

Predicción de Resultados Cognitivos a los 2 Años en Infantes Muy Prematuros Usando Métodos de Aprendizaje Automático

Andrea K. Bowe; Gordon Lightbody; Anthony Staines; Deirdre M. Murray; Mikael Norman

JAMA Network Open. 2023;6(12):e2349111. doi:10.1001/jamanetworkopen.2023.49111

Resumen

IMPORTANCIA La intervención temprana puede mejorar los resultados cognitivos de infantes muy prematuros (MPT) pero es intensa en recursos. Identificar a aquellos que más necesitan intervención temprana es importante

OBJETIVO Evaluar un modelo para empleo en infantes MPT para predecir el rezago cognitivo a los 2 años de edad empleando los datos clínicos y sociodemográficos rutinariamente disponibles.

DISEÑO, ÁMBITO, Y PARTICIPANTES Este estudio pronóstico estuvo basado en el Registro de Calidad Neonatal Sueco. La cobertura nacional de datos neonatales se logró en 2011, y el registro de datos de Seguimiento se inició el 1º de Enero de 2015, finalizando la inclusión el 31 de Septiembre, 2022. Una variedad de modelos de aprendizaje fue probada (luego de entrenamiento) para predecir rezago cognitivo. Los infantes sobrevivientes de las unidades neonatales en Suecia con EG < 32 semanas y los datos completos de los puntajes de la escala cognitiva Bayley III a los dos años de edad corregida, fueron evaluados. Los infantes con anomalías congénitas mayores fueron excluidos.

EXPOSICIONES Se examinaron para su predictibilidad un total de 90 variables (conteniendo información sociodemográfica y clínica sobre las condiciones, investigaciones y tratamientos iniciados durante la gestación, parto, y la admisión en UCIN).

PRINCIPALES MEDIDAS Y RESULTADOS El principal resultado fue la función cognitiva a los 2 años, categorizada como tamizaje positivo para rezago cognitivo (puntaje de índice cognitivo <90) o exhibiendo desarrollo cognitivo típico (puntaje ≥90).

RESULTADOS Un total de 1062 niños/as (PN mediana (RIQ) 880 g [720- 1100]; 566 [53.3%] varones) fueron incluidos en el proceso de modelaje, de los cuales 231 (21.8%) presentaban rezago cognitivo. Un modelo de regresión logística conteniendo 26 características predictivas logró un área bajo la curva operativa del receptor de 0.77 (IC 95% 0.71-0.83). Las 5 características más importantes para rezago cognitivo fueron lenguaje familiar no-escandinavo, duración prolongada de la internación, bajo peso de nacimiento, egreso a otro destino diferente del hogar, y no recibir leche materna al alta. Al egreso de UCIN, el modelo completo pudo identificar correctamente 605 de 650 infantes que podrían presentar rezago cognitivo a los 24 meses (sensibilidad 0.93) y 1081 de 2350 que no lo presentarían (especificidad 0,46).

CONCLUSIONES Y RELEVANCIA Los hallazgos de este estudio sugieren que el modelo predictivo en el cuidado neonatal podría facilitar la intervención temprana y dirigida para todos los infantes prematuros en mayor riesgo de desarrollar déficit cognitivo.

Puntos Clave

Pregunta ¿Pueden los datos neonatales fácilmente disponibles identificar infantes MPT que exhibirán rezago cognitivo más tarde en la vida?

Hallazgos En este estudio pronóstico de resultados cognitivos a los 2 años de seguimiento entre 1062 infantes nacidos MPT, un modelo de regresión logística conteniendo 26 características neonatales identificó 93% de infantes MPT que fueron positivos en el tamizaje para rezago cognitivo a los 2 años de seguimiento, con una especificidad de 46%.

Significado El empleo de este modelo podría apuntar a aquéllos infantes MPT en más alto riesgo de rezago cognitivo para recibir intervención temprana y efectiva.

Introducción

Cada año, 2.2 millones de infantes nacen MPT, esto es a una EG menor de 32 semanas (1). Junto a la mayor sobrevivencia, hay más niños/as, más niños/as están en riesgo para déficits a largo plazo asociados con prematuridad, como ser dificultades cognitivas (2-5). El CI medio, a través de la niñez, para infantes MPT ha sido reportado casi 1 DS más bajo que sus pares de término, equivalente a la diferencia media de aproximadamente 13 puntos de CI (6). Este hallazgo tiene importantes implicancias para el logro académico, estado socioeconómico, y salud física y mental (7-12). Desde 1990, la tasa de mortalidad infantil de PT en países de altos ingresos ha disminuido a razón de 4% por año, pero no se han logrado mejoras significativas en resultados cognitivos (2, 6, 13).

Las vías causales subyacentes a la asociación entre nacimiento MPT y déficit cognitivo permanecen poco claras, haciendo difícil implementar estrategias primarias de prevención (6). La prevención secundaria, en la forma de intervención temprana, ha sido hasta ahora la única manera efectiva de avanzar (14). Sin embargo, muchas de las intervenciones estudiadas fueron altamente intensas en cuanto a recursos- iniciando en el hospital, seguido de frecuentes visitas domiciliarias a las familias, por períodos de hasta 3 años (14). La identificación temprana y precisa de los infantes que necesitan estas intervenciones se ha vuelto cada vez más importante.

Los factores biológicos y ambientales, así como los eventos clínicos del período perinatal, contienen información pronóstica acerca de la función cognitiva en infantes PT menores de 5 años (15). Varios modelos pronósticos han sido reportados, pero pocos han podido dar cuenta de completa de las relaciones e interacciones no-lineales entre predictores o han validado sus modelos sobre datos no vistos (16). El objetivo de este estudio fue desarrollar un algoritmo explicable para uso en infantes MPT para predecir rezago cognitivo (RC) a los 2 años de edad. Con este propósito, describimos las características perinatales de aquellos que presentaron y los que no presentaron RC a los 2 años. Un enfoque basado en datos para identificar predictores relevantes del resultado cognitivo fue usado, y una variedad de modelos de aprendizaje automático fueron armados y subsiguientemente testeados en un paquete de prueba no visto. Finalmente, las características predictivas más importantes fueron identificadas.

Métodos

La autoridad de Revisión Ética Sueca aprobó este estudio pronóstico el 28 de Marzo, 2022. Esta autoridad también liberó de consentimiento informado de padres o cuidadores, que fueron informados que los datos perinatales y de seguimiento fueron registrados, con una posibilidad de retirarse en cualquier momento (menos de 5 familias en el período de estudio). El Modelo de Reporte Transparente de Predicción Multivariable para el Pronóstico Individual o Diagnóstico (TRIPOD) de guías para reportes fue seguido en el informe del estudio (17).

Datos

Los datos fueron del Registro de Calidad Neonatal Sueco (SNQ). El SNQ es un registro de base poblacional de toda la nación que captura a todos los infantes ingresados a cuidado neonatal. El registro provee información clínica detallada sobre condiciones, investigaciones, y tratamientos iniciados durante la gestación, parto, e ingreso a UCIN. La cobertura a nivel nacional de datos neonatales fue alcanzada en 2011, y el registro de datos de seguimiento se inició el 1º de Enero, 2015, finalizando la inclusión el 31 de Septiembre, 2022. Los detalles de reporte de datos, completitud y validez del registro han sido publicados previamente (18).

Como parte del Programa de Seguimiento nacional lanzado en 2015, los sobrevivientes de alto riesgo del cuidado intensivo neonatal recibieron la recomendación de tener evaluación del ND, incluyendo Bayley-III, programada a la edad corregida de 2 años. El objetivo primario del seguimiento fue tamizar para diferentes déficits, particularmente en quienes tuvieron EG28 semanas (19). Los resultados de estas evaluaciones debían ser reportados al SNQ.

Población en estudio

Fueron elegibles para inclusión en este estudio niños/as MPT de 2 años con datos del puntaje cognitivo en Bayley-III. Los niños/as con evaluación cognitiva, pero con anomalía congénita mayor según definición del Registro Europeo de Anomalías Congénitas y Gemelares (20) (n=142) fueron excluidos. Hubo 59 niños/as MPT que concurren a la evaluación Bayley-III, pero no pudieron completar la prueba (Tabla 1).

Resultado

El resultado fue la capacidad cognitiva a los 2 años de seguimiento, medida con Bayley-III. Bayley-III no ha sido estandarizado para la población sueca. Sin embargo, en controles suecos sanos nacidos a término (n=366) y evaluados a los 30 meses de edad, el puntaje medio del índice cognitivo fue 104 (10.6) (21). En base a ese estudio, el programa de seguimiento SNQ aplica un puntaje límite de menos de 90 (aproximadamente un corte de -1 DS) para identificar niños/as que puedan beneficiarse de ser más evaluados o intervención (19). En este estudio, los niños/as con un puntaje de índice cognitivo menor de 90 fueron categorizados como portadores de RC y aquellos con un puntaje de 90 o más como desarrollo cognitivo típico (DCT).

Análisis estadístico

Estadísticas descriptivas

Las características de los participantes fueron comparadas usando pruebas χ^2 de Pearson y exacta de Fisher (datos categóricos), *t* Welch de dos muestras (para datos continuos normalmente

distribuidos), y Wilcoxon rank sum (datos con distribución anormal). Las pruebas de hipótesis fueron de dos colas, y la significancia a priori fue establecida en $P < .05$.

Preparación de datos

Los datos se manejaron con un paquete estadístico R, versión 4.1.1 (R Project for Statistical Computing) (22). Un grupo de datos de estudio conteniendo 97 características potencialmente predictivas, elegidas en base a los contenidos de SNQ, literatura previa, hipótesis plausibles, y aportes del autor senior, fue construida. (eTabla 1 en el Suplemento 1), dejando 90 características categorizadas como gestación; nacimiento; resucitación neonatal; enfermedades neonatales respiratoria, neurológica, infección y otras; y factores sociodemográficos. Los valores restantes perdidos (e Tabla 2 en Suplemento 1) fueron imputados usando el paquete valores-perdidos, un método randomizado de imputación forestal de datos (23, 24).

Un enfoque “todo es relevante”, tratando de identificar todas las características relevantes para la clasificación, fue el utilizado, y se empleó el algoritmo Boruta (25, 26). Para cada característica, se creó una característica “sombra” para “arrastrar” randomizadamente los valores de la característica original, destruyendo toda posible asociación con el resultado. Un clasificador forestal random fue llevado sobre las características originales y las sombras. El valor de importancia de cada característica original fue comparado con un umbral- definido como el máximo valor de importancia registrado entre las variables sombra. Si la importancia de una característica era superior a ese umbral, era registrado en un vector de impacto. Este proceso fue reiterativo. Las características que significativamente sobresalían al umbral eran confirmadas como importantes, pero aquellos que significativamente se comportaron por debajo del umbral fueron rechazados y retirados para las siguientes iteraciones. El algoritmo se detuvo cuando se llegó al número de 600 iteraciones y las restantes características fueron consideradas como tentativas (25, 26)

Características correlativas

Los coeficientes de correlación Pearson fueron calculados y ploteados en una forma pareada para todas las características identificadas como importantes o tentativas. Las características con coeficientes superiores a 0.70 fueron examinadas para ver redundancia. Para elegir qué características retener, se consideraron la robustez, dimensionalidad, correlación con el resultado, opinión experta, y efecto sobre la predicción.

Modelaje

El paquete de datos fue estratificado por el resultado y dividido randomizadamente en un set conteniendo 70% de los datos y un set de pruebas conteniendo 30%. Las clases de resultados en el paquete de datos fueron desbalanceados, que arriesgaron producir un modelo predictivo con sesgo hacia la clase mayoritario. Para examinar esto, la técnica de sobremuestreo sintético de la minoría (SMOTE) fue aplicado al set de datos en curso (27, 28).

Usar todas las características relevantes identificadas por Boruta, una foresta randomizada, regresión logística, máquina vector de soporte, y algoritmo de máquina de gradiente de impulso fueron encarrilados y se seleccionaron hiperparámetros óptimos en una grilla de validación cruzada usando el paquete de intercalación (modelos A-D) (29). Estos 4 algoritmos fueron elegidos

porque han sido ampliamente utilizados en la investigación del cuidado de la Salud, son más factibles de ser reconocidos por los clínicos que los usarían, son eficientes, y tienen potencial de explicabilidad (30, 31).

Tabla 1. Características evaluadas como importantes para la tentativa de futura función cognitiva en niños muy prematuros (EG<32 SDG)

Característica	N o Valido	Total (N=1062)	Desarrollo cog-nitivo típico(n = 831)	Retraso Cognitivo (n = 231)	Valor P
Embarazo					
Diabetes gest No (%)	1062	14 (1.3)	8 (1.0)	6 (2.6)	.09 ^a
Nacimiento					
EG promedio (DS) S _{semana}	1062	26.5 (2.2)	26.7 (2.2)	26.0 (2.2)	<.001 ^b
Peso nacer mediana (RIQ) g	1055	880 (720 to 1100)	900 (760 to 1125)	800 (660 to 980)	<.001 ^c
Pz peso al nacer Mediana (RIQ)	1056	-1.1 (-2.3 to -0.3)	-1.0 (-2.2 to -0.2)	-1.2 (-2.5 to -0.4)	.16 ^c
Perimetro cefálico promedio (DS), cm	883	24.7 (2.5)	24.9 (2.5)	24.2 (2.5)	<.001 ^b
Genero No. (%)					
Masculino	1062	566 (53.3)	414 (49.8)	152 (65.8)	<.001 ^e
Femenino	1062	496 (46.7)	417 (50.2)	79 (34.2)	
Apgar mediana (RIQ)					
5 minutos	1036	8 (6 to 9)	8 (6 to 9)	7 (5 to 9)	.001 ^c
10 minutos	1033	9 (8 to 10)	9 (8 to 10)	9 (7 to 10)	.01 ^c
Resucitación Neonatal					
Intubación en ATQ No. (%)	1053	371 (35.2)	265 (32.1)	106 (46.7)	<.001 ^e
Respiratorio					
Displasia broncopulm. No. (%) ^c	1062	637 (60.0)	475 (57.2)	162 (70.1)	<.001 ^e
Oxígeno extra mediana (RIQ) ^d	1062	49 (15 to 80)	45 (12 to 74)	67 (34.5 to 98.5)	<.001 ^c
Duración de CPAP Mediana (RIQ), d	1038	26 (9 to 39)	25 (8 to 38)	30 (14 to 44)	<.001 ^c
Duración de ventilación Mecánica mediana (RIQ), d	1062	4 (0 to 17)	3 (0 to 14)	12 (0 to 26)	<.001 ^c
Duración ventilación Mecánica convencional Mediana (RIQ), D	1062	1 (0 to 6)	1 (0 to 5)	3 (0 to 9)	<.001 ^c
Duración de VAFO Mediana (RIQ),d	1062	0 (0 to 8)	0 (0 to 5)	3 (0 to 15)	<.001 ^c
Esteroides postnatales (sistémicos) No. %	1062	341 (32.1)	238 (28.6)	103 (44.6)	<.001 ^e
Infección					
Duración de antibióticos Mediana (RIQ) ^d	1062	14 (7 to 27)	12 (6 to 25)	20 (8 to 33)	<.001 ^c
Neurología					
HIV grado, No. (%)					
No confirmado HIV					
1	1045	745 (71.3)	602 (73.5)	143 (63.3)	
2	1045	129 (12.3)	100 (12.2)	29 (12.8)	
3	1045	97 (9.3)	73 (8.9)	24 (10.6)	<.001 ^e
4	1045	36 (3.4)	24 (2.9)	12 (5.3)	
No HIV, hemorragia Intracraneal No. (%)	1045	38 (3.6)	20 (2.4)	18 (8.0)	
LMPVc No. (%)	1062	12 (1.1)	6 (0.7)	6 (2.6)	.03 ^a
Otros tratamientos					
Catéter arterial umbilical No. (%)	1062	17 (1.6)	8 (1.0)	9 (3.9)	.004 ^d
Tratamiento de insulina para hiperglicemia No (%)	1062	861 (81.1)	679 (81.7)	182 (78.8)	.36 ^d
Tratamiento ROP Duración de administración Inotrópicos Mediana (RIQ), D	1062	131 (12.3)	88 (10.6)	43 (18.6)	.002 ^d
No. Transfusión plasma mediana (RIQ)	1000	124 (12.4)	84 (10.8)	40 (17.9)	.007 ^d
	1062	0 (0 to 0)	0 (0 to 0)	0 (0 to 1.5)	<.001 ^c
	1050	2 (1 to 12)	1 (1 to 12)	2 (1 to 20)	<.001 ^c

Continúa

Continuación tabla 1

Tabla 1. Características evaluadas como importantes para la tentativa de futura función cognitiva en niños muy prematuros (EG<32 SDG) (continuación)

Característica	No Valido	Total (N=1062)	Desarrollo cognitivo típico (n = 831)	Retraso Cognitivo (n = 231)	Valor P
Estancia hospital y egreso					
Duración de cuidados en hospital y casa, mediana (RIQ), d	1062	75 (56 to 96)	73 (54 to 94)	85 (65 to 108)	<.001 ^c
Duración de hospitalización mediana (RIQ), d	1062	75 (56 to 96)	72 (53 to 93)	83 (65.5 to 108)	<.001 ^e
Egreso al hogar, No. (%)		927 (87.3)	746 (89.8)	181 (78.4)	<.001 ^d
Recibiendo leche materna al egreso, No. (%)					
Nada	913	320 (35.0)	235 (32.2)	85 (46.2)	
Parcialmente	913	355 (38.9)	284 (39.0)	71 (38.6)	<.001 ^d
Completamente	913	238 (26.1)	210 (28.8)	28 (15.2)	
Características sociodemográficas					
Sueco u otro lenguaje escandinavo como lenguaje familiar, No. (%)	914	641 (70.1)	541 (74.8)	100 (52.4)	<.001 ^d
Educación de padres, No. (%)					
0-3 y	859	8 (0.9)	5 (0.7)	3 (1.6)	
3-6 y	859	3 (0.3)	2 (0.3)	1 (0.5)	
7-9 y	859	50 (5.8)	31 (4.6)	19 (10.4)	<.001 ^a
10-12 y	859	322 (37.5)	243 (35.9)	79 (43.2)	
>12 y	859	476 (55.4)	395 (58.4)	81 (44.3)	
BSID-III^f					
Edad corregida al test Promedio (DS), A.	1024	2.2 (0.2)	2.2 (0.2)	2.2 (0.2)	.59 ^b
Índice puntaje cognitivo Promedio (DS)	1062	96.8 (16.1)	103.0 (10.5)	73.5 (10.0)	<.001 ^b

Abreviaturas: BSID- Escala de Bayley del desarrollo del niño y del infante; CPAP, presión continua de las vías aéreas; LMPVc, leucomalacia periventricular cística; EG, edad gestacional; VAFO, ventilación oscilatoria de alta frecuencia; HIV, hemorragia intraventricular; ROP, retinopatía del prematuro. DS, desviación estándar. Y, año. RIQ, rango intercuartílico.

a Prueba exacta de Fisher

b Prueba de t de Welch para dos muestras

c Prueba de Wilcoxon

d Prueba X2 Pearson

e Clasificación de enfermedades y problemas de salud relacionados con el código (p27.1) registrado como suplemento de oxígeno a las 36 SDG de edad postmenstrual.

f Razones para no completar la prueba de Bayley (59) fueron: el niño declino o no pudo participar (n=23); inatención, hiperactividad o fatiga (6); barrera del lenguaje (1); Padres declinaron (4); razones de recursos administrativos (6); Niño no disponible (13); razones no claras o no recordadas (6)

Evaluación

La validación interna fue realizada examinando la exactitud en una validación cruzada de 10 cruces. Este proceso incluyó dividir los datos alineados en 10 partes conteniendo igual número de observaciones. Cada iteración del modelo fue alineada usando el 90% (9 partes) de los datos y validados en el remanente 10%. Este procedimiento fue repetido 10 veces usando un diferente

10% para validación cada vez y los resultados fueron promediados en las etapas. La decisión de umbral se estableció en una probabilidad de 0.5.

Los modelos finales con hiperparámetros sintonizados fueron luego validados externamente probando su desempeño en el paquete de datos de prueba no vista. La exactitud, la exactitud balanceada, sensibilidad, y especificidad cuando se compararon. El área ROC (AUROC) fue usado para examinar el desempeño a través de todos los umbrales de decisión posibles (32). El proceso de modelaje está resumido en la Figura 1.

Explicabilidad

La explicabilidad de los modelos con mejor desempeño en la validación externa fue explorada examinando la importancia de la característica y su efecto. Se crearon lotes de importancia de la característica usando el grupo de prueba de datos y el método de permutación en el paquete VIP (33). El desempeño basal del modelo fue medido usando AUROC. La característica de interés fue luego barajado al azar, y el desempeño del modelo fue medido nuevamente. La diferencia entre 2 medidas fue empleada como una medida de la importancia de la característica. Para cada característica, se simuló 10 veces el barajado y la importancia fue promediado en las simulaciones. El efecto de las características y los métodos más detallados son descriptos en eMétodos en el Suplemento 1.

Resultados

Características de la población en estudio

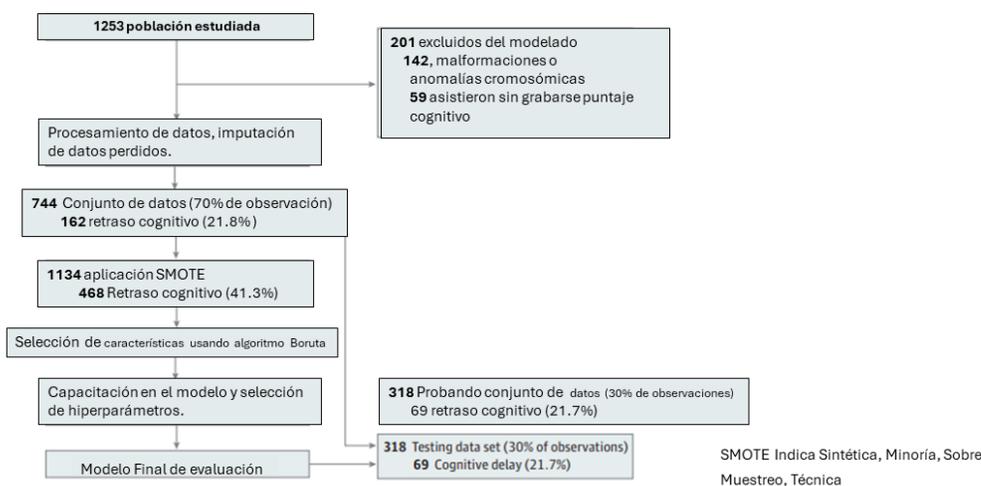
Las características perinatales, neonatales y familiares de los 1062 niños/as (mediana PN [RIQ] 880 [720-1100] g; 566 [53.3%] varones y 496 [46.7%] mujeres) incluídas en el proceso de modelaje, se muestran en la Tabla 1. Una descripción de la cohorte a través de las 90 características incluídas en el proceso de selección está contenida en la eTabla 3 en el suplemento 1. Un total de 231 niños/as (21.8%) resultaron positivos para RC en el tamizaje. La proporción de varones fue más alta en el grupo RC que en el de DCT (152 [65.8%] vs 414 [49.8%]; $P < .001$). Los infantes en el grupo de RC tenía media (DS) de EG más baja (26.0 [2.2] vs 26.7 [2.2] semanas; $P < .001$) y menor mediana [RIQ] de PN (800 [660-980] vs 900 [760- 1125] g; $P < .001$) que el grupo DCT. Una mayor proporción de niños/as con RC fueron intubados durante la resucitación neonatal (106 [46.7%] vs 265 [32.1%]; $P < .001$). También hubo diferencias significativas entre los grupos en términos de complicaciones respiratorias y requerimientos de oxígeno.

La proporción de DBP fue significativamente más elevada en el grupo con RC que en el DCT (162 [70.1%] vs 475 [57.2%]; $P < .001$), así como también la proporción que requirió esteroides postnatales (103 [44.6%] vs 238 [28.6%]; $P < .001$). La mediana de duración de CPAP y ventilación convencional o de alta frecuencia fue mayor en el grupo RC.

En general, la incidencia de hemorragia intraventricular (HIV) de cualquier grado fue superior (83 [36.7%] vs 217 [26.5%]), los grados más altos fueron más frecuentemente observados en el grupo RC que en el DCT. La incidencia de hemorragia intracraneana no-HIV fue 11% (12 de 1062), y la incidencia de leucomalacia quística periventricular fue 1.6% (17 de 1062), con una incidencia más elevada en el grupo RC.

Los infantes en el grupo RC tuvieron media (DS) de días de estadía hospitalaria mayor (83 [42.5] vs 72 [40.0] días; $P < .001$), y una menor proporción de estos infantes recibieron algún volumen de leche humana al egreso (99 [53.8%] vs 494 [67.8%]). Menor proporción de padres del grupo RC reportaron más de 12 años de educación (81[44.3%] vs 395 [58.4%]; $P < .001$) y lengua escandinava como lengua familiar (100 [52.4%] vs 541 [74.8%]; $P < .001$).

Figura 1. Descripción general del proceso del modelado



Paquetes de datos encadenados y de prueba

El paquete de datos contenía 744 infantes, de los cuales 162 (21.8%) fueron positivos para RC en el tamizaje. Después de aplicar SMOTE, el paquete encadenado consistió de 1134 participantes, de los cuales 468 (41.3%) presentó RC. El paquete de prueba estuvo compuesto de 318 infantes, de los cuales 69 (21.7%) presentó RC. SMOTE no fue aplicado al paquete de prueba.

Selección de Característica

El algoritmo confirmó 27 de 90 características como importantes, con otras 4 marcadas como tentativas (eFigura 1 en el Suplemento 1). Entre estas 31 características, había 12 con coeficientes de correlación mayores que 0.70 (eFigura en Suplemento 1). Cinco características fueron reducidas para reducir la colinearidad (Apgar a los 5 minutos, duración total de cualquier tipo de ventilación mecánica, duración de oxígeno suplementario, duración de cuidado en hospital y domicilio combinados, y puntaje z para peso de nacimiento), dejando 26 características detalladas en eTabla 4 en Suplemento 1. La duración de la hospitalización y de CPAP. Así como PN, EG, y circunferencia craneana, estuvieron altamente correlacionadas, pero el desempeño predictivo fue mejor con todas las características incluidas. El puntaje z de PN fue removido porque la característica derivaba de EG y PN juntos y el comportamiento predictivo fue mejor cuando se incluyeron las características fuente.

Armado del Modelo y Evaluación

Los resultados de validación interna se muestran en la eTabla 5 en el Suplemento 1. Como se muestra en la Tabla 2 y en la eFigura 3 en el Suplemento 1, en los datos no vistos, todos los

modelos lograron un AUROC superior a 0.70. El Modelo D (potenciación del gradiente automático) tuvo una exactitud de 0.77 para identificar infantes teniendo RC o DCT a los 2 años de seguimiento, con una sensibilidad de 0.55 y una especificidad de 0.83 en un umbral de decisión de 0.55 y una especificidad de 0.83 en un umbral de decisión de 0.5. El Modelo B (regresión logística) tuvo una exactitud de 0.76, con una sensibilidad de 0.48 y una especificidad de 0.84. El Modelo A (foresta al azar) tuvo una exactitud de 0.74, y el modelo C (máquina de vector de apoyo) tuvo una exactitud de 0.70. Pese a que el Modelo D alcanzó la exactitud más elevada de 0.77 en el umbral de 0.55, la curva AUROC sugirió que el modelo B podría alcanzar la sensibilidad más alta con la alteración de ese umbral. El Modelo B, de regresión logística conteniendo 26 características, alcanzó un AUROC de 0.77 (IC 95%, 0.71-0,83).

Aplicación del Modelo B

El umbral de decisión del modelo B fue descendido a 0.20 para mejorar la sensibilidad. En la Tabla 3 se muestra una tabla de contingencia mostrando una aplicación hipotética. Entre los 3000 infantes MPT sin malformaciones mayores, se esperaría que aproximadamente 650 presentarían RC a los 24 meses. Al egreso de UCIN, el modelo B podría identificar correctamente 605 de 650 infantes que podrían presentar RC a los 24 meses (sensibilidad, 0.93), y 1081 de 2350 que podrían no presentar (especificidad, 0.46). Entre 1874 infantes que tuvieron predicción de RC, 605 (32.3%) evolucionaron a presentarlo. Entre 1126 infantes con predicción de no presentar RC, 1081 (96%) no lo presentarían.

Tabla 2. Validación externa sobre datos de prueba invisible para modelos A y D

Modelo	Algoritmo	No. De Características	Hiperparámetro valor final	Sensibilidad	Especificidad	AUROC (ic 95%)	Precisión equilibrada	Exactitud
A	Aleatorizado	26	Mtry (6)	0.49	0.81	0.74 (0.68-0.81)	0.65	0.74
B	Regresión logística	26	NA	0.48	0.84	0.77 (0.71-0.83)	0.66	0.76
C	Vector de soporte Mecanizado	26	C(2), Sigma (0.05)	0.64	0.72	0.75 (0.69-0.81)	0.68	0.70
D	Aumento de gradiente mecanizado	26	Ntrees (800), Interaction depth (15), Shrinkage (0.1), n.minobsinnode (10)	0.55	0.83	0.76 (0.70-0.82)	0.69	0.77

Abreviaturas AUROC, Área Debajo,Receptor, Operando, Característica, Curva. Ntrees (800) etc... Sin traducción por ser concepto técnico.

Importancia de Característica

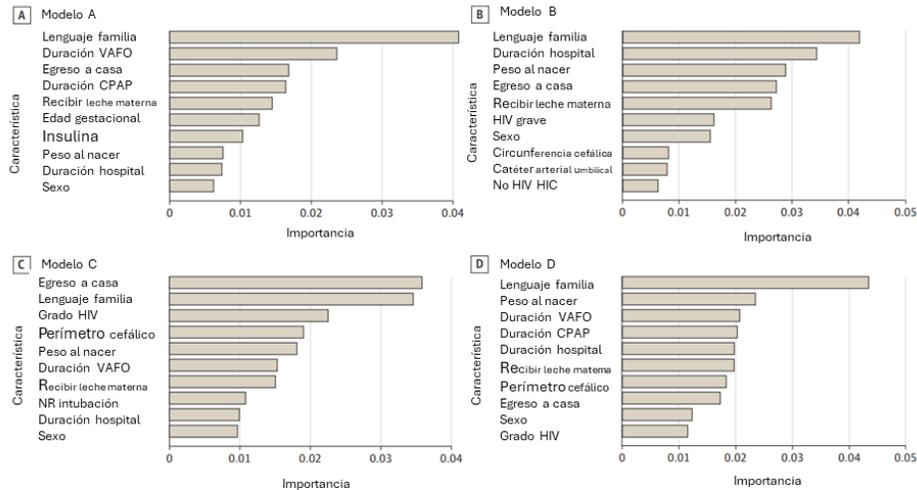
Los grupos de importancia característica para los modelos A a D se muestran en la Figura 2. Las 10 características más importantes para predicción en el modelo B fueron (en orden de importancia): lengua familiar, duración de la hospitalización, PN, si el infante fue externado a su casa, si el infante estaba recibiendo leche humana al egreso hospitalario, grado de HIV, sexo, circunferencia craneana, uso de catéter arterial umbilical, y hemorragia intracraneana no-HIV. Lengua familiar, PN, sexo, duración de hospitalización, si el infante recibía leche humana al egreso, y si egresaba directo a su hogar se incluyeron entre las 10 características más importantes en todos los modelos. Las 5 características más importantes para predecir RC en el modelo B fueron lengua familiar no- Escandinava, duración prolongada de hospitalización, bajo PN, egreso hacia otro destino que no sea el hogar, y el infante sin recibir leche humana al egreso hospitalario. Los grupos de efecto de características se encuentran en las eFiguras 4 y 5 en el Suplemento 1.

Tabla 3. Tabla de contingencia para modelos B en un umbral de decisión de 0.20^a

Predicción	Desarrollo cognitivo a los 24 m		Total
	Retraso cognitivo	No retraso cognitivo	
Retraso cognitivo	605	1269	1874
No retraso cognitivo	45	1081	1126
Total	650	2350	3000

^aNúmeros están basados sobre población hipotética de 3000 niños muy prematuros sin malformaciones mayores.

Figura 2. Gráfico de las características de importancia para los modelos de la A a D



Niños no egresados a su casa fueron egresados a otro hospital o casa de cuidados. CPAP indica presión continua de vías aéreas; VAFO, ventilación de alta frecuencia oscilatoria; HIC, hemorragia intracraneana; HIV, hemorragia intraventricular; NR reanimación neonatal.

Discusión

Usando datos clínicos que están fácilmente disponibles, hemos mostrado que un modelo predictivo conteniendo 26 características podría identificar al egreso de UCIN el 93% de infantes MPT sobrevivientes que podrían mostrar tamizaje positivo para RC a los 24 meses, con una especificidad correspondiente de 46%. Tal modelo podría ser incorporado en un registro de salud electrónico de la unidad para permitir alertas para el requerimiento de intervención temprana dirigida.

La intervención temprana puede mejorar el resultado cognitivo de un infante PT en aproximadamente la mitad de un DS en una prueba cognitiva estandarizada en la edad pre-escolar (equivalente a 7.5 puntos de CI) (14). Para 25% de infantes con RC en este estudio, una mitad de DS de mejora en Bayley-III podría traer su función cognitiva al rango normal, comparable con sus pares desarrollándose típicamente.

Alternativamente, si estos niños/as no son identificados al nacer, pueden pasar los chequeos tempranos del desarrollo (34) y perder la ventana de oportunidad de la intervención temprana. Sus dificultades cognitivas pueden no surgir hasta que aparecen con el fallo académico o las dificultades conductuales en los primeros años escolares (35).

En el umbral de decisión propuesto, una especificidad de 46% resultaría en una proporción sustancial de resultados de tamizaje falsos-positivos. Sin embargo, la aceptabilidad de resultados falsos-positivos entre familias es factible que sea elevada dado que todos los infantes positivos en el tamizaje reciben intervención efectiva, y el número estimado a tratar para evitar 1 individuo con cognición debajo del rango normal, es bajo ($n=12$). Los programas basados en evidencia, favoreciendo la relación parental-infante y enriqueciendo el ambiente extienden sus beneficios a dominios diferentes a la cognición, tales como mejores habilidades motrices y sociales del infante, así como ansiedad y depresión parental reducidas (14, 36-38). Según esto, cualquier sobretratamiento, podría ser considerado costo-efectivo.

El enfoque estadístico usado en este estudio fue diseñado para optimizar la predicción, no para investigar relaciones causales. Importantes predictores, como la intubación al nacer, uso de catéteres umbilicales, tratamiento con insulina para hiperglucemia, duración del soporte ventilatorio y hospitalización, y si el infante recibía o no leche humana, son proclives a reflejar la severidad de la enfermedad después de nacer.

Sin embargo, estos factores son modificables, y no puede ser excluido que la duración del tiempo pasado en UCIN, donde los procedimientos invasivos, ventilación artificial, y otras exposiciones sensoriales nocivas ocurren regularmente, pueden tener un efecto en la evolución cognitiva posterior (39). El cuidado del desarrollo, diseñado para ajustar el ambiente neonatal para reducir el estrés neonatal y promover el crecimiento neuronal, puede ser benéfico (40), como también pueden los esfuerzos para reducir la duración de la ventilación invasiva y facilitar el egreso al hogar. El egreso temprano para infantes PT estables ha demostrado ser seguro y mejorar el bienestar familiar (41, 42).

Limitaciones

Nuestro estudio tiene importantes limitantes. La población en estudio consistió de sobrevivientes con puntaje cognitivo a los 2 años de edad en Bayley-III. Esta población no representa a todos/as los infantes en Suecia. Pese a que el programa de seguimiento fue lanzado en 2015, la realización de las evaluaciones con Bayley-III ha sido limitada por restricciones organizacionales y de recursos, particularmente por falta de capacidad de testeo en áreas remotas. La mayoría de los niños incluidos nacieron EPT (<28 semanas) y residían en las 3 áreas urbanas más densamente pobladas (43). Los sesgos de selección pueden por lo tanto haber introducido una sobreestimación de las tasas de RC en infantes MPT, y los resultados pueden no ser válidos para todos/as los niños MPT. Sin embargo, los infantes EPT evaluados con Bayley-III en 2016 a 2019 no tuvieron diferencia significativa de aquellos no evaluados en cuanto a sexo, EG, PN, PEG, o incidencia de estadios HIV 3 a 4 en el período neonatal (M.N. datos no publicados, 1º de Noviembre, 2023)

El resultado usado en este estudio fue la función cognitiva, pero los infantes MPT con frecuencia presentan déficits en múltiples dominios (44). Pese a que las trayectorias cognitivas individuales variarán por factores socioambientales y familiares, la mayoría de los sobrevivientes EPT con bajos puntajes Bayley-III a los 2 años de edad tendrán CI por debajo del promedio más tarde en la infancia (45, 46). Sin embargo, una proporción significativa de categorizados/as como desarrollados típicamente a los 2 años de edad experimentarán empeoramiento posterior de la función cognitiva (47).

El nivel de educación parental y la lengua de la familia fueron características importantes en este estudio, y la información más detallada sobre el ambiente socioeconómico pueden haber mejorado el comportamiento del modelo (48). La lengua familiar extranjera puede ser surrogante de bajo nivel socioeconómico o inmigración. La administración de Bayley-III es fuertemente dependiente del lenguaje, y los niños/as haciendo la prueba en lengua no- nativa o a través de un intérprete pueden estar en desventaja, con aparente pobre desempeño (49).

Conclusiones

Este estudio pronóstico encontró que es posible identificar RC antes del egreso de UCIN en infantes MPT. Futuro trabajo debiera incluir mayor validación del modelo en infantes MPT. Otros grupos de investigadores usando datos neurofisiológicos, microbioma, o imágenes deberían considerar incluir estas 26 características clínicas identificadas aquí en próximos modelos predictivos.

REFERENCIAS

1. Walani SR. Global burden of preterm birth. *Int J Gynaecol Obstet.* 2020;150(1):31-33. doi:10.1002/ijgo.13195
2. Bhutta AT, Cleves MA, Casey PH, Cradock MM, Anand KJS. Cognitive and behavioral outcomes of school-aged children who were born preterm: a meta-analysis. *JAMA.* 2002;288(6):728-737. doi:10.1001/jama.288.6.728
3. Larroque B, Ancel PY, Marret S, et al; EPIPAGE Study group. Neurodevelopmental disabilities and special care of 5-year-old children born before 33 weeks of gestation (the EPIPAGE study): a longitudinal cohort study. *Lancet.* 2008;371(9615):813-820. doi:10.1016/S0140-6736(08)60380-3
4. Wolke D, Meyer R. Cognitive status, language attainment, and prereading skills of 6-year-old very preterm children and their peers: the Bavarian Longitudinal Study. *Dev Med Child Neurol.* 1999;41(2):94-109.
5. van Noort-van der Spek IL, Franken MCJP, Weisglas-Kuperus N. Language functions in preterm-born children: a systematic review and meta-analysis. *Pediatrics.* 2012;129(4):745-754. doi:10.1542/peds.2011-1728
6. Twilhaar ES, Wade RM, de Kieviet JF, van Goudoever JB, van Elburg RM, Oosterlaan J. Cognitive outcomes of children born extremely or very preterm since the 1990s and associated risk factors: a meta-analysis and metaregression. *JAMA Pediatr.* 2018;172(4):361-367. doi:10.1001/jamapediatrics.2017.5323
7. Lager A, Bremberg S, Vågerö D. The association of early IQ and education with mortality: 65 year longitudinal study in Malmö, Sweden. *BMJ.* 2009;339:b5282. doi:10.1136/bmj.b5282
8. Alesi M, Rappo G, Pepi A. Emotional profile and intellectual functioning: a comparison among children with borderline intellectual functioning, average intellectual functioning, and gifted intellectual functioning. *SAGE Open.* Published online July 13, 2015. doi:10.1177/2158244015589995
9. Forrest LF, Hodgson S, Parker L, Pearce MS. The influence of childhood IQ and education on social mobility in the Newcastle Thousand Families birth cohort. *BMC Public Health.* 2011;11:895. doi:10.1186/1471-2458-11-895
10. Furnham A, Cheng H. Childhood cognitive ability predicts adult financial well-being. *J Intell.* 2016;5(1):3. doi: 10.3390/jintelligence5010003
11. Whalley LJ, Deary IJ. Longitudinal cohort study of childhood IQ and survival up to age 76. *BMJ.* 2001;322(7290):819. doi:10.1136/bmj.322.7290.819
12. Twilhaar ES, de Kieviet JF, van Elburg RM, Oosterlaan J. Academic trajectories of very preterm born children at school age. *Arch Dis Child Fetal Neonatal Ed.* 2019;104(4):F419-F423.
13. Sefidkar R, Zayeri F, Kazemi E, Salehi M, Dehnad A, Hafizi M. A trend study of preterm infant mortality

rate in developed and developing countries over 1990 to 2017. *Iran J Public Health*. 2021;50(2):369-375. doi:10.18502/ijph.v50i2.5353

14. Spittle AJ, Orton J, Doyle LW, Boyd R. Early developmental intervention programs post hospital discharge to prevent motor and cognitive impairments in preterm infants. *Cochrane Database Syst Rev*. 2007;(2):CD005495. doi:10.1002/14651858.CD005495.pub2
15. Linsell L, Malouf R, Morris J, Kurinczuk JJ, Marlow N. Prognostic factors for poor cognitive development in children born very preterm or with very low birth weight: a systematic review. *JAMA Pediatr*. 2015;169(12): 1162-1172. doi:10.1001/jamapediatrics.2015.2175
16. Crilly CJ, Haneuse S, Litt JS. Predicting the outcomes of preterm neonates beyond the neonatal intensive care unit: what are we missing? *Pediatr Res*. 2021;89(3):426-445. doi:10.1038/s41390-020-0968-5
17. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med*. 2015;162(1):W1-73. doi: 10.7326/M14-0698
18. Norman M, Källén K, Wahlström E, Håkansson S; SNQ Collaboration. The Swedish Neonatal Quality Register— contents, completeness and validity. *Acta Paediatr*. 2019;108(8):1411-1418. doi:10.1111/apa.14823
19. Swedish Neonatal Society. National guidelines for follow-up after neonatal intensive care. Swedish Neonatal Society. 2015. Accessed July 12, 2023. [https://neo.barnlakarforeningen.se/riktlinjer/wardprogram/uppfoljning-avneonatala-riskbarn-2/](https://neo.barnlakarforeningen.se/riktlinjer/vardprogram/uppfoljning-avneonatala-riskbarn-2/)
20. European Commission. EUROCAT Guide 1.5 Chapter 3.3 EUROCAT Subgroups of Congenital Anomalies. 2022. Accessed November 2, 2023. https://eu-rd.platform.jrc.ec.europa.eu/system/files/public/eurocat/Guide_1.5_Chapter_3.3.pdf
21. Serenius F, Källén K, Blennow M, et al; EXPRESS Group. Neurodevelopmental outcome in extremely preterm infants at 2.5 years after active perinatal care in Sweden. *JAMA*. 2013;309(17):1810-1820. doi:10.1001/jama. 2013.3786
22. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; 2021. Accessed December 5, 2023. <https://www.R-project.org>
23. Stekhoven DJ, Bühlmann P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*. 2012;28(1):112-118. doi:10.1093/bioinformatics/btr597
24. Hong S, Lynn HS. Accuracy of random-forest-based imputation of missing data in the presence of non-normality, non-linearity, and interaction. *BMC Med Res Methodol*. 2020;20(1):199. doi:10.1186/s12874-020-01080-1
25. Kursa MB, Rudnicki WR. Feature selection with the boruta package. *J Stat Softw*. 2010;36:1-13. doi:10.18637/jss.v036.i11
26. Degenhardt F, Seifert S, Szymczak S. Evaluation of variable selection methods for random forests and omics data sets. *Brief Bioinform*. 2019;20(2):492-503. doi:10.1093/bib/bbx124
27. Chawla N, Bowyer K, Hall L, Kegelmeyer W. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res*. 2002;16:341-348. doi:10.1613/jair.953
28. Blagus R, Lusa L. SMOTE for high-dimensional class-imbalanced data. *BMC Bioinformatics*. 2013;14:106. doi: 10.1186/1471-2105-14-106
29. Kuhn M. Variable selection using the caret package. 2009. Accessed June 6, 2023. <https://citeseerx.ist.psu.edu/document?repid=rep1&type=p>
30. Bohr A, Memarzadeh K. The rise of artificial intelligence in healthcare applications. *Artif Intell Healthcare*. 2020:25–60.
31. Deo RC. Machine learning in medicine. *Circulation*. 2015;132(20):1920-1930. doi:10.1161/CIRCULATIONAHA.115.001593
32. Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian J Intern Med*. 2013;4(2):627-635.

33. Greenwell B, Boehmke B. Variable importance plots: an introduction to the vip package. *R J.* 2020;12:343-366. doi:10.32614/RJ-2020-013
34. Bowe AK, Hourihane J, Staines A, Murray DM. The predictive value of the ages and stages questionnaire in late infancy for low average cognitive ability at age 5. *Acta Paediatr.* 2022;111(6):1194-1200. doi:10.1111/apa.16309
35. Bowe AK, Staines A, Murray DM. Below average cognitive ability: an under researched risk factor for emotional-behavioural difficulties in childhood. *Int J Environ Res Public Health.* 2021;18(24):12923. doi:10.3390/ijerph182412923
36. Benzies KM, Magill-Evans JE, Hayden KA, Ballantyne M. Key components of early intervention programs for preterm infants and their parents: a systematic review and meta-analysis. *BMC Pregnancy Childbirth.* 2013; 13(Suppl 1)(suppl 1):S10. doi:10.1186/1471-2393-13-S1-S10
37. Herd M, Whittingham K, Sanders M, Colditz P, Boyd RN. Efficacy of preventative parenting interventions for parents of preterm infants on later child behavior: a systematic review and meta-analysis. *Infant Ment Health J.* 2014;35(6):630-641. doi:10.1002/imhj.21480
38. Ritchie K, Bora S, Woodward LJ. Social development of children born very preterm: a systematic review. *Dev Med Child Neurol.* 2015;57(10):899-918. doi:10.1111/dmcn.12783
39. Soleimani F, Azari N, Ghiasvand H, Shahrokhi A, Rahmani N, Fatollahierad S. Do NICU developmental care improve cognitive and motor outcomes for preterm infants? a systematic review and meta-analysis. *BMC Pediatr.* 2020;20(1):67. doi:10.1186/s12887-020-1953-1
40. Khurana S, Kane AE, Brown SE, Tarver T, Dusing SC. Effect of neonatal therapy on the motor, cognitive, and behavioral development of infants born preterm: a systematic review. *Dev Med Child Neurol.* 2020;62(6): 684-692. doi:10.1111/dmcn.14485
41. Brooten D, Kumar S, Brown LP, et al. A randomized clinical trial of early hospital discharge and home follow-up of very-low-birth-weight infants. *N Engl J Med.* 1986;315(15):934-939. doi:10.1056/NEJM198610093151505
42. Sáenz P, Cerdá M, Díaz JL, et al. Psychological stress of parents of preterm infants enrolled in an early discharge programme from the neonatal intensive care unit: a prospective randomised trial. *Arch Dis Child Fetal Neonatal Ed.* 2009;94(2):F98-F104. doi:10.1136/adc.2007.135921
43. Swedish Neonatal Quality Register. Annual report from the Swedish Neonatal Quality Register 2021 (in Swedish). June 2021. Accessed October 30, 2023. [https://www.medscinet.com/PNQ/uploads/website/Neonatalv%C3%A5rdsregistrets%20%C3%85rsrapport%202021%20\(final%20version\).pdf](https://www.medscinet.com/PNQ/uploads/website/Neonatalv%C3%A5rdsregistrets%20%C3%85rsrapport%202021%20(final%20version).pdf)
44. van Baar AL, van Wassenaeer AG, Briët JM, Dekker FW, Kok JH. Very preterm birth is associated with disabilities in multiple developmental domains. *J Pediatr Psychol.* 2005;30(3):247-255. doi:10.1093/jpepsy/jsi035
45. Lowe J, Bann CM, Dempsey AG, et al; Eunice Kennedy Shriver National Institute of Child Health and Human Development Neonatal Research Network. Do Bayley-III composite scores at 18-22 months corrected age predict full-scale IQ at 6-7 years in children born extremely preterm? *J Pediatr.* Published online August 25, 2023. doi:10.1016/j.jpeds.2023.113700
46. Serenius F, Ewald U, Farooqi A, et al; Extremely Preterm Infants in Sweden Study Group. Neurodevelopmental outcomes among extremely preterm infants 6.5 years after active perinatal care in Sweden. *JAMA Pediatr.* 2016;170(10):954-963. doi:10.1001/jamapediatrics.2016.1210
47. Koller H, Lawson K, Rose SA, Wallace I, McCarton C. Patterns of cognitive development in very low birth weight children during the first six years of life. *Pediatrics.* 1997;99(3):383-389. doi:10.1542/peds.99.3.383
48. Benavente-Fernández I, Synnes A, Grunau RE, et al. Association of socioeconomic status and brain injury with neurodevelopmental outcomes of very preterm children. *JAMA Netw Open.* 2019;2(5):e192914. doi:10.1001/jamanetworkopen.2019.2914
49. Lowe JR, Nolen TL, Vohr B, Adams-Chapman I, Duncan AF, Watterberg K. Effect of primary language on

developmental testing in children born extremely preterm. Acta Paediatr. 2013;102(9):896-900.
doi:10.1111/apa.12310

