exclusiva para el uso no comercial de la obra, con fines estrictamente académicos y no comerciales. Autorizo además a la Universidad Católica de Cuenca, para que realice la publicación de éste proyecto de titulación en el Repositorio Institucional de conformidad a lo dispuesto en el artículo 144 de la Ley Orgánica de Educación Superior.

Azogues, **13 de mayo de 2026**

F: 

**Joselyn Lizbeth González Minchala**

**C.I. 0302471420**