

Visitas domiciliarias a escala y desarrollo infantil

M. Caridad Araujo
Marta Dormal
Sally Grantham-McGregor
Fabiola Lazarte
Marta Rubio-Codina
Norbert Schady

Visitas domiciliarias a escala y desarrollo infantil

M. Caridad Araujo
Marta Dormal
Sally Grantham-McGregor
Fabiola Lazarte
Marta Rubio-Codina
Norbert Schady

Catalogación en la fuente proporcionada por la
Biblioteca Felipe Herrera del

Banco Interamericano de Desarrollo

Visitas domiciliarias a escala y desarrollo infantil / M. Caridad Araujo, Marta Dormal,
Sally Grantham-McGregor, Fabiola Lazarte, Marta Rubio-Codina, Norbert Schady.
p. cm. — (Documento de trabajo del BID ; 1277)

Incluye referencias bibliográficas.

1. Child development-Government policy-Peru. 2. Child care services-Government
Policy-Peru. 3. Home visits (Education)-Peru. 4. Children-Nutrition-Peru. I. Araujo,
Maria Caridad. II. Dormal, Marta. III. Grantham-McGregor, Sally M. IV. Lazarte,
Fabiola. V. Rubio-Codina, Marta. VI. Schady, Norbert Rüdiger, 1967- VII. Banco
Interamericano de Desarrollo. División de Protección Social y Salud. VIII. Serie.
IDB-WP-01277

<http://www.iadb.org>

Copyright © 2021 Banco Interamericano de Desarrollo. Esta obra se encuentra sujeta a una licencia Creative Commons IGO 3.0 Reconocimiento-NoComercial-SinObrasDerivadas (CC-IGO 3.0 BY-NC-ND) (<http://creativecommons.org/licenses/by-nc-nd/3.0/igo/legalcode>) y puede ser reproducida para cualquier uso no-comercial otorgando el reconocimiento respectivo al BID. No se permiten obras derivadas.

Cualquier disputa relacionada con el uso de las obras del BID que no pueda resolverse amistosamente se someterá a arbitraje de conformidad con las reglas de la CNUDMI (UNCITRAL). El uso del nombre del BID para cualquier fin distinto al reconocimiento respectivo y el uso del logotipo del BID, no están autorizados por esta licencia CC-IGO y requieren de un acuerdo de licencia adicional.

Después de un proceso de revisión por pares, y con el consentimiento previo y por escrito del BID, una versión revisada de esta obra podrá reproducirse en cualquier revista académica, incluyendo aquellas referenciadas por la Asociación Americana de Economía a través de EconLit, siempre y cuando se otorgue el reconocimiento respectivo al BID, y el autor o autores no obtengan ingresos de la publicación. Por lo tanto, la restricción a obtener ingresos de dicha publicación sólo se extenderá al autor o autores de la publicación. Con respecto a dicha restricción, en caso de cualquier inconsistencia entre la licencia Creative Commons IGO 3.0 Reconocimiento-NoComercial-SinObrasDerivadas y estas declaraciones, prevalecerán estas últimas.

Note que el enlace URL incluye términos y condiciones adicionales de esta licencia.

Las opiniones expresadas en esta publicación son de los autores y no necesariamente reflejan el punto de vista del Banco Interamericano de Desarrollo, de su Directorio Ejecutivo ni de los países que representa.



scl_sph@iadb.org

www.iadb.org/SocialProtection

Visitas Domiciliarias a Escala y Desarrollo Infantil

M. Caridad Araujo¹
Marta Dormal²
Sally Grantham-McGregor³
Fabiola Lazarte⁴
Marta Rubio-Codina¹
Norbert Schady⁵

¹Banco Interamericano de Desarrollo

²Harvard Graduate School of Education, Harvard University

³University College London

⁴Instituto de Investigación Nutricional, Perú

⁵Banco Mundial

* Los autores agradecen al Gobierno de Perú, en particular al Ministerio de Economía y Finanzas, al Ministerio de Desarrollo e Inclusión Social y al programa Cuna Más, por las valiosas discusiones y por el apoyo a esta evaluación. También queremos agradecer a todas las familias, al personal de Cuna Más y a los equipos en el terreno que participaron en el estudio. Por último, un agradecimiento especial a Rafael Novella, Romina Tomé, Erika Dunkelberg, Beatriz Ore, María Adelaida Martínez y Julieth Parra por su excepcional asistencia técnica y de investigación en varias etapas del proyecto. Los puntos de vista que se presentan en este documento no representan al Banco Interamericano de Desarrollo, a su junta directiva ni a los países que representan.

Resumen

La mayor parte de la evidencia sobre los impactos de las políticas para la promoción del desarrollo infantil se refiere a programas piloto implementados en circunstancias estrictamente controladas. Este estudio aporta evidencia novedosa sobre los efectos de un programa de visitas domiciliarias implementado a escala. El programa que estudiamos, Cuna Más en Perú, se inició de cero en 2012 y, en tan solo tres años, ya ofrecía visitas domiciliarias semanales a más de 67.000 niños en zonas rurales. La estrategia de identificación proviene de la asignación aleatoria de distritos a las condiciones de tratamiento y control. Mostramos que, después de aproximadamente dos años, los niños asignados aleatoriamente al tratamiento tienen puntajes de desarrollo infantil que son 0,10 desviaciones estándar más altos que los niños asignados al grupo de control. La relación beneficio-costos estimada de la intervención es de 5,4.

Palabras clave: visitas domiciliarias, desarrollo infantil, implementación a escala, evaluación experimental, Programa Nacional Cuna Más, Perú.

Códigos JEL: I38, J24, O15

1. Introducción

Existe una extensa literatura en economía, educación y medicina que ha demostrado que el desarrollo en la primera infancia tiene efectos a largo plazo en el desempeño en la edad adulta. La calidad de la estimulación en el hogar y la medida en que los adultos son sensibles, consistentes y receptivos en sus interacciones con los niños son determinantes críticos del desarrollo de los bebés y los niños pequeños.¹

En este estudio, analizamos los impactos de un programa de visitas domiciliarias dirigido a madres económicamente desfavorecidas con hijos menores de 3 años. El programa que estudiamos, Cuna Más en Perú, buscaba mejorar el desarrollo infantil a través de tutorías entre pares realizadas a las madres por parte de miembros de su misma comunidad (de ahora en adelante, "facilitadoras" dado que en su mayoría eran mujeres). Estas facilitadoras utilizaban un currículo estructurado centrado en la estimulación cognitiva y del lenguaje a través del juego.

Cuna Más fue creado de cero por una nueva administración en 2012. En tres años, ya ofrecía visitas domiciliarias a más de 67.000 familias de bajos ingresos en zonas rurales de 586 distritos.² Esto lo convierte en uno de los programas de este tipo más grandes en el mundo.³ La ampliación de cobertura del programa fue gradual, con distritos asignados aleatoriamente a varias fases de expansión.

Usamos la implementación gradual del programa para estimar sus efectos en el desarrollo infantil por métodos experimentales. Específicamente, comparamos los niños de los distritos asignados aleatoriamente a las visitas domiciliarias en la primera fase con los niños que fueron asignados a la última fase de implementación. En el momento de la encuesta de seguimiento, los niños de la primera fase —nuestro grupo de tratamiento— habían sido elegibles para las visitas domiciliarias durante aproximadamente 2 años,

¹ Revisiones de literatura realizadas por economistas incluyen Almond y Currie (2011); Almond, Currie y Duque (2018); Attanasio (2015); Heckman y Mosso (2014). A lo largo del documento usamos la palabra 'niño' para referirnos a niñas y niños indistintamente.

² Existen 1.854 distritos en Perú. Como comentamos más adelante, los servicios de visitas domiciliarias de Cuna Más no estaban pensados para llegar a todos los distritos.

³ En Estados Unidos, el programa *Early Head Start* (EHS) cubre a unos 250.000 niños a través de cientos de proveedores individuales (<https://eclkc.ohs.acf.hhs.gov/programs/article/about-early-head-start-program>, consultado el 1 de abril de 2018). El EHS tiene distintas modalidades. En 2013, el 42% de los niños estaban en la modalidad domiciliaria, en la que los servicios se prestan en el hogar de la familia mediante una facilitadora y a través de actividades grupales (Mayoral 2014). En los países en desarrollo, otros programas extensos de visitas domiciliarias incluyen el programa *Creciendo con Nuestros Hijos* en Ecuador, que cubre a aproximadamente 200.000 niños cada año (comunicación directa con el Ministerio de Inclusión Económica y Social, 2018), y el programa *Criança Feliz* en Brasil, que alcanza a 233.000 familias (comunicación directa con el Ministerio de Desarrollo Social, 2018).

mientras que los de la última fase —nuestro grupo de control— todavía no eran elegibles para el programa.

Mostramos que Cuna Más mejoró modesta pero significativamente el desarrollo infantil. El análisis de intención de tratamiento arroja impactos estimados de 0,10 desviaciones estándar para un indicador global del desarrollo infantil. Las estimaciones de "tratamiento sobre los tratados", que tienen en cuenta el hecho de que una proporción sustancial de familias asignadas aleatoriamente a la primera fase de expansión no recibieron visitas domiciliarias, ofrecen impactos de 0,15 desviaciones estándar. Un cálculo somero sugiere que la relación beneficio-costos de la intervención, que fue de 5,4, fue positiva y de tamaño considerable.

Consistente con las mejoras en el desarrollo infantil, encontramos que Cuna Más cambió las prácticas de crianza en los hogares: los padres asignados aleatoriamente al grupo de tratamiento jugaban más a menudo con sus hijos y eran menos propensos a adoptar estrategias de crianza punitivas (como pegar o gritar al niño). Asimismo, entre los niños de 36 meses o más (edad en la que dejan de ser elegibles para las visitas domiciliarias, pero pasan a ser elegibles para la educación preescolar pública), aquellos que fueron asignados aleatoriamente al tratamiento de Cuna Más tuvieron una probabilidad 8 puntos porcentuales mayor de ser inscritos en el centro de preescolar, con base a una tasa de inscripción en preescolar del 50% en el grupo de control.

Este estudio tiene varias contribuciones. En primer lugar, nuestro trabajo contribuye a la poca literatura existente sobre el impacto de los programas de visitas domiciliarias. En comparación con otras intervenciones en la primera infancia —por ejemplo, transferencias monetarias o el acceso a la educación preescolar—, se sabe menos sobre la eficacia de los programas que buscan mejorar las prácticas de crianza directamente. Esto, a pesar de que los programas de visitas domiciliarias son a menudo la única forma de llegar a los niños de las zonas rurales dispersas de los países en desarrollo —precisamente donde viven los niños más vulnerables y que tienden a tener los mayores retrasos en su desarrollo.

En segundo lugar, nuestro trabajo se suma a la literatura sobre el mejor uso de recursos locales para lograr cambios de comportamiento. Las facilitadoras de Cuna Más, en su mayoría mujeres, pertenecían a las mismas comunidades que las familias con las que iban a trabajar. Esto se hizo para crear confianza entre las familias y las facilitadoras, y para reducir los costos del programa. Sin embargo, el nivel educativo promedio solía ser

bastante bajo en las zonas en las que operaba Cuna Más y, en consecuencia, también lo era el de las facilitadoras: el 56% no había completado la educación secundaria.

Mostramos que, a pesar de los bajos niveles de educación de las facilitadoras, el programa tuvo un efecto positivo, aunque modesto, en el desarrollo de los niños.⁴

En tercer lugar, y tal vez lo más importante, nuestro estudio habla de los desafíos que plantea el escalamiento de intervenciones prometedoras.⁵ Como comentamos a continuación, la intervención que estudiamos se basó en un ensayo de eficacia muy pequeño en Jamaica (con 64 niños que recibieron visitas domiciliarias) y en un programa implementado a escala piloto en Colombia (con 720 niños que recibieron visitas). En cambio, en sus primeros tres años de implementación, el programa Cuna Más estaba realizando visitas domiciliarias a casi 100 veces más niños que los del programa piloto colombiano. Hay muchas cosas que pueden fallar en la transición de una prueba piloto pequeña, que cubre sólo unas docenas o unos cientos de niños, a una intervención a escala que cubre a decenas de miles. Analizamos algunas de las dificultades de implementación a las que se enfrentó Cuna Más, y especulamos sobre cómo éstas pueden haber afectado a los impactos del programa.

El resto del documento se organiza de la siguiente manera. En la sección 2 se analiza el programa de visitas domiciliarias Cuna Más y se explica en qué medida se basa en programas anteriores. La sección 3 analiza el diseño y los datos de la evaluación. Presentamos los resultados en la sección 4 y concluimos en la sección 5.

⁴ La tutoría y el coaching estructurados entre pares (en contraposición a los informales) se han utilizado ampliamente en educación y salud; por ejemplo, en programas en los que profesores o enfermeras "expertos" observan y proporcionan *feedback* a otros profesores o enfermeras (a menudo principiantes). La tutoría y el coaching son también un elemento importante en varios programas para jóvenes. Por ejemplo, en el programa *Becoming A Man*, que pretendía reducir la delincuencia y la deserción escolar entre jóvenes desfavorecidos de Chicago, los mentores que trabajaban con los jóvenes solían ser contratados en barrios muy similares a aquellos en los que iban a trabajar (Heller et al. 2017). Sin embargo, en estos programas los mentores tienden a ser altamente calificados. También existen muchos ejemplos de programas que utilizan facilitadores o promotores locales para cambiar comportamientos, incluidos los programas para promover el uso de agua limpia (Zwane y Kremer 2007), la lactancia materna (véase Chapman et al. 2010; Haroon et al. 2013; Rollins et al. 2016), la adopción de mosquiteros tratados con insecticida y otras estrategias para prevenir la malaria (véase Salam et al. 2014 para un análisis), o los comportamientos saludables para los padres de niños pequeños (véase Prost et al. 2013 para un análisis y meta-análisis), pero estos programas generalmente no utilizan un plan de estudios estructurado ni se basan en visitas domiciliarias regulares. Véase también Björkman et al. (2019) para un ejemplo reciente en la literatura económica. Para cambiar los comportamientos de los padres de forma escalable y sostenible, Cuna Más diseñó una intervención que, por un lado, era más estructurada que la mayoría de los programas que dependen de facilitadores locales, pero que, por otro lado, se basaba en facilitadores menos cualificados que en la mayoría de los programas de coaching y tutoría entre pares.

⁵ En cuanto a los desafíos del escalamiento, véanse, especialmente, Al-Ubaydli, List y Suskind (2017) y la colección de artículos en List, Suskind y Supplee (2021); también, Banerjee et al. (2017).

2. El programa de visitas domiciliarias Cuna Más

Los programas de visitas domiciliarias que buscan mejorar el desarrollo infantil han sido del interés de hacedores de política pública e investigadores en países en desarrollo durante más de dos décadas. Existe un pequeño número de evaluaciones aleatorias de estos programas, publicadas principalmente en la literatura médica (véase Aboud y Yousafzai 2015 para una revisión). Todos estos estudios se refieren a ensayos de eficacia o a pilotos a pequeña escala, en lugar de a intervenciones a gran escala.⁶

La evaluación más conocida de un programa de visitas domiciliarias en un país en desarrollo comenzó en 1986 en Kingston, Jamaica. El programa buscaba mejorar el desarrollo de los niños mediante la estimulación psicosocial y el juego. Las visitas eran realizadas por promotoras comunitarias de salud (paraprofesionales). Los niños del grupo de tratamiento recibieron visitas domiciliarias semanales a partir de los 9-24 meses de edad durante 2 años.

La intervención de Jamaica se ha evaluado en varios experimentos aleatorios controlados. El más ambicioso es un experimento con 129 niños que, hasta la fecha, han sido objeto de estudio durante más de 20 años (estudio longitudinal). Los resultados iniciales mostraron impactos muy grandes a corto plazo (después de 24 meses) en desarrollo cognitivo, aproximadamente de 0,85 desviaciones estándar (Grantham-McGregor et al. 1991; Walker et al. 2000). Unos veinte años después de la finalización del programa, los niños que habían sido asignados aleatoriamente al grupo de tratamiento tenían más años de escolaridad, mejores puntuaciones en exámenes estandarizados, mayor probabilidad de tener trabajo y sus ingresos eran un 25% más altos que los de los niños asignados al grupo de control. También tenían mayores niveles de desarrollo cognitivo, personal-social y salud mental (Gertler et al. 2014; Walker et al. 2011).

El modelo jamaicano —ahora denominado *Reach Up* (Walker et al. 2018)— ha sido adaptado y probado en varios países. Una de esas adaptaciones se hizo en Colombia. Allí, los investigadores implementaron un programa con un currículo muy similar al que se usó en Jamaica. Sin embargo, las visitas domiciliarias en Colombia fueron realizadas por

⁶ Existe también literatura sobre programas de visitas domiciliarias en los Estados Unidos. Olds (2010) hace una revisión de las evaluaciones del programa *Family-Nurse Partnership* (NFP). Love et al. (2013) muestran los resultados de una evaluación experimental del programa *Early Head Start* (EHS) que presta servicios a niños muy vulnerables. Algunos centros de EHS incluyen servicios prestados en el hogar.

madres comunitarias (en lugar de promotoras comunitarias del sector salud). Estas madres, conocidas como "madres líderes", habían sido elegidas por sus comunidades para servir de enlace con los administradores locales de un programa nacional de transferencias monetarias. Fueron elegidas para ser facilitadoras de la adaptación del modelo jamaicano en Colombia porque, en palabras de los investigadores, eran "influyentes y tenían muchos contactos en sus comunidades". Se trabajó con personal de la comunidad para que la intervención fuera escalable y potencialmente sostenible más allá de la vida del proyecto. Cerca de 720 familias recibieron visitas domiciliarias durante 18 meses. La evaluación aleatoria de este piloto encontró impactos positivos en el desarrollo infantil a corto plazo, de alrededor de 0,18 desviaciones estándar (Attanasio et al. 2014). Sin embargo, no se encontraron impactos significativos 2 años después de la finalización del programa (Andrew et al. 2018).

El programa Cuna Más se basó en los experimentos de Jamaica y, sobre todo, de Colombia.⁷ Se centró en familias con niños de 0 a 36 meses y estaba focalizado en distritos con altos niveles de pobreza y desnutrición crónica infantil, en población principalmente rural.⁸ En estos distritos también funcionaba otro programa asistencial (Juntos, un programa de transferencias monetarias). La implementación de cada programa era completamente independiente y no hubo ningún intento de coordinarlos de forma explícita (más allá del hecho de que estaban dirigidos a los mismos distritos). Juntos se creó mucho antes que Cuna Más y la proporción de beneficiarios de Juntos en los grupos de tratamiento y control era muy similar.⁹

⁷ Véase Araujo, Rubio-Codina y Schady (2021) para un análisis de las diferencias de diseño e implementación entre el ensayo de eficacia de Jamaica, la intervención piloto de Colombia y el programa a escala Cuna Más.

⁸ Para ser elegibles, los distritos tenían que tener una tasa de pobreza del 50% o más, una tasa de desnutrición crónica en niños de 0 a 5 años del 30% o más, y más del 50% de la población viviendo en centros poblados rurales. (Perú define los centros poblados rurales como aquellos con una población máxima de 2.000 habitantes o menos de 400 viviendas). En estos distritos, Cuna Más sólo interviene en los centros poblados rurales. Cuna Más también ofrece servicios de cuidado infantil en centros en zonas urbanas de bajos ingresos —véase Araujo, Dormal y Schady (2019). Sin embargo, no hay solapamiento en la cobertura de estos dos servicios: Cuna Más no contaba con centros de cuidado infantil en ninguno de los distritos que conforman la base del análisis en este estudio.

⁹ En la línea de base, el 52% de los hogares del grupo de tratamiento y el 47% de los del grupo de control recibieron transferencias. Si las transferencias monetarias y las visitas domiciliarias son complementarias, parte del impacto de Cuna Más que estimamos podría ser, en principio, resultado de la interacción entre Juntos y Cuna Más. Esto se debe a que, estrictamente hablando, evaluamos el impacto de las visitas domiciliarias en un contexto en el que los hogares ya eran elegibles para un programa de transferencias monetarias. Sin embargo, aportamos varios elementos que sugieren que es poco probable que ésta sea la razón principal de los impactos positivos de Cuna Más. En primer lugar, las evaluaciones experimentales de programas como Juntos indican que, por sí solas, las transferencias monetarias tienen efectos positivos, pero sólo modestos, en el desarrollo infantil (véase Paxson y Schady 2010 sobre Ecuador; y Macours, Schady y Vakis 2012 sobre Nicaragua). En segundo lugar, la regresión del nivel de desarrollo infantil sobre la asignación aleatoria al tratamiento Cuna Más, una variable que toma el valor de 1 si el hogar recibió Juntos en la línea de base, la interacción entre ambas y efectos fijos de trío muestra que el impacto de la interacción es de -0,040 (con un error estándar de 0,063). Por último, exploramos si la asignación

Cuna Más se diseñó para realizar visitas domiciliarias semanales basadas en el juego, de una hora de duración, a cargo de paraprofesionales con bajo nivel educativo. Como en Colombia, las facilitadoras procedían de las mismas comunidades que las familias con las que trabajaban. Eran principalmente mujeres y cada una trabajaba con 10 familias. Las facilitadoras recibían un pago modesto por su trabajo y, para la mayoría de ellas, éste era su único trabajo remunerado. Aunque en principio se les exigía haber completado la escuela secundaria, en la práctica sólo el 44% tenían secundaria completa. Por su parte, a las supervisoras, también mujeres en su mayoría, se les requerían estudios superiores y cada una de ellas trabajaba con 10 facilitadoras. Dos veces al mes, las supervisoras tenían que trabajar individualmente con cada una de las facilitadoras que supervisaban, acompañándolas en sus visitas, observándolas, dándoles retroalimentación y ayudando en la planificación de actividades para cada niño.

Aunque Cuna Más se basó en el programa piloto de Colombia, hubo diferencias importantes entre los dos programas. Cuna Más era realmente una intervención a escala —con todos los desafíos de implementación que ello conlleva. No había madres líderes en Perú, por lo que Cuna Más tuvo que desarrollar sus propios criterios de selección para las facilitadoras.¹⁰ Algunos requerimientos de la intervención se relajaron: la capacitación previa al servicio fue mucho más limitada en Perú, tanto para las facilitadoras —4 días en Perú, 3 semanas en Colombia— como para las supervisoras —9 días en Perú, 6 semanas en Colombia.

Las facilitadoras de ambos países utilizaron un currículo estructurado de actividades apropiadas al nivel de desarrollo de cada niño, muy rico en materiales de juego —principalmente juguetes caseros de bajo costo, bloques, libros ilustrados, rompecabezas y

aleatoria al grupo de tratamiento de Cuna Más aumentó la proporción de familias que recibían Juntos. No encontramos evidencia de que éste fuera el caso. En una regresión logística multinomial con cuatro categorías (recibir transferencias monetarias en la línea de base y en la de seguimiento, no recibir transferencias monetarias en ninguno de los dos periodos, recibir transferencias en la línea de base pero no en la de seguimiento, y recibir transferencias en la línea de seguimiento pero no en la línea de base), los efectos marginales del tratamiento Cuna Más en las cuatro categorías son pequeños e insignificantes. En el caso de pasar de no recibir a recibir las transferencias de Juntos, el efecto marginal asociado al tratamiento de Cuna Mas es de -0,028 (con un error estándar de 0,018).

¹⁰ Las facilitadoras fueron seleccionadas considerando las sugerencias de las comunidades. Específicamente, los líderes de cada comunidad propusieron candidatos que cumplieran los siguientes criterios: tener 21 años o más, saber leer y escribir, tener experiencia previa trabajando con niños, hablar con fluidez las lenguas locales y ser respetados por los demás miembros de la comunidad. Las supervisoras de Cuna Más revisaron las candidatas propuestas y las evaluaron, en algunos casos junto con los líderes de la comunidad. Existía un manual y una lista de chequeo que debían utilizar en este proceso (la lista incluía un examen de salud mental y una evaluación de las habilidades comunicativas e interpersonales). No sabemos hasta qué punto estos procedimientos se aplicaron de forma consistente en la selección de las facilitadoras.

juegos de clasificación y correspondencias, entre otros. Las actividades se organizaban en fichas de actividades, ordenadas por dificultad, en las que se sugería el rango de edad del niño (de entre 1 y 2 meses de edad) en el que sería apropiado utilizarlas. De esta forma se facilitaba el andamiaje; esto es, la práctica de enseñar a un niño algo nuevo partiendo de las cosas que ya sabe. En general, las actividades y los juegos favorecían el desarrollo de las capacidades cognitivas (por ejemplo, la permanencia de los objetos, la causalidad, la resolución de problemas) y del lenguaje (imitación de gestos y vocalización, introducción de nuevas palabras). Para los niños mayores, había actividades adicionales para enseñar conceptos (tamaño, cantidad, color, forma y posición, entre otros) y funciones ejecutivas tempranas (como la atención y la persistencia). Todos los materiales se adaptaron al contexto local y todas las imágenes de los juegos de clasificación y correspondencias, los libros ilustrados y los rompecabezas se ilustraron de forma que reflejaran la diversidad de las culturas andina y amazónica de Perú.¹¹

Las visitas domiciliarias de Cuna Más se basaban en demostración y práctica. Durante la visita, la facilitadora demostraba cómo interactuar con el niño, animaba y orientaba a la persona cuidadora para que participara en juegos y conversaciones con su hijo y a que respondiera a sus vocalizaciones y acciones. La facilitadora demostraba cómo elogiar al niño y celebraba sus esfuerzos y logros como forma de promover su autoestima y su desarrollo socioemocional. Se desaconsejaba el castigo verbal o físico para redirigir los comportamientos del niño. A lo largo de la visita se hacía especial hincapié en escuchar a la persona cuidadora, pedirle su opinión, darle ánimos y elogiar tanto a la persona cuidadora como al niño.

Una visita típica de Cuna Más iniciaba con el saludo y seguimiento al estado de la familia y lavando las manos del niño. A continuación, la facilitadora hacía participar a la persona cuidadora y al niño en una sesión de juego estructurada en tres "momentos" distintos:

- (i) *Vida en familia*, que incluía una actividad basada en el juego que podía integrarse en las rutinas diarias (por ejemplo, "palabras que puedo aprender durante la hora del baño" o "clasificar la ropa por tipo y tamaño mientras lavo la ropa");

¹¹ Para aumentar la apropiación de la intervención por parte de las familias y las facilitadoras, se elaboraron dos conjuntos de materiales independientes, uno por región.

- (ii) *Jugando aprendo*, que introducía un juego o actividad orientada a lograr una tarea cognitiva o motriz específica, a menudo utilizando un juguete o material diseñado para dicha actividad (por ejemplo, un rompecabezas); y
- (iii) *Cuéntame un cuento*, que incluía una actividad de lenguaje, la mayoría de las veces utilizando un libro de imágenes o un material similar con imágenes sobre las que conversar.

La metodología y la estructura de cada "momento" era la misma. Primero, la facilitadora preguntaba cómo habían avanzado la persona cuidadora y el niño con las actividades de la visita anterior; e, idealmente, la persona cuidadora y el niño mostraban su progreso. A continuación, la facilitadora presentaba la nueva actividad de la semana, describiéndola, demostrándola y observando cómo el niño y la persona cuidadora la repetían y la adaptaban. La visita terminaba con una canción, un repaso de las actividades a realizar durante la semana y motivación para que se incorporaran a las rutinas diarias.

3. La evaluación de Cuna Más

A. Diseño de la evaluación

Durante el diseño de Cuna Más, el gobierno de Perú comprendió que, debido a las limitaciones operativas y presupuestarias, no podía llegar a toda la población elegible para el programa de forma inmediata. Aprovechamos la expansión gradual de la intervención para diseñar e implementar una evaluación experimental.

Específicamente, los distritos elegibles para Cuna Más se ordenaron según su índice de pobreza.¹² Luego se seleccionaron sesenta distritos aleatoriamente, de uno en uno, y se emparejaron con dos distritos que tuvieran un nivel de pobreza similar, uno por encima y otro por debajo. De este modo, se construyeron 60 "tríos" de distritos. Dos de los distritos fueron asignados aleatoriamente a la primera fase de visitas domiciliarias de Cuna Más —el grupo de tratamiento— y el tercero a la última fase —el grupo de control. Cuna Más no inició sus operaciones en ninguno de estos distritos antes de la encuesta de

¹² La tasa de pobreza fue calculada por el Instituto Nacional de Estadística e Informática de Perú (INEI). Hubo algunas restricciones adicionales al construir la muestra. Concretamente, el marco de la muestra excluyó los distritos en cualquiera de los siguientes grupos: (a) todos los distritos de las regiones que formaron parte de una implementación piloto de Cuna Más realizada en 2012; (b) distritos en los que el programa predecesor de Cuna Más (llamado Wawa Wasi) había implementado intervenciones rurales de crianza en grupo; (c) regiones con tan pocos distritos elegibles que la implementación era incierta; y (d) distritos que, según los datos del censo más reciente, no tuviesen al menos dos centros poblados rurales, cada uno con un mínimo de 21 niños de 0 a 24 meses de edad.

seguimiento de la evaluación.¹³

En cada uno de los 180 distritos de la muestra de evaluación, seleccionamos los dos centros poblados rurales con el mayor número de niños entre 0 y 24 meses de edad (según el último censo de población). Luego, realizamos un barrido censal para identificar a todos los niños de este rango de edad en los 360 centros poblados de la muestra. De esa lista, se eligieron aleatoriamente 17 niños en cada centro poblado. En los centros poblados en los que había menos de 17 niños, todos fueron incluidos.¹⁴

B. Datos sobre los hogares, las madres y los niños

La información de línea de base se recolectó entre abril y julio de 2013; la implementación del programa en los 120 distritos de tratamiento comenzó en agosto de 2013, con algunas variaciones en la fecha exacta en que cada centro poblado recibió visitas domiciliarias por primera vez. La encuesta de seguimiento se realizó entre mayo y diciembre de 2015. Los cuestionarios de las líneas de base y de seguimiento de los hogares recogieron información sobre las características del hogar, los comportamientos de los padres y el desarrollo de los niños. La persona encuestada era generalmente la madre del niño.

i. Los hogares

Recopilamos información sobre las características básicas de los hogares, incluido el material principal del piso de la vivienda, el número de dormitorios, el número de bienes en el hogar,¹⁵ el número de miembros del hogar, si el hogar tenía agua corriente y si estaba conectado al sistema de alcantarillado (dos variables distintas).

ii. Las madres

La información sobre las madres incluía su edad y los años de escolaridad alcanzados. En la línea de base, se recogió una versión adaptada de la Escala de Depresión del Centro de Estudios Epidemiológicos (CESD-R-10; Radloff 1977; Björgvinsson et al. 2013). Se ha

¹³ Inicialmente, Cuna Más tenía previsto evaluar dos tratamientos: uno que ofrecía visitas domiciliarias semanales y otro que las complementaba con reuniones grupales quincenales. En la práctica, la implementación del segundo tratamiento no se produjo porque exigía una organización y unos recursos adicionales de los que el programa no disponía en su momento. Por ese motivo, fusionamos los dos grupos de tratamiento en uno solo para análisis.

¹⁴ DGSE-MIDIS et al. (2015) presenta más detalles sobre el diseño de la evaluación, la estrategia de muestreo y los cálculos de potencia. Rubio-Codina, Tomé y Araujo (2016) comparan las características de los distritos de esta muestra con la media nacional.

¹⁵ Se recopiló información sobre si un hogar tenía los siguientes bienes: televisión en color, internet, reproductor de DVD, teléfono móvil, equipo de música, ordenador, nevera, horno, lavadora, plancha y batidora. Como resultado, la cantidad de bienes en el hogar puede oscilar entre 0 y 11. En nuestros datos, el 20,1% de los niños viven en hogares sin bienes y el 0,6% viven en hogares que declaran tener los 11 bienes.

encontrado que la depresión materna se asocia negativamente con el desarrollo infantil en muchos contextos (Sohr-Preston y Scaramella 2006), incluso en los países en desarrollo (Walker et al. 2007). También en la línea de base, se administró a las madres el Test de Vocabulario en Imágenes de Peabody (TVIP; Dunn et al. 1986), una prueba de lenguaje receptivo. El nivel de vocabulario de la madre ha mostrado ser un buen predictor del desarrollo del niño, incluso después de controlar por los años de escolaridad. Este patrón ha sido documentado en muchos estudios, incluso en Ecuador, un país vecino a Perú (Schady 2011).

iii. Los niños

La información sobre los niños incluía su sexo, edad, altura, peso y si el niño vivía con ambos padres. Para medir el desarrollo infantil, se utilizó la tercera edición del Cuestionario de Edades y Etapas (ASQ-3, *Ages and Stages Questionnaire, Third Edition*; Squires et al. 2009). El ASQ-3 mide el desarrollo en cinco dimensiones: resolución de problemas (o cognición), comunicación (o lenguaje), motricidad fina, motricidad gruesa y desarrollo personal-social. Esta herramienta se ha aplicado ampliamente en América Latina, incluso en trabajos anteriores sobre la calidad de los servicios de cuidado infantil de Cuna Más (Araujo, Dormal y Schady, 2019). En el Apéndice A se presentan más detalles sobre el ASQ-3 y su administración.

En las líneas de base y de seguimiento, se preguntó a los encuestados sobre la frecuencia con la que ellos u otra persona del hogar participaban en diversas prácticas de crianza con el niño en cuestión. Estas preguntas se extrajeron de subescalas seleccionadas de los Indicadores de Cuidado Familiar (FCI, *Family Care Indicators*; Kariger et al. 2012) de UNICEF y del cuestionario de Observación del Ambiente del Hogar (HOME, *Home Observation of the Home Environment*; Caldwell y Bradley 2003). Agrupamos estos comportamientos en cuatro categorías: (i) actividades de juego, que incluyen leer o jugar con los niños; (ii) castigos físicos severos, que incluyen golpear, azotar o pellizcar al niño; (iii) castigos verbales severos, que incluyen gritar o regañar al niño; (iv) respuestas positivas a los niños, que incluye abrazarlos o elogiarlos. También recogimos datos sobre el número de materiales de juego disponibles para los niños. Por último, durante la línea de seguimiento, preguntamos a los padres de los niños de 36 meses o más si su hijo estaba matriculado en un centro de educación preescolar.

C. Características en la línea de base

La Tabla 1 presenta las estadísticas descriptivas de los 4.685 niños de la muestra y sus familias.¹⁶ En la línea de base, el niño promedio tenía poco más de un año de edad y la mitad de ellos eran niñas. El 86% de los niños de la muestra vivían con ambos padres. Utilizamos las tablas de la Organización Mundial de la Salud (OMS) para calcular la proporción de niños con desnutrición crónica (talla para la edad <2 desviaciones estándar del valor de la población de referencia), con malnutrición (peso para la talla <2 desviaciones estándar) y con sobrepeso (peso para la talla >2 desviaciones estándar). La Tabla 1 muestra que una proporción considerable de los niños, el 37%, sufría desnutrición crónica. En cambio, la proporción de niños con malnutrición era muy baja (el 1% de la muestra), al igual que la proporción de niños con sobrepeso (el 4% de la muestra).

El manual del ASQ-3 proporciona puntos de corte, basados en una población de referencia (normalizada) en los Estados Unidos, para clasificar a los niños en uno de los tres grupos siguientes: niños con "desarrollo típico", aquellos que se encuentran en una "zona de observación" y aquellos que presentan "posibles rezagos en el desarrollo". La Tabla 1 indica que, utilizando estos puntos de corte "externos", el 26% de los niños presentaba un posible rezago en al menos una dimensión del ASQ-3, el 18% en al menos 3 dimensiones y el 3% tenía posibles rezagos en las 5 dimensiones.

La mayoría de los niños, el 91%, había realizado una actividad de juego con un adulto en el hogar en la semana anterior a la encuesta. Esto quizá no sea sorprendente, ya que la lista de actividades de juego es amplia: incluye leer, contar cuentos, cantar, jugar, dibujar o dar un paseo con los niños. El 15% de los niños habían sido sometidos a castigos físicos severos, una proporción considerable si se tiene en cuenta que todos los niños de la muestra tenían 2 años de edad o menos en la línea de base. Al 31% de los niños se les había gritado o regañado. La proporción de niños que habían sido elogiados por su buen comportamiento al menos una vez en la última semana era muy alta, el 95%.

En promedio, las madres tenían 28 años al inicio del estudio y habían completado 7 años de escolaridad. Las madres también tenían un nivel de vocabulario muy bajo. Los

¹⁶ Excluimos a 582 niños (el 11% del número total de niños con datos de línea de base y seguimiento) de la muestra de análisis porque vivían en hogares que no eran monolingües de habla hispana. Los niños de estos hogares aprendían una lengua indígena (quechua, aymara o una de las muchas lenguas que se hablan en la selva amazónica oriental), y en muchos casos ésta era su primera (o única) lengua, así como la de sus madres. El ASQ-3 y el TVIP podrían no tener un desempeño adecuado con niños y adultos en hogares en los que se habla un idioma distinto del español.

desarrolladores de la prueba TVIP proporcionan tablas de conversión para calcular los puntajes normalizados de vocabulario. En la población de referencia para la prueba (en México y Puerto Rico), la puntuación media es 100 y la desviación estándar es 15. La puntuación media del TVIP de las madres incluidas en la muestra de estudio fue de 66, más de 2 desviaciones estándar por debajo de la media de la población de referencia, y también por debajo de los puntajes observados en una población de madres de ingresos bajos en zonas rurales de Ecuador (media de 73; Schady 2011). El 13% de las madres tenía puntuaciones CESD-R-10 superiores o iguales a 10, el punto de corte estándar para la depresión.

El Gráfico 1 muestra los gradientes de desarrollo infantil por educación de la madre. Para ello, y para el resto del análisis, primero convertimos los puntajes brutos del ASQ-3 de cada dimensión en puntajes z después de ajustar la edad del niño por métodos no paramétricos.¹⁷ Los primeros cinco paneles del gráfico corresponden a las dimensiones individuales del ASQ-3 y el último panel corresponde a una suma que pondera por igual las cinco dimensiones. Los valores que se muestran son para niños del grupo de control en la línea de seguimiento. En cada panel, se muestran las barras correspondientes al desarrollo de los niños de madres con diferentes niveles educativos: educación primaria incompleta o menos (30% de las madres del grupo de control); educación primaria completa (27%); educación secundaria incompleta (19%) y educación secundaria completa o más (24%).

El Gráfico 1 muestra claros gradientes en el desarrollo infantil por educación de la madre. Los gradientes más pronunciados se observan en el desarrollo cognitivo: la diferencia entre los hijos de madres con mayor y menor escolaridad es de 0,56 desviaciones estándar. Esta es también la dimensión en la que la mayor proporción de niños del grupo de control, el 41%, presentaba "posibles rezagos en el desarrollo" en la línea de seguimiento. Los gradientes en otras dimensiones son menos pronunciados,

¹⁷ En principio, los puntajes de desarrollo infantil se pueden estandarizar por edad calculando la media y la desviación estándar del puntaje bruto de los niños en un rango de edad acotado (por ejemplo, cada mes de edad), restando la media del puntaje bruto de un niño y dividiendo por la desviación estándar. En la práctica, sin embargo, este procedimiento puede ser sensible a valores extremos si el número de niños en cada mes de edad es pequeño. Rubio-Codina et al. (2016) proponen un método de tres pasos que es menos sensible a los valores extremos. Tal y como se hizo en estudios anteriores sobre Cuna Más (Araujo, Dormal y Schady 2019), aplicamos este método rigurosamente: (1) En primer lugar, para obtener el puntaje medio específico por edad, \hat{f}_i , para cada dominio en el ASQ-3, estimamos regresiones no paramétricas del puntaje bruto del niño, Y_i , sobre la edad en días, X_i : $Y_i=f(X_i) + \varepsilon_i$; (2) a continuación, para obtener la desviación estándar específica por edad del puntaje, estimamos regresiones no paramétricas del cuadrado de los residuos, ε_i , sobre la edad del niño en días, $\varepsilon_i^2 = g(X_i) + \nu_i$, y tomamos la raíz cuadrada de los valores ajustados \hat{g}_i ; (3) finalmente, para calcular el puntaje z ajustado por edad para cada dominio, ZY_i , tomamos el puntaje bruto de cada niño, restamos la media obtenida en (1), y dividimos por la desviación estándar obtenida en (2): $ZY_i = (Y_i - \hat{f}_i) / \sqrt{\hat{g}_i}$.

especialmente en el desarrollo personal-social.¹⁸ Observamos que, por su diseño, las familias de la muestra de evaluación eran muy pobres; es probable que los gradientes socioeconómicos en una muestra de niños representativa a nivel nacional fueran mayores.

En el apéndice B se ofrecen más detalles sobre la muestra de evaluación. En este apéndice, mostramos que las niñas tenían puntajes más altos en el ASQ-3 que los niños, alrededor de 0,20 desviaciones estándar, un resultado que es consistente con otros reportados en la literatura (Bando, López-Boo, y Li 2016). Los niños con un estado nutricional deficiente, en particular los niños con baja talla para su edad, muestran niveles inferiores de desarrollo. El puntaje de TVIP de la madre y la riqueza del hogar (medida por el número de bienes en el hogar) se correlacionan positivamente con los puntajes del ASQ-3. Por otro lado, la depresión materna no predice el desarrollo infantil en nuestra muestra.

El Apéndice B también muestra que las características en la línea de base de los niños tratados y de los del grupo de control están balanceadas. La pérdida de muestra (tasa de atrición) entre la línea de base y la de seguimiento fue baja y es similar en los grupos de tratamiento (8,6% de pérdida) y de control (9,3% de pérdida). La diferencia en las tasas de pérdida entre los niños tratados y los del grupo de control no es significativa (valor P de 0,46), y las características observables de aquéllos que abandonaron el estudio en ambos grupos son estadísticamente indistinguibles.

D. Tasa de participación y dosificación del programa

Una tasa de participación imperfecta es un problema en cualquier programa de la vida real y Cuna Más no ha sido una excepción. El 34% de las familias asignadas aleatoriamente al tratamiento no recibió ninguna visita. (Tampoco hubo contaminación del grupo de control). Además, entre los hogares que recibieron al menos una visita, la mayoría no recibió el número total de visitas previstas. En otras palabras, tanto la participación como la dosificación fueron inferiores a lo previsto.

El panel A del Gráfico 2 muestra la densidad del número de visitas domiciliarias recibidas por las familias evaluadas del grupo de tratamiento, limitando la muestra a las familias que recibieron al menos una visita. El promedio de visitas recibidas fue de 64, y las familias situadas en los percentiles 10, 25, 75 y 90 de la distribución recibieron 27, 44, 85 y 99 visitas respectivamente. El panel B se centra en la proporción de visitas recibidas con

¹⁸ En todos los paneles del Gráfico 1, la diferencia entre los hijos de madres con niveles de educación más altos y más bajos es significativa al nivel del 95%.

respecto al número de visitas que debería haber recibido una familia, lo que denominamos dosis efectiva.¹⁹ El gráfico, que incluye solo a las familias que recibieron al menos una visita, muestra que, para algo menos de la mitad de las familias, la dosis efectiva se situó entre el 70% y el 100%. Sin embargo, muchas familias recibieron muy pocas visitas: una de cada diez recibió el 30% o menos de las visitas que deberían haber recibido.

No disponemos de datos para establecer formalmente por qué las familias recibieron menos visitas de las previstas. Nuestras propias visitas sobre el terreno y las conversaciones que mantuvimos con los administradores del programa, los supervisores y las facilitadoras sugieren que hubo una variedad de factores en juego. Debido a las dificultades de suministro, las facilitadoras no siempre tenían el material (como libros y rompecabezas) que debían tener. La rotación del personal de Cuna Más fue elevada: un 50% al año, en promedio, entre facilitadoras y supervisoras. Esto provocó interrupciones frecuentes en las visitas y, en algunos casos, el abandono de las familias del programa.

Para comprender mejor las diferencias en participación y dosis efectiva recibida, analizamos si existen características específicas de los niños, los hogares y los centros poblados que predigan sistemáticamente el número de visitas domiciliarias. Para ello, limitamos la muestra a los niños asignados aleatoriamente al grupo de tratamiento y analizamos la relación entre diferentes covariables y la probabilidad de que un hogar reciba al menos una visita; el número de visitas recibidas (incluidos los ceros); y, para aquellos que recibieron al menos una visita, la dosis efectiva.²⁰ Para facilitar la interpretación, convertimos las variables continuas en variables dicotómicas.²¹ Reportamos los resultados de tres especificaciones: la primera incluye cada característica de una en una, la segunda añade efectos fijos a nivel distrito y la tercera incluye todas las características a la vez, así

¹⁹ Para este cálculo, primero establecemos la fecha en la que la primera familia de un distrito tratado recibió su primera visita domiciliaria, la fecha en la que se entrevistó a la primera familia de ese distrito en la encuesta de seguimiento y la fecha en la que un niño cumplió 36 meses de edad (que es la fecha en la que dejó de ser elegible para las visitas domiciliarias). El número previsto de visitas domiciliarias para cada niño viene dado por la diferencia entre la primera visita en ese distrito y la fecha que ocurriese primero entre las dos siguientes: la fecha de inicio de la encuesta de seguimiento en el distrito o la fecha en que el niño cumplió 36 meses de edad. Calculamos la dosis efectiva como la proporción entre el número de visitas que una familia recibió y el número de visitas previstas.

²⁰ Todas las características se refieren a los datos recogidos en la línea de base o a aquéllos que son invariables en el tiempo (sexo del niño, distancia a la capital del distrito). Una excepción es la edad de los niños, donde utilizamos la edad de los niños en la línea de seguimiento, ya que esto es lo que determina si un niño seguía siendo elegible para recibir visitas domiciliarias.

²¹ En general, dividimos los datos en valores por encima o por debajo de la media de la variable correspondiente. Dos excepciones son los años de escolaridad de la madre, donde comparamos a los hijos de las madres que tenían al menos estudios de secundaria completa con los demás niños, y la depresión materna, donde construimos una variable que toma el valor de 1 para las madres con valores de CESD-R-10 superiores o iguales a 10, el punto de corte estándar para la depresión autodeclarada (según la población de referencia).

como efectos fijos de distrito.

Los resultados se presentan en la Tabla 2. El resultado más importante de la tabla es que los niños en peor situación económica tenían más probabilidades de recibir visitas que aquéllos en mejor situación económica. Los niños con desnutrición crónica y los que tenían puntajes ASQ-3 inferiores a la media tenían alrededor de 6 puntos porcentuales más de probabilidad de recibir al menos una visita. El vocabulario materno y el número de bienes en el hogar se asociaron a un menor número de visitas. Lo más llamativo es que los hijos de madres con educación secundaria completa tenían una probabilidad 12 puntos porcentuales menor de recibir al menos una visita domiciliaria que los demás niños; en promedio, recibieron 7 visitas menos. La última fila de la tabla muestra que las familias de los centros poblados situados más lejos de la capital del distrito tenían más probabilidad de recibir visitas domiciliarias.²²

No sabemos por qué los hogares más favorecidos recibieron menos visitas. Es posible que las facilitadoras hicieran un esfuerzo especial para llegar a los hogares más pobres, o que los hogares más favorecidos fueran más propensos a rechazar el programa. También es posible que fuera más fácil contratar o retener a las facilitadoras en los centros poblados más pobres (quizás porque el costo de oportunidad de trabajar como facilitadora era menor).²³ En cualquier caso, los resultados de la Tabla 2 muestran que las visitas domiciliarias fueron una forma eficaz de llegar a muchos de los hogares más pobres, incluso dentro de comunidades generalmente muy pobres. Esto es alentador, ya que, en muchos otros contextos, suele ocurrir que los hogares más favorecidos son los primeros en recibir un programa o una prestación.

²² La información sobre el tiempo de viaje a la capital del distrito fue proporcionada por un líder de la comunidad, en una encuesta aparte. Las regresiones de la última fila de la Tabla 2 se basan en el tiempo de viaje, en minutos. El encuestado también indicó si el viaje a la capital del distrito se hacía generalmente a pie, en carro, en autobús, en moto o en barco (por agua). Obtenemos resultados muy similares a los de la Tabla 2 si calculamos por separado el tiempo de viaje superior e inferior a la media para cada modo de transporte, y luego dividimos la muestra en consecuencia.

²³ Controlamos por efectos fijos de distrito porque la asignación aleatoria se produjo a nivel de distrito. Sin embargo, para las regresiones de la Tabla 2, también podemos incluir efectos fijos a nivel de centro poblado, en lugar de distrito. Cuando lo hacemos, la magnitud de los coeficientes de las diferentes mediciones del estatus socioeconómico disminuye sustancialmente. Por ejemplo, en la regresión del número de visitas recibidas, el coeficiente del indicador de madres que tienen al menos educación secundaria completa es de -7,4 (2,2) sin efectos fijos, de -6,8 (1,8) con efectos fijos de distrito y de -4,2 (1,5) con efectos fijos de centro poblado. En el caso de los hogares con bienes superiores a la media, el coeficiente es de -4,1 (2,4) sin efectos fijos, y de -3,6 (2,2) y -0,78 (1,6) en las regresiones con efectos fijos de distrito y de centro poblado, respectivamente. En otras palabras, entre el 44% (en el caso de la educación materna) y el 93% (en el caso de la riqueza) de la variación en el número de visitas recibidas se debe a las diferencias entre centros poblados, en lugar de diferencias al interior de ellos. Las diferencias en el número de visitas entre los centros poblados podrían ocurrir por una serie de razones, incluyendo problemas para atraer o retener a las facilitadoras, u otros desafíos de implementación.

4. Resultados

A. Estrategia empírica

Nuestros resultados sobre los impactos de Cuna Más se basan en regresiones de la siguiente forma:

$$(1) Y_{ihdt} = \alpha_r + \beta_1 T + \beta_2 Y_{ihdt-1} + \beta_3 X_{ihdt-1} + \varepsilon_{ihdt}$$

donde Y es un indicador del desarrollo infantil o de una práctica de crianza, e i indexa los niños, h los hogares, d los distritos, r los tríos y t el tiempo (línea de base o línea de seguimiento). El parámetro de interés es β_1 , que mide si existen diferencias entre los niños que fueron asignados aleatoriamente a los grupos de tratamiento y control.

Presentamos los resultados de dos especificaciones de intención de tratamiento. La especificación (1) incluye sólo los efectos fijos de tríos, α_r ; en la especificación (2), añadimos los controles de la Tabla 2.²⁴ También presentamos los resultados de dos especificaciones de tratamiento sobre los tratados. En la primera de ellas, la especificación (3), el "tratamiento" toma el valor de 1 si una familia recibió al menos una visita domiciliaria, y en la otra, la especificación (4), el "tratamiento" se refiere al número de visitas recibidas. En ambos casos, instrumentamos el tratamiento con la asignación aleatoria al grupo de tratamiento o de control.²⁵

El tamaño de la muestra es de 4.685.²⁶ Reportamos errores estándar ajustados por conglomerados a nivel de distrito, y los valores P correspondientes. Además, para las regresiones OLS, reportamos los valores P ajustados por pruebas de hipótesis múltiples utilizando el procedimiento de reducción de Romano y Wolf (2005), los valores P basados en la inferencia aleatorizada (Rosenbaum 2002), y los resultados de una prueba de significancia conjunta basada en la aleatorización, conforme a Young (2019).²⁷

²⁴ Parametrizamos los controles como variables dicotómicas, como en la Tabla 2, y según lo recomendado en Athey e Imbens (2012), con una excepción: incluimos el valor total del ASQ-3 (o de cada dimensión individual) en la línea de base como un polinomio de tercer orden del puntaje continuo. Sin embargo, los resultados son muy similares si, en cambio, incluimos un indicador con valor de 1 para aquéllos con puntajes del ASQ-3 superiores a la media, como hacemos con los demás controles.

²⁵ La primera etapa de estas regresiones es muy significativa: un coeficiente de 0,67, con un error estándar de 0,01, cuando el tratamiento se define como recibir al menos una visita, y de 4,31, con un error estándar de 0,09, cuando el tratamiento se define como el número de visitas que recibe una familia.

²⁶ Faltan datos sobre una o más de las variables de la Tabla 1 para el 14,2% de los niños de la muestra. En estos casos, sustituimos el valor que falta por la mediana de la variable en la muestra e incluimos variables dicotómicas para los niños a los que les faltan estas variables.

²⁷ Todos ellos están disponibles como archivos ADO en el programa Stata. Para la corrección Romano-Wolf, véase el comando `rwolf`, programado por Clarke (2016); para la inferencia basada en la aleatorización, véase el comando `ritest`, programado por Hess (2019); y para la prueba de insignificancia conjunta, véase `randcmd`, programado por Young (2020).

B. Impacto del programa en el desarrollo infantil

La Tabla 3 muestra las estimaciones de los impactos de Cuna Más sobre el desarrollo infantil, basadas en (1). Centramos la discusión de resultados en la especificación sin controles, pero notamos que los resultados de las regresiones OLS con controles tienden a ser más precisos.

Considerando todas las dimensiones del ASQ-3, los resultados de la intención de tratamiento indican que Cuna Más mejoró el desarrollo infantil en 0,10 desviaciones estándar. Las estimaciones de tratamiento sobre tratados de la columna (3) muestran que los niños de las familias que recibieron al menos una visita tuvieron puntajes de desarrollo infantil que fueron 0,15 desviaciones estándar más altos; mientras que las estimaciones de la columna (4) muestran que, por cada 10 visitas recibidas, el desarrollo infantil mejoró en 0,023 desviaciones estándar. En la especificación OLS sin controles, los valores P son de 0,049 cuando ajustamos los errores estándar por conglomerados a nivel de distrito y de 0,122 cuando corregimos por inferencia aleatorizada.

El resto de las filas en la tabla indican que Cuna Más tuvo efectos positivos en cognición y lenguaje, pero no en las otras dimensiones. Esto coincide con los resultados de otras evaluaciones de impacto de programas piloto a pequeña escala de visitas domiciliarias en países en desarrollo (véase la revisión de Grantham-McGregor y Smith 2016). Los impactos de intención de tratamiento son de 0,10 desviaciones estándar para el desarrollo cognitivo y de 0,11 desviaciones estándar para el lenguaje. Los impactos son significativos o casi significativos incluso cuando se corrigen los valores P con pruebas de hipótesis múltiples e inferencia aleatorizada, con valores P que van de 0,011 a 0,062. El valor P en la prueba de total irrelevancia, según Young (2019), es de 0,056. Las regresiones de variables instrumentales en la columna (4) de la tabla muestran que por cada 10 visitas domiciliarias se mejoró el desarrollo cognitivo en 0,022 desviaciones estándar y el desarrollo del lenguaje en 0,025 desviaciones estándar.

Un elemento para considerar de los resultados de la Tabla 3 es que el ASQ-3 recoge algunos datos por observación directa y otros a través de la información proporcionada por la madre. Esto plantea la posibilidad de que el impacto de Cuna Más que estimamos pudiera estar motivado, en parte, por un sesgo de información o de deseabilidad —las madres que recibieron visitas en el hogar podrían ser más propensas a informar de que su hijo puede realizar una tarea (como decir una palabra determinada) que las del grupo de control, incluso cuando no haya diferencias reales en los resultados de los

niños. Para ver si esto era un problema real, aprovechamos el hecho de que la dimensión de lenguaje del ASQ-3 incluye algunos ítems que se recogieron mediante reporte materno y otros mediante observación directa. Específicamente, construimos dos índices agregados de lenguaje por separado —uno para los ítems recolectados mediante reporte de la madre y el otro para los ítems observados directamente— y realizamos regresiones por separado para cada uno de ellos. Los resultados de estas regresiones se presentan en la Tabla 4. La tabla muestra que, si acaso, los impactos estimados de las visitas domiciliarias son mayores en los ítems recogidos por observación directa (0,13 desviaciones estándar) que en los reportados por la madre (0,07 desviaciones estándar). Consideramos que esto es una prueba contundente de que los sesgos de información o deseabilidad no explican los impactos estimados de Cuna Más en el desarrollo infantil, al menos en el desarrollo de lenguaje.

A continuación, estudiamos si existe heterogeneidad en los efectos del programa, complementando la ecuación (1) con la interacción entre una determinada covariable y el indicador de tratamiento. Incluimos una interacción a la vez. La Tabla 5 presenta los coeficientes del tratamiento, la covariable en cuestión (por ejemplo, si el niño tenía más de 36 meses en la línea de seguimiento o el sexo del niño) y la interacción correspondiente. La tabla muestra que no hay evidencia de heterogeneidad significativa en ninguna de las dimensiones que medimos, aunque los errores estándar en los impactos de la interacción son generalmente bastante grandes.

C. Impacto del programa en las prácticas de crianza

Analizamos los impactos de Cuna Más en las prácticas de crianza. Para ello, utilizamos las respuestas de las personas cuidadoras a las preguntas sobre actividades de juego, castigos físicos severos, castigos verbales severos y crianza positiva.²⁸ Todas las respuestas se refieren a la semana anterior a la encuesta. Observamos que los comportamientos de crianza que medimos están estrechamente ligados al desarrollo infantil en muchos contextos.²⁹ También disponemos de datos sobre el número de

²⁸ Hay siete preguntas sobre actividades de juego (que incluyen, por ejemplo, si alguien le leyó al niño o jugó con él), tres preguntas sobre castigos físicos severos (que incluyen, por ejemplo, si el niño fue azotado o golpeado con un objeto sólido como un cinturón), dos preguntas sobre castigos verbales severos (que incluyen si alguien gritó o insultó al niño) y cuatro preguntas sobre crianza positiva (que incluyen, por ejemplo, si la persona cuidadora elogió o abrazó al niño).

²⁹ Las referencias incluyen a Bradley y Corwin (2005), Britto et al. (2017), Engle et al. (2011), Galasso, Weber y Fernald (2019), Hamadami et al. (2010) y Hamadami et al. (2014).

materiales de juego (por ejemplo, libros para colorear o juguetes para construcción) disponibles para el niño en casa.

Parametrizamos las prácticas de crianza de dos maneras. Primero, creamos variables dicotómicas que toman el valor de 1 si la persona cuidadora indicó que ocurrió al menos una de las acciones de una categoría determinada —por ejemplo, en el caso de los castigos físicos severos, si el niño fue golpeado, o azotado o pellizcado al menos una vez en la última semana. En segundo lugar, creamos variables que corresponden al número de acciones dentro de una categoría que se llevaron a cabo en la última semana; por ejemplo, si el niño fue golpeado y azotado en la última semana la variable obtiene un valor de 2, mientras que si el niño sólo fue azotado la variable obtiene un valor de 1. Por simplicidad, nos referimos a estas dos variables como los márgenes *extensivos* e *intensivos*, respectivamente, de las prácticas de crianza.

La Tabla 6 muestra que se produjeron cambios sustanciales en algunos comportamientos de crianza, especialmente en la forma de disciplinar a los niños. Los niños del grupo de tratamiento tenían menos probabilidades de estar expuestos a castigos físicos severos (una reducción de 7 puntos porcentuales, de una proporción del 35% en el grupo de control) o a castigos verbales severos (una reducción de 6 puntos porcentuales, de una proporción del 72% en el grupo de control).³⁰ En las actividades de juego, también hubo cambios en el margen intensivo. Todos estos impactos son significativos incluso después de tener en cuenta las pruebas de hipótesis múltiples (valores $P < 0,01$) o la inferencia aleatorizada (valores $P < 0,01$). Por otro lado, no observamos incrementos en la crianza positiva o en el número de materiales de juego en el hogar.

Por último, la Tabla 6 muestra que, entre los niños de 3 años o más, los que fueron asignados aleatoriamente al grupo de tratamiento tenían 8 puntos porcentuales más de probabilidad de ser matriculados en un centro de educación preescolar, partiendo de una proporción del 50% en el grupo de control. Se trata de un impacto considerable y puede ser especialmente importante si se tiene en cuenta la literatura tanto de los países desarrollados como de los que están en vías de desarrollo, incluyendo los de América Latina, que muestra que asistir a la educación preescolar tiene impactos sustanciales y

³⁰ No hay aumentos en el margen extensivo en las actividades de juego o en las prácticas positivas de crianza, pero observamos que los niveles en el grupo de control en estos comportamientos son muy altos: 90% y 93%, respectivamente.

causales en puntuaciones de exámenes y otros resultados, incluso en la edad adulta.³¹

D. Relación beneficio-costo

¿Debemos considerar los impactos de Cuna Más en el desarrollo infantil como "pequeños" o "grandes"? Como primera aproximación a esta pregunta, comparamos la magnitud de los impactos de Cuna Más con las asociaciones entre la educación materna y el desarrollo infantil de los niños del grupo de control en la línea de seguimiento. En una regresión simple del puntaje del ASQ-3 sobre los años de escolaridad alcanzados por las madres, sin ningún control, el coeficiente es de 0,06 (con un error estándar de 0,01). Según esta estimación, los impactos del tratamiento sobre los tratados equivalen a un aumento de 2,5 años de escolaridad de las madres, en una población en la que la escolaridad media era de 7 años.

A continuación, calculamos la relación beneficio-costo de Cuna Más y comparamos estas estimaciones con las del programa piloto en Colombia.³² Para calcular los beneficios del programa, procedemos de la siguiente manera. En primer lugar, observamos que los niños de Colombia recibieron visitas durante 18 meses, en promedio, mientras que los de Perú fueron elegibles durante 24 meses. También observamos que en Colombia prácticamente todos los niños (97,2%) asignados al tratamiento recibieron al menos una visita. Como se ha comentado anteriormente, ese no fue el caso en Perú. Asumimos que los costos del programa fueron aproximadamente cero para las familias que no recibieron ninguna visita en Perú, por lo que en Perú utilizamos las estimaciones del tratamiento de los tratados como impacto del programa. Por último, notamos que, en Colombia, el indicador de desarrollo infantil sólo incluía el desarrollo cognitivo, el lenguaje y la motricidad fina. Por lo tanto, para que los resultados sean comparables entre los dos países, construimos un índice agregado que sólo considera estas tres dimensiones del ASQ-3. El impacto estimado de Cuna Más sobre este indicador, sin covariables, es de 0,19 desviaciones estándar.

Para monetizar estos impactos del programa, recurrimos a los datos de Jamaica. El impacto de las visitas domiciliarias en un indicador global de desarrollo infantil después de

³¹ Revisiones de la literatura sobre la educación preescolar en Estados Unidos incluyen a Cascio (2009), Cascio y Schanzenbach (2013), Cunha y Heckman (2010), Duncan y Magnusson (2013) y Ludwig y Phillips (2007). Sobre América Latina, véase Berlinski, Galiani y Gertler (2009), Berlinski, Galiani y Manacorda (2008), y Bastos, Botton y Cristia (2017).

³² No podemos calcular la relación beneficio-costo del ensayo de eficacia de Jamaica ya que desconocemos los costos de las visitas domiciliarias allí.

24 meses en Jamaica fue de 0,88 desviaciones estándar. En Perú, el impacto comparable fue de 0,19 desviaciones estándar y en Colombia, tras ajustar por el hecho de que el programa piloto sólo duró 18 meses, fue de 0,24 desviaciones estándar. Por tanto, los impactos a corto plazo de las visitas domiciliarias en Colombia y Perú fueron del 27% (0,24/0,88) y el 21% (0,19/0,88), respectivamente, de los observados en Jamaica. En Jamaica, los impactos a corto plazo sobre el desarrollo infantil supusieron un aumento del 25% en los ingresos laborales. Por lo tanto, estimamos que, para quienes tengan trabajo, el aumento previsto de los ingresos laborales sería del 6,8% en Colombia (0,27*0,25) y del 5,3% (0,21*0,25) en Perú.³³

Utilizamos las encuestas laborales de Colombia y Perú para calcular el Valor Presente Neto (VPN) del incremento esperado de los ingresos futuros de los beneficiarios de las visitas domiciliarias.³⁴ Al realizar estos cálculos, asumimos que las personas pueden obtener ingresos laborales cuando tienen entre 25 y 65 años de edad. De forma conservadora, también asumimos que los beneficios de la intervención aplican sólo a quienes trabajan, por lo que ajustamos el aumento de los ingresos laborales por la tasa de participación en cada franja de edad. Utilizamos una tasa de descuento del 3%, y estimamos que el VPN del aumento de los ingresos a lo largo de la vida por participante es de 3.418 dólares estadounidenses en Colombia y de 3.210 en Perú. Por último, utilizamos los datos sobre los costos anuales por niño, que se estiman en 500 dólares en Colombia y 300 en Perú. Combinando los datos sobre beneficios y costos, obtenemos una relación beneficio-costos de 5,4 en Perú y de 4,6 en Colombia.

5. Conclusión

Las intervenciones en la primera infancia pueden tener un gran impacto, pero, con la excepción de la abundante literatura sobre el programa *Head Start* en Estados Unidos, la

³³ El incremento previsto de los ingresos laborales en Colombia y Perú debe considerarse meramente indicativo, por muchas razones. En primer lugar, el instrumento para medir desarrollo infantil no fue el mismo en todos los países (se usaron las escalas de Desarrollo Mental de Griffiths en Jamaica, la Escala Bayley del Desarrollo Infantil —escala Bayley III— en Colombia, y el ASQ-3 en Perú). En segundo lugar, nuestras estimaciones presuponen que, si el impacto de un programa sobre el desarrollo infantil en el país A (Colombia o Perú) es una proporción x del impacto de un programa similar en el país B (Jamaica), entonces el impacto previsto sobre los salarios en el país A también será una proporción x del impacto sobre los salarios observado en el país B. En la práctica, la relación entre los cambios en el desarrollo y los cambios en los salarios puede ser menos directa; por ejemplo, podría ser no lineal, o podría depender de otros factores que varían según el país, como la disponibilidad y la calidad de los servicios educativos.

³⁴ Se utilizaron las encuestas nacionales más cercanas a la fecha de la evaluación: en Colombia, la Gran Encuesta Integrada de Hogares (GEIH) de 2011, y en Perú, la Encuesta Nacional de Hogares (ENAH) de 2018.

mayoría de la evidencia proviene de ensayos de eficacia o de programas piloto implementados minuciosamente. Como es bien sabido, hay una serie de razones por las que los impactos encontrados en estos programas piloto pueden no traducirse en impactos comparables cuando las intervenciones se llevan a escala (List, Suskind y Supplee, 2021).

En este estudio, utilizamos un ensayo controlado aleatorio para estimar los impactos de un programa de visitas domiciliarias que, en el momento de la evaluación, abarcaba a 67.000 familias. Mostramos que los niños que recibieron visitas domiciliarias tuvieron resultados de desarrollo que fueron 0,15 desviaciones estándar más altos que los que no las recibieron. La relación beneficio-costos estimada del programa es considerable: 5,4. Los niños asignados aleatoriamente al tratamiento también tenían menos probabilidades de estar expuestos a castigos físicos o verbales severos, y tenían mayor probabilidad de ser matriculados en un centro de educación preescolar al graduarse de Cuna Más. Enfatizamos que estos impactos positivos responden a un programa que se puso en marcha desde cero y que se escaló en un periodo de tiempo muy corto, con todas las dificultades de implementación que ello conlleva.

¿Serán duraderos los impactos de Cuna Más? Por supuesto, es demasiado pronto para saberlo. Los resultados de Jamaica, así como los de los programas piloto implementados a pequeña escala en Estados Unidos, son sorprendentes.³⁵ Por otro lado, los resultados del programa de visitas domiciliarias en Colombia, que tuvo impactos a corto plazo más similares en magnitud a los que estimamos en Perú, son menos alentadores. Nuestros resultados señalan la importancia de dar seguimiento a niños como los de la evaluación de Cuna Más a lo largo de su trayectoria escolar y, de ser posible, hasta la edad adulta.

³⁵ Véase Heckman et al. (2010) sobre el programa preescolar Perry, y García et al. (2020) sobre el proyecto Abecedarian de Carolina (*Carolina Abecedarian Project*).

Bibliografía

- Aboud, Frances E., y Aisha K. Yousafzai. 2015. "Global Health and Development in Early Childhood". *Annual Review Psychology* 66: 433-457.
- Almond, Douglas, y Janet Currie. 2011. "Human Capital Development before Age Five." En *Handbook of Labor Economics* (Vol. 4, pp. 1315-1486). Elsevier.
- Almond, Douglas, Janet Currie, y Valentina Duque. 2018. "Childhood Circumstances and Adult Outcomes: Act II." *Journal of Economic Literature* 56(4): 1360-1444.
- Al-Ubaydli, Omar, John A. List, y Dana J. Suskind. 2017. "What Can We Learn from Experiments? Understanding the Threats to the Scalability of Experimental Results." *American Economic Review* 107(5): 282-86.
- Andrew, Alison, Orazio Attanasio, Emla Fitzsimons, Sally Grantham-McGregor, Costas Meghir, y Marta Rubio-Codina. 2018. "Impacts 2 Years After a Scalable Early Childhood Development Intervention to Increase Psychosocial Stimulation in the Home: A Follow-up of a Cluster Randomised Controlled Trial in Colombia." *PLoS Medicine* 15(4): e1002556. Disponible en: <https://doi.org/10.1371/journal.pmed.1002556>.
- Araujo, M. Caridad, Marta Dormal, y Norbert Schady. 2019. "Child Care Quality and Child Development." *Journal of Human Resources* 54(3): 656-82.
- Athey, Susan, y Guido W. Imbens. 2017. "The Econometrics of Randomized Experiments." En: Banerjee, Abhijit, and Esther Duflo, (eds.), *Handbook of Economic Field Experiments* (Vol. 1, pp. 73-140). Elsevier.
- Attanasio, Orazio. 2015. "The Determinants of Human Capital Formation During the Early Years of Life: Theory, Measurement, and Policies." *Journal of the European Economic Association* 13(6): 949-997.
- Attanasio, Orazio, Camila Fernández, Emla Fitzsimons, Sally Grantham-McGregor, Costas Meghir, y Marta Rubio-Codina. 2014. "Using the Infrastructure of a Conditional Cash Transfer Program to Deliver a Scalable Integrated Early Child Development Program in Colombia: Cluster Randomized Controlled Trial". *British Medical Journal* 349: g5785.
- Bando, Rosangela, Florencia López-Bóo, y Xia Li. 2016. "Sex-Differences in Language and Personal-social Skills in Early Childhood." IDB-WP-714. Washington, D.C.: Banco Interamericano de Desarrollo.
- Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukerji, Marc Shotland, y Michael Walton. 2017. "From Proof of Concept to Scalable Policies: Challenges and Solutions, with an Application". *Journal of Economic Perspectives* 31(4): 73-102.
- Bastos, Paulo, Nicolas Bottan, y Julian Cristia. 2017. "Access to Preprimary Education and Progression in Primary School: Evidence from Rural Guatemala." *Economic Development and Cultural Change* 65(3): 521-47.
- Berlinski, Samuel, Sebastian Galiani, y Paul Gertler. 2009. "The Effect of Pre-Primary Education on Primary School Performance." *Journal of Public Economics* 93(1): 219-34.
- Berlinski, Samuel, Sebastian Galiani, y Marco Manacorda. 2008. "Giving Children a Better Start: Preschool Attendance and School-Age Profiles." *Journal of Public Economics* 92(5-6): 1416-40.
- Björgvinsson, Thröstur, Sarah J. Kertz, Joe S. Bigda-Peyton, Katrina L. McCoy, y Idan M. Aderka. 2013. "Psychometric Properties of the CES-D-10 in a Psychiatric Sample." *Assessment* 20(4): 429-36.

- Björkman Nyqvist, Martina, Andrea Guariso, Jakob Svensson, y David Yanagizawa-Drott. 2019. "[Reducing Child Mortality in the Last Mile: Experimental Evidence on Community Health Promoters in Uganda.](#)" *American Economic Journal: Applied Economics* 11(3): 155-92.
- Bradley, Robert, y Robert Corwyn. 2005. "Caring for Children Around the World: A View from HOME." *International Journal of Behavioral Development* 29(6): 468–78.
- Britto, Pia R., Stephen L. Lye, Kerrie Proulx, Aisha K. Yousafzai, Stephen G. Matthews, Tyler Vaivada et al. 2017. "Nurturing Care: Promoting Early Childhood Development." *The Lancet* 389(10064): 91–102.
- Caldwell Bettye M. y Robert H. Bradley. 2001. HOME Administration Manual. Third Edition. University of Arkansas University, Print Design, Little Rock.
- Cascio, Elizabeth. 2009. "Do Investments in Universal Education Pay Off? Long-Term Effects of Introducing Kindergartens into Public Schools." NBER Working Paper 14951.
- Cascio, Elizabeth, y Diane Whitmore Schanzenbach. 2013. "The Impacts of Expending Access to High-Quality Preschool Education." NBER Working Paper 19735.
- Chapman, Donna J., Katherine Morel, Alex Kojo Anderson, Grace Damio, y Rafael Perez-Escamilla. 2010. "Breastfeeding Peer Counselling: From Efficacy to Scale-Up." *Journal of Human Lactation* 26(3): 314-26.
- Clarke, Damien. 2016. "RWOLF: Stata Module to Calculate Romano-Wolf Stepdown p-values for Multiple Hypothesis Testing." Statistical Software Components S458276, Boston College Department of Economics.
- Cunha, Flavio, y James Heckman. 2010. "Investing in Our Young People." NBER Working Paper 16201.
- DGSE-MIDIS, DGPP-MEF, BID. 2015. "Documento de Línea de Base para la Evaluación de Impacto del Programa Nacional Cuna Más – Servicio de Acompañamiento a Familias", Lima, febrero de 2015, Documento del MIDIS. Disponible en: http://www.midis.gob.pe/dgsye/evaluacion/documentos/CUNAMAS-AF_30_de_Enero.pdf. Consultado en marzo 23, 2017.
- Duncan, Greg J., y Katherine Magnuson. 2013. "Investing in Preschool Programs." *Journal of Economic Perspectives* 27(2): 109-32.
- Dunn, Lloyd, Delia Lugo, Eligio Padilla, y Leota Dunn. 1986. *Test de Vocabulario en Imágenes Peabody* (Circle Pines, MN: American Guidance Service).
- Engle, Patrice, Lia C.H. Fernald, Harold Alderman, Jere Behrman, Chloe O’Gara, Aisha Yousafzai et al. 2011. "Strategies for Reducing Inequalities and Improving Developmental Outcomes for Young Children in Low-Income and Middle-Income Countries." *The Lancet* 378(9799): 1339–53.
- Fernald, Lia C.H., Patricia Kariger, Melissa Hidrobo, y Paul J. Gertler. 2012. "Socioeconomic Gradients in Child Development in Very Young Children: Evidence from India, Indonesia, Peru, and Senegal." *Proceedings of the National Academy of Science* 109(2): 17273-80.
- Galasso, Emmanuela, Ann Weber, y Lia C.H. Fernald. 2019. "Dynamics of Child Development: Analysis of a Longitudinal Cohort in a Very Low Income Country." *The World Bank Economic Review* 33(1): 140-59.
- García, Jorge Luis, James J. Heckman, Duncan Ermini Leaf, y María José Prados. 2020. "Quantifying the Life-Cycle Benefits of an Influential Early Childhood Program." *Journal of Political Economy* 128(7): 2502-41.
- Gertler, Paul, James Heckman, Rodrigo Pinto, Adriana Zanolini, Christel Vermeersch, Susan Walker, Susan Chang, y Sally Grantham-McGregor. 2014. "Labor Market

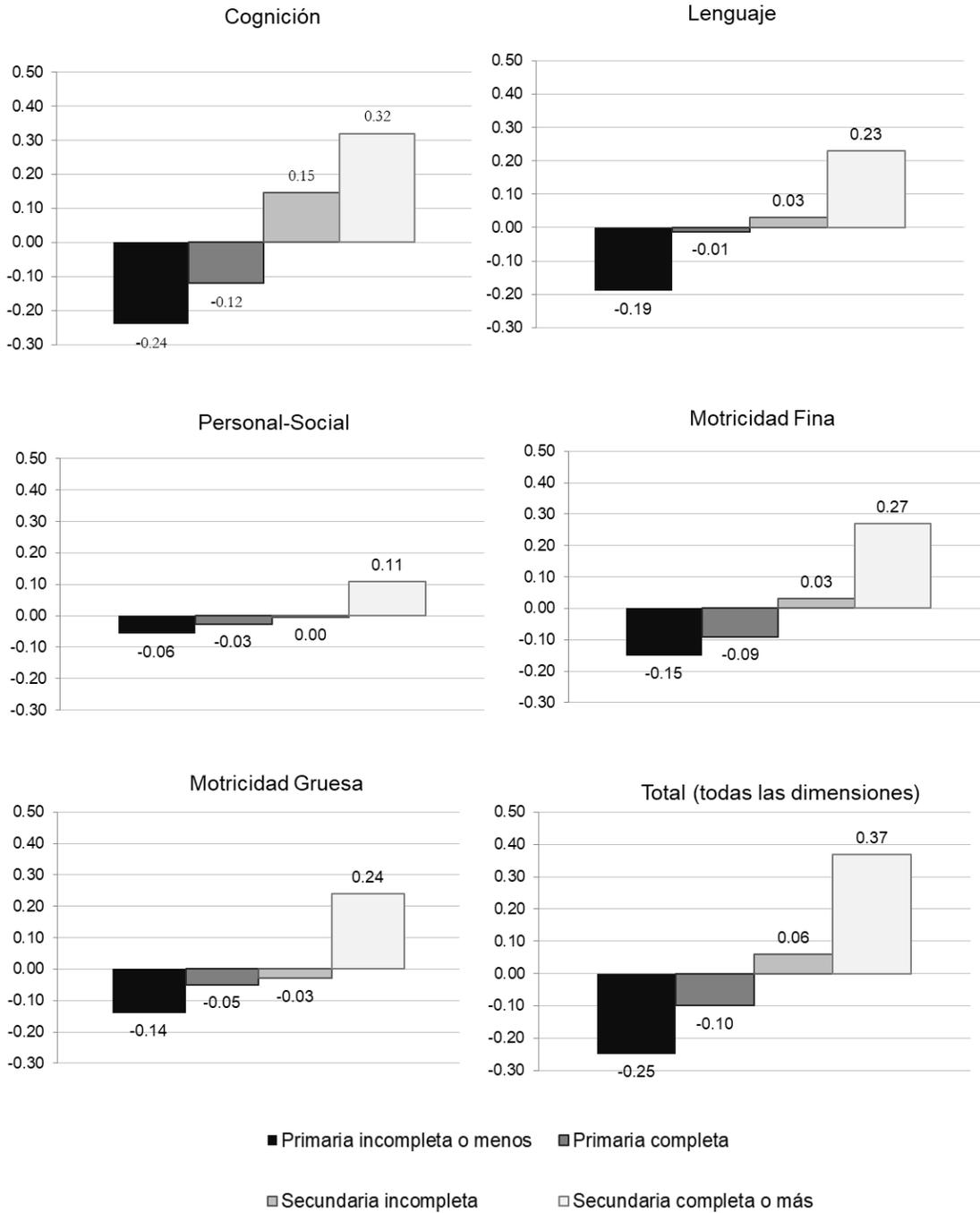
- Returns to an Early Childhood Stimulation Intervention in Jamaica.” *Science* 344(6187): 998-1001.
- Grantham-McGregor, Sally M., Christine A. Powell, Susan P. Walker, y John H. Himes. 1991. “Nutritional Supplementation, Psychosocial Stimulation, and Mental Development of Stunted Children: the Jamaican Study.” *The Lancet* 338(8758): 1-5.
- Grantham-McGregor, Sally, y Joanne A. Smith. 2016. “Extending the Jamaican Early Childhood Development Intervention.” *Journal of Applied Research on Children: Informing Policy for Children at Risk* 7(2): 4.
- Hamadani, Jena, Fahmida Tofail, Afroza Hilaly, Syed N. Huda, Patrice L. Engle, y Sally Grantham-McGregor. (2010). “Use of Family Care Indicators and Their Relationship with Child Development in Bangladesh.” *Journal of Health, Population, and Nutrition* 28(1) 23-33.
- Hamadani, Jena Fahmida Tofail, Syed N. Huda, Dewan S. Alam, Deborah A. Ridout, Orazio Attanasio, y Sally Grantham-McGregor. 2014. “Cognitive Deficit and Poverty in the First 5 Years of Childhood in Bangladesh.” *Pediatrics* 134(4): e1001-e1008.
- Haroon, Sarah, Jai k. Das, Rehana A. Salam, Aamer Imdad, y Zulfiqar A. Bhutta. 2013. “Breastfeeding Promotion Interventions and Breastfeeding Practices: A Systematic Review.” *BMC Public Health* 13(3): 1-18.
- Heckman, James J., y Stefano Mosso. 2014. “The Economics of Human Development and Social Mobility.” *Annual Review of Economics* 6(1): 689-733.
- Heckman, James J, Seong Hyeok Moon, Rodrigo Pinto, Peter A Savelyev, y Adam Yavitz. 2010. “The Rate of Return to the High/Scope Perry Preschool Program.” *Journal of Public Economics* 94(1–2): 114-28.
- Heller, Sara B., Anuj K. Shah, Jonathan Guryan, Jens Ludwig, Sendhil Mullainathan, y Harold A. Pollack. 2017. “Thinking, Fast and Slow? Some Field Experiments to Reduce Crime and Dropout in Chicago.” *Quarterly Journal of Economics* 132(1): 1-54.
- Hess, Simon. 2019. “RITEST: Stata Module to Perform Randomization Inference and Permutation Tests.” Statistical Software Components S458599, Boston College Department of Economics.
- Imbens, Guido, y Donald Rubin. 2015. *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge: Cambridge University Press.
- Imbens, Guido, y Jeffrey Wooldridge. 2009. “Recent Developments in the Econometrics of Program Evaluation.” *Journal of Economic Literature* 47(1): 5-86.
- Kariger, Patricia, Edward A. Frongillo, Patricia Engle, Pia R. Britto, Sara M. Sywulka, y Purmina Menon. 2012. “Indicators of Family Care for Development for Use in Multicountry Surveys.” *Journal of Health, Population and Nutrition* 30(4): 472–486.
- List, John, Dana Suskind, y Lauren Supplee, Eds. (2021). *The Scale-Up Effect in Early Childhood and Public Policy: Why Interventions Lose Impact at Scale and What We Can Do About It*. Routledge.
- Love, John M., Rachel Chazan-Cohen, Helen Raikes, y Jeanne Brooks-Gunn. 2013. “What Makes a Difference: Early Head Start Evaluation Findings in a Developmental Context”. *Monographs of the Society for Research in Child Development* 78(1): vii–viii.
- Ludwig, Jens, y Deborah A. Phillips. 2007. “The Benefits and Costs of Head Start.” NBER Working Paper 12973.
- Macours, Karen, Norbert Schady, y Renos Vakis. 2012. “Cash Transfers, Behavioral Changes, and Cognitive Development in Early Childhood: Evidence from a Randomized Experiment.” *American Economic Journal: Applied Economics* 4(2):

247-73.

- Mayoral, Maria V. 2014. "Fact Sheet: Early Head Start." *Zero to Three National Center for Infants, Toddlers, and Families*. Disponible en: <http://www.buildinitiative.org/Portals/0/Uploads/Documents/Fact%20Sheet%20-%20Early%20Head%20Start%20-%20ZTT.pdf>. Consultado en abril de 2019.
- Olds, David L. 2010. "The Nurse-Family Partnership: From Trials to Practice." En Arthur Reynolds, Arthur Rolnick, Michelle Englund, and Judy Temple, Eds. *Childhood Programs and Practices in the First Decade of Life: A Human Capital Integration* (pp. 49-75). Cambridge: Cambridge University Press.
- Paxson, Christina, y Norbert Schady. 2010. "Does Money Matter? The Effects of Cash Transfers on Child Health and Development in Rural Ecuador." *Economic Development and Cultural Change* 59(1): 187-229.
- Prost, Audrey, Tim Colbourn, Nadine Seward, Kishwar Azad, Arri Coomarasamy, Andrew Copas et al. 2013. "Women's Groups Practising Participatory Learning and Action to Improve Maternal and Newborn Health in Low-Resource Settings: A Systematic Review and Meta-Analysis." *The Lancet* 381(9879): 1736-46.
- Radloff, Lenore Sawyer. 1997. "The CES-D scale: A Self-Report Depression Scale for Research in the General Population." *Applied Psychological Measurement* 1(3): 385-401.
- Rollins, Nigel C., Nita Bhandari, Nemat Hajeighbay, Susan Horton, Chessa Lutter, Jose Martinez et al. "Why Invest, and What it Will Take to Improve Breastfeeding Practices?" *The Lancet* 387(10017): 491-504.
- Romano, Joseph P., y Michael Wolf. 2005. "Stepwise Multiple Testing as Formalized Data Snooping." *Econometrica* 73(4): 1237-82.
- Rosenbaum, Paul R. 2002. *Observational Studies*. 2nd ed. New York: Springer
- Rubio-Codina, Marta, M. Caridad Araujo, Orazio Attanasio, Pablo Muñoz, y Sally Grantham-McGregor. 2016. "Concurrent Validity and Feasibility of Short Tests Currently Used to Measure Early Childhood Development in Large Scale Studies". *PLoS One* 11(8): e0160962.
- Rubio-Codina, Marta, Romina Tomé, y M. Caridad Araujo. 2016. "Una fotografía sobre el bienestar y el desarrollo de los niños del Programa Nacional Cuna Más." IDB-TN-1093. Washington, D.C.: Banco Interamericano de Desarrollo.
- Salam, Rehana, Jai K. Das, Zohra Lassi, y Zulfiqar Bhutta. 2014. "Impact of Community-Based Interventions for the Prevention and Control of Malaria on Intervention Coverage and Health Outcomes for the Prevention and Control of Malaria." *Infectious Diseases and Poverty* 3(25): 1-15.
- Schady, Norbert. 2011. "Parental Education, Vocabulary, and Cognitive Development in Early Childhood: Longitudinal Evidence from Ecuador." *American Journal of Public Health* 101(12): 2299-2307.
- Squires, Jane, Diane Bricker, Elizabeth Twombly, Robert Nickel, Jantina Clifford, Kimberly Murphy, Robert Hoselton, Lawanda Potter, Linda Mounts, y Jane Farrell. 2009. *Ages & Stages English Questionnaires, Third Edition (ASQ-3): A Parent-Completed, Child-Monitoring System*. Baltimore, MD: Paul H. Brookes Publishing Co.
- Sohr-Preston, Sara L., y Laura V. Scaramella. 2006. "Implications of Timing of Maternal Depressive Symptoms for Early Cognitive and Language Development." *Clinical Child and Family Psychology Review* 9(1): 65-83.
- Walker, Susan P., Susan M. Chang, Joanne A. Smith, y Helen Baker-Henningham. 2018. "The Reach Up Early Childhood Parenting Program: Origins, Content, and Implementation." *Zero to Three* 38(4):37-43.

- Walker, Susan P., Susan M. Chang, Marcos Vera-Hernández, y Sally Grantham-McGregor. 2011. "Early Childhood Stimulation Benefits Adult Competence and Reduces Violent Behavior." *Pediatrics* 127(5): 849–57.
- Walker, Susan P., Sally M. Grantham-McGregor, Christine A. Powell, y Susan M. Chang. 2000. "Effects of Growth Restriction in Early Childhood on Growth, IQ, and Cognition at Age 11 to 12 Years and the Benefits of Nutritional Supplementation and Psychosocial Stimulation." *The Journal of Pediatrics* 137(1): 36-41.
- Walker, Susan P., Theodore D. Wachs, Julie Meeks Gardner, Betsy Lozoff, Gail A. Wasserman, Ernesto Pollitt, Julie A. Carter, e International Child Development Steering Group. 2007. "Child Development: Risk Factors for Adverse Outcomes in Developing Countries." *The Lancet* 369(9556): 145-57.
- Young, Alwyn. 2019. "Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results." *Quarterly Journal of Economics* 134(2): 557-98.
- Young, Alwyn. 2020. "RANDCMD: Stata Module to Compute Randomization Inference p-values." Statistical Software Components S458774, Boston College Department of Economics.
- Zwane, Alix Peterson, y Michael Kremer. 2007. "What Works in Fighting Diarrheal Diseases in Developing Countries?" A Critical Review. *World Bank Research Observer* 22(1): 1-24.

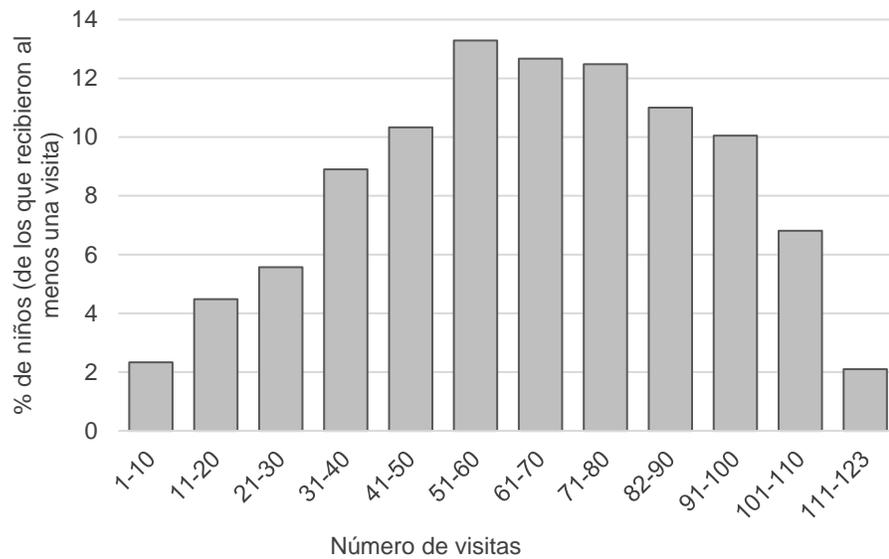
Gráfico 1. Gradientes en el desarrollo infantil por educación de la madre



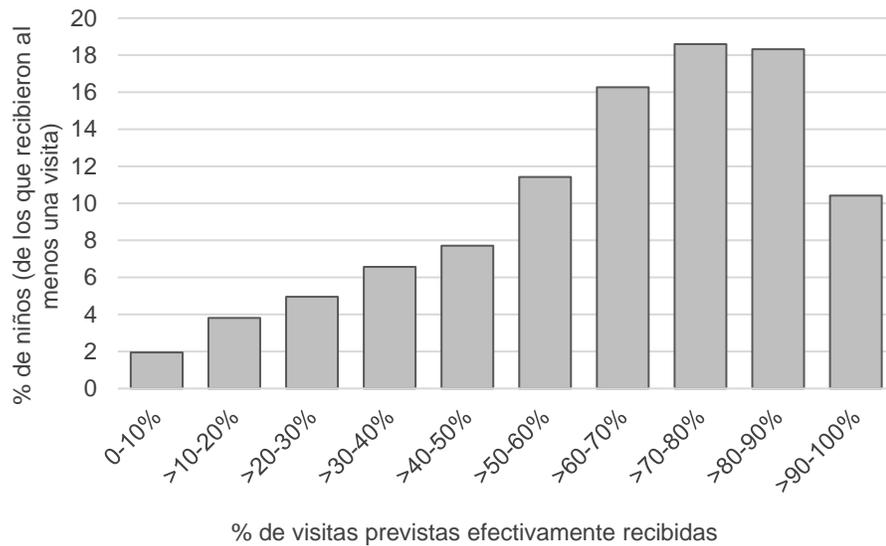
Nota: Dimensiones individuales y puntaje total en el ASQ-3. Sólo se incluyen los niños del grupo de control en la línea de seguimiento. En cada panel, las barras corresponden al rendimiento promedio de los hijos de madres con diferentes niveles educativos: primaria incompleta o menos (30% de las madres del grupo de control); primaria completa (27%); secundaria incompleta (19%); y secundaria completa o superior (24%).

Gráfico 2. Tasa de participación y dosis efectiva

Panel A: Número de visitas domiciliarias



Panel B: Dosis efectiva



Nota: Datos procedentes de registros administrativos y de las encuestas.

Tabla 1. Estadísticas descriptivas de los niños y las familias

	N	Media / Proporción	Desviación Estándar
Niños			
Edad en meses	4.685	12,8	6,6
Proporción de mujeres	4.685	0,49	
Proporción sin rezago en el ASQ-3	4.685	0,18	
Proporción con rezago en al menos 1 dimensión del ASQ-3	4.685	0,26	
Proporción con rezago en al menos 2 dimensiones del ASQ-3	4.685	0,25	
Proporción con rezago en al menos 3 dimensiones del ASQ-3	4.685	0,18	
Proporción con rezago en al menos 4 dimensiones del ASQ-3	4.685	0,10	
Proporción con rezago en las 5 dimensiones del ASQ-3	4.685	0,03	
Puntaje z de talla para la edad	4.559	-1,66	1,07
Proporción con desnutrición crónica (talla para la edad < -2 SD)	4.559	0,37	
Puntaje z del peso para la talla	4.559	0,35	0,92
Proporción con malnutrición (peso para la talla < -2 SD)	4.561	0,01	
Proporción con sobrepeso (peso para la talla > 2 SD)	4.561	0,04	
Proporción que realizaron una actividad de juego o más en la última semana	4.685	0,91	
Proporción que recibieron castigos físicos severos en la última semana	4.685	0,15	
Proporción que recibieron castigos verbales severos en la última semana	4.685	0,31	
Proporción elogiados por buen comportamiento en la última semana	4.685	0,95	
Proporción viviendo con ambos padres	4.685	0,86	
Madres			
Edad	4.685	28,5	7,6
Años de escolaridad	4.453	6,9	4,1
Puntaje TVIP	4.363	65,8	15,3
Proporción de madres con un puntaje CESD-R ≥ 10	4.672	0,13	
Hogares			
Número de miembros	4.685	5,5	2,1
Número de habitaciones	4.685	2,5	1,5
Número de bienes (0-11)	4.685	2,6	2,3
Proporción de hogares con agua corriente	4.685	0,60	
Proporción de hogares conectados al alcantarillado	4.685	0,26	
El material principal del piso es tierra	4.685	0,73	0,44

Nota: Los posibles rezagos en el desarrollo se calcularon siguiendo los puntos de corte previstos en el manual del ASQ-3, y los indicadores de nutrición siguiendo las directrices de la Organización Mundial de la Salud. La depresión materna se midió con una versión adaptada del CESD-R-10 (Radloff 1977; Björgvinsson et al. 2013). TVIP se refiere al Test de Vocabulario en Imágenes Peabody (Dunn, Padilla y Dunn 1986). Los bienes incluyen: televisión en color, internet, reproductor de DVD, teléfono móvil, equipo de música, ordenador, nevera, horno, lavadora, plancha y licuadora.

Tabla 2: Tasa de participación y dosis efectiva

	Al menos una visita			Número de visitas			Visitas efectivas (%)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Características de los niños									
Edad >36 meses	-0,079***	-0,085***	-0,095***	-17,801***	-18,566***	-19,135***	2,636**	1,262	1,375
	(0,022)	(0,020)	(0,020)	(1,843)	(1,752)	(1,800)	(1,198)	(1,099)	(1,161)
Sexo	0,007	0,006	0,009	-1,114	-1,384	-0,973	-0,257	-0,586	-0,441
	(0,019)	(0,016)	(0,017)	(1,407)	(1,195)	(1,197)	(0,990)	(0,926)	(0,984)
Puntaje del ASQ-3 en la línea de base superior a la mediana	-0,058***	-0,041***	-0,033*	-2,392	-2,256*	-1,151	1,066	0,357	0,808
	(0,021)	(0,015)	(0,017)	(1,563)	(1,176)	(1,251)	(1,157)	(0,901)	(0,958)
Desnutrición crónica	0,055**	0,042**	0,052**	1,043	0,336	2,442*	0,676	0,455	-0,280
	(0,024)	(0,019)	(0,020)	(1,782)	(1,364)	(1,377)	(1,146)	(0,942)	(0,995)
Características de las familias									
Educación secundaria completa o superior	-0,120***	-0,094***	-0,057***	-7,391***	-6,830***	-4,017**	-0,320	-1,433	-1,109
	(0,031)	(0,024)	(0,020)	(2,201)	(1,794)	(1,659)	(1,250)	(1,173)	(1,300)
Puntaje TVIP de la madre superior a la mediana	-0,051*	-0,028	-0,009	-4,817**	-3,354**	-1,685	-1,094	-0,423	0,032
	(0,027)	(0,022)	(0,021)	(1,954)	(1,674)	(1,570)	(1,239)	(1,136)	(1,238)
Edad de la madre superior a la mediana	0,023	0,004	-0,015	1,517	0,541	-0,175	1,613	2,009**	1,943*
	(0,019)	(0,017)	(0,018)	(1,454)	(1,292)	(1,406)	(1,223)	(0,983)	(1,049)
Puntaje CESD-R de la madre >=10	0,015	0,003	-0,017	1,918	0,456	-0,622	1,315	0,070	-0,082
	(0,026)	(0,020)	(0,022)	(2,208)	(1,864)	(1,827)	(1,870)	(1,533)	(1,671)
Riqueza del hogar superior a la mediana	-0,059*	-0,036	-0,009	-4,115*	-3,623*	-1,827	-1,117	-0,799	-0,757
	(0,034)	(0,029)	(0,027)	(2,437)	(2,152)	(2,118)	(1,588)	(1,473)	(1,550)
Características de los distritos y los centros poblados									
Distancia a la capital superior a la mediana	0,086*			4,687			0,629		
	(0,051)			(3,560)			(2,025)		
Regresión multivariable	N	N	S	N	N	S	N	N	S
Efectos fijos de distrito	N	S	S	N	S	S	N	S	S

Nota: Sólo los niños del grupo de tratamiento. Las especificaciones (7)-(9) incluyen sólo a los niños con al menos una visita. El N va de 1.825 a 3.192 dependiendo de la covariable de interés y de la especificación. Los errores estándar se ajustan a nivel de distrito (entre paréntesis). *, ** y *** indican los niveles de significancia del 10%, 5% y 1%, respectivamente.

Tabla 3: Impactos de Cuna Más en el desarrollo infantil

	Intención de tratamiento		Tratamiento sobre los tratados	
	(1)	(2)	(3)	(4)
Todas las dimensiones	0,101 (0,051) [0,049] <0,122>	0,109 (0,045) [0,017] <0,063>	0,151 (0,077) [0,052]	0,023 (0,012) [0,052]
Cognición	0,095 (0,040) [0,018] {0,056} <0,051>	0,095 (0,036) [0,008] {0,029} <0,027>	0,142 (0,059) [0,018]	0,022 (0,009) [0,019]
Lenguaje	0,109 (0,043) [0,011] {0,038} <0,049>	0,111 (0,039) [0,005] {0,023} <0,036>	0,162 (0,064) [0,012]	0,025 (0,010) [0,012]
Personal-social	0,068 (0,052) [0,193] {0,363} <0,325>	0,081 (0,051) [0,116] {0,277} <0,250>	0,102 (0,079) [0,200]	0,016 (0,012) [0,201]
Motricidad Fina	0,068 (0,046) [0,141] {0,363} <0,213>	0,068 (0,041) [0,099] {0,277} <0,176>	0,101 (0,069) [0,144]	0,016 (0,011) [0,144]
Motricidad Gruesa	-0,021 (0,045) [0,646] {0,658} <0,715>	-0,012 (0,044) [0,782] {0,788} <0,802>	-0,031 (0,067) [0,648]	-0,005 (0,010) [0,648]
Efectos fijos de trío	X	X	X	X
Controles (características socioeconómicas)		X		
Valor P aleatorizado, prueba múltiple de Westfall-Young de la significancia del tratamiento	0.056	0.037		

Nota: El N es igual a 4.685 en todas las especificaciones. La tabla presenta los coeficientes, los errores estándar (entre paréntesis), los valores p [entre corchetes], los valores p de Romano-Wolf {entre corchetes rizados} y los valores p basados en la inferencia aleatorizada <entre corchetes angulares> de las regresiones de la variable dependiente en la primera columna sobre el tratamiento. Las especificaciones (1) y (2) corresponden a la intención de tratamiento y los resultados se refieren a los niños asignados aleatoriamente al tratamiento. En la especificación (3), el "tratamiento" se define como los niños que recibieron al menos una visita domiciliaria y está instrumentado con la asignación aleatoria al tratamiento. Del mismo modo, en la especificación (4), el "tratamiento" se define como el número total de visitas recibidas/10 y se instrumenta con la asignación aleatoria al tratamiento. Las especificaciones (1), (3) y (4) incluyen sólo los efectos fijos de trío. La especificación (2) añade los siguientes controles de la línea de base del niño y del hogar: edad del niño superior a 36 meses, sexo, un polinomio de tercer orden en el puntaje total del ASQ-3 en la línea de base, desnutrición crónica, educación materna secundaria o superior, edad materna por encima de la mediana, puntaje TVIP materno por encima de la mediana, depresión materna (CESD-R-10>=10), riqueza del hogar por encima de la mediana, distancia a la capital del distrito por encima de la mediana. Todos los errores estándar están ajustados a nivel de distrito.

Tabla 4: Impactos de Cuna Más en el desarrollo del lenguaje, ítems por reporte materno versus ítems por observación directa

	Intención de tratamiento		Tratamiento sobre los tratados	
	(1)	(2)	(3)	(4)
Lenguaje, todos los ítems	0,109 (0,042) [0,011] {0,038} <0,049>	0,111 (0,039) [0,005] {0,023} <0,036>	0,162 (0,064) [0,012]	0,025 (0,010) [0,012]
Lenguaje, ítems por observación directa	0,131 (0,046) [0,005] {0,028} <0,044>	0,129 (0,043) [0,003] {0,014} <0,027>	0,196 (0,069) [0,005]	0,03 (0,011) [0,005]
Lenguaje, ítems por reporte materno	0,065 (0,032) [0,041] {0,139} <0,130>	0,068 (0,031) [0,029] {0,100} <0,108>	0,097 (0,048) [0,045]	0,015 (0,007) [0,045]
Efectos fijos de trío	X	X	X	X
Controles (características socioeconómicas)		X		
Valor P aleatorizado, prueba múltiple de Westfall-Young de la significancia del tratamiento	0,039	0,016		

Nota: El N es igual a 4.685 en todas las especificaciones. La tabla presenta los coeficientes, los errores estándar (entre paréntesis), los valores p [entre corchetes], los valores p de Romano-Wolf {entre corchetes rizados} y los valores p basados en la inferencia de la aleatorización <entre corchetes angulares> de las regresiones de la variable dependiente en la primera columna sobre el tratamiento. Las especificaciones (1) y (2) corresponden a la intención de tratamiento y los resultados se refieren a los niños asignados aleatoriamente al tratamiento. En la especificación (3), el "tratamiento" se define como los niños que recibieron al menos una visita domiciliar y está instrumentado con la asignación aleatoria al tratamiento. Del mismo modo, en la especificación (4), el "tratamiento" se define como el número total de visitas recibidas/10 y se instrumenta con la asignación aleatoria al tratamiento. Las especificaciones (1), (3) y (4) incluyen sólo los efectos fijos de trío. La especificación (2) añade los siguientes controles de la línea de base del niño y del hogar: edad del niño superior a 36 meses, sexo, un polinomio de tercer orden en el puntaje total del ASQ-3 en la línea de base, desnutrición crónica, educación materna secundaria o superior, edad materna por encima de la mediana, puntaje TVIP materno por encima de la mediana, depresión materna (CESD-R-10>=10), riqueza del hogar por encima de la mediana, distancia a la capital del distrito por encima de la mediana. Todos los errores estándar están ajustados a nivel de distrito.

Tabla 5: Heterogeneidad de los impactos de Cuna Más

	Edad >36 meses	Sexo	ASQ-3 en línea de base superior a la mediana	Desnutrición crónica	Educación secundaria completa o superior	TVIP de la madre superior a la mediana	Edad de la madre superior a la mediana	CESD-R de la madre >=10	Riqueza del hogar superior a la mediana	Distancia a la capital del distrito superior a la mediana
Covariable	-0,006 (0,060)	0,180*** (0,049)	0,405*** (0,063)	-0,351*** (0,062)	0,458*** (0,073)	0,303*** (0,056)	0,036 (0,050)	-0,026 (0,067)	0,333*** (0,063)	-0,180** (0,083)
Tratamiento	0,027 (0,073)	0,138** (0,061)	0,122* (0,062)	0,100* (0,055)	0,071 (0,056)	0,096 (0,063)	0,100 (0,064)	0,096* (0,053)	0,054 (0,060)	0,125 (0,081)
Covariable* Tratamiento	0,107 (0,074)	-0,073 (0,060)	-0,001 (0,073)	0,005 (0,076)	0,060 (0,089)	-0,007 (0,071)	0,001 (0,064)	0,044 (0,084)	0,070 (0,081)	-0,041 (0,105)
Efectos fijos de trío	X	X	X	X	X	X	X	X	X	

Nota: El N es igual a 4.685 en todas las especificaciones. La tabla presenta los coeficientes y los errores estándar ajustados a nivel de distrito (entre paréntesis) de las regresiones del puntaje total de ASQ-3 sobre cada covariable, tratamiento y término de interacción. Las especificaciones incluyen los efectos fijos de trío (y ninguna otra covariable), excepto para la distancia a la capital del distrito. *, ** y *** indican los niveles de significancia del 10, 5 y 1 por ciento, respectivamente.

Tabla 6: Impactos de Cuna Más en las prácticas de crianza

	Variable dependiente: la práctica ocurrió al menos una vez				Variable dependiente: variedad de prácticas					
	Proporción (grupo de control)	Intención de tratamiento		Tratamiento sobre los tratados		Promedio, # de ítems, [rango] (grupo de control)	Intención de tratamiento		Tratamiento sobre los tratados	
		(1)	(2)	(3)	(4)		(5)	(6)	(7)	(8)
Actividades de juego	0,9	0,012 (0,018) [0,495] {0,586} <0,555>	0,013 (0,017) [0,447] {0,585} <0,5327>	0,019 (0,027) [0,497]	0,003 (0,004) [0,496]	3,47 [0-7]	0,366 (0,147) [0,014] {0,039} <0,044>	0,379 (0,125) [0,003] {0,009} <0,020>	0,545 (0,219) [0,014]	0,085 (0,034) [0,014]
Maltrato físico severo	0,35	-0,072 (0,018) [0,000] {0,000} <0,001>	-0,072 (0,018) [0,000] {0,000} <0,001>	-0,107 (0,028) [0,000]	-0,017 (0,004) [0,000]	0,43 [0-3]	-0,105 (0,028) [0,000] {0,001} <0,002>	-0,104 (0,028) [0,000] {0,002} <0,002>	-0,157 (0,042) [0,000]	-0,024 (0,007) [0,000]
Maltrato verbal severo	0,72	-0,064 (0,019) [0,001] {0,000} <0,007>	-0,065 (0,017) [0,000] {0,000} <0,002>	-0,096 (0,028) [0,001]	-0,015 (0,004) [0,001]	0,98 [0-2]	-0,086 (0,033) [0,010] {0,039} <0,036>	-0,089 (0,031) [0,004] {0,009} <0,016>	-0,128 (0,050) [0,011]	-0,02 (0,008) [0,011]
Crianza positiva	0,93	0,008 (0,009) [0,368] {0,586} <0,452>	0,008 (0,009) [0,362] {0,586} <0,442>	0,013 (0,014) [0,368]	0,002 (0,002) [0,368]	1,23 [0-4]	0,021 (0,042) [0,612] {0,829} <0,676>	0,019 (0,039) [0,627] {0,856} <0,681>	0,032 (0,062) [0,612]	0,005 (0,010) [0,612]
Materiales de juego	0,94	0,017 (0,012) [0,178] {0,428}	0,015 (0,012) [0,205] {0,474}	0,025 (0,018) [0,181]	0,004 (0,003) [0,180]	3,56 [0-8]	0,041 (0,169) [0,809] {0,829}	0,021 (0,140) [0,880] {0,869}	0,061 (0,252) [0,809]	0,009 (0,039) [0,809]

	Variable dependiente: la práctica ocurrió al menos una vez				Variable dependiente: variedad de prácticas					
	Proporción (grupo de control)	Intención de tratamiento		Tratamiento sobre los tratados		Promedio, # de ítems, [rango] (grupo de control)	Intención de tratamiento		Tratamiento sobre los tratados	
		(1)	(2)	(3)	(4)		(5)	(6)	(7)	(8)
		<0,271>	<0,289>			<0,842>	<0,894>			
Preescolar (>=36 meses en la línea de seguimiento)	0,50	0,083 (0,023) [0,000] {0,004} <0,003>	0,081 (0,023) [0,001] {0,003} <0,002>	0,128 (0,036) [0,000]	0,021 (0,006) [0,000]					
Efectos fijos de trío		X	X	X	X	X	X	X	X	
Controles (características socioeconómicas)			X				X			
Valor P aleatorizado, prueba múltiple de Westfall-Young de la relevancia del tratamiento [1-5]		0,004	0,002			0,001	0,001			

Nota: El N es igual a 4.685 en todas las especificaciones, excepto para los materiales de juego (N=4.625) y la asistencia a educación preescolar (N=3.422). La tabla presenta los coeficientes, los errores estándar (entre paréntesis), los valores p [entre corchetes], los valores p de Romano-Wolf {entre corchetes rizados} y los valores p basados en la inferencia aleatorizada <entre corchetes angulares> de las regresiones de la variable dependiente de la primera columna sobre el tratamiento. Las especificaciones (1), (2), (5) y (6) son las de intención de tratamiento y los resultados se refieren a los niños asignados aleatoriamente al tratamiento. En las especificaciones (3) y (7) de tratamiento sobre los tratados, el "tratamiento" se define como los niños que recibieron al menos una visita domiciliar y se instrumenta con la asignación aleatoria al tratamiento. Del mismo modo, en las especificaciones (4) y (8), el "tratamiento" se define como el número total de visitas recibidas/10 y se instrumenta con la asignación aleatoria al tratamiento. Las especificaciones incluyen sólo los efectos fijos de trío, excepto las especificaciones (2) y (6), que añaden las siguientes características del niño y del hogar en la línea de base: edad del niño superior a 36 meses, sexo, un polinomio de tercer orden en el puntaje total del ASQ-3 en la línea de base, desnutrición crónica, educación materna secundaria o superior, edad materna por encima de la mediana, puntaje TVIP materno por encima de la mediana, depresión materna (CESD-R-10>=10), riqueza del hogar por encima de la mediana, distancia a la capital del distrito por encima de la mediana. Todos los errores estándar están ajustados a nivel de distrito.

Apéndice A: Adaptación y administración del ASQ-3

Adaptación. El ASQ-3 es una herramienta de evaluación, compuesta por cuestionarios específicos para cada edad y completados por el cuidador principal—la madre en la mayoría de los casos. Como herramienta de evaluación, está diseñada para identificar a los niños con riesgo de rezago en el desarrollo y, por lo tanto, es más sensible en la medición del desarrollo en el extremo inferior de la distribución. Nuestro objetivo era identificar los cambios en los niveles de desarrollo debidos a la intervención en toda la distribución de habilidades de cada dimensión del desarrollo y para ello administramos también los tres primeros ítems de los cuestionarios subsiguientes cada vez que el niño alcanzaba la puntuación máxima en una escala. Dado que los elementos de la prueba se organizan en orden creciente de dificultad, esto amplía el techo de la prueba. Además, debido al bajo nivel de educación materna en las zonas rurales de Perú, la prueba se administró realizando una entrevista en lugar de ser autocompletada por la madre o el cuidador principal. Se instruyó a los encuestadores para que administraran ítems específicos mediante la observación directa del niño y otros mediante el reporte de la madre. El método de administración variaba según la dimensión evaluada y la edad del niño. Tanto en la línea de base como en la de seguimiento, los ítems de las dimensiones de cognición y de motricidad fina se administraron en su mayoría por observación directa, mientras que los de la escala personal-social se recogieron principalmente por reporte de la madre. La dimensión de motricidad gruesa, por otro lado, se administró principalmente por reporte de la madre en la línea de base, y una mezcla de reporte de la madre y observación directa en la línea de seguimiento. Los ítems de la dimensión de lenguaje se obtuvieron a través de ambos métodos tanto en la línea de base como en la de seguimiento. Se han utilizado adaptaciones similares del ASQ en otros estudios en países de ingresos medios o bajos (Fernald et al. 2012). Araujo et al. (2019) analizan las características psicométricas del test en su administración para la evaluación de Cuna Más.

Capacitación. La encuesta de línea de base fue realizada por un equipo de 76 encuestadores, que fueron formados por 3 capacitadores durante 14 días, incluyendo 5 días de prácticas. Para la línea de seguimiento, el equipo de encuestadores encargado de la encuesta se redujo a 20, con una capacitación igualmente larga (13 días) y más prácticas de administración.

Puntaje (puntajes brutos). Para cada ítem del ASQ-3, se asignan 10, 5 o 0 puntos a las opciones de respuesta "sí", "a veces" o "todavía no", respectivamente y siguiendo el manual. Para cada dimensión, el puntaje bruto total es la suma de los puntos de todos los ítems de la dimensión. Los ítems que faltan se sustituyen por la media de la escala, sin embargo, si faltan más de dos ítems en una escala, ésta no se puntúa. Como a algunos niños se les administraron tres ítems adicionales cada vez que alcanzaban la puntuación máxima en una dimensión, los puntajes brutos de cada escala se sitúan en el rango de 0 a 90.

Administración. El ASQ-3 se administró en la vivienda del niño, tras una encuesta al hogar. En la línea de base, la encuesta se administró entre abril y agosto de 2013, cuando los niños tenían entre 1 y 24 meses de edad y antes de comenzar a recibir las visitas domiciliarias. La encuesta en la línea de base se administró a 5.209 niños, de los cuales 5.156 registraron información completa en las cinco escalas del ASQ-3. La encuesta de seguimiento se administró entre junio y diciembre de 2015, cuando los niños tenían entre 25 y 55 meses de edad, después de recibir visitas domiciliarias durante aproximadamente dos años.

Apéndice B: Características adicionales de la muestra

La Tabla B1 compara las características en la línea de base de los niños asignados aleatoriamente al tratamiento y al grupo de control. Para cuantificar la magnitud de las diferencias entre los grupos de tratamiento y de control en cada variable, calculamos la diferencia estandarizada recomendada por Imbens y Wooldridge (2009): $\Delta = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_1^2 + S_2^2}}$. La Tabla B1 muestra

que la media de la diferencia estandarizada es de 0,025, y la diferencia más grande es de 0,10. La regla general que se recomienda en Imbens y Rubin (2015) es que esta diferencia no debe superar el 0,25.

La Tabla B2 presenta las correlaciones bivariadas entre el desarrollo infantil en la línea de seguimiento y las características de los niños, las madres y sus familias en la línea de base. Centrándose en el índice agregado del ASQ-3 de las cinco dimensiones y comenzando con las características en la línea de base de los niños, la tabla muestra que las niñas tienen puntajes de ASQ-3 más altos que los niños, alrededor de 0,20 desviaciones estándar, un resultado que es consistente con otros presentados en la literatura (Bando, López-Boo y Li 2016); como era de esperar, los puntajes en la línea de base del ASQ-3 predicen los puntajes del ASQ-3 en la línea de seguimiento; los niños con un estado nutricional deficiente, en particular los niños con baja talla para la edad, tienen menores niveles de desarrollo. Con respecto a las características de las madres, la Tabla B2 muestra que por cada año completo de escolarización se obtienen 0,06 desviaciones estándar más en los puntajes del ASQ-3, mientras que un aumento de una desviación estándar en el TVIP materno está asociado con 0,22 desviaciones estándar más en los puntajes del ASQ-3. En cambio, la depresión materna no predice el desarrollo infantil en nuestra muestra. Por último, en lo que respecta a las características del hogar, el nivel de riqueza (medido por el número de bienes del hogar), tener agua corriente dentro de la vivienda, estar conectado a la red de alcantarillado y que el piso del hogar no sea de tierra, son elementos que se correlacionan positivamente con los puntajes del ASQ-3. Los niños en hogares de mayor tamaño tienden a tener niveles de desarrollo más bajos.

La Tabla B3 presenta los resultados de las regresiones de una variable dicotómica que toma el valor de 1 para aquellos niños que ya no participaban en el estudio en la línea de seguimiento sobre cada una de las características enumeradas en la primera columna y la interacción entre la característica y la asignación aleatoria al tratamiento (16 regresiones distintas). En estas regresiones, el coeficiente sobre el efecto principal —por ejemplo, el coeficiente sobre el sexo del niño de una regresión de la variable dicotómica “niños que abandonaron el estudio” sobre el sexo y la interacción entre el sexo y el tratamiento— indica si determinadas características tienen más probabilidades de relacionarse con la tasa de atrición —en el ejemplo anterior, si las niñas tienen mayor probabilidad de abandonar el estudio que los niños. Por otra parte, el coeficiente del término de interacción —por ejemplo, la interacción entre el sexo y el tratamiento— indica si los niños del grupo de tratamiento que abandonan el estudio son diferentes de los niños del grupo de control que abandonan el estudio —en el ejemplo anterior, si las niñas tratadas tienen más probabilidades de abandonar el estudio que las niñas de control.

La tabla muestra que los niños más pequeños y los hijos de madres más jóvenes, de madres con mayor nivel educativo y de madres con puntuaciones más altas en el TVIP tienen más probabilidades de abandonar el estudio que los demás niños; en cuanto a las características del

hogar, los niños que viven en casas con un piso que no sea de tierra tienen mayores tasas de atrición, mientras que el número de dormitorios en el hogar está correlacionado negativamente con la atrición. Sin embargo, todas las diferencias entre los que abandonan el estudio y los demás niños son de pequeña magnitud. Más importante aún, la interacción con el tratamiento nunca es significativa al nivel del 95% o superior. En otras palabras, aunque los niños que abandonan el estudio son diferentes de los que no lo hacen, estas diferencias son moderadas. Además, no hay indicios de que las características observables de los niños tratados que abandonan el estudio y de los niños del grupo de control que abandonan el estudio sean diferentes entre sí.

Tabla B1. Balance en la línea de base de los niños asignados al tratamiento vs. al grupo de control

	Grupo de tratamiento		Grupo de control		Diferencia estandarizada
	Media/ proporción	SD	Media/ proporción	SD	$\Delta = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_1^2 + S_2^2}}$
<u>Niños</u>					
Edad en meses	12,78	6,60	12,81	6,59	0,00
Proporción de mujeres	0,50	0,50	0,49	0,50	0,01
Puntaje z del ASQ total	-0,04	1,00	0,08	1,00	-0,08
Puntaje z de la talla para la edad	-1,67	1,06	-1,64	1,09	-0,02
Puntaje z del peso para la talla	0,34	0,93	0,37	0,90	-0,02
Actividades de juego (al menos una vez)	0,91	0,28	0,90	0,29	0,02
Castigo físico severo (al menos una vez)	0,14	0,35	0,17	0,37	-0,05
Castigo verbal severo (al menos una vez)	0,32	0,47	0,29	0,45	0,05
Crianza positiva (al menos una vez)	0,94	0,23	0,96	0,20	-0,04
Proporción que vive con ambos padres	0,86	0,35	0,86	0,35	0,00
<u>Madres</u>					
Edad	28,57	7,6	28,37	7,61	0,02
Años de escolaridad	6,94	4,05	6,69	4,02	0,04
Puntaje TVIP	66,05	15,47	65,27	14,87	0,04
Proporción con un puntaje CESD-R >=10	0,13	0,33	0,13	0,34	0,00
<u>Hogares</u>					
Número de miembros	5,49	2,05	5,57	2,09	-0,03
Número de habitaciones	2,52	1,49	2,53	1,49	-0,00
Número de bienes (0-11)	2,70	2,27	2,49	2,23	0,07
Vivienda conectada al acueducto	0,60	0,49	0,59	0,49	0,01
Vivienda conectada al alcantarillado	0,28	0,45	0,22	0,44	0,10
El material principal del piso es tierra	0,74	0,44	0,72	0,45	0,03

Nota: Características en la línea de base de los niños en el grupo de control y de tratamiento. La diferencia estandarizada recomendada por Imbens y Wooldridge (2009) es igual a: $\Delta = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_1^2 + S_2^2}}$

Tabla B2. Asociaciones bivariadas entre las características observables y el desarrollo del niño en la línea de seguimiento

	Todas las dimensiones	Cognición	Lenguaje	Personal-Social	Motricidad Fina	Motricidad Gruesa
	(1)	(2)	(3)	(5)	(4)	(6)
Mujer	0.196*** (0.052)	0.065 (0.052)	0.113** (0.052)	0.233*** (0.051)	0.243*** (0.051)	-0.035 (0.052)
Puntaje z del ASQ total	0.232*** (0.025)	0.164*** (0.026)	0.196*** (0.025)	0.096*** (0.026)	0.152*** (0.026)	0.124*** (0.026)
Puntaje z de la talla para la edad	0.172*** (0.024)	0.130*** (0.024)	0.124*** (0.024)	0.058** (0.024)	0.125*** (0.024)	0.105*** (0.024)
Puntaje z del peso para la talla	0.087*** (0.029)	0.036 (0.029)	0.059** (0.029)	0.063** (0.029)	0.086*** (0.029)	0.030 (0.029)
Actividades de juego (al menos una vez)	0.237*** (0.088)	0.170* (0.088)	0.191** (0.088)	0.117 (0.088)	0.167* (0.088)	0.103 (0.088)
Castigo físico severo (al menos una vez)	-0.005 (0.069)	0.113 (0.069)	-0.027 (0.069)	-0.027 (0.069)	-0.079 (0.069)	0.003 (0.069)
Castigo verbal severo (al menos una vez)	0.048 (0.057)	0.056 (0.057)	0.030 (0.057)	-0.017 (0.057)	0.026 (0.057)	0.056 (0.057)
Crianza positiva (al menos una vez)	0.303** (0.127)	0.316** (0.127)	0.149 (0.127)	0.217* (0.127)	0.190 (0.127)	0.081 (0.127)
Proporción que vive con ambos padres	0.094 (0.075)	0.062 (0.075)	0.104 (0.075)	0.067 (0.075)	-0.001 (0.075)	0.066 (0.075)
Edad de la madre	0.001 (0.003)	-0.001 (0.003)	0.001 (0.003)	0.003 (0.003)	0.003 (0.003)	-0.001 (0.003)
Años de escolaridad de la madre	0.064*** (0.006)	0.059*** (0.006)	0.043*** (0.007)	0.017*** (0.007)	0.047*** (0.007)	0.034*** (0.007)
Puntaje z del TVIP de la madre	0.220*** (0.026)	0.197*** (0.027)	0.128*** (0.027)	0.070*** (0.027)	0.155*** (0.027)	0.144*** (0.027)
Madre tiene depresión	-0.015 (0.076)	-0.050 (0.076)	0.009 (0.076)	-0.053 (0.076)	0.002 (0.076)	0.044 (0.076)
Número de miembros	-0.031** (0.012)	-0.010 (0.012)	-0.023* (0.012)	-0.012 (0.012)	-0.050*** (0.012)	-0.003 (0.012)
Número de habitaciones	0.021 (0.017)	0.016 (0.017)	0.017 (0.017)	0.016 (0.017)	0.007 (0.017)	0.010 (0.017)
Número de bienes (0-11)	0.084*** (0.011)	0.068*** (0.011)	0.051*** (0.012)	0.036*** (0.012)	0.066*** (0.011)	0.044*** (0.012)
Vivienda conectada al acueducto	0.308*** (0.052)	0.264*** (0.052)	0.220*** (0.052)	0.092* (0.053)	0.300*** (0.052)	0.092* (0.053)
Vivienda conectada al alcantarillado	0.214*** (0.062)	0.189*** (0.062)	0.073 (0.062)	0.074 (0.062)	0.241*** (0.062)	0.096 (0.062)
El material principal del piso es tierra	-0.126** (0.058)	-0.174*** (0.057)	-0.005 (0.058)	-0.046 (0.058)	-0.184*** (0.057)	0.015 (0.058)

Nota: Sólo niños en el grupo de control. N va de 1.388 a 1.493 dependiendo de la característica de la línea de base de interés. Cada línea corresponde a una regresión del puntaje z del ASQ-3 (sobre la característica de la línea de base de interés). *, ** y *** indican significancia estadística al 10%, 5% y 1%, respectivamente.

Tabla B3. Probabilidad de pérdida de muestra (tasa de atrición) en función de las características observables

Características en la línea de base	Atrición	Atrición*tratamiento
<u>Niños</u>		
Edad en meses	-0,001* (0,001)	-0,000 (0,001)
Proporción de niñas	0,013 (0,011)	-0,012 (0,013)
Puntaje z total del ASQ-3	-0,001 (0,007)	-0,005 (0,009)
Puntaje z de la talla para la edad	0,003 (0,006)	0,007 (0,005)
Puntaje z del peso para la talla	0,000 (0,010)	-0,003 (0,010)
Actividades de juego (al menos una vez)	0,015 (0,016)	-0,009 (0,011)
Castigo físico severo (al menos una vez)	0,010 (0,018)	0,017 (0,023)
Castigo verbal severo (al menos una vez)	0,026* (0,016)	-0,019 (0,018)
Crianza positiva (al menos una vez)	0,034* (0,020)	-0,012 (0,010)
Proporción que vive con ambos padres	-0,025* (0,014)	-0,012 (0,011)
<u>Madres</u>		
Edad	-0,003*** (0,001)	-0,000 (0,000)
Años de escolaridad	0,006*** (0,002)	-0,002 (0,001)
Puntaje TVIP	0,025*** (0,007)	-0,011 (0,009)
Proporción con un puntaje CESD-R >=10	0,011 (0,018)	0,016 (0,023)
<u>Hogares</u>		
Número de miembros	0,000 (0,002)	-0,003 (0,002)
Número de habitaciones	-0,008* (0,004)	-0,003 (0,003)
Número de bienes (0-11)	0,007** (0,003)	-0,003 (0,003)
Vivienda con agua corriente	0,006 (0,012)	-0,010 (0,013)
Vivienda conectada al alcantarillado	0,036* (0,019)	-0,025 (0,021)
El material principal del piso es tierra	-0,35*** (0,012)	-0,002 (0,010)

Nota: Regresiones de la variable dependiente de la primera columna con respecto a una variable dicotómica que toma el valor 1 si el niño abandonó la muestra y la interacción entre esta variable y el tratamiento. Coeficientes y errores estándar (entre paréntesis). Errores estándar ajustados por conglomerados a nivel de distrito. Los valores de *, ** y *** indican los niveles de significancia del 10%, 5% y 1%, respectivamente.