VALIDEZ PREDICTIVA DEL AGES AND STAGES QUESTIONNAIRE (ASQ®)

UN CUESTIONARIO DE CRIBADO DEL DESARROLLO PSICOMOTOR INFANTIL BASADO EN EL REPORTE DE PADRES O CUIDADORES PRINCIPALES

DRA. LUISA SCHONHAUT BERMAN

DOCTORADO METODOLOGÍA DE LA INVESTIGACIÓN BIOMÉDICA Y SALUD PÚBLICA

DIRECTORES DE TESIS:
XAVIER DEMESTRE G. Y PAMELA SERÓN S.

TUTOR:
XAVIER BONFILL C.

SANTIAGO DE CHILE
NOVIEMBRE 2021

AUTORÍA

DRA. LUISA SCHONHAUT BERMAN

Médico Pediatra

Magister en Salud Pública,

en Educación en Ciencias de la Salud y Epidemiología Clínica

Profesor Asociado Adjunto

Facultad de Medicina Clínica Alemana Universidad del Desarrollo
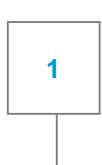
AGRADECIMIENTOS

Esta tesis es el resultado de un proceso de investigación que comenzó hace casi 15 años, en la búsqueda de instrumentos objetivos para evaluar el desarrollo de los niños y poder ofrecerles una intervención oportuna; focalizándonos en grupos de mayor riesgo, como eran los niños que nacían "un poco" prematuros, o cercanos a término.

Esta búsqueda me llevó a cruzarme con el Dr. Xavier Demestre, quien había comenzado una línea de investigación similar con niños prematuros tardíos del SCIAS - Hospital de Barcelona, en una población con características socioeconómicas y culturales similares a los niños que seguimos en Clínica Alemana. Comenzamos así a trabajar y publicar juntos, integrándome en su equipo, junto a la Dra. Silvia Martínez-Nadal. Los lazos que establecimos fueron más allá de lo laboral, construyendo una hermosa amistad. La generosidad de Xavier ha sido infinita, motivándome a hacer el Doctorado en Barcelona, y acompañándome en cada uno de los pasos necesarios para completarlo.

El tercer pilar de esta tesis se lo debo a la Dra. Pamela Serón, quien fue mi profesora en el Magister de Epidemiología Clínica de la Universidad de la Frontera, en Temuco. Pamela me involucró en las revisiones sistemáticas, metaanálisis y metodología de la investigación y participó como tutora de este camino. Sin duda, gracias a Xavier Demestre y Pamela Serón, fue posible entrelazar las tres ramas de la investigación que dieron forma a esta tesis.

En este punto, debo agradecer a Xavier Bonfill, quien participó como tutor, guiándome con su tranquilidad y amabilidad.

Agradezco también a cada uno de los coautores, de cada uno de los trabajos que conforman la tesis, cada uno ha tenido un rol preponderante en este camino. Destaco a Iván Armijo, con quien comenzamos el vuelo de la investigación clínica - metodológica hace ya 14 años, y a Andrés Maturana, con quien pasamos largas horas analizando los aspectos metodológicos, los resultados y la redacción de los manuscritos.

En especial, agradezco a mi familia. A mis padres, que siempre me han apoyado y acompañado en el maravilloso recorrido de la vida. A mis tres maravillosos hijos: Ilan, Mijael y Susana, quienes han compartido y vibrado conmigo, interesándose en mis descubrimientos, análisis y logros. Y a mi esposo Roberto, quien ha transitado incondicionalmente junto a mí, impulsándome en mis sueños personales, académicos y profesionales.

ÍNDICE

ABREVIATURAS

CCDRP:

Cuestionarios de Cribado del Desarrollo basados en el Reporte de Padres o Cuidadores

ASQ:

Ages and Stages Questionnaire

PEDS:

Parents' Evaluations of Developmental Status

PEDS:DM:

Parents' Evaluations of Developmental Status: Developmental Milestones

SWYC:

Survey of Well-being of Young Children

DSM:

Desarrollo Psicomotor

AAP:

Academia Americana de Pediatría

OR:

Odds Ratio

ROC:

Receiver Operating Characteristic o Característica Operativa del Receptor

AUC:

Area under the Curve o Área bajo la curva

QUADAS-2:

Quality of Diagnostic Accuracy Studies (versión 2)

PRESENTACIÓN

La presente tesis reúne y consolida el resultado de tres trabajos que forman parte de una línea de investigación que se inició hace más de 10 años en forma paralela en Santiago de Chile y Barcelona, Cataluña. Esta línea de investigación tiene una vertiente fundamentalmente clínica, relacionada con la validación de un Cuestionario de Cribado del Desarrollo Psicomotor Infantil basado en el Reporte de Padres o Cuidadores principales (CCDRP), como es el Ages and Stages Questionnaire (ASQ). Secundariamente, posee una vertiente metodológica que coloca énfasis en las medidas de asociación para análisis de validez predictiva.

En las siguientes páginas se pone en contexto la importancia de la monitorización del desarrollo psicomotor durante los primeros 5 años de vida, a través de instrumentos estandarizados, y la confiabilidad de los CCDRP. Se analizan los estudios de validación del ASQ en distintos contextos culturales y clínicos, y se realiza una reseña acerca de los antecedentes de su validación en Santiago de Chile y en Barcelona.

Si bien el uso del ASQ es cada vez más difundido a nivel internacional, hay aún algunas brechas de conocimiento, entre las cuales, para elaborar este trabajo de tesis, se consideraron las siguientes:

✖ El ASQ ha sido utilizado en Barcelona para conocer el pronóstico de los niños que nacen prematuros tardíos, con y sin patología perinatal; no obstante, no había sido analizada la confiabilidad del instrumento en esta población.

✖ En Chile, el ASQ fue validado en niños de distintas edades; no obstante, faltaba evidencia acerca de su validez predictiva a más largo plazo.

✖ Si bien hay estudios que analizan la capacidad de los CCDRP para predecir el rendimiento cognitivo o académico en la edad escolar, este conocimiento no había sido sistematizado.

Por lo tanto, el propósito de la tesis es ampliar la evidencia acerca de la validez y confiabilidad de los CCDRP, específicamente del ASQ, y analizar su capacidad predictiva a largo plazo.

Establecidos los tres objetivos generales de cada una de las investigaciones que se entrelazan para dar repuesta al propósito de la tesis, se presenta la metodología utilizada para responder a cada uno de ellos, los resultados y se discute el alcance de ellos en la toma de decisiones y proyecciones.

Los trabajos que constituyen esta tesis son tres:

1. El primero evalúa la validez y confiabilidad del ASQ en español en Barcelona y su capacidad predictiva para niños entre los 2 y 4 años.

2. El segundo analiza la validez predictiva del ASQ validado en Chile, para las dificultades cognitivas en la etapa escolar.

3. El tercero, a través de una revisión sistemática de la literatura, estudia la capacidad del ASQ y otros CCDRP para predecir las dificultades cognitivas o académicas a largo plazo, en población general.

El análisis de los 3 objetivos, en su conjunto, permite la consolidación del conocimiento.

RESUMEN

Antecedentes: Los Cuestionarios de Cribado del Desarrollo Basados en el Reporte de Padres o Cuidadores (CCDRP) han ganado cada vez mayor difusión. De estos, el más conocido a nivel mundial es el Ages and Stages Questionnaire (ASQ), instrumento que se ha utilizado en Cataluña y validado en Chile, existiendo una brecha en cuanto al conocimiento de su capacidad predictiva a más largo plazo.

Objetivo: Determinar la confiabilidad y validez predictiva del ASQ para el cribado de dificultades del desarrollo.

Métodos: Se realizaron tres trabajos de investigación:

1. Una cohorte prospectiva de 321 niños nacidos en el SCIAS - Hospital de Barcelona, con el objeto de evaluar la confiabilidad de los cuestionarios aplicados a los 24 meses de edad corregida y 48 meses de edad cronológica del ASQ-3 en español y la correlación entre ambas evaluaciones.

2. Una cohorte de 227 niños controlados en Clínica Alemana de Santiago de Chile, con el objetivo de estudiar la validez predictiva del ASQ aplicado a los 8 y 18 meses de edad cronológica en niños nacidos a término y edad corregida en prematuros, o de 30 meses de edad cronológica, para detectar los niños que tendrán déficit cognitivo en la etapa escolar.

3. Finalmente se realizó una revisión sistemática de la literatura para analizar la capacidad del ASQ y los CCDRP para predecir bajo rendimiento cognitivo o académico en niños de población general.

Resultados: En Barcelona se evidenció que el ASQ es un instrumento confiable para el cribado del desarrollo psicomotor (alfa de Cronbach para el puntaje global 0,78/0,79), existiendo una correlación positiva y significativa entre las evaluaciones

realizadas a los 24 y 48 meses. La presencia de dos o más dominios en zona de riesgo a los 24 meses fue predictor de déficit del desarrollo a los 48 meses (Odds Ratio (OR) =140 [95% IC 14,85;3575,65]).

En Chile, la capacidad predictiva del ASQ para identificar aquellos niños que tendían menor rendimiento cognitivo en la etapa escolar fue satisfactoria (área bajo la curva (AUC) 0,77 [IC95% 0,65- 0,89]) y comparable con la prueba diagnóstica de referencia [p=0.58]. Considerando al menos un dominio en zona de riesgo, la sensibilidad fue de 66,7% y la especificidad de 71,7%.

Finalmente, a través de una revisión sistemática de la literatura se identificaron ocho cohortes evaluadas con ASQ que reportaron asociaciones en general "positivas" entre la evaluación temprana del desarrollo y el rendimiento cognitivo o académico posterior, con AUC entre 0.66 y 0.87, y OR superiores a 3. Un efecto en espejo se encontró al revisar la sensibilidad y especificidad de las cohortes. La heterogeneidad de los estudios no permitió realizar un metaanálisis ni análisis de subgrupos.

Conclusiones: El ASQ es confiable y tiene una adecuada correlación y capacidad para predecir las dificultades del desarrollo de niños entre los dos y cuatro años en Barcelona. A más largo plazo, en Chile, se demostró que las evaluaciones realizadas a los 8, 18 y 30 meses tienen adecuada capacidad predictiva de las dificultades cognitivas en la etapa escolar.

Estos resultados fueron corroborados a través de una revisión sistemática de la literatura. Se requieren estudios adicionales para conocer el impacto de las distintas adaptaciones, los diferentes contextos socioculturales, factores de riesgo biológicos y las edades de evaluación, en la capacidad predictiva del ASQ, como de los CCDRP.

ABSTACT

Introduction: Parent/caregiver completing developmental screening questionnaires (DSQs) have become increasingly widespread. Worldwide, the most well-known is the Ages and Stages Questionnaire (ASQ), an instrument that has been used in Catalonia and validated in Chile; however, there is a gap in the knowledge of its predictive capacity in the longer term.

Objective: To determine the reliability and predictive validity of the ASQ for the screening of developmental difficulties.

Methods: Three research studies were carried out. A prospective cohort of 321 children born at the SCIAS - Hospital in Barcelona, in order to evaluate the reliability of the questionnaires administered at 24 months of corrected age and 48 months of chronological age of the ASQ-3 in Spanish and the correlation between both evaluations. A cohort of 227 children seen at Clínica Alemana de Santiago, Chile, to study the predictive validity of the ASQ applied at 8 and 18 months of chronological age in full-term children and corrected age in preterm, or 30 months of chronological age, to detect children who will have cognitive deficits at school. Finally, a systematic review of the literature was carried out to analyze the ability of the ASQ, and other DSQs, to predict low cognitive or academic performance in children of the general population.

Results: In Barcelona, it was observed that ASQ is a reliable instrument for psychomotor development screening (Cronbach's alpha for the global score 0.78/0.79), with a positive and significant correlation between the assessments made at 24 and 48 months. The presence of 2 or more domains in the risk zone at

24 months was predictive of developmental deficit at 48 months (Odds Ratio (OR) =140 [95% CI 14.85;3575.65]). In Chile, the predictive ability of the ASQ to identify children with lower cognitive performance at school was satisfactory (area under the curve (AUC) 0.77 [95% CI 0.65- 0.89]) and comparable with the reference diagnostic test [p=0.58]. Considering at least one domain in the risk zone, it presented a 66.7% sensitivity and 71.7% specificity. Finally, through a systematic review of the literature, we identified 8 cohorts evaluated with the ASQ that reported overall "positive" associations between early developmental screening and later cognitive or academic performance, with AUC between 0.66 and 0.87, and OR higher than 3. We observed a mirror effect when reviewing the sensitivity and specificity of the cohorts. The heterogeneity of the studies did not allow meta-analysis or subgroup analysis.

Conclusions: The ASQ is reliable and has an adequate correlation and capacity to predict developmental difficulties between two and four years of age in Barcelona. In the longer term, in Chile, it was demonstrated that the assessments performed at 8, 18, or 30 months have an adequate predictive capacity of cognitive difficulties in the school stage. These results were corroborated through a systematic review of the literature. Further studies are needed to determine the impact of different adaptations, sociocultural contexts, biological risk factors, and assessment ages on the predictive ability of the ASQ, such as parent-repot DSQs.

INTRODUCCIÓN

Los primeros 5 años de vida son reconocidos como un período crítico para la fundación de habilidades cognitivas y aprendizaje escolar [1]. Las deficiencias en el desarrollo psicomotor (DSM) tienen un importante impacto en la sociedad en términos de costos en salud, soporte educacional y servicios de apoyo [2, 3].

Los déficits del desarrollo engloban un grupo heterogéneo de condiciones crónicas definidas por la presencia de dificultades en el funcionamiento físico, conductual o cognitivo, detectadas en los primeros años de vida, y asociadas a discapacidad y/o problemas cognitivos o del aprendizaje en la etapa escolar [4-7]. Se estima que 1 de cada 6 niños podría tener alguna dificultad en el DSM [8-10], siendo esta condición más prevalente en niños que nacen con factores de riesgo ambientales, como es la pobreza multidimensional; o con factores biológicos, como son las condiciones genéticas, la prematuridad y las enfermedades perinatales, condiciones que se potencian entre sí [11-16].

La identificación de los niños con dificultades del DSM representa un verdadero desafío para los profesionales de salud, especialmente en niños que no tienen factores de riesgo, debido a que la impresión clínica es subjetiva y poco sensible para este fin [17, 18]. Por este motivo, la Academia Americana de Pediatría (AAP), al igual que otras sociedades científicas, recomiendan la monitorización del desarrollo mediante escalas estandarizadas en los controles de supervisión de

salud de lactantes y preescolares en forma regular y seriada, debido a la posibilidad de emergencia de dificultades en áreas específicas a lo largo de la infancia [19, 20].

A nivel mundial se han desarrollado distintas pruebas de cribado del DSM, que se diferencian en su adecuación cultural, énfasis y modalidades de aplicación. Entre estas, destacan los cuestionarios de cribado del desarrollo basados en el reporte de padres o cuidadores (CCDRP), que corresponden a los instrumentos recomendados en la última guía de supervisión de salud de la AAP del 2020 [20]: Ages and Stages Questionnaires (ASQ) [21]; Parents' Evaluations of Developmental Status (PEDS) con su actualización PEDS: Developmental Milestones (PEDS:DM) [22, 23] y Survey of Well-being of Young Children (SWYC) [24]. Entre las distintas escalas de cribado del desarrollo, la más utilizada en EEUU es el ASQ [25, 26], instrumento que ha sido adaptado, validado y utilizado en distintos países, incluyendo Chile y España[27].

El ASQ fue desarrollado en la Universidad de Oregón en los años ochenta y actualizado el 2009, como ASQ-3. Este instrumento se ha adaptado y utilizado en varios países, en distintos contextos culturales y socioeconómicos, y en niños con distintos riesgos biológicos [28-31]. Los estudios de validación del ASQ en Chile (ASQ-Cl) incluyeron distintos análisis de confiabilidad y validez convergente y de constructo [29, 31, 32], además de un análisis preliminar de su validez para predecir el coeficiente cognitivo en los primeros años de educación escolar en una muestra de niños de nivel socioeconómico medio alto, atendidos en Clínica Alemana [33].

Por otro lado, en el Hospital SCIAS - Hospital de Barcelona, que también atiende familias pertenecientes al estrato socioeconómico medio alto, se empezó a aplicar la versión del ASQ tercera edición (ASQ-3) en español, traducido por la editorial Brookes Publishing, en niños nacidos prematuros tardíos y niños nacidos a término, demostrándose la validez de constructo del instrumento, al discriminar entre grupos de mayor riesgo de presentar dificultades del desarrollo [28, 30].

En este contexto, el propósito de esta tesis es determinar la validez y confiabilidad del ASQ, como CCDRP y analizar su capacidad predictiva a largo plazo.

Generalidades del Cribado del Desarrollo Infantil

La importancia de la pesquisa temprana del déficit del DSM radica en la posibilidad de implementar una intervención oportuna y efectiva [34-36]. Por este motivo, distintas sociedades científicas recomiendan monitorizar el desarrollo durante los controles de supervisión de salud de lactantes y preescolares, usando instrumentos estandarizados y válidos a edades específicas. Es así como la AAP recomienda la aplicación de escalas de cribado a las edades de 9, 18 y 30 a 36 meses, y luego reevaluar entre los 4 y 5 años [19, 20]. En el programa de salud infantil de Chile se realiza el cribado del desarrollo a las edades de 8, 18 y 36 meses [37]. Por su parte, en el programa de salud de Inglaterra, se aplican las escalas de cribado del desarrollo entre los 24 y 30 meses [38].

Esta recomendación es especialmente relevante en niños de mayor riesgo, como son los prematuros, debido a que este grupo es menos estable en su desarrollo, pudiendo emerger dificultades en la etapa escolar, especialmente en las áreas cognitivas [7, 39–41]. Los prematuros son aquellos niños que nacen antes de las 37 semanas de gestación: se clasifican como tardíos (34-36 semanas y 6 días), moderados (32-33 semanas y 6 días) y extremos (≤32 semanas). Al evaluar el desarrollo DSM en niños prematuros, se recomienda corregir la edad, restando a la edad cronológica, el número de semanas de nacimiento antes de las 40 semanas de gestación [42]. Si bien hay debate al respecto, en prematuros se recomienda la aplicación de las escalas de evaluación de desarrollo según edad corregida hasta los 2 años de vida [43, 44].

En la evaluación comprensiva del DSM se han determinado 3 etapas fundamentales: la vigilancia del desarrollo, la monitorización del desarrollo con escalas de cribado estandarizadas, y, por último, las pruebas diagnósticas de dificultades del desarrollo [19] (Tabla 1).

La Vigilancia del desarrollo es el proceso informal de indagación de preocupaciones de los padres, observación de la conducta y de las habilidades de los niños e identificación de los factores de riesgo, realizada por un profesional durante el control de supervisión de salud [19, 20]. Este es un proceso continuo y flexible, basado fundamentalmente en el juicio clínico, conocimiento y entrenamiento del profesional; con la desventaja de ser poco confiable y carecer de estandarización y de validez [18, 45]. En una revisión retrospectiva, Corrigan y cols demostraron que tan sólo el 6% de los niños con dificultades de aprendizaje habían

sido detectados a través de la vigilancia del desarrollo realizada en la etapa preescolar [46].

Por su parte, la monitorización del desarrollo se realiza a través de un cribado con escalas estandarizadas que pueden ser administradas por un profesional entrenado, o bien basarse en el reporte de padres o cuidadores, como son los CCDRP. Su propósito es identificar en forma objetiva aquellos niños con posibles deficiencias en el DSM, es decir, niños que no adquieren las habilidades en las ventanas de tiempo esperadas [19, 20]. El uso de CCDRP ha incrementado de forma importante en los últimos años a nivel mundial [25, 26, 47], logrando cada vez mayor aceptación entre los profesionales, debido a su bajo costo, el poco tiempo que demanda y el involucramiento de los adultos responsables del cuidado y estimulación de los niños [48, 50].

La AAP y otras sociedades científicas, plantean que, un niño que fracasa en el cribado del desarrollo debe ser evaluado con una escala diagnóstica, con el fin de confirmar o descartar la deficiencia del DSM y establecer un plan de intervención oportuno [19, 20, 51]. A diferencia de las escalas de cribado, las escalas diagnósticas son de mayor costo ya que requieren de una batería de instrumentos que deben ser aplicados por profesionales entrenados. Estas escalas son ampliamente utilizadas tanto en la atención clínica como en la investigación, al considerarse como patrón de referencia en la validación concurrente de las pruebas de cribado [52]. Entre ellas, la más utilizada para la evaluación de niños menores de 4 años es la Escala de Bayley de desarrollo infantil [53].

Tabla 1: Proceso de cribado de dificultades del Desarrollo Psicomotor

PROCEDIMIENTO	GRUPO BLANCO	OBJETIVOS	MÉTODOS
Vigilancia Clínica del desarrollo	Todos los niños en los controles de supervisión de salud.	Promoción de la salud y el desarrollo, del cuidado paterno, identificación de factores de riesgo y elementos de sospecha de dificultades del desarrollo.	Proceso de identificación de niños en riesgo de presentar dificultades en su desarrollo fundamentadas en el juicio clínico y logro de hitos, sin emplear pruebas estandarizadas.
Cribado del Desarrollo	Todos los niños a edades específicas.	Identificar aquellos niños que deberían recibir una evaluación diagnóstica más especializada, verificar las preocupaciones de los padres.	Empleo de escalas estandarizadas para detectar aquellos niños con riesgo de déficit del desarrollo.
Pruebas Diagnósticas	Niños con sospecha de déficit o cribado alterado.	Proveer una detallada descripción y categorización del desarrollo del niño para establecer plan de manejo y seguimiento.	Herramientas estandarizadas de evaluación y diagnóstico de las dificultades del desarrollo.

Figura adaptada de Schonhaut L y Armijo I.
Aplicabilidad del Ages & Stages Questionnaires para el tamizaje del desarrollo psicomotor.
Revista Chilena de Pediatría 2014;85 (1):12-21

Cuestionarios de Desarrollo basados en el Reporte de Padres o Cuidadores

Los CCDRP recomendados por la AAP 2020 [20] son el ASQ, PEDS, PEDS: DM y SWYC. Ni en Chile ni en España existen pruebas validadas de reporte de padres, por lo que, en ambos países se comenzó a utilizar y validar las pruebas recomendadas por la AAP.

El Parents' Evaluations of Developmental Status (PEDS) [22] consiste en una serie de 10 preguntas que indagan sobre la preocupación parental en relación a diferentes aspectos del desarrollo de los niños, incluyendo los dominios de lenguaje expresivo y comunicación, motricidad fina y gruesa, autoayuda, comportamiento, desarrollo socioemocional, desempeño académico y salud en general. Este cuestionario ha sido traducido a varios idiomas y se ha utilizado ampliamente a nivel mundial en población general y muestras clínicas [54, 55]. El año 2008 se validó una versión modificada PEDS: Developmental Milestones (PEDS: DM) que incorpora preguntas específicas acerca del desempeño de los niños [23].

El Survey of Well-being of Young Children (SWYC) [24] es una serie de 12 cuestionarios breves para edades específicas, que incluye los dominios cognitivo, lenguaje y habilidades motoras, validado el 2013 en EEUU. Este cuestionario ha sido validado en poblaciones nativas y, a nivel internacional, fue traducido al portugués [56, 57].

El Ages and Stages Questionaire (ASQ), se encuentra en su tercera validación del 2009 (ASQ-3), que consiste en una serie de 21 cuestionarios para ser aplicados diferencialmente a niños desde los 2 meses hasta los 5 años. Cada formulario consta de 5 dominios, de 6 preguntas cada uno, que evalúan diferentes ámbitos del DSM: como son comunicación, motricidad fina, motricidad gruesa, resolución de problemas y relaciones interpersonales. Al final, hay una sección de 7 preguntas abiertas destinadas a indagar posibles preocupaciones de los padres.

El ASQ ha sido adaptado en varios idiomas, países, contextos culturales y en muestras de niños con distintas características clínicas y demográficas [27, 58–60]. Se ha implementado para el monitoreo de niños con desarrollo típico, y también en el seguimiento de niños con factores de riesgo biológico como son los antecedentes de encefalopatía hipóxico-isquémica o prematuridad [31, 61–65].

El ASQ originalmente fue creado para ser completado por los padres o cuidadores principales, no obstante, de acuerdo con las diferentes realidades, ha sido adaptado con distintas variaciones, como es la asistencia de trabajadores de campo en comunidades de alta vulnerabilidad social, como fue la validación de Colombia, India o Nepal [66-68] y la aplicación por los educadores de nivel preescolar en Brasil [69].Por otro lado, se ha completado en distintos escenarios, como son las visitas domiciliarias, salas de espera, jardines infantiles y por vía electrónica [66, 67, 69, 70], sin que se haya determinado cómo estos contextos influyen en la confiabilidad de las respuestas. Otras actualizaciones del instrumento incluyen formas abreviadas, en que sólo se utilizan algunos de los dominios, o la forma extendida (EASQ), que incorpora más preguntas por dominio [69, 71].

Otro aspecto en discusión es la definición de riesgo de déficit del DSM según ASQ. Si bien la validación original recomienda considerar a un niño en déficit si tiene al menos un área bajo el punto de corte [21], otros investigadores han demostrado una mayor especificidad si se consideran 2 o más dominios bajo el punto de corte o con el puntaje total [33, 72, 73], mientras que si se incorpora la dimensión de preguntas abiertas aumenta la sensibilidad [73].

Precisión de los Cuestionarios del Desarrollo Basadas en el Reporte de Padres

En estudios concurrentes, los CCDRP reportan valores de sensibilidad y especificidad del orden del 70%, lo que cumple con las recomendaciones de la AAP para un cribado del desarrollo [19]. En una revisión sistemática de las validaciones del ASQ 24 -30 meses, Velikonja y cols. reportaron que los estudios de validez concurrente del ASQ mostraron en general valores positivos, siendo estos resultados más consistentes en las versiones que respetan la forma original que en las versiones adaptadas a otros idiomas y culturas [58].

Estudios que comparan el ASQ con el PEDS describen una discordancia substancial entre ambos [74-76]. Por su parte, el SWYC mostró una moderada a alta correlación con todas las formas del ASQ-3 [24]. En una investigación reciente, Sheldrick y cols. compararon los 3 CCDRP, reportando adecuada especificidad, pero baja sensibilidad para la detección de dificultades concurrentes del desarrollo, en una muestra de niños de 9 a 60 meses, sin poder demostrar claras ventajas de uno de los cuestionarios sobre los demás [77].

Son escasos los estudios prospectivos que analizan la capacidad predictiva de los CCDRP para detectar a aquellos niños que tendrán dificultades cognitivas o académicas posteriormente [33, 64, 78, 79]. En una revisión sistemática que analizó la precisión de las escalas de evaluación de comportamiento y lenguaje, Sim y cols. demostraron una excelente validez predictiva, siendo mejor los CCDRP comparados con las escalas administradas en forma directa por profesionales [80], lo

que puede ser explicado porque los padres supervisan constantemente a sus hijos y son quienes mejor los conocen. Por otra parte, se han publicado dos metaanálisis sobre la validez predictiva de pruebas diagnósticas del desarrollo psicomotor en prematuros extremos para identificar a aquellos niños que tendrán problemas cognitivos en la edad escolar, describiéndose baja sensibilidad, pero elevada especificidad y correlación, especialmente de los dominios mentales [81, 82].

A pesar de la amplia evidencia que respalda el uso de CCDRP, aún no se han incorporado en la actividad rutinaria de los pediatras ni en los programas de supervisión de salud, salvo en Estados Unidos e Inglaterra [20, 38]

La evaluación del Desarrollo Psicomotor en Chile y en España

En Chile, desde 1990, existe un Programa de Estimulación y Evaluación del DSM, el cual cubre a los menores que se atienden en el sistema público de salud. Este programa considera la evaluación del desarrollo motor y cognitivo mediante dos pruebas desarrolladas en Chile en la década de 1970 a 1980: la Escala de Evaluación del Desarrollo Psicomotor (EEDP) y el Test de Desarrollo Psicomotor (TEPSI), aplicadas por enfermeras u otros profesionales capacitados, en los controles de salud de los 8, 18 y 36 meses [37]. En la actualidad se está validando un nuevo instrumento, el Test de Aprendizaje y Desarrollo Infantil (TADI), que también debe ser aplicado por profesionales entrenados.

Por otro lado, en los controles de supervisión de salud del sector privado, es el pediatra quien realiza la vigilancia del DSM, generalmente basado en su juicio clínico [17].

A diferencia de Chile, en España no hay políticas universales en relación con el cribado del desarrollo. La tabla Haizea-Llevant fue validada a partir del estudio del Dr. Emilio Fernández Álvarez y ha sido cada vez más utilizada en atención primaria, como medida objetiva de pesquisa de niños de 0 a 5 años con posibles dificultades del desarrollo [83]. Hace unos años, el grupo de seguimiento de prematuros tardíos de la Sociedad Española de Neonatología (SEN-34-36), recomendó la aplicación del ASQ-3 en español al menos a los 2 años de vida [44].

En la Clínica Alemana de Santiago y en SCIAS - Hospital de Barcelona de Cataluña, se siguieron cohortes para evaluar el ASQ como instrumento de detección de dificultades del desarrollo en niños nacidos a término y prematuros. Ambos establecimientos tienen en común que son instituciones privadas que atienden a una población de nivel socioeconómico medio alto, con padres, en su mayoría, profesionales con estudios universitarios. Ambos hospitales atienden un número de partos entre 2000 y 3500 anuales, y cuentan con un programa de seguimiento para los prematuros, en que se ha implementado el ASQ, con gran adherencia por parte de las familias.

Validación del ASQ en Santiago de Chile y Barcelona

Entre los años 2008 y 2011 se realizó una adaptación a la lingüística chilena del ASQ-3, para las edades de 8 y 18 meses. Para conocer la validez y confiabilidad del ASQ, se incorporaron 1896 niños atendidos en distintos centros del sistema público y privado de Atención de Salud. Al comparar los puntajes medios de la muestra chilena con la versión original del ASQ-3, destacaron diferencias significativas solo en las áreas de motricidad gruesa a los 8 meses, y resolución de problemas a los 18 meses. El ASQ mostró tener una buena consistencia interna, con una alfa de Cronbach de 0,66 – 0,85, tanto para sus dominios como para el puntaje total. La estabilidad temporal y acuerdo interjuez fueron alta (r Pearson´s 0,73 – 0,94; correlación intraclase 0,68 – 0,93). En el análisis de validez de constructo, se encontró que los prematuros extremos presentaron significativamente mayor tasa de fracaso en el ASQ en todos los dominios comparados con los niños nacidos a término y hubo diferencias significativas con los niños con retraso del desarrollo conocido [29].

En una sub-muestra de 306 niños de 8, 18 y 30 meses de distinta edad gestacional, atendidos en Clínica Alemana, se estudió la validez concurrente del ASQ-Cl, tomando como referencia la escala de Bayley-III, destacando una sensibilidad de 73% y una especificidad de 81% [31]. En un primer análisis de la capacidad del ASQ para predecir el rendimiento cognitivo en los primeros años de educación escolar, se incluyeron 174 evaluaciones (correspondientes a 123 niños), se reportó un valor predictivo positivo que varió, según la edad de evaluación, entre 21,4 y 45,5%, mientras que el valor predictivo negativo fue 92,0-93,2% [33].

Por su parte, en el SCIAS - Hospital de Barcelona, se comenzó a aplicar en el año 2013 el ASQ-3 en español en una cohorte de niños de 2 y 4 años nacidos prematuros tardíos y a término. Para conocer la validez de constructo se analizó la densidad de probabilidades de riesgo de déficit del DSM de acuerdo con el puntaje total, encontrado diferencias significativas entre los grupos analizado [28]. En el análisis por dominio destacó el antecedente de patología neonatal, específicamente los problemas respiratorios, se asociaron a mayor prevalencia de déficit en el área del lenguaje [30]. Sin embargo, no había sido estudiada la validez y confiabilidad del ASQ-3 en esta población.

Justificación

La importancia de contar con pruebas de cribado del DSM para niños menores de 5 años se basa en la neuroplasticidad cerebral, que nos da la oportunidad de intervenir tempranamente y minimizar los daños a largo plazo. La gran ventaja de las CCDRP se basa en su versatilidad, costo-eficacia, factibilidad y empoderamiento de los adultos responsables [29, 48, 49, 84]. Entre estos cuestionarios, el más utilizado y validado a nivel mundial es el ASQ [27, 59, 60], habiéndose iniciado investigaciones paralelamente en Santiago de Chile y Barcelona, Cataluña [28, 31].

Contar con fundamentos acerca de la confiabilidad y validez del ASQ en distintos contextos clínicos y culturales, y conocer su capacidad predictiva a largo plazo es fundamental para la toma de decisiones políticas y clínicas en cuanto a su uso.

Si bien el uso del ASQ es cada vez más difundido a nivel internacional, aún hay algunas brechas de conocimiento, entre las cuales, para fundamentar este trabajo de tesis, se consideraron las siguientes:

1. El ASQ ha sido utilizado en Barcelona, para conocer el pronóstico de los niños que nacen prematuros tardíos con y sin patología perinatal. No obstante, no había sido analizada su confiabilidad en esta población.

2. En Chile, si bien el ASQ fue validado en niños de distintas edades, faltaba evidencia acerca de su validez predictiva.

3. No había suficiente evidencia acerca de la capacidad predictiva a largo plazo del ASQ y demás CCDRP recomendados por la AAP en población general.

En este contexto, se ha formulado la siguiente pregunta de investigación como hilo conductor para este trabajo de tesis, desarrollado como compendio de 3 publicaciones:

¿Cuál es la confiabilidad y validez del ASQ en muestras de bajo riesgo biológico y socioeconómico en Cataluña y Chile?

¿Cuál es la capacidad del ASQ y otros CCDRP para identificar aquellos niños con mayor riesgo de presentar dificultades académicas o cognitivas a largo plazo?

Consistentemente, el propósito de la tesis es determinar la validez y confiabilidad del ASQ, como CCDRP, y analizar su capacidad predictiva a largo plazo, en población general.

OBJETIVOS

Objetivo general Estudio Nº1:

Evaluación de la confiabilidad de los cuestionarios de 24 y 48 meses del ASQ-3 en español, en una muestra de nivel socioeconómico medio alto de un Hospital de Barcelona

Objetivos específicos

1. Comparar la confiabilidad del ASQ aplicado en niños que nacen prematuros y niños nacidos a término.

2. Analizar la correlación entre las evaluaciones de ASQ completadas a los 24 y 48 meses.

3. Analizar la capacidad del ASQ completado a los 2 años, de predecir dificultades del desarrollo a los 4 años.

Objetivo General Estudio Nº2:

Estudiar la validez predictiva del ASQ-3 aplicado a los 8, 18 o 30 meses en una muestra de nivel socioeconómico medio-alto de Santiago de Chile para detectar niños que tendrán déficit cognitivo en la etapa escolar.

Objetivos específicos

1. Comparar la capacidad predictiva del ASQ-3 con el Bayley-III, considerado patrón de referencia para el diagnóstico de dificultades del desarrollo psicomotor.

2. Identificar los dominios predictores del desempeño cognitivo, tanto en ASQ-3 como en el Bayley-III.

3. Comparar la validez predictiva del ASQ según edad de evaluación.

Objetivo General Estudio Nº3:

Determinar la capacidad del ASQ y otros CCDRP para predecir bajo rendimiento cognitivo o académico a largo plazo, en niños de población general.

Objetivos específicos

1. Conocer la literatura publicada acerca de validez predictiva de los CCDRP.

2. Evaluar las formas en que los distintos estudios analizan la capacidad predictiva de los cuestionarios.

METODOLOGÍA

Esta tesis se ha configurado como compendio de 3 publicaciones. La metodología, que se detalla a continuación fue la utilizada para responder los objetivos de cada una de las investigaciones, que se tradujo en tres publicaciones.

Metodología Estudio N°1: Evaluar la confiabilidad de los cuestionarios de 24 y 48 meses del ASQ-3 en español en una muestra de nivel socioeconómico medio alto de un hospital de Barcelona

Estudio de cohorte prospectiva de 158 niños nacidos a término y 163 prematuros tardíos (edad gestacional de 340/7 a 366/7 semanas) en el SCIAS - Hospital de Barcelona. Se siguieron 2 cohortes: la primera de niños nacidos entre el 1 de enero y 31 de diciembre 2009, y la segunda de niños nacidos entre el 1 de enero y 31 de diciembre 2011.

Criterios de inclusión

Se seleccionó una muestra pareada por fecha de nacimiento de niños nacidos a término con los prematuros tardíos. Se incluyó solo niños aparentemente sanos, seguidos en sus controles por los mismos pediatras de la institución de salud.

Criterios de Exclusión

Se excluyeron los recién nacidos que presentaran síndromes mal formativos, enfermedad genética o metabólica conocida. En el grupo de los nacidos a término se excluyeron además los niños que requirieron hospitalización en unidad de cuidados intensivos en el periodo neonatal.

Procedimiento

A las edades de 4 años, en la primera cohorte, y 2 y 4 años en la segunda cohorte, se solicitó a los padres o cuidadores principales de los niños que completaran el formulario de ASQ-3 traducido al español por la editorial Brookes Publishing.

La evaluación a los 2 años se realizó según edad corregida en los prematuros y según edad cronológica en los niños nacidos a término. La evaluación a los 4 años se consideró de acuerdo a la edad cronológica.

Los padres completaron el formulario correspondiente en sus casas. Un funcionario de salud llevó el cuestionario a los padres y los instruyó sobre el procedimiento necesario para completarlo. Todos los padres completaron el Consentimiento Informado. El Proyecto fue aprobado por el Comité de Ética de Investigación del Hospital de Barcelona.

Análisis estadístico

Se realizó un análisis descriptivo de la muestra. Se calcularon promedios y desviación estándar (DS). Estos valores se compararon con los publicados en la validación original del ASQ-3. Se consideró significativa una diferencia mayor a 5 puntos, considerando que el ASQ es una escala de rango, cuyos puntajes van de 5 en 5 puntos [85].

El cribado fue considerado positivo si el rendimiento era inferior a -2DS en al menos un dominio de los evaluados, de acuerdo con las recomendaciones del manual de utilización del ASQ.

Para determinar la confiabilidad, se estudió la consistencia interna a través del coeficiente de alfa de Cronbach y la Correlación de Pearson para cada uno de los dominios y el puntaje total. Se analizó la capacidad predictiva del ASQ-3 evaluado a los 2 meses para predecir menor rendimiento a los 48 meses, usando Odds Ratio (OR) y análisis de regresión logística. Se compararon los resultados en niños nacidos a término y prematuros tardíos.

Los análisis fueron realizados sobre la plataforma R. El flujograma metodológico del estudio 1 se puede ver en la siguiente figura:

Ilustración 1: Flujograma metodológico Estudio 1
PT: prematuros tardíos, AT: a término, OR: Odds Ratio

Metodología Estudio N°2: Evaluar la validez predictiva del ASQ-3 aplicado a los 8, 18 o 30 meses, para detectar a niños que tendrán déficit cognitivo en la etapa escolar en Chile

Estudio longitudinal, de seguimiento de una muestra de 306 niños atendidos en una Clínica privada de Santiago, Chile. Los participantes fueron reclutados entre abril 2008 y abril 2011. Se definió una muestra por cuoteo de niños nacidos a término y prematuros. A la edad de 8, 18 o 30 meses, los niños fueron evaluados con ASQ-3 validado para Chile, y en forma concurrente se les aplicó la escala de Bayley-III [31]. Las evaluaciones se realizaron a los 8 y 18 meses de edad corregida en prematuros, y según edad cronológica en niños nacidos a término, y a los 30 meses cronológicos para todos los niños. Los niños fueron posteriormente evaluados entre los 6 y 9 años de vida con la Escala de Inteligencia para Niños de Wechsler (WISC-III), validada en Chile [86].

Criterios de inclusión

Niños aparentemente sanos, sin antecedentes de enfermedad genética o neurológica diagnosticada.

Criterios de exclusión

Historia de patología intercurrente que pueda haber afectado la trayectoria del desarrollo, como, por ejemplo: meningitis, patología tumoral del SNC y/o accidente con traumatismo encéfalo craneano severo. Además, se excluyeron los niños que tenían más de 9 años al momento de ser contactados.

Procedimiento

El WISC-III fue aplicado por psicólogas capacitadas en dicha evaluación y ciegas a los antecedentes clínicos del niño, como la edad gestacional y los resultados obtenidos previamente en las evaluaciones de ASQ-3 y la escala de Bayley-III. Las evaluaciones fueron realizadas en una consulta o en el domicilio de los pacientes, y se les entregó a los padres un informe del desempeño logrado por los niños. Los niños que debido a dificultades severas del desarrollo no puedan ser evaluados a través del WISC, fueron considerados como positivos para déficit cognitivo.

El estudio fue aprobado por el Comité de Ética de la Facultad de Medicina Clínica Alemana Universidad del Desarrollo. Los padres completaron un Consentimiento Informado y se les dio una retroalimentación de la evolución de los niños.

Análisis estadístico

Se construyeron curvas ROC (Receiver Operating Characteristic o Característica Operativa del Receptor) y áreas bajo la curva (AUC) para déficit cognitivo, de acuerdo con el total de evaluaciones y segmentados según la edad de aplicación de las pruebas (8, 18 y 30 meses). Para comparar las AUC se aplicó el método de DeLong. Mediante análisis de regresión, se evaluó la capacidad de los cuestionarios para predecir el déficit cognitivo. Los análisis fueron realizados con la plataforma R, con la implementación del módulo pROC.

El flujograma metodológico del estudio 2 se puede ver en la figura:
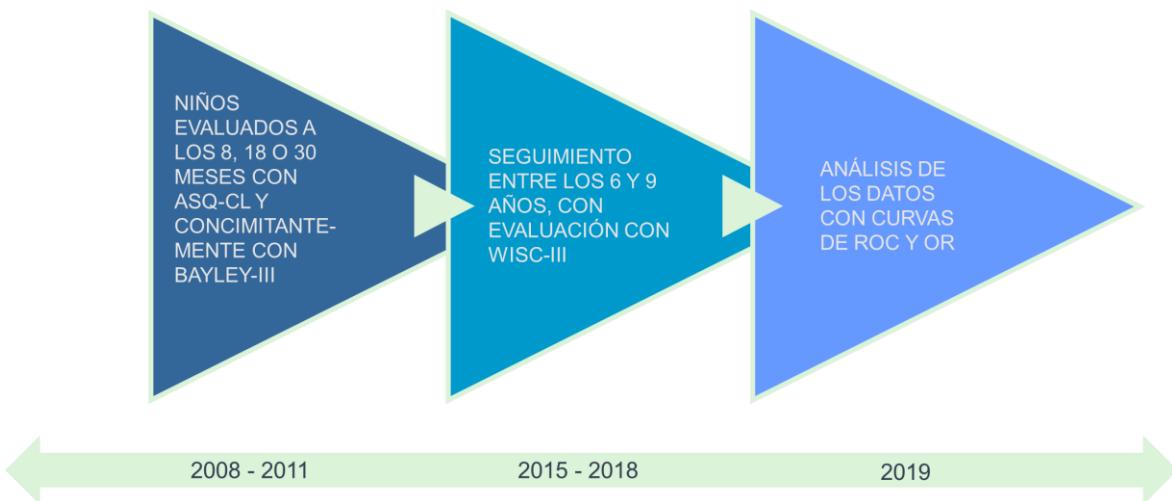
Ilustración 2: Flujograma metodológico Estudio 2
ASQ-CL: Ages and Stages 3ª edición, ASQ-3 validado para Chile,
ROC: Característica Operativa del Receptor, OR: Odds Ratio

Metodología Estudio N°3: Determinar la capacidad del ASQ y otros CCDRP para predecir las dificultades cognitivas o académicas a largo plazo, en población general.

Estudio de revisión sistemática de la literatura.

Fuente de los datos

Se realizó una búsqueda sistemática en las bases de datos de Cochrane, Medline Pubmed, CINAHL, Embase, Web of Science, Scielo y Scopus databases (hasta el 5 de junio del 2020) para identificar la literatura publicada. Se consideraron los siguientes términos para identificar la población: *Infant, Child, Preschool*. Para identificar manuscritos publicados en relación al test índice: *Surveys and Questionnaires, Developmental screening, Ages and Stages, Parents Evaluation of Developmental Status, Survey of Wellbeing of Young Children, parents' evaluation*. Y para identificar el test de referencia: *Intelligence Test, Developmental Disabilities, Intellectual Disability, Intelligence, Academic Performance, intellectual quotient*.
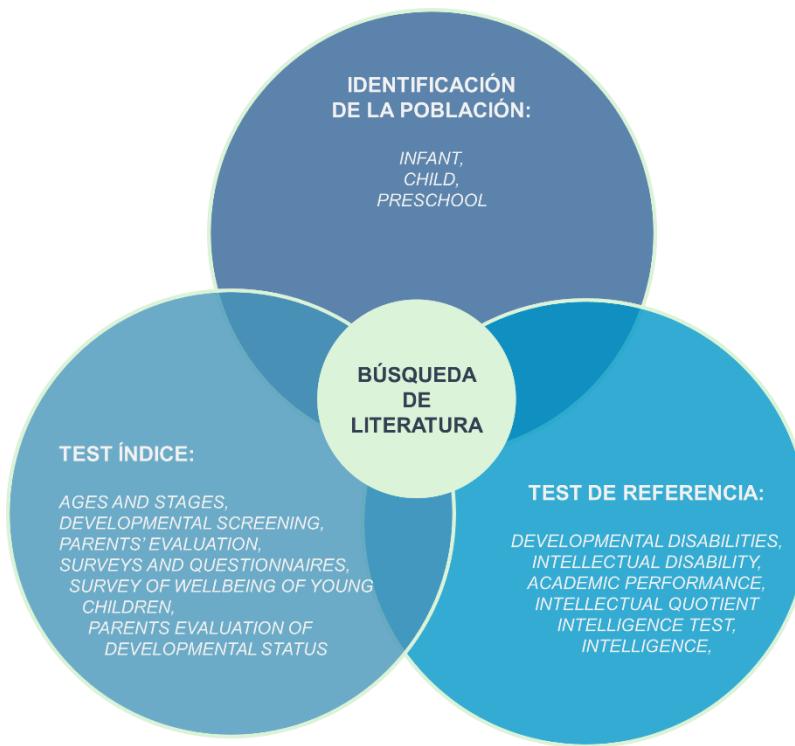
Ilustración 3: Diagrama ilustrativo de los criterios incluidos en la búsqueda sistemática de la literatura para el Estudio 3

Selección de los estudios

En la primera selección se eligieron 3 tipos de estudios:

1. Estudios que mostraban una evaluación con CCDRP en niños menores de 5 años.

2. Estudios que describieran evaluaciones cognitivas y/o académicas.

3. Estudios que describían la asociación entre el resultado del CCCRP y el desempeño académico o cognitivo posterior.

Se incluyeron estudios publicados en español, inglés u otra lengua romance.

A partir de los estudios seleccionados se identificaron cohortes que describieran evaluaciones seriadas con CCDRP y evaluaciones académicas o cognitivas en la edad escolar (en niños de 5 o más años y con la condición de que haya pasado al menos 1 año desde la evaluación con CCDRP).

Para identificar las cohortes, se realizó un cribado de los títulos y resúmenes de los estudios obtenidos a partir de la búsqueda de manuscritos. Luego se revisaron las citas de los manuscritos seleccionados y se revisó dónde estaban citados. Para completar la búsqueda, se conectaron los autores de los CCDRP, a modo de identificar estudios adicionales que cumplieran con los criterios de inclusión.

Extracción de datos

Se construyeron fichas para la extracción de los datos. Se contactó a los autores de cada una de las cohortes para corroborar la información extraída y completar los posibles datos faltantes.

La evaluación de riesgo de sesgos y análisis de la aplicabilidad de los resultados de las cohortes se realizó con la escala de Quality of Diagnostic Accuracy Studies versión 2 (QUADAS-2).

Análisis de los datos

Se realizó un análisis cualitativo de los datos obtenidos. Cuando fue posible, se analizaron las medidas de asociación y, en base a los datos de sensibilidad, especificidad y valores predictivos, se construyó una tabla de 2 x 2 y se calculó el SROC y Forest Plot, usando el programa Rev-Man 5.0.

El estudio fue registrado en PROSPERO (International Prospective Register of Systematic Reviews) con el número de registro CRD42020183883.

El flujograma metodológico del estudio 3 se puede ver en la siguiente figura:

Ilustración 4: Flujograma metodológico Estudio 3

CCDRP: Cuestionarios de Cribado del Desarrollo basados en el Reporte de Padres o Cuidadores, OR: Odds Ratio, ROC: Característica Operativa del Receptor, S: Sensibilidad, E: Especificidad:

RESULTADOS

A continuación, se resumen los principales resultados de cada uno de los estudios incluidos en la tesis (presentados en los anexos).

Resultados Estudio N°1: Evaluar la confiabilidad de los cuestionarios de 24 y 48 meses del ASQ-3 en español, en una muestra de nivel socioeconómico medio alto de un Hospital de Barcelona

La primera cohorte incluyó 179 niños evaluados a la edad de 4 años. La segunda cohorte incluyó 152 niños evaluados a los 2 y 4 años. Para el estudio se consideraron 473 evaluaciones realizadas en 331 niños.

Los puntajes de ASQ-3 24 y 48 meses de la muestra de Barcelona (BCN) fueron comparables con la muestra normativa de la validación original [21], salvo motricidad fina, que fue inferior a los 24 meses y superior a los 48 meses.

El coeficiente alfa de Cronbach fue bajo para cada uno de los dominios analizados en forma independiente, pero aceptablemente bueno para el puntaje global (0,78/0,79). El coeficiente de correlación de Pearson fue consistentemente positivo entre dominios y puntaje total, mientras que la correlación entre los distintos dominios fue significativa, pero baja.

La correlación entre los puntajes de 24 y 48 meses fue positiva y significativa para el puntaje total y para la mayoría de los dominios, siendo más significativa en el grupo de PT que en los AT. Tener al menos un dominio del desarrollo en rango de déficit a los 24 meses no se correlacionó con menores puntajes o riesgo de déficit a los 48 meses; no obstante, la presencia de 2 o más dominios en rango de déficit a los 2 años, fue predictor de déficit del desarrollo a los 48 meses, con un OR de 140 [95% IC 14,85;3575,65], no se encontró diferencia por dominio, sino que en la suma de ellos.

ESTUDIO 1:

Schonhaut L., Martinez Nadal S., Demestre X., Armijo I.

Reliability and agreement of ages and stages questionnaires®

Results in late T preterm and term-born infants at 24 and 48 months

Factor de Impacto:

2.079

Reliability and agreement of ages and stages questionnaires®: Results in late preterm and term-born infants at 24 and 48 months

Luisa Schonhaut[a,*], Silvia Martinez-Nadal[b], Ivan Armijo[c], Xavier Demestre[b]

[a] Clinica Alemana, Faculty of Medicine, Universidad del Desarrollo, Santiago, Chile
[b] SCIAS, Hospital de Barcelona, Barcelona, Spain
[c] Faculty of Psychology, Universidad del Desarrollo, Santiago, Chile

ABSTRACT

Aim: To evaluate the reliability of the Ages and Stages Questionnaires (ASQ-3) 24 and 48 month intervals translated to Spanish by Brookes Publishing, and the agreement between both questionnaires, comparing late preterm (LPI) and term-born infants (terms).

Methods: Two cohort samples of healthy LPI and terms that were born in a private hospital in Barcelona, Spain. Internal consistency was analyzed by Cronbach's alpha scores and Pearson product- moment correlation between the domain scores and the overall score. The agreement was analyzed using Pearson's correlations between the two questionnaires, and the odds ratio (OR) for positive screening at 48 months, given a positive screen in 24 month assessment.

Results: A total of 473 evaluations were analyzed, representing 331 children. Cronbach's alpha scores for the motor domains on both intervals were low, but acceptable compared with the overall score; a strong positive correlation between the domain and overall score were obtained in the majority of the domains. The correlation between the 24 and 48 month total scores were positive, especially for LPI. Having at least 2 domains in the referral zone at 24 months was associated with an OR of 140 [95% CI 14.85; 3575.65] for positive screen at 48 months.

Conclusion: The Spanish ASQ-3 24 and 48 month intervals appear to be a reliable for developmental screening and for the follow-up of children, especially for LPI. Having two or more domains in the referral zone at 24 months screening is a significant predictor of developmental delay risk at 48-month questionnaire.

1. Introduction

Early childhood neurological development is a determining factor in behavior and learning throughout life [1]. There is evidence suggesting that early detection of developmental delay (DD) allows for timely, effective intervention [2,3]. Early intervention can allow improving the learning and behavior alterations that LPI can present, modifying the natural history of these alterations [4,5]. The identification of children with subtle DD represents a real challenge for pediatricians, as it has been shown that clinical impression is subjective and not sensitive to this target [6]. For this reason, the American Academy of Pediatrics (AAP) recommends the application of standardized developmental screening tools during the health monitoring of infants and preschool children at specific ages (9, 18, and 30–36 months), and especially when certain well-known risk factors are at play, such as premature birth, and genetic or metabolic problems [7]; while in European

countries, like the Healthy Child Programme of England, recommend the application of screening tests and developmental surveillance, by the time the child is one year old and between two and two-and-a-half years old [8].

Despite the existence of health monitoring policies that emphasize the importance of standardized screening tools for developmental assessment, fewer than 50% of pediatricians in the United States (U.S.) routinely use them [9]. A number of development screening tools have been developed; they may be distinguished by the conditions they assess, their cultural adaptation, and their means of application. Of note among these are the parent-completed developmental screening tools, which have seen increasing use given their low cost, their ease of use, and the empowerment they provide to the adults responsible for the care and stimulation of the children [10,11].

One of the parent-completed developmental screening tool in wide use in the U.S, is the *Ages and Stages Questionnaires* (ASQ) which was

---

* Corresponding author.
  E-mail address: lschonhaut@alemana.cl (L. Schonhaut).

developed at the University of Oregon in the 1980s and then revised in 2009, as the ASQ3 [12,13]. These questionnaire has been adapted for use in several countries [14,15,16,17,18,19,20] including the region of Galicia in Spain, were the ASQ completed by the educators was studied [21]. The ASQ has been shown to be valid and reliable both in screening children born at full term and in the monitoring of children with biological risk factors [22,23].

In primary care in Spain, the pediatrician is responsible for monitoring child development in health controls, following clinical protocols [24], as a result of which there is a risk of under-diagnosis of children with subtle problems of development, as it has been noted in published reports6. The late preterm infant (LPI) follow-up group of the Spanish Neonatology Society (SEN34-36) has proposed the use of simplified neurodevelopment questionnaires, with the use of ASQ-3 at two years of age [25].

In the Hospital de Barcelona-SCIAS, the ASQ-3, translated into Spanish by Brookes Publishing, was introduced for use in 2013 in an LPI cohort at 24 and 48 months and for term-born infants. We selected the age of 2 years based on the recommendation other European countries, like England [8], and the policies of SEN34–36 in Spain [25]. The follow up at 48 months was chosen based in fact that this age is considered the last step for the detection of subtle delays that can benefit from intervention [26]. In our previous studies it was shown that the ASQ-3 allowed for the identification of those at risk [27,28], however, no reliability studies for the ASQ-3 had been carried out in this population. In addition, few studies have analyzed the effects of biological risk factors, such as premature birth, on the psychometric properties of the test and on the developmental pathways of the children.

The aim of the present study, then, was to evaluate the reliability of the Ages and Stages Questionnaires (ASQ-3) 24 and 48 month intervals translated to Spanish by Brookes Publishing, and the agreement between both questionnaires, comparing LPI and terms, in an upper middle class sample in Barcelona, Spain.

2. Methods

2.1. Population

A cohort study was carried out including LPI (GA of $34^{0/7}$ to $36^{6/7}$ weeks) and term-born infants (GA of $37^{0/7}$ to $41^{6/7}$ weeks) born in the private hospital of a healthcare insurance company with a neonatal intensive care unit, in Barcelona, Spain. The socio-economic level of the participants was considered middle-high due to the social characteristics of the population analyzed, assuming that this population could afford an expensive private healthcare center.

We followed two cohort samples: the first of children born from January 1 to December 31, 2009, and the second of children born from January 1 to December 31, 2011. Inclusion criteria were: LPI and term born in the period whose parents were locatable and, after phone contact, agreeable to participation. For the term group, we selected a sample of children born in the hospital at full- term gestational age matched by date of birth with LPI. We included only apparently healthy term-born infants who were followed up by pediatricians belonging to our insurance group.

Exclusion criteria were children with malformative syndromes and with known genetic or metabolic diseases and, in the term group, we also excluded those who were admitted to Neonatal Intensive Care Unit during the neonatal period.

2.2. Measures

Ages and Stages Questionnaires® Third Edition translated to Spanish by Brookes Publishing (ASQ-3) [13] is a validated, parent-completed developmental screening tool. Twenty-one questionnaires are available from 1 to 66 months of age. Parents answer 30 questions covering five domains of development, including communication, gross motor, fine motor, problem solving, and personal-social domain. Each domain contains six questions that can be answered with a 'yes' (10 points), 'sometimes' (5 points), or 'not yet' (0 points); which are summed for a domain total.

Infants were one standard deviation below the mean in any domain were considered on the monitoring zone and if they have two standard deviations below the mean in any domain have positive screen and were considered in referral zone, or at risk of DD, in accordance with the ASQ manual [13].

Parents were instructed to try activities with their children to facilitate accurate evaluation and were offered the option of completing the questionnaire by e-mail or with a home visit by a person other than a healthcare professional. The healthcare professional brought the document to parents and provided advice as needed.

Parents of included preschoolers completed the ASQ-3 48 month interval for the first cohort, and the 24 and 48 month intervals for the second cohort.

All the parents signed an informed consent form. The project was approved by the hospital's teaching and ethics commission.

2.3. Statistical analysis

For descriptive analysis we first assessed the background characteristics of the study samples. We compared the demographic characteristics of the LPI and term groups using central tendency measures and proportions.

We compared mean scores and cutoff scores of ASQ-3 24 and ASQ-3 48 months with those from the U.S. normative data [13]. A difference of > 5 points was considered significant, considering that the ASQ-3 is a range scale with scores increasing in intervals of five [16].

For reliability analysis, internal consistency was measured with the Cronbach alpha coefficient for each of the five domains and for the overall test. Criteria for Cronbach's alpha included: > 0.9 Excellent; > 0.8 Good; > 0.7 Acceptable; > 0.6 Questionable; > 0.5 Poor; and, < 0.5 Unacceptable [29]. Additionally, the Pearson product-moment correlation coefficients between the overall score and the total scores for each domain were analyzed.

The agreement between the two interval questionnaires was measured separately for the LPI and terms; we analyzed the correlations of the scores between both intervals separately, and, finally, we examined the capacity of positive screen at 24 month to predict poorer performance at 48 months, using logistic regression analysis.

3. Results

1) General information

The first cohort included 179 children who were assessed with ASQ-3 at the age of 48 months. In the second cohort a total of 152 children were assessed at 24 months; of these, 142 were followed up and reassessed at the age of 48 months. The total, then, consisted of 473 evaluations made of 331 children (see Fig. 1).

The bio-demographic characteristics of the sample are presented in Table 1. In the LPI group the median gestational age was 36 weeks (range 34–36) while in the term group it was 39 weeks (range 37–41). There were significant differences in birth weight; 61.9% of the LPI group were hospitalized in the Neonatal Intensive Care Unit while no children from the term group were admitted (in accordance with the exclusion criteria for the study), and the percentages of caesarean sections and twin births were significantly higher for the LPI group compared to the term group. The 84,5% of mothers of the term sample and 77,3% of the LPI sample report university education.

2) Comparison of study sample ASQ-3 mean scores with published U.S. normative data
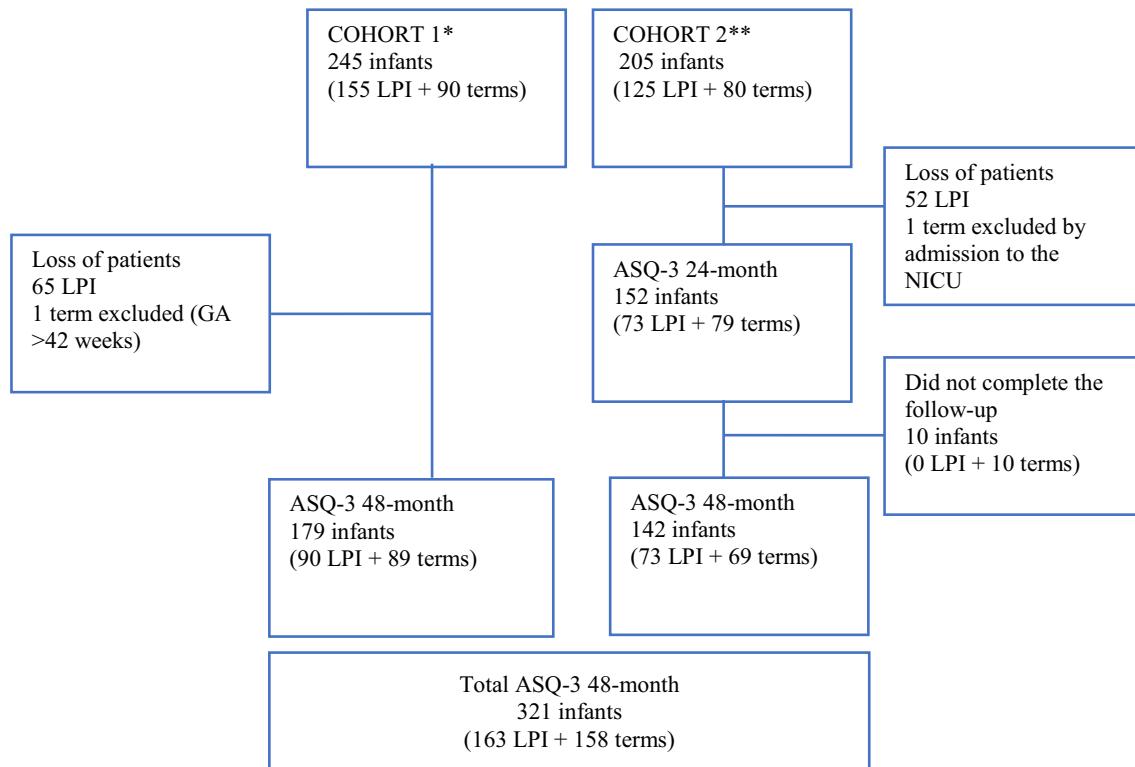
Fig. 1. Flowchart showing the number of eligible children in the final study population.
*COHORT 1: infants born at term and late preterm between 1st January and 31st December 2009.
**COHORT 2: infants born at term and late preterm between 1st January and 31st December 2011.
LPI: late preterm infants; terms: term-born infants; NICU: Neonatal Intensive Care Unit.

Table 1
Analysis of bio-demographic and perinatal variables of the sample.

	Terms	LPI	P
N	168	163	
Birth weight (g) X ± SD	3260.8 ± 421.0	2464.71 ± 420.2	< 0.0001
Gestational age (w) (median and rank)	39,2 (37–41)	36 (34–36)	< 0.0001
Male gender (%)	96 (57.1)	96 (58.9)	NS
Mother's age (y) X ± SD	37.2 ± 3.5	38.2 ± 4.3	0.03
Father's age (y) X ± SD	38.6 ± 5.2	39.6 ± 5.2	0.08
University mother (%)	142 (84.5)	126 (77.3)	NS
Single parent (%)	3 (1.8)	5 (3.1)	NS
Caesarean section (%)	71 (42.3)	100 (61.3)	0.0004
Twins (%)	3 (1.8)	64 (39.3)	< 0.0001
Neonatal Intensive Care Unit admission (%)	0 (−)	101 (61.9)	< 0.0001

Terms: term-born infants. LPI: late preterm infants. P = p-value.
X = mean; SD = standard deviation.

Table 2 presents the scores and the cutoff for ASQ-3 24 and 48 months for the Barcelona sample in comparison to the U.S. normative data [13]. The scores for ASQ-3 at 24 months are similar to those for the U.S., except for the fine motor domain, which were lower. In contrast, at 48 months the scores for the fine motor domain were higher than those reported in the U.S. normative data. There were no differences in the other domains. However, important differences in cutoff scores were found in fine motor domain at 24 months and in all domains at 48 months, due to the lower standard deviation in Barcelona sample compared with U.S. sample. Cutoff scores were higher in Barcelona sample at 48 months.

It will be noted that in both samples the scores between the 24 and 48 month questionnaires showed a rising trend, except for the gross motor domain in the study sample, which remained stable, and the fine

motor domain in the U.S. normative data, in which the scores decreased.

3) Reliability: Cronbach's alpha coefficient

The Cronbach alpha coefficient for the sample's ASQ-3 24 and 48 month scores was computed for each domain and for the overall score. Item deletion did not improve alpha coefficients in any domain. Scores ranged from 0.35 (gross motor at 48 months) to 0.68 (communication at 24 and 48 months). These results indicate that internal consistency ranged from unacceptable to questionable in all domains. But when the overall analysis was made, we obtained Cronbach alphas of 0.78 and 0.79, which indicate acceptable-to-good internal consistency (Table 3).

4) Reliability: Pearson product- moment correlation coefficients between the domain scores and the overall score for each questionnaire

Fig. 2 shows the Pearson product-moment correlation coefficients between the overall and the total scores for each domain, for both ASQ-3 at 24 and 48 months. A strong positive correlation between the domain scores and overall score was obtained, except for fine and gross motor, where the correlations were moderate. The correlations between the various domains was significant, but weakly so. We did not find differences in the correlation pattern between domains and the overall score when we reviewed the ASQ-3 24 and 48 months form separately.

5) Agreement: Correlational analysis at 24 and 48 months

The correlation between the scores for ASQ-3 24 and 48 months was positive and significant for the overall score and for the majority of the

Table 2
Comparison of study sample ASQ-3 mean scores and cutoff scores at 24 and 48 months with US normative data.

		n	Communication		Gross motor		Fine motor		Problem solving		Personal-social	
			Mean/SD	Monitoring zone/ referral zone	Mean/SD	Monitoring zone/ referral zone	Mean/SD	Monitoring zone/ referral zone	Mean/SD	Monitoring zone/ referral zone	Mean/SD	Monitoring zone/ referral zone
ASQ 24 months												
Barcelona sample	ASQ-3 Spanish	152	50 ± 11.4	38.6/27.2	53 ± 7.6	45.4/37.8	**45.5 ± 10**	**35.5/25.5**	48.8 ± 9.6	39.2/29.6	48.8 ± 9.6	39.2/29.6
USA 2009 [13]	ASQ-3	1443	51.2 ± 13	38.2/25.2	54.7 ± 8.3	46.4/38.1	**51.7 ± 8.2**	**43.4/35.2**	49 ± 10	39.6/29.8	51 ± 10	41.3/31.5
ASQ 48 months												
Barcelona sample	ASQ-3 Spanish	321	57.2 ± 6.5	**50.7/44.2**	54.4 ± 7	**47.4/40.4**	52.4 ± 8.7	43.7/35	56.1 ± 6.6	**49.5/42.9**	55.1 ± 6.6	**48.5/41.9**
USA 2009 [13]	ASQ-3	672	52.9 ± 11.1	41.8/30.7	52.7 ± 9.9	42.7/32.8	45.3 ± 14.7	30.6/15.8	52.7 ± 10.7	42.0/31.3	50.3 ± 11.8	38.5/26.6

n = sample size; ASQ-3 = Ages and Stages Questionnaire Third Edition.
SD = standard deviation.
Bold numbers indicate differences greater than 5 points.

domains. In the LPI group these correlations ran from moderate to high (ranging between 0.51 and 0.72) except for the personal-social domain in which the correlation was not significant. Meanwhile, in the term group the correlations were low (communication, gross motor, personal-social, and overall score) or not significant (fine motor and problem solving) (Fig. 3).

6) Agreement: Risk of developmental delay at 48 months according to risk at 24 monthss

Having a positive screening at ASQ-3 24 months, was not associated with a greater risk of positive screen nor with lower mean scores at 48 months; nevertheless, having two or more domains in the referral zone was associated with the risk of DD at 48 months (OR 140 [95% CI 14.85; 3575.65]) as well as a lower ASQ-3 scores (Table 4 and Fig. 4). We did not find differences between the domains, but rather in the sum. It is of note that there were only children in the LPI cohort with two or more domains below the cut-off on the 24 month interval.

4. Discussion

The scores obtained on the ASQ-3 in this Barcelona sample were similar to those reported in the U.S. normative data. The exceptions to this were in the fine motor domain, with children's scores at 24 months lower than those reported for the U.S. by Squires et al. [13]. By contrast children's scores at 48 months, were higher. In the other domains, the differences were generally small and not clinically relevant. However, the cutoff scores were higher in all domains at 48 months for Barcelona sample, but lower for fine motor at 24 months.

On reviewing the reliability of the ASQ-3 we found questionable internal consistency in the analysis of domains, as reported as well by Lopes et al. in Portugal and Alvik et al. in Norway [14,30]. Although these results might cast a shadow of doubt on the reliability of the test, we should stress that the size of alpha depends on the number of items in the scale, and each domain consists of only 6 questions [31]. Additionally, the items in each domain may ask for very different skills, and would not be expected to be highly correlated such as items in the personal social domain[13.] By the same token, if we consider the overall score, with thirty questions, Cronbach's alpha rises to 0.78/0.79, which is in the acceptable-to-good reliability. In a systematic review of ASQ-3 2–2.5 years, Velikonja et al. demonstrated that the English version of the questionnaire, based on the medians of the five ASQ-3 domains, had positive values for internal consistency (Cronbach's alpha > 0.70), with variability in the results of the domain analysis and less reliability in the translated/adapted versions of the ASQ-3 [32].

Another way of analyzing the reliability of a questionnaire is by the Pearson product- moment correlation between the domain scores and the overall score. With this analysis we found correlations between the domains and the overall score to be moderate-to-high, as was found by Lopes et al. in the validation of ASQ in Portugal [14], while the correlation among the domains was low, which is to be expected, given that they mark out separate domains.

In our study we analyzed a cohort of children who were assessed at 24 and 48 months of age, and we found that the scores obtained at 48 months were higher than those at 24 months. Correlations were significant between overall scores of both questionnaires, with greater correlation for the LPI than for the term. Few studies have examined the developmental trajectories of children using ASQ, and the methodologies employed have been diverse. Valla et al. [33] found, using ASQ in Norway, that a majority of children showed positive and stable normative developmental pathways from 4 to 24 months of age, while a small percentage of children had decreasing scores or other patterns of development.

Our study was not designed to establish development pathways, as there are only two measurements made by the ASQ-3. Nonetheless, it is possible to assess the stability of the risk of DD, with a very significant

Table 3

Reliability analysis for each domain and overall score using Alpha Cronbach.

	Communication	Gross motor	Fine motor	Problem solving	Personal-social	Overall score
24-month (n = 152)	0.68 (0.60–0.74)	0.43 (0.31–0.56)	0.37 (0.24–0.51)	0.37 (0.22–0.53)	0.51 (0.39–0.63)	0.79 (0.74–0.84)
48-month (n = 321)	0.68 (0.63–0.73)	0.35 (0.25–0.46)	0.54 (0.46–0.61)	0.5 (0.42–0.58)	0.46 (0.37–0.54)	0.78 (0.75–0.82)

n = sample size.

OR when there were two or more domains in the referral zone on ASQ-3 at 24 months, which only occurred in the LPI group. In this regard, we did not find differences between the domains.

Previous studies have reported an increase in the psychometric properties of the test in higher biological risk samples [22], which may be explained in part by the effect of extreme values and the increased stability of the delays. Hornmann et al. [34] found that overall development of moderately-late preterm children had stability patterns comparable to term-born children at 4 to 5 years of age, probably because when the children advance in their age, the DD have achieved greater stability. Studies have shown that ASQ-3 has a high negative predictive value; that is, if a child does have a negative screen, they are less likely to manifest risk in development over the medium and long term. The positive predictive value, however, is low [35,36,37], probably because the more subtle problems in development may be resolved with time while there is a lesser likelihood of other problems appearing, especially with increasing age in the child.

The strongest point of this study is that it analyzed two child cohorts of terms and LPI—with high levels of and patient participation and adherence. However, one of the limitations of the study is that it did not consider the possible interventions undergone by the children. Another limitation is that the follow-up sample was small, increasing the possibility of false negative results. For these reasons it seems opportune to continue the follow-up of children for a longer time period in order to learn more about the developmental pathway and about the impact of early detection of subtle developmental problems on learning and school performance. In addition, these studies should be replicated with larger more economically diverse families.

5. Conclusions

The ASQ-3 24 and 48 months in Spanish appears to be a reliable tool for developmental screening and follow-up of children between 2 and 4 years of age, especially in the LPI group. A positive agreement was found for the two questionnaires analyzed. Having two or more domains in the referral zone on the 24 months screening is a significant predictor of the risk of DD risk at 48 months of age.
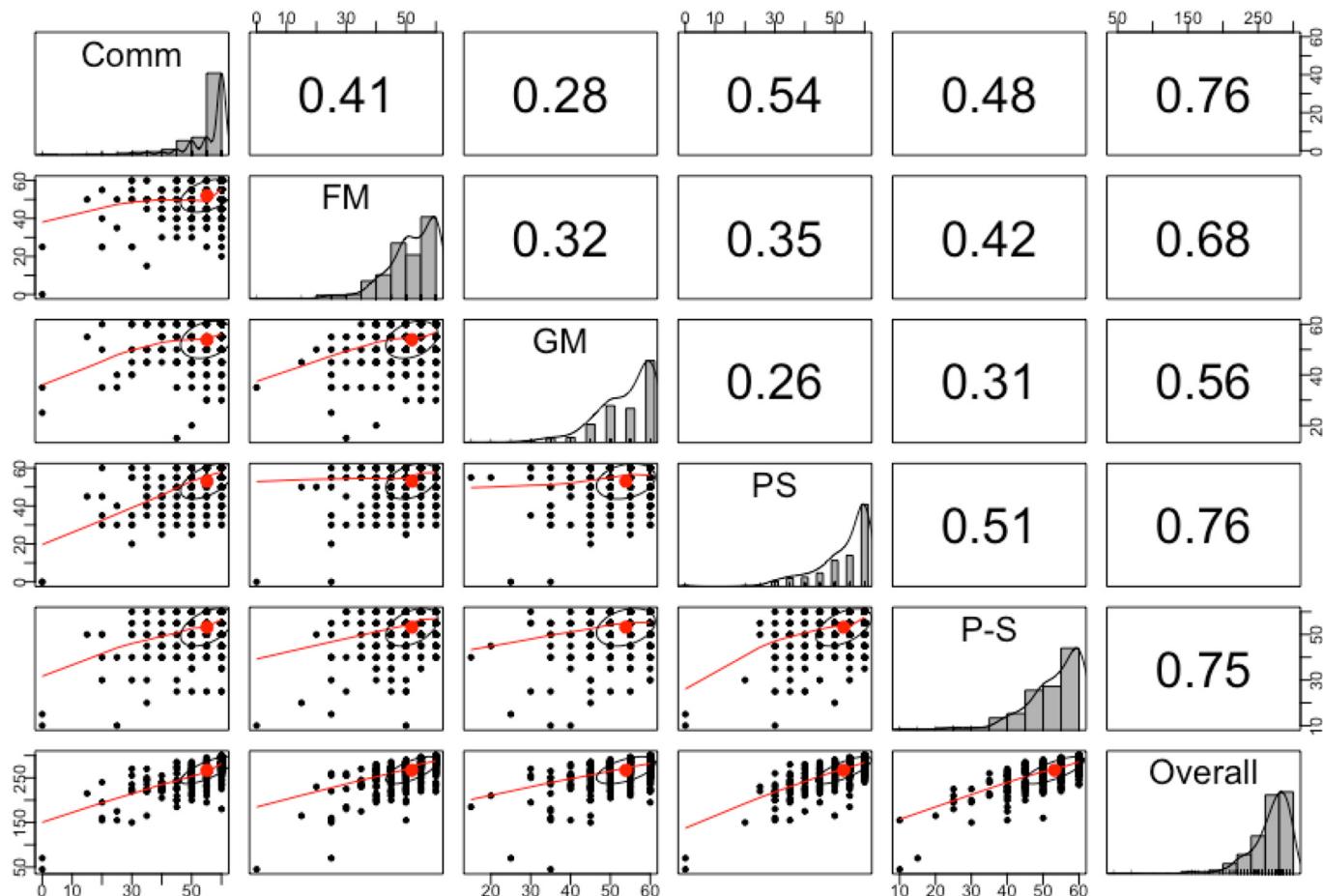

Fig. 2. Pearson product- moment correlation matrix between the domain scores and the overall score on 24- and 48-month questionnaires. All correlations are significant at p < 0.001.

Comm: communication; FM: fine motor; GM: gross motor; PS: problem solving; P-S personal-social.

Terms

48 month

LPI
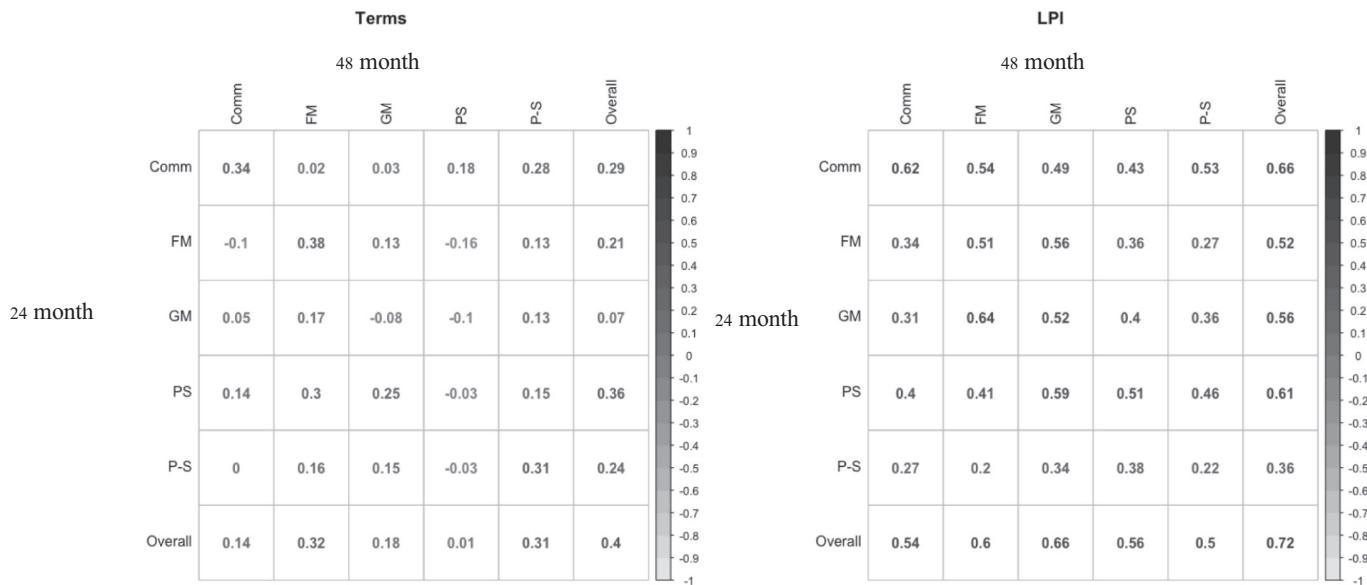
48 month

Fig. 3. Correlation for domains and overall scores at 24 and 48 months in term-born and late preterm infants.
Comm: communication; FM: fine motor; GM: gross motor; PS: problem solving; P-S personal-social; terms: term-born infants; LPI: late preterm infants.

Table 4
Odds ratio for risk on developmental delay at 48 month, according 24 month assessment, in term-born and late preterm infants.

	Terms	95% CI	LPI	95% CI
1 domain in referral zone	5.89	[0.68; 43.32]	5.60	[0.24; 69.85]
2 or more domains in referral zone			140.00***	[14.85; 3575.65]

Developmental delay risk = at least 1 domain in referral zone.
Referral zone: domain score > −2 SD below the mean score.
LPI: late preterm infants; terms: term-born infants; CI: confidence interval.
In the footnote is the definition for the asterisk: *p > 0.05.
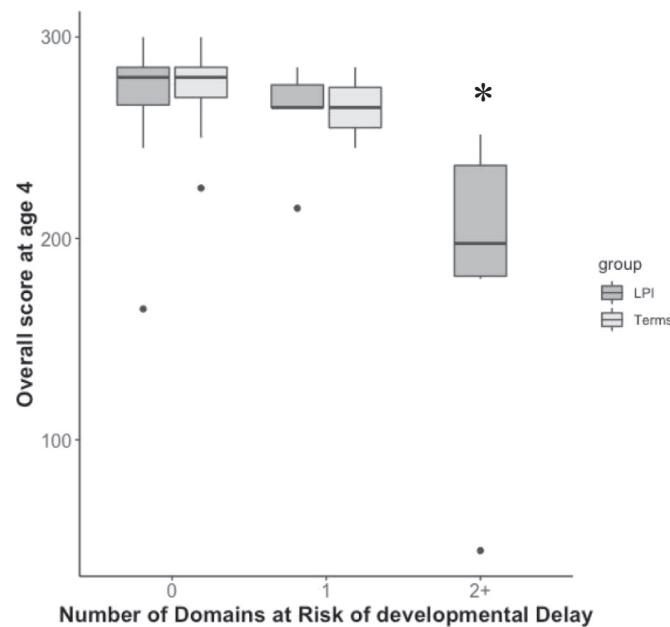
Fig. 4. Comparison of ASQ-3 48-month overall score according to number of domains under cut-off point at 24 months, in term-born and late preterm infants.
0: no domains in referral zone; 1: one domain in referral zone; 2+: two or more domains in referral zone.
*p < 0.05. LPI: late preterm infants; terms: term-born infants.

References

[1] M.J. Guralnick, Effectiveness of early intervention for vulnerable children: a developmental perspective, Am, J. Ment. Retard. 102 (1998) 319–345.
[2] L.M. Anderson, C. Shinn, M.T. Fullilove, S.C. Scrimshaw, J.E. Fielding, J. Normand, V.G. Carande-Kulis, Task force on community preventive services, The effectiveness of early childhood development programs. A systematic review, Am. J. Prev. Med. 24 (2003) 32–46.
[3] J. Orton, A. Spittle, L. Doyle, P. Anderson, R. Boyd, Do early intervention programmes improve cognitive and motor outcomes for preterm infants after discharge? A systematic review, Dev. Med. Child Neurol. 51 (2009) 851–859.
[4] G. Cioni, E. Inguaggiato, G. Sgandurra, Early intervention in neurodevelopmental disorders: underlying neural mechanisms, Dev. Med. Child Neurol. 58 (Suppl. 4) (2016) 61–66.
[5] M.J. Guralnick, Why early intervention works: a systems perspective, Infants Young Child. 24 (2011) 6–28.
[6] R.C. Sheldrick, S. Merchant, E.C. Perrin, Identification of developmental-behavioral problems in primary care: a systematic review, Pediatrics 128 (2011) 356–363.
[7] American Academy of Pediatrics. Council on Children with Disabilities. Section on Developmental Behavioral Pediatrics. Bright Futures Steering Committee. Medical Home Initiatives for Children with Special Needs Project Advisory Committee, Identifying infants and young children with developmental disorders in the medical home: an algorithm for developmental surveillance and screening, Pediatrics 118

(2006) 405–420.

[8] Department of Health, Healthy Child Programme: pregnancy and the first five years of life, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/167998/Health_Child_Programme.pdf, (2009) , Accessed date: 15 November 2018.

[9] L. Radecki, N. Sand-Loud, K.G. O'Connor, S. Sharp, L.M. Olson, Trends in the use of standardized tools for developmental screening in early childhood: 2002–2009, Pediatrics 128 (2011) 14–19.

[10] N. Sand, M. Silverstein, F.P. Glascoe, V.B. Gupta, T.P. Tonniges, K.G. O'Connor, Pediatricians' reported practices regarding developmental screening: do guidelines work? Do they help? Pediatrics 116 (2005) 174–179.

[11] A. Schonwald, K. Horan, N. Huntington, Developmental screening: is there enough time? Clin. Pediatr. (Phila) 48 (2009) 648–655.

[12] J. Squires, D. Bricker, L. Potter, Paul Brookes (Ed.), Ages and stages questionnaires user's guide, 2nd edition, 1999 (Baltimore).

[13] J. Squires, D. Bricker, Ages & Stages Questionnaires, (ASQ-3): A Parent-completed Child-monitoring System, Third edition, Paul Brookes Publishing Company, Stanford, 2009.

[14] S. Lopes, P. Graça, S. Teixeira, A.M. Serrano, J. Squires, Psychometric properties and validation of Portuguese version of Ages & Stages Questionnaires (3rd edition): 9, 18 and 30 Questionnaires, Early Hum. Dev. 91 (2015) 527–533.

[15] E. Catino, M. di Trani, F. Giovannone, F. Manti, L. Nunziata, F. Piccari, V. Sirchia, L. Vannuci, C. Sogos, Screening for developmental disorders in 3 and 4 year old Italian children: a preliminary study, Front Pediatr. 5 (2017) 181.

[16] J.M. Kerstjens, A.F. Bos, E.M. ten Vergert, G. de Meer, P.R. Butcher, S.A. Reijneveld, Support for the global feasibility of the Ages and Stages Questionnaire as developmental screener, Early Hum. Dev. 85 (2009) 443–447.

[17] H. Janson, J. Squires, Parent-completed developmental screening in a Norwegian population sample: a comparison with US normative data, Acta Paediatr. 93 (2004) 1525–1529.

[18] I. Armijo, L. Schonhaut, M. Cordero, Validation of the Chilean version of the Ages and Stages questionnaire (ASQ-CL) in community health settings, Early Hum. Dev. 91 (2015) 671–676.

[19] K.H. Heo, J. Squires, P. Yovanoff, Cross-cultural adaptation of a pre-school screening instrument: comparison of Korean and US population, J. Intellect. Disabil. Res. 52 (2008) 195–206.

[20] J.W. Small, H. Hix-Small, E. Vargas-Baron, K.P. Marks, Comparative use of the Ages and Stages Questionnaires in low- and middle-income countries, Dev. Med. Child Neurol. (2018), https://doi.org/10.1111/dmcn.13938.

[21] J.A. Sarmiento Campos, J. Squires, J. Ponte, Universal developmental screening: preliminary studies in Galicia. Spain, Early Child Dev. Care 1 (2009) 1–11.

[22] L. Schonhaut, I. Armijo, M. Schönstedt, J. Alvarez, M.A. Cordero, Validity of the Ages and Stages Questionnaires in term and preterm infants, Pediatrics 131 (2013) e1468–e1474.

[23] A. Klamer, A. Lando, A. Pinborg, G. Greisen, Ages and Stages Questionnaire used to measure cognitive deficit in children born extremely preterm, Acta Paediatr. 94 (2005) 1327–1329.

[24] M.J. Álvarez Gómez, J. Soria Aznar, J. Galbe Sánchez-Ventura, Importancia de la vigilancia del desarrollo psicomotor por el pediatra de Atención Primaria: revisión del tema y experiencia de seguimiento en una consulta en Navarra, Rev. Pediatr. Aten. Prim. 11 (2009) 65–87.

[25] J.A. Hurtado Suazo, M. García-Reymundo, M.J. Calvo Aguilar, G. Ginovart Galiana, A. Jimenez Moya, M. Tirado Aguinagalde, X. Demestre Guasch, Recomendaciones para el manejo perinatal y seguimiento del recién nacido prematuro tardío, An. Pediatr. (Barc.) 81 (2014) (327-e1-7).

[26] S.A. Rosenberg, D. Zhang, C.C. Robinson, Prevalence of developmental delays and participation in early intervention services for young children, Pediatrics 121 (2008) e1503–e1509.

[27] X. Demestre, L. Schonhaut, J. Morillas, S. Martínez-Nadal, C. Vila, F. Raspall, P. Sala, Riesgo de déficits en el desarrollo en los prematuros tardíos: evaluación a los 48 meses mediante el Ages & Stages Questionnaires®, An. Pediatr. (Barc.) 84 (2016) 39–45.

[28] S. Martínez-Nadal, X. Demestre, L. Schonhaut, S.R. Muñoz, P. Sala, Impact of neonatal morbidity on the risk of developmental delay in late preterm infants, Early Hum. Dev. 116 (2018) 40–46.

[29] D. George, P. Mallery, Allyn, Bacon (Eds.), SPSS for Windows Step by Step 11.0 Update, 4th ed., 2003 New York.

[30] A. Alvik, B. Grøholt, Examination of the cut-off scores determined by the Ages and Stages Questionnaire in a population-based sample of 6-month-old Norwegian infants, BMC Pediatr. 11 (2011) 117.

[31] D.L. Streiner, Starting at the beginning: an introduction to coefficient alpha and internal consistency, J. Pers. Assess. 80 (2003) 99–103.

[32] T. Velikonja, J. Edbrooke-Childs, A. Calderon, M. Sleed, A. Brown, J. Deighton, The psychometric properties of the Ages & Stages Questionnaires for ages 2–2.5: a systematic review, Child Care Health Dev. 43 (2017) 1–17.

[33] L. Valla, M.S. Birkeland, DH, KS, Developmental pathways in infants from 4 to 24 months, Child Care Health Dev. (2017) 1–10.

[34] J. Hornman, A.F. de Winter, J.M. Kertjens, A.F. Bos, S.A. Reijneveld, Stability of developmental problems after school entry of moderately-late preterm and early preterm-born children, J. Pediatr. 187 (2017) 73–79.

[35] L. Schonhaut, M. Pérez, A.M. Castilla, S. Castro, P. Salinas, I. Armijo, Predictive value of Ages & Stages Questionnaires for cognitive performance during early child education, Rev. Chil. Pediatr. 88 (2017) 35–40.

[36] R. Lamsal, D.J. Dutton, J.D. Zwicker, Using the ages and stages questionnaire in the general population as a measure for identifying children not at risk of a neurodevelopmental disorder, BMC Pediatr. 18 (2018) 122.

[37] M. Halbwachs, J.B. Muller, S. Nguyen The Tich, G. Gascoin, A. Chauty-Frondas, B. Branger, V. Rouger, J.C. Roze, C. Flamant, Predictive value of the parent-completed ASQ for school difficulties in preterm-born children < 35 weeks' GA at five years of age, Neonatology 106 (2014) 311–316.

Resultados Estudio N°2: Evaluar la validez predictiva del ASQ-3 aplicado a los 8, 18 o 30 meses, para detectar a los niños que tendrán déficit cognitivo en la etapa escolar en Chile

Se incluyeron 227 niños para el análisis de validez predictiva, lo que corresponde al 75% de la muestra original. Se catalogó a 15 niños (6,6%) con déficit cognitivo.

La capacidad predictiva del ASQ para identificar a aquellos niños con déficit cognitivo fue satisfactoria, con área bajo la curva (AUC) 0,77 [IC95% 0,65- 0,89] y comparable con el AUC generado por la escala de Bayley, AUC 0,80 [IC95% 0,68- 0,93] [p=0.58]. La sensibilidad, considerando al menos un dominio en zona de riesgo, fue 0,67 para ASQ-3 y 0,53 para Bayley-III, mientras que la especificidad de 0,72 y 0,88 respectivamente.

Con relación a la edad de evaluación, no hubo diferencias significativas en las AUC al comparar ASQ-CI y Bayley-III, ni al comparar las 3 edades de evaluación, pudiendo observarse una mayor AUC en la evaluación realizada a los 30 meses.

En el análisis univariado, el tener déficit en algún dominio del ASQ-CI representó un OR de 5,07 [IC95% 1,72- 16,84, p 0.004], mientras que con el test de Bayley-III el OR fue de 4,91 [IC95% 1,67-14,79 p0.003] de déficit cognitivo. En el análisis por dominio del ASQ-CI, el tener déficit en los dominios de comunicación, motricidad gruesa y resolución de problemas, se asociaron significativamente a déficit cognitivo. Mientras que con la escala de Bayley III, todos los dominios fueron significativos.

ESTUDIO 2:

Schonhaut L., Pérez M., Armijo I., Maturana A.

Comparison between Ages & Stages Questionnaire and Bayley Scales, to predict cognitive delay in school age

Factor de Impacto: 2.079

Comparison between Ages & Stages Questionnaire and Bayley Scales, to predict cognitive delay in school age[☆]

Luisa Schonhaut[a,b,*], Marcela Pérez[a,b], Iván Armijo[c], Andrés Maturana[a,b]

[a] Clinica Alemana, Santiago, Chile
[b] Faculty of Medicine, Universidad del Desarrollo, Santiago, Chile
[c] Faculty of Psychology, Universidad del Desarrollo, Santiago, Chile

ABSTRACT

Objective: To compare the predictive value of the Spanish Ages & Stages Questionnaire third edition adapted for Chilean population (ASQ-Cl) and the Bayley Scale of Infant and Toddler Development 3rd edition (Bayley-III) for cognitive delay at school age, and to identify the domain predictors.

Methodology: Data were collected from 306 term and preterm children of medium-high socio-economic level enrolled in a prospective cohort study. Developmental outcomes at 8, 18 and 30 months were assessed via the ASQ-Cl and Bayley-III; at 6–8 years cognitive development was assessed using the Wechsler Intelligence Scale for Children (WISC-III). The area under the curve (AUC), sensitivity, specificity and predictive values were calculated, and logistic regression analysis was used.

Results: Of 227 children studied, 6.6% had cognitive delay. ASQ-Cl and Bayley-III generate equivalent AUC [0.77 and 0.80]. Sensitivity 67% and 53%; specificity of 72% and 88%, positive predictive value of 14% and 24%, negative predictive values of 97% and 96% respectively. Greater predictive validity was obtained at 30 months assessment. Deficit in the communication and gross motor skills and problem-solving domains of the ASQ-Cl and all the Bayley-III domains were significantly associated with cognitive delay.

Conclusions: ASQ-Cl can be used to identify children at risk for cognitive delay at 6–8 years of age, being comparable with the Bayley-III. Some domains of ASQ-Cl and all domains of Bayley-III were significant predictors for cognitive delay. These results support the use of ASQ-Cl as a screening tool for developmental delay.

1. Introduction

Early detection and intervention for children with developmental delays (DD) is recognized as an essential part of health care to optimize outcomes for children and families [1,2].

To ensure the identification of children at risk of DD (children not developing adequately and/or acquiring skills in the expected time frame), the American Academy of Pediatrics (AAP) recommends the application of standardized developmental screening tests at 9, 18, and 30 month follow-up visits, using standardized, valid and reliable tools [3]. The Public Health System recommendations in Chile are similar

[4]. When developmental screening identifies a child as being at high risk of DD, a diagnostic developmental evaluation should be pursued [3].

The use of parent-completed screening test, such as the Ages and Stages Questionnaire (ASQ), has increased in recent years in the USA [5,6]. The ASQ was updated in 2009, as ASQ-3 [7]. This questionnaire has been validated in several countries with promising results [8–11]. The Bayley Scale of Child Development (Bayley) has been considered the gold standard as a developmental assessment tool, both clinically and in research [12]. Studies that analyze the concurrent validity of ASQ and Bayley have shown high specificity but variable sensitivity

values [13–16]. According to the AAP, sensitivity and specificity levels of 70% to 80% have been considered acceptable for developmental screening tests [3]. These values are lower than those commonly used in other screening tests, due to the challenges implicit in the developmental assessment and the absence of clearly defined therapies.

The ASQ-3 was validated in Chile, in a representative sample of children from different socioeconomic strata, showing that the psychometric properties make it appropriate for its application in this cultural setting (ASQ-Cl) [17].

In a subsample of healthy children of different gestational ages, from a medium high socioeconomic setting, the concurrent validity of the assessment at 8, 18 and 30 months of the ASQ-Cl was evaluated, using the Bayley Scales of Infant and Toddler Development 3th edition (Bayley-III) as the reference standard; the sensitivity and specificity were 73% and 81% respectively [18]. In a preliminary analysis of 123 children ASQ-Cl was found highly predictive for the lowest cognitive coefficient during the first years of school education, but cognitive delay was not analyzed in this sample [19], nor comparisons were made with the predictive capacity of Bayley. Studies have analyzed the predictive value for cognitive delay of ASQ and Bayley of ASQ separately describing adequate values of sensitivity and specificity for cognitive delay, but with shorter of follow-up periods [20,21] or in high risk groups, as the extremely preterm born children [22,23]. No published study has yet compared ASQ with Bayley for predicting long term cognitive delay.

The aim of the present study is to compare the predictive value of the 8, 18 and 30-month assessment with ASQ-Cl and Bayley-III for cognitive delay at age 6–8 years, and to identify the domain predictors.

2. Methods

2.1. Study population

A convenience sample of 306 children who attended a pediatric ambulatory clinic in Santiago, Chile, was recruited from April 2008 to April 2011. The sample comprised term and preterm infants from families of medium-high socio-economic level and 95% of the mothers had > 13 years of scholar education. The sample was evaluated at 8, 18 months (corrected gestational age for preterm), and at 30 months. ASQ-Cl and Bayley-III test were applied concurrently.

Each child was evaluated only at one point in time with both tests. The specifications of the original cohort and the application of the ASQ-Cl and Bayley-III tests in the studied cohort are previously published [18].

Subsequently, every mother was contacted by the research team when their children were between 6 and 8 years of age, for an evaluation with the Wechsler Intelligence Scale for Children (WISC-III). Children with a history of intercurrent disease that could affect development, such as meningitis, central nervous system tumor pathology, severe cranial trauma; and children who were > 9 years old at the time of recruitment were excluded.

In our hospital, premature infants < 32 weeks or < 1500 g at birth, had access to early intervention as a part of the follow-up program. At the time of the study this was not available for moderate and late preterm born children.

2.2. Instruments

The Ages and Stages Questionnaire Third Edition translated to Spanish and adapted for Chilean population (ASQ-Cl) [7,17]: is an instrument in which parents rate their child's current skills and development, from 1 to 66 months of age. In this test twenty-one questionnaires are available but only the 8, 18 and 30 months have been validated for the local population in Chile [17,18]. Parents answer 30 questions covering 5 domains of development, including communication, gross motor, fine motor, problem-solving, and socio-emotional

skills. Infants were two standard deviations below the mean in any domain have positive screen and were considered in referral zone, or at risk of DD [7]. Parents completed the ASQ-Cl before de Bayley III assessment.

The Bayley Scale of Infant and Toddler Development 3rd edition (Bayley III) [12]: is a comprehensive developmental assessment, for children ages 1 to 42 months. Three subscales were administered (cognitive, language, and motor) by an accredited occupational therapist, who was blinded to the ASQ results. A child was considered to have a positive screen if the score is below 80 points in at least one domain (equivalent to < -2 SD for the study sample).

The Wechsler Intelligence Scale for Children, Third Edition (WISC-III) [24]: is an instrument that evaluates the intellectual capacity of children from 6 to 17 years. The instrument uses 13 sub-tests, of which 6 are verbal and 7 manual and is reported in performance intellectual quotient and verbal intellectual quotient scores. The WISC-III was applied by a group of trained psychologists and blinded to the child's clinical and developmental history, gestational age, and the results previously obtained in the ASQ-Cl and the Bayley-III scale. A score of < 85 points (equivalent to < -1.5 SD) in verbal and/or performance scales was defined as cognitive delay according to the Chilean validation [25], Parents were informed of the results.

Children unable to complete the test because of severe developmental difficulties were considered as having a cognitive delay (or true positives for the analysis).

All three tests were applied and scored independently in every child and unaware of the results from other tests. Additionally, the researchers interpreted the results of each test independently and unaware of other scores.

Written informed consent was obtained. Research Ethics Board of Clinical Alemana and Universidad del Desarrollo approved the study.

2.3. Statistical analysis

Descriptive analysis was done using chi-square and *t*-test for categorical and continuous variables respectively.

Receiver operating characteristic (ROC) and area under the curve (AUC) were calculated for ASQ-Cl and Bayley-III total scores, using as reference cognitive delay.

The analyses were done for the total group and segmented according to the age of application of the tests (8, 18 and 30 months). Bayley-III and ASQ-Cl AUC were compared using the method described by DeLong [26]. Although an absolute standard for the interpretation of AUC does not exist, it is considered that values of 0.5 indicate a random result; between 0.6 and 0.7 are acceptable; between 0.7 and 0.9 are good and > 0.9 are excellent [27]. Sensitivity, specificity, positive and negative predictive values were calculated.

Logistic regression analysis was used to test the significance of deficit in each domain, to predict cognitive delay. All the analyses were performed on the R platform, using the pROC module [28,29].

3. Results

Loss to follow-up rate in this study was 25.2% (Fig. 1). Differences between included and not included children are shown in Table 1. Included population had a lower gestational age, a higher percentage of males and twins, and a trend towards lower scores in ASQ-Cl and Bayley-III scale (not reaching statistical significance).

Analyses were conducted on 227 children assessed at 6–8 years of age. Of all the evaluations with ASQ-Cl and Bayley-III scale, 85 were performed at 8 months, 75 at 18 months and 67 evaluations were done at 30 months of age. Table 2 shows no differences in biodemographic characteristics between the children evaluated at 8, 18 and 30 months.

Of the 227 children, 221 were evaluated with WISC-III. Because of severe deficits in their development, 6 cases could not be evaluated and were therefore considered positive for the analysis reaching a total of
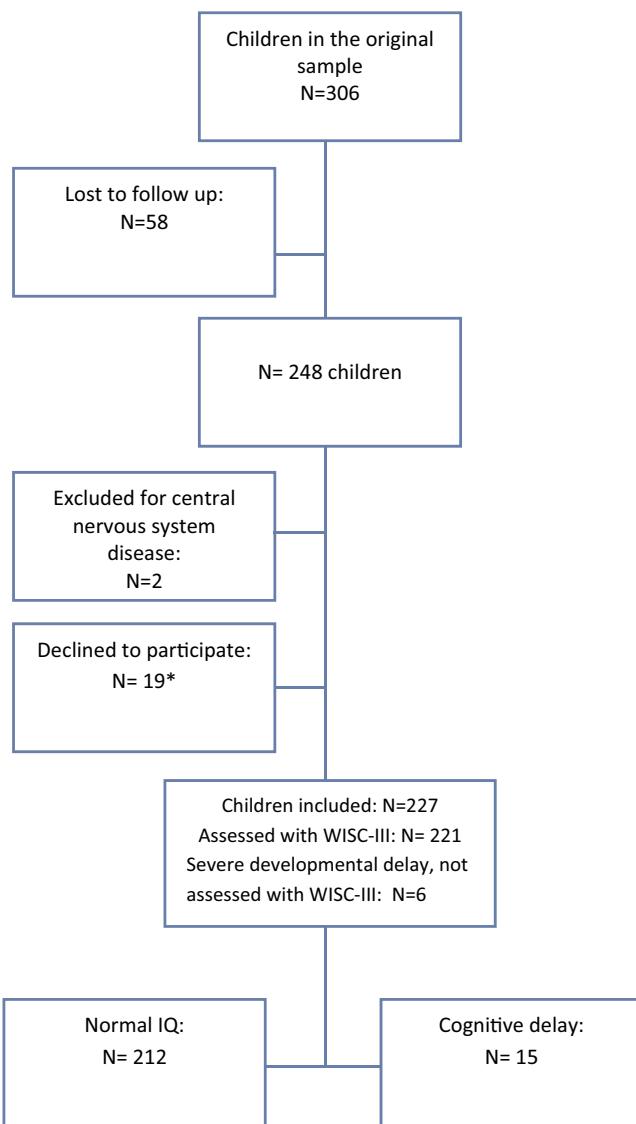
Fig. 1. Recruitment flowchart of eligible children in the study population.
* The parents of 4 children who declined to participate declared being already intervened for school performance difficulties.

Table 1
Biodemographic data. Comparison between included and non-included children.

		Included	Non-included	P[a]
		227	79	
Age of assessment[b]; n (%)	8	85 (37)	25 (32)	0.468
	18	75 (33)	25 (32)	
	30	67 (30)	29 (37)	
Male gender	n (%)	128 (56)	33 (42)	**0.025**[c]
Twin birth	n (%)	84 (37)	16 (20)	**0.006**[c]
Gestational age (weeks)	M (SD)	34.5 (3.91)	36.34 (3.68)	**0.000**[c]
ASQ-Cl total score	M (SD)	232.49 (37.62)	241.58 (39.95)	0.070
Bayley-III total score	M (SD)	297.55 (29.31)	302.73 (29.72)	0.178

ASQ-Cl = Ages and Stages Questionnaire Third Edition translated to Spanish and adapted for Chilean population; Bayley Scales of Infant and Toddler Development 3rd edition.
n = number of children included; M = media; SD = standard deviation.
[a] Chi-square for categorical data and *t*-student for continues variables.
[b] Corrected age at 8 and 18 months for preterm born children.
[c] p > 0.05

15 children with cognitive delay (6.6%) (Fig. 1).

Fig. 2 shows the predictive accuracy of the ASQ-Cl and Bayley-III for cognitive delay. Both tests generate good and equivalent AUC 0.77 [IC95% 0.65–0.89] and 0.80 [CI 95% 0.68–0.93] (p = 0.580) (Fig. 2). Considering at least one domain in the risk zone, the sensitivity was 67% for ASQ-Cl and 53% for Bayley-III, while the specificity was 72% and 88% respectively; positive predictive value of 14% and 24%, negative predictive values of 97% and 96% respectively (Table 3).

No significant differences were found, when comparing the 3 evaluation points in the AUC when comparing ASQ-Cl and Bayley-III, observing a greater AUC in the evaluation done at 30 months (Fig. 3).

Univariate analysis was done considering results at all evaluation time points for the sample as predictors for cognitive delay; having at least positive screen in any ASQ-Cl domain represented an odds ratio of 5.07 [CI 95% 1.73–16.84], p = 0.004 while for the Bayley-III test the odds ratio was 8.18 [CI 95% 2.74–24.42] p < 0.001. In ASQ-Cl the analysis by domain showed that having a risk zone performance in communication, gross motor skills and problem-solving domains, were significantly associated with cognitive delay. The deficit in any domain of the Bayley-III scale, showed significance for cognitive delay (Table 4).

4. Discussion

In the present study we found that both ASQ-Cl and Bayley-III tests were good predictors of cognitive delay, without significant differences between them (AUC 0.77 and 0.80 respectively). These results could be expected given the good correlation and concurrent validity between the two tests previously reported [18]. A recent systematic review, showed that parent-report screening tools for language, achieved higher sensitivity, specificity and negative predictive values than direct child assessment [30]. Another advantage of using parent-report screening tools such as ASQ is its low cost, and the fact that they do not require trained personnel for their application. It also empowers and involves the parents in the neurodevelopment of their children [31,32].

We found modest sensitivity values for ASQ-Cl and Bayley-III (67% and 53%, respectively), adequate specificity (72 and 88%, respectively), low positive predictive values (14% and 24%) and excellent negative predictive values (97% and 96%). These results are comparable with previous studies [20,21]. However, we must highlight the longer follow-up period in the present cohort. It is known that the time span between testing and end point measurement, may influence its predictive capacity [33]. Charkaluk et al. studied the predictive value of the 36-month ASQ assessment for cognitive delay at age 5 to 6 years in the general population, reporting sensitivity values of 77% and a specificity of 68% [20]. Kerstjens et al. reported sensitivity and specificity of 89 and 80%, respectively, to predict the need for special education within 1 year of follow-up, in children evaluated at 4 years with ASQ from Netherland's general population [10]. Halbwachs et al. analyzed the predictive capacity of ASQ at 18, 24, or 36 months for severe learning difficulties at 5 years of age in a sample of preterm children, obtaining AUC between 0.66 and 0.77 [21].

In the analysis by age of assessment, we found good predictive capacity in all ages in which the ASQ-Cl and Bayley-III were applied, being greater at 30 months, which coincides with the study by Halbwachs et al. with ASQ and Doyle LW et al. with the Bayley scale [21] [22]. There are currently no published studies that compare both tests.

In our study the domains that best predicted cognitive delay were communication, gross motor and problem-solving skills of ASQ-Cl. All the Bayley-III domains predicted accurately the cognitive delay. In a cohort of children with low risk in which the trajectory of development with ASQ was evaluated in 11 intervals between 4 months and 4 years and then the intellectual quotient was measured between 6 and 11 years, Piek et al. demonstrated a relationship between early motor development and later cognitive function [34]. While in the study of

Table 2
Demographic characteristics of the study sample according to age of assessment.

		8 months[b] N = 85	18 months[b] N = 75	30 months N = 67	p[a]
Gestational age group; n (%)[b]	At term born > 37[0]	32 (38)	28 (37)	18 (27)	NS
	Moderately and late preterm 32[0]–36[+6]	33 (39)	32 (43)	31 (46)	
	Extremely preterm < 32[0] and/or < 1500 g	20 (24)	15 (20)	18 (27)	
Birth weight	M (SD)	2820.81 (891.56)	2529.17 (832.48)	2557.05 (869.68)	NS
Male gender	n (%)	54 (64)	43 (57)	31 (46)	NS
Twin birth	n (%)	25 (29)	26 (35)	33 (49)	NS
Hospitalized newborn period	n (%)	44 (52)	39 (52)	42 (63)	NS
Mother age at delivery	M (SD)	31 (5)	32 (4)	32 (4)	NS
Mothers with > 13 years of scholar education	n (%)	81 (95)	72 (96)	65 (97)	NS

ASQ-Cl = Ages and Stages Questionnaire Third Edition translated to Spanish and adapted for Chilean population; Bayley Scales of Infant and Toddler Development 3rd edition.

n = number of children included; M = media; SD = standard deviation; NS = not significant.

[a] Chi-square for categorical data and ANOVA for continues variables.
[b] Corrected age at 8 and 18 months for children born before term.
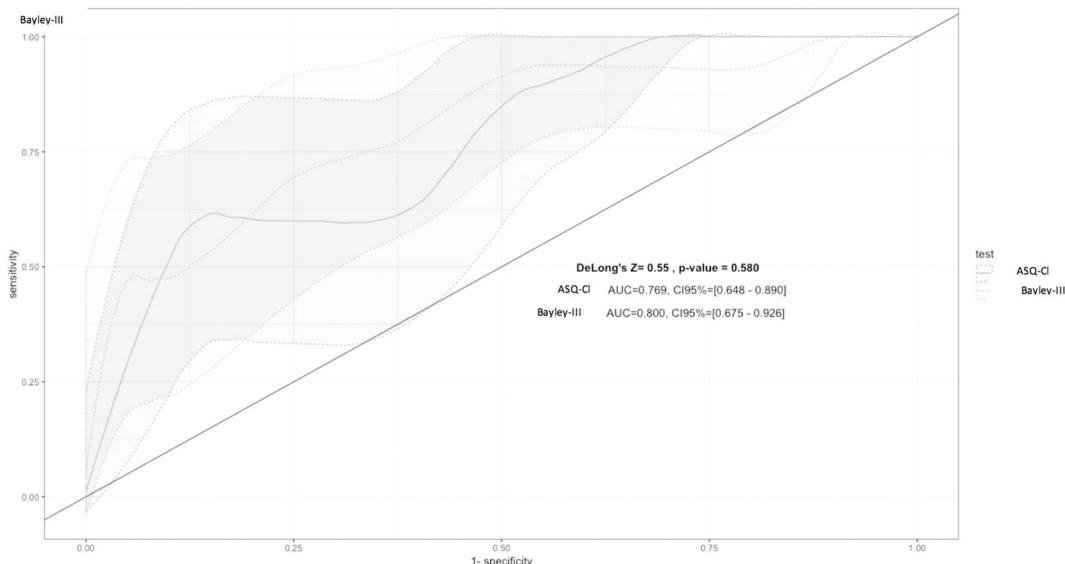
Fig. 2. Comparison of the ASQ-Cl and Bayley-III ROC curves for predicting cognitive delay.

Table 3
Comparison of psychometric properties of the ASQ-Cl and the Bayley-III for predicting cognitive delay.

	ASQ-Cl[a]	Bayley-III[b]
Sensitivity	67 (39–87)	53 (27–78)
Specificity	72 (65–78)	88 (82–92)
Positive predictive value	14 (7–25)	24 (11–42)
Negative predictive value	97 (92–99)	96 (92–98)

ASQ-Cl = The Ages and Stages Questionnaire Third Edition translated to Spanish and adapted for Chilean population; Bayley III = Bayley Scales of Infant and Toddler Development 3rd edition.

[a] Children that scored two standard deviations below the mean in any domain were considered as a positive screen.
[b] Children that scored below 80 points in at least one domain were considered to have a positive screen.

Peyre et al. early language skills more strongly predict later intellectual quotient disabled children at 5–6 years old [35]. In a meta-analysis that considered studies performed in children born extremely premature, Luttikhuizen dos Santos et al. described that mental developmental index scores in Bayley were strongly predictive for poor cognitive performance [23].

When studying the predictive capacity of questionnaires, it is important to consider that prediction is difficult because of numerous factors that can modify the natural history of children's development, which include: rapid developmental change, biologic or environmental variables, developmental interventions, and the fact that testing itself has an impact on the developmental trajectory [33,36,37]. The enriched environment and access to early interventions could have improved the outcome in our studied cohort, explaining in part the low positive predictive value.

One of the limitations of our study is that we only measure cognitive development without evaluating other aspects such as motor or socioemotional performance, aspects that escaped the objectives of this study. On the other hand, we could not analyze in depth the intensity of the interventions received by children and their impact on subsequent cognitive performance. The small number of children with deficit probably diminishes the power of our results, and the external validity is limited as the participants were from medium to high socioeconomic level in Chile.

Due to the mentioned limitations we should probably consider these results as preliminary. Nevertheless, our study, with high adherence rate and long follow-up period, shows promising evidence suggesting a strong predictive validity of a parent-completed screening test with a professionally administered cognitive test for cognitive delay at early
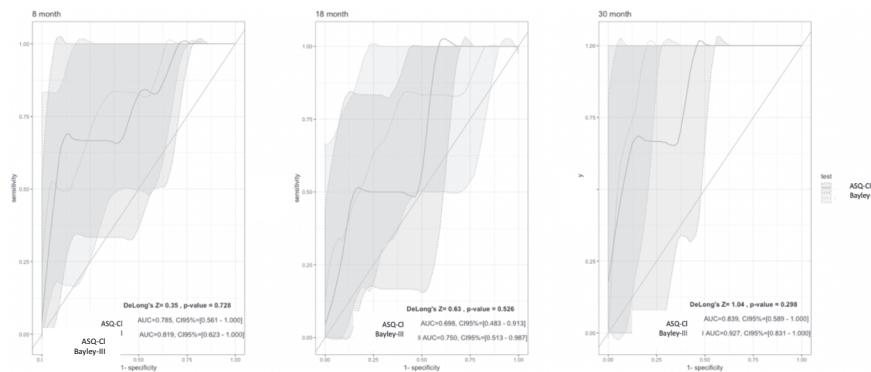
Fig. 3. Comparison of psychometric properties of ASQ-CI and Bayley-III for predicting cognitive delay according to age of assessment.

Table 4

Univariate analysis of the different developmental domains for predicting cognitive delay in the total sample ($n = 227$ children).

Predictor	OR	95% CI	p
ASQ-Cl *positive screen*[a]			
Communication	9.42	2.82; 30.22	< 0.001
Gross motor	5.22	1.63; 15.81	0.003
Fine motor	2.35	0.34; 9.79	0.291
Problem solving	4.17	1.07; 13.75	0.024
Social-personal	0.58	0.03; 3.13	0.615
At least one domain at risk zone	5.07	1.73; 16.84	0.004
Bayley-III Scales*delay*[b]			
Cognitive	26.25	4.00; 172.27	< 0.001
Language	7.65	2.28; 25.700	0.001
Motor	6.12	1.87; 20.10	0.003
At least one domain at risk zone	8.18	2.74; 24.42	< 0.001

[a] Domain at risk zone was defined as a performance under the cutoff point ($> -2DS$) in The Ages and Stages Questionnaire Third Edition translated to Spanish and adapted for Chilean population (ASQ-Cl).

[b] Delay was defined as a performance under < 80 points in the Bayley Scales of Infant and Toddler Development 3rd edition (Bayley-III).

school age (6 to 8 years). Additional studies with a more heterogeneous and diverse group of children and families are needed to confirm these results.

5. Conclusions

Our results suggest that, the 8, 18 and 30-month ASQ-Cl assessment, could be used to identify children at risk of cognitive delay at 6 to 8 years of age, being comparable with the Bayley-III scale, traditionally considered as the gold standard for developmental assessment. The elevated negative predictive value indicates that if a child has a normal evaluation early in life, there is a high probability of not having cognitive difficulties in early school age. On the other hand, a low positive predictive value suggests that even a child with an early evaluation showing deficit, still has the possibility of an adequate cognitive development later in life in the adequate enriched environment. It is probably important to continue to monitor cognitive development throughout childhood with special attention to those children with positive screen in communication, gross motor and problem-solving domains in the ASQ, and any delay in the Bayley Scale.

Funding source

Financial disclosure

Declaration of competing interest

Acknowledgments

References

[1] L.M. Anderson, C. Shinn, M.T. Fullilove, S.C. Scrimshaw, J.E. Fielding, J. Normand, et al., The effectiveness of early childhood development programs, Am. J. Prev. Med. 24 (3) (2003) 32–46.

[2] J. Orton, A. Spittle, L. Doyle, P. Anderson, R. Boyd, Do early intervention programmes improve cognitive and motor outcomes for preterm infants after discharge? A systematic review, Dev. Med. Child Neurol. 51 (11) (2009) 851–859.

[3] Council on Children With Disabilities, Section on Developmental Behavioral Pediatrics, Bright Futures Steering Committee, Medical Home Initiatives for Children With Special Needs Project Advisory Committee, Identifying infants and young children with developmental disorders in the medical home: an algorithm for developmental surveillance and screening, Pediatrics 118 (1) (2006) 405–420.

[4] Ministerio de Salud de Chile, Manual para el Apoyo y Seguimiento del Desarrollo Psicosocial de los Niños y Niñas de 0 a 6 Años [Internet], Available from, 2008. https://scielo.conicyt.cl/scielo.php?pid=S0370-41062008000700006&script=sci_arttext.

[5] A.H. Hirai, M.D. Kogan, V. Kandasamy, C. Reuland, C. Bethell, Prevalence and variation of developmental screening and surveillance in early childhood, JAMA Pediatr. 172 (9) (2018) 857–866.

[6] L. Radecki, N. Sand-Loud, K.G. O'Connor, S. Sharp, L.M. Olson, Trends in the use of standardized tools for developmental screening in early childhood: 2002-2009, Pediatrics 128 (1) (2011) 14–19.

[7] J. Squires, D. Bricker, Ages and Stages Questionnaires User's Guide, Third edit, PAUL H. Brookes Publishing Co, Baltimore, USA, 2009.

[8] K.H. Heo, J. Squires, P. Yovanoff, Cross-cultural adaptation of a pre-school screening instrument: comparison of Korean and US populations, J. Intellect. Disabil. Res. 52 (2008) 195–206.

[9] S. Lopes, P. Graça, S. Teixeira, A.M. Serrano, J. Squires, Psychometric properties and validation of Portuguese version of Ages & Stages Questionnaires (3rd edition): 9, 18 and 30 questionnaires, Early Hum. Dev. 91 (9) (2015) 527–533.

[10] J.M. Kerstjens, A.F. Bos, E.M.J. ten Vergert, G. de Meer, P.R. Butcher, S.A. Reijneveld, Support for the global feasibility of the Ages and Stages Questionnaire as developmental screener, Early Hum. Dev. 85 (7) (2009) 443–447.

[11] L. Schonhaut, S. Martinez-Nadal, I. Armijo, X. Demestre, Reliability and agreement of ages and stages questionnaires®: results in late preterm and term-born infants at 24 and 48 months, Early Hum. Dev. 128 (1) (2019) 55–61.

[12] C.A. Albers, A.J. Grieve, Test review: Bayley, N. (2006). Bayley Scales of Infant and Toddler Development–Third Edition. San Antonio, TX: Harcourt Assessment, J. Psychoeduc. Assess. 25 (2) (2007) 180–190.

[13] L. Gollenberg a, C.D. Lynch, L.W. Jackson, B.M. McGuinness, M.E. Msall, Concurrent validity of the parent-completed Ages and Stages Questionnaires, 2nd ed. with the Bayley Scales of Infant Development II in a low-risk sample, Child Care Health Dev. 36 (4) (2010) 485–490.

[14] S. Veldhuizen, J. Clinton, C. Rodriguez, T.J. Wade, J. Cairney, Concurrent validity of the Ages and Stages Questionnaires and Bayley Developmental Scales in a general population sample, Acad. Pediatr. 15 (2) (2015) 231–237.

[15] L.J.P. Steenis, M. Verhoeven, D.J. Hessen, A.L. van Baar, Parental and professional assessment of early child development: the ASQ-3 and the Bayley-III-NL, Early Hum. Dev. 91 (3) (2015) 217–225.

[16] P.K. Agarwal, L. Shi, L.M. Daniel, P. Hong, Y. Poh, C. Khoo, Prospective evaluation of the Ages and Stages Questionnaire 3rd edition in very-low-birthweight infants, Dev. Med. Child Neurol. (2016) 1–6.

[17] I. Armijo, L. Schonhaut, M. Cordero, Validation of the Chilean version of the Ages and Stages Questionnaire (ASQ-CL) in community health settings, Early Hum. Dev. 91 (12) (2015) 671–676.

[18] L. Schonhaut, I. Armijo, M. Schönstedt, J. Alvarez, M. Cordero, Validity of the ages and stages questionnaires in term and preterm infants, Pediatrics 131 (5) (2013) e1468–e1474.

[19] L. Schonhaut, M. Pérez, A.M. Castilla, S. Castro, P. Salinas, I. Armijo, Predictive value of Ages & Stages Questionnaires for cognitive performance during early child education, Rev. Chil. Pediatr. 88 (1) (2017) 35–40.

[20] M.-L. Charkaluk, J. Rousseau, J. Calderon, J.Y. Bernard, A. Forhan, B. Heude, et al., Ages and Stages Questionnaire at 3 years for predicting IQ at 5–6 years, Pediatrics 139 (4) (2017) e20162798.

[21] M. Halbwachs, J.-B. Muller, S. Nguyen The Tich, G. Gascoin, A. Chauty-Frondas, B. Branger, et al., Predictive value of the parent-completed ASQ for school difficulties in preterm-born children < 35 weeks' GA at five years of age, Neonatology 106 (4) (2014) 311–316.

[22] L.W. Doyle, P.G. Davis, B. Schmidt, P.J. Anderson, Cognitive outcome at 24 months is more predictive than at 18 months for IQ at 8-9 years in extremely low birth weight children, Early Hum. Dev. 88 (2) (2012) 95–98.

[23] E.S. Luttikhuizen dos Santos, J.F. de Kieviet, M. Königs, R.M. van Elburg, J. Oosterlaan, Predictive value of the Bayley Scales of Infant Development on development of very preterm/very low birth weight children: a meta-analysis, Early Hum. Dev. 89 (7) (2013) 487–496.

[24] D. Wechsler, Wechsler Intelligence Scale for Children (WISC). Wechsler IQ Test [Internet], Available from, 2018. https://wechsleriqtest.com/wechsler-intelligence-scale-for-children/.

[25] V. Ramírez, R. Rosas, Estandarización del WISC-III en Chile: Descripción del Test, Estructura Factorial y Consistencia Interna de las Escalas Standarization of WISC-III in Chile: Test Description, Factorial Structure, and Internal Consistency of the Scales, Psykhe 16 (1) (2007) 91–109.

[26] E.R. DeLong, D.M. DeLong, D.L. Clarke-Pearson, Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach, Biometrics 44 (3) (1988) 837–845.

[27] J.A. Swets, J.A. Swets, Measuring the accuracy of diagnostic systems, Science 240 (1998) 1285–1293.

[28] Team RC, A language and environment for statistical computing. R Foundation for Statistical Computing [Internet], R Foundation for Statistical Computing, Vienna, Austria, (2012) Available from http://www.r-project.org/.

[29] N. Turck, L. Vutskits, P. Sanchez-Pena, X. Robin, A. Hainard, M. Gex-Fabry, et al., pROC: an open-source package for R and S+ to analyze and compare ROC curves, BMC Bioinformatics 8 (2011) 12–77.

[30] F. Sim, L. Thompson, L. Marryat, N. Ramparsad, P. Wilson, Predictive validity of preschool screening tools for language and behavioural difficulties: a PRISMA systematic review, PLoS One 14 (2019) 1–31.

[31] L. Sices, D. Drotar, A. Keilman, H.L. Kirchner, D. Roberts, T. Stancin, Communication about child development during well-child visits: impact of parents evaluation of developmental status screener with or without an informational video, Pediatrics 122 (5) (2008) e1091–e1099.

[32] S. Kendall, A. Nash, A. Braun, G. Bastug, E. Rougeaux, H. Bedford, Acceptability and understanding of the Ages & Stages Questionnaires®, Third Edition, as part of the Healthy Child Programme 2-year health and development review in England: parent and professional perspectives, Child Care Health Dev. 45 (2) (2019) 251–256.

[33] G.P. Aylward, Developmental screening and assessment: what are we thinking? J. Dev. Behav. Pediatr. 30 (2) (2009) 169–173.

[34] J.P. Piek, L. Dawson, L.M. Smith, N. Gasson, The role of early fine and gross motor development on later motor and cognitive ability, Hum. Mov. Sci. 27 (5) (2008) 668–681.

[35] H. Peyre, M.-L. Charkaluk, A. Forhan, B. Heude, F. Ramus, Do developmental milestones at 4, 8, 12 and 24 months predict IQ at 5-6 years old? Results of the EDEN mother-child cohort, Eur. J. Paediatr. Neurol. 21 (2) (2017) 272–279.

[36] L. Valla, M.S. Birkeland, D. Hofoss, K. Slinning, Developmental pathways in infants from 4 to 24 months, Child Care Health Dev. 43 (4) (2017) 546–555.

[37] G. Cioni, E. Inguaggiato, G. Sgandurra, Early intervention in neurodevelopmental disorders: underlying neural mechanisms, Dev. Med. Child Neurol. 58 (Suppl. 4) (2016) 61–66.

Resultados Estudio N°3: Determinar la capacidad del ASQ y otros CCDRP para predecir las dificultades cognitivas o académicas a largo plazo, en población general.

Luego de revisar 2.277 citas, se seleccionaron 74 estudios, de los cuales 32 cumplieron los criterios de inclusión, correspondiendo a 10 cohortes. Todas las cohortes incluidas fueron evaluadas con ASQ, cuestionario que pasó por distintos procesos de validación y adaptación. Las cohortes provenían de Sudamérica (Chile y Colombia), Europa (Cataluña, Holanda, Francia y Noruega), Asia (Corea del Sur y el norte de India), y Oceanía (Australia) [28–31, 33, 64, 68, 79, 85, 87–91]. Los autores de 7 de las cohortes colaboraron con la revisión de las fichas de extracción de datos.

En 8 cohortes se pudo obtener la asociación entre ASQ y rendimiento académico o cognitivo en la etapa escolar. Las asociaciones fueron en general positivas, en especial si se consideraba el cuestionario en su totalidad, con AUC entre 0,66 y 0,87, y OR superiores a 3. Un efecto espejo se encontró al revisar la sensibilidad y especificidad de las cohortes. La heterogeneidad de los estudios no permitió realizar un metaanálisis.

En la evaluación de los estudios incluidos, destacó que el riesgo de sesgo en la selección de pacientes fue en general baja. Hubo diferencias entre las cohortes con la definición de puntos de corte para la definición de déficit cognitivo o en la adaptación del cuestionario.

ESTUDIO 3:

Schonhaut L., Maturana A., Cepeda O., Serón P.

Predictive Validity of Developmental Screening Questionnaires for Identifying Children with Later Cognitive or Educational Difficulties- A Systematic Review

Predictive Validity of Developmental Screening Questionnaires for Identifying Children With Later Cognitive or Educational Difficulties: A Systematic Review

Luisa Schonhaut [1,2]*, Andres Maturana [1,2,3], Olenkha Cepeda [3] and Pamela Serón [4]

[1] *Departamento de Pediatría, Clínica Alemana, Santiago, Chile,* [2] *Facultad de Medicina Clínica Alemana, Universidad del Desarrollo, Santiago, Chile,* [3] *Departamento de Desarrollo Académico e Investigación, Clínica Alemana, Santiago, Chile,* [4] *Departamento Medicina Interna y Centro de excelencia CIGES, Universidad de La Frontera, Temuco, Chile*

Context: Parent/caregiver completing developmental screening questionnaires (DSQs) for children before 5 years of age is currently recommended. The DSQs recommended by the American Academy of Pediatrics (AAP) are the Ages and Stages Questionnaires (ASQ), Parents' Evaluation of Developmental Status (PEDS), and the Survey of Well-being of Young Children (SWYC). Nevertheless, their predictive validity has not been well-established.

Objective: To assess in the current literature, the value of AAP-recommended DSQs (ASQ, PEDS, SWYC) administered between 0 and 5 years of age, for predicting long-term cognitive achievement and/or school performance (CA/SP), after 1 year or more of evaluation and at/or after age 5 years, in the general population.

Data Sources: Cochrane, MEDLINE PubMed, CINAHL, EMBASE, Web of Science, Scielo, and Scopus databases (until March 2021).

Study Selection: Two authors selected the studies. Forward and backward citation follow-up was done; authors of DSQ were contacted to identify additional studies.

Data Extraction: Cohorts were identified, and authors of selected studies were contacted to corroborate and complete extracted data.

Results: Thirty-two publications, corresponding to 10 cohorts, were included. All cohorts used ASQ. Only cohort using PEDS was identified but did not meet the inclusion criteria. No cohorts conducted with SWYC were identified. Associations between ASQ and CA/SP were extracted for eight cohorts. The odds ratios were ≥ 3, and the area under the curve was 0.66–0.87. A trade-off between sensitivity and specificity was observed.

Limitations: Heterogeneity in population characteristics and in DSQ adaptations.

Conclusions: A positive association between ASQ and later CA/SP was found in different social, cultural, and economic settings. Additional studies are necessary to determine the impact factors in the predictive capacity of DSQs.

Systematic Review Registration: PROSPERO, identifier: CRD42020183883.

Keywords: screening tools, developmental screening questionnaires, cognition, educational difficulties, Ages and Stages Questionnaires (ASQ)

HIGHLIGHTS

- ASQ is the most widely used developmental screening questionnaire in follow-up cohorts of young children.
- Positive association between ASQ and later cognitive and/or school performance was found.
- Trade-off between sensitivity and specificity could be explained by different scoring methods.

INTRODUCTION

It is estimated that one in six children has a developmental disability, defined by problems in cognitive, behavioral, language, learning, or physical performance, which are often more prevalent in children with biological risk factors such as prematurity (1–3). Considering that development is a continuum and that the first 5 years of life are recognized as a critical period for subsequent cognitive performance and school success, it is accepted that these disabilities begin in early childhood, under the definition of developmental delay (DD) (4). Early detection of DD allows for timely and effective interventions (5, 6). For this reason, early screening and referral of developmental difficulties are a critical element in the routine health supervision of children to guarantee that children have adequate conditions for optimal learning (7, 8).

Considering that the accuracy of healthcare providers for detecting DD is low when they rely on judgment or surveillance alone (9, 10), the current recommendation is to use standardized, valid, and reliable tools for screening at specific ages (7). The new guidelines from the American Academy of Pediatrics (AAP) focus on parent/caregiver-completed developmental screening questionnaires (DSQs) for children before 5 years of age (11). If screening results suggest delayed development or if parents have concerns, the child should be referred to a comprehensive developmental evaluation, which includes the application of a developmental diagnostic assessment (7). The Bayley Scale of Infant and Toddler Development is currently one of the most used tools with this purpose.

The DSQs recommended in the updated clinical report of the AAP are the Ages and Stages Questionnaires (ASQ), subsequently updated as ASQ-3 (12); Parents' Evaluation

of Developmental Status (PEDS) and its complement Developmental Milestones (PEDS:DM) (13, 14); and the Survey of Well-being of Young Children (SWYC) (15). These questionnaires report values of sensitivity and specificity levels of 70–80%, thresholds recommended by the AAP statement in developmental screening tests (7, 11). The use of DSQs has increased in recent years because of their acceptable psychometric properties, versatility, cost-effectiveness, and parent empowerment (16–19). These questionnaires have been validated in a range of cultural and linguistic contexts and are widely used around the world in general populations and clinical samples (20–23).

In a recent study, Sheldrick et al., compared the three recommended DSQs, reporting adequate specificity and sensitivity for detecting concurrent severe DD (>70%) but low sensitivity to mild delays (24–62%) among children aged 9 months to 5 years, with no one questionnaire emerging superior (24). Despite numerous DSQ studies that analyzed concurrent validity (25–28), the predictive validity of these questionnaires has not been well-established, probably due to its complexity (29).

As background information, there are systematic literature reviews that analyze the predictive validity of developmental screening tools and developmental diagnostic assessment. In an extremely premature population, Wong et al., (30) reported a global sensitivity of 55% and a specificity of 84% of developmental assessments for identifying those children who will have cognitive problems later at school age. Luttikhuizen dos Santos et al., (31) reported that the mental coefficient of the Bayley test correlated significantly with subsequent cognitive functioning, $r = 0.61$.

In the general population, Sim et al., (32) demonstrated robust predictive validity of later disorders of language and socioemotional functioning, particularly when parent-report tools were used. In a recent publication, Cairney et al., (33) analyzed the predictive value of preschool developmental assessment on later educational outcomes in high-income countries, showing a consistent association between relatively poor early child development and later educational difficulties. They report ASQ as having the best correlation despite including only one study using ASQ in their review (33). Although these studies suggest robust predictive ability of the DSQs, none of the published studies analyzed the DSQs as a whole. We are not aware of any other publication to date that systematically reviews studies exploring an association between DSQ and later cognitive or educational performance.

Abbreviations: DSQ, developmental screening questionnaires; ASQ, Ages and Stages Questionnaires; DD, developmental delay; AAP, American Academy of Pediatrics; PDS, Parents' Evaluation of Developmental Status; PEDS, Parents' Evaluation of Developmental Status; DM, Developmental Milestones; SWYC, Survey of Well-being of Young Children; QUADAS-2, Quality of Diagnostic Accuracy Studies; AUC, area under the curve; OR, Odds Ratio.

The objective of this review is to assess in the current literature the value of AAP recommended DSQs (ASQ, PEDS, SWYC) administered between 0 and 5 years of age, for predicting long-term cognitive achievement and/or school performance after 1 year or more of evaluation and at/or after age 5 years, in the general population.

METHODS

Protocol and Registration
Our systematic review protocol was registered in advance with PROSPERO (International Prospective Register of Systematic Reviews) on July 5, 2020 (registration no. CRD42020183883).

Eligibility Criteria
Included studies were in English and Spanish languages from peer-reviewed articles of cohort studies, which included two or more serial developmental evaluations with at least one DSQ before 5 years of age and at least one evaluation of intelligence or academic performance during school age (at 5 years of age or later and with at least 1 year between evaluations). In the first selection, we included three types of studies: those with an early developmental assessment, based on DSQ administered before 5 years of age; studies that conducted a developmental assessment at school age with intelligence or academic performance assessments in cohorts previously assessed with any DSQ; and finally, those that described the association between DSQ and school age assessment.

We included cohorts assessed with DSQ (ASQ, PEDS, and SWYC) applied in general populations, in any condition (whether completed by parents, education professionals, with or without assistance in completing it). We accepted those cases in which adjustments of the original test have been made to local conditions (including language translations, sociocultural adaptations, and/or validation process).

We excluded studies in which the developmental screening was performed after 5 years of age; studies that included concurrent evaluations or with <1-year difference between the screening test and the learning/intelligence evaluation; studies focused on children with known conditions or disease that severely affects development and cognition, such as genetic and/or metabolic diseases. We excluded prevalence and case–control studies, because of potentially overestimating the properties of the test, and case series (34).

Data Sources
A systematic search was carried out in Cochrane, MEDLINE PubMed, CINAHL, EMBASE, Web of Science, Scielo, and Scopus databases (until March 13, 2021) to identify the literature published. For the systematic search, we used the following terms: "infant," "child, preschool" for population identification. The index tests were identified using the terms: "surveys and questionnaires," "developmental screening," "Ages and Stages," "Parents Evaluation of Developmental Status," "Survey of Well-being of Young Children," "parents' evaluation." Finally, the terms used to identify the reference test were "intelligence test", "developmental disabilities," "intellectual disability," "intelligence," "academic performance," "intellectual quotient."

To complete the search, the authors of the DSQ were contacted to identify additional studies that met the inclusion criteria.

Study Selection
A multiple-stage process was used to identify the studies and the cohorts behind them. First, two authors screened the titles and abstracts of studies retrieved from the electronic search for possible inclusion based on the predefined inclusion criteria. Second, forward and backward citation follow-up for each of the previously identified studies was done using Google Scholar–related references. The full text of all relevant studies identified was evaluated to select studies for final inclusion.

To identify and match the cohorts in the different publications reported separately, authors, site, and characteristics of the studied populations were considered. Although each cohort could have several published studies, only those that contributed data for either early developmental assessment with DSQ and/or academic or cognitive tests were included in the review.

Data Extraction
All information included was either published or extracted from published cohorts with the help of the authors. A data extraction form was completed for each cohort. The authors of the different cohorts were contacted to verify the cohorts and to corroborate the information extracted and to request additional information necessary to complete the data: author, study design, site, population, sampling method, sample size, age at DSQ, and cognitive/academic assessment and scoring method.

When children had more than one evaluation, each DSQ assessment was considered as a separate point for the analysis. When there was more than one simultaneous assessment of academic or cognitive performance, the cognitive assessment was considered as the most objective.

Evaluation of Risk of Bias
Two reviewers independently evaluated the risk of bias in each study using the Quality of Diagnostic Accuracy Studies version 2 (QUADAS-2) checklist. Each study was given a grade of "low," "high," or "unclear" for risk of bias and concerns regarding their applicability (35). Any disagreement between reviewers was resolved by consensus.

Data Synthesis
A qualitative analysis of the results was performed and summarized. The population characteristics, type of reference standard, index test, and reported comparison measures were summarized for each cohort [area under the curve (AUC), sensitivity, specificity, positive and negative predictive values, odds ratio (OR), correlation coefficients]. When necessary, the findings from the comparison measures were recalculated based on the more exact information provided by the corresponding authors. Based on sensitivity, specificity, and predictive values, 2 × 2 tables were constructed, and the summary receiver operating
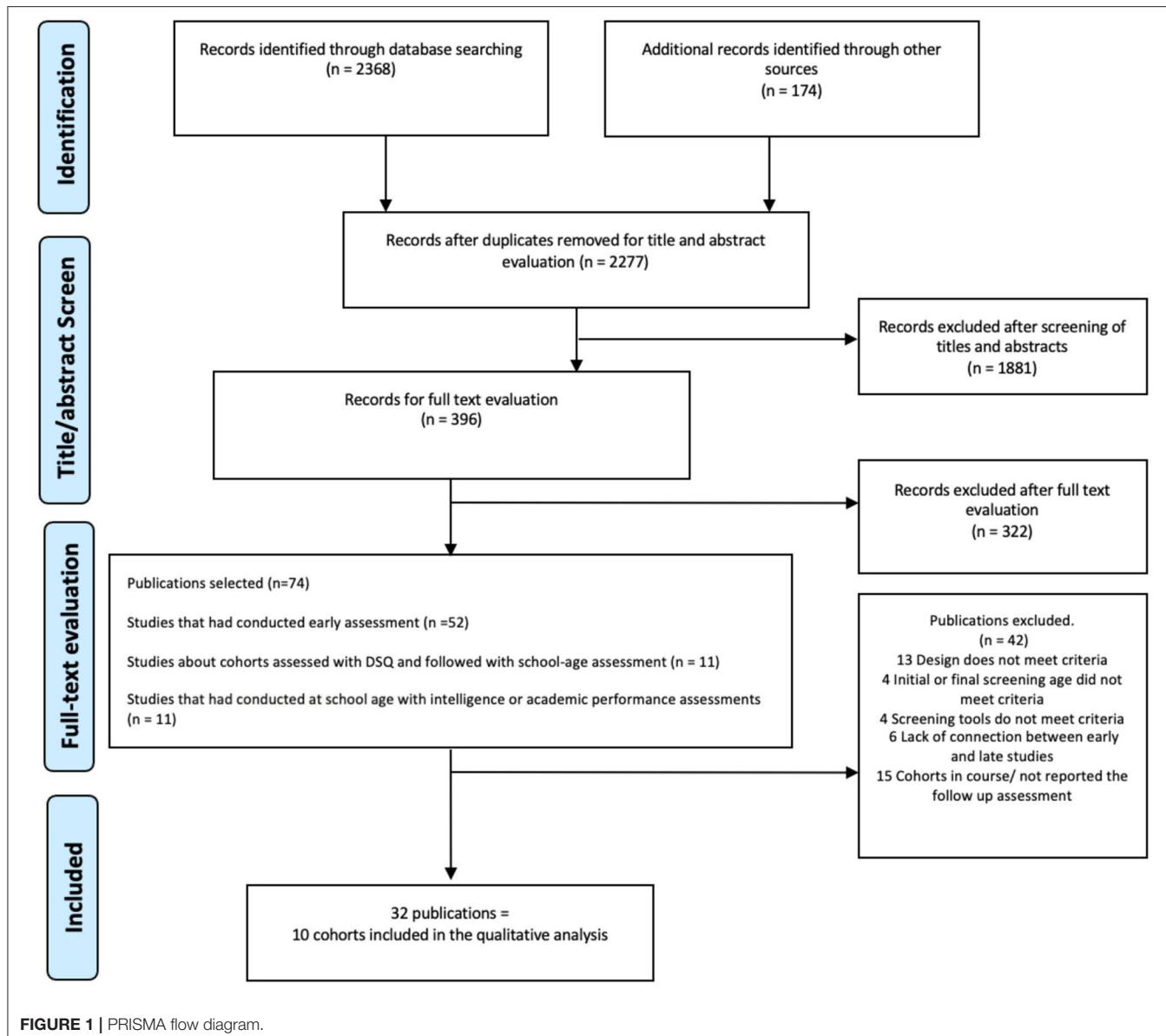
FIGURE 1 | PRISMA flow diagram.

characteristic and forest plot were calculated using RevMan 5.0. Because of significant heterogeneity, no summary measures were calculated.

RESULTS

The literature search yielded 2,277 citations after excluding duplicates. A total of 396 studies were selected for full text review, selecting 74 studies. Of these, 32 publications met the inclusion criteria, corresponding to 10 cohort studies (**Figure 1**). The cohorts studied came from South America: Chile (19, 27, 36, 37) and Colombia (26, 38, 39); Europe: Catalonia, Spain (40–43), the Netherlands (28, 44–47); France Poitiers and Nancy (48) and France Loire (49–51) and Norway (52–54); Asia: South Korea (55–57) and North India (58–62); Oceania: Australia (63).

Authors from seven of the 10 included cohorts reviewed and completed the data extraction form.

All the 10 cohorts included ASQ assessments. Only one prospective study using PEDS was identified but did not meet the inclusion criteria for this review because of the age of the children at first assessment (64). No cohorts conducted with SWYC were identified.

Two of the cohorts used abbreviated forms of ASQ, including only some domains (Norway and Australia) and one, an extended form of the test (Colombia). Except for the cohorts from Spain, France, and South Korea that used the official translated ASQ versions, the rest used locally translated and adapted versions. This information could not be obtained for the Australian cohort. All relevant characteristics are presented (**Table 1**).

Comparison measures between ASQ and cognitive/academic performance assessments in school age were extracted for eight

TABLE 1 | Qualitative summary of included cohorts.

City/Country	Sample characteristics	Initial evaluation N and gestational age distribution	Index test: Developmental assessment tool/Delay threshold	DSQ assessment age	N included/N evaluations in follow-up	Reference standard for cognitive/academic assessment/Delay threshold	Age at evaluation
Australia (63)	Different SES	Total: 50 AT : 35 MLPT: 10 EP: 5	Abbreviated ASQ-2. Gross motor and fine motor domains/Threshold for delay not defined, continuous scores for Gross Motor and Fine Motor domains were considered	4–48 months (11 ASQ forms)	33/301	WISC-IV/Threshold for delay not defined, continuous scores were considered	6–12 years
Catalonia, Spain (40–43)	Middle-high SES	Total: 179 AT: 89 LP: 90 LP	ASQ-3 Spanish edition/Score >2 SDs below the mean in any of ASQ domains	48 months	133	Standardized school test of the Education Department of Catalonia. Children scored low on at least in one of the competences measured by the test: communicative-linguistic and mathematics, determined by defined norms	8–9 years
Chile (19, 27, 36, 37)	Middle-high SES	Total: 306 AT: 119 MLPT: 124 EP: 63	ASQ-3 Chilean validation/Score > 2 SDs below the mean in any of ASQ domains	8–18–30 months	232/283	WISC-III/A score of <85 points (equivalent to <-1.5 SD) in verbal and/or performance scales	6–9 years
Colombia (26, 38, 39)	Poorest SES	Total: 770 AT: 653 PT: 117	Extended ASQ (EASQ)/Threshold for delay not defined, continuous scores were considered	6–42 months (16 ASQ forms)	470	WISC-V and school achievement.**/Threshold for delay not defined, continuous scores were considered	6–8 years
France, Loire (49–51)	46.5% upper SES	PT GA<35: 3197 GA Median 32 (IQR 30-33)	ASQ- 2 French edition/Overall ASQ scores. ROC curves were drawn to establish the optimal cutoff values	18, 24, and 36 months	1,775/4,626	GSA/Children belonging to the first decile of the GSA score (<38) were considered to have severe school difficulties.	5 years
France, Poitiers and Nancy (48)	16.9% of the families have financial difficulties	Total: 1,225 AT: 1,156 PT: 69 GA Median 40 (IQR 39-40)	ASQ- 2 French edition/Overall ASQ scores. ROC curves were drawn to establish the optimal cutoff values	36 months	939	WPSSI-III/Score of <85 in verbal, performance or full-scale	5–6 years
Netherlands (28, 44–47)	Population sample	Total: 1,983 AT: 544 MLPT: 927 EP: 512	ASQ−2 translated and adapted to Dutch/Score > 2 SDs below the mean in any of ASQ domains	48 months	1,286 (5 years) 378 (7 years)	Special education, medical childcare centers or having special educational needs in mainstream education (enrollment in special education or having special educational needs in mainstream education as criteria for developmental disability.) WISC-III NL, parental report of executive functioning, attention (TEACH-NL) and memory (AVLT), and visuomotor integration (Beery) tested/The 10th percentile, defined as a z score below 1.28, was the cutoff.	5 years 7 years

(Continued)

TABLE 1 | Continued

City/Country	Sample characteristics	Initial evaluation N and gestational age distribution	Index test: Developmental assessment tool/Delay threshold	DSQ assessment age	N included/N evaluations in follow-up	Reference standard for cognitive/academic assessment/Delay threshold	Age at evaluation
North India (58–62)	Low and middle SES	422 GA not reported	ASQ-3 "home procedure" translated and culturally adapted to Hindi/ Score below the 25th percentile any of ASQ domains	12–36 months (11 ASQ forms)	350	WISC-IV[INDIA] (index scores from three out of four subtests)/ index scores from three out of four subtests, defining general ability z-score	6–9 years
Norway (52–54)	Population sample	114,500 GA not reported	Abbreviated ASQ validated for Norwegian population: gross motor, fine motor and communication scales. Communication scale was modified as an extended form and/threshold for delay not defined, continuous scores were considered	18, 36, and 60 months	8,371	Subscale on writing within the communication domain in the vineland adaptive behavior scale-II/threshold for delay not defined, continuous scores were considered	8 years
South Korea (55–57)	Population sample	Total: 1,475 AT: 1,395 PT: 80 GA Mean 38.7 ± 1.7	Korean ASQ/Not defined	6–12–24–36–48 months	697	WPPSI-R/ A score of 89–80 was classified as low average, 79–70 as borderline, and 69 and below as indicating intellectual deficiency	5 years

SES, socioeconomic status; GA, Gestational age; AT, Born at term (38–42 weeks GA); PT, Preterm (<37 weeks GA); LP, Late preterm (34–36 weeks GA); MLPT, Moderately and late preterm (32–36 weeks GA); EP, Early Preterm (<32 weeks GA); ASQ, Ages and Stages Questionnaire; WPPSI, Wechsler Preschool and Primary Scale of Intelligence; WISC, The Wechsler Intelligence Scale for Children; GSA, The teacher-completed Global School Adaptation; IQR, Interquartile range.

**School achievement was assessed using the arithmetic (calculations) and reading comprehension subtests in the Woodcock-Muñoz Test of Achievement (WM-III), the Spanish version of the Woodcock-Johnson; and a subset of 75 words from the Test de Vocabulario en Imágenes de Peabody (TVIP), the Spanish version of the Peabody Picture Vocabulary Test-Revised.

TABLE 2 | Comparison measures and main results of included cohorts.

City/Country	AUC	OR (univariate)	Correlation/other association measures
Australia (63)	NR	NR	Regression analysis: gross motor trajectory set of predictors added 19.5% of the variance for IQ. Fine Motor not significant
Catalonia, Spain (40–43)	Total sample: 0.73	6.5 [IC 95%, 1.9–22.2]	NR
Chile (19, 27, 36, 37)	Total sample: 0.8* 8 months: 0.77 18 months: 0.75 30 months: 0.87	6.38* [IC95% 2.1–19.3]	NR
Colombia (26, 38, 39)	NR	NR	Pearson correlations on internally standardized scores Overall score for total sample r = 0.1 6–18 months: problem solving r = 0.19, other domains not significant 19–30 months: not significant 31–42 months: problem solving and communication r = 0.31; gross Motor: r = 0.25
France, Loire (49–51)	18 months: 0.66 24 months: 0.72 36 months: 0.77	3.8* [3.0–4.8]	NR
France, Poitiers, and Nancy (48)	0.78	6.7 [IC95% 3.8–12.0]	NR
Netherlands (28, 44–47)	NR	12.9* [IC95% 6.7–25.2]	NR
Norway (52–54)	NR	Communication domain: 3.2–9.8 depending on developmental trajectories	NR

AUC, Area under the curve; OR, Odds Ratio; IQ, Intellectual Quotient; NR, Not reported and not possibility of calculation with the available data.
*Values calculated in base to the information sent by the authors.
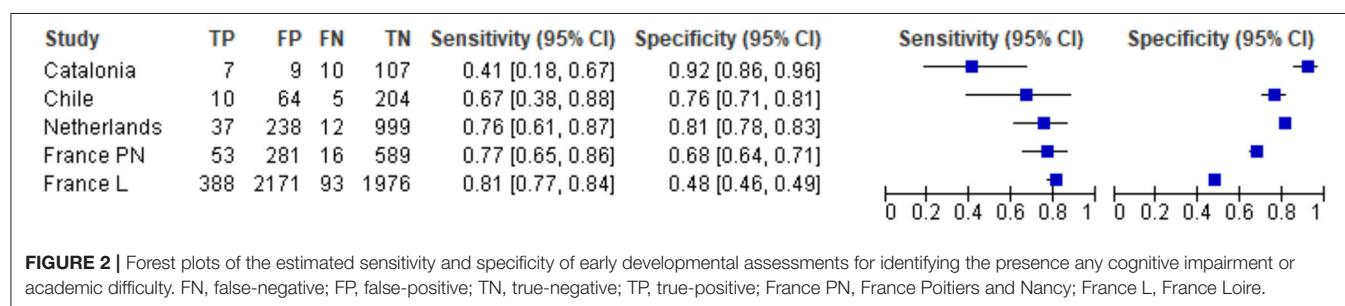
FIGURE 2 | Forest plots of the estimated sensitivity and specificity of early developmental assessments for identifying the presence any cognitive impairment or academic difficulty. FN, false-negative; FP, false-positive; TN, true-negative; TP, true-positive; France PN, France Poitiers and Nancy; France L, France Loire.

of the cohorts (**Table 2**). In the five cohorts that report results based on the entire ASQ, a positive association was shown. Using the extended ASQ, the Colombian cohort reported a low global correlation at 6–8 years of age, with higher correlations for the Problem Solving and Communication domains, whereas in the Chilean cohort, all domains independently were significant predictors of long-term cognitive difficulties, except for personal–social. In studies that analyzed abbreviated forms of ASQ, positive associations were found for communication trajectories in Norway, and for the gross motor trajectories but not for fine motor trajectories in Australia, no other domains were analyzed.

The extracted or calculated AUC ranged between 0.66 and 0.87, and the ORs were all >3 (**Table 2**). In five cohorts, a 2 × 2 table was constructed, allowing the calculation of sensitivity and specificity, showing a trade-off between them (**Figures 2**, **3**).

Risk-Of-Bias Assessment

The assessment of each of the cohort studies for each dimension of the QUADAS is detailed in **Table 3**. The risk of bias in patient selection for most cohorts was low. However, external validity was limited, because special inclusion criteria based on gestational age or socioeconomic status were used in some cohorts. In relation to index test interpretation and applicability, we found some issues of concern due to differences in the scoring method and adaptations of the test. Another source of bias was due to significant dropout rate or follow-up of specific subgroups.

DISCUSSION

Our search identified 10 cohorts including children from early age who were all assessed with ASQ and followed to school age. Eight of these cohorts describe comparison measures showing
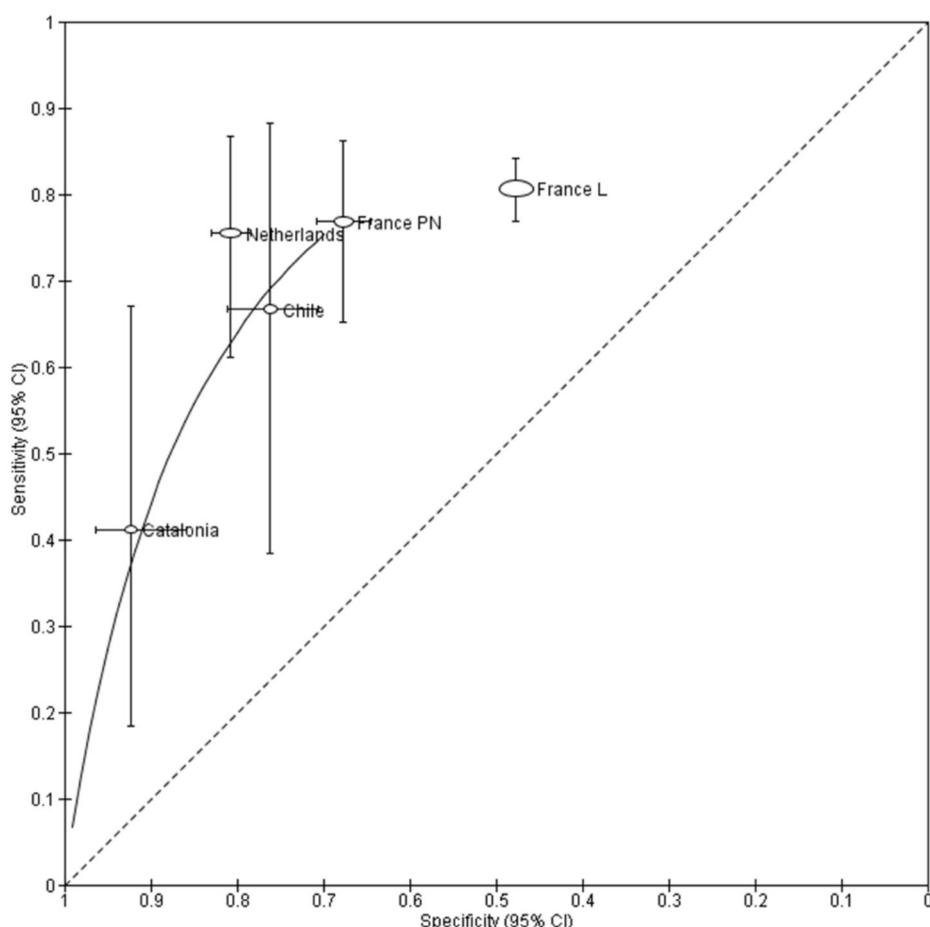
FIGURE 3 | Summary receiver operating characteristic curve of early developmental screening questionnaires for identifying any cognitive impairment or academic difficulty. Each marker displays the study and is scaled according to the sample size; the line represents the confidence interval. France PN, France Poitiers & Nancy; France L, France Loire.

adequate capacity to predict later cognitive achievement/school performance. The ORs reported were >3, and the AUC was high (0.66–0.87), showing trade-offs in sensitivity and specificity, which could be explained by the different scoring methods and thresholds used (28, 36, 42, 48, 50); the optimal cutoff point, for a screening test, is the one that yields sensitivity and specificity values >70% (7).

This review is in line with the results reported in both the reviews by Cairney et al., (33) and Sim et al., (32) that showed a consistent association between different developmental screening assessment tools and later educational performance. They reported better predictive capacity especially when using a parent-reported assessment than direct child assessment.

Our review expands these results by including studies using adapted/translated versions of ASQ, which increases the evidence supporting its widespread applicability. Some groups have adapted the form of application of the test, such as the "Home Procedure" model in India, abbreviated form in Norway and Australia, and extended ASQ in Colombia (26, 53, 58, 63). All these modifications could potentially impact the psychometric

properties of the test, as shown by Velikonja et al., (25) in the analysis of ASQ concurrent validity studies. The heterogeneity regarding age at evaluation could also impact the results. In only two cohorts, a trend to improved predictive properties of the tests with assessment age was observed (37, 50). The heterogeneity among the studies did not allow conclusions in the domain analysis, as only some of the cohorts were included analysis by domain, and two cohorts used abbreviated forms of the ASQ, including only specific domains of the test (36, 52, 63).

The cohorts emerge from different socioeconomic, clinical, and cultural backgrounds. Some cohorts were population-based, whereas others corresponded to samples with specific socioeconomic or biological characteristics, which could compromise external validity of this data. It has been shown that the prevalence of DD increases with biological and psychosocial adversity (22, 65). In extremely premature infants, the predictive validity of developmental diagnostic tests has been well-established (30, 31). These variables can modify the developmental trajectories of children and,

TABLE 3 | Risk of bias and concerns about applicability.

	Risk of bias				Concerns about applicability		
Country/city	Patient selection*	Index test**	Reference standard***	Flow and timing****	Patient selection*	Index test**	Reference standard***
Australia	Low	Not clear	Low	Low	Not clear	High	Low
Catalonia, Spain	Low	Low	Low	Low	High	Low	Low
Chile	High	Low	Low	Low	High	Low	Low
Colombia	Low	Not clear	Low	High	High	High	Low
France, Loire	Low	High	Low	High	High	Low	Low
Francia, Poitiers, and Nancy	Low	High	Low	Low	Low	Low	Low
Netherlands	Low	Low	Low	Low	Low	Low	Low
North India	Low	Low	Low	Low	High	Low	High
Norway	Low	Not clear	Low	Not clear	Low	High	High
South Korea	Low	Not clear	Low	Not clear	Low	Low	Low

*France PN, France Poitiers and Nancy; France L, France Loire; ASQ, Ages and Stages Questionnaires; EASQ, Extended ASQ; SES, Socioeconomic Status. ***Patient Selection**, High risk of bias: Non-consecutive or non-random sampling methods. In Chile the sample was an opportunity sample. In Australia: newspaper and radio announcements, and snowballing techniques considered as low risk of bias. High concerns regarding applicability: Special inclusion criterion based on gestational age or socioeconomic status characteristics of the population. Only children from medium and high SES in Chile and Catalonia. Only children from medium-low SES in Colombia and India. Only preterm in Francia L. ****Index Test**, High risk of bias: Scoring methods and thresholds were defined after reviewing the reference standard. In France P&N and France L the cutoff points for ASQ were defined based on the ROC curve of the reference standard. High concerns regarding applicability: Special adaptations that can't extrapolated to other population groups. Colombia adapted ASQ- EASQ. In Norway and Australia include only some scales of ASQ. *****Reference Standard**, High risk of bias: Inappropriate test used for population under study or if assessors were not blinded to results of early developmental test. Not reported in any cohort. High concerns regarding applicability: Nonuniversal tests. In Netherlands a series of parameters including Physical conditions and school support requirements. In Norway only one domain was used. ******Flow and Timing**, High risk of bias: Participants received different assessments, if all children were not included in follow-up or if dropout rate were >35%. In France L the dropout rate was 55%. In Colombia true positives were excluded.*

consequently, the predictive capacity of the questionnaires (44, 66).

Another factor that could alter developmental trajectories is the interventions carried out in children, data not reported by any of the cohorts. Only in the study from Catalonia was there evidence of a lack of association between the evaluation carried out at the age of 4 years and referral to support programs in development (42). It is described that in real world, referral rates for early intervention among children with positive screens ranged from 10–86% (67, 68).

of this review, they contribute to understanding the scope of developmental screening in early stages of life (69, 70).

One of the key strengths of this review is the systematic and comprehensive literature search that is highly sensitive in capturing all available data relevant to the research question in different social, cultural, and economic settings. The presented analysis was based in cohorts and not individual studies with potentially overlapping populations with the additional advantage of having contacted a significant number of authors to corroborate and better extract data.

STRENGTHS AND LIMITATIONS

The limitations of this review include great heterogeneity in population characteristics and in the way DSQ was used, such as thresholds considered and special adaptation of the questionnaires. Therefore, any summary result resulting from meta-analysis would be uninterpretable and will not allow any subgroup analysis. In addition, the variability of both initial and outcome assessments makes the mathematical synthesis of results difficult. In addition, several current ongoing cohorts are being studied and will need to be included in the future. There are currently no published studies of cohorts using SWYC and PEDS:DM as they are relatively new. Only one prospective study using PEDS was identified but did not meet the inclusion criteria for this review (64). Other studies analyzed the predictive validity of some DSQ for adaptive skills and behavior or social–emotional problems at school age. Although this is outside the purpose

CONCLUSIONS

ASQ is the most widely used DSQ in follow-up cohorts. Associations between early ASQ assessment and later cognitive achievement/school performance have been established, suggesting it is a promising tool in early child assessment in different social, cultural, and economic settings. Additional studies are needed to determine the impact of different settings, prematurity, developmental interventions, age at assessment, and test adaptations in the predictive capacity of DSQ.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

LS: conceptualized and designed the study, designed the data collection instruments and search strategy, collected and reviewed the data, carried out the initial analyses, drafted the initial manuscript, and reviewed and revised the manuscript. AM: designed the study, reviewed the data collection instruments, collected and reviewed the data, reviewed, and revised the manuscript. OC: conceptualized, designed and performed the search strategy and reviewed and revised the manuscript. PS: reviewed and revised the data collection instruments and the final manuscript. All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

ACKNOWLEDGMENTS

REFERENCES

1. Boyle CA, Boulet S, Schieve LA, Cohen RA, Blumberg SJ, Yeargin-Allsopp M, et al. Trends in the prevalence of developmental disabilities in US children, 1997-2008. *Pediatrics.* (2011) 127:1034–42. doi: 10.1542/peds.2010-2989

2. Rosenberg SA, Zhang D, Robinson CC. Prevalence of developmental delays and participation in early intervention services for young children. *Pediatrics.* (2008) 121:e1503–9. doi: 10.1542/peds.2007-1680

3. Allotey J, Zamora J, Cheong-See F, Kalidindi M, Arroyo-Manzano D, Asztalos E, et al. Cognitive, motor, behavioural and academic performances of children born preterm: a meta-analysis and systematic review involving 64 061 children. *BJOG.* (2018) 125:16–25. doi: 10.1111/1471-0528.14832

4. Shevell M. Global developmental delay and mental retardation or intellectual disability: conceptualization, evaluation, and etiology. *Pediatr Clin North Am.* (2008) 55:1071–84. doi: 10.1016/j.pcl.2008.07.010

5. Nordhov SM, Rønning JA, Dahl LB, Ulvund SE, Tunby J, Kaaresen PI. Early intervention improves cognitive outcomes for preterm infants: randomized controlled trial. *Pediatrics.* (2010) 126:e1088–94. doi: 10.1542/peds.2010-0778

6. Rao N, Sun J, Chen EE, Ip P. Effectiveness of early childhood interventions in promoting cognitive development in developing countries: a systematic review and meta-analysis. *HK J Paediatr.* (2017) 22:14–25. Available online at: http://www.hkjpaed.org/details.asp?id=1103&show=1234

7. Council on Children With Disabilities, Section on Developmental Behavioral Pediatrics, Bright Futures Steering Committee, Medical Home Initiatives for Children With Special Needs Project Advisory Committee. Identifying infants and young children with developmental disorders in the medical home: an algorithm for developmental surveillance and screening. *Pediatrics.* (2006) 118:405–20. doi: 10.1542/peds.2006-1231

8. Williams PG, Okamoto J, Lieser D, Del Conte B, Donoghue E, Earls M, et al. The pediatrician's role in optimizing school readiness. *Pediatrics.* (2016) 138:e20162293. doi: 10.1542/peds.2016-2293

9. Sheldrick RC, Merchant S, Perrin EC. Identification of developmental-behavioral problems in primary care: a systematic review. *Pediatrics.* (2011) 128:356–63. doi: 10.1542/peds.2010-3261

10. Thomas RE, Spragins W, Mazloum G, Cronkhite M, Maru G. Rates of detection of developmental problems at the 18-month well-baby visit by family physicians' using four evidence-based screening tools compared to usual care: a randomized controlled trial. *Child Care Health Dev.* (2016) 42:382–93. doi: 10.1111/cch.12333

11. Lipkin PH, Macias MM. Promoting optimal development: identifying infants and young children with developmental disorders through developmental surveillance and screening. *Pediatrics.* (2020) 145:e20193449. doi: 10.1542/peds.2019-3449

12. Squires J, Bricker D. *Ages and Stages Questionnaires User's Guide.* 3rd ed. Baltimore, USA: Paul H Brookes Publishing Co (2009).

13. Glascoe FP. The value of "parents" evaluations of developmental status' in detecting and addressing children's developmental and behavioral problems. *Diagnostique.* (1998) 23:185–203. doi: 10.1177/073724779802300401

14. Brothers KB, Page Glascoe F, Robertshaw NS. PEDS: developmental milestones - an accurate brief tool for surveillance and screening. *Clin Pediatr.* (2008) 47:271–9. doi: 10.1177/0009922807309419

15. Sheldrick RC, Perrin EC. Evidence-based milestones for surveillance of cognitive, language, and motor development. *Acad Pediatr.* (2013) 13:577–86. doi: 10.1016/j.acap.2013.07.001

16. Lipkin PH, Macias MM, Chen B, Coury D, Gottschlich EA. Trends in pediatricians' developmental screening: 2002–2016. *Pediatrics.* (2020) 145:e20190851. doi: 10.1542/peds.2019-0851

17. Schonwald A, Horan K, Huntington N. Developmental screening: is there enough time? *Clin Pediatr.* (2009) 48:648–55. doi: 10.1177/0009922809334350

18. Schonwald A, Huntington N, Chan E, Risko W, Bridgemohan C. Routine developmental screening implemented in urban primary care settings: more evidence of feasibility and effectiveness. *Pediatrics.* (2009) 123:660–8. doi: 10.1542/peds.2007-2798

19. Armijo I, Schonhaut L, Cordero M. Validation of the chilean version of the ages and stages questionnaire (ASQ-CL) in community health settings. *Early Hum Dev.* (2015) 91:671–6. doi: 10.1016/j.earlhumdev.2015.10.001

20. Small JW, Hix-Small H, Vargas-Baron E, Marks KP. Comparative use of the ages and stages questionnaires in low- and middle-income countries. *Dev Med Child Neurol.* (2019) 61:431–43. doi: 10.1111/dmcn.13938

21. Marks KP, Madsen Sjö N, Wilson P. Comparative use of the ages and stages questionnaires in the USA and Scandinavia: a systematic review. *Dev Med Child Neurol.* (2019) 61:419–30. doi: 10.1111/dmcn.14044

22. Woolfenden S, Eapen V, Williams K, Hayen A, Spencer N, Kemp L, et al. Systematic review of the prevalence of parental concerns measured by the parents' evaluation of developmental status (PEDS) indicating developmental risk. *BMC Pediatr.* (2014) 14:1–13. doi: 10.1186/1471-2431-14-231

23. Rousseau M, Dionne C, Savard RT, Schonhaut L, Londono M. Translation and cultural adaptation of the ages and stages questionnaires (ASQ) worldwide. *J Dev Behav Pediatr.* (2021) 42:490–501. doi: 10.1097/DBP.0000000000000940

24. Sheldrick RC, Marakovitz S, Garfinkel D, Carter AS, Perrin EC. Comparative accuracy of developmental screening questionnaires. *JAMA Pediatr.* (2020) 174:366–74. doi: 10.1001/jamapediatrics.2019.6000

25. Velikonja T, Calderon A, Sleed M, Deighton J. The psychometric properties of the ages & stages questionnaires for ages 2-2. 5: a systematic review. *Child Care Health Dev.* (2017) 43:1–17. doi: 10.1111/cch.12397

26. Rubio-Codina M, Araujo MC, Attanasio O, Mu P. Concurrent validity and feasibility of short tests currently used to measure early childhood development in large scale studies. *PLoS ONE.* (2016) 11:e0160962. doi: 10.1371/journal.pone.0160962

27. Schonhaut L, Armijo I, Schönstedt M, Alvarez J, Cordero M. Validity of the ages and stages questionnaires in term and preterm infants. *Pediatrics.* (2013) 131:e1468–74. doi: 10.1542/peds.2012-3313

28. Kerstjens JM, Bos AF. ten Vergert EMJ, de Meer G, Butcher PR, Reijneveld SA. Support for the global feasibility of the ages and stages questionnaire as developmental screener. *Early Hum Dev.* (2009) 85:443–7. doi: 10.1016/j.earlhumdev.2009.03.001

29. Marks K, Glascoe FP, Aylward GP, Shevell MI, Lipkin PH, Squires JK. The thorny nature of predictive validity studies on screening tests for developmental-behavioral problems. *Pediatrics.* (2008) 122:866–8. doi: 10.1542/peds.2007-3142

30. Wong HS, Santhakumaran S, Cowan FM, Modi N. Developmental assessments in preterm children: a meta-analysis. *Pediatrics.* (2016) 138:e20160251. doi: 10.1542/peds.2016-0251

31. Luttikhuizen dos Santos ES, de Kievit JF, Königs M, van Elburg RM, Oosterlaan J. Predictive value of the bayley scales of infant development on development of very preterm/very low birth weight children: a meta-analysis. *Early Hum Dev.* (2013) 89:487–96. doi: 10.1016/j.earlhumdev.2013.03.008

32. Sim F, Thompson L, Marryat L, Ramparsad N, Wilson P. Predictive validity of preschool screening tools for language and behavioural difficulties: a PRISMA systematic review. *PLoS ONE.* (2019) 14:e0211409. doi: 10.1371/journal.pone.0211409

33. Cairney DG, Kazmi A, Delahunty LM. The predictive value of universal preschool developmental assessment in identifying children with later educational difficulties : a systematic review. *PLoS ONE.* (2021) 16:e0247299. doi: 10.1371/journal.pone.0247299

34. Bossuyt PM, Leeflang MM. Chapter 6: developing criteria for including studies. In: Bossuyt P, editor. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 0.4 [updated September 2008].* The Cochrane Collaboration (2008). Available online at: https://methods.cochrane.org/sdt/handbook-dta-reviews

35. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* (2011) 155:529–38. doi: 10.7326/0003-4819-155-8-201110180-00009

36. Schonhaut L, Pérez M, Armijo I, Maturana A. Comparison between ages & stages questionnanaire and bayley scales, to predict cognitive delay in school age. *Early Hum Dev.* (2020) 141:104933. doi: 10.1016/j.earlhumdev.2019.104933

37. Schonhaut BL, Pérez RM, Castilla FAM, Castro MS, Salinas AP, Armijo RI. Predictive value of ages & stages questionnaires for cognitive performance during early child education. *Rev Chil Pediatr.* (2017) 88:35–40. doi: 10.1016/j.rchipe.2016.08.008

38. Rubio-Codina M, Grantham-McGregor S. Predictive validity in middle childhood of short tests of early childhood development used in large scale studies compared to the Bayley-III, the family care indicators, height-for-age, and stunting: a longitudinal study in Bogota, Colombia. *PLoS ONE.* (2020) 15:e0231317. doi: 10.1371/journal.pone.0231317

39. Rubio-Codina M, Araujo MC, Attanasio O, Grantham-McGregor S. *Validez concurrente y viabilidad de pruebas cortas comúnmente usadas para medir el desarrollo infantil temprano en estudios a gran escala: Metodología y resultados.* Banco Interamericano del Desarrollo Division de protección social y salud (2016). Available online at: https://publications.iadb.org/bitstream/handle/11319/7823/Validez-concurrente-y-viabilidad-de-pruebas-cortas-comunmente-usadas-para-medir-el-desarrollo-infantil-temprano-en-estudios-a-gran-escala-metodologia-y-resultados.pdf (accessed April 18, 2021).

40. Demestre X, Schonhaut L, Morillas J, Martínez-Nadal S, Vila C, Raspall F, et al. Development deficit risks in the late premature newborn: evaluation at 48 months using the ages stages questionnaires®. *An Pediatr.* (2016) 84:39–45. doi: 10.1016/j.anpede.2015.09.023

41. Schonhaut L, Martinez-Nadal S, Armijo I, Demestre X. Reliability and agreement of ages and stages questionnaires®: results in late preterm and term-born infants at 24 and 48 months. *Early Hum Dev.* (2019) 128:55–61. doi: 10.1016/j.earlhumdev.2018.11.008

42. Martínez-Nadal S, Schonhaut L, Armijo I, Demestre X. Predictive value of the ages and stages questionnaire® for school performance and school intervention in late preterm and term-born children. *Child Care Health Dev.* (2021) 47:103–111. doi: 10.1111/cch.12814

43. Martínez-Nadal S, Demestre X, Schonhaut L, Muñoz SR, Sala P. Impact of neonatal morbidity on the risk of developmental delay in late preterm infants. *Early Hum Dev.* (2018) 116:40–6. doi: 10.1016/j.earlhumdev.2017.11.001

44. Hornman J, de Winter AF, Kerstjens JM, Bos AF. Stability of developmental problems after school entry. *J Pediatr.* (2017) 187:73–9. doi: 10.1016/j.jpeds.2017.05.022

45. Kerstjens JM, de Winter AF, Bocca-Tjeertes IF, ten Vergert EMJ, Reijneveld SA, Bos AF. Developmental delay in moderately preterm-born children at school entry. *J Pediatr.* (2011) 159:92–8. doi: 10.1016/j.jpeds.2010.12.041

46. Cserjesi R, van Braeckel KNJA, Butcher PR, Kerstjens JM, Reijneveld SA, Bouma A, et al. Functioning of 7-year-old children born at 32 to 35 weeks' gestational age. *Pediatrics.* (2012) 130:e838–46. doi: 10.1542/peds.2011-2079

47. Dotinga BM, Eshuis MS, Bocca-Tjeertes IF, Kerstjens JM, van Braeckel KNJA, Reijneveld SA, et al. Longitudinal growth and neuropsychological functioning at age 7 in moderate and late preterms. *Pediatrics.* (2016) 138:e20153638. doi: 10.1542/peds.2015-3638

48. Charkaluk M, Rousseau J, Calderon J, Bernard JY. Ages and stages questionnaire at 3 years for predicting IQ at 5–6 years. *Pediatrics.* (2017) 139:e20162798. doi: 10.1542/peds.2016-2798

49. Flamant C, Branger B, Tich SNT, de la Rochebrochard E, Savagner C, Berlie I, et al. Parent-completed developmental screening in premature children: a valid tool for follow-up programs. *PLoS ONE.* (2011) 6:e20004. doi: 10.1371/journal.pone.0020004

50. Halbwachs M, Muller JB, Tich SNT, Gascoin G, Chauty-Frondas A, Branger B, et al. Predictive value of the parent-completed ASQ for school difficulties in preterm-born children <35 weeks' GA at five years of age. *Neonatology.* (2014) 106:311–6. doi: 10.1159/000363216

51. Halbwachs M, Muller JB, The Tich SN, de La Rochebrochard E, Gascoin G, Branger B, et al. Usefulness of parent-completed ASQ for neurodevelopmental screening of preterm children at five years of age. *PLoS ONE.* (2013) 8:e71925. doi: 10.1371/journal.pone.0071925

52. Fufen F, Synnve S, Mari VW, Patricia E, Ragnhild BN, Espen R, et al. Predicting literacy skills at 8 years from preschool language trajectories: a population-based cohort study. *J Speech Lang Hear Res.* (2020) 63:2752–62. doi: 10.1044/2020_JSLHR-19-00286

53. Wood ME, Frazier JA, Nordeng HME, Lapane KL. Longitudinal changes in neurodevelopmental outcomes between 18 and 36 months in children with prenatal triptan exposure: findings from the norwegian mother and child cohort study. *BMJ Open.* (2016) 6:e011971. doi: 10.1136/bmjopen-2016-011971

54. Schjølberg S, Psychol C, Eadie P, Zachrisson HD, Øyen A-S, Prior M. Predicting language development at age 18 months: data from the norwegian mother and child cohort study. *J Dev Behav Pediatr.* (2011) 32:375–83. doi: 10.1097/DBP.0b013e31821bd1dd

55. Bhang SY, Ha E, Park H, Ha M, Hong YC, Kim BN, et al. Maternal stress and depressive symptoms and infant development at six months: the mothers and children's environmental health (MOCEH) prospective study. *Korean Med Sci.* (2016) 31:843–51. doi: 10.3346/jkms.2016.31.6.843

56. Shah S, Jeong KS, Park H, Hong YC, Kim Y, Kim B, et al. Environmental pollutants affecting children's growth and development: collective results from the MOCEH study, a multi-centric prospective birth cohort in Korea. *Environ Int.* (2020) 137:105547. doi: 10.1016/j.envint.2020.105547

57. Jeong KS, Park H, Ha E, Shin J, Hong YC, Ha M, et al. High maternal blood mercury level is associated with low verbal IQ in children. *J Korean Med Sci.* (2017) 32:1097–104. doi: 10.3346/jkms.2017.32.7.1097

58. Kvestad I, Taneja S, Kumar T, Bhandari N, Strand TA, Hysing M. The assessment of developmental status using the ages and Stages questionnaire-3 in nutritional research in north Indian young children. *Nutr J.* (2013) 12:1–11. doi: 10.1186/1475-2891-12-50

59. Kvestad I, Hysing M, Shrestha M, Ulak M, Thorne-Lyman AL, Henjum S, et al. Vitamin B-12 status in infancy is positively associated with development and cognitive functioning 5 y later in Nepalese children. *Am J Clin Nutr.* (2017) 105:1122–31. doi: 10.3945/ajcn.116.144931

60. Kvestad I, Taneja S, Upadhyay RP, Hysing M, Bhandari N, Strand TA. Vitamin B12, folate, and cognition in 6- To 9-year-olds: a randomized controlled trial. *Pediatrics.* (2020) 145:e20192316. doi: 10.1542/peds.2019-2316

61. Chowdhury R, Taneja S, Kvestad I, Hysing M, Bhandari N, Strand TA. Vitamin D status in early childhood is not associated with cognitive development and linear growth at 6 – 9 years of age in North Indian children : a cohort study. *Nutr J.* (2020) 19:1–9. doi: 10.1186/s12937-020-00530-2

62. Chowdhury R, Taneja S, Bhandari N, Kvestad I, Strand TA, Bhan MK. Vitamin-D status and neurodevelopment and growth in young north Indian children: a secondary data analysis. *Nutr J.* (2017) 16:1–8. doi: 10.1186/s12937-017-0285-y

63. Piek JP, Dawson L, Smith LM, Gasson N. The role of early fine and gross motor development on later motor and cognitive ability. *Hum Mov Sci.* (2008) 27:668–81. doi: 10.1016/j.humov.2007.11.002

64. Wake M, Gerner B, Gallagher S. Does parents' evaluation of developmental status at school entry predict language, achievement, and quality of life 2 years later? *Ambul Pediatr.* (2005) 5:143–9. doi: 10.1367/A04-162R.1

65. Potijk MR, Kerstjens JM, Bos AF, Reijneveld S, de Winter AF. Developmental delay in moderately preterm-born children with low socioeconomic status: risks multiply. *J Pediatr.* (2013) 163:1289–95. doi: 10.1016/j.jpeds.2013.07.001

66. Aylward GP. Developmental screening and assessment: what are we thinking? *J Dev Behav Pediatr.* (2009) 30:169–73. doi: 10.1097/DBP.0b013e31819f1c3e

67. Sheldrick RC, Breuer DJ, Hassan R, Chan K, Polk DE, Benneyan J, et al. System dynamics model of clinical decision thresholds for the detection of developmental-behavioral disorders. *Implement Sci.* (2016) 11:156. doi: 10.1186/s13012-016-0517-0

68. Wallis KE, Rivera LBD, Guthrie W, Bennett AE, Mandell DS, Miller JS. Provider responses to positive developmental screening: disparities in referral practices? *J Dev Behav Pediatr.* (2021) 42:23–31. doi: 10.1097/DBP.0000000000000855

69. Sananes R, Goldberg CS, Newburger JW, Hu C, Trachtenberg F, Gaynor JW, et al. Six-year neurodevelopmental outcomes for children with single-ventricle physiology. *Pediatrics.* (2021) 147:e2020014589. doi: 10.1542/peds.2020-014589

70. Piek JP, Barrett NC, Smith LM, Rigoli D, Gasson N. Do motor skills in infancy and early childhood predict anxious and depressive symptomatology at school age? *Hum Mov Sci.* (2010) 29:777–86. doi: 10.1016/j.humov.2010.03.006

DISCUSIÓN

Esta tesis integra tres trabajos de investigación realizados por separado, en distintas poblaciones y con objetivos independientes, los que se encuentran entrelazados por aspectos clínicos y metodológicos comunes. El aspecto clínico estudiado fue el análisis de la capacidad predictiva del ASQ para detectar dificultades del desarrollo, en distintos contextos culturales. Desde el punto de vista metodológico se usaron distintas medias de asociación como Odds Ratio, curvas de ROC y correlaciones. Este tema se consideró relevante, dado que existente abundante evidencia acerca de la importancia de la detección temprana de dificultades del desarrollo, debido a que la neuroplasticidad cerebral infantil nos brinda oportunidades de intervención precoz y efectiva [92].

La pregunta de investigación, ¿cuál es la confiabilidad y validez del ASQ en muestras de bajo riesgo biológico y socioeconómico en Cataluña y Chile, y cuál es su capacidad para identificar aquellos niños con mayor riesgo de presentar dificultades académicas o cognitivas a largo plazo?, fue respondida a través de los objetivos en cada uno de los trabajos:

1. Evaluar la confiabilidad de los cuestionarios de 24 y 48 meses del ASQ-3 en español en una muestra de nivel socioeconómico medio alto de un Hospital de Barcelona [90].

2. Estudiar la validez predictiva del ASQ-3 aplicado a los 8, 18 o 30 meses para detectar niños que tendrán déficit cognitivo en la etapa escolar, en una muestra de nivel socioeconómico medio alto Santiago de Chile [91].

3. Determinar la capacidad del ASQ y demás CCDRP, para predecir bajo rendimiento cognitivo o académico en la población general [93].

En el primer trabajo se demostró que el ASQ es un instrumento confiable para el cribado del DSM, tanto en niños nacidos a término como en prematuros tardíos, y que existe una correlación positiva y significativa entre las evaluaciones realizadas a los 24 meses y los 48 meses. La presencia de 2 o más dominios en zona de riesgo fue predictor de déficit del desarrollo a los 48 meses, no así cuando hay sólo un dominio en zona de riesgo [90].

En el segundo estudio se demostró que el ASQ tiene una adecuada capacidad para identificar aquellos niños que tenían menor rendimiento cognitivo en la etapa escolar, comparable con la prueba de Bayley, considerado como escala de referencia. El tener déficit en algún dominio en el cuestionario ASQ-CI o en el test de Bayley-III, fue predictor de déficit cognitivo (con OR cercano a 5). Para el ASQ-CI los dominios que fueron significativos fueron comunicación, motricidad gruesa y resolución de problemas; mientras que con la escala de Bayley III, todos los dominios fueron significativos [91].

Finalmente, en el tercer trabajo, a través de una revisión sistemática de la literatura, se identificaron cohortes que siguieron niños evaluados precozmente con

CCDRP, como es el ASQ, y luego fueron seguidos y evaluados desde el punto de vista de rendimiento académico y/o desarrollo cognitivo. En las cohortes que reportaron medidas de asociación, como curvas de ROC, OR o correlaciones, destacándose asociaciones en general positivas, en especial si se consideraba el cuestionario no modificado.

Desde el punto de vista metodológico, se utilizaron los diseños apropiados para responder a cada pregunta, con el menor riesgo de sesgo posible, a modo de disponer de evidencia de alto nivel.

Para el primer trabajo, que evaluó la validez del ASQ en español en Barcelona y su capacidad predictiva entre los 2 y 4 años, utilizando un diseño de cohorte prospectiva, similar al diseño del segundo trabajo, cuyo objetivo fue evaluar la validez predictiva del ASQ validado en Santiago de Chile, para la detección de dificultades cognitivas en la etapa escolar. En el último trabajo, se realizó una revisión sistemática de pruebas diagnósticas para identificar la literatura publicada en relación con la validez predictiva del ASQ y otros CCDRP, de este modo, se pudo conocer el estado del arte en el tema y las medidas de asociación empleadas por los distintos investigadores, logrando establecer asociaciones entre el desarrollo infantil temprano y el desempeño académico y cognitivo posterior.

En el análisis de las medidas de asociación, al usar datos categóricos, es decir presencia o no de cribado positivo, frente a presencia o no de dificultades académicas, el uso de OR fue el método de elección. Las curvas de ROC fueron

utilizadas cuando se utilizaron resultados categóricos de la variable dependiente y continuos de la independiente, permitiendo obtener análisis de sensibilidad, especificidad y valores predictivos (los que dependen del punto de corte seleccionado); mientras que las correlaciones son las recomendadas cuando ambas variables son continuas. Si bien en la revisión sistemática se pudo ver la congruencia entre las distintas medias de asociación, desde el punto de vista clínico, el uso de OR y valores predictivos son de más clara interpretación y, a la larga, permiten reunir los datos en un metaanálisis.

En las siguientes líneas se profundizará en cada uno de los estudios, se discute sobre las relaciones entre las 3 publicaciones y se contrastan los hallazgos con la bibliografía publicada. Para terminar con las fortalezas, limitaciones, los aprendizajes logrados y las proyecciones. Además, se comenta sobre estudios relacionados que se desarrollaron durante el mismo periodo, pero no forman parte de la tesis, como lo es el estudio de validez predictiva del ASQ para dificultades académicas en Barcelona.

En el primer estudio de cohorte se demostró que el ASQ-3 en español es confiable para la evaluación del desarrollo al considerar el puntaje global, pero con una consistencia interna cuestionable en el análisis de dominios, comparable a lo reportado en la literatura [58, 94, 95]. Por otro lado, encontramos correlaciones entre los dominios y el puntaje total de moderadas a elevadas, coincidiendo con lo reportado por Lopes y cols. en la validación del ASQ de Portugal [94], mientras que la correlación entre los dominios fue baja, lo que es esperable, dado que apuntan a distintas habilidades del desarrollo.

La correlación fue significativa entre los puntajes obtenidos en la evaluación de los 24 y 48 meses, destacando un OR muy significativo cuando habían 2 o más áreas bajo el punto de corte en el ASQ-3 de los 24 meses. En este aspecto, no encontramos diferencias significativas entre los dominios, sino que, en la suma de estos, lo que podría guardar relación con las trayectorias del desarrollo, es que, en la mayoría de los niños de bajo riesgo biológico, es estable [7,96].

Este resultado se condice con la definición de deficiencia global del desarrollo para niños menores de 5 años, que corresponde a la presencia de un déficit significativo (es decir, un rendimiento inferior a -2 DS) en 2 o más dominios del desarrollo. Es sabido que la deficiencia global del desarrollo es un continuo con la discapacidad intelectual a largo plazo [97]. Por este motivo, distintos investigadores han revisado la definición de déficit en ASQ, demostrándose una mayor especificidad al considerar 2 o más dominios bajo el punto de corte o al utilizar el puntaje total [33, 72, 73]. La cohorte de Barcelona fue seguida a más largo plazo, y se evidenció que la evaluación con ASQ-3 realizada a los 48 meses tiene una capacidad aceptable para predecir el menor rendimiento académico a la edad de 8 a 9 años (AUC 0,73, OR, 6.5 [IC 95%, 1.9-22.2]) [88].

Coincidiendo con los hallazgos del estudio de Barcelona, en el segundo estudio de cohorte, se reportó que el ASQ-3 validado en Chile tiene una adecuada capacidad para predecir el menor rendimiento cognitivo en la edad escolar, siendo estos resultados comparables con los obtenidos con la escala de Bayley, considerado como prueba de referencia. La equivalente predictibilidad podría ser esperable dada la buena correlación entre ambas, demostrada en estudios

previos[31]. Al considerar al menos un dominio en zona de riesgo, encontramos regulares valores de sensibilidad y adecuada especificidad, con excelentes valores predictivos negativos. En el análisis por edad de evaluación, se describió una mejor capacidad predictiva a los 30 meses, comparado con las evaluaciones realizadas a menor edad, lo que coincide con lo reportado por Halbwachs y cols. con ASQ [64] y Doyle y cols. con la escala de Bayley [98]. En el análisis por dominios, lo que resultó más significativo fue: comunicación, motricidad gruesa y resolución de problemas.

A modo de aportar a la comprensión más global de la capacidad predictiva del ASQ y otros CCDRP recomendados por la AAP, el análisis se completó con una revisión sistemática de pruebas diagnósticas. En 2 revisiones sistemáticas realizadas en población general y recientemente publicadas, se ha mostrado una asociación robusta entre las evaluaciones del desarrollo y el rendimiento académico y/o cognitivo en la etapa escolar y se ha insinuado que cuando se utilizan CCDRP, la capacidad predictiva es superior que las evaluaciones directas realizadas por profesionales [80, 99], no obstante, en ambos estudios hubo poca representación de CCDRP.

A pesar de encontrar diferencias en las características de las cohortes y las medidas de asociación empleadas [28–31, 33, 64, 68, 79, 85, 87–91], en su mayoría, las asociaciones, fueron positivas, con un efecto espejo entre sensibilidad y especificidad, que podría ser explicados por los diferentes puntos de corte y criterios utilizados en la definición de déficit del desarrollo [64, 79, 85, 88, 91].

Es importante considerar que la predicción en sí es dificultosa porque existen numerosos factores que pueden modificar la historia natural del desarrollo, incluyendo los rápidos cambios que experimentan los niños, las variables biológicas y ambientales, además de las intervenciones que pueden tener los niños a partir de las evaluaciones [100, 101]. Uno de los posibles riesgos de la aplicación de los cribados de desarrollo es la sobre-derivación [33, 102, 103], aunque clínicamente, podría ser más relevante la infra-derivación o sub-pesquisa de niños con dificultades.

Desde ese punto de vista, es debatible la corrección de la edad gestacional hasta los 2 años de edad en niños prematuros. Los detractores argumentan que al ser menos exigentes se pierden oportunidades de intervención; al respecto, es importante mencionar que el ASQ tiene ventanas de tiempo para su aplicación, y mientras menos prematuros son los niños, y a mayor edad, se estrechan las ventanas de edad corregida y cronológica. En el estudio de Barcelona, al comparar las correlaciones entre la evaluación realizada a los 2 de edad corregida y 4 años cronológicos, los resultados fueron más consistentes en los PT que en los niños nacidos AT [90].

Estudios han demostrado que el principal valor de la monitorización del desarrollo con ASQ, se basa en su elevado valor predictivo negativo, mientras que el valor predictivo positivo es más bien bajo [64, 88, 91, 104], probablemente porque los problemas más sutiles del desarrollo pueden resolverse en el tiempo, mientras que existe menor posibilidad de que aparezcan otras dificultades a medida que crecen las exigencias que se hace a los niños, especialmente en poblaciones de mayor riesgo biológico, como son los prematuros [7].

De los estudios analizados, sólo en la cohorte de Barcelona se reportó la derivación a intervención temprana, evidenciando una baja asociación entre la evaluación con ASQ y la participación en programas de apoyo al desarrollo [88]. Este resultado no nos sorprende, ya que a nivel internacional se describe que entre el 10 y el 86% de los niños con cribados positivos son derivados a una evaluación más especializada [105, 106]. Sin duda, la derivación es superior en condiciones en que las dificultades del desarrollo son clínicamente evidentes, comparado con cuando solo son pesquisadas por un cuestionario. Es necesaria mayor evidencia para conocer su real impacto de las dificultades sutiles del desarrollo en el desempeño cognitivo posterior [100].

Limitaciones, fortalezas y aprendizajes

Las limitaciones de los estudios realizados en Barcelona y Santiago de Chile fueron las muestras pequeñas, en niños que crecieron en ambientes enriquecidos culturalmente, y que no fueron analizadas las posibles intervenciones en su desarrollo. Por otro lado, sólo se evaluó el desempeño cognitivo, sin considerar otros aspectos del desarrollo como es la motricidad o el desempeño socio emocional, aspectos que escaparon a los objetivos de los estudios, y que sin duda son de gran relevancia en el desarrollo de los niños. No obstante, la fortaleza de ambos estudios se basa en el estrecho seguimiento de los niños y la elevada adherencia a las evaluaciones.

La principal limitación de la revisión sistemática fue la gran heterogeneidad de las poblaciones incluidas, tanto a nivel biológico (como las distintas edades de los niños o el antecedente de prematuridad) como sociocultural, las distintas adaptaciones del ASQ (desde formas abreviadas a formas extendidas) y las diferentes formas de definición de déficit del desarrollo: lo que no permitió un metaanálisis y menos aún, el análisis por subgrupos.

Adicionalmente, debemos considerar que actualmente hay cohortes en curso con resultados aún no publicados, que no pudieron ser incluidas. Sin embargo, la principal fortaleza del estudio, fue la búsqueda sistemática y exhaustiva de la literatura que permitió identificar la información disponible en relación con la pregunta de investigación y el contacto con los autores de las cohortes, para corroborar la información obtenida.

Otro aspecto discutible, es el enfoque utilizado, que fue de pruebas diagnósticas en lugar de pronóstico [107]. La elección de la metodología de pruebas diagnósticas diferidas, se fundamentó en el hecho que el desarrollo infantil temprano (antes de los 5 años de vida) y la discapacidad intelectual, son un continuo, y la mayoría de los niños tienen trayectorias estables en su desarrollo [97]. Durante el periodo en que se realizó esta tesis, fueron publicados otros estudios de validez predictiva de pruebas de cribado, con distintos matices; ambos utilizaron las medidas de sensibilidad y especificidad y OR, y Cairney, además, regresiones lineares [80, 99]. En otros 2 metaanálisis sobre la capacidad predictiva de pruebas diagnósticas en prematuros extremos, Wong y cols. consideró curvas de ROC, sensibilidad y especificidad [81], mientras que Luttikhuizen, Dos Santos y cols., utilizó

un modelo predictivo con correlaciones [82]. Sin duda, esta es una historia que se está empezando a escribir, y está marcando una pauta para las investigaciones de validez predictiva que se encuentran en curso.

Reflexiones y definición de líneas de investigación

Durante la pandemia de COVID-19 tomaron especial relevancia los CCDRP, debido a la necesidad de empoderar a los padres y desarrollar atención vía remota. Los CCDRP que están liderando la oferta son aquellos recomendados por la AAP. En Santiago de Chile y Barcelona, paralelamente, comenzaron a adaptarse, validarse y utilizarse el ASQ hace más de 10 años, en poblaciones de características socioeconómicas y culturales similares. En España, se proyecta continuar la investigación, ampliando la muestra para el seguimiento de los prematuros tardíos a nivel nacional. En Clínica Alemana de Santiago, se firmó un convenio con la editorial Brookes Publishing y, a partir del año 2020, se incorporó el ASQ en versión online, para la monitorización del desarrollo de todos los lactantes y preescolares que acuden a su control de supervisión de salud, a edades específicas. Esta innovación ha tenido un impacto clínico muy significativo, y a su vez, permitirá continuar el seguimiento de los niños a largo plazo, para agregar nueva evidencia acerca de las trayectorias del desarrollo y la capacidad predictiva de los cuestionarios; incluso en poblaciones de bajo riesgo, en que las dificultades pueden ser más sutiles.

CONCLUSIONES

Las conclusiones derivadas de cada uno de los estudios contribuyen a la práctica clínica, al demostrarse que, en Barcelona, el ASQ e confiable, y la evaluación realizada a los 2 años tiene capacidad para predecir las dificultades del desarrollo a los 4 años. A más largo plazo, en Chile, se demostró que las evaluaciones realizadas a los 8, 18 y 30 meses, tienen una adecuada capacidad predictiva de las dificultades cognitivas en la etapa escolar.

Estos resultados fueron corroborados a través de una revisión sistemática de la literatura. El elevado valor predictivo negativo para el déficit cognitivo y/o académico, apoya su uso como prueba de cribado del desarrollo, lo que ha sido reportado por otros autores [104]. Se demostró un efecto espejo entre la sensibilidad y la especificidad, lo que depende fundamentalmente del criterio y punto de corte utilizado.

La ventaja de usar CCDRP, como es el ASQ, se debe a su bajo costo: no requiere de personal entrenado para su aplicación y empodera a los padres en la vigilancia del neurodesarrollo de sus hijos. Esto permite la monitorización continua, identificando el momento en que un niño presenta dificultades, para poder implementar una intervención oportuna.

Sin duda, se requieren estudios adicionales para conocer el impacto de las distintas adaptaciones, las condiciones biológicas y socioculturales, de las intervenciones tempranas y de las distintas edades de evaluación, en la capacidad predictiva de los CCDRP.
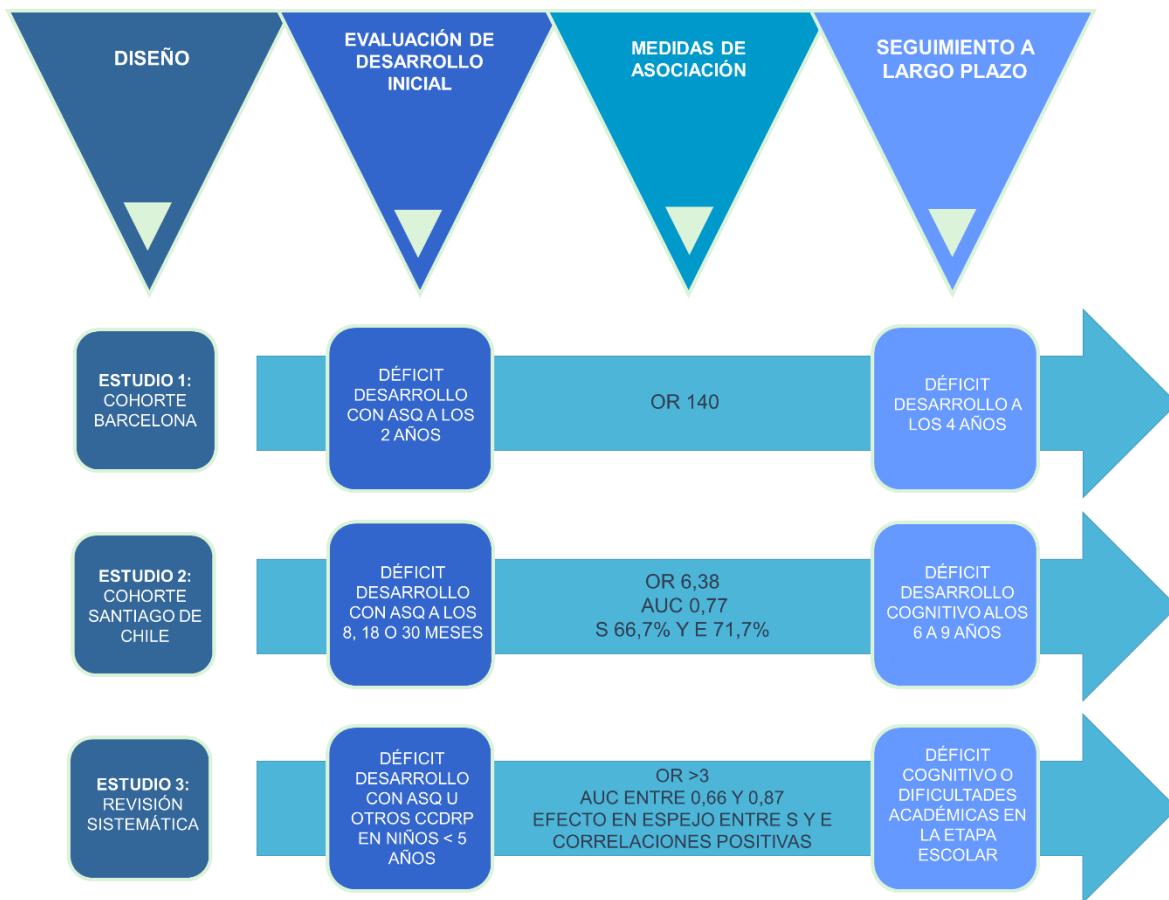
Ilustración 5: Infografía con principales conclusiones tesis

CCDPR: Cuestionarios de Cribado de Desarrollo basados en el Reporte de Padres o Cuidadores

ASQ: Ages and Stages Questionnaire, OR: Odds Ratio,

AUC: Área bajo la curva, S: Sensibilidad, E: Especificidad

REFERENCIAS Y BIBLIOGRAFÍA

1.   Guralnick MJ. Effectiveness of early intervention for vulnerable children: a developmental perspective. Am J Ment Retard. 1998;102(4):319-45. doi: 10.1352/0895-8017(1998)102<0319:eoeifv>2.0.co;2.

2.   Gallaher MM, Christakis DA, Connell FA. Health care use by children diagnosed as having developmental delay. Arch Pediatr Adolesc Med. 2002;156(3):246-51. doi: 10.1001/archpedi.156.3.246.

3.   Boulet SL, Boyle CA, Schieve LA. Health care use and health and functional impact of developmental disabilities among US children, 1997-2005. Arch Pediatr Adolesc Med. 2009;163(1):19-26. doi: 10.1001/archpediatrics.2008.506.

4.   Kodric J, Sustersic B, Paro-Panjan D. Relationship between neurological assessments of preterm infants in the first 2 years and cognitive outcome at school age. Pediatr Neurol. 2014;51(5):681-7. doi: 10.1016/j.pediatrneurol.2014.07.024.

5.   Shevell M, Majnemer A, Platt RW, Webster R, Birnbaum R. Developmental and functional outcomes in children with global developmental delay or developmental language impairment. Dev Med Child Neurol. 2005;47(10):678-83. doi: 10.1017/S0012162205001386.

6.   Squarza C, Picciolini O, Gardon L, Ravasi M, Giannì ML, Porro M, et al. Seven years cognitive functioning and early assessment in extremely low birth weight children. Front Psychol. 2017;8:1257. doi:10.3389/fpsyg.2017.01257

7.   Hornman J, de Winter AF, Kerstjens JM, Bos AF, Reijneveld SA. Stability of Developmental Problems after School Entry of Moderately-Late Preterm and Early Preterm-Born Children. J Pediatr. 2017;187:73-79. doi: 10.1016/j.jpeds.2017.05.022..

8.   Rosenberg SA, Zhang D, Robinson CC. Prevalence of developmental delays and participation in early intervention services for young children. Pediatrics. 2008;121(6):e1503-9. doi: 10.1542/peds.2007-1680.

9.   Boyle CA, Boulet S, Schieve LA, Cohen RA, Blumberg SJ, Yeargin-Allsopp M, Visser S, Kogan MD. Trends in the prevalence of developmental disabilities in US children, 1997-2008. Pediatrics. 2011;127(6):1034-42. doi: 10.1542/peds.2010-2989.

10. Maenner MJ, Blumberg SJ, Kogan MD, Christensen D, Yeargin-Allsopp M, Schieve LA. Prevalence of cerebral palsy and intellectual disability among children identified in two U.S. National Surveys, 2011-2013. Ann Epidemiol. 2016;26(3):222-6. doi: 10.1016/j.annepidem.2016.01.001.

11. Morse SB, Zheng H, Tang Y, Roth J. Early school-age outcomes of late preterm infants. Pediatrics. 2009;123(4):e622-9. doi: 10.1542/peds.2008-1405.

12. Allotey J, Zamora J, Cheong-See F, Kalidindi M, Arroyo-Manzano D, Asztalos E, van der Post J, Mol BW, Moore D, Birtles D, Khan KS, Thangaratinam S. Cognitive, motor, behavioural and academic performances of children born preterm: a meta-analysis and systematic review involving 64 061 children. BJOG. 2018;125(1):16-25. doi: 10.1111/1471-0528.14832.

13. Kerstjens JM, de Winter AF, Bocca-Tjeertes IF, Bos AF, Reijneveld SA. Risk of developmental delay increases exponentially as gestational age of preterm infants decreases: a cohort study at age 4 years. Dev Med Child Neurol. 2012;54(12):1096-101. doi: 10.1111/j.1469-8749.2012.04423.x.

14. Potijk MR, Kerstjens JM, Bos AF, Reijneveld SA, de Winter AF. Developmental delay in moderately preterm-born children with low socioeconomic status: risks multiply. J Pediatr. 2013;163(5):1289-95. doi: 10.1016/j.jpeds.2013.07.001.

15. Richards JL, Chapple-McGruder T, Williams BL, Kramer MR. Does neighborhood deprivation modify the effect of preterm birth on children's first grade academic performance? Soc Sci Med. 2015;132:122-31. doi: 10.1016/j.socscimed.2015.03.032.

16. Schonhaut L, Armijo I, Pérez M. Gestational age and developmental risk in moderately and late preterm and early term infants. Pediatrics. 2015;135(4):e835-41. doi: 10.1542/peds.2014-1957.

17. Schonhaut L, Álvarez J, Salinas P. El pediatra y la evaluación del desarrollo psicomotor. Rev Chil Pediatr.2008;79(1):26–31.

18. Sheldrick RC, Merchant S, Perrin EC. Identification of developmental-behavioral problems in primary care: A systematic review. Pediatrics. 2011;128(2):356–63. doi: 10.1542/peds.2010-3261.

19. Council on Children With Disabilities; Section on Developmental Behavioral Pediatrics; Bright Futures Steering Committee; Medical Home Initiatives for Children With Special Needs Project Advisory Committee. Identifying infants and young children with developmental disorders in the medical home: an algorithm for developmental surveillance and screening. Pediatrics. 2006;118(1):405-20. doi: 10.1542/peds.2006-1231.

20. Lipkin PH, Macias MM; Council on children with disabilities, section on developmental and behavioral pediatrics. Promoting Optimal Development: Identifying Infants and Young Children With Developmental Disorders Through Developmental Surveillance and Screening. Pediatrics. 2020;145(1):e20193449. doi: 10.1542/peds.2019-3449.

21. Squires J, Bricker D. Ages and Stages Questionnaires User's Guide. third edit. Baltimore, USA: PAUL H. Brookes Publishing Co; 2009.

22. Glascoe FP. The Value of "Parents" Evaluations of Developmental Status' in Detecting and Addressing Children's Developmental and Behavioral Problems. Diagnostique. 1998;23(4):185–203. doi: 10.1177/073724779802300401

23. Brothers KB, Glascoe FP, Robertshaw NS. PEDS: developmental milestones--an accurate brief tool for surveillance and screening. Clin Pediatr (Phila). 2008;47(3):271-9. doi: 10.1177/0009922807309419.

24. Sheldrick RC, Perrin EC. Evidence-based milestones for surveillance of cognitive, language, and motor development. Academic Pediatrics. 2013;13(6):577–86. doi: 10.1016/j.acap.2013.07.001.

25. Lipkin PH, Macias MM, Baer Chen B, Coury D, Gottschlich EA, Hyman SL, Sisk B, Wolfe A, Levy SE. Trends in Pediatricians' Developmental Screening: 2002-2016. Pediatrics. 2020;145(4):e20190851. doi: 10.1542/peds.2019-0851.

26. Radecki L, Sand-Loud N, O'Connor KG, Sharp S, Olson LM. Trends in the use of standardized tools for developmental screening in early childhood: 2002-2009. Pediatrics. 2011;128(1):14–9. doi: 10.1542/peds.2010-2180.

27. Rousseau M, Dionne C, Savard RT, Schonhaut L, Londono M. Translation and Cultural Adaptation of the Ages and Stages Questionnaires (ASQ) Worldwide: A Scoping Review. J Dev Behav Pediatr. 2021;42(6):490-501. doi: 10.1097/DBP.0000000000000940.

28. Demestre X, Schonhaut L, Morillas J, Martínez-Nadal S, Vila C, Raspall F, Sala P. Riesgo de déficits en el desarrollo en los prematuros tardíos: evaluación a los 48 meses mediante el Ages & Stages Questionnaires®. An Pediatr (Barc). 2016;84(1):39-45. doi: 10.1016/j.anpedi.2015.02.017.

29. Armijo I, Schonhaut L, Cordero M. Validation of the Chilean version of the Ages and Stages Questionnaire (ASQ-CL) in Community Health Settings. Early Hum Dev. 2015;91(12):671-6. doi: 10.1016/j.earlhumdev.2015.10.001

30. Martínez-Nadal S, Demestre X, Schonhaut L, Muñoz SR, Sala P. Impact of neonatal morbidity on the risk of developmental delay in late preterm infants. Early Hum Dev. 2018 Jan;116:40-46. doi: 10.1016/j.earlhumdev.2017.11.001.

31. Schonhaut L, Armijo I, Schönstedt M, Alvarez J, Cordero M. Validity of the ages and stages questionnaires in term and preterm infants. Pediatrics. 2013;131(5):e1468–74. doi: 10.1542/peds.2012-3313

32. Schonhaut B L, Armijo R I, Millán K T, Herreros A J, Hernández R K, Salgado V AM, et al. Comparison of traditional psychomotor development evaluation versus a self-administered test? Rev Chil Pediatr. 2010;81(6):498-505. doi:10.4067/S0370-41062010000600003

33. Schonhaut B. L, Pérez R. M, Castilla F. AM, Castro M. S, Salinas A. P, Armijo R. I. Validez del Ages & Stages questionnaires para predecir el desempeño cognitivo en los primeros años de educación escolar. Rev Chil Pediatr.2017;88(1):35–40. doi.org/10.1016/j.rchipe.2016.08.008

34. Nordhov SM, Rønning JA, Dahl LB, Ulvund SE, Tunby J, Kaaresen PI. Early intervention improves cognitive outcomes for preterm infants: randomized controlled trial. Pediatrics. 2010;126(5):e1088-94. doi: 10.1542/peds.2010-0778.

35. Anderson LM, Shinn C, Fullilove MT, Scrimshaw SC, Fielding JE, Normand J, Carande-Kulis VG; Task Force on Community Preventive Services. The effectiveness of early childhood development programs. A systematic review. Am J Prev Med. 2003;24(3 Suppl):32-46. doi: 10.1016/s0749-3797(02)00655-4.

36. Rao N, Sun J, Chen EE, Ip P. Effectiveness of early childhood interventions in promoting cognitive development in developing countries: A systematic review and meta-analysis. Hong Kong Journal of Paediatrics. 2017;22(1):14–25.

37. Ministerio de Salud. Santiago Chile. Normas técnicas de evaluación y estimulación del desarrollo psicomotor en el niño y la niña menor de 6 años. [Internet]. 2004. Available from: http://dx.doi.org/10.1016/j.cirp.2016.06.001%0Ahttp://dx.doi.org/10.1016/j.powtec.2016.12.055%0Ahttps://doi.org/10.1016/j.ijfatigue.2019.02.006%0Ahttps://doi.org/10

.1016/j.matlet.2019.04.024%0Ahttps://doi.org/10.1016/j.matlet.2019.127252%0Ahttp://dx.doi.o

38. Kendall S, Nash A, Braun A, Bastug G, Rougeaux E, Bedford H. Acceptability and understanding of the Ages & Stages Questionnaires®, Third Edition, as part of the Healthy Child Programme 2-year health and development review in England: Parent and professional perspectives. Child Care Health Dev. 2019;45(2):251-256. doi: 10.1111/cch.12639.

39. Leversen KT, Sommerfelt K, Elgen IB, Eide GE, Irgens LM, Júlíusson PB, Markestad T. Prediction of outcome at 5 years from assessments at 2 years among extremely preterm children: a Norwegian national cohort study. Acta Paediatr. 2012;101(3):264-70. doi: 10.1111/j.1651-2227.2011.02504.x

40. Roberts G, Anderson PJ, Doyle LW; Victorian Infant Collaborative Study Group. The stability of the diagnosis of developmental disability between ages 2 and 8 in a geographic cohort of very preterm children born in 1997. Arch Dis Child. 2010;95(10):786-90. doi: 10.1136/adc.2009.160283.

41. Munck P, Niemi P, Lapinleimu H, Lehtonen L, Haataja L; PIPARI Study Group. Stability of cognitive outcome from 2 to 5 years of age in very low birth weight children. Pediatrics. 2012;129(3):503-8. doi: 10.1542/peds.2011-1566.

42. Engle WA; American Academy of Pediatrics Committee on Fetus and Newborn. Age terminology during the perinatal period. Pediatrics. 2004 Nov;114(5):1362-4. doi: 10.1542/peds.2004-1915.

43. Mckinnon K, Huertas-Ceballos A. Developmental follow-up of children and young people born preterm, NICE guideline 2017. Arch Dis Child Educ Pract Ed. 2019;104(4):221-223. doi: 10.1136/archdischild-2017-314044.

44. Hurtado Suazo JA, García Reymundo M, Calvo Aguilar MJ, Ginovart Galiana G, Jiménez Moya A, Trincado Aguinagalde MJ, Demestre Guasch X. Recomendaciones para el manejo perinatal y seguimiento del recién nacido prematuro tardío. An Pediatr (Barc). 2014;81(5):327.e1-7. Spanish. doi: 10.1016/j.anpedi.2014.06.006.

45. Thomas SA, Cotton W, Pan X, Ratliff-Schaub K. Comparison of systematic developmental surveillance with standardized developmental screening in primary care. Clin Pediatr (Phila). 2012;51(2):154-9. doi: 10.1177/0009922811420711.

46. Corrigan N, Stewart M, Scott M, Fee F. Predictive value of preschool surveillance in detecting learning difficulties. Arch Dis Child. 1996;74(6):517-21. doi: 10.1136/adc.74.6.517.

47. Hirai AH, Kogan MD, Kandasamy V, Reuland C, Bethell C. Prevalence and Variation of Developmental Screening and Surveillance in Early Childhood. JAMA Pediatrics. 2018;172(9):857–66. doi: 10.1001/jamapediatrics.2018.1524.

48. Dobrez D, Lo Sasso A, Holl J, Shalowitz M, Leon S, Budetti P. Estimating the cost of developmental and behavioral screening of preschool children in general pediatric practice. Pediatrics. 2001;108(4):913–22. doi: 10.1542/peds.108.4.913.

49. Schonwald A, Horan K, Huntington N. Developmental screening: is there enough time? Clin Pediatr (Phila). 2009 Jul;48(6):648-55. doi: 10.1177/0009922809334350.

50. Schonhaut L, Salinas P, Armijo I, Schönstedt M, Álvarez J, Manríquez M. Validación de un Cuestionario Autoadministrado para la Evaluación del Desarrollo Psicomotor. Rev Chil Pediatr. 2009;80(6): 513-519. doi: 10.4067/S0370-41062009000600003

51. Pizur-Barnekow K, Erickson S, Johnston M, Bass T, Lucinski L, Bleuel D. Early identification of developmental delays through surveillance, screening, and diagnostic evaluation. Infants Young Child.. 2010;23(4):323–30. doi: 10.1097/IYC.0b013e3181f422a4

52. Bellman M, Byrne O, Sege R. Developmental assessment of children. BMJ. 2013;346: e8687. doi: 10.1136/bmj.e8687

53. Bayley N. Bayley Scales of Infant and Toddler Development Screening Test: Third Edition. 2006.

54. Woolfenden S, Eapen V, Williams K, Hayen A, Spencer N, Kemp L. A systematic review of the prevalence of parental concerns measured by the Parents' Evaluation of Developmental Status (PEDS) indicating developmental risk. BMC Pediatr. 2014;14:231. doi: 10.1186/1471-2431-14-231.

55. Simon AE, Pastor PN, Avila RM, Blumberg SJ. Socioeconomic disadvantage and developmental delay among US children aged 18 months to 5 years. J Epidemiol Community Health. 2013;67(8):689-95. doi: 10.1136/jech-2013-202610.

56. Whitesell NR, Sarche M, Trucksess C; Tribal Early Childhood Research Center SWYC Community Of Learning. The survey of well-being of young children: results of a feasibility study with American Indian and Alaska native communities. Infant Ment Health J. 2015;36(5):483-505. doi: 10.1002/imhj.21526.

57. Moreira RS, Magalhães L de C, Siqueira CM, Alves CRL. Cross-cultural adaptation of the child development surveillance instrument "Survey of Wellbeing of Young Children (SWYC)" in the Brazilian context. J Hum Growth Dev. 2019;29(1):28-38. doi: 10.7322/jhgd.145001.

58. Velikonja T, Edbrooke-Childs J, Calderon A, Sleed M, Brown A, Deighton J. The psychometric properties of the Ages & Stages Questionnaires for ages 2-2.5: a systematic review. Child Care Health Dev. 2017;43(1):1-17. doi: 10.1111/cch.12397.

59. Marks KP, Madsen Sjö N, Wilson P. Comparative use of the Ages and Stages Questionnaires in the USA and Scandinavia: a systematic review. Dev Med Child Neurol. 2019;61(4):419-430. doi: 10.1111/dmcn.14044.

60. Small JW, Hix-Small H, Vargas-Baron E, Marks KP. Comparative use of the Ages and Stages Questionnaires in low- and middle-income countries. Dev Med Child Neurol. 2019 Apr;61(4):431-443. doi: 10.1111/dmcn.13938.

61. Flamant C, Branger B, Nguyen The Tich S, de la Rochebrochard E, Savagner C, Berlie I, Rozé JC. Parent-completed developmental screening in premature children: a valid tool for follow-up programs. PLoS One. 2011;6(5):e20004. doi: 10.1371/journal.pone.0020004.

62. Lindsay NM, Healy GN, Colditz PB, Lingwood BE. Use of the Ages and Stages Questionnaire to predict outcome after hypoxic-ischaemic encephalopathy in the

neonate. J Paediatr Child Health. 2008 Oct;44(10):590-5. doi: 10.1111/j.1440-1754.2008.01388.x.

63. Halbwachs M, Muller JB, Nguyen The Tich S, de La Rochebrochard E, Gascoin G, Branger B, Rouger V, Rozé JC, Flamant C. Usefulness of parent-completed ASQ for neurodevelopmental screening of preterm children at five years of age. PLoS One. 2013;8(8):e71925. doi: 10.1371/journal.pone.0071925.

64. Halbwachs M, Muller JB, Nguyen The Tich S, Gascoin G, Chauty-Frondas A, Branger B, Rouger V, Roze JC, Flamant C. Predictive value of the parent-completed ASQ for school difficulties in preterm-born children <35 weeks' GA at five years of age. Neonatology. 2014;106(4):311-6. doi: 10.1159/000363216.

65. Klamer A, Lando A, Pinborg A, Greisen G. Ages and Stages Questionnaire used to measure cognitive deficit in children born extremely preterm. Acta Paediatr. 2005;94(9):1327-9. doi: 10.1111/j.1651-2227.2005.tb02095.x.

66. Kvestad I, Taneja S, Kumar T, Bhandari N, Strand TA, Hysing M; Study Group. The assessment of developmental status using the Ages and Stages questionnaire-3 in nutritional research in north Indian young children. Nutr J. 2013;12:50. doi: 10.1186/1475-2891-12-50.

67. Shrestha M, Strand TA, Ulak M, Chandyo RK, Ranjitkar S, Hysing M, Shrestha L, Kvestad I. The feasibility of the Ages and Stages Questionnaire for the assessment of child development in a community setting in Nepal. Child Care Health Dev. 2019;45(3):394-402. doi: 10.1111/cch.12654.

68. Rubio-Codina M, Araujo MC, Attanasio O, Muñoz P, Grantham-McGregor S. Concurrent Validity and Feasibility of Short Tests Currently Used to Measure Early Childhood Development in Large Scale Studies. PLoS One. 2016;11(8):e0160962. doi: 10.1371/journal.pone.0160962.

69. Filgueiras A, Pires P, Maissonette S, Landeira-Fernandez J. Psychometric properties of the Brazilian-adapted version of the Ages and Stages Questionnaire in public child daycare centers. Early Hum Dev. 2013;89(8):561-76. doi: 10.1016/j.earlhumdev.2013.02.005.

70. Yovanoff P, Squires J, McManus S. Adaptation From Paper–Pencil to Web-Based Administration of a Parent-Completed Developmental Questionnaire for Young Children. Infants Young Child. 2013;26(4):318–32. DOI:10.1097/IYC.0b013e31829f95f3

71. Fernald LC, Kariger P, Hidrobo M, Gertler PJ. Socioeconomic gradients in child development in very young children: evidence from India, Indonesia, Peru, and Senegal. Proc Natl Acad Sci U S A. 2012;109 (Suppl 2):17273-80. doi: 10.1073/pnas.1121241109.

72. Kapci EG, Kucuker S, Uslu RI. How Applicable Are Ages and Stages Questionnaires for Use With Turkish Children? Top Early Child Spec Educ. 2010;30(3):176–88. doi:10.1177/0271121410373149

73. Hornman J, Kerstjens JM, de Winter AF, Bos AF, Reijneveld SA. Validity and internal consistency of the Ages and Stages Questionnaire 60-month version and the effect of three scoring methods. Early Hum Dev. 2013;89(12):1011-5. doi: 10.1016/j.earlhumdev.2013.08.016.

74. Sices L, Stancin T, Kirchner L, Bauchner H. PEDS and ASQ developmental screening tests may not identify the same children. Pediatrics. 2009;124(4):e640-7. doi: 10.1542/peds.2008-2628.

75. Limbos MM, Joyce DP. Comparison of the ASQ and PEDS in screening for developmental delay in children presenting for primary care. J Dev Behav Pediatr. 2011;32(7):499-511. doi: 10.1097/DBP.0b013e31822552e9.

76. Çelen Yoldaş T, Karakaya J, Özdemir G, Engin Erdal A, Özmert EN. Comparison of the Parents' Evaluation of Developmental Status and Ages and Stages Questionnaire Developmental Screening Tests in a Eurasian Country. J Dev Behav Pediatr. 2021;42(6):450-456. doi: 10.1097/DBP.0000000000000912.

77. Sheldrick RC, Marakovitz S, Garfinkel D, Carter AS, Perrin EC. Comparative Accuracy of Developmental Screening Questionnaires. JAMA Pediatr. 2020;174(4):366-374. doi: 10.1001/jamapediatrics.2019.6000.

78. Peyre H, Charkaluk ML, Forhan A, Heude B, Ramus F; EDEN Mother–Child Cohort Study Group. Do developmental milestones at 4, 8, 12 and 24 months predict IQ at 5-6 years old? Results of the EDEN mother-child cohort. Eur J Paediatr Neurol. 2017;21(2):272-279. doi: 10.1016/j.ejpn.2016.11.001.

79. Charkaluk ML, Rousseau J, Calderon J, Bernard JY, Forhan A, Heude B, Kaminski M; EDEN Mother–Child Cohort Study Group. Ages and Stages Questionnaire at 3 Years for Predicting IQ at 5-6 Years. Pediatrics. 2017;139(4):e20162798. doi: 10.1542/peds.2016-2798.

80. Sim F, Thompson L, Marryat L, Ramparsad N, Wilson P. Predictive validity of preschool screening tools for language and behavioural difficulties: A PRISMA systematic review. PLoS One. 2019;14(2):e0211409. doi: 10.1371/journal.pone.0211409.

81. Wong HS, Santhakumaran S, Cowan FM, Modi N; Medicines for Neonates Investigator Group. Developmental Assessments in Preterm Children: A Meta-analysis. Pediatrics. 2016;138(2):e20160251. doi: 10.1542/peds.2016-0251.

82. Luttikhuizen dos Santos ES, de Kieviet JF, Königs M, van Elburg RM, Oosterlaan J. Predictive value of the Bayley scales of infant development on development of very preterm/very low birth weight children: a meta-analysis. Early Hum Dev. 2013;89(7):487-96. doi: 10.1016/j.earlhumdev.2013.03.008.

83. Álvarez Gómez MJ, Aznar JS, Sánchez-ventura JG. Importancia de la vigilancia del desarrollo psicomotor por el pediatra de Atención Primaria: revisión del tema y experiencia de seguimiento en una consulta en Navarra. Rev Pediatr Aten Primaria 2009;11(41):65–87.

84. Schonwald A, Huntington N, Chan E, Risko W, Bridgemohan C. Routine developmental screening implemented in urban primary care settings: more evidence of feasibility and effectiveness. Pediatrics. 2009;123(2):660-8. doi: 10.1542/peds.2007-2798.

85. Kerstjens JM, Bos AF, ten Vergert EM, de Meer G, Butcher PR, Reijneveld SA. Support for the global feasibility of the Ages and Stages Questionnaire as developmental screener. Early Hum Dev. 2009;85(7):443-7. doi: 10.1016/j.earlhumdev.2009.03.001.

86. Ramírez V, Rosas R. Estandarización del WISC-III en Chile : Descripción del Test , Estructura Factorial y Consistencia Interna de las Escalas. Psykhe. 2007;16(1):91–109. doi: 10.4067/S0718-22282007000100008

87. Rubio-Codina M, Grantham-McGregor S. Predictive validity in middle childhood of short tests of early childhood development used in large scale studies compared to the Bayley-III, the Family Care Indicators, height-for-age, and stunting: A longitudinal study in Bogota, Colombia. PLoS One. 2020;15(4):e0231317. doi: 10.1371/journal.pone.0231317.

88. Martínez-Nadal S, Schonhaut L, Armijo I, Demestre X. Predictive value of the Ages and Stages Questionnaire® for school performance and school intervention in late preterm- and term-born children. Child Care Health Dev. 2021;47(1):103-111. doi: 10.1111/cch.12814.

89. Jin F, Schjølberg S, Wang MV, Eadie P, Nes RB, Røysamb E, Tambs K. Predicting Literacy Skills at 8 Years From Preschool Language Trajectories: A Population-Based Cohort Study. J Speech Lang Hear Res. 2020;63(8):2752-2762. doi: 10.1044/2020_JSLHR-19-00286.

90. Schonhaut L, Martinez-Nadal S, Armijo I, Demestre X. Reliability and agreement of ages and stages questionnaires®: Results in late preterm and term-born infants at 24 and 48 months. Early Hum Dev. 2019;128:55-61. doi: 10.1016/j.earlhumdev.2018.11.008.

91. Schonhaut L, Pérez M, Armijo I, Maturana A. Comparison between Ages & Stages Questionnaire and Bayley Scales, to predict cognitive delay in school age. Early Hum Dev. 2020;141:104933. doi: 10.1016/j.earlhumdev.2019.104933.

92. Cioni G, Inguaggiato E, Sgandurra G. Early intervention in neurodevelopmental disorders: underlying neural mechanisms. Dev Med Child Neurol. 2016;58 Suppl 4:61-6. doi: 10.1111/dmcn.13050.

93. Schonhaut L,Maturana A,Cepeda O, Serón P. Predictive Validity of Developmental Screening Questionnaires for Identifying Children With Later Cognitive or Educational Difficulties: A Systematic Review. Front.Pediatr 2021;9:698549. doi:10.3389/fped.2021.698549P

94. Lopes S, Graça P, Teixeira S, Serrano AM, Squires J. Psychometric properties and validation of Portuguese version of Ages & Stages Questionnaires (3rd edition): 9, 18 and 30 Questionnaires. Early Hum Dev. 2015;91(9):527-33. doi: 10.1016/j.earlhumdev.2015.06.006.

95. Alvik A, Grøholt B. Examination of the cut-off scores determined by the Ages and Stages Questionnaire in a population-based sample of 6 month-old Norwegian infants. BMC Pediatr. 2011;11:117. doi: 10.1186/1471-2431-11-117.

96. Valla L, Birkeland MS, Hofoss D, Slinning K. Developmental pathways in infants from 4 to 24 months. Child Care Health Dev. 2017;43(4):546-555. doi: 10.1111/cch.12467.

97. Shevell M. Global developmental delay and mental retardation or intellectual disability: conceptualization, evaluation, and etiology. Pediatr Clin North Am. 2008;55(5):1071-84, xi. doi: 10.1016/j.pcl.2008.07.010.

98.   Doyle LW, Davis PG, Schmidt B, Anderson PJ. Cognitive outcome at 24 months is more predictive than at 18 months for IQ at 8-9 years in extremely low birth weight children. Early Hum Dev. 2012;88(2):95-8. doi: 10.1016/j.earlhumdev.2011.07.013.

99.   Cairney DG, Kazmi A, Delahunty L, Marryat L, Wood R. The predictive value of universal preschool developmental assessment in identifying children with later educational difficulties: A systematic review. PLoS One. 2021;16(3):e0247299. doi: 10.1371/journal.pone.0247299.

100.   Aylward GP. Developmental screening and assessment: what are we thinking? J Dev Behav Pediatr. 2009;30(2):169-73. doi: 10.1097/DBP.0b013e31819f1c3e.

101.   Marks K, Glascoe FP, Aylward GP, Shevell MI, Lipkin PH, Squires JK. The thorny nature of predictive validity studies on screening tests for developmental-behavioral problems. Pediatrics. 2008;122(4):866-8.

102.   Hix-Small H, Marks K, Squires J, Nickel R. Impact of implementing developmental screening at 12 and 24 months in a pediatric practice. Pediatrics. 2007;120(2):381-9. doi: 10.1542/peds.2006-3583.

103.   Sheldrick RC, Garfinkel D. Is a Positive Developmental-Behavioral Screening Score Sufficient to Justify Referral? A Review of Evidence and Theory. Acad Pediatr. 2017;17(5):464-470. doi: 10.1016/j.acap.2017.01.016.

104.   Lamsal R, Dutton DJ, Zwicker JD. Using the ages and stages questionnaire in the general population as a measure for identifying children not at risk of a neurodevelopmental disorder. BMC Pediatr. 2018;18(1):122. doi: 10.1186/s12887-018-1105-z.

105.   Sheldrick RC, Breuer DJ, Hassan R, Chan K, Polk DE, Benneyan J. A system dynamics model of clinical decision thresholds for the detection of developmental-behavioral disorders. Implement Sci. 2016;11(1):156. doi: 10.1186/s13012-016-0517-0.

106.   Wallis KE, Davis Rivera LB, Guthrie W, Bennett AE, Mandell DS, Miller JS. Provider Responses to Positive Developmental Screening: Disparities in Referral Practices? J Dev Behav Pediatr. 2021;42(1):23-31. doi: 10.1097/DBP.0000000000000855.

107.   Bossuyt PM, Leeflang MM. Chapter 6: Developing Criteria for Including Studies. In: Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 0.4 [updated September 2008]. The Cochrane Collaboration, 2008.