

**Resultados de la evaluación de impacto del Servicio de Acompañamiento a Familias del  
Programa Nacional Cuna Más**

Septiembre 12, 2016

Documento preparado para el Gobierno del Perú<sup>1</sup>

---

<sup>1</sup> En la elaboración de este reporte colaboraron Viviana Cruzado de la Vega y Denice Cavero Arguedas del Ministerio de Economía y Finanzas en la sección 1, y María Caridad Araujo, Marta Dormal y Marta Rubio-Codina del Banco Interamericano de Desarrollo, en las secciones 2-9. Cualquier pregunta o comentario puede hacerse llegar a [mcaraujo@iadb.org](mailto:mcaraujo@iadb.org).

## ***Resumen ejecutivo***

En 2012, el Gobierno de Perú creó el Servicio de Acompañamiento a Familias (SAF). Parte del Programa Nacional Cuna Más, el SAF está dirigido a población rural, en zonas con altos niveles de pobreza y elevadas tasas de desnutrición. El SAF ofrece visitas domiciliarias semanales a familias con niños menores de 36 meses de edad. Durante estas visitas, a cargo de facilitadoras de la comunidad que han sido capacitadas por el Programa, se trabaja con la madre y el niño en un currículo estructurado que les introduce a una serie de actividades, adecuadas para la edad del niño. El SAF trabaja con las madres y las familias para mejorar sus prácticas de crianza y la calidad de las interacciones que se ofrecen a los niños pequeños. A través de más y mejores actividades de juego en el hogar, el fin último del SAF es promover el desarrollo infantil.

Aprovechando el proceso gradual de expansión del Programa, el SAF contó con una evaluación de impacto experimental. Esta evaluación permite comparar a niños en localidades que fueron aleatoriamente asignadas a recibir el Programa en su primera fase con aquellos que viven en lugares que entrarían al Programa en fases posteriores. Para la evaluación de impacto se recogieron dos encuestas sobre una muestra de más de cinco mil niños: una primera encuesta en 2013, antes del inicio del Programa; y otra en 2015, cuando los niños asignados al grupo de tratamiento habían recibido el Programa por algo más de dos años. En cada una de estas encuestas se obtuvo información socioeconómica sobre el hogar y sus miembros, se caracterizó la calidad del ambiente en el hogar en cuanto a las oportunidades de estimulación que se ofrece a los niños, se tomaron medidas antropométricas—peso y talla, y se evaluó el desarrollo infantil en cinco áreas: cognitiva, comunicación, personal-social y motricidad fina y gruesa.

Los principales resultados de la evaluación de impacto del SAF revelan que el Programa tiene impactos significativos sobre el desarrollo. Estos impactos son más robustos y más significativos para las dimensiones cognitiva y de comunicación. Aunque menos significativos estadísticamente, también se observan impactos sobre las áreas de motricidad fina y personal-social. El impacto del SAF sobre los niños que recibieron al menos una visita del Programa es de una magnitud de entre 0,094 y 0,122 desviaciones estándar para las áreas en las cuales éste es significativo. Estos no son impactos pequeños. Equivalen a cerrar el 18% del gradiente socioeconómico en resolución de problemas y el 35% en comunicación, para los niños de esta muestra.<sup>2</sup>

En una sub-muestra de niños—más pequeños, indígenas en menor proporción y que recibieron una mayor intensidad de tratamiento—y a quienes se administró una prueba diagnóstica, se identifican impactos sobre desarrollo cognitivo y lenguaje receptivo de 0,295 y 0,193 desviaciones estándar, respectivamente, en esta medición. Para esta sub-muestra, los impactos del SAF son de mayor magnitud entre niños de hogares más pobres y con madres con un menor nivel educativo.

La evaluación sugiere que el impacto del SAF en desarrollo infantil opera a través de una mayor frecuencia de actividades de juego y una mayor disponibilidad de juguetes caseros para los niños intervenidos. Entre los posibles canales de impacto también se observan menos prácticas violentas de disciplina y una mayor tasa de matrícula en educación inicial.

En apenas cuatro años de implementación, a julio de 2016, el SAF llega a más de 82 mil niños, en más de 580 distritos. Se trata de una experiencia inédita a nivel internacional de llevar a esa escala un programa de esta índole. Esto supuso importantes desafíos de implementación. Por ejemplo, entre los niños asignados al grupo de tratamiento en la muestra de la evaluación, apenas 64% recibieron el servicio. Entre ellos, existe una gran heterogeneidad en el número efectivo de visitas recibidas a lo largo de los dos años de intervención que fueron evaluados.

---

2 El gradiente socioeconómico se calcula mediante un análisis de regresión. La variable dependiente es el nivel de desarrollo del niño estandarizado. El gradiente corresponde al coeficiente de una variable dicotómica igual a 1 si el niño está en el quintil más rico del índice de riqueza (en comparación con el quintil más pobre, la categoría omitida) y se interpreta como la distancia en desviaciones estándar del puntaje entre los niños de estos dos quintiles. El impacto del SAF en resolución de problemas y en comunicación se expresa como porcentaje de este gradiente.

Además de aspectos de implementación, el Programa tiene todavía un importante margen de mejora en lo que se refiere a la calidad de las visitas en sí mismas. Específicamente, los datos sobre calidad de las visitas sugieren que el Programa tiene un amplio margen de mejora en esta área, tanto en las dimensiones que implican una mejor preparación de la facilitadora como en su capacidad de lograr un mayor involucramiento de la madre y el niño durante la visita.

## 1. Introducción

### 1.1. Antecedentes

Los niños que viven en situación de pobreza están expuestos a múltiples factores de riesgo — condiciones de salubridad precarias, deficiencias nutricionales, entornos de aprendizaje poco estimulantes, entre otros— que limitan su potencial de desarrollo cognitivo, de lenguaje, socioemocional y físico (Walker et al., 2011). En América Latina y el Caribe las brechas socioeconómicas en el desarrollo se hacen evidentes desde muy temprano, incluso desde el primer año de vida, y son crecientes con la edad (Berlinski y Schady, 2015; Rubio-Codina et al., 2015; Schady et al., 2015). Además, en la región estas brechas son mayores en magnitud que en otras partes del mundo (Paxson y Schady, 2007, comparan Ecuador con Estados Unidos; López-Boo, 2014 contrasta Perú con Etiopía, India y Vietnam).

Los efectos de la adversidad durante los primeros años perduran en el tiempo y pueden tener consecuencias durante la infancia, la adolescencia y la edad adulta pues afectan el aprendizaje y la trayectoria escolar, los comportamientos de riesgo, las decisiones de empleo, la salud y otras variables clave para la acumulación de capital humano y el bienestar en general (Walker et al., 2011; para una revisión de la literatura: Currie et al., 2010; Shonkoff, Boyce y McEwen, 2009, entre otros). Al mismo tiempo, se ha documentado la existencia de factores protectores que median para mitigar los efectos de la adversidad en el desarrollo infantil; entre ellos, la lactancia, la educación materna, así como la exposición a interacciones de mayor calidad y a entornos de aprendizaje estimulantes y ricos en vocabulario (Walker et al., 2011; Rubio-Codina, Attanasio y Grantham-McGregor, 2016b). En consecuencia, el acceso a servicios de desarrollo infantil de calidad, focalizados en poblaciones muy vulnerables, tienen el potencial de reducir el tamaño de la brecha socioeconómica en el desarrollo desde los primeros años de vida (Engle et al., 2011).

Esta evidencia ha motivado que en años recientes, en Perú y otros países de América Latina y el Caribe, los programas de desarrollo infantil adquieran mayor importancia dentro de la agenda de la política social. Esto se ha reflejado también en el crecimiento del gasto público destinado a la primera infancia. En el caso específico de Perú, el país invierte aproximadamente el 0,4% del PIB en los menores de 5 años. Con este nivel de gasto, el país se acerca al promedio regional, aunque por debajo del nivel del 0,7% del PIB de los países de la OECD (Berlinski y Schady, 2015).

Perú ha experimentado mejoras notables en algunos indicadores de salud y bienestar de la primera infancia. Por ejemplo, la mortalidad infantil descendió de 55 por cada 1000 nacidos en 1990 a 17 por cada 1000 nacidos en 2014 y la desnutrición crónica se redujo del 37% al 18% entre 1991 y 2012. Destaca además el hecho de que estas reducciones fueron especialmente pronunciadas entre los grupos de peor estatus socioeconómico (Berlinski y Schady, 2015). Pese a estos logros, persisten grandes desafíos, en particular en lo que se refiere a los niveles de desarrollo infantil. Schady et al. (2015) comparan el desarrollo del lenguaje en niños de 4-6 años de edad en cinco países de América Latina y encuentran los resultados más bajos entre niños de las zonas rurales de Perú, a la par con los de Nicaragua. Esto es preocupante dado que los resultados en Perú son significativamente peores que los de Colombia o Ecuador, dos países con un nivel de ingreso per cápita parecido al peruano.

El Gobierno del Perú, en el marco de la política de desarrollo e inclusión social, desde el año 2012 ha impulsado una respuesta articulada a la problemática de la pobreza a través de la Estrategia Nacional “Incluir para Crecer” cuya rectoría se encuentra a cargo del Ministerio de Desarrollo e Inclusión Social (MIDIS)<sup>3</sup>. Esta estrategia, basada en un enfoque de ciclo de vida y de articulación horizontal (entre programas sociales) y vertical (entre los tres niveles de gobierno),

---

<sup>3</sup> Los cinco ejes estratégicos de la estrategia Incluir para Crecer son: (i) nutrición infantil, para niños de 0-3 años, incluyendo la madre gestante; (ii) desarrollo infantil temprano, para niños de 0-5 años; (iii) desarrollo integral de la niñez y la adolescencia, para niños y adolescentes entre 6 y 17 años; (iv) inclusión económica, dirigido a personas en edad productiva, entre 18 y 64 años; y, (v) protección al adulto mayor, que se concentra en la población mayor a 65 años.

es implementada de manera progresiva a través de diferentes programas que tienen criterios de focalización y priorización, a fin de garantizar su efectividad y eficacia.

Para promover el desarrollo infantil de los menores de 36 meses, se creó en 2012 el Programa Nacional Cuna Más (PNCM), adscrito al MIDIS, con una estrategia diferenciada según ámbito urbano y rural a fin de dar una respuesta acorde al contexto de la población. Así, a partir del año 2012 el PNCM brinda el Servicio de Cuidado Diurno (SCD) sobre la experiencia del Programa Nacional Wawa Wasi en ámbitos urbanos y crea el Servicio de Acompañamiento a Familias (SAF)<sup>4</sup>, enfocado en ámbitos rurales donde no existía un programa de cobertura nacional.

La expansión gradual del SAF ha permitido que este servicio pueda contar con una evaluación de impacto experimental. Se trata de una oportunidad única—probablemente la primera a nivel mundial—para evaluar un programa de visitas domiciliarias para la promoción del desarrollo infantil implementado a escala.

En línea con los resultados esperados del Programa, el objetivo principal de la evaluación es medir la efectividad del SAF en el logro de su resultado específico, es decir, el impacto sobre el nivel del desarrollo infantil de los menores de 36 meses que viven en situación de pobreza y pobreza extrema en las dimensiones motora, cognitiva (resolución de problemas), de lenguaje y socioemocional<sup>5</sup>. Dado que la acción directa del SAF se dirige a la madre, se espera observar también cambios en las prácticas de la madre y en la disponibilidad de actividades, materiales y espacios de juego y aprendizaje en la rutina diaria. Son estos los canales de transmisión para lograr la mejora en el desarrollo del niño.

## 1.2. Marco Institucional

La evaluación de impacto del SAF se enmarca en el proceso de implementación de la reforma del Presupuesto por Resultados (PpR), liderado por la Dirección General de Presupuesto Público (DGPP) del Ministerio de Economía y Finanzas (MEF). Asimismo, esta evaluación es de suma relevancia para el PNCM, considerando que el Decreto de creación del PNCM<sup>6</sup>, en el artículo 1º, especifica que, según los resultados de la evaluación de impacto, se propondrá su formalización como política pública, en caso de ser una intervención efectiva. En tal sentido, dicha evaluación fue incluida en el Plan Anual de Evaluaciones del MIDIS del año 2013<sup>7</sup>. Además, la evaluación del SAF también ha sido una de las actividades de la matriz de políticas acordadas con el BID para apoyar las labores del MIDIS en el diseño y evaluación de sus programas, en el marco de sus operaciones de préstamo con el país.

Durante 2012, paralelamente al diseño del Programa Presupuestal de Cuna Más y del SAF, se inició el diseño de la evaluación de impacto del SAF. El diseño metodológico e implementación de la evaluación de impacto fue coordinado por un Equipo Técnico en el que participaron la Dirección de Calidad de Gasto Público de la DGPP del MEF, la División de Protección Social y Salud del Banco Interamericano de Desarrollo (BID), la Dirección General de Seguimiento y Evaluación (DGSE) del MIDIS y PNCM. La DGSE y el PNCM participaron como contraparte técnica de la evaluación. El equipo técnico brindó aportes técnicos durante todas las etapas de la evaluación (diseño metodológico, diseño de instrumentos, capacitación, levantamiento de información y análisis). En el anexo 1 se detallan las funciones y actividades que asumió cada entidad en el desarrollo de la evaluación

---

4 Para facilitar la presentación, en el resto de este documento se usarán el acrónimo *SAF* y el término *Programa* como sinónimos. Se reconoce que el PNCM ofrece dos servicios: el SAF en zonas rurales y el Servicio de Cuidado Diurno en zonas urbanas. En este documento, las referencias al Programa se refieren *siempre* exclusivamente al SAF.

5 Cabe recalcar que los resultados de interés no se enfocan en la salud o nutrición, sino en el desarrollo de las habilidades cognitivas y no cognitivas en la infancia temprana.

6 Decreto Supremo N° 003-2012-MIDIS

7 Resolución Ministerial 044-2013-MIDIS

La metodología de evaluación de impacto empleó un diseño experimental, aprovechando las restricciones en la capacidad operativa del PNCM para cubrir toda su población objetivo a nivel nacional. Entre abril y julio de 2013, con financiamiento del MEF, el Instituto Nacional de Estadística (INEI) realizó el operativo de campo de la encuesta de línea de base (LB), la Encuesta de Salud y Desarrollo de la Primera Infancia. Los instrumentos de medición del desarrollo y otras variables de interés para el estudio (ver anexo 2) así como el análisis de datos de la encuesta se describen en la nota metodológica<sup>8</sup> e informe de LB (DGSE-MIDIS et al., 2015), elaborados por el Equipo Técnico de la evaluación. Entre mayo y noviembre de 2015, dos años después de la LB, el INEI realizó la línea de seguimiento (LS), también financiada por el MEF. Esto permitió revisitar a los hogares de los niños de la muestra de evaluación.<sup>9</sup>

Las labores de preparación, capacitación de encuestadores y aplicadores y ejecución de ambas encuestas, de LB y de LS, fueron coordinadas estrechamente entre el Equipo Técnico y el INEI. Asimismo, personal del MEF y del MIDIS participó en la supervisión y monitoreo del operativo de campo de ambas encuestas.

Entre 2013 y 2015, el Equipo Técnico, a partir del Sistema de Información del SAF (SISAF) ha realizado un monitoreo continuo del estado de afiliación de la muestra de evaluación. El SAF tuvo problemas durante el arranque de sus operaciones y retrasó el inicio de la intervención en algunos lugares con respecto a lo que se había programado inicialmente para la evaluación. Las estimaciones finales, producto de la aplicación del diseño y metodología de la evaluación, han sido realizadas por la División de Protección Social y Salud del BID.

### 1.3. El PNCM

El PNCM, adscrito al MIDIS, fue creado el 23 de marzo de 2012 mediante Decreto Supremo N° 003-2012-MIDIS, con el objetivo de mejorar el nivel de desarrollo físico, cognitivo, lenguaje y socioemocional de niñas y niños menores de 36 meses en condición de pobreza y pobreza extrema a través de dos servicios:

- a. El SAF, enfocado en ámbito rural, ofrece visitas domiciliarias semanales en las cuales una facilitadora trabaja con la madre y el niño en actividades de estimulación sicosocial.
- b. El SCD, enfocado en centros poblados urbanos, brinda un servicio de cuidado para niños de 6 a 36 meses de edad.

El PNCM opera sobre la base de un modelo de cogestión Estado-Comunidad. La cogestión comunal tiene sustento en la participación social, el voluntariado y la vigilancia social. La participación social se entiende como el proceso mediante el cual los individuos asumen responsabilidades en la ejecución de programas y en la toma de decisiones que les afectan. El voluntariado se refiere al rol que cumplen los ciudadanos en el proceso de gobierno y al compromiso personal en proyectos comunales de desarrollo. La vigilancia social se define como el ejercicio de control del ciudadano en la gestión pública.

Los criterios de focalización para los dos servicios del PNCM se describen a continuación.<sup>10</sup> El SCD se focaliza en distritos que cumplan con estas dos condiciones: (a) una incidencia de pobreza total (extrema y no extrema) mayor o igual a 19.1%<sup>11</sup><sup>12</sup>; y (b) al menos un centros

8 La nota metodológica está disponible en el siguiente enlace:

[http://www.midis.gob.pe/dgsye/evaluacion/documentos/CUNAMAS-AF\\_30\\_de\\_Enero.pdf](http://www.midis.gob.pe/dgsye/evaluacion/documentos/CUNAMAS-AF_30_de_Enero.pdf)

9 La metodología de evaluación de impacto y el trabajo de campo se explican en las secciones 1.5 y 2.1.

10 Aprobados mediante Resolución Ministerial N° 148-2013-MIDIS.

11 Umbral definido de acuerdo al promedio nacional de incidencia de pobreza en el área urbana indicado en la Encuesta Nacional de Hogares – ENAHO 2010.

12 De acuerdo al Mapa de Pobreza Provincial y Distrital 2009 del INEI.

poblado urbano<sup>13</sup>. El SAF se focaliza en distritos que cumplen con los siguientes criterios: (a) una incidencia de pobreza total (extrema y no extrema) mayor o igual al 50% <sup>14</sup>; (b) predominantemente rurales<sup>15</sup>; (c) un porcentaje de desnutrición crónica en menores de 5 años mayor o igual a 30%<sup>16</sup>; y (d) al 2012, estar en el ámbito de intervención del Programa Nacional de Apoyo directo a los más pobres “JUNTOS”.<sup>17</sup> Para el caso del SAF, adicionalmente a los criterios señalados, el servicio se focaliza en las familias que habitan los centros poblados rurales (definidos como aquellos con menos de 2,000 habitantes o menos de 400 viviendas). En julio de 2016, el SAF atendía a 82 mil niños en 586 distritos de 18 departamentos.

El PNCM ha logrado un avance importante en su cobertura a nivel nacional, sin embargo su expansión ha sido gradual en sus dos servicios. Al 2015, el PNCM brindó el SCD a 53,493 niñas y niños de 6 a 36 meses de edad y el SAF a un total de 81,874 familias con niñas y niños menores de 0 a 36 meses de edad. Con ello, alcanzó una cobertura de 31% de su población objetivo en ámbito rural y del 12% en ámbito urbano. La ejecución presupuestal para lograr dicha meta en el 2015 fue de 292 millones de soles.

El documento de Diseño del Programa Presupuestal (anexo 2), estableció los siguientes resultados e indicadores asociados al desempeño del Programa:

#### **[Figura 1. Resultados e indicadores asociados al desempeño del Programa]**

El resto de este documento se concentra exclusivamente en el SAF y su evaluación de impacto.

#### **1.4. El SAF**

Como se mencionó anteriormente, el SAF opera a través de un modelo de cogestión con la comunidad. El PNCM aporta el modelo de atención y asume el financiamiento en tanto que la comunidad ofrece su aporte de trabajo voluntario. Los Comités de Gestión de la comunidad suscriben un convenio de cooperación con el PNCM previo al inicio del servicio. De este modo, el PNCM promueve la participación de la comunidad para la prestación del servicio, brindando a su vez asistencia técnica y capacitación de un equipo técnico multidisciplinario que se encuentra en las Unidades Territoriales.

En su diseño original, el SAF se propuso brindar dos servicios:

**1. Sesiones Individualizadas:** Realizadas en el hogar, a partir de actividades de juego con las madres<sup>18</sup> y sus niños. Son desarrolladas por los facilitadores del PNCM que son miembros reconocidos por la comunidad, que cumplen con el perfil establecido<sup>19</sup> y son seleccionados por el PNCM. Son capacitados y supervisados permanentemente por los Acompañantes Técnicos, profesionales cuya responsabilidad es brindar acompañamiento técnico a los facilitadores y conducir las sesiones de socialización e inter aprendizaje. Las sesiones individualizadas se realizan mediante visitas al hogar con una frecuencia semanal y

---

13 Se considera como Centro Poblado Urbano a aquel con al menos 2,000 habitantes y 400 viviendas.

14 De acuerdo a los parámetros del Mapa de Pobreza Provincial y Distrital 2009 del INEI.

15 Distritos con más del 50% de la población reside en Centros Poblados Rurales (aquellos con 2,000 habitantes o menos de 400 viviendas).

16 El porcentaje debe corresponder al promedio en zona rural según la Encuesta Demográfica y de Salud Familiar 2011. Se refiere a datos de 2009 y a definiciones de la Organización Mundial de la Salud.

17 Según datos oficiales a diciembre de 2012.

18 A lo largo del documento se usa el término *madre* para referirse a la madre o cuidador principal del niño, en aquellos casos en que ella no lo fuera.

19 Edad mínima de 21 años, nivel educativo mínimo primaria completa, de preferencia padre o madre de familia, poseer experiencia en trabajo comunitario con niños y familias y manejar la lengua materna de los niños y las familias de la comunidad donde vive o presta el servicio, así como manejo de las diferentes formas de comunicación (gestual, corporal) entre otras. Fuente: RDE-010-2013 MIDIS PNCM Directiva 002 Lineamientos\_Técnicos del SAF del PNCM

una duración promedio de 60 minutos. Para el desarrollo de esta actividad, las familias usuarias reciben materiales de juego y materiales de promoción de prácticas de salud y nutrición.

**2. Sesiones de socialización e inter aprendizaje:** Realizadas en espacios comunales con el objetivo de generar espacios de encuentro con madres, padres y familias y con madres gestantes. El objetivo de estas sesiones es fortalecer las prácticas de cuidado y aprendizaje trabajadas en las sesiones individualizadas. Tienen una duración aproximada de hora y media y se realizan con una frecuencia mensual aunque la participación de las familias se realiza de manera trimestral debido a la dispersión de los ámbitos de atención. La organización de las sesiones está a cargo del Acompañante Técnico y del Acompañante Comunitario. Éste último es además el encargado de facilitar los procesos de gestión comunitaria.

El diseño del SAF se inspiró en experiencias previas implementadas en Perú, como el Programa Integral con Base en la Familia, un programa no escolarizado dirigido a niños menores de 5 años e implementado por el Ministerio de Educación; o Qatari Wawa, del ahora desaparecido programa Wawa Wasi. No existen evaluaciones rigurosas en Perú de programas de este tipo.

El SAF también incorporó en su diseño lecciones de experiencias a nivel internacional de programas de promoción del desarrollo infantil a través de visitas domiciliarias. La evidencia internacional sobre el impacto de estas experiencias tampoco es particularmente abundante. En América Latina y el Caribe, las evaluaciones más rigurosas de este tipo de intervenciones son las del programa piloto de estimulación sicosocial creado y evaluado por Sally Grantham-McGregor y colegas en Jamaica y de su primera implementación a mediana escala en Colombia (Attanasio et al. 2014). Se han identificado impactos a corto, mediano y largo plazo del piloto de Jamaica sobre múltiples variables de resultado, incluyendo el cociente intelectual e indicadores de aprovechamiento y desempeño escolar (Grantham-McGregor et al. 1991, Grantham-McGregor et al. 1997, Walker et al. 2000, Walker et al. 2005). Incluso 22 años después de haber recibido la intervención, los adultos que de niños se beneficiaron de ella tienen un mayor cociente intelectual, mejores resultados en pruebas de lenguaje, matemáticas y conocimientos generales, más años de educación, menos prevalencia de depresión, mejores habilidades sociales, menos comportamientos violentos y salarios 25% mayores a quienes no recibieron el programa (Gertler et al. 2014). El estudio en Colombia, que contextualizó la intervención y ajustó su diseño tal que se replicaran condiciones de implementación similares a las de una intervención a escala, también halló impactos significativos en desarrollo cognitivo y de lenguaje en el corto plazo. La experiencia de ambos proyectos influyó sustancialmente en el diseño del SAF, que adoptó muchos de los elementos del currículo de Grantham-McGregor empleado en Colombia.

El SAF se enfoca en las siguientes prácticas:

#### **[Figura 2. Prácticas principales del Programa]**

Desde que se inició la implementación del SAF, se han desarrollado múltiples esfuerzos para generar las condiciones que demanda su operación en territorio. Entre ellos, se destaca: la convocatoria y contratación de equipos técnicos en las Unidades Territoriales, incorporando nuevos perfiles; el fortalecimiento de capacidades de los equipos técnicos y actores comunales responsables de la prestación del servicio; la provisión de equipamiento y materiales; monitoreo y acompañamiento técnico a los equipos.

Por ser un servicio nuevo, la implementación del SAF no solo implicó el desarrollo de una normativa sino también el diseño de un currículo y material contextualizado, dirigidos a familias en un ámbito multicultural así como la generación de condiciones operativas adecuadas para trabajar en el ámbito rural.

## 1.5. La evaluación de impacto del SAF

El principal objetivo de la evaluación es medir el efecto del SAF sobre el nivel del desarrollo de los niños en las dimensiones cognitiva, de lenguaje, motora y socioemocional. Asimismo, y dado que la acción directa del Programa se dirige a la madre, se esperaría observar también incrementos en su conocimiento sobre el desarrollo infantil y en aquellas prácticas de crianza que lo favorecen. Estos elementos serían los canales de acción—es decir, los mecanismos—para lograr una mejora en el desarrollo del niño.

Como se mencionó en la sección 1.2., se aprovechó de las restricciones en la capacidad operativa del Programa para cubrir a toda la población objetivo de manera inmediata, la evaluación tiene un diseño experimental con asignación aleatoria del tratamiento.

El marco muestral para la evaluación está constituido por 531 distritos focalizados por el Programa en base a los datos del Censo de Población y Vivienda (CPV) de 2007.<sup>20</sup> Estos distritos cuentan con 31.828 centros poblados rurales y se ubican en 12 regiones. Se dejaron fuera del marco muestral de la evaluación las siguientes zonas:

- Distritos de la región de Ayacucho y la provincia de Celendín en Cajamarca, que formaron parte del piloto de la intervención en el año 2012.
- Regiones con muy poca población objetivo: Arequipa (2 distritos), Ica (1), Lambayeque (1), Lima (1), Pasco (6) y Ucayali (2), por dificultades de implementación.
- Distritos donde el Programa ha venido implementando el servicio Qatari Wawa, heredado del ahora desaparecido programa Wawa Wasi, orientado a ámbitos rurales y basado en sesiones grupales con los padres para promover buenas prácticas de crianza.

Dada la capacidad de expansión inicial del Programa y a partir de ejercicios para establecer el efecto mínimo detectable de esta intervención, se estableció trabajar con una muestra de 180 distritos. De éstos, 120 deberían contar con el servicio en la primera fase de expansión del Programa, desde inicios de 2013, y los 60 restantes, se incorporarían 24 meses después, luego del levantamiento de la LS.

Para determinar la muestra de distritos y centros poblados a incluir en la evaluación se siguieron los siguientes pasos:

- Se eliminaron centros poblados que no contaban con un mínimo de 21 niños, por razones relacionadas con la complejidad y el costo del recojo de datos.
- Se ordenó los 195 distritos según su tasa de pobreza, en base al Mapa de Pobreza del año 2009 elaborado por el Instituto Nacional de Estadística e Informática (INEI), y se armaron 65 tríos de distritos con similar nivel de pobreza (*pairwise matching*).<sup>21</sup>
- Se eligió aleatoriamente 60 tríos. En cada trío, se asignaron aleatoriamente dos distritos al grupo de tratamiento y uno al grupo de control.
- En cada uno de estos 180 distritos se seleccionó a los dos centros poblados con la mayor cantidad de niños entre 0 y 24 meses<sup>22</sup>, con lo cual se tuvo una muestra de 360 centros poblados. Adicionalmente, se seleccionó al tercer centro poblado con más niños como reemplazo en caso de que hubiese menos de 5 niños en los dos primeros.
- Se realizó un barrido censal de los centros poblados de la muestra para identificar a todos los hogares con niños de entre 1 y 24 meses, que serían elegibles para recibir el SAF. Una vez obtenido este padrón se seleccionó aleatoriamente 17 hogares por cada centro poblado

---

20 La conformación del marco muestral de 531 distritos se basa en la nota metodológica para la evaluación de impacto del Programa, mencionada en la sección 1.2.

21 Para mayor detalle sobre este método, véase Bruhn y McKenzie (2009).

22 Dado que el Programa atiende a niños hasta 36 meses de edad, se acotó la muestra de evaluación a niños hasta 24 meses para que todos los participantes tuvieran la oportunidad de recibir por lo menos 12 meses de intervención.

para la encuesta de LB. En centros poblados con 17 niños o menos, todos fueron incluidos en la muestra.<sup>23</sup>

Por diseño, la muestra de evaluación no es representativa de la población nacional de niños menores de 36 meses, ni permite hacer inferencias a nivel distrital o provincial (es decir, no tiene validez externa). Dicha muestra ha sido, no obstante, diseñada para producir validez interna de los resultados, en el sentido de producir comparabilidad entre los grupos de tratamiento y control. DGSE-MIDIS et al. (2015) documenta el diseño de la evaluación y la selección de la muestra en mayor detalle.<sup>24</sup>

## 2. Descripción de la muestra de estudio

### 2.1. Indicadores y muestras de análisis

La encuesta de (LB) estuvo a cargo del INEI, y se llevó a cabo entre abril y julio de 2013 en los hogares incluidos por selección aleatoria en los centros poblados de la muestra de evaluación. La encuesta se propuso cubrir una muestra de al menos 5.400 niños de 1 a 24 meses, localizados en 12 departamentos, 180 distritos y 360 centros poblados. Para alcanzar esta meta, se programó encuestar 17 viviendas por centro poblado. Empleando los registros de población, se estimó encontrar a alrededor de 5.784 niños. Según los resultados de la encuesta, del total de viviendas programadas el 97,1% reportó encuestas completas y sólo el 0,4% reportó información incompleta. La tasa de no respuesta neta fue del 1,5%. La muestra final de la LB presentó información sobre 5.859 niños en el rango de edad de 1 a 24 meses al momento de la entrevista, en 5.620 hogares.

El operativo de LB recogió la siguiente información: (a) características sociodemográficas del hogar y sus miembros; (b) calidad del ambiente en el hogar en términos de estimulación, empleando algunos indicadores de la escala *Family Care Indicators* (Frongillo et al., 2003) así como preguntas sobre prácticas de cuidado, disciplina y formas de felicitar al niño; (c) medición de peso y talla de los niños en la muestra y de sus madres; y (d) medición del desarrollo infantil. Para la medición del desarrollo se administraron las cinco escalas—comunicación, motricidad fina, resolución de problemas, motricidad gruesa y socio-individual—de la tercera versión del Cuestionario de Edades y Etapas o *Ages and Stages Questionnaires-3* (ASQ-3; Squires et al., 2009), en el hogar y con la presencia de la madre. Un equipo de sicólogas peruanas trabajó en la adaptación de la prueba al contexto local, incluyendo una versión de la prueba en quechua<sup>25</sup>.

---

23 La elección del conglomerado se basó en estimaciones de la correlación intragrupo usando datos del estudio Niños del Milenio. Este estudio tiene mediciones en 9 distritos predominantemente rurales del nivel de vocabulario receptivo de niños entre 5 y 6 años de edad, usando el Test de Vocabulario en Imágenes Peabody (TVP). Dado que las estimaciones arrojan una correlación intragrupo similar a nivel de distrito y de centro poblado (0.05), se eligió el distrito como conglomerado para reducir las probabilidades de contaminación y desgaste muestral.

Durante el diseño de la evaluación, se calcularon los efectos mínimos detectables (EMD), es decir, los tamaños de impactos en desviaciones estándar, DE, que podrían detectarse dependiendo de las cantidades de conglomerados y niños por conglomerado que se consideren, con un nivel de 5% y un poder de 80%. Para los cálculos de poder de esta evaluación se consideró un EMD de 0.15 desviaciones estándar. Se trata de un EMD más bajo de aquellos identificados en evaluaciones previas de programas de visitas domiciliarias a escala piloto, como el de Jamaica o Colombia. El supuesto con respecto al tamaño del efecto del Programa fue más conservador debido a las características de la intervención en Perú, en particular por la escala de implementación y la menor intensidad de capacitación de supervisoras y facilitadoras.

24 La validez interna se refiere a la credibilidad con la cual el diseño experimental permite identificar un impacto causal. La validez externa se refiere a la posibilidad de extrapolar los resultados de un experimento a otros contextos.

25 En la práctica, en la LB, la traducción quechua del ASQ-3 sólo se administró al 7% de las familias de la muestra. Dado esto, los módulos a administrar en la LS no se tradujeron al quechua.

El ASQ-3 es una prueba de tamizaje para niños de entre 1 y 66 meses de edad, consistente en 21 cuestionarios que se aplican a distintos rangos de edad. Cada cuestionario incluye seis preguntas o ítems por área del desarrollo, que reciben un puntaje de 10, 5 o 0 si el niño puede realizar cierta acción, si sólo la hace a veces, o si aún no la realiza, respectivamente. Los cuestionarios están diseñados para ser auto-completados por las madres. Sin embargo, dados los bajos niveles educativos de las madres en la muestra, se capacitó a un equipo de entrevistadores para que visitaran a las familias y completaran los cuestionarios por entrevista con la madre, o por administración del ítem al niño. Con el fin de estandarizar el proceso de administración, el equipo de sicólogas definió instrucciones específicas para la administración y puntuación de cada ítem<sup>26</sup>. Además, para cada escala en cada cuestionario, se añadieron los tres ítems no coincidentes en la escala correspondiente del cuestionario para la edad inmediatamente superior. Bajo la hipótesis que los ítems están ordenados en orden creciente de dificultad dentro de un mismo cuestionario y entre cuestionarios sucesivos, esta estrategia pretendía ampliar la capacidad de la prueba de medir las habilidades en cada área del desarrollo para aquellos niños en el extremo superior de la distribución, reduciendo el número de observaciones que potencialmente podrían estar censuradas. Los tres ítems adicionales se aplicaron sólo para aquellos niños que realizaron correctamente los primeros seis ítems de alguna escala (es decir, que alcanzaron el puntaje máximo en la escala).

Dado que su administración es más sencilla y rápida que la de otras pruebas de desarrollo para niños tan pequeños, el ASQ-3 o versiones modificadas como el EASQ, han sido comúnmente empleadas para la evaluación del desarrollo infantil a escala en estudios de investigación en Perú y en otros países (Ángeles et al., 2012; Fernald et al., 2012; Bernal, 2015).

La LS se llevó a cabo entre los meses de junio y diciembre de 2015 y recogió la misma información que la LB, con excepción de la escala de identificación de síntomas de depresión materna y algunas preguntas adicionales (en particular las relacionadas a la salud de la madre durante el embarazo del niño de la muestra).

La muestra empleada para esta evaluación está conformada por los 5.339 niños que se encontraron en la LS de los 5.858 niños en la encuesta de LB con medición del desarrollo completa (mediante la prueba ASQ-3). Esta es la muestra de análisis y se usará el término *muestra de la evaluación* para referirse a esta muestra de aquí en adelante.

Así, la LS tuvo una cobertura del 92% de la muestra de la LB, o en otras palabras una tasa de atracción (pérdida) del 8%. De los niños encontrados en seguimiento, 3.530 (66%) forman parte del grupo de tratamiento y 1.809 (34%) del grupo de control, con lo que se mantienen las proporciones originales de niños asignados a un grupo y a otro. En realidad, como era de esperar dado el diseño experimental de la evaluación, la tasa de atracción no se encuentra asociada con la asignación al grupo de tratamiento.<sup>27</sup>

Con el propósito de contar con una medición más precisa de los niveles de desarrollo infantil, en paralelo con la encuesta de seguimiento y para una sub-muestra de niños de la muestra de la evaluación, se recogió también la tercera versión de la prueba *Bayley Scales of Infant and Toddler Development* (Bayley-III; Bayley, 1969; 2006). La prueba de Bayley-III es un instrumento de diagnóstico del desarrollo infantil y está considerado como el patrón oro para su medición en menores de 42 meses de edad (Fernald et al. 2009; Frongillo et al. 2013). Además, la prueba ha demostrado sensibilidad en la identificación de impactos de programas piloto de visitas domiciliarias similares al SAF en contextos tan diversos como Bangladesh (Nahar et al. 2012) y Colombia (Attanasio et al. 2014).

---

26 Como resultado de este proceso, la mayor parte de los ítems se obtienen por administración en las escalas de motricidad fina, motricidad gruesa, y resolución de problemas y por reporte materno en las escalas de comunicación y socio-individual.

27 Se hizo un análisis para evaluar si la pérdida de muestra está asociada o no a variables observables de los niños y de sus hogares en LB. Los resultados están disponibles bajo solicitud.

El Bayley-III evalúa cinco áreas del desarrollo: cognitivo, lenguaje (receptivo y expresivo), motricidad (fina y gruesa), socio-emocional y comportamiento adaptativo. No obstante, dadas varias restricciones de tiempo y presupuestarias, para la evaluación del SAF se administraron exclusivamente las escalas de desarrollo cognitivo, lenguaje (expresivo y receptivo) y motricidad fina. Estas escalas se evalúan mediante la administración de una serie de ítems (actividades) al niño. Esta medición se realizó entre agosto y septiembre de 2015 por sicólogas profesionales, en condiciones cercanas a las óptimas: la prueba de Bayley-III se administró en una sala habilitada para este propósito en la comunidad (por ejemplo, en la posta de salud, en la escuela o en un local comunitario) y en presencia de la madre del niño. El Bayley-III fue adaptado al contexto peruano por especialistas locales y extensamente piloteado. Lazarte (2016) resume en detalle el proceso de adaptación, capacitación y supervisión de la administración de la prueba.

La prueba de Bayley-III se completó para 1.038 niños. Esta será denominada la *sub-muestra de Bayley* de aquí en adelante. Esta cifra representó el 95% de la muestra objetivo que se buscó evaluar con Bayley-III.<sup>28</sup>

La sub-muestra de Bayley fue definida a partir de la muestra de la evaluación de acuerdo con los siguientes criterios:

- a. Se excluyeron los tríos de distritos en los cuales al menos uno de los centros poblados de tratamiento había sido, posterior a la LB, identificado como no elegible al Programa<sup>29</sup>. Este criterio buscó concentrar el esfuerzo de recolección en centros poblados donde el Programa tenía la intención y/o capacidad de operar.
- b. Se excluyeron los tríos de distritos con al menos un centro poblado en el que más del 30% de los hogares respondieron la encuesta de LB en una lengua indígena. Este criterio buscó limitar la administración de la prueba a niños hispano-hablantes pues el Bayley-III no ha sido traducido al quechua.
- c. Dentro de los distritos que cumplían con los criterios de inclusión a y b, se incluyó en la sub-muestra de Bayley a aquellos niños que, en el momento de la LS tuvieran 42 meses de edad o menos, dado que esta es la edad límite para la cual se puede administrar el Bayley-III.

## 2.2. Características de las muestras y balance

En esta sección comparamos las características socioeconómicas y variables de resultado en LB en la muestra de la evaluación entre los grupos de control y tratamiento a nivel de niño, hogar y comunidad. Dado el diseño experimental de la evaluación, estos grupos deberían ser comparables en la gran mayoría de variables observables y no-observables.<sup>30</sup> Las características observables que resulten no estar balanceadas entre estos dos grupos se incluirán como controles en las regresiones que estiman el impacto del Programa.

La Tabla 1 presenta los promedios de estas variables para la muestra de la evaluación, así como para los grupos de control y tratamiento. En la última columna, se reporta el p-valor de la hipótesis nula ( $H_0$ : la diferencia en medias entre los grupos de tratamiento y control es igual a cero), corrigiendo los errores estándar a nivel de distrito, que es el nivel de conglomerado al que se realizó la aleatorización al grupo de tratamiento o control.

---

28 Se calcularon los EMD con un nivel de 5% y un poder de 80%. Usando los datos de la LB, se estimó un coeficiente de correlación intragrupos (CCI) para los puntajes del ASQ-3 del 0,08, controlando por características observables de las familias tales como el índice de riqueza y la composición sociodemográfica del hogar. Con base a un supuesto del 7% para la tasa de atracción, los EMD resultaron ser de una magnitud de 0,227.

29 De los 180 centros poblados asignados a tratamiento, posterior a la LB, el Programa identificó que 40 de ellos no eran elegibles para recibir el SAF, principalmente por dos motivos: o porque contaban con centros del Servicio de Cuidado Diurno o porque las condiciones de seguridad o conflicto social hacían que la operación del SAF no fuera factible.

30 Para mayor detalle sobre la encuesta de LB, véase DGSE-MIDIS et al. (2015).

La Tabla 1 presenta algunas estadísticas descriptivas sobre las características en el momento de la medición de la LB de los niños de la muestra final de evaluación. En LB, en promedio, los niños de la evaluación tenían entre 12 y 13 meses de edad. 49% de ellos son mujeres. Destacan los niveles elevados de desnutrición crónica (38%) y de desnutrición global (8%) en esta muestra. En cuanto a las características del hogar, la madre o cuidador principal de estos niños, en promedio, había completado 6,57 años de escolaridad y el 11% contestó la encuesta de LB en una lengua indígena. En lo que se refiere a la tenencia de activos, a las condiciones de la vivienda y el acceso a servicios básicos, claramente se trata de hogares muy vulnerables: apenas el 8% tiene una refrigeradora, el 41% cocina con gas, el 50% cuenta con un televisor y el 60% posee teléfono celular. Una cuarta parte de las viviendas de los niños de la muestra tienen pisos distintos de tierra y tres cuartas partes cuentan con electricidad. Un 59% de las viviendas tiene acceso a agua y un 25% posee servicio higiénico (en ambos casos, los datos se refieren a acceso dentro de la vivienda).

A nivel de los centros poblados, los resultados muestran que los niños viven en comunidades alejadas, pues el tiempo promedio para llegar a la capital distrital en los medios más frecuentes, en auto y a pie, es de una hora y cerca de una hora y media, respectivamente. Destaca que varios servicios públicos y programas sociales están presentes en la mayoría de los centros poblados de la muestra: el programa Juntos está presente en todas las comunidades, y en casi la totalidad de ellas se reporta tener un centro de educación inicial del Ministerio de Educación (94%) y recibir los servicios de desayuno escolar Qali Warma (97%) y Pensión 65 (98%). Además, el 85% de las comunidades reporta tener un centro o una posta de salud. En cuanto a la actividad económica, la agricultura es una de las tres principales actividades económicas en 85% de estas comunidades. El 39% de las comunidades tienen una farmacia y el 12% un banco o caja.

**[Tabla 1. Características y balance de la muestra de la evaluación (N=5.339)]**

La última columna de la Tabla 1 muestra que—si bien no se observan diferencias significativas con respecto a la edad, el sexo, y el estatus nutricional de los niños—sí existen algunas diferencias significativas entre los grupos de tratamiento y control con respecto a los puntajes de desarrollo, la educación de la madre, y un índice de riqueza del hogar construido a partir de características de infraestructura de la vivienda y activos del hogar usando componentes principales. En particular, los niños en el grupo de control muestran menores niveles de desarrollo cognitivo pero mayores niveles de desarrollo personal-social y motricidad fina. Las diferencias en puntajes promedio son, no obstante, muy pequeñas en magnitud<sup>31</sup>. Asimismo, los niños en el grupo de control tienen, en promedio, madres con menos años de educación (aunque la diferencia es de magnitud muy pequeña) y una mayor proporción de madres indígenas, definidas según la lengua en la cual contestaron la encuesta de LB. En promedio, los hogares en el grupo de control parecen también ser más pobres, según su índice de riqueza—si bien esta diferencia es significativa al 10%. A nivel de los centros poblados, existen diferencias con respecto a la actividad económica entre los centros poblados de estos dos grupos: una mayor cantidad de centros poblados en el grupo de control reportan que la agricultura es una de las tres actividades principales (el 89% en comparación con el 83% en el grupo de tratamiento). También existe una diferencia con respecto al tiempo de traslado a la capital distrital a pie entre los dos grupos (1,8 horas para el grupo de control y 1,2 horas para el grupo de tratamiento, significativa al 10%), aunque esta variable se refiere únicamente a un subconjunto de centros poblados en la muestra que identificaron estos como los medios de traslado más frecuentes.

---

31 Rubio-Codina, Tomé y Araujo (2016) comparan los puntajes de la muestra de LB del ASQ-3 en la muestra de la evaluación con los de la población de referencia que se tomó para normar esta prueba y que es representativa de la población de los Estados Unidos (estandarización externa). Esta comparación debe interpretarse con cautela pues los puntos de corte que se usan para la estandarización externa del ASQ-3 no son necesariamente trasladables a poblaciones distintas de la población para la cual fueron calculados dadas las diferencias tanto socioeconómicas como culturales que se han tomado en cuenta para modificar/adaptar la prueba.

La Tabla 2 replica este ejercicio con la sub-muestra de Bayley. En la LB, los niños de la sub-muestra de Bayley tenían en promedio 8 meses y 49% de ellos son mujeres. Con respecto al estatus nutricional, los resultados muestran prevalencias de desnutrición crónica y global altas, aunque un poco menores que las de la muestra de la evaluación (34 % y 6%, respectivamente). Consistente con la estrategia de selección de esta sub-muestra, solamente el 1% de los niños tiene una madre indígena. El nivel de educación promedio de las madres es de 6,7 años, muy similar al de la muestra de la evaluación. También son muy parecidas las características de las viviendas entre ambas muestras. Con respecto al balance de la sub-muestra de Bayley, existen diferencias entre los grupos de tratamiento y control con respecto a la proporción de mujeres (hay más niñas en el grupo de control), a los puntajes del ASQ-3 en la LB en las áreas de motricidad fina y personal-social, ambos mayores en el grupo de control, y a la proporción de hogares con agua en la vivienda, menor en el grupo de control. A nivel de los centros poblados, no existen diferencias significativas entre los grupos de tratamiento y control. La única excepción es el porcentaje de estos, que cuentan con una comisaría (el 5% en el grupo de control en comparación con el 17% en el grupo de tratamiento).

#### **[Tabla 2. Características y balance de la sub-muestra de Bayley (N=1.038)]**

Es importante destacar que la comparación de medias descrita no está corrigiendo estadísticamente por el hecho que muchas de estas variables están correlacionadas entre sí y provienen de una misma muestra (*multiple hypothesis testing*). La aplicación de este método, que va más allá del alcance de este documento, reduciría el número de diferencias estadísticamente significativas entre el grupo de tratamiento y el grupo de control. Esta consideración aplica a todas las comparaciones de promedios a lo largo del reporte, tanto con los datos de la LB como con los indicadores de resultado en la LS (estimaciones de impacto).

#### **2.3. Comparación de la muestra de la evaluación y la sub-muestra de Bayley**

De cara a la interpretación de los resultados de la evaluación de impacto, es importante documentar las diferencias entre la muestra de la evaluación y la sub-muestra de Bayley. Es importante tomar en cuenta estas diferencias en la interpretación de los resultados de la evaluación y reconocer que la sub-muestra de Bayley fue diseñada para contar con una medición precisa y de muy alta calidad del desarrollo infantil en LS, sin necesariamente ser representativa de la población objetivo del Programa.

La Tabla 3 documenta aquellos aspectos en los cuales las dos muestras de análisis difieren de forma estadísticamente significativa. Específicamente, y como se esperaría por su diseño, la sub-muestra de Bayley se concentra en niños de menor edad y tiene una menor proporción de familias indígenas. Los niños en la sub-muestra de Bayley también tienen tasas de desnutrición significativamente menores a las de los niños de la muestra de la evaluación, así como puntajes mayores para resolución de problemas y motricidad fina en LB. Por el contrario, los niños de la sub-muestra de Bayley tienen puntajes de LB significativamente menores que los niños de la muestra de la evaluación en las otras áreas del ASQ-3 (comunicación, personal-social y motricidad gruesa). Los niños de la sub-muestra de Bayley tienen madres con más años de educación y con un menor índice de riqueza.

#### **[Tabla 3. Comparación de la muestra de la evaluación y la sub-muestra de Bayley]**

### **3. Metodología para el análisis de impacto**

Se identifica el impacto del Programa a través de dos estimadores: el de intención de tratamiento (*intent-to-treat* o ITT) y el de tratamiento en los efectivamente tratados (*treatment-on-the-treated* o

TOT). A continuación se describe la especificación econométrica de cada uno de estos estimadores de impacto.<sup>32</sup>

### 3.1. Intención de tratamiento (ITT)

Para estimar el ITT, se define el indicador de tratamiento  $T_i$  igual a 1 si el niño fue asignado al grupo de tratamiento y 0 en el caso contrario. Dado que la asignación a los grupos de tratamiento o control se hizo de manera aleatoria a nivel de distrito, la estimación no sesgada del impacto del Programa sobre un indicador de resultados  $Y_i$  consiste en la comparación del promedio de este indicador para niños en el grupo de tratamiento ( $T=1$ ) y niños en el grupo de control ( $T=0$ ), controlando por aquellas variables no balanceadas en la LB e identificadas en la sección 2.2. Adicionalmente, se controla por algunas características básicas del niño y por el valor de la variable dependiente en la LB para reducir la varianza residual.

En particular, se estima el siguiente modelo:

$$Y_i = \alpha T_i + \beta \vec{X}_i + \gamma G_i + \delta E_i + \varepsilon_i \quad (1)$$

donde  $Y_i$  corresponde al indicador de la variable de resultado de interés (por ejemplo, el puntaje del niño en el ASQ-3);  $T_i$  es el indicador de tratamiento;  $\vec{X}_i$  es un vector de características del niño y del hogar;  $G_i$  es un efecto fijo del trío de distritos según el cual se hizo la asignación aleatoria del tratamiento; y  $E_i$  es un efecto fijo de encuestador. Al igual que en las pruebas de balance, los errores estándar de la estimación se corrigen a nivel de conglomerado (en este caso, distrito).

Las principales variables dependientes de interés corresponden a los puntajes del ASQ-3 y del Bayley-III. Para el caso del ASQ-3 se analizan las cinco áreas por separado, un puntaje total que combina las cinco áreas y, por comparabilidad con el Bayley-III, un puntaje que combina las áreas cognitiva (resolución de problemas), comunicación y motricidad fina. Ambos puntajes están calculados por análisis factorial. Asimismo, para el Bayley-III se analizan las cuatro áreas medidas por separado y el puntaje total de las cuatro áreas, agregadas también por análisis factorial. Para cada área del desarrollo de cada prueba, los puntajes crudos se estandarizaron internamente usando la media y desviación estándar edad-específicas en la muestra, calculadas con métodos no paramétricos. Por construcción, estos puntajes tienen una media de 0 y desviación estándar de 1. Esta estandarización facilita la comparación de puntajes entre grupos, independientemente de la edad del niño. En estas regresiones, el vector  $\vec{X}_i$  incluye la edad del niño en la LS (en meses cumplidos), el sexo y los puntajes en las cinco áreas del ASQ-3 en la LB. Adicionalmente, todas las especificaciones incluyen la educación de la madre del niño, una variable dicotómica que indica si la madre es indígena y el índice de riqueza, estas son las variables no balanceadas en la LB entre el grupo de tratamiento y el de control<sup>33</sup>.

Adicionalmente se consideran los impactos del Programa para una serie de indicadores intermedios que pueden ayudar a explicar los mecanismos de acción del mismo, tales como las prácticas de crianza, la calidad del ambiente en el hogar en términos de la prevalencia de materiales o actividades que promuevan el desarrollo, o los conocimientos sobre desarrollo infantil. También se considera el impacto del Programa en otros indicadores de interés, como por ejemplo la matrícula en la educación inicial y el estatus nutricional de los niños.

32 El ITT se interpreta como el cambio promedio en la variable de resultado cuando un individuo escogido al azar pasa aleatoriamente de ser participante a ser no participante. Por otro lado, el TOT mide el efecto promedio del tratamiento en el subconjunto de individuos que fueron efectivamente tratados (Bernal y Peña, 2010).

33 En todas las regresiones, los valores faltantes de las variables independientes (5% para educación de la madre, y menos del 1% para los puntajes del ASQ-3 en la LB) se reemplazan por la media (variables continuas) o la mediana (variables dicotómicas) de la muestra y se introduce una variable dicotómica en la regresión que toma el valor 1 para los casos en los que hubo reemplazo.

Es importante recordar que los errores estándar y p-valores de los estimadores de impacto deberían corregirse, dada la correlación existente entre todas las variables que entran en la estimación así como el número de pruebas que se están empleando como variables de resultado. Tal y como se indicó en la sección 2.2, esto se conoce como test de hipótesis múltiples (*multiple hypothesis testing*) y requiere la evaluación de variables altamente correlacionadas—o que forman parte de una misma hipótesis—de forma conjunta. Por ejemplo, todas las áreas del ASQ-3 son aspectos de un constructo más amplio—el desarrollo infantil—y deberían ser evaluadas conjuntamente. Esta corrección puede aumentar el p-valor hasta en tantas veces como variables (hipótesis) se prueben conjuntamente (en este caso, 5 escalas de desarrollo), en el caso más conservador. Como resultado, ciertos p-valores pueden dejar de ser estadísticamente significativos una vez se ajusten por el test de hipótesis múltiples. Si bien la aplicación de este método va más allá del alcance de este documento, el calcular un índice que combine todos estos puntajes es una forma alternativa de controlar por esto, aunque menos rigurosa desde un punto de vista estadístico.

### 3.2. Tratamiento en los tratados (TOT)

El TOT estaría expresado por la siguiente regresión:

$$Y_i = \alpha A_i + \beta \vec{X}_i + \gamma G_i + \delta E_i + \varepsilon_i \quad (2)$$

donde se define el indicador de tratamiento alternativo  $A_i$ , igual a 1 si el niño recibe al menos una visita del SAF en el período de intervención y 0 en el caso contrario. Alternativamente,  $A_i$  también puede ser el número de visitas efectivamente recibido por la familia. Se presentan los resultados de la estimación en ambos casos.

Si bien el número de visitas recibidas por hogar no está reportado por las familias sino que proviene de las bases de datos administrativas del Programa, el SISAF, la estimación del modelo (2) de TOT enfrenta un posible problema de endogeneidad. A diferencia del indicador  $T_i$ , que representa la asignación *aleatoria* a los grupos de tratamiento y control, el haber recibido visitas (o el número de las visitas efectivamente recibidas por cada niño), representado por el indicador  $A_i$ , puede *no* ser exógeno. En este caso, el estimador de TOT estaría sesgado. Por ejemplo, es factible pensar que los niños con madres más interesadas en su desarrollo y en consecuencia en hogares con mayores niveles de estimulación y mayores niveles de desarrollo inicial, reciban más visitas ya que la madre muestra más predisposición a recibir a la facilitadora o la facilitadora disfruta más de las visitas en estos hogares pues es más fácil trabajar en ellos. En este caso, el estimador de TOT estaría sobre-estimando el impacto real del Programa sobre los niveles de desarrollo.

Para evitar este tipo de sesgos, se estima el TOT por el método de variables instrumentales, instrumentando la variable endógena,  $A_i$ , con la asignación al grupo de tratamiento,  $T_i$ . Específicamente, el estimador de interés se obtiene en dos etapas. En una primera etapa, se estima la probabilidad de recibir visitas como función de la asignación aleatoria al tratamiento, la variable instrumental  $T_i$ , y las misma variables observables de los modelos (1) y (2), tal y como muestra la ecuación (3):

$$A_i = \alpha' T_i + \beta' \vec{X}_i + \gamma' G_i + \delta' E_i + \nu_i \quad (3)$$

En una segunda etapa, se estima la regresión (2') donde  $\hat{A}_i$  es el valor estimado de  $A_i$ , calculado en (3):

$$Y_i = \alpha \hat{A}_i + \beta \vec{X}_i + \gamma G_i + \delta E_i + \varepsilon_i \quad (2')$$

La asignación al tratamiento es un instrumento *válido*, ya que es exógeno dada su naturaleza aleatoria, y *relevante*, dado que está altamente correlacionado con el número de visitas efectivamente recibidas por los niños ( $r = 0,65$  para la muestra de la evaluación y  $r = 0,84$  para la sub-muestra de Bayley, cuando se define el indicador de tratamiento alternativo igual a 1 si el niño recibe al menos una visita del SAF) tal y como se observa en el coeficiente  $\alpha'$  en la primera etapa<sup>34</sup>.

### 3.3 Implementación efectiva del SAF en el grupo de tratamiento

Los datos del SISAF documentan que efectivamente la intervención no llegó a un porcentaje importante de los niños en la muestra de la evaluación que habían sido originalmente asignados al grupo de tratamiento. En realidad, el 36% de los niños originalmente asignados al grupo de tratamiento no recibieron el Programa. Por esta razón, se espera que las regresiones de IIT y TOT revelen estimadores de impacto cuantitativamente diferentes.

Las causas por las que este 36% de niños que deberían haber sido tratados no fueron tratados se puede desagregar tal y como se describe en la Figura 3:

- a. El 14% corresponde a niños en los 40 centros poblados de tratamiento que, posterior a la LB, fueron identificados como no elegibles para recibir el SAF por las razones descritas en la sección 2.1.
- b. Otro 5% son niños que nunca recibieron el SAF porque, dados los retrasos en la implementación del Programa, cuando éste llegó a su centro poblado ya tenían más de 24 meses de edad y habían superado la edad máxima de elegibilidad para recibir el servicio.
- c. Un 0,36% de los casos son de niños que no recibieron el SAF porque viven en centros poblados de tratamiento a donde no llegó el Programa.
- d. Las razones por las cuales el 17% restante de niños asignados a tratamiento no recibieron la intervención son varias y están probablemente relacionadas a la decisión de la familia de no participar y/o a que el Programa efectivamente no llegó hasta ellos.

**[Figura 3. Niños asignados al grupo de tratamiento, según su participación en el Programa (N= 3.870)]**

Si bien no existen datos administrativos que permitan desagregar las razones por las que este 17% o equivalentemente, 638 niños asignados a tratamiento no recibieron el Programa, es posible usar los datos de la LB para explorar si existen características observables que expliquen en qué medida estos 638 niños y sus familias son distintos de los 2.469 asignados a tratamiento que sí adoptaron el Programa.

Tal y como muestra la Tabla 4, existían algunas diferencias estadísticamente significativas entre los niños asignados a tratamiento que posteriormente adoptaron el Programa y aquellos que no lo hicieron en base a sus características en la LB.

**[Tabla 4. Características de los asignados a tratamiento que participan y no participan en el Programa]**

Específicamente, los niños que no participan en el Programa son ligeramente mayores y tienen una menor prevalencia de desnutrición crónica. No difieren de los otros niños en su nivel de desarrollo. Estos niños son hijos de madres más educadas, con un mejor nivel de vocabulario y viven en hogares con más activos y mayores niveles de riqueza. En otras palabras, los niños asignados a tratamiento que no participan del Programa parecen estar ligeramente mejor en todas las diferentes dimensiones que se pueden observar. Dado esto, calcular el estimador de TOT sin corregir por la participación efectiva al Programa usando el método de variables instrumentales,

---

34 Los resultados de las estimaciones de esta primera etapa están disponibles contactando a las autoras.

estaría posiblemente sobre-estimando el impacto del Programa para esta sub-muestra. Desde el punto de vista de la progresividad del servicio, es positivo que los niños más vulnerables no hayan sido quienes sistemáticamente se hubieran quedado excluidos de este tratamiento.

Entre los niños que sí recibieron el SAF, los datos administrativos del Programa documentan que existe una gran heterogeneidad en la intensidad de tratamiento. La Figura 4 ilustra la proporción de niños que recibió diferentes categorías de número de visitas (y el equivalente en número de meses de intervención), en base a los datos del SISAF. Casi la mitad (44,4%) de los niños que participó en el Programa recibió al menos 49 visitas en el período de la evaluación (el equivalente de 12 meses), es decir, entre agosto de 2013 y diciembre de 2015. Para poner estos números en perspectiva, se puede pensar que un niño promedio que, en la LB tenía un poco más de 12 meses, en ese lapso debió haber recibido al menos 24 meses completos de intervención hasta su graduación del SAF (lo cual a 4 visitas por mes, corresponde a un total de 96 visitas)<sup>35</sup>.

**[Figura 4. Número de visitas recibidas, grupo de tratamiento de la muestra de la evaluación (N = 3.870)]**

#### **4. Resultados principales: desarrollo infantil**

Esta sección presenta el impacto del Programa en los indicadores de desarrollo infantil considerados y está dividida en dos partes. La primera se concentra en los resultados para la muestra de la evaluación, que emplean el ASQ-3 como indicador de resultado. En la segunda parte se presentan los resultados para la sub-muestra de Bayley. En cada parte se reportan primero la intención de tratamiento (ITT) y la estimación del tratamiento en los tratados (TOT).

##### **4.1. Impactos del SAF sobre desarrollo infantil para la muestra de la evaluación**

La Tabla 5 muestra la estimación del ITT sobre la muestra de la evaluación. Los resultados documentan efectos positivos y significativos sobre casi todas las escalas del desarrollo que evalúa la prueba. Los impactos son de una magnitud de entre 0,06 y 0,08 desviaciones estándar, en otras palabras, haber sido asignado al grupo de tratamiento se traduce en puntajes entre 0,06 y 0,08 desviaciones estándar mayores en las áreas del ASQ-3. Específicamente, en las áreas de resolución de problemas y de comunicación, ambas significativas al 5%, se encuentran impactos de 0,064 y 0,079 desviaciones estándar, respectivamente. Además, en las áreas de motricidad fina y personal-social se identifican impactos positivos y significativos al 10% de 0,061 y 0,066 desviaciones estándar, respectivamente. Al pie de la tabla se reporta el impacto sobre dos medidas agregadas. Una de ellas es el puntaje para todas las dimensiones del ASQ-3 calculado por análisis factorial. La otra es un puntaje que combina únicamente aquellas dimensiones que se evaluaron con Bayley-III, es decir, comunicación, resolución de problemas y motricidad fina, y las agrega también con análisis factorial. En los dos casos el impacto es significativo al 5% y de una magnitud de 0,06 desviaciones estándar.

**[Tabla 5. Efecto ITT sobre el ASQ-3 (N=5.339)]**

La Tabla 6 presenta los resultados de la estimación del TOT mediante variables instrumentales. La parte superior de la tabla define el tratamiento como una variable dicotómica de valor 1 si el niño recibió alguna visita y 0 si no lo hizo. La parte inferior de la misma tabla define el tratamiento como el número de visitas efectivamente recibido. Estas estimaciones reflejan impactos de mayor magnitud, de 0,098 desviaciones estándar para resolución de problemas y 0,122 para comunicación, ambos significativos al 5%. En las áreas de motricidad fina y personal-social, los impactos son de 0,094 y 0,102 desviaciones estándar, respectivamente y significativos al 10%. En esta estimación tampoco se detecta un impacto significativo del Programa sobre el

<sup>35</sup> La sistematización de la información operativa a través del SISAF experimentó dificultades de diversa índole durante los primeros meses de operación del Programa. Por esta razón, es probable que exista un sub-registro de visitas que efectivamente ocurrieron en ese período.

área de motricidad gruesa. Los impactos sobre los puntajes agregados son alrededor de 50% más grandes en magnitud que los reportados para la estimación del ITT. Las magnitudes y significancia estadística de los coeficientes en la parte inferior de la tabla son consistentes con estos resultados: una visita adicional se traduce en un incremento de 0,001 desviaciones estándar en el área de resolución de problemas, y 0,002 desviaciones estándar para comunicación (ambos significativos al 5%). Los impactos para motricidad fina y personal-social son también de 0,001 y 0,002 desviaciones estándar respectivamente, pero significativos al 10%. Con respecto a los puntajes agregados, la magnitud de los coeficientes es de 0,001 desviaciones estándar, significativos al 5% para el total de las 5 áreas y al 1% para el índice que incluye resolución de problemas, comunicación y motricidad fina.

#### **[Tabla 6. Efecto TOT sobre el ASQ-3 (N=5.339)]**

En resumen, estas estimaciones identifican un impacto significativo del Programa en cuatro de las cinco áreas del desarrollo: resolución de problemas, comunicación, motricidad fina y personal-social. Estos impactos se mantienen también en los dos puntajes agregados. La magnitud de los impactos fluctúa entre 0,06 y 0,08 desviaciones estándar para la estimación del ITT y para la estimación del TOT por variables instrumentales presenta magnitudes 50% mayores, o entre 0,09 y 0,12 desviaciones estándar. Esta corrección es importante dado que la participación en el Programa del grupo de tratamiento fue imperfecta, es decir, apenas el 64% de los hogares asignados al tratamiento lo recibieron efectivamente.

No se identifica un impacto diferente de cero para el área de motricidad gruesa, lo cual es consistente con el espíritu del currículo del Programa que tiene un mayor énfasis en las áreas cognitiva (resolución de problemas) y de comunicación.

#### **4.2. Impactos del SAF sobre desarrollo infantil para la sub-muestra de Bayley**

En general, los impactos del Programa para la sub-muestra de Bayley son consistentes con los observados para la muestra de la evaluación, aunque mayores en magnitud. Vale la pena recordar que, como se explicó en la sección 2.3 y se ilustró en la Tabla 3, la sub-muestra de Bayley tiene una composición diferente que la muestra de la evaluación. Específicamente, se concentra en niños de menor edad, excluye a las localidades que el Programa determinó que no serían elegibles para el SAF y se acota a poblaciones con menor presencia indígena. Por estas características, en la sub-muestra de Bayley existe una mayor participación en el SAF, 83% comparado con 64% para la muestra de la evaluación. Además, en la sub-muestra de Bayley hay un porcentaje más grande de niños que en LS continuaban recibiendo el servicio pues no habían cumplido los 36 meses de edad. Apenas 43% de los niños de la sub-muestra de Bayley tenían más de 36 meses en LS (comparados con 68% de aquellos en la muestra de la evaluación). Esto quiere decir que los impactos estimados sobre la sub-muestra de Bayley son más de corto plazo que los impactos estimados sobre la muestra de la evaluación. O en otras palabras, consideran en menor medida una posible reducción paulatina del tamaño del impacto que podría darse una vez que el niño se gradúa del Programa.

Para contrastar la muestra de la evaluación con la sub-muestra de Bayley, en el panel de la izquierda de la Tabla 7 se replica la estimación del ITT usando como variable de resultado el ASQ-3 (presentados antes en la Tabla 5) y en el panel de la derecha se los compara con los mismos resultados acotados para los niños que se encontraron en la LS y que también están en la sub-muestra de Bayley. Esta estimación revela un impacto estadísticamente significativo al 1% de 0,177 desviaciones estándar en resolución de problemas y de 0,075 desviaciones estándar, significativo al 10%, para el agregado de las tres dimensiones del ASQ-3 que se evaluaron con Bayley-III. Los impactos sobre las otras áreas del desarrollo no son significativos. Dado el tamaño de los errores estándar de los coeficientes de estas regresiones, no se puede rechazar que éstos sean iguales a los estimados con la muestra completa de la evaluación y reportados en el panel de la izquierda.

**[Tabla 7. Comparación del efecto ITT sobre el ASQ-3, para la muestra de la evaluación y la sub-muestra de Bayley]**

La Tabla 8 presenta los resultados de las estimaciones de impacto usando el Bayley-III como variable de resultado. En la sub-muestra de Bayley, los impactos del SAF se concentran en dos de las cuatro áreas del desarrollo evaluadas, el área cognitiva y el lenguaje receptivo. El coeficiente de la estimación del ITT sobre desarrollo cognitivo es de 0,247 desviaciones estándar, significativo al 1%. El impacto sobre lenguaje receptivo es de 0,167 desviaciones estándar, significativo al 5%. Estos impactos se traducen en un efecto de 0,139 desviaciones estándar sobre el puntaje agregado para Bayley-III (significativo al 10%).

**[Tabla 8. Efecto ITT sobre el Bayley-III (N=1.038)]**

Además de las diferentes características de la sub-muestra de Bayley, el hecho de que el Bayley-III sea un instrumento de diagnóstico y no de tamizaje y capture de manera más precisa el desarrollo infantil puede también explicar las diferencias en magnitud que presentan los coeficientes. Al mismo tiempo, dado el tamaño de los errores estándar de los coeficientes, éstos no son tan diferentes de aquellos medidos con el ASQ-3.

Por último, la Tabla 9 presenta los resultados de la estimación del TOT mediante variables instrumentales para la sub-muestra de Bayley. Al igual que para la muestra de la evaluación, la magnitud de los impactos es mayor. Sin embargo y como era de esperar, la diferencia no es tan grande como la observada antes. Los impactos del Programa sobre desarrollo cognitivo son de 0,295 desviaciones estándar (19% mayores a los del ITT) y sobre lenguaje receptivo son de 0,193 desviaciones estándar (15% mayores a los del ITT), ambos significativos al 1%. Esto se traduce en un impacto de 0,165 desviaciones estándar sobre el puntaje agregado (significativo al 10%). No sorprende que para la sub-muestra de Bayley los puntajes del ITT y del TOT sean más cercanos en magnitud dado que, por diseño, en esta muestra la intensidad de tratamiento es mayor.

**[Tabla 9. Efecto TOT sobre el Bayley-III (N=1.038)]**

## **5. Heterogeneidad de los impactos**

Detrás de los parámetros estimados en las regresiones anteriores, pueden esconderse impactos heterogéneos para diferentes sub-grupos de la población de niños y familias beneficiarios del Programa. En esta sección se explora la existencia de impactos heterogéneos. Específicamente, se replican las estimaciones del ITT incluyendo una interacción entre la variable de asignación al grupo de tratamiento y un conjunto de características del niño y su familia que se esperaría pudieran estar asociadas a impactos diferenciados del Programa, entre ellas:

- La edad, definida como una variable dicotómica que adquiere el valor de 1 si el niño tenía menos de 36 meses de edad en LS y 0 si tenía 36 meses o más.
- El sexo del niño.
- El nivel de riqueza del hogar, definido como una variable dicotómica que adquiere el valor de 1 entre los niños de hogares del 20% más pobre (el quintil 1 del índice de riqueza en la LB) y 0 entre los niños de los quintiles 2, 3, 4 y 5.
- La educación materna, definida como una variable dicotómica que adquiere el valor de 1 entre niños cuyas madres tienen educación primaria incompleta o menos y 0 si es que han completado la primaria o han adquirido más años de educación.
- La desnutrición crónica medida en LB, usando los lineamientos de la Organización Mundial de la Salud<sup>36</sup>.
- Una variable dicotómica que adquiere el valor de 1 si el niño de la evaluación es el primer hijo de la madre y 0 si no lo es.

---

36 Los lineamientos utilizados se consultaron en la siguiente página web de la Organización Mundial de la Salud: <http://www.who.int/childgrowth/es/>

- Una variable dicotómica que adquiere el valor de 1 si el niño de la evaluación es hijo único y 0 si tiene hermanos.
- Una variable dicotómica que adquiere el valor de 1 si el niño de la evaluación es hijo de una madre que está en el 10% más alto de un puntaje que indica síntomas depresivos, construido mediante el método de análisis factorial con los indicadores de reporte materno que se incluyeron en la encuesta de LB y 0 si está en el 90% restante de esa distribución.
- Una variable dicotómica que adquiere el valor de 1 si el niño de la evaluación es hijo de una madre que reporta estar en el 10% más alto de un puntaje de violencia doméstica construido mediante el método de análisis factorial con los indicadores de reporte materno que se incluyeron en la LB y 0 si está en el 90% restante de esa distribución.
- Una variable dicotómica que adquiere el valor de 1 si el niño recibió lactancia materna exclusiva durante sus primeros 6 meses de vida.

La Tabla 10 resume los principales resultados del análisis de impactos heterogéneos. La tabla tiene dos paneles. En el panel superior, se recogen los resultados de la dimensión cognitiva de las pruebas (resolución de problemas en el caso de ASQ-3). En el panel inferior, se recogen los resultados para comunicación (lenguaje receptivo en el caso de Bayley-III). Dentro de cada panel, las columnas de la izquierda hacen referencia a la muestra de la evaluación y las de la derecha a la sub-muestra de Bayley. En las primeras dos columnas se reportan los valores de los parámetros para las interacciones y sus p-valores. Inmediatamente después se encuentra el principal parámetro de interés: el p-valor de la prueba de la igualdad de estos coeficientes. Como referencia, la última columna resume el porcentaje de niños de la muestra que se incluyen en cada uno de los grupos analizados.

Los principales resultados de la Tabla 10 son los siguientes. En la muestra completa de la evaluación, hay muy poca evidencia de impactos heterogéneos. Los resultados en la sub-muestra de Bayley muestran más impactos heterogéneos significativos. Puede que esto se deba a que se trata de una medición más fina, o que esté relacionado con aspectos de composición de la muestra. En el área de desarrollo cognitivo, se observa que el Programa tiene impactos significativamente mayores en las niñas, en los hijos de hogares más pobres, y en los hijos de madres con menor educación. En el área de lenguaje receptivo, se observan impactos significativamente mayores entre los hijos de hogares más pobres, entre los hijos de madres con menor educación y entre aquellos que recibieron lactancia materna exclusiva durante sus primeros seis meses de vida. Estos resultados son alentadores pues parecen sugerir que el impacto del Programa es progresivo, es decir, de mayor magnitud entre sub-grupos de beneficiarios más vulnerables. Dicho esto, es importante tener en cuenta que los coeficientes de estas interacciones que miden heterogeneidad en los impactos pueden estar sesgados—salvo tal vez para la interacción de género. La muestra de la evaluación no se diseñó para identificar impactos heterogéneos: por ejemplo, no se asignó tratamiento de forma aleatoria a madres con mayores y menores niveles educativos. Por ello, toda evidencia sobre heterogeneidad de impactos es simplemente indicativa y no robusta, estadísticamente hablando.

**[Tabla 10. Heterogeneidad en el efecto sobre el ASQ-3 y el Bayley-III]**

## **6. Prácticas, conocimientos y otras variables de resultado**

La teoría de cambio que se encuentra detrás del modelo de intervención del SAF propone que los impactos de este tipo de intervención sobre el desarrollo infantil operan, entre otras cosas, mediante más y mejores conocimientos y prácticas de crianza en las familias, los mismos que se reflejan en mayores oportunidades de juego y aprendizaje para los niños. En esta sección se analizan los resultados de la estimación del modelo de ITT sobre un conjunto de variables intermedias que podrían verse afectadas por la intervención—la calidad del ambiente en el hogar, los conocimientos, entre otras—y que, a su vez, estarían asociadas con mejores resultados de desarrollo durante los primeros años de vida.

Para caracterizar la calidad del ambiente del hogar o el nivel de estimulación que se ofrece a los niños, se administró una versión ligeramente modificada de los Indicadores de Cuidado Familiar o *Family Care Indicators* (FCI, Frongillo et al., 2003), desarrollados por UNICEF. El FCI, más corto y sencillo de administrar, ha sido validado contra el *Home Observation for Measurement of the Environment* (HOME, Caldwell y Bradley, 2003) en Bangladesh (Hamadani et al., 2004) y ha mostrado sensibilidad a intervenciones de desarrollo infantil dentro y fuera de la región (ver Attanasio et al., 2013 y Attanasio et al., 2014 para Colombia; y Tofail et al., 2013 para Bangladesh). La versión del FCI administrada incluyó:

- a. El número de juguetes con los que el niño usualmente juega, clasificados según el origen de los materiales de juego: comprados, hechos en el hogar, y objetos del hogar.
- b. El número de juguetes con los que el niño usualmente juega, clasificados según su uso: juguetes con los que se toca música, materiales para pintar/escribir, juguetes para armar/construir, juguetes que requieren mucho movimiento físico, muñecos para el juego de roles, libros para colorear, juguetes para aprender formas y colores, y libros infantiles de imágenes.
- c. El número de libros para adultos en el hogar, sin incluir libros escolares ni libros para niños.
- d. El número de revistas y periódicos.
- e. La frecuencia con la que el niño realizó alguna de las siguientes actividades de juego con una persona mayor de 15 años en la semana previa a la entrevista: leer libros, contar historias, cantar, salir de paseo, jugar con sus juguetes, dibujar/pintar/escribir, jugar a nombrar objetos/colores/contar.
- f. Si el niño se quedó solo o al cuidado de un menor de 10 años en el hogar y cuántas veces lo hizo; y
- g. Formas o métodos de disciplina y de reconocimiento de buenos comportamientos.

Las variables (a)-(d) se obtuvieron por observación directa y la información en (e)-(g) por reporte de la madre. Las variables (b) y (e) se agregaron usando análisis factorial y se normalizaron a puntajes Z con media 0 y desviación estándar de 1 para facilitar su interpretación.

También con el objetivo de evaluar la calidad del ambiente del hogar, se recogieron un subconjunto de los ítems de las escalas de receptividad y aceptación de la herramienta HOME (*Home Observation for Measurement of the Environment*; Bradley, 1993) relacionados con la calidez y la dureza de las interacciones entre padres e hijos observadas durante la entrevista. Para evaluar esta escala las entrevistadoras registran lo observado durante la encuesta en relación a la forma en que las madres y sus parejas tratan a los niños. Las escalas de receptividad y aceptación del instrumento original constan de 11 y 7 ítems respectivamente. En esta encuesta se incluyó un total de 11 de esos ítems. Los ítems de estas dos escalas se agregaron usando análisis factorial y, al igual que en los casos anteriores, se normalizaron a puntajes Z.

Con el objeto de indagar en qué medida el Programa había tenido un impacto en los conocimientos de las familias sobre el desarrollo infantil, la encuesta de seguimiento incluyó dos baterías de preguntas enfocadas en este tema. La primera de ellas recoge un subconjunto de los ítems de una escala más larga y compleja, el *Inventario de Conocimientos sobre el Desarrollo de los Bebés* (*Knowledge of Infant Development Inventory* o KIDI, por sus siglas en inglés; MacPhee, 2002). Este cuestionario fue desarrollado en los Estados Unidos con el fin de evaluar el conocimiento de un adulto sobre los procesos de desarrollo y las normas de comportamiento infantil para niños de 0 a 5 años, en el contexto de informar la investigación sobre las prácticas de crianza y la evaluación de programas de apoyo a los padres. La escala original consta de 58 ítems. En esta encuesta se administraron 22, los más relevantes para el grupo etario de interés. También el Programa incluyó en el cuestionario un conjunto de 11 preguntas sobre prácticas de cuidado y aprendizaje que se esperaría que las madres hubieran adquirido gracias al SAF. Como en todos los casos anteriores, los ítems de estos dos conjuntos de preguntas se agregaron mediante análisis factorial y los totales se expresan como puntajes Z.

Otra variable de resultado que se explora en esta sección se relaciona con la participación de los niños en servicios de educación inicial. Se esperaría que los niños que se gradúan del SAF

tuvieran acceso a los servicios de prescolar, de tal manera de mantener y continuar potenciando los logros de desarrollo y aprendizaje que el Programa promueve en la primera infancia. A nivel agregado y a partir de la encuesta de la comunidad que se recogió en seguimiento, se conoce que un 94% de los centros poblados de la muestra de la evaluación cuentan con un servicio de educación inicial (definido como un centro o jardín del Ministerio de Educación y sin contar otras modalidades de servicio ni los jardines de proveedores privados). Tomando en cuenta ese contexto, se definió una variable dicotómica que tiene un valor de 1 si el niño asiste a un servicio de esa índole en LS y 0 si no lo hace. La construcción de esta variable se acotó a los niños de más de 36 meses en LS y que podrían matricularse en los servicios de prescolar del Ministerio de Educación.

Además de estas variables de resultado, se explora también si el Programa tuvo algún impacto en los indicadores nutricionales, específicamente el peso para la edad, la talla para la edad, el peso para la talla y el índice de masa corporal. Usando los lineamientos de la Organización Mundial de la Salud<sup>37</sup>, también se calcularon como variables de resultado las prevalencias de la desnutrición global, crónica y aguda.

La Tabla 11 presenta los resultados de impacto del SAF sobre todas estas variables. Se observan algunos resultados consistentes con una mejor calidad del ambiente en el hogar. Por ejemplo, la tabla documenta un efecto positivo y significativo al 1% de 0,201 desviaciones estándar sobre la frecuencia de actividades de juego en el hogar. También se destaca un incremento significativo al 1% de 8 puntos porcentuales (sobre una media de 12,7% para el grupo de control) en la presencia de juguetes hechos en casa entre aquellos con los que habitualmente juega el niño. Hay un efecto, muy modesto y significativo al 10%, de 1,7 puntos porcentuales en la frecuencia de juguetes comprados (sobre una media de 90,5% para el grupo de control). El incremento en las actividades de juego que un adulto realiza con el niño es muy relevante puesto que denota un aumento en la cantidad y calidad de interacciones entre el niño y adultos en el hogar—es decir, un cambio en prácticas de crianza que se considera una de las condiciones necesarias para la sostenibilidad de los impactos. No es claro por qué la asignación al tratamiento podría estar asociada con una menor presencia de libros para adultos ni periódicos, variables que aparecen como negativas y estadísticamente significativas. En ambos casos, los coeficientes son significativos al 5% y muestran una reducción de 3,8 y 3,2 puntos porcentuales, respectivamente, con respecto a promedios de 18,8% y 12,7% para el grupo de control. Se observa que un área donde existe espacio de mejora para el Programa es tanto en la promoción de una mayor variedad de materiales de juego como en el uso de objetos disponibles en el hogar para actividades de juego pues en ninguna de estas dos variables hay impactos significativos en la estimación del ITT.

En lo que se refiere a las prácticas de disciplina, la estimación del ITT documenta una reducción de prácticas violentas como los gritos e insultos de 3,8 puntos porcentuales, estadísticamente significativa al 1%, y con respecto a un promedio de 30,9% para el grupo de control. También se reduce la frecuencia de los golpes con correa, que caen en 2,4 puntos porcentuales con respecto a un promedio de 13,2% para el grupo de control (este coeficiente es significativo al 5%). Además, hay un impacto sobre la frecuencia con la cual los padres premian un comportamiento positivo mediante felicitaciones verbales o lenguaje no verbal como los aplausos. Se detecta un impacto de 8,4 puntos porcentuales significativo al 1%, sobre un promedio de 41,5%. Un área en la cual existe espacio de mejora en cuanto a prácticas de disciplina tiene que ver con el uso de recompensas o premios para incentivar el buen comportamiento, que actualmente parece estar negativa y significativamente afectado por la asignación al Programa (con un impacto de 2,7 puntos porcentuales con respecto a una media de 11,6% y significativo al 1%).

Es importante tomar en cuenta que estas variables tienen algunas limitaciones. Específicamente, las variables como el número de juguetes que se recogen por observación pueden medirse con poca precisión, por ejemplo, si la madre no le muestra al encuestador todos los juguetes con los cuales normalmente juega el niño. De la misma manera, las variables que se recogen por

---

37 Ibid

entrevista a la madre como la frecuencia de actividades de juego o las prácticas de disciplina pueden ser reportadas de manera estratégica por parte de las madres que han recibido el SAF y que han adquirido una noción sobre la importancia de estas variables para el desarrollo de sus hijos.

En lo que se refiere a los indicadores de calidad del ambiente del hogar de la escala HOME (dureza y calidez) y a las variables de conocimiento sobre el desarrollo infantil (tanto las de la escala KIDI como las definidas por el Programa), no se observan impactos significativos de la asignación al tratamiento. En cuanto a participación en la educación inicial, es interesante comprobar que el Programa tuvo un impacto significativo al 1% de 0,085. En otras palabras, con respecto a una matrícula promedio para el grupo de control de 50,7%, ésta aumentó en 8,5 puntos porcentuales entre los niños asignados a tratamiento y en edad de asistir al prescolar.

Finalmente, la tabla muestra que—a partir de la estimación de ITT—se observan efectos modestos sobre las variables nutricionales: únicamente una reducción de la desnutrición global en 0,012 puntos porcentuales sobre una media de 0,079 para el grupo de control (significativa al 5%) y un incremento significativo de 0,036 desviaciones estándar (significativo al 10%) al puntaje Z de talla para la edad.

**[Tabla 11. Efecto ITT en otros indicadores de resultados, muestra de la evaluación (N=5.339)]**

## **7. Diferencias entre tratamientos**

El diseño original de la evaluación de impacto del SAF contempló la implementación de dos tratamientos alternativos (T1 y T2). Uno de ellos (T1) consistía exclusivamente en las visitas domiciliarias semanales que ofrece el Programa. Un segundo tratamiento (T2) complementaba las visitas con sesiones grupales quincenales, bajo la conducción de la Acompañante Técnica, que servían tres propósitos: (a) profundizar los mensajes transmitidos por las facilitadoras en las visitas domiciliarias, (b) fortalecer la capacitación continua para las facilitadoras mediante la demostración de las Acompañantes Técnicas sobre cómo transmitir estos mensajes, y (c) generar espacios de interacción entre las madres de la comunidad y sus hijos. En abril de 2014, tras 9 meses de implementación del SAF en los distritos de la evaluación, el Programa identificó que la organización de las sesiones grupales no estaba funcionando y su organización imponía costos excesivos de tiempo y dedicación de parte de las Acompañantes Técnicas, que además limitaban su capacidad de cumplir con sus otras responsabilidades. Asimismo, las sesiones grupales no contaban con un contenido ni una guía curricular prestablecida, ni se estaban llevando a cabo con la frecuencia esperada en todas las localidades. Ante esa situación, el Programa decidió dejar de promover las sesiones grupales y, para propósitos de la evaluación, se optó por consolidar los dos grupos de tratamiento en uno solo.

Lamentablemente, el Programa no cuenta con datos administrativos que permitan cuantificar el número de sesiones grupales que se llevaron a cabo en los centros poblados de tratamiento ni tampoco en cuántos de ellos éstas se han seguido organizando con alguna periodicidad después de abril de 2014. Por esta razón no se puede documentar empíricamente con qué frecuencia e intensidad se llegó a implementar esta estrategia de tratamiento diferenciado.

Es importante destacar que si se compara la intensidad de las visitas domiciliarias entre los dos grupos de tratamiento usando los datos administrativos del Programa, no se detectan diferencias estadísticamente significativas. En otras palabras, los números promedio de visitas recibidas por los niños de T1 y T2 tanto en la muestra de la evaluación como en la sub-muestra de Bayley son parecidos. Esto sugiere que, más allá de la corta implementación de las sesiones grupales, no hubo otras diferencias sistemáticas en cómo se implementó el SAF en los distritos en estos dos grupos.

Usando la asignación aleatoria a estos dos grupos de tratamiento, es posible replicar las regresiones de ITT para explorar si existe algún efecto diferenciado entre los niños de hogares

asignados a estos dos grupos de tratamiento y probar si los coeficientes correspondientes a estos dos grupos son significativamente distintos uno de otro. La Tabla 12 resume estos resultados para todas las medidas de desarrollo infantil reportadas antes. No se identifica ninguna diferencia significativa en los impactos del Programa entre estos dos grupos de tratamiento, lo cual es consistente con la corta, poco frecuente y escasamente estructurada implementación de las sesiones grupales.

**[Tabla 12. Efecto ITT sobre el ASQ-3 y el Bayley-III, según el tipo de tratamiento]**

## 8. Calidad

Detrás de los impactos estimados para el SAF están condiciones heterogéneas en lo que se refiere a la implementación del Programa. Existen variables estructurales, como la duración y la frecuencia de las visitas o la preparación de las facilitadoras, que pueden afectar la efectividad de la intervención. También existen variables de proceso, sobre cómo se implementan las visitas domiciliarias, que deberían estar correlacionadas con los resultados de éstas sobre los cambios de comportamiento en la familia y, en última instancia, con los resultados de desarrollo infantil.

Con el objetivo de lograr una mejor comprensión sobre la calidad de las visitas domiciliarias del SAF, la evaluación se complementó con un ejercicio adicional: se recogió una medida de la calidad de las visitas desarrolladas por la facilitadora para aquellos niños del grupo de tratamiento en la sub-muestra de Bayley y menores de 36 meses de edad (tal que todavía estuvieran recibiendo visitas domiciliarias). Entre agosto y octubre de 2015, se filmaron 554 visitas, correspondientes a 176 facilitadoras, en 100 centros poblados de 9 departamentos. De ellas, 229 correspondían a niños de la sub-muestra de Bayley y 325 a otros niños que trabajaban con las mismas facilitadoras pero no eran parte de la sub-muestra.

Para el efecto, se realizó una filmación de una visita a cargo de esta facilitadora con este niño y su familia. También se filmaron al menos 2 visitas *adicionales* de esta misma facilitadora con dos niños y familias más, a los que ella atendiera regularmente. Estas filmaciones se emplearon para administrar una escala que se ha usado en Estados Unidos para la medición de la calidad de las visitas domiciliarias de promoción del desarrollo infantil, conocida como Escala de Calificación de las Visitas Domiciliarias o HOVRS, por sus siglas en inglés (*Home Visits Rating Scale*, HOVRS A+V2, Roggman et al., 2010).

El HOVRS es un instrumento observacional que tiene como objetivo evaluar la efectividad de la visita domiciliaria. Tiene dos dominios. El primero, Prácticas de la visitadora domiciliaria, consta de cuatro dimensiones: (a) Receptividad hacia la familia, (b) Relación con la familia, (c) Facilitación de la interacción entre el cuidador y el niño, y (d) No intrusión y colaboración. El segundo dominio, Involucramiento de la familia, consta de tres dimensiones: (a) Interacción entre el cuidador y el niño, (b) Involucramiento del cuidador, y (c) Involucramiento del niño. El HOVRS puede administrarse por observación directa o codificarse a través de un video. Se puntúa en una escala de 1 a 7, en donde un 1 se interpreta como una visita de calidad inadecuada, un 3 como adecuada, un 5 es bueno y un 7 excelente. En el pasado, se ha empleado tanto para propósitos de investigación como para la supervisión y acompañamiento de facilitadoras.

Para este ejercicio, cada video fue codificado dos veces, por dos codificadores separados, asignados aleatoriamente a esta tarea. Cuando existían discrepancias en los puntajes de codificación, el video recibió una tercera codificación. En la práctica, 39% de los videos fueron codificados tres veces.

Vale la pena destacar las diferencias entre los dos dominios del HOVRS. El primero se enfoca en aspectos como la preparación de la facilitadora, su capacidad de adaptarse y responder a las circunstancias individuales de la madre y el niño, si existe una relación cálida y positiva entre la facilitadora y madre, y la capacidad de la facilitadora de promover interacciones de calidad entre la madre y el niño. Por el contrario, el segundo dominio del HOVRS se concentra en la calidad

de las interacciones entre la madre y el niño—es decir, qué tan receptivas, atentas y cálidas son; y en el nivel de interés y participación de la madre y el niño durante la visita.

Las Figuras 5 y 6 reportan la distribución de puntajes de HOVRS para las 176 facilitadoras cuyas visitas fueron filmadas y codificadas con este instrumento. Se diferencia en tres colores diferentes las zonas de la distribución correspondientes a puntajes de calidad inadecuada, adecuada y buena.

La Figura 5 se concentra en el primer dominio: Prácticas de la visitadora domiciliaria (o la facilitadora, en el caso de SAF). En las cuatro escalas que conforman este dominio, destaca una mayor concentración de los puntajes en la zona azul, de calidad adecuada. En dos de las cuatro escalas (Receptividad hacia la familia y No intrusión/colaboración) no hay ni una sola facilitadora en la muestra con puntajes en la región de la distribución de calidad buena. En Receptividad hacia la familia, una densidad considerable del histograma se concentra en la zona de calidad inadecuada. En las dos dimensiones restantes (Relación con la familia y Facilitación de la interacción entre cuidador y niño), existe un número minúsculo de facilitadoras que puntúan en la zona de buena calidad. Para Relación con la familia, Facilitación de la interacción entre cuidador y niño y No intrusión/colaboración, la proporción de facilitadoras observadas en la región de calidad inadecuada es baja.

Los puntajes de HOVRS sugieren que, en general, las facilitadoras del Programa tienen mucho margen de mejora en lo que se refiere a su nivel de preparación para las visitas. Por ejemplo, los bajos puntajes en la primera escala del HOVRS, Receptividad hacia la familia, revelan que las facilitadoras pueden mejorar en la planeación de la visita, en su capacidad de interesarse y preguntar a la familia por sus intereses y también su habilidad de tomar en cuenta esta información y de responder a ella. De igual manera, los bajos puntajes en la escala de Facilitación de la interacción entre cuidador y niño muestran que el Programa necesita fortalecer la capacidad de las facilitadoras de empoderar a las madres en su rol y en sus interacciones con sus hijos, de responder a las interacciones entre madre y niño y de emplear los materiales del entorno y aquellos que ella ha traído a la visitas para estimular más y mejores interacciones. Para obtener más puntajes de calidad buena en el área de Relación con la familia, el Programa podría fortalecer las capacidades de los facilitadores para desarrollar relaciones cálidas, de confianza y respetuosas con la familia, demostrando interacciones positivas no solo con la madre y el niño que son el foco de la visita pero también con otros miembros que estén presentes durante ella. También parece ser importante trabajar en las habilidades de las facilitadoras para desarrollar estrategias colaboradoras con las madres, sin que éstas se perciban intrusivas o rígidas (No intrusión/colaboración).

#### **[Figura 5. Puntajes del índice del HOVRS “Prácticas de la facilitadora” (N= 176)]**

La Figura 6 se concentra en el segundo dominio del instrumento, Involucramiento de la familia. Destaca que en este dominio se observa una mayor variabilidad de los puntajes y también, para las tres escalas que lo comprenden, una mayor densidad de la distribución de calidad en las zonas de calidad adecuada y buena (y, por consiguiente, muy pocas observaciones de calidad inadecuada). El área en la cual se observa un mayor número de facilitadoras que puntúan en la zona de calidad buena es la de Involucramiento del niño. Por el contrario, las áreas de Interacción cuidador-niño e Involucramiento del cuidador se parecen más a las del primer dominio y concentran la mayor cantidad de observaciones en la zona de calidad adecuada de la distribución.

Para el caso de este dominio, es claro que las visitas logran atraer la participación y el interés de los niños (Involucramiento del niño). Sin embargo, al igual que se identificó en el dominio anterior, existe margen de mejora en dos escalas. Por ejemplo, se podría fortalecer las visitas de una manera que éstas demuestren interacciones de mayor calidez, atención y receptividad entre la madre y el niño (Interacción cuidador-niño). También se podría promover una participación más cercana y activa de la madre en la visita, respondiendo a su interés (o falta de él), involucrándola e incluso ofreciéndole un rol físicamente más próximo al niño durante la visita (Involucramiento del cuidador).

## [Figura 6. Puntajes del índice del HOVRS “Involucramiento de la familia” (N= 176)]

Como un complemento al análisis presentado, es posible explorar en qué medida los niveles de calidad de las visitas se encuentran estadísticamente asociados a los resultados del SAF. En esta sección, se presentan los resultados de una regresión que busca estimar precisamente esta asociación. Como variable dependiente, se incluyen los puntajes estandarizados para las áreas de desarrollo con impactos más significativos y robustos en este ejercicio, es decir, resolución de problemas y comunicación (en ASQ-3) y cognitiva y de lenguaje receptivo (en Bayley-III). Como controles se incluye el sexo y la edad del niño, la educación de la madre, una variable dicotómica que indica si la madre es indígena, el índice de riqueza, algunas características de la facilitadora (su nivel educativo y edad) y las variables dicotómicas que representan los identificadores de los entrevistadores o sicólogos que administraron las pruebas. Los puntajes de HOVRS son netos del efecto de codificador, es decir, se trabaja con los residuos de una regresión de los puntajes crudos del HOVRS sobre los identificadores de los codificadores y se calcula el promedio de ellos entre las dos codificaciones.

La variable de calidad de la visita que se le asigna a cada niño no corresponde a la calidad de la visita observada y codificada para *ese niño*. La estimación de una regresión con esta medida de calidad puede sufrir de dos problemas potenciales. Primero, por tratarse de una sola observación de la calidad de las visitas que produce esa facilitadora, puede estar medida de manera demasiado imprecisa y ser incapaz de capturar cualquier asociación estadística entre la calidad de las visitas y el desarrollo infantil. Una alternativa sería calcular el promedio de *todas* las visitas observadas y codificadas para esa facilitadora. Como se comentó antes, el ejercicio de medición de calidad buscó medir al menos tres visitas por cada una de las facilitadoras de la muestra. Sin embargo, al incluir en ese promedio también el puntaje de la visita del niño cuya medida de desarrollo es la variable independiente de esta estimación, se corre un potencial problema de sesgo por variables no observadas. Es muy probable que algunas características no observadas de ese niño (como su temperamento o la sensibilidad de su madre para reconocer las necesidades del niño y responder a ellas) puedan estar asociadas tanto a su nivel de desarrollo como a la facilidad con la que se desempeña la visita y en consecuencia su calidad. Por esta razón, la medida de la calidad de la visita que se incluye en esta estimación corresponde al promedio para la facilitadora del niño cuyo desarrollo se mide como variable dependiente, pero excluyendo del promedio la medida de calidad para la vista de *ese niño en particular*. Los puntajes de HOVRS se expresan como puntajes Z con desviación estándar 1 y promedio 0 para facilitar su interpretación.

Los resultados de la estimación descrita en los párrafos anteriores se presentan en la Tabla 13. En esta regresión se observa que el primer dominio del HOVRS, Prácticas de la facilitadora, tiene un asociación significativa al 5% con resolución de problemas. El coeficiente de Involucramiento de la familia es significativo al 1%. Específicamente, un aumento de una desviación estándar en la calidad de las visitas en términos de las Prácticas de la Facilitadora se asocia con un incremento de 0,103 desviaciones estándar en resolución de problemas. Un incremento de una desviación estándar en la calidad de las visitas en Involucramiento de la Familia se asocia con un incremento de 0,169 desviaciones estándar en resolución de problemas. En el primer dominio, Prácticas de la facilitadora, el impacto proviene principalmente del coeficiente más significativo para Facilitación de la interacción entre el cuidador y el niño (0,096, significativo al 5%) y algo menos significativo, de Relación con la familia (0,088, significativo al 10%). En el segundo dominio, Involucramiento de la familia, los impactos son estadísticamente significativos al 1% para dos de las tres áreas y al 5% para una de ellas, y de una magnitud parecida, que fluctúa en el rango de entre 0,134 y 0,155 de una desviación estándar.

La estimación que emplea como variable independiente el puntaje de comunicación tiene menos coeficientes significativos. No obstante, destaca que el puntaje del segundo dominio, Involucramiento de la familia, es de 0,103 desviaciones estándar y significativo al 5% y que se encuentran coeficientes significativos al 10% para Interacción cuidador-niño e Involucramiento del cuidador (de una magnitud de 0,105 y 0,104 desviaciones estándar, respectivamente). Un patrón de coeficientes bastante parecido al de la regresión con resolución de problemas como

variable dependiente se encuentra en la estimación que emplea el puntaje total del ASQ-3 como variable dependiente (no reportada pero disponible para el lector interesado). Las estimaciones con los puntajes de Bayley-III revelan resultados menos interesantes. Probablemente por ser una muestra más pequeña, o por la composición etaria de esta sub-muestra, no se logra detectar estadísticamente un impacto significativo de la calidad de las visitas ni sobre el desarrollo cognitivo ni sobre el lenguaje receptivo. Hay dos excepciones, se estima una asociación de 0,092 y significativa al 5% entre el Involucramiento del niño en la visita y su desarrollo cognitivo y también una asociación estadísticamente significativa al 10% de 0,095 entre Facilitación de la interacción entre cuidador y niño y lenguaje receptivo.

En resumen, el análisis de la calidad de las vistas de SAF sugiere que ésta es un elemento importante para lograr cambios en la trayectoria del desarrollo infantil. Al mismo tiempo, el Programa tiene un amplio margen de mejora en esta área, tanto en las dimensiones que implican una mejor preparación de la facilitadora como en su capacidad de lograr un mejor involucramiento de la madre y el niño durante la visita.

#### **[Tabla 13. Asociación entre la calidad de la visita y el desarrollo infantil]**

Además de los datos del HOVRS, para estas 554 visitas domiciliarias que fueron filmadas, se recogieron dos medidas adicionales para documentar su calidad:

1. La misma persona que codificó HOVRS también codificó la visita en base a un protocolo de observación del contenido de la visita, el *Home Visit Content and Characteristics Form*, HVCCF (Boller et al., 2009). Este instrumento documenta el contenido de una visita domiciliaria—por ejemplo, aspectos como el uso de tiempo, elementos de distracción, las actividades realizadas durante la misma y las características de los participantes, tales como el número de adultos y niños presentes, lengua, fortalezas y desafíos de la familia. Diseñado originalmente para ser administrado como un complemento al HOVRS, también se puede aplicar en forma independiente. Al igual que el HOVRS, se puede administrar por observación directa o a partir de una filmación.
2. Además, durante estas 554 filmaciones, estuvo presente en la visita la Acompañante Técnico de SAF a cargo de esa familia y facilitadora, quien durante la visita rellenó una ficha de chequeo diseñada pensando en el monitoreo y acompañamiento continuo de este tipo de Programa. Este instrumento fue construido a partir de las fichas de supervisión que usan el SAF y otros programas de visitas domiciliarias.

El análisis de los datos del HVCCF y de la lista de chequeo, sobre qué documentan éstos con respecto a la calidad de las visitas, cómo se asocian con los niveles de desarrollo infantil y en qué medida las diferentes medidas de calidad recogidas a través de estos distintos instrumentos se relacionan entre sí se documentarán en un estudio complementario a la presente evaluación de impacto. A la fecha de elaboración de este informe, el análisis de estos datos se encuentra todavía en una etapa inicial.

#### **9. Conclusiones**

Este documento resume los resultados de la evaluación de impacto del Servicio de Acompañamiento a Familias del Programa Nacional Cuna Más. Se trata de la primera evaluación experimental de un programa de visitas domiciliarias para la promoción del desarrollo infantil implementado a una escala como la del SAF.

La recolección de los datos aquí analizados se llevó a cabo en 2013 y en 2015, es decir, antes de la implementación del SAF y dos años después. La evaluación aprovechó de la expansión gradual del SAF para utilizar una asignación aleatoria al tratamiento a nivel de distrito.

Los impactos se estiman en dos muestras. La muestra de la evaluación comprende 5.339 niños que, a línea de base, tenían entre 1 y 24 meses de edad y que recibirían el SAF durante un mínimo de 12 meses. También se estiman los impactos del SAF sobre una sub-muestra de 1.038 niños a

quienes, en línea de seguimiento, fue posible evaluar usando una escala diagnóstica de medición del desarrollo muy rigurosa.

La evaluación de impacto muestra que el SAF tiene impactos significativos sobre el desarrollo, en las áreas cognitiva, de comunicación de motricidad fina y personal-social, siendo los más robustos y significativos los de las dos primeras áreas. El impacto del SAF sobre los niños que recibieron al menos una visita del Programa es de una magnitud de entre 0,094 y 0,122 desviaciones estándar para las áreas en las cuales éste es significativo. En la sub-muestra de niños—más pequeños, indígenas en menor proporción y que recibieron una mayor intensidad de tratamiento—y a quienes se administró una prueba diagnóstica, se identifican impactos sobre desarrollo cognitivo y lenguaje receptivo de 0,295 y 0,193 desviaciones estándar, respectivamente. En la sub-muestra, los impactos del SAF son de mayor magnitud entre niños de hogares más pobres y con madres con menos años de escolaridad.

Cuando se exploran algunos posibles mediadores de los impactos del SAF, se identifica que una mayor frecuencia de actividades de juego, una mayor disponibilidad de juguetes caseros, menos prácticas violentas de disciplina y una mayor tasa de matrícula en educación inicial entre los niños intervenidos.

Dada su escala y velocidad de implementación, la evaluación también documenta que el SAF tiene importantes áreas de mejora en materia operativa. Por ejemplo, entre los niños que debieron recibir el SAF en la muestra de la evaluación, apenas 64% recibieron el servicio. Incluso entre quienes reciben el servicio, existe heterogeneidad en el número efectivo de visitas recibidas.

El SAF ha logrado resultados importantes, pero necesita continuar consolidando la calidad de las visitas domiciliarias que entrega, en particular a través de una mejor preparación de la facilitadora y en su capacidad de lograr un mayor involucramiento de la madre y el niño en las actividades de la visita.

## Referencias

- Ángeles, G., GaDEen, P., Galiani, S., Gertler, P., Herrera, A., Kariger, P. y Seira, E. (2011). “Evaluacion de Impacto del Programa Estancia Infantiles para Apoyar a Madres Trabajadoras. Informe Final de la Evaluación de Impacto.” *Instituto Nacional de Salud Pública*. México.
- Attanasio O., Fernández C, Fitzsimons E, Grantham-McGregor S, Meghir C, Rubio-Codina M. (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
- Attanasio O., Fernández C, Fitzsimons E, Grantham-McGregor S, Meghir C, Rubio-Codina M. (2013) “Enriching the home environment of low-income families in Colombia: a strategy to promote child development at scale”. *Early Childhood Matters*. Van Leer Foundation.
- Berlinski, S. y Schady, N. (eds.) (2015). *The Early Years: Child Well-being and the Role of Public Policy Development in the Americas Series*. New York: Palgrave MacMillan y Washington, DC: Inter-American Development Bank.
- Bernal, R. (2015). “The Impact of a Vocational Education Program for Childcare Providers on Children’s Well-being”. Forthcoming in *Economics of Education Review*.
- Bernal, R. y X. Peña (2010). “Guía práctica para la evaluación de impacto”, Universidad de Los Andes, Facultad de Economía: Bogotá, Colombia.
- Boller, K., Vogel, C., Cohen, R., Aikens, N., and Hallgren, K. (2009). *Home Visit Characteristics and Content Form*. Princeton, NJ: Mathematica Policy Research
- Bradley, R.H. (1993): “Children’s Home Environments, Health, Behavior, and Intervention Efforts: A Review Using the HOME Inventory as a Marker Measure.” *Genetic, Social, and General Psychology Monographs*, 119: 437-490.
- Bruhn, M. y D. McKenzie. (2009). “In Pursuit of Balance: Randomization in Practice in Development Field Experiments”, *American Economic Journal: Applied Economics* 2009, 1:4, 200-232.
- Caldwell, B., y Bradley, R. (2003). Home Observation for Measurement of the Environment: Administration manual. Little Rock, AR: University of Arkansas.
- Currie, J.; Stabile, M.; Manivong, P. y Roos, L. (2010). Child Health and Young Adult Outcomes. *The Journal of Human Resources*, 45(3):517-548.
- Engle, P.; Fernald, L.; Alderman, H.; Behrman, J.; O’Gara, C.; Yousafzai, A. et al. (2011). Strategies for reducing inequalities y improving developmental outcomes for young children in low and middle income countries. *The Lancet*, 378(9799):1339-1353.
- 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 interno del MIDIS.
- Fernald, L., Kariger, P., Hidrobo, M. y Gertler, P. (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.

- Fernald, L.C.H., Kariger, P.K., Engle, P. y Raikes, A. (2009). *Examining child development in low-income countries: a toolkit for the assessment of children in the first five years of life*. Washington D.C.: The World Bank.
- Frongillo, E. A., Tofail, F., Hamadani, J. D., Warren, A. M. and Mehrin, S. F. (2014), Measures and indicators for assessing impact of interventions integrating nutrition, health, and early childhood development. *Annals of the New York Academy of Sciences*, 1308: 68–88. doi: 10.1111/nyas.12319
- Frongillo, E.A., Sywulka, S.M. y Kariger, P. (2003). “UNICEF psychosocial care indicators project. Final report to UNICEF”. mimeo Cornell University.
- Gertler, Paul, James Heckman, Rodrigo Pinto, Arianna Zanolini, Christel Vermeersch, Susan Walker, Susan M. Chang, and Sally Grantham-McGregor. (2014). “Labor market returns to an early childhood stimulation intervention in Jamaica”. *Science* 30 May 2014: 344 (6187), 998-1001.
- Grantham-McGregor, S.M., C.A. Powell, S.P. Walker, y J.H. Himes. (1991). “Nutritional supplementation, psychosocial stimulation, and mental development of stunted children: the Jamaican Study”, *The Lancet*, Vol 338, Issue 8758, pp. 1-5.
- Grantham-McGregor, S.M., S.P. Walker, S.M. Chang y C.A. Powell. (1997). “Effects of early childhood supplementation with and without stimulation on later development in stunted Jamaican children”, *The American Journal of Clinical Nutrition*, 66(2), pp. 247-53.
- Hamadani, J.D., Tofail, F., Huda, S.N, F., Alam, D.S., Ridout, D.A., Attanasio, O. y Grantham-McGregor, S. (2014). “Cognitive deficit and poverty in the first 5 years of childhood in Bangladesh”. *Pediatrics*, 134(4):e1001-e1008.
- Lazarte, F. (2016). *Informe final de las actividades realizadas para el recojo de información del estudio de Bayley III*. Reporte de consultoría no publicado.
- Lopez-Boo, F. (2014). “Socio-Economics Status and Early Childhood Cognitive Skills: Is Latin America Different?”. Young Lives Working Paper 127.
- MacPhee, D. (2002). *Knowledge of Infant Development Inventory Survey of Child Care Experiences Manual*. Mimeo.
- Nahar B, Hossain MI, Hamadani JD, Ahmed T, Huda SN, Grantham-McGregor SM, (2012). “Effects of a community-based approach of food and psychosocial stimulation on growth and development of severely malnourished children in Bangladesh: a randomised trial”. *Eur J Clin Nutr*, 66:701-9
- Paxson, C. y Schady, N. (2007). Cognitive Development among Young Children in Ecuador: The Roles of Wealth, Health, and Parenting. *Journal of Human Resources*, (42)1: 49-84.
- Roggman, L. A., Cook, G. A., Innocenti, M. S., Jump Norman, V. K., Christiansen, K., Boyce, L. K., Aikens, N., Boller, K., Paulsell, D., & Hallgren, K. (2010). *Home Visit Rating Scales—Adapted and Extended (HOVRS-A+)* version 2. Adapted from Roggman, L. A., et al. (2008). Home Visit Rating Scales. In L. Roggman, L. Boyce, and M. Innocenti, *Developmental Parenting* (pp. 209-217). Baltimore: Paul H. Brookes Publishing.
- Rubio-Codina, M. Tomé, R. y Araujo, M.C. (2016). *Los primeros años de vida de los niños peruanos: Una fotografía sobre el bienestar y el desarrollo de los niños del Programa Nacional Cuna Más*. Nota Técnica del Banco Interamericano de Desarrollo. IDB-TN-1093. Septiembre.

Rubio-Codina, M.; Attanasio, O.; Meghir, C.; Varela, N. y Grantham- McGregor, S. (2015). The Socioeconomic Gradient of Child Development: Cross- Sectional Evidence from Children 6–42 Months in Bogota. *Journal of Human Resources*, 50(2): 464-483.

Rubio-Codina, M.; Attanasio, O. y Grantham-McGregor, S. (2016b). Mediating Pathways in the Socio-Economic Gradient of Child Development: Evidence from Children 6-42 Months in Bogota. *International Journal of Behavioural Development*,

Schady, N.; Behrman, J.; Araujo, M.C.; Azuero, R.; Bernal, R.; Bravo, D.; Lopez-Boo, F.; Macours, K.; Marshall, D.; Paxson, C. y Vakis, R. (2015). Wealth gradients in early childhood cognitive development in five Latin American Countries. *Journal of Human Resources*, 50(2): 446-463.

Shonkoff, J.P.; Boyce, W.T. y McEwen, B.S. (2009). Neuroscience, molecular biology, and the childhood roots of health disparities: Building a new framework for health promotion and disease prevention. *JAMA*, 301:2252–2259.

Squires J., Twombly E., Bricker D., y Potter L. (2009). Ages and stages questionnaires (ASQ): A parent-completed child monitoring system (3rd ed.). Baltimore, MD: Brookes.

Tofail, F., Hamadani, J.D., Mehrin, F., Ridout, D.A., Huda, S.N. y Grantham-McGregor, S. (2013). “Psychosocial Stimulation Benefits Development in Nonanemic Children but Not in Anemic, Iron-Deficient Children”, *The Journal of Nutrition*, doi: 10.3945/jn.112.160473.

Walker S.P., S.M. Grantham-McGregor, C.A. Powell, S.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”, *Journal of Pediatrics*, 137(1): 36-41.

Walker S.P., S.M. Chang, C.A. Powell, y S.M. Grantham-McGregor. (2005). “Effects of early childhood psychosocial stimulation and nutritional supplementation on cognition and education in growth-stunted Jamaican children: prospective cohort study”, *The Lancet*, Vol. 366, Issue 9499, pp. 1804-1807.

Walker, S.; Wachs, T.; Grantham-McGregor, S.; Black, M.; Nelson, C.; Huffman, S.; Baker-Henningham, H.; Chang, S.; Hamadani, J.; Lozoff, B.; Meeks Gardner, J.; Powell, C.; Rahman, A. y Richter, L. (2011). Inequality in early childhood: risk and protective factors for early child development. *The Lancet*, 378(9799):1325-1338.

## Tablas y Figuras

**Figura 1. Resultados e indicadores asociados al desempeño del Programa**

OBJETIVOS		INDICADORES DE DESEMPEÑO
Resultado Final	Niños y niñas con competencias básicas	Los y las estudiantes de II ciclo-EBR obtienen los logros de aprendizaje esperados en Comunicación Integral y Pensamiento Lógico Matemático
Resultado Específico	Mejora en el nivel de desarrollo infantil de los niños y niñas menores de 36 meses que viven en situación de pobreza y pobreza extrema	Porcentaje de niños y niñas menores de 36 meses de edad que logran los hitos esperados para su edad en las dimensiones (i) física (motricidad fina y gruesa), (ii) cognitiva (resolución de problemas), (iii) lenguaje (comunicación pre verbal y verbal) y (iv) socioemocional (relación con otros y autonomía básica)

*Fuente:* Anexo 2 del Programa Presupuestal PNCM.

**Figura 2. Prácticas principales del Programa**

Tipo	Práctica
Cuidado Saludable	(i) Alimentación: Lactancia materna exclusiva, alimentación complementaria, consumo de micronutrientes (ii) Higiene y Aseo: lavado de manos y consumo de agua segura (iii) Asistencia oportuna al control de CRED
Aprendizaje	(i) Comunicación (ii) Interacción (iii) Juego

**Tabla 1. Características y balance de la muestra de la evaluación (N=5.339)**

	Muestra completa			Control			Tratamiento			P-valor diferencia
	N	Media	DE	N	Media	DE	N	Media	DE	
<b><i>Características del niño</i></b>										
Edad (meses)	5.339	12,786	6,662	1.809	12,701	6,637	3.530	12,830	6,676	0,446
Proporción de mujeres	5.339	0,492	0,500	1.809	0,491	0,500	3.530	0,493	0,500	0,961
Desnutrición crónica	5.186	0,376	0,484	1.745	0,380	0,486	3.441	0,373	0,484	0,714
Desnutrición global	5.190	0,081	0,273	1.746	0,085	0,279	3.444	0,079	0,270	0,450
ASQ resolución de problemas	5.335	46,037	17,641	1.809	47,430	18,142	3.526	45,322	17,337	0,007
ASQ comunicación	5.335	43,538	18,658	1.809	43,231	18,502	3.526	43,695	18,739	0,458
ASQ motricidad fina	5.335	44,612	16,564	1.809	46,639	17,411	3.526	43,572	16,016	0,000
ASQ personal-social	5.335	45,386	15,890	1.809	46,462	16,063	3.526	44,834	15,774	0,011
ASQ motricidad gruesa	5.335	43,122	18,234	1.809	43,402	18,141	3.526	42,978	18,283	0,522
<b><i>Características del cuidador y del hogar</i></b>										
Años de educación (madre)	5.065	6,573	4,076	1.707	6,293	4,032	3.358	6,715	4,091	0,038
Proporción de indígena (madre)	5.339	0,110	0,313	1.809	0,159	0,365	3.530	0,085	0,279	0,012
Índice de riqueza	5.339	0,027	0,997	1.809	-0,066	0,985	3.530	0,074	1,000	0,058
Proporción con televisor	5.339	0,496	0,500	1.809	0,454	0,498	3.530	0,517	0,500	0,023
Proporción con refrigeradora	5.339	0,078	0,267	1.809	0,072	0,259	3.530	0,080	0,272	0,507
Proporción con cocina a gas	5.339	0,411	0,492	1.809	0,371	0,483	3.530	0,431	0,495	0,070
Proporción con celular	5.339	0,599	0,490	1.809	0,587	0,492	3.530	0,606	0,489	0,537
Proporción con piso distinto de tierra	5.339	0,259	0,438	1.809	0,258	0,438	3.530	0,259	0,438	0,971
Proporción con electricidad	5.339	0,763	0,426	1.809	0,745	0,436	3.530	0,771	0,420	0,432
Proporción con agua (dentro de la vivienda)	5.339	0,592	0,492	1.809	0,585	0,493	3.530	0,595	0,491	0,840
Proporción con servicio higiénico (dentro de la vivienda)	5.339	0,248	0,432	1.809	0,205	0,403	3.530	0,270	0,444	0,033
<b><i>Características de los centros poblados</i></b>										
Tiempo de traslado a la capital distrital a pie (horas)	166	1,359	2,369	44	1,778	3,369	122	1,207	1,879	0,082
Tiempo de traslado a la capital distrital en auto (horas)	168	1,089	0,948	62	1,090	0,946	106	1,088	0,953	0,524
Tiene centro o posta de salud	360	0,831	0,376	120	0,800	0,402	240	0,846	0,362	0,184
Tiene centro de educación inicial del MinEd	360	0,944	0,229	120	0,958	0,201	240	0,938	0,243	0,317
Recibe Vaso de leche	360	0,978	0,148	120	0,975	0,157	240	0,979	0,143	0,764
Recibe Comedor Popular	360	0,383	0,487	120	0,383	0,488	240	0,383	0,487	1,000
Recibe desayuno escolar Qali Warma	359	0,967	0,180	119	0,975	0,157	240	0,962	0,190	0,468
Recibe almuerzo escolar Qali Warma	357	0,891	0,312	119	0,882	0,324	238	0,895	0,307	0,647
Tiene Servicio de Cuidado Diurno de Cuna Más	360	0,114	0,318	120	0,108	0,312	240	0,117	0,322	0,814
Recibe Beca 18	360	0,564	0,497	120	0,575	0,496	240	0,558	0,498	0,729
Recibe Juntos	360	1,000	0,000	120	1,000	0,000	240	1,000	0,000	-
Recibe Pensión 65	360	0,975	0,156	120	0,967	0,180	240	0,979	0,143	0,366
Tiene farmacia	360	0,394	0,489	120	0,375	0,486	240	0,404	0,492	0,529
Tiene banco o caja municipal/rural	360	0,119	0,325	120	0,117	0,322	240	0,121	0,327	0,887
Tiene comisaría	359	0,181	0,386	119	0,151	0,360	240	0,196	0,398	0,164
Agricultura es una de 3 actividades principales	360	0,853	0,355	120	0,892	0,312	240	0,833	0,373	0,082

Notas: Características en la LB de los 5.339 niños que fueron encuestados en la LS y que completaron el ASQ-3. El indicador de tratamiento es igual a 1 si el niño fue asignado al grupo de tratamiento, 0 de lo contrario. Todas las estimaciones incluyen un efecto fijo del trío de distritos según el cual se hizo la asignación aleatoria del tratamiento y corrigen los errores estándar a nivel de distrito. La proporción de madres indígenas corresponde a los hogares que respondieron la encuesta de LB en una lengua indígena. El índice de riqueza se construye a partir de las características de infraestructura de la vivienda y activos del hogar presentados en la tabla, usando componentes principales. Las características de los centros poblados se refieren a los 360 centros de la muestra de la evaluación (120 en control y 240 en tratamiento). El tiempo de traslado, a pie y en auto, solo incluye aquellos centros poblados para los cuales estos son los métodos de transporte más frecuentes.

**Tabla 2. Características y balance de la sub-muestra de Bayley (N=1.038)**

	Muestra completa			Control			Tratamiento			P-valor diferencia
	N	Media	DE	N	Media	DE	N	Media	DE	
<b>Características del niño</b>										
Edad (meses)	1.038	7,867	3,978	343	7,708	3,789	695	7,945	4,069	0,254
Proporción de mujeres	1.038	0,493	0,500	343	0,534	0,500	695	0,473	0,500	0,004
Desnutrición crónica	1.000	0,341	0,474	325	0,314	0,465	675	0,354	0,479	0,220
Desnutrición global	1.000	0,061	0,239	325	0,068	0,252	675	0,058	0,233	0,555
ASQ resolución de problemas	1.038	45,626	17,956	343	46,822	18,236	695	45,036	17,799	0,235
ASQ comunicación	1.038	48,015	15,950	343	48,396	16,549	695	47,827	15,654	0,721
ASQ motricidad fina	1.038	45,703	17,457	343	48,309	18,005	695	44,417	17,048	0,015
ASQ en personal-social	1.038	43,362	15,637	343	45,175	15,746	695	42,468	15,517	0,008
ASQ motricidad gruesa	1.038	39,075	17,641	343	40,102	17,842	695	38,568	17,531	0,287
<b>Características del cuidador y del hogar</b>										
Años de educación (madre)	994	6,696	4,106	329	6,520	4,259	665	6,783	4,029	0,432
Proporción de indígena (madre)	1.038	0,013	0,115	343	0,012	0,108	695	0,014	0,119	0,579
Índice de riqueza	1.038	0,036	0,975	343	-0,039	0,972	695	0,073	0,975	0,418
Proporción con televisor	1.038	0,496	0,500	343	0,475	0,500	695	0,506	0,500	0,558
Proporción con refrigeradora	1.038	0,056	0,230	343	0,061	0,240	695	0,053	0,225	0,479
Proporción con cocina a gas	1.038	0,406	0,491	343	0,370	0,484	695	0,423	0,494	0,338
Proporción con celular	1.038	0,651	0,477	343	0,668	0,472	695	0,643	0,479	0,645
Proporción con piso distinto de tierra	1.038	0,250	0,433	343	0,280	0,450	695	0,235	0,424	0,367
Proporción con electricidad	1.038	0,752	0,432	343	0,738	0,441	695	0,760	0,428	0,746
Proporción con agua (dentro de la vivienda)	1.038	0,620	0,486	343	0,534	0,500	695	0,663	0,473	0,043
Proporción con servicio higiénico (dentro de la vivienda)	1.038	0,273	0,446	343	0,254	0,436	695	0,282	0,450	0,690
<b>Características de los centros poblados</b>										
Tiempo de traslado a la capital distrital a pie (horas)	57	1,123	2,085	13	0,801	1,164	44	1,218	2,290	0,255
Tiempo de traslado a la capital distrital en auto (horas)	56	1,104	0,957	21	1,258	0,985	35	1,012	0,942	0,315
Tiene centro o posta de salud	120	0,808	0,395	39	0,769	0,427	81	0,827	0,380	0,414
Tiene centro de educación inicial del MinEd	120	0,950	0,219	39	0,974	0,160	81	0,938	0,242	0,177
Recibe Vaso de leche	120	0,992	0,091	39	1,000	0,000	81	0,988	0,111	0,316
Recibe Comedor Popular	120	0,375	0,486	39	0,333	0,478	81	0,395	0,492	0,548
Recibe desayuno escolar Qali Warma	120	0,975	0,157	39	0,974	0,160	81	0,975	0,156	1,000
Recibe almuerzo escolar Qali Warma	120	0,933	0,250	39	0,897	0,307	81	0,951	0,218	0,204
Tiene Servicio de Cuidado Diurno de Cuna Más	120	0,050	0,219	39	0,077	0,270	81	0,037	0,190	0,318
Recibe Beca 18	120	0,583	0,495	39	0,538	0,505	81	0,605	0,492	0,368
Recibe Juntos	120	1,000	0,000	39	1,000	0,000	81	1,000	0,000	-
Recibe Pensión 65	120	0,983	0,129	39	0,974	0,160	81	0,988	0,111	0,565
Tiene farmacia	120	0,442	0,499	39	0,462	0,505	81	0,432	0,498	0,705
Tiene banco o caja municipal/rural	120	0,075	0,264	39	0,051	0,223	81	0,086	0,283	0,387
Tiene comisaría	119	0,134	0,343	38	0,053	0,226	81	0,173	0,380	0,017
Agricultura es una de 3 actividades principales	120	0,842	0,367	39	0,872	0,339	81	0,827	0,380	0,526

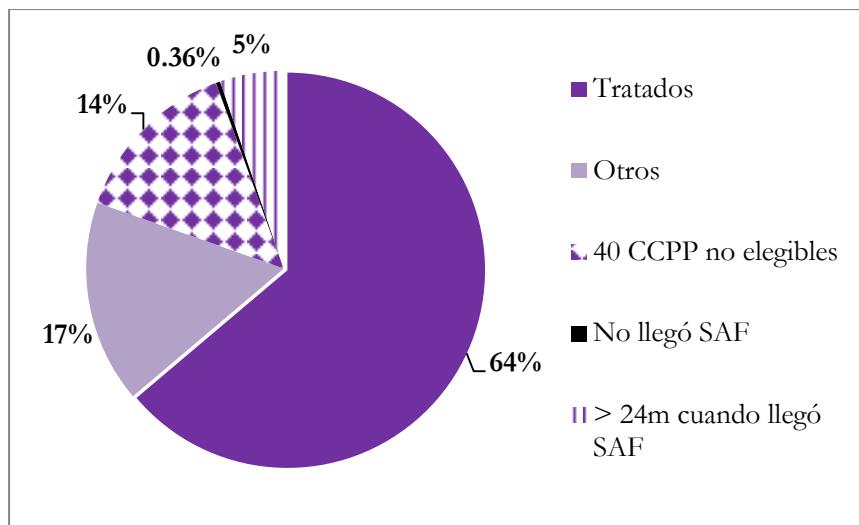
Notas: Características en la LB de los 1.038 niños que completaron la prueba de Bayley-III. El indicador de tratamiento es igual a 1 si el niño fue asignado al grupo de tratamiento, 0 de lo contrario. Todas las estimaciones incluyen un efecto fijo del trío de distritos según el cual se hizo la asignación aleatoria del tratamiento y corregir los errores estándar a nivel de distrito. La proporción de madres indígenas corresponde a los hogares que respondieron la encuesta de LB en una lengua indígena. El índice de riqueza se construye a partir de las características de infraestructura de la vivienda y activos del hogar presentados en la tabla, usando componentes principales. Las características de los centros poblados se refieren a los centros de la muestra de la evaluación en los cuales se encuentran los 1.038 niños de la sub-muestra de Bayley. El tiempo de traslado, a pie y en auto, solo incluye aquellos centros poblados para los cuales estos son los métodos de transporte más frecuentes.

**Tabla 3. Comparación de la muestra de la evaluación y la sub-muestra de Bayley**

	Muestra de la evaluación (N=5.339)			Sub-muestra de Bayley (N=980)			P-valor diferencia
	N	Media	DE	N	Media	DE	
<b>Características del niño</b>							
Edad (meses)	5.339	12,786	6,662	980	7,898	3,972	0,000
Proporción de mujeres	5.339	0,492	0,500	980	0,492	0,500	0,224
Desnutrición crónica	5.186	0,376	0,484	948	0,347	0,476	0,000
Desnutrición global	5.190	0,081	0,273	948	0,058	0,234	0,008
ASQ resolución de problemas	5.335	46,037	17,641	980	45,816	17,855	0,093
ASQ comunicación	5.335	43,538	18,658	980	48,026	15,985	0,000
ASQ motricidad fina	5.335	44,612	16,564	980	45,770	17,428	0,000
ASQ en personal-social	5.335	45,386	15,890	980	43,332	15,646	0,087
ASQ motricidad gruesa	5.335	43,122	18,234	980	39,056	17,729	0,000
<b>Características del cuidador y del hogar</b>							
Años de educación (madre)	5.065	6,573	4,076	939	6,652	4,093	0,012
Proporción de indígena (madre)	5.339	0,110	0,313	980	0,012	0,110	0,019
Índice de riqueza	5.339	0,027	0,997	980	0,021	0,963	0,052
Proporción con televisor	5.339	0,496	0,500	980	0,496	0,500	0,048
Proporción con refrigeradora	5.339	0,078	0,267	980	0,055	0,228	0,452
Proporción con cocina a gas	5.339	0,411	0,492	980	0,399	0,490	0,121
Proporción con celular	5.339	0,599	0,490	980	0,643	0,479	0,004
Proporción con piso distinto de tierra	5.339	0,259	0,438	980	0,245	0,430	0,141
Proporción con electricidad	5.339	0,763	0,426	980	0,754	0,431	0,283
Proporción con agua (dentro de la vivienda)	5.339	0,592	0,492	980	0,615	0,487	0,118
Proporción con servicio higiénico (dentro de la vivienda)	5.339	0,248	0,432	980	0,268	0,443	0,110

Notas: Características en la LB de los 5.339 niños que fueron encuestados en la LS y que completaron el ASQ-3, y de los 980 niños que fueron encuestados en la LS, que completaron el ASQ-3, y que completaron la prueba de Bayley-III. Todas las estimaciones incluyen un efecto fijo del trío de distritos según el cual se hizo la asignación aleatoria del tratamiento y corrigen los errores estándar a nivel de distrito. La proporción de madres indígenas corresponde a los hogares que respondieron la encuesta de LB en una lengua indígena. El índice de riqueza se construye a partir de las características de infraestructura de la vivienda y activos del hogar presentados en la tabla, usando componentes principales.

**Figura 3. Niños asignados al grupo de tratamiento, según su participación en el Programa (N= 3.870)**



*Notas:* la Figura 3 se refiere a los 3.870 niños de la muestra de la evaluación asignados al grupo de tratamiento. Las proporciones presentan los motivos por los cuales algunos niños no participaron en la intervención (a excepción de la proporción de los 64% de niños tratados). “CP” se refiere a centros poblados y “24m” a 24 meses de edad.

**Tabla 4. Características de los asignados a tratamiento que participan y no participan en el Programa**

	Participan (N=2,469)			No Participan (N=638)			P-valor diferencia
	N	Media	DE	N	Media	DE	
<b><i>Características del niño</i></b>							
Edad (meses)	2,469	12,219	6,612	637	13,496	6,693	0,000
Proporción de mujeres	2,469	0,500	0,500	637	0,487	0,500	0,816
Desnutrición crónica	2,417	0,388	0,488	600	0,322	0,468	0,003
Desnutrición global	2,418	0,075	0,264	601	0,068	0,252	0,656
ASQ resolución de problemas	2,467	45,124	17,381	637	46,162	18,418	0,184
ASQ comunicación	2,467	43,778	18,375	637	44,199	19,342	0,623
ASQ motricidad fina	2,467	43,646	16,303	637	43,776	16,322	0,619
ASQ en personal-social	2,467	44,544	15,648	637	45,495	16,362	0,260
ASQ motricidad gruesa	2,467	42,262	18,111	637	43,375	18,802	0,125
<b><i>Características del cuidador y del hogar</i></b>							
Años de educación (madre)	2,333	6,491	3,901	610	7,810	4,398	0,000
Proporción de indígena (madre)	2,469	0,081	0,272	637	0,064	0,246	0,299
Índice de riqueza	2,469	0,041	0,963	637	0,217	1,093	0,005
Proporción con televisor	2,469	0,507	0,500	637	0,540	0,499	0,164
Proporción con refrigeradora	2,469	0,070	0,255	637	0,102	0,303	0,010
Proporción con cocina a gas	2,469	0,416	0,493	637	0,466	0,499	0,080
Proporción con celular	2,469	0,591	0,492	637	0,666	0,472	0,031
Proporción con piso distinto de tierra	2,469	0,246	0,431	637	0,339	0,474	0,000
Proporción con electricidad	2,469	0,786	0,410	637	0,743	0,438	0,102
Proporción con agua (dentro de la vivienda)	2,469	0,587	0,492	637	0,644	0,479	0,142
Proporción con servicio higiénico (dentro de la vivienda)	2,469	0,246	0,431	637	0,339	0,474	0,010

Notas: Características en la LB de los 2,469 niños de la muestra de evaluación asignados a tratamiento y que adoptaron el Programa, en comparación con los 638 niños que fueron asignados a tratamiento, que no adoptaron el Programa, y para los cuales los motivos de no participación son desconocidos (la categoría “otros” de la Figura 3). Todas las estimaciones incluyen un efecto fijo del trío de distritos según el cual se hizo la asignación aleatoria del tratamiento y corrigen los errores estándar a nivel de distrito. La proporción de madres indígenas corresponde a los hogares que respondieron la encuesta de LB en una lengua indígena. El índice de riqueza se construye a partir de las características de infraestructura de la vivienda y activos del hogar presentados en la tabla, usando componentes principales.

**Figura 4. Número de visitas recibidas, grupo de tratamiento de la muestra de la evaluación (N=3.870)**



Notas: la Figura 4 muestra el rango de número de visitas recibidas por los 3.870 niños de la muestra de la evaluación asignados al grupo de tratamiento, y su equivalente en número de meses de intervención (1 mes corresponde a 4 visitas). Los números en la parte derecha de la Figura se refieren al porcentaje de niños del grupo de tratamiento que se encuentran en cada uno de los rangos del número de visitas.

**Tabla 5. Efecto ITT sobre el ASQ-3 (N=5.339)**

Área	Impacto	EE	P-valor
Resolución de problemas	0,064*	0,032	0,047
Comunicación	0,079*	0,032	0,015
Motricidad fina	0,061+	0,036	0,093
Personal-social	0,066+	0,037	0,077
Motricidad gruesa	-0,008	0,031	0,804
Total (todas las áreas)	0,064*	0,028	0,026
Total (RP+Com+MF)	0,063*	0,024	0,010

Notas: El indicador de tratamiento es igual a 1 si el niño fue asignado a tratamiento, 0 de lo contrario. Los puntajes totales se calcularon por análisis factorial. Los controles incluyen la edad del niño, el sexo, los puntajes en LB de las 5 áreas del ASQ-3, los identificadores de los entrevistadores, la educación de la madre, una variable dicotómica igual a 1 si el hogar respondió la encuesta de LB en una lengua indígena, y el índice de riqueza. Todas las estimaciones incluyen un efecto fijo del trío de distritos según el cual se hizo la asignación aleatoria del tratamiento y corrigen los errores estándar a nivel de distrito. Coeficientes significativos al + p<0.10, \* p<0.05, \*\* p<0.01.

**Tabla 6. Efecto TOT sobre el ASQ-3 (N=5.339)**

Área	Impacto	EE	P-valor
<b>Tratamiento: igual a 1 si el niño recibe al menos una visita del SAF</b>			
Resolución de problemas	0,098*	0,048	0,042
Comunicación	0,122*	0,049	0,013
Motricidad fina	0,094+	0,055	0,089
Personal-social	0,102+	0,057	0,074
Motricidad gruesa	-0,012	0,047	0,801
Total (todas las áreas)	0,098*	0,043	0,024
Total (RP+Com+MF)	0,097**	0,037	0,009
<b>Tratamiento: el número de visitas recibidas por el niño</b>			
Resolución de problemas	0,001*	0,001	0,042
Comunicación	0,002*	0,001	0,013
Motricidad fina	0,001+	0,001	0,090
Personal-social	0,002+	0,001	0,074
Motricidad gruesa	0,000	0,001	0,801
Total (todas las áreas)	0,001*	0,001	0,025
Total (RP+Com+MF)	0,001**	0,001	0,009

*Notas:* La parte superior de la tabla define el tratamiento como una variable dicotómica de valor 1 si el niño recibió alguna visita y 0 si no lo hizo. La parte inferior define el tratamiento como el número de visitas efectivamente recibido. Los puntajes totales se calcularon por análisis factorial. Los controles incluyen la edad del niño, el sexo, los puntajes d LB en las 5 áreas del ASQ-3, los identificadores de los entrevistadores, la educación de la madre, una variable dicotómica igual a 1 si el hogar respondió la encuesta de LB en una lengua indígena, y el índice de riqueza. Todas las estimaciones incluyen un efecto fijo del trío de distritos según el cual se hizo la asignación aleatoria del tratamiento y corrigen los errores estándar a nivel de distrito. Coeficientes significativos al + p<0.10, \* p<0.05, \*\* p<0.01.

**Tabla 7. Comparación del efecto ITT sobre el ASQ-3, para la muestra de la evaluación y la sub-muestra de Bayley**

Área	Muestra de la evaluación (N=5.339)			Sub-muestra de Bayley (N=980)		
	Impacto	EE	P-valor	Impacto	EE	P-valor
Resolución de problemas	0,064*	0,032	0,047	0,177**	0,062	0,006
Comunicación	0,079*	0,032	0,015	0,078	0,059	0,185
Motricidad fina	0,061+	0,036	0,093	-0,035	0,064	0,586
Personal-social	0,066+	0,037	0,077	0,042	0,074	0,577
Motricidad gruesa	-0,008	0,031	0,804	0,006	0,070	0,937
Total (todas las áreas)	0,064*	0,028	0,026	0,070	0,053	0,194
Total (RP+Com+MF)	0,063*	0,024	0,010	0,075+	0,043	0,087

*Notas:* El indicador de tratamiento es igual a 1 si el niño fue asignado a tratamiento, 0 de lo contrario. Los puntajes totales se calcularon por análisis factorial. Los controles incluyen la edad del niño, el sexo, los puntajes de LB en las 5 áreas del ASQ-3, los identificadores de los entrevistadores, la educación de la madre, una variable dicotómica igual a 1 si el hogar respondió la encuesta de LB en una lengua indígena, y el índice de riqueza. Todas las estimaciones incluyen un efecto fijo del trío de distritos según el cual se hizo la asignación aleatoria del tratamiento y corrigen los errores estándar a nivel de distrito. Coeficientes significativos al + p<0.10, \* p<0.05, \*\* p<0.01.

**Tabla 8. Efecto ITT sobre el Bayley-III (N=1.038)**

Área	Impacto	EE	P-valor
Desarrollo cognitivo	0,247**	0,08	0,002
Lenguaje receptivo	0,162*	0,06	0,011
Lenguaje expresivo	-0,111	0,07	0,129
Motricidad Fina	0,034	0,07	0,622
<b>Total (todas las áreas)</b>	<b>0,139+</b>	<b>0,07</b>	<b>0,066</b>

*Notas:* El indicador de tratamiento es igual a 1 si el niño fue asignado a tratamiento, 0 de lo contrario. El puntaje total se calculó por análisis factorial. Los controles incluyen la edad del niño, el sexo, los puntajes de LB en las 5 áreas del ASQ-3, los identificadores de los entrevistadores, la educación de la madre, una variable dicotómica igual a 1 si el hogar respondió la encuesta de LB en una lengua indígena, y el índice de riqueza. Todas las estimaciones incluyen un efecto fijo del trío de distritos según el cual se hizo la asignación aleatoria del tratamiento y corrigen los errores estándar a nivel de distrito. Coeficientes significativos al + p<0.10, \* p<0.05, \*\* p<0.01.

**Tabla 9. Efecto TOT sobre el Bayley-III (N=1.038)**

Área	Impacto	EE	P-valor
<b>Tratamiento: igual a 1 si el niño recibe al menos una visita del SAF</b>			
Desarrollo cognitivo	0,295**	0,088	0,001
Lenguaje receptivo	0,193**	0,071	0,007
Lenguaje expresivo	-0,132	0,083	0,115
Motricidad Fina	0,040	0,079	0,610
<b>Total (todas las áreas)</b>	<b>0,165+</b>	<b>0,086</b>	<b>0,054</b>
<b>Tratamiento: el número de visitas recibido por el niño</b>			
Desarrollo cognitivo	0,004**	0,001	0,001
Lenguaje receptivo	0,002**	0,001	0,006
Lenguaje expresivo	-0,002	0,001	0,114
Motricidad Fina	0,001	0,001	0,609
<b>Total (todas las áreas)</b>	<b>0,002+</b>	<b>0,001</b>	<b>0,053</b>

*Notas:* La parte superior de la tabla define el tratamiento como una variable dicotómica de valor 1 si el niño recibió alguna visita y 0 si no lo hizo. La parte inferior define el tratamiento como el número de visitas efectivamente recibido. El puntaje total se calculó por análisis factorial. Los controles incluyen la edad del niño, el sexo, los puntajes en las 5 áreas del ASQ-3, los identificadores de los entrevistadores, la educación de la madre, una variable dicotómica igual a 1 si el hogar respondió la encuesta de LB en una lengua indígena, y el índice de riqueza. Todas las estimaciones incluyen un efecto fijo del trío de distritos según el cual se hizo la asignación aleatoria del tratamiento y corrigen los errores estándar a nivel de distrito. Coeficientes significativos al + p<0.10, \* p<0.05, \*\* p<0.01.

Tabla 10. Heterogeneidad en el efecto sobre el ASQ-3 y el Bayley-III, Panel A

Variable	Muestra de la evaluación (N= 5.339)				Sub-muestra de Bayley (N=1.038)			
	Coef.	P-Valor	Prob > F	% en Control	Coef.	P-Valor	Prob > F	% en Control
<b>PANEL A:</b>				<b>Resolución de problemas</b>		<b>Desarrollo cognitivo</b>		
<36 meses (edad)	0,031	0,585	0,476	32,780	0,262*	0,014	0,858	56,560
>= 36 meses (edad)	0,079*	0,041		67,220	0,235*	0,035		43,440
Mujer	0,083+	0,055	0,470	49,090	0,345**	0,001	0,064	53,350
Hombre	0,045	0,257		50,910	0,144+	0,094		46,650
Quintil 1 de riqueza (25% más pobre)	0,136+	0,084	0,302	21,450	0,554**	0,000	0,021	21,870
Quintiles 2-5	0,046	0,189		78,550	0,187*	0,025		78,130
Primaria incompleta (o menos)	0,037	0,455	0,479	34,160	0,476**	0,001	0,034	34,110
Primaria completa (o más)	0,077*	0,037		65,840	0,149+	0,074		65,890
Desnutrido crónico	0,045	0,303	0,635	36,650	0,269**	0,006	0,791	29,740
No desnutrido crónico	0,071+	0,076		63,350	0,240**	0,008		70,260
Primogénito	0,122*	0,032	0,180	28,160	0,235+	0,066	0,904	28,920
No primogénito	0,038	0,285		71,840	0,252**	0,004		71,080
No tiene hermanos	0,076	0,239	0,780	20,060	0,192	0,191	0,650	23,020
Tiene uno o más hermanos	0,058+	0,082		79,940	0,265**	0,002		76,980
Madre/responsable 10% superior indice de depresión (FA)	0,078	0,410	0,870	10,200	0,227	0,300	0,927	9,500
Madre/responsable 90% inferior indice de depresión (FA)	0,062+	0,055		89,800	0,249**	0,003		90,500
Madre/responsable 10% superior indice de violencia (FA)	0,143	0,116	0,334	10,340	0,291	0,301	0,853	8,490
Madre/responsable 90% inferior indice de violencia (FA)	0,055+	0,090		89,660	0,240**	0,002		91,510
Nino recibió lactancia exclusiva por 6 meses o más	0,059+	0,089	0,793	75,520	0,253**	0,003	0,794	73,670
Nino recibió lactancia exclusiva por un maximo de 6 meses	0,078	0,239		24,480	0,218+	0,077		26,330

Tabla 10. Heterogeneidad en el efecto sobre el ASQ-3 y el Bayley-III, Panel B

Variable	Muestra de la evaluación (N= 5.339)				Sub-muestra de Bayley (N=1.038)			
	Coef.	P-Valor	Prob > F	% en Control	Coef.	P-Valor	Prob > F	% en Control
<b>PANEL B:</b>				<b>Comunicación</b>				<b>Lenguaje receptivo</b>
<36 meses (edad)	0,028	0,614	0,291	32,780	0,245**	0,004	0,103	56,560
>= 36 meses (edad)	0,099*	-0,013		67,220	0,058	0,494		43,440
Mujer	0,098*	0,017	0,478	49,090	0,188*	0,020	0,630	53,350
Hombre	0,060	0,160		50,910	0,134	0,138		46,650
Quintil 1 de riqueza (25% más pobre)	0,044	0,542	0,588	21,450	0,435**	0,008	0,062	21,870
Quintiles 2-5	0,087*	0,015		78,550	0,102	0,140		78,130
Primaria incompleta (o menos)	0,075	0,155	0,917	34,160	0,340**	0,005	0,043	34,110
Primaria completa (o más)	0,081*	0,028		65,840	0,082	0,224		65,890
Desnutrido crónico	0,103*	0,033	0,529	36,650	0,234*	0,028	0,451	29,740
No desnutrido crónico	0,064	0,125		63,350	0,139+	0,067		70,260
Primogénito	0,105+	0,063	0,551	28,160	0,232+	0,052	0,462	28,920
No primogénito	0,068+	0,065		71,840	0,132+	0,064		71,080
No tiene hermanos	0,063	0,347	0,775	20,060	0,207+	0,065	0,686	23,020
Tiene uno o más hermanos	0,083*	0,018		79,940	0,148+	0,067		76,980
Madre/responsable 10% superior indice de depresión (FA)	-0,014	0,876	0,247	10,200	0,237	0,208	0,652	9,500
Madre/responsable 90% inferior indice de depresión (FA)	0,089**	0,007		89,800	0,153*	0,017		90,500
Madre/responsable 10% superior indice de violencia (FA)	0,223*	0,012	0,078	10,340	0,224	0,486	0,820	8,490
Madre/responsable 90% inferior indice de violencia (FA)	0,063+	0,058		89,660	0,151*	0,016		91,510
Nino recibió lactancia exclusiva por 6 meses o más	0,067+	0,059	0,479	75,520	0,221**	0,002	0,078	73,670
Nino recibió lactancia exclusiva por un maximo de 6 meses	0,114+	0,060		24,480	-0,007	0,952		26,330

Notas: Cada línea representa un modelo en el cual se incluye una interacción entre el indicador de tratamiento y cada una de las variables presentadas en la Tabla. “Prob > F” muestra el p-valor de la prueba de la igualdad de las dos interacciones. Los puntajes “FA” fueron calculados por análisis factorial. Los controles incluyen la edad del niño, el sexo, los puntajes de LB en las 5 áreas del ASQ-3, los identificadores de los entrevistadores, la educación de la madre, una variable dicotómica igual a 1 si el hogar respondió la encuesta de LB en una lengua indígena, y el índice de riqueza. Todas las estimaciones incluyen un efecto fijo del trío de distritos según el cual se hizo la asignación aleatoria del tratamiento y corrigen los errores estándar a nivel de distrito. Coeficientes significativos al + p<0.10, \* p<0.05, \*\* p<0.01.

**Tabla 11. Efecto ITT en otros indicadores de resultados, muestra de la evaluación (N=5.339)**

	N	Impacto	EE	P-valor	Media (control)
<b><i>Family Care Indicators (FCI)</i></b>					
Actividades de juego (FA, puntaje Z)	5.339	0,201**	0,035	0,000	0,000
Juguetes caseros	5.268	0,082**	0,016	0,000	0,127
Juguetes comprados	5.268	0,017+	0,010	0,095	0,905
Juguetes con objetos del hogar	5.268	-0,012	0,030	0,676	0,562
Libros para adulto	5.298	-0,038*	0,019	0,048	0,557
Periódicos	5.298	-0,032*	0,013	0,015	0,188
Materiales de juego (FA, puntaje Z)	5.268	0,049	0,040	0,226	0,000
<b><i>Disciplina y formas de felicitar al niño (semana pasada)</i></b>					
Gritos e insultos	5.339	-0,038**	0,014	0,010	0,309
Golpes con correa (o algún otro objeto)	5.339	-0,024*	0,011	0,023	0,132
Felicitar verbalmente o aplaudir	5.339	0,084**	0,015	0,000	0,415
Dar una recompensa	5.339	-0,027**	0,009	0,003	0,116
Abrazar, consentir, ser cariñoso	5.339	-0,018	0,015	0,239	0,458
Reirse juntos	5.339	0,001	0,014	0,971	0,208
<b><i>Calidad del hogar</i></b>					
Hogar duro (FA, puntaje Z)	5.339	0,037	0,035	0,283	0,000
Hogar Frio (FA, puntaje Z)	5.339	-0,016	0,046	0,724	0,000
<b><i>Conocimiento sobre desarrollo infantil</i></b>					
Puntaje KIDI (FA, puntaje Z)	5.211	0,001	0,044	0,988	0,000
Puntaje Cuna Más (FA, puntaje Z)	5.211	-0,053	0,046	0,246	0,000
<b><i>Educación Inicial</i></b>					
Participación en educación inicial	3.641	0,085**	0,023	0,000	0,507
<b><i>Nutrición</i></b>					
Peso por edad (puntaje Z)	5.328	0,032	0,020	0,120	-0,813
Peso por talla (puntaje Z)	5.323	0,016	0,025	0,531	0,275
IMC por edad (puntaje Z)	5.321	0,005	0,025	0,836	0,456
Talla por edad (puntaje Z)	5.329	0,036+	0,020	0,065	-1,725
Desnutrición global (peso por edad)	5.328	-0,012*	0,005	0,022	0,079
Desnutrición crónica (talla por edad)	5.329	-0,015	0,011	0,181	0,385
Desnutrición aguda (peso por talla)	5.323	-0,001	0,002	0,681	0,005

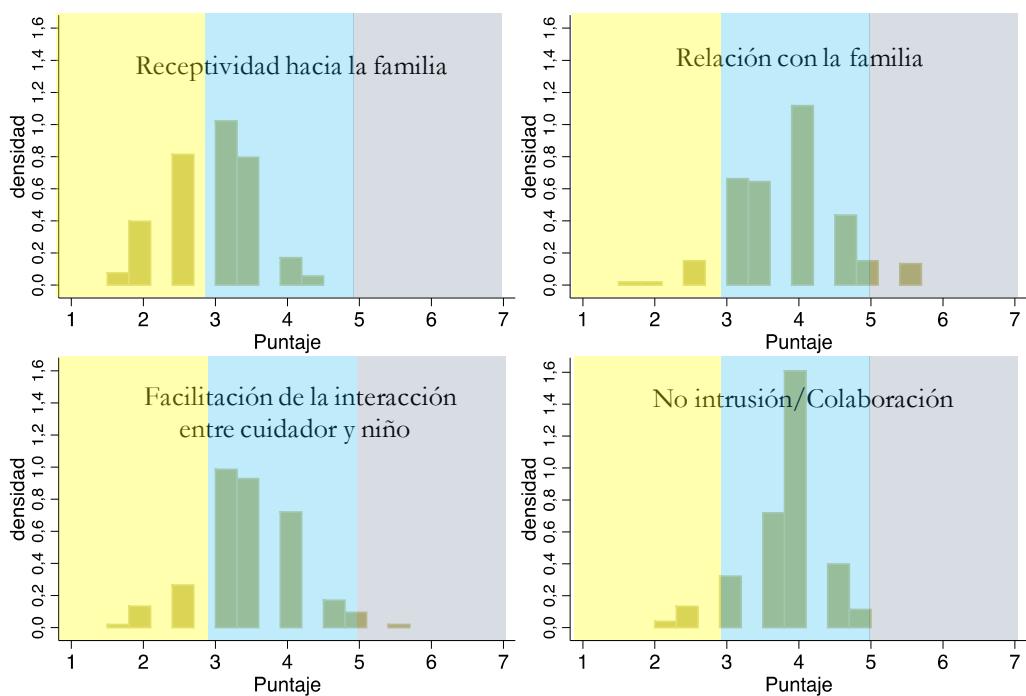
Notas: El indicador de tratamiento es igual a 1 si el niño fue asignado al grupo de tratamiento, cero de lo contrario. Los puntajes "FA" fueron calculados por análisis factorial. El indicador de las actividades de juego se calcula con base a la frecuencia con la que el niño realizó alguna de 7 actividades con una persona mayor de 15 años en la semana previa a la entrevista. Los indicadores sobre el origen de los juguetes (caseros, comprados, objetos del hogar), los libros para adultos y los periódicos son variables dicotómicas que toman el valor 1 cuando el hogar tiene uno o más de estos. El indicador para los materiales de juegos se calcula con base a la disponibilidad de 8 variedades de juguetes en el hogar. Los indicadores de disciplina y formas de felicitar al niño son variables dicotómicas que toman el valor 1 cuando la madre realizó la forma de disciplinar o felicitar al niño la semana previa a la entrevista. La calidad del hogar se mide a través de un subconjunto de los ítems de las escalas de receptividad y aceptación de la herramienta *Home Observation for Measurement of the Environment* (HOME; Caldwell y Bradley, 2003), y el conocimiento sobre el desarrollo infantil a través del *Knowledge of Infant Development Inventory* (KIDI; MacPhee, 2002), o de preguntas definidas por el mismo Programa. Educación inicial es una variable dicotómica que tiene un valor de 1 si el niño asiste a un centro o jardín del Ministerio de Educación en la LS, e incluye solo a los niños que tienen más de 36 meses en la LS. Los indicadores nutricionales se construyeron usando los lineamientos de la Organización Mundial de la Salud. Los controles incluyen la edad del niño, el sexo, el puntaje del indicador de resultado en la LB (cuando sea aplicable), la educación de la madre, una variable dicotómica igual a 1 si el hogar respondió la encuesta de LB en una lengua indígena, y el índice de riqueza. Todas las estimaciones incluyen un efecto fijo del trío de distritos según el cual se hizo la asignación aleatoria del tratamiento y corrigen los errores estándar a nivel de distrito. Coeficientes significativos al + p<0.10, \* p<0.05, \*\* p<0.01.

**Tabla 12. Efecto ITT sobre el ASQ-3 y el Bayley-III, según el tipo de tratamiento**

Área	T1 (visitas)			T2 (visitas y sesiones grupales)			Prob > F
	Impacto	EE	P-valor	Impacto	EE	P-valor	
<b>Muestra de la evaluación (N=5.339)</b>							
Resolución de problemas	0,056	0,039	0,147	0,071*	0,035	0,042	0,695
Comunicación	0,106**	0,039	0,007	0,053	0,035	0,136	0,147
Motricidad fina	0,051	0,042	0,223	0,070+	0,040	0,084	0,652
Personal-social	0,053	0,045	0,234	0,078+	0,042	0,062	0,576
Motricidad gruesa	0,004	0,038	0,923	-0,019	0,033	0,569	0,530
Total (todas las áreas )	0,066+	0,035	0,057	0,061*	0,030	0,046	0,870
Total (RP+Com+MF)	0,066*	0,030	0,026	0,060*	0,026	0,025	0,814
<b>Sub-muestra de Bayley (N=1.038)</b>							
Desarrollo cognitivo	0,239**	0,087	0,008	0,255**	0,082	0,003	0,847
Lenguaje receptivo	0,172*	0,074	0,023	0,151*	0,073	0,042	0,789
Lenguaje expresivo	-0,101	0,088	0,254	-0,120	0,085	0,161	0,841
Motricidad Fina	0,020	0,085	0,809	0,047	0,075	0,527	0,753
Total (todas las áreas)	0,136	0,087	0,124	0,142+	0,085	0,099	0,952

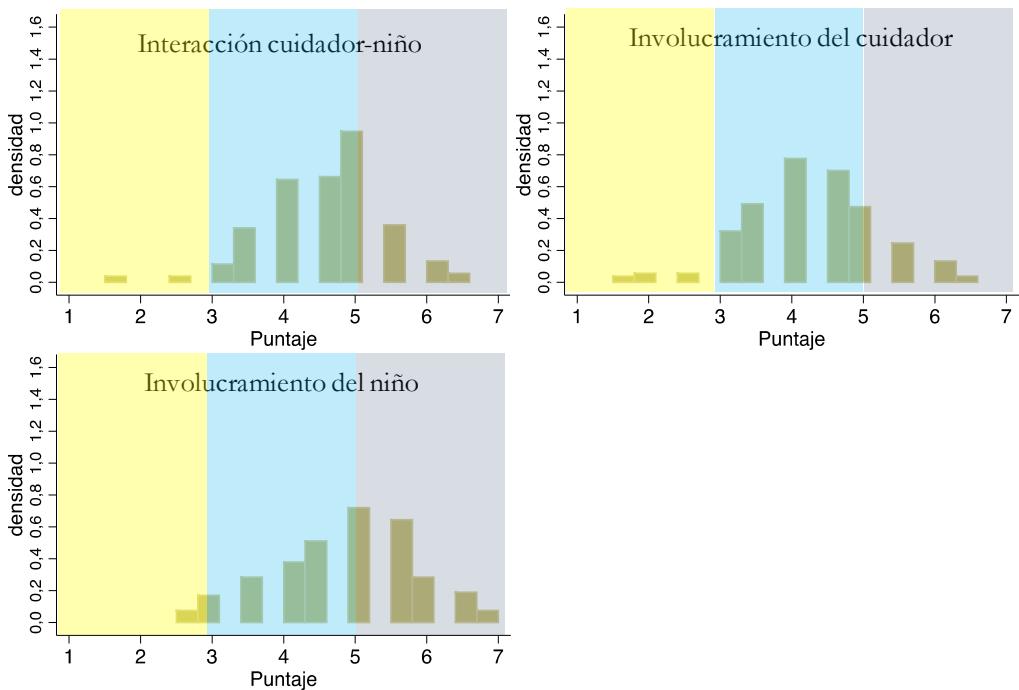
*Notas:* En el panel de la izquierda, el indicador de tratamiento es igual a 1 si el niño fue asignado al grupo de tratamiento con solo visitas (T1), 0 de lo contrario. En el panel de la derecha, el indicador de tratamiento es igual a 1 si el niño fue asignado al grupo de tratamiento con visitas y sesiones grupales (T2), 0 de lo contrario. Los puntajes totales se calcularon por análisis factorial. Los controles incluyen la edad del niño, el sexo, los puntajes de LB en las 5 áreas del ASQ-3, los identificadores de los entrevistadores, la educación de la madre, una variable dicotómica igual a 1 si el hogar respondió la encuesta de LB en una lengua indígena, y el índice de riqueza. Todas las estimaciones incluyen un efecto fijo del trío de distritos según el cual se hizo la asignación aleatoria del tratamiento y corrigen los errores estándar a nivel de distrito. Coeficientes significativos al + p<0.10, \* p<0.05, \*\* p<0.01.

**Figura 5. Puntajes del índice del HOVRS “Prácticas de la facilitadora” (N= 176)**



Notas: puntajes promedios de las dimensiones del dominio “Prácticas de la facilitadora” del HOVRS para 176 facilitadoras de la muestra de medición de calidad de las visitas.

**Figura 6. Puntajes del índice del HOVRS “Involucramiento de la familia” (N= 176)**



Notas: puntajes promedios de las dimensiones del dominio “Involucramiento de la familia” del HOVRS para 176 facilitadoras de la muestra de medición de calidad de las visitas.

**Tabla 13. Asociación entre la calidad de la visita y el desarrollo infantil**

	Coef.	EE	P-valor
<b>Asociación con el ASQ-3 (N= 460)</b>			
<b>Resolución de problemas</b>			
<b>Dominio 1. Prácticas de la facilitadora</b>	<b>0,103*</b>	<b>0,048</b>	<b>0,034</b>
1. Receptividad hacia la familia	0,090	0,057	0,117
2. Relación con la familia	0,088+	0,052	0,094
3. Facilitación de la interacción entre cuidador y niño	0,096*	0,046	0,041
4. No intrusión/Colaboración	0,066	0,043	0,135
<b>Dominio 2. Involucramiento de la familia</b>	<b>0,169**</b>	<b>0,046</b>	<b>0,001</b>
5. Interacción cuidador-niño	0,155**	0,044	0,001
6. Involucramiento del cuidador	0,143**	0,042	0,001
7. Involucramiento del niño	0,134*	0,053	0,013
<b>Total</b>	<b>0,146**</b>	<b>0,048</b>	<b>0,003</b>
<b>Comunicación</b>			
<b>Dominio 1. Prácticas de la facilitadora</b>	<b>0,053</b>	<b>0,053</b>	<b>0,323</b>
1. Receptividad hacia la familia	0,030	0,055	0,591
2. Relación con la familia	0,048	0,050	0,336
3. Facilitación de la interacción entre cuidador y niño	0,071	0,045	0,120
<b>4. No intrusión/Colaboración</b>	<b>0,015</b>	<b>0,058</b>	<b>0,802</b>
<b>Dominio 2. Involucramiento de la familia</b>	<b>0,103*</b>	<b>0,049</b>	<b>0,041</b>
5. Interacción cuidador-niño	0,105+	0,053	0,051
6. Involucramiento del cuidador	0,104+	0,054	0,059
7. Involucramiento del niño	0,058	0,043	0,179
<b>Total</b>	<b>0,083</b>	<b>0,052</b>	<b>0,113</b>
<b>Asociación con el Bayley-III (N= 367)</b>			
<b>Desarrollo Cognitivo</b>			
<b>Dominio 1. Prácticas de la facilitadora</b>	<b>0,011</b>	<b>0,046</b>	<b>0,816</b>
1. Receptividad hacia la familia	0,000	0,059	0,998
2. Relación con la familia	-0,001	0,044	0,988
3. Facilitación de la interacción entre cuidador y niño	0,056	0,044	0,212
4. No intrusión/Colaboración	-0,029	0,041	0,486
<b>Dominio 2. Involucramiento de la familia</b>	<b>0,037</b>	<b>0,049</b>	<b>0,457</b>
5. Interacción cuidador-niño	0,001	0,045	0,975
6. Involucramiento del cuidador	-0,007	0,048	0,886
7. Involucramiento del niño	0,092*	0,045	0,046
<b>Total</b>	<b>0,026</b>	<b>0,046</b>	<b>0,579</b>
<b>Lenguaje Receptivo</b>			
<b>Dominio 1. Prácticas de la facilitadora</b>	<b>0,039</b>	<b>0,054</b>	<b>0,477</b>
1. Receptividad hacia la familia	0,055	0,057	0,331
2. Relación con la familia	0,036	0,051	0,481
3. Facilitación de la interacción entre cuidador y niño	0,095+	0,053	0,078
4. No intrusión/Colaboración	-0,075	0,049	0,130
<b>Dominio 2. Involucramiento de la familia</b>	<b>0,046</b>	<b>0,048</b>	<b>0,342</b>
5. Interacción cuidador-niño	-0,005	0,053	0,926
6. Involucramiento del cuidador	0,065	0,054	0,227
7. Involucramiento del niño	0,054	0,050	0,283
<b>Total</b>	<b>0,046</b>	<b>0,051</b>	<b>0,369</b>

*Notas:* El puntaje del HOVRS corresponde al promedio (expresado en puntaje Z) para la facilitadora del niño cuyo desarrollo se mide como variable dependiente, pero excluyendo del promedio la medida de calidad para la vista de *ese* niño en particular. Los controles incluyen la edad del niño, el sexo, los puntajes de LB en las 5 áreas del ASQ-3, los identificadores de los entrevistados, la educación de la madre, una variable dicotómica igual a 1 si el hogar respondió la encuesta de LB en una lengua indígena, el índice de riqueza, y la edad y educación de la facilitadora. Todas las estimaciones corrigen los errores estándar a nivel de distrito. Coeficientes significativos al + p<0.10, \* p<0.05, \*\* p<0.01.

## Anexo 1. Roles y funciones del equipo técnico de la evaluación

INSTITUCIÓN	ORGANO DE LÍNEA	FUNCIONES
<b>Banco Interamericano de Desarrollo (BID)</b>		<ul style="list-style-type: none"> <li>Desarrollo de la propuesta inicial del diseño metodológico de la evaluación de impacto en coordinación con el Equipo Técnico.</li> <li>Brindar aportes técnicos durante el desarrollo de la evaluación (diseño de los instrumentos, capacitación, levantamiento de información, entre otros).</li> <li>Financiar y gestionar recursos para estudios y actividades complementarias para el desarrollo de la evaluación (desarrollo de instrumentos, digitación y análisis de la información, recolección de datos sobre calidad del servicio, operativo Bayley, entre otros).</li> <li>Elaborar el informe de la línea de base conjuntamente con el Equipo Técnico.</li> <li>Elaborar el informe final de la evaluación de impacto recogiendo los comentarios del Equipo Técnico.</li> </ul>
<b>Ministerio de Economía y Finanzas</b>	Dirección de Calidad del Gasto Público	<ul style="list-style-type: none"> <li>Conducir el recojo de la línea de base y de la línea de seguimiento en coordinación con el Equipo Técnico.</li> <li>Brindar aportes técnicos para el diseño metodológico de la evaluación de impacto y para el análisis de los resultados de la misma.</li> <li>Brindar aportes técnicos en la elaboración de informes y documentación a lo largo de la evaluación de impacto.</li> <li>Elaborar el informe de la línea de base conjuntamente con el Equipo Técnico.</li> <li>Coordinar, financiar y supervisar actividades complementarias a la evaluación de impacto.</li> </ul>
<b>Ministerio de Desarrollo e Inclusión Social</b>	Dirección de Seguimiento y Evaluación	<ul style="list-style-type: none"> <li>Convocar al Equipo Técnico para las reuniones y coordinaciones que sean necesarias.</li> <li>Brindar aportes técnicos para el diseño metodológico de la evaluación de impacto y para el análisis de los resultados de la misma.</li> <li>Elaborar el informe de la línea de base conjuntamente con el Equipo Técnico.</li> <li>Coordinar con el Programa Nacional Cuna Más y con otras direcciones del MIDIS la provisión de toda la información necesaria para la evaluación de impacto.</li> <li>Velar por el cumplimiento de los lineamientos de evaluación de la directiva 007-2012-MIDIS.</li> <li>Financiar y supervisar actividades complementarias a la evaluación de impacto.</li> </ul>

## Anexo 2. Indicadores de resultado de la evaluación

Tipos de variables	Área / Dimensión	Indicador
Resultado final	Desarrollo del niño	Motora gruesa
		Motora fina
		Comunicación
		Cognitiva
		Socioemocional
Resultados intermedios	Calidad del ambiente en el hogar	Motora gruesa
		Motora fina
		Comunicación
		Cognitiva
		Socioemocional
	Interacciones	Puntaje en la escala de calidad del ambiente del hogar (HOME)
		Adultos que leen libros, cuentan cuentos o les cantan a los niños (número de días)
		Adultos que utilizan juguetes para jugar con los niños (número de días)
		Adultos que salen a pasear con los niños (número de días)
		Adultos que hacen dibujar o pintar a los niños (número de días)
	Receptividad de la madre	Adultos que juegan a nombrar/contar objetos o colores con los niños (número de días)
		Madres que dan de lactar a sus niños en un clima de interacción afectiva
		Madres que alimentan a sus hijos de manera responsiva
		Madres que responden verbalmente ante sonidos o palabras de afecto de sus hijos
	Cuidado	Madres que responden con afecto a las señales de llanto del niño
		Número de días que el niño es dejado a cargo de otro niño durante la semana previa a la encuesta
	Alimentación e Higiene	Número de días que el niño es dejado solo durante la semana previa a la encuesta
		Niños mayores de 6 meses que reciben alimentos al menos 5 veces al día
		Niños mayores de 6 meses que reciben almuerzos con consistencia adecuada
		Niños que reciben los alimentos en su propio plato
		Niños mayores de 6 meses que consumen frutas y verduras
		Niños mayores de 6 meses que consumen alimentos de origen animal
		Madres que realizan prácticas adecuadas de lavado de manos
		Madres que utilizan agua segura para consumo humano
		Disponibilidad de espacios físicos y materiales en el hogar que favorecen el desarrollo infantil
		Madres que habilitan un espacio de juego adecuado para sus niños
		Niños que disponen de un espacio para guardar sus juguetes
		Número de libros, revistas y/o periódicos en el hogar
		Número de juguetes hechos en casa
		Número de juguetes comprados/fabricados
		Número de juguetes para armar o construir
		Número de juguetes para pintar o escribir
		Número de juguetes que requieren movimiento físico
		Número de juguetes para el juego de roles
		Número de libros infantiles de cuentos y/o con imágenes para colorear
		Número de juguetes para aprender formas y colores
		Número de juguetes que producen música o imágenes audiovisuales