Evaluación de Impacto del Programa 1000 Días San Miguel sobre las medidas antropométricas de los niños

María Sol González[1]
María Emma Santos[2]

Resumen

El trabajo evalúa el impacto del Programa de Acompañamiento Familiar Mil Días, una iniciativa pionera de desarrollo infantil temprano en Argentina, implementado en el municipio de San Miguel, Provincia de Buenos Aires, desde agosto de 2015, sobre los indicadores antropométricos de niños en situación de vulnerabilidad. El programa, de enfoque multidimensional, busca acompañar a los niños y sus familias durante los primeros "1000 días" de vida, período que abarca desde la gestación (270 días) hasta los dos años de edad (730 días). Si bien este es el foco de la intervención, las madres y los niños pueden incorporarse en cualquier momento dentro de dicho lapso. La duración prevista de la intervención es de aproximadamente un año, durante el cual se procura revertir la vulnerabilidad que motivó el ingreso. La principal herramienta de intervención es el acompañamiento familiar -primero en los hogares y luego en los Centros de Atención Primaria de Salud (CAPS)-, junto con la estimulación infantil en los Centros de Desarrollo Infanto-Familiar (CDIF), además de otros beneficios definidos por los criterios de admisión.

La evaluación estima los efectos a partir de dos estrategias de Diferencias en Diferencias. La primera adapta los datos a la estructura canónica tradicional 2×2. La segunda se apoya en un modelo de Diferencias en Diferencias con múltiples períodos y adopción heterogénea del tratamiento, siguiendo la propuesta de Callaway y Sant'Anna (2021). Los resultados muestran un impacto significativo y elevado en el z-score de peso-para-la-edad, del orden de una desviación estándar, robusto a las distintas metodologías de estimación. También se encuentra un impacto en el z-score de peso-para-la-talla, aunque menos robusto, mientras que no se observa impacto en talla-para-la-edad.

Palabras clave: primera infancia, mil días, vulnerabilidad, acompañamiento familiar, evaluación de impacto, enfoque multidimensional.

[1] Becaria Doctoral del Instituo de Ciencias para la Familia, Universidad Austral- CONICET.
[2] Investigadora adjunta del Instituto de Investigaciones Económicas y Sociales del Sur, Universidad Nacional del Sur, CONICETy OPHI

1. Introducción

Los primeros años de vida constituyen una etapa fundamental para el desarrollo integral, en la que se establecen las bases del crecimiento físico, cognitivo, emocional y social. Durante la gestación y los primeros años postnatales, el cerebro humano experimenta una transformación notable, alcanzando aproximadamente el 80% de su tamaño adulto a los 3 años y hasta el 90% a los 5 años (Araujo & López Boo, 2010; Heckman, 2006). Este proceso, dinámico, integral y acumulativo, está influenciado por las experiencias ambientales (Berlinski & Schady, 2015; Papalia & Martorell, 2017).

La evidencia empírica indica que tanto las habilidades cognitivas como las no cognitivas (como la perseverancia, el autocontrol y la autoestima) se forjan en estos primeros años (Cunha & Heckman, 2007). La existencia de períodos críticos y sensibles implica que las inversiones tempranas generan efectos acumulativos que potencian competencias futuras, mientras que su ausencia o retraso limita el desarrollo pleno.

Desde un punto de vista económico, la inversión en la primera infancia presenta retornos sociales y económicos más elevados que cualquier otro momento del ciclo vital. García, Heckman, Duncan y Prados (2016), han demostrado que los programas dirigidos a niños desfavorecidos desde el nacimiento hasta los 5 años pueden generar retornos anuales de hasta un 13%, superando ampliamente los obtenidos cuando la intervención se retrasa. Asimismo, estudios basados en los programas Perry Preschool y Abecedarian (Conti & Heckman, 2012) han mostrado que las intervenciones tempranas no solo mejoran el rendimiento académico y las capacidades cognitivas, sino que también inciden positivamente en indicadores socioeconómicos como el empleo y la reducción de la delincuencia.

En este marco, el trabajo presenta la primera evaluación de impacto cuantitativa sobre indicadores antropométricos de un programa dirigido a la primera infancia en situación de alta vulnerabilidad: el "Programa de Acompañamiento Familiar Mil Días" (en adelante, Mil Días), implementado desde agosto de 2015 en el municipio de San Miguel, provincia de Buenos Aires, Argentina. Este programa acompaña a embarazadas, madres con niños menores de 2 años y niños de hasta esa edad en situación de extrema vulnerabilidad, durante sus primeros "1000 días" de vida -desde la gestación (270 días) hasta los dos primeros años (730 días). El objetivo último del programa es la reducción de la morbi-mortalidad materno-infantil. La principal herramienta de intervención es el acompañamiento familiar, primero bajo la forma de visitas domiciliarias y luego en los Centros de Atención primaria de Salud (CAPS), así como también la estimulación de los niños en los Centros de Desarrollo Infantil y Familiar (CDIF), pero el programa ofrece otros beneficios según los criterios de vulnerabilidad por los que se admiten los participantes. Aunque su foco es el período de los primeros 1000 días, la incorporación al programa puede producirse en cualquier momento dentro de ese lapso, y la participación efectiva suele extenderse aproximadamente un año.

Al momento de su puesta en marcha, el programa no contempló la realización de una evaluación de impacto, lo que planteó importantes desafíos metodológicos:

i. los únicos indicadores disponibles antes, durante y después de la participación fueron los antropométricos;
ii. la implementación no fue aleatorizada, lo que exigió el uso de metodologías cuasiexperimentales; y
iii. la información de beneficiarios no se encontraba sistematizada, requiriendo la construcción y depuración exhaustiva de la base de datos.

Para superar estas limitaciones, se diseñó ex post un grupo de comparación y se aplicaron dos estrategias metodológicas complementarias: (i) la adaptación de los datos a un modelo canónico tradicional de Diferencias en Diferencias 2×2 y (ii) la metodología no paramétrica propuesta por Callaway y Sant'Anna (2021), especialmente desarrollada para programas con tratamiento temporal heterogéneo

Los resultados presentados en este trabajo se obtuvieron de una base de datos construida desde cero, con información antropométrica de talla y peso, desde el nacimiento, con mediciones tomadas al menos semestralmente, de niños que participaron en Mil Días y de niños que no lo hicieron. En todos los casos, estas mediciones fueron realizadas por profesionales de la salud en los CAPS del municipio. Las variables de interés para la evaluación de impacto son los puntajes z de peso-para-la-edad, peso-para-la-talla y talla-para-la-edad. Estos indicadores fueron calculados a partir de los datos brutos mediante el algoritmo de libre acceso desarrollado por la Organización Mundial de la Salud (OMS).

Las medidas antropométricas constiuyen un indicador objetivo de gran relevancia. No solo proporcionan información sobre el bienestar actual del niño, sino que también sirven como buenos predictores del crecimiento físico y el desarrollo cognitivo a largo plazo. También permiten la detección temprana de problemas de desarrollo (UNICEF, 2013; 2020; OMS, 2020; Ramos-Padilla, Carpio-Arias, Delgado-López, 2020; Victora, Adair, Fall, Hallal, Martorell, Richter et al., 2008; Dewey y Begum, 2011; Flores-Huerta, 2006; Hoddinott et al., 2013; Berlinski y Schady, 2015).

Finalmente, los objetivos y hallazgos de la investigaicon se alinean con los Objetivos de Desarrollo Sostenible (ODS) establecidos por las Naciones Unidas: poner fin a la pobreza (ODS 1), lograr hambre cero (ODS 2) y garantizar una vida sana y promover el bienestar para todos (ODS 3). Particularmente, la investigación se vincula directamente con las metas 3.1 y 3.2, orientadas a reducir la mortalidad materna y eliminar las muertes evitables de recién nacidos y niños menores de 5 años. Esta evaluación aporta evidencia concreta y rigurosa para fortalecer políticas públicas que garanticen los derechos fundamentales de la infancia y reduzcan las brechas de desigualdad desde la gestación hasta los dos primeros años de vida.

El trabajo se estructura de la siguiente manera. En la Sección 2 se presenta una breve descripción del programa Mil Días. En la Sección 3, se detallan los desafíos metodológicos que se debieron abordar para realizar la evaluación de impacto. En la Sección 4 se describe la base de datos construida para el Grupo Tratamiento y el Grupo Control. En la Sección 5 se presenta el modelo canónico de Diferencias-en-Diferencias. En la Sección 6 se describe cómo se adecuaron los datos de 1000 Días para implementar este modelo. En la Sección 7 se presenta la metodología de Callaway y Sant'Anna y en la Sección 8 se describe cómo se adecuaron los datos para implementar dicha metodología. En la Sección 9 se presentan y discuten los resultados de ambas estrategias metodológicas en perspectiva comparada. Finalmente, en la Sección 10 se sintetizan las principales conclusiones.

2. Descripción del Programa "Mil Días"

El Programa de Acompañamiento Familiar Mil Días constituye una iniciativa pionera en su tipo en Argentina, implementada en agosto de 2015 en el municipio de San Miguel, provincia de Buenos Aires. Fue diseñado por la Coordinación de Primera Infancia, Niñez y Familia, creada un año antes bajo la órbita de la Secretaría de Salud y Bienestar Familiar. Tal como su nombre lo indica, el programa busca acompañar a niños y familias en

situación de extrema vulnerabilidad durante los primeros “1000 días” de vida, que por diversas razones tienen dificultades para acceder al sistema de salud y alcanzar un nivel mínimo de bienestar familiar. Este período comprende desde la gestación (270 días) hasta los dos primeros años de vida (730 días). Si bien este es el período objetivo de intervención, las madres y los niños pueden incorporarse en cualquier momento dentro de ese tiempo. La duración prevista de la intervención es de aproximadamente un año, durante el cual se procura revertir la vulnerabilidad que motivó el ingreso. No obstante, este plazo puede variar según la gravedad de cada caso.

El diseño del programa se inspiró en experiencias de otros países de América Latina, tales como *Chile Crece Contigo* (Chile, desde 2006), *Uruguay Crece Contigo* (Uruguay, desde 2012), *Programa Unidos* (Colombia, desde 2011) y *Familia Paranaense* (Brasil, desde 2012). A través de diversos instrumentos de intervención, *Mil Días* busca garantizar una crianza segura y saludable, promoviendo habilidades y capacidades en los cuidadores (madre o tutor), fortaleciendo los recursos familiares y consolidando vínculos y redes de apoyo. El objetivo ultimo del programa es la reducción de la morbi-mortalidad materno-infantil.

El programa se implementa en los ocho barrios más pobres del municipio, seleccionados a partir de los datos del Censo 2010 y relevamientos propios realizados por la Coordinación de Primera Infancia, Niñez y Familia en 2014-2015. La principal vía de derivación es la Unidad Materno Infantil (UMI) del Hospital Municipal Raúl F. Larcade y los Centros de Atención Primaria de la Salud (CAPS). Otra vía de detección son los “Puntos de Infancia”, que son oficinas territoriales móviles que se instalan en los barrios vulnerables para acercar información sobre programas y servicios municipales, realizar campañas de sensibilización y facilitar trámites sanitarios y sociales. Estos puntos surgen para llegar a familias que, por su situación, no concurren habitualmente a los CAPS, UMI u oficinas municipales.

Cuando una embarazada o madre de un niño menor de 2 años es atendida en estos espacios, se le realiza una encuesta para determinar si es candidata al programa. En caso afirmativo, se la postula a través del Sistema de Acompañamiento Familiar (SAF), un registro unificado de familias vulnerables que permite emitir alertas tempranas ante situaciones de riesgo.

Una vez postulado el caso, la Coordinación aplica una segunda encuesta para verificar la información y evaluar 51 criterios de vulnerabilidad de ingreso, agrupados en dimensiones de salud, vivienda, situación socioeconómica y violencia familiar. La admisión depende de la disponibilidad de vacantes y de la aceptación por parte de la familia; de no haber cupo, el hogar ingresa a una lista de espera.

Una vez que la familia ingresa al programa reciben un paquete de intervención cuyo eje principal es el acompañamiento familiar, que comienza con visitas domiciliarias y, a partir del sexto mes, se complementa con encuentros individuales en los CAPS. Durante el período de intervención, se adopta un enfoque multidimensional e intersectorial, coordinando acciones con prestadores de salud y distintas áreas de gobierno para garantizar un abordaje integral de cada caso.

Además de las visitas domiciliarias, el programa promueve la asistencia de los niños a los CDIF, con el fin de asegurar que reciban estimulación acorde a su edad. Las familias que enfrentan casos de desnutrición severa o en las que ninguno de sus miembros tiene empleo reciben cajas de alimentos especialmente diseñadas por nutricionistas. El programa también garantiza la provisión de leche, vacunas y medicamentos esenciales cuando es necesario. Asimismo, organiza talleres mensuales -denominados internamente *mateadas*-

en los que se realizan actividades de formación para las familias y se fortalecen los lazos comunitarios. Finalmente, el programa lleva adelante operativos móviles de salud, que incluyen control de embarazadas, madres y niños, vacunación, atención de enfermería, oftalmología, ecografías, odontología y servicios de la seguridad social (ANSES).

El egreso se produce cuando se han revertido las condiciones que motivaron el ingreso y se ha reducido el riesgo biopsicosocial, según la evaluación conjunta del equipo técnico, la coordinación y los acompañantes familiares. En caso de ser necesario, la permanencia puede extenderse más allá de los dos años de edad; entre 2015 y 2019, esto ocurrió en el 25% de los beneficiarios.

En trabajos previos (González y Santos, 2023a, 2023b) se ha demostrado que el programa efectivamente alcanza a una población altamente vulnerable, es decir, está bien focalizado. También se encontró que los participantes en situaciones de vulnerabilidad extrema (violencia doméstica, desnutrición, niños indocumentados, madres con trastornos psiquiátricos) tienen mayor probabilidad de abandonar el programa, y que los casos más complejos tienden a graduarse más lentamente. De hecho, los datos muestran que el 50% de los casos que egresaron exitosamente permanecieron en el programa más de un año. Asimismo, se obsreva que, cuando las madres adolescentes logran permanecer en el programa, se gradúan más rápido que las madres no adolescentes. Por último, un dato relevante es que la percepción de transferencias monetarias complementarias se asocia a una menor probabilidad de abandono.

3. Desafíos metodológicos para realizar la evaluación de impacto de 1000 Días-San Miguel

La evaluación del impacto del Programa 1000 Días presentó desafíos metodológicos significativos debido a la naturaleza heterogénea de la intervención. Específicamente, no existe un período común de tratamiento entre los participantes: cada niño ingresó (y egresó) al (del) programa en diferentes momentos del tiempo entre agosto de 2015 y mayo de 2019, y a distintas edades. Esta característica fundamental se puede desglosar en los siguientes puntos:

1. **Variabilidad máxima en los períodos temporales de tratamiento, tanto a nivel calendario como cronológico:** A nivel de tiempo calendario, el programa se implementó en forma escalonada en diferentes barrios entre 2015 y 2017 (cuatro en 2015, tres más en 2016 y el último en 2017). Existen, por ejemplo, evaluaciones de impacto de programas que se fueron implementando de manera escalonada en el tiempo en diferentes áreas geográficas (Angrist y Pischke, 2015). Sin embargo, el tiempo calendario no es la única fuente de heterogeneidad: cada niño ingresa al programa en una edad cronológica diferente, en cualquier momento entre el nacimiento y los 2 años de edad.[3]
2. **Focalización:** 1000 Días no se trata de un programa que brinde tratamiento masivo a todo un grupo poblacional que cumpla con ciertos requisitos. Por lo tanto, no es posible asumir que todos los niños con una característica "x" en el barrio "z" recibieron el tratamiento a partir de un momento “t” en el que se puso en marcha el programa en ese barrio, y todos los que no tenían esa

[3]En sentido estricto, el niño puede ingresar al programa desde la gestación hasta los 2 años de edad, pero durante la gestación, el beneficiario del programa no es el niño sino la embarazada, y no ha sido posible vincular el registro de la embarazada con el registro del niño después del nacimiento. Por lo tanto, el análisis se centra en los niños a partir de su nacimiento.

característica "x" no lo recibieron. Este tipo de estrategia se ha utilizado en algunos estudios de impacto de la Asignación Universal por Hijo en Argentina (Garganta y Gasparini, 2015; Garganta, Gasparini, Marchionni y Tapata; 2017), pero no es aplicable para la evaluación de este programa.

3. **Incertidumbre en la manifestación de los efectos**: Un elemento adicional que le da particularidad al programa es que, al tratarse de una política de acompañamiento familiar, basado en visitas domiciliarias y en la asistencia de los niños a centros de desarrollo, es incierto el momento a partir del cual se alcanza el efecto total sobre los resultados nutricionales, en este caso. Este efecto podría manifestarse inmediatamente después del ingreso al programa, en algún momento durante el tratamiento o recién después de la graduación. Esto no ocurre, por ejemplo, con políticas de transferencias de ingresos, cuyo impacto puede ser más inmediato. En otras palabras, es probable que el efecto promedio del tratamiento del programa varíe con el tiempo de exposición. Esta situación, es típica en los estimadores de estudios de eventos (Callaway y Sant'Anna , 2021, p. 201).

Considerando estas particularidades, se adoptaron dos estrategias metodológicas complementarias. La primera, consiste en adaptar los datos a la estructura del Modelo Canónico Tradicional de Diferencias en Diferencias 2x2, permitiendo una primera aproximación al impacto del programa. La segunda, se basa en utilizar la metodología recientemente desarrollado por Callaway y Sant'Anna (2021) de Diferencias en Diferencias con múltiples períodos de tiempo, que permite que el tratamiento haya ocurrido en diferentes momentos. Este modelo es particularmente adecuado para capturar la heterogeneidad temporal del programa. Antes de detallar cada una de las estrategias metodológicas, se describe la base de datos.

4. Descripción de la base de datos

4.1. Proceso de construcción de la base de datos y relevancia de la antropometría para la evaluación de impacto

Para la evaluación de impacto se contruyó una base conformada por 513 niños beneficiarios del programa (Grupo de Tratamiento, GT) y 457 niños seleccionados como Grupo de Control (GC). Esta base específica se denominó *Base de 1000 Días para Evaluación de Impacto* (BAMIDEI) y contiene las trayectorias antropométricas de peso, talla y, en algunos casos, perímetro cefálico, medidas con una frecuencia al menos semestral, para ambos grupos.

La elección de las variables antropométricas para realizar la evaluación se debe en parte a la limitación de los datos disponibles, ya que estas eran las únicas variables a partir de las cuales se podía recuperar y construir información a lo largo del tiempo. Sin embargo, considerando los objetivos del programa, los indicadores antropométricos son, sin duda, un indicador intrínsecamente valioso. Sirven como un indicador objetivo del bienestar humano, así como un predictor fiable de muchos resultados relevantes en etapas posteriores de la vida.

Con los datos antropométricos del GT y el GC, se calcularon las puntuaciones z de peso-para-la-edad, talla-para-la-edad y peso-para-la-talla, a través del paquete igrowup_stata desarrollado por la Organización Mundial de la Salud (OMS, 2007). Estos z-scores permiten comparar el estado nutricional y de crecimiento de los niños respecto de una población de referencia sana. De acuerdo con los criterios de la OMS, los niños con un puntaje z de peso-para-la-edad más de 2 desviaciones estándar (DE) por debajo de la

mediana de la población de referencia se consideran con bajo peso; los niños con un puntaje z de peso-para-la-talla más de 2 DE por debajo de la mediana de la población de referencia se consideran emaciados y los niños con un puntaje z de talla-parala-edad más de 2 DE por debajo de la mediana de la población de referencia se consideran con retraso del crecimiento. Si sus puntajes z están por debajo de 3 DE, los casos se consideran en situación de desnutrición grave.

De esta forma, el GT incluyó a todos los niños que ingresaron al programa hasta el 31 de mayo de 2019, descartando un solo caso desde el principio por no contar con su fecha de nacimiento, lo que impide calcular su edad, entre otros aspectos. No obstante, las trayectorias antropométricas de estos niños se reconstruyeron retrospectivamente desde el nacimiento y se extendieron hasta julio de 2021. Para 303 de ellos se dispuso además de la encuesta de ingreso al programa (línea de base), que aportó información detallada sobre la madre y/o el niño, incluyendo condiciones habitacionales, socioeconómicas, laborales, educativas, de salud y vinculares. Esto permitió complementar los registros antropométricos con un panorama integral del contexto familiar. En los restantes 209 casos solo se contó con información proveniente del Sistema de Monitoreo de Acompañamiento Familiar (SAF), un registro único de beneficiarios que consolida el historial de cada familia y permite generar alertas tempranas ante situaciones de riesgo.

Para definir el GC, se seleccionaron niños con características lo más similares posible a los del GT en términos de localización (barrios), cohortes de nacimiento (nacidos entre agosto de 2015 y mayo de 2019) y centros de salud en los que fueron atendidos, pero que no participaron en el programa. La selección se centró en los mismos barrios en los que opera el programa ya que 1000 Días se implementa en los ocho barrios más vulnerables del municipio, y elegir niños de otros barrios no garantizaría un nivel comparable de vulnerabilidad. Los niños del GC fueron seleccionados manteniendo la misma distribución por barrio que los niños del GT, que nacieron en el mismo período y asistieron a los mismos centros de salud, pero no participaron en el programa.

Para obtener los datos del GC, se contactó primero a los Centros de Atención Primaria de Salud (CAPS) de cada barrio para solicitar un listado de niños que cumplieran con los criterios especificados. Los centros proporcionaron un listado de niños que cumplían con las características requeridas (en términos de cohorte de nacimiento, equilibrio de género y no haber participado en 1000 Días) con base en los datos iniciales de un software donde se registran los casos. El listado contenía únicamente información identificativa (nombre, apellido y DNI). Posteriormente, se gestionó una autorización especial para acceder al sistema central de salud del municipio y extraer los datos antropométricos de estos niños. En el caso del GC, las mediciones antropométricas se extienden desde el nacimiento hasta septiembre de 2021.

Los datos disponibles del GC se limitan a fecha de nacimiento, sexo, barrio, peso, talla (y, en algunos casos, perímetro cefálico) y la fecha de cada medición. El sistema de salud no contiene ningún tipo de información socio-económica como podrían ser las características de la madre (edad, educación), vivienda, empleo de los adultos del hogar, etc. De todas formas, tampoco se cuenta con este tipo de información para 209 casos del GT. Esto limitó mucho el conjunto de variables de control que pudo utilizarse en las estimaciones del efecto de tratamiento.

Cabe señalar que la recopilación de datos antropométricos, tanto del GT como del GC, requirió un trabajo manual, ya que cada registro debía buscarse individualmente en el sistema de salud. Se necesitaron seis meses para completar esta tarea. Las mediciones antropométricas fueron tomadas en cada momento -y cargadas en el sistema de salud- por

los médicos pediatras de los CAPS a medida que los niños asistían a los controles de rutina de niño sano o por enfermedad. El que las mediciones hayan sido realizadas por médicos garantiza su confiabilidad.

A continuación, se presenta la Tabla 1 que muestra la distribución por sexo, cohorte de nacimiento y barrio del total de los grupos de tratamiento y control (antes de la depuración de la base según los criterios que se explicarán más adelante). Allí puede verse que los grupos están relativamente bien balanceados con respecto a estas variables. Sin embargo, se observa una proporción mayor de niños varones en relación con las mujeres en el GT en comparación con el GC, así como también una participación mayor en el GT de niños nacidos entre julio y diciembre de 2016, en tanto que en el GC se observa una participación relativa mayor de niños nacidos más tarde, entre enero y junio de 2018, y entre enero y mayo de 2019. También se observa una mayor participación de niños de los barrios Mariló/San Ambrosio/Trujui y de Sarmiento en el GC en comparación con el GT. Pero estas diferencias son lo suficientemente pequeñas como para considerar que, dadas las restricciones en la construcción de la base de datos, el balance entre los grupos es adecuado.

Tabla 1: Distribución por sexo, cohorte de nacimiento y barrio del GT y GC iniciales. En número de niños y porcentaje

	Grupo de tratamiento		**Grupo de control**		**Test de Dif de proporciones**
	Casos	%	Casos	%	
Sexo					
Varones	288	56%	228	50%	**
Mujeres	224	44%	229	50%	
Nacidos entre…					
jul-dic 2013	3	1%	0	0%	No sig.
ene-jun 2014	2	0%	**0**	0%	No sig.
jul-dic 2014	2	0%	**0**	0%	No sig.
ene-jun 2015	22	4%	**0**	0%	***
jul-dic 2015	52	10%	39	9%	No sig.
ene-jun 2016	63	12%	48	11%	No sig.
jul-dic de 2016	79	15%	54	12%	No sig.
ene-jun 2017	72	14%	70	15%	No sig.
jul-dic de 2017	68	13%	66	14%	No sig.
ene-jun 2018	67	13%	73	16%	No sig.
jul-dic de 2018	63	12%	57	12%	No sig.
ene-mayo 2019	19	4%	50	11%	***
Barrio					
Barrufaldi	65	13%	56	12%	No sig.
Don Alfonso	53	10%	44	10%	No sig.
Mariló/San Ambrosio/Trujui	63	12%	62	14%	No sig.
Mitre	89	17%	77	17%	No sig.
Obligado	69	13%	55	12%	No sig.
San Miguel Centro	57	11%	47	10%	No sig.
Santa Brígida	85	17%	78	17%	No sig.
Sarmiento	31	6%	38	8%	No sig.
Total	**512**	**100%**	**457**	**100%**	**No sig.**

Fuente: Elaboración propia a partir de los datos de BAMIDEI.

La Tabla 2 resume la incidencia de los z-score para diferentes rangos de relevancia, para cada uno de los tres índices antropométricos, para el GT y el GC calculados con la primera medición con la que se cuenta para cada niño.[4] Allí puede verse que la incidencia de z-scores negativos al nacer, pero sin caer en la condición de bajo-peso es de 44.42% del total en el GT, en tanto que en GC esto es de 35.41%. También se observa que un 7.23% de los niños que pasaron por 1.000 Dias, presentaban bajo peso, y un 9.50% bajo peso severo. Estos valores eran de 2.63% y 3.83%para el GC. Asimismo, las incidencias de z-scores negativos en los indicadores de talla-para-la-edad y peso-para-la-talla al nacer presentan valores similares a los de peso-para-la-edad.

Tabla 2: Z-score de peso-para-la-edad, peso-para-la-talla y talla-para la primera observación – Distribución en Porcentajes por rangos del Z-score.

	Peso-para-la-edad		Talla-para-la-edad		Peso-para-la-talla	
	GT	GC	GT	GC	GT	GC
Z-score >=0	38.84	58.13	27.85	35.77	59.85	69.68
-2<Z-score <0	44.42	35.41	46.00	45.50	33.33	27.87
-3<Z-score <=-2 (desnutrición)	7.23	2.63	13.56	11.68	3.89	2.20
Z-score <=-3 (desnutrición severa)	9.50	3.83	12.59	7.06	2.92	0.24
Total	100%	100%	100%	100%	100%	100%
N	**484**	**418**	**413**	**411**	**411**	**409**

Fuente: Elaboración propia utilizando el paquete igrowup stata (OMS, 2007) y los datos de BAMIDEI.

En síntesis, estos resultados iniciales justifican la aplicación de un enfoque cuasi-experimental de Diferencias en Diferencias (DD) para estimar rigurosamente el efecto del programa sobre los indicadores antropométricos.

4.2. Depuración de la base de datos para la implementación de la evaluación de impacto

Para poder implementar las dos estrategias metodológicas de DD de evaluación de impacto es necesario que las mediciones antropométricas de los niños satisfagan ciertos requisitos mínimos de frecuencia, lo cual llevó a que se redujera el tamaño tanto del GT como del GC para el análisis. La Tabla 3 describe el tamaño inicial de cada grupo, los criterios por los cuales hubo que ir descartando unidades observadas (niños) y cómo se va achicando el conjunto de observaciones como consecuencia de cada criterio de descarte. El tamaño inicial para los tres índices antropométricos fue de 512 niños para el GT y de 457 para el GC. Primero hubo que descartar, para cada índice, niños que no tuvieran mediciones en ese índice. Cabe señalar que las variables utilizadas para el descarte de unidades fueron los z-scores. En algunos casos, aunque en una cantidad marginal, hay z-scores con valores faltantes aun cuando hubiera medición de peso y/o talla. Esto se debe a que el paquete de cálculo de los z-scores descarta aquellos resultados que resulten biológicamente imposibles o excesivamente alejados de los rangos esperados

[4] Alternativamente, se elaboró la misma tabla, pero utilizando la medición de peso y talla a momento del nacimiento. Se reporta la tabla para la primera medición porque cuando se utiliza la medición al momento del nacimiento, el tamaño de la muestra se reduce mucho, especialmente para los indicadores que involucran la talla. De todas formas, las conclusiones de incidencia que se obtienen son similares.

en la población de referencia. Dichos valores atípicos—que pueden deberse a errores de medición o a datos extremos—se tratan como faltantes para evitar que influyan de manera indebida en los análisis posteriores (OMS, 2007).[5]

Descartar unidades con índices antropométricos faltantes implicó resignar un considerable tamaño muestral especialmente para los dos indicadores que involucran la talla del niño, perdiéndose casi 99 niños sin mediciones de talla-para-la-edad en el GT y 46 en el GC, y 101 en peso-para-la talla en el GT y 48 en el GC. En peso-para-la-edad la pérdida de fue de 28 niños para el GT y 39 para el GC.

También se debieron descartar niños para los cuales hubiera sólo una medición. Naturalmente, nada se podía hacer con ellos, dado que las metodologías DD requieren al menos dos observaciones (antes-después) para cada individuo.[6] Allí se perdieron 61 niños sin mediciones suficientes de peso-para-la-edad en el GT y 62 en el GC, 32 niños sin mediciones suficientes de talla-para-la edad en el GT y 70 en el GC, y 33-73 niños en GT y GC respectivamente sin mediciones suficientes del índice de peso-para-la-talla.

Luego hubo que descartar niños del GT que no tuvieran mediciones pre-tratamiento, así como también niños que, teniendo mediciones pre-tratamiento, no tuvieran mediciones posteriores al ingreso al tratamiento. Esto implicó perder otros 86 niños para el índice de peso-para-la-edad y nada menos que 214 niños en mediciones de talla para la edad, casi en su totalidad por falta de medición de talla pre-tratamiento. Así, el GT con mediciones utilizables de peso-para-la-edad quedó conformado por 337 niños, un 66% del tamaño original del grupo, en tanto que el GC para este mismo índice quedó conformado por 356 niños, un 78% del tamaño original del grupo. El tamaño final de los grupos para los índices que involucran mediciones de talla fue aún mucho menor: 167 niños en el GT y 341 en GC en talla-para-la-edad, y 161 niños en GT y 336 en GC en peso-para-la-talla; estos tamaños representan un tercio del tamaño original del GT y tres cuartos del tamaño original del GC.

Por último, en lo que respecta a las mediciones de Perímetro Cefálico, estas no fueron consideradas, ya que se contó con muy pocas observaciones: de los 512 niños iniciales en el GT, solo 65 tenían datos disponibles, mientras que en el GC, de 457 niños, únicamente 31 presentaban mediciones.

A partir de los valores de la Tabla 3 es realmente llamativo y desafortunado que para un tercio de los niños que pasaron por 1000 Días no se registraran adecuadamente los datos antropométricos de ningún índice y para otro tercio del grupo el registro de los datos antropométricos fuera incompleto, documentándose adecuadamente el peso, pero no la talla. Este problema fue más leve para los niños del GC: para el 22% no hubo registros antropométricos adecuados de ningún tipo y para un 3% adicional se registró el peso pero no la talla. La expresión 'registros adecuados' se refiere a que haya un registro sistemático y continuado del peso y la talla. Cabe resaltar que la construcción de la base de datos estuvo sujeta a los datos de las mediciones tal como habían sido registrados en el SAF para el GT (también complementada con la información del sistema de salud), y en el sistema de salud para el GC.

Para implementar la segunda estrategia metodológica, la de Callaway y Sant' Anna (2021), la depuración de los datos deja unos tamaños del GT y GC aun menores,

[5] Este es el caso de sólo 16 mediciones de peso (que quedan con z-score faltante) correspondientes a 6 niños y 13 mediciones de talla peso (que quedan con z-score faltante) correspondientes a 3 niños.

[6] También se incluyen en este criterio de descarte tres niños para los cuales la fecha de ingreso y egreso del programa coincidían.

reflejados en la Tabla 4. Esto se debe a la adecuación de los datos que hubo que hacer para implementar esta metodología. Esencialmente consiste en considerar una medición de cada niño por cada bimestre de vida desde el nacimiento. En la implementación principal se retuvo la primera medición de cada bimestre. En un ejercicio de robustez, se retuvo la última observación de cada bimestre. En esta metodología hubo que incluir un criterio adicional de depuración, el cual es que se debieron descartar los niños cuya medición pre-tratamiento correspondiera al mismo bimestre en el que ingresaron al programa.[7] De todas formas, este criterio adicional solo implica una reducción adicional marginal de la muestra. Cuando se utiliza la primera observación del bimestre, el GT y el GC del índice de peso-para-la-edad quedan en 229 y 356 niños respectivamente, los del índice de talla-para-la-edad quedan en 112 niños el GT y 339 niños el GC, y los de peso-para-la-talla en 112 niños el GT y 339 niños el GC.

Tabla 3: Proceso de depuración de datos metodología 2X2. En número de niños.

Criterios de filtrado	**Peso para la edad (zwei)**		**Talla para la edad (zlen)**		**Peso para la talla (zwfl)**	
	GT	**GC**	**GT**	**GC**	**GT**	**GC**
Muestra inicial	513	457	513	457	513	457
Eliminando…						
Mediciones faltantes para el indicador	484	418	413	411	411	409
Solo una medición (o fecha de ingreso = salida)	423	356	381	341	378	336
Sin mediciones previas al tratamiento	344	356	170	341	162	336
Con mediciones previas al tratamiento pero sin mediciones durante o después del tratamiento	337	356	167	341	161	336
FINAL	**337**	**356**	**167**	**341**	**161**	**336**

Fuente: Elaboración propia utilizando los datos de BAMIDEI.

[7] Al retener con una sola medición por bimestre, los niños que tienen su primera medición en el mismo bimestre en el que ingresan, quedan como si no tuvieran mediciones pre-tratamiento.

Tabla 4: Proceso de depuración de datos metodología C&S (2021) -manteniendo la primera medición de cada bimestre- En número de niños

Criterios de filtrado	**Peso para la edad (zwei)**		**Talla para la edad (zlen)**		**Peso para la talla (zwfl)**	
	GT	**GC**	**GT**	**GC**	**GT**	**GC**
Muestra inicial	513	457	513	457	513	457
Eliminando…						
Mediciones faltantes para el indicador	484	418	413	411	411	409
Solo una medición (o fecha de ingreso = salida)	416	353	378	339	372	334
Sin mediciones previas al tratamiento	340	353	169	339	161	334
Con mediciones previas al tratamiento pero sin mediciones durante o después del tratamiento	332	353	165	339	159	334
Con mediciones previas al tratamiento en el mismo bimestre de entrada al programa	229	353	112	339	107	334
FINAL	**229**	**353**	**112**	**339**	**107**	**334**

Fuente: Elaboración propia utilizando los datos de BAMIDEI

4.3. Balance de los grupos post-depuración y frecuencias de las mediciones

Una vez depurada la base de datos para la implementación de las dos estrategias metodológicas, se re-evaluó el balance del GT y GC, distinguiendo entre peso-para-la-edad y talla-para-la-edad, dado que el tamaño que quedó de cada grupo para cada uno de estos dos índices es significativamente distinto. Se obvia, por brevedad y similitud con los grupos de talla-para-la-edad, el de los grupos de peso-para-la-talla.

En las Tablas 5 y 6 se presenta la distribución por sexo, cohorte de nacimiento y barrio luego de la depuración de la base, tanto para la metodología 2x2 (Tabla 5), como la de Callaway y Sant´Anna, manteniendo la primera (Tabla 6). Como se señaló, también se realizó como ejercicio de robustez el análisis manteniendo la última observación.

Como resultado se encontró que las diferencias por sexo, cohorte de nacimiento y barrio son lo suficientemente pequeñas como para considerar que, dadas las restricciones en la construcción de la base de datos, el balance entre los grupos es adecuado y que estas diferencias pueden controlarse incorporando estas variables en las estimaciones. Una situación muy similar se observa en el caso de la base empleada para la metodología Callaway y Sant´Anna. En síntesis, la depuración de la base de datos no afectó el balance de la misma, y de todos modos se incorporan las variables observables como controles para las estimaciones.

Tabla 5. Base depurada metodología 2x2. En número de niños y porcentajes

	Peso-para-la-edad				**Test Dif. De propor.**	**Talla-para-la-edad**				**Test Dif. De propor.**
	GT		**GC**			**GT**		**GC**		
	Casos	**%**	**Casos**	**%**		**Casos**	**%**	**Casos**	**%**	
Sexo										
Varones	186	55%	174	49%	*	87	52%	168	49%	No sig.
Mujeres	151	45%	182	51%		80	48%	173	51%	No sig.
Nacidos entre…										
jul-dic 2013	0	0%	0	0%	No sig.	0	0%	0	0%	No sig.
ene-jun 2014	2	1%	0	0%	No sig.	2	1%	0	0%	**
jul-dic 2014	0	0%	0	0%	No sig.	0	0%	0	0%	No sig.
ene-jun 2015	10	3%	0	0%	***	1	1%	0	0%	No sig.
jul-dic 2015	29	9%	20	6%		6	4%	19	6%	No sig.
ene-jun 2016	53	16%	30	8%	***	13	8%	27	8%	No sig.
jul-dic de 2016	34	10%	40	11%	No sig.	16	10%	40	12%	No sig.
ene-jun 2017	53	16%	59	17%	No sig.	31	19%	58	17%	No sig.
jul-dic de 2017	49	15%	53	15%	No sig.	32	19%	49	14%	No sig.
ene-jun 2018	44	13%	60	17%	No sig.	26	16%	57	17%	No sig.
jul-dic de 2018	51	15%	48	13%	No sig.	32	19%	46	13%	*
ene-mayo 2019	12	4%	46	13%	***	8	5%	45	13%	***
Barrio										
Barrufaldi	44	10%	35	12%	No sig.	21	13%	34	10%	No sig.
Don Alfonso	32	12%	41	9%	No sig.	19	11%	39	11%	No sig.
Mariló/San Ambrosio/Trujui	43	14%	46	12%	No sig.	28	17%	43	13%	No sig.
Mitre	57	19%	64	16%	No sig.	16	10%	62	18%	**
Obligado	62	14%	47	17%	*	32	19%	43	13%	*
San Miguel Centro	33	11%	38	9%	No sig.	19	11%	37	11%	No sig.
Santa Brígida	50	15%	52	14%	No sig.	24	14%	50	15%	No sig.
Sarmiento	16	10%	33	4%	**	8	5%	33	10%	*
Total	337	100%	356	100%		167	100%	341	100%	

Fuente: Elaboración propia utilizando los datos de BAMIDEI

Tabla 6: Base depurada metodología C&S (2021)- manteniendo la primera medición de cada bimestre- En número de niños y porcentajes

	Peso-para-la-edad				**Test Dif. De propor.**	**Talla-para-la-edad**				**Test Dif. De propor.**
	GT		**GC**			**GT**	**GC**	**GT**		
	Casos	**%**	**Casos**	**%**		**Casos**	**%**	**Casos**	**%**	
Sexo										
Varones	128	56%	174	49%	No sig.	60	54%	171	50%	No sig.
Mujeres	101	44%	179	51%		52	46%	168	50%	
Nacidos entre…										
jul-dic 2013	0	0%	0	0%	No sig.	0	0%	0	0%	No sig.
ene-jun 2014	2	1%	0	0%	*	2	2%	0	0%	**
jul-dic 2014	0	0%	0	0%	No sig.	0	0%	0	0%	No sig.
ene-jun 2015	10	4%	0	0%	***	1	1%	0	0%	*
jul-dic 2015	20	9%	19	5%	No sig.	4	4%	18	5%	No sig.
ene-jun 2016	38	17%	30	8%	***	11	10%	27	8%	No sig.
jul-dic de 2016	21	9%	40	11%	No sig.	14	13%	40	12%	No sig.
ene-jun 2017	42	18%	59	17%	No sig.	25	22%	58	17%	No sig.
jul-dic de 2017	31	14%	53	15%	No sig.	18	16%	49	14%	No sig.
ene-jun 2018	29	13%	59	17%	No sig.	18	16%	56	17%	No sig.
jul-dic de 2018	33	14%	47	13%	No sig.	18	16%	46	14%	No sig.
ene-mayo 2019	3	1%	46	13%	***	1	1%	45	13%	***
Barrio										
Barrufaldi	34	15%	33	9%	***	11	10%	33	10%	No sig.
Don Alfonso	27	12%	41	12%	No sig.	17	15%	39	12%	No sig.
Mariló/San Ambrosio/Trujui	27	12%	46	13%	No sig.	20	18%	43	13%	No sig.
Mitre	36	16%	64	18%	No sig.	12	11%	62	18%	*
Obligado	44	19%	46	13%	***	20	18%	42	12%	No sig.
San Miguel Centro	18	8%	38	11%	No sig.	12	11%	37	11%	No sig.
Santa Brígida	30	13%	52	15%	No sig.	13	12%	50	15%	No sig.
Sarmiento	13	6%	33	9%	No sig.	7	6%	33	10%	No sig.
Total	229	100%	353	100%		112	100%	339	100%	

Fuente: Elaboración propia utilizando los datos de BAMIDEI

Si bien los tamaños del GT y GC se redujeron luego de la depuración de los datos en términos de cantidad de niños, la cantidad de mediciones válidas que quedó es lo suficientemente buena como para poder implementar las estrategias de evaluación de impacto. La Tabla 7 presenta la cantidad de niños de cada grupo con cada posible frecuencia de mediciones de peso y talla en la base de datos para el modelo 2x2, y la Tabla 8 presenta lo mismo pero para la base de datos de C&S reteniendo la primera medición de cada bimestre (la base que retiene la última medición de cada bimestre presenta una distribución de frecuencias prácticamente igual). Tanto en las mediciones de peso-para-la-edad como de talla-para-la edad, la distribución relativa de frecuencia de niños con cada cantidad total de mediciones indica que en el GT hay una baja proporción de niños con sólo dos mediciones (7% y 5% en peso y talla respectivamente en la base de 2x2; 7% y 8% en peso y talla respectivamente en la base de C&S) y alta proporción de niños con 3 o más mediciones, y es de destacar la alta proporción de niños con 9 y más mediciones.

En cambio, en el GC, esto está invertido. Esto es razonable, en tanto es de esperar que los niños de GT hayan sido controlados con mucha mayor frecuencia que los niños del GC.

Tabla 7: z-score de Peso-para-la-edad y Talla-para-la-edad. Cantidad total de controles según GT y GC. Base modelo 2x2. En número de niños y porcentaje.

Cantidad de controles totales	Peso-para-la-edad					Talla-para-la-edad				
	N° Total de niños	Grupo de tratamiento		Grupo de control		N° Total de niños	Grupo de tratamiento		Grupo de control	
		Niños	%	Niños	%		Niños	%	Niños	%
2	89	24	7.12%	65	18.26%	82	9	5.39%	73	21.41%
3	107	38	11.28%	69	19.38%	99	16	9.58%	83	24.34%
4	99	39	11.57%	60	16.85%	74	17	10.18%	57	16.72%
5	96	46	13.65%	50	14.04%	69	26	15.57%	43	12.61%
6	97	46	13.65%	51	14.33%	66	27	16.17%	39	11.44%
7	68	43	12.76%	25	7.02%	34	15	8.98%	19	5.57%
8	40	31	9.20%	9	2.53%	30	22	13.17%	8	2.35%
9	38	27	8.01%	11	3.09%	20	11	6.59%	9	2.64%
10	22	13	3.86%	9	2.53%	14	8	4.79%	6	1.76%
11	10	8	2.37%	2	0.56%	6	5	2.99%	1	0.29%
12	9	6	1.78%	3	0.84%	4	1	0.60%	3	0.88%
13	4	3	0.89%	1	0.28%	6	6	3.59%	0	0.00%
14	3	3	0.89%	0	0.00%	3	3	1.80%	0	0.00%
15	4	4	1.19%	0	0.00%	-	0	0.00%	0	0.00%
16	3	2	0.59%	1	0.28%	-	0	0.00%	0	0.00%
17	2	2	0.59%	0	0.00%	-	0	0.00%	0	0.00%
20	1	1	0.30%	0	0.00%	1	1	0.60%	0	0.00%
21	1	1	0.30%	0	0.00%	-	0	0.00%	0	0.00%
Total	693	337	100.0%	356	100.0%	508	167	100.0%	341	100%

Fuente: Elaboración propia utilizando los datos de BAMIDEI

Tabla 8: z-score de Peso-para-la-edad y Talla-para-la-edad. Cantidad total de controles según GT y GC. Base C&S - manteniendo la primera medición de cada bimestre. En número de niños y porcentajes.

Cantidad de controles totales	Peso-para-la-edad					Talla-para-la-edad				
	N° Total de niños	Grupo de tratamiento		Grupo de control		N° Total de niños	Grupo de tratamiento		Grupo de control	
		Casos	%	Casos	%		Casos	%	Casos	%
2	90	16	6.99%	74	20.96%	90	9	8.04%	81	23.89%
3	107	33	14.41%	74	20.96%	99	14	12.50%	85	25.07%
4	104	34	14.85%	70	19.83%	78	15	13.39%	63	18.58%
5	94	37	16.16%	57	16.15%	69	22	19.64%	47	13.86%
6	80	40	17.47%	40	11.33%	52	20	17.86%	32	9.44%
7	47	29	12.66%	18	5.10%	31	14	12.50%	17	5.01%
8	31	20	8.73%	11	3.12%	19	10	8.93%	9	2.65%
9	13	7	3.06%	6	1.70%	5	3	2.68%	2	0.59%
10	10	7	3.06%	3	0.85%	5	2	1.79%	3	0.88%
11	3	3	1.31%	0	0.00%	1	1	0.89%	0	0.00%
12	1	1	0.44%	0	0.00%	1	1	0.89%	0	0.00%
13	1	1	0.44%	0	0.00%	1	1	0.89%	0	0.00%
14	1	1	0.44%	0	0.00%	0				
Total	582	229	100%	353	100%	451	112	100%	339	100%

Fuente: Elaboración propia utilizando los datos de BAMIDEI

En el marco del análisis de impacto del programa, también resulta relevante conocer la distribución de observaciones según el momento de medición: 1. Pre y durante tratamiento; 2. Pre, durante y post tratamiento, y 3. Pre y post tratamiento, para los indicadores de peso-para-la-edad y talla-para-la-edad para el grupo de tratamiento (GT). Al respecto, como puede verse en las Tablas 9 y 10, la mayor cantidad de observaciones se concentra en el grupo 2 (Pre, durante y post tratamiento), seguido de Pre y durante tratamiento. Esto es así para los tres indicadores antropométricos independientemente de la base utilizada (el de peso-para-talla es similar al de talla-para-la-edad).

Tabla 9: Número de observaciones con mediciones pre, durante y post tratamiento. Base modelo 2x2. En número de niños**.**

Número total de mediciones	**Peso-para-la-edad**				**Talla-para-la-edad**			
	Con mediciones…			**Número total de niños GT**	**Con mediciones…**			**Número total de niños GT**
	Pre y durante tratamiento	**Pre, durante y post tratamiento**	**Pre y post tratamiento**		**Pre y durante tratamiento**	**Pre, durante y post tratamiento**	**Pre y post tratamiento**	
2	17	0	7	24	6	0	3	9
3	17	6	15	38	10	3	3	16
4	17	16	6	39	9	7	1	17
5	15	27	4	46	10	16	0	26
6	16	26	4	46	12	13	2	27
7	12	30	1	43	7	8	0	15
8	5	25	1	31	4	18	0	22
9 o mas	21	49	0	70	12	23	0	35
Total	**120**	**179**	**38**	**337**	**70**	**88**	**9**	**167**

Fuente: Elaboración propia utilizando los datos de BAMIDEI

Tabla 10: Número de observaciones con mediciones pre, durante y post tratamiento. Base modelo C&S. En número de niños**.**

Número total de mediciones	**Peso-para-la-edad**				**Talla-para-la-edad**			
	Con mediciones…			**Número total de niños GT**	**Con mediciones…**			**Número total de niños GT**
	Pre y durante tratamiento	**Pre, durante y post tratamiento**	**Pre y post tratamiento**		**Pre y durante tratamiento**	**Pre, durante y post tratamiento**	**Pre y post tratamiento**	
2	13	0	3	16	8	0	1	9
3	16	6	11	33	9	4	1	14
4	15	14	5	34	9	5	1	15
5	17	17	3	37	9	11	2	22
6	12	28	0	40	4	16	0	20
7	7	22	0	29	4	10	0	14
8	5	15	0	20	4	6	0	10
9 o mas	6	13	1	20	2	6	0	8
Total	**91**	**115**	**23**	**229**	**49**	**58**	**5**	**112**

Fuente: Elaboración propia utilizando los datos de BAMIDEI

5. Primera estrategia metodológica: el modelo canónico de Diferencias en Diferencias 2x2

La primera estrategia metodológica consistió en adaptar los datos a la estructura del modelo canónico de Diferencias en Diferencias (DD) 2x2. Este modelo evalúa el impacto causal de un programa calculando la diferencia entre el cambio en la variable de interés antes y después del tratamiento (primera diferencia), en un grupo de tratamiento frente a un grupo control (segunda diferencia), como se muestra en la Tabla 11.

Tabla 11: Evaluación de impacto por medio de la metodología de diferencias en diferencias

	Grupo Tratamiento	**Grupo Control**	***Diferencias***
Antes	$\bar{Y}_{TRAT}^{PRE}$	$\bar{Y}_{CONTROL}^{PRE}$	$(\bar{Y}_{TRAT}^{PRE} - \bar{Y}_{CONTROL}^{PRE})$
Después	$\bar{Y}_{TRAT}^{POST}$	$\bar{Y}_{CONTROL}^{POST}$	$(\bar{Y}_{TRAT}^{POST}$ -$\bar{Y}_{CONTROL}^{POST})$
Diferencias	$(\bar{Y}_{TRAT}^{POST} - \bar{Y}_{TRAT}^{PRE})$	$(\bar{Y}_{CONTROL}^{POST} - \bar{Y}_{CONTROL}^{PRE})$	$(\bar{Y}_{TRAT}^{POST} - \bar{Y}_{TRAT}^{PRE}) - (\bar{Y}_{CONTROL}^{POST} - \bar{Y}_{CONTROL}^{PRE})$

DD=$(\bar{Y}_{TRAT}^{POST} - \bar{Y}_{TRAT}^{PRE}) - (\bar{Y}_{CONTROL}^{POST} - \bar{Y}_{CONTROL}^{PRE})$ (1)

Donde:

$\bar{Y}_{TRAT}^{PRE}$ y $\bar{Y}_{TRAT}^{POST}$: es el resultado de interés para el GT antes y después del tratamiento y

$\bar{Y}_{CONTROL}^{PRE}$ y $\bar{Y}_{CONTROL}^{POST}$: es el resultado de interés para el GC antes y después del tratamiento.

Esa simple diferencia en diferencias es igual al coeficiente de interacción estimado de una variable ficticia del grupo de tratamiento y una variable ficticia del período posterior al tratamiento en la siguiente regresión (Angrist y Pischke, 2015):

$$Y_{it} = \beta_0 + \beta_i TRAT_i + \beta_t POST_t + \beta_{2x2} TRAT_i \times POST_t + e_{it} \quad (2)$$

Donde:

- Y_{it}, es la variable de resultado para el individuo i en el tiempo t
- $TRAT_i$, es una variable una dummy que toma valor 1 (0) si el individuo i pertenece al grupo de tratamiento (control). El subíndice i denota la variabilidad de la variable dummy entre los distintos individuos. $TRAT_i$ Controla las diferencias fijas entre las unidades que se comparan.
- $POST_t$ es una dummy que toma valor 1 (0) en el período post-tratamiento (pre-tratamiento). El subíndice t denota la variablidad de la variable dummy con el tiempo. $POST_t$ controla el hecho de que las condiciones cambian con el tiempo para todos, ya sea que reciban tratamiento o no.
- $TRAT_i \times POST_t$ es un término de interacción. Su coeficiente (β_{2x2}) es el efecto causal de Diferencias en Diferencias.

El supuesto principal detrás del modelo de DD, es el de *tendencias paralelas*, que asume que, en ausencia del tratamiento, los resultados promedio para los grupos tratado y de control habrían seguido trayectorias paralelas a lo largo del tiempo (Angrist y Pischke, 2015; Abadie, 2005). La validez de este supuesto -fundamental para evitar sesgos en la estimación- puede evaluarse graficamente o, de forma más rigurosa, a través de una

regresión que incluya interacciones entre los indicadores de tiempo y el indicador del grupo tratado para el período previo al tratamiento, tal como en la expresión (3):

$$Y_{it} = \beta_0 + \beta_i TRAT_i + \sum_{t \epsilon Pre} \delta_t (TRAT_i \times TIEMPO_t) + e_{it} \quad (3)$$

Si los coeficientes δ_t de los términos de interacción no son estadísticamente significativos, se considera que el supuesto de tendencias paralelas se sostiene (Angrist, & Pischke, 2009).

6. Adecuación de los datos del programa 1.000 Dias para la implementación del DD 2x2

6.1. Imputación de edad de ingreso

Para adaptar los datos a la estructura 2x2, el primer paso fue definir un periodo de "antes" y "después" del tratamiento. Dadas las características del programa evaluado, se definieron dichos periodos en función del *tiempo cronológico* o *edad* del niño, en lugar del *tiempo calendario* (mes y año); controlando el tiempo calendario con variables dummy que se adicionan a la especificación. Definir el "antes" y el "después" en términos de la edad del niño tiene sentido considerando que los primeros 1000 Días de vida constituyen una ventana crítica de acción en el desarrollo infantil (Heckman, 2006, 2014). En otras palabras, la edad en la que ingresó y/o egresó el niño al programa es presumiblemente mucho más importante que el tiempo calendario en el que esto ocurrió.

Sin embargo, mientras que el "antes" y el "después" son datos observables para GT (ya que se dispone de la fecha y la edad de entrada y salida del programa de cada niño), por definición, estos momentos no se observan en el GC, ya que nunca fueron tratados. Para abordar esto, se indagaron técnicas que permitieran imputar razonablemente una edad de ingreso y una de egreso del programa a cada niño del GC, con el fin de definir los periodos de tratamiento "antes" y "después". Vale la pena señalar que se imputó una edad tanto de ingreso como de egreso porque si bien parece razonable definir el "después" luego del ingreso al programa, también se consideró como ejercicio de robustez el "después" luego del egreso del programa. Se implementaron así diferentes modelos de imputación que estiman los momentos potenciales del ingreso y egreso del grupo control, a partir de las siguientes variables 'explicativas': fecha de nacimiento, sexo y barrio del grupo de tratamiento; en el caso del egreso también se incluyó el momento de ingreso estimado.

Cabe señalar que esta estrategia de, ante la variabilidad de tiempos de tratamiento, imputar un momento de ingreso (y egreso) del programa para adecuar los datos de los que disponemos a un modelo DD 2X2, es una novedad metodológica y, de acuerdo con la revisión metodológica realizada, inédita. Este enfoque puede representar un aporte metodológico interesante. Naturalmente, debe evaluarse comparando los resultados obtenidos con los obtenidos utilizando otras estrategias de estimación.

En concreto, se emplearon dos técnicas alternativas, cada una con diversas variantes de cálculo: la imputación simple, utilizando la técnica de *Predictive Mean Matching* (PMM) combinada con *K-Nearest Neighbors* (KNN); y la imputación múltiple, implementada mediante las técnicas de *Multiple Imputation by Chained Equations* (MICE) o Método Multivariado de Ecuaciones Encadenadas y el Método Monotónico. Adicionalmente, se empleó la técnica de bootstrap para replicar las estimaciones, lo que especifica que las estimaciones posteriores de los parámetros del modelo se obtendrán mediante muestreo con reemplazo. Esto permite obtener estimaciones más robustas de los errores estándar e intervalos de confianza (Efron & Tibshirani, 1993). Además, esta técnica es útil en

situaciones donde las distribuciones son asimétricas, lo cual es común en la imputación de datos faltantes (Efron & Tibshirani, 1993). Cabe señalar que el análisis principal se realizó en función de los resultaods obtenidos en la imputación múltiple con MICE, combinada con PMM y KNN, reforzada con técnicas de bootstrap.

Una vez imputadas las edades de ingreso y egreso de los niños del grupo control, se realizó una prueba de diferencia de medias entre los GT y GC para verificar que quedaran emparejados. Los resultados muestran que la edad de ingreso imputada para el GC no difiere estadísticamente de la observada en el GT. De manera similar, en los análisis de robustez, considerando el egreso como referencia, la diferencia en las medias de edad entre ambos grupos fue pequeña y no estadísticamente significativa. Estos hallazgos respaldan la validez de la estrategia de imputación y la comparabilidad de los grupos para los fines de la evaluación de impacto.

6.2 Test de tendencias paralelas

Antes de aplicar el modelo 2x2, se verificó el cumplimiento del supuesto de tendencias paralelas mediante un análisis gráfico y un ejercicio de regresión. En este último, las variables dependientes fueron los z-scores de peso-para-la-edad, talla-para-la-edad y peso-para-la-talla. Como variables explicativas se incluyeron la pertenencia al grupo de tratamiento y la interacción entre esta y los efectos del tiempo de tratamiento, capturados a través de la edad del niño previa al ingreso al programa. La ausencia de significatividad en dichas interacciones indica la existencia de trayectorias paralelas entre los grupos antes de la intervención.

Los resultados muestran que el supuesto de tendencias paralelas se cumple en los indicadores de peso-para-la-edad y peso-para-la-talla, pero no en talla-para-la-edad (Tabla 12).

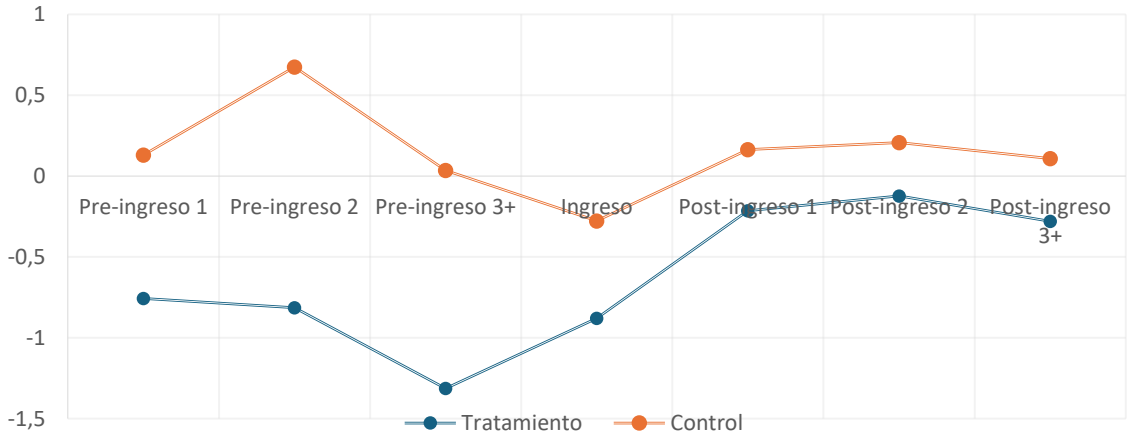
Tabla 12: Test de tendencias paralelas hasta el momento del ingreso

	Peso-para-la-edad	Talla-para-la-edad	Peso-para-la-talla
Tratado	- 1.103***	0.232	-0.661***
	(0.188)	(0.198)	(0.148)
Tratado*Edad	-0.005	-0.067***	-0.004
	(0.021)	(0.020)	(0.015)
Constante	0.228*	-0.731***	0.682***
	(0.132)	(0 .113)	(0.084)
P-valor	0.002	0.004	0.000
R cuadrado	0.017	0.014	0.037
R cuadrado aj.	0.014	0.012	0.035
bservaciones	693	508	487

Fuente: Elaboración propia utilizando los datos de BAMIDEI

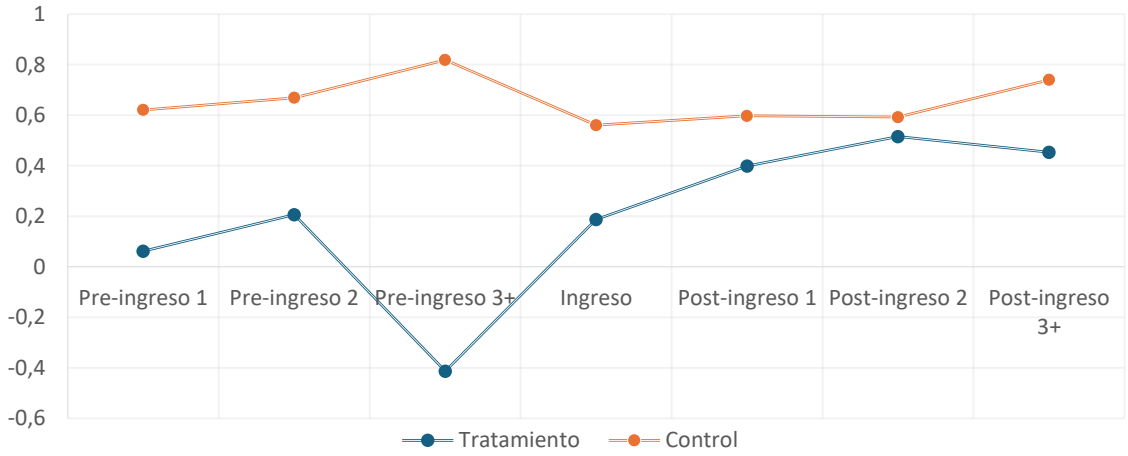
A continuación, se presenta el análisis gráfico cuyos resultados complementan esos hallazgos: antes de la intervención, el GT y GC, presentan pendientes similares en peso-para-la-edad y peso-para-la-talla, aunque con niveles distintos, reflejando la mayor vulnerabilidad inicial del grupo tratado. Tras el ingreso al programa, se observa una convergencia progresiva hacia el grupo de control, lo que sugiere mejoras atribuibles a la intervención (Gráfico 1 y 2). Eso no se verifica en talla-para-la-edad (Gráfico 3).

Gráfico 1: Z-score: Peso-para-la-edad



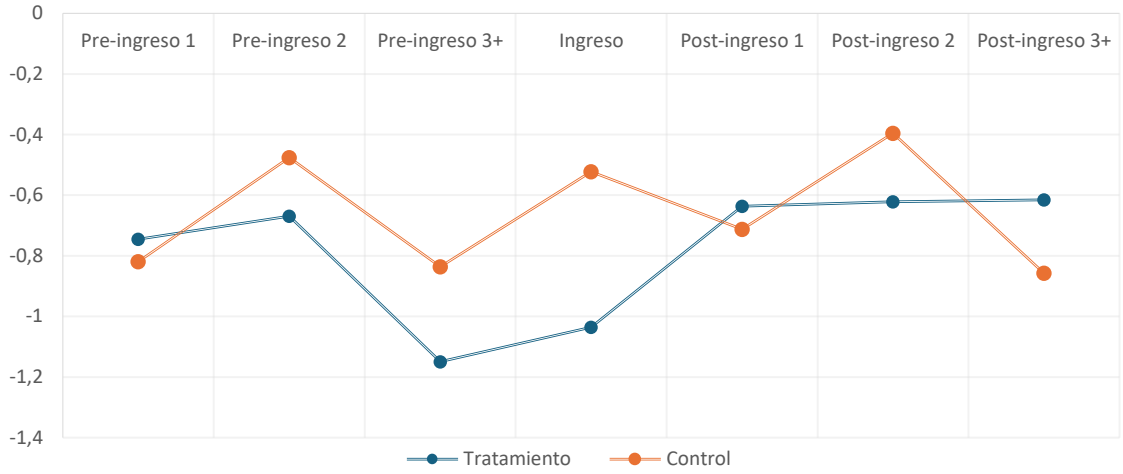
Fuente: Elaboración propia utilizando los datos de BAMIDEI

Gráfico 2: Z-score: Peso-para-la-talla



Fuente: Elaboración propia utilizando los datos de BAMIDEI

Gráfico 3: Z-score: Talla-para-la-talla



Fuente: Elaboración propia utilizando los datos de BAMIDEI

6.3. Especificación del Modelo DD 2x2 al Programa 1.000 Dias con los datos ajustados

Habiendo adecuado los datos del programa 1000 Días a la estructura de un modelo de DD 2x2, y habiendo verificado que satisface el supuesto de tendencias paralelas para la variable peso-para-la-edad y peso-para-la-talla, se especifica la siguiente regresión, en la cual el tiempo "antes y después" está determinado por la edad del niño al ingreso al programa (observada para el GT e imputada para el GC):

$$Y_{ie} = \beta_0 + \beta_i TRAT_i + \beta_e POST_e + \beta_{2x2} TRAT_i \times POST_e + \sum_{t=2016}^{2019} D_t + \sum_{b=1}^{7} B_b + e_{ie} \quad (4)$$

Donde:

- Y_{ie} representa el resultado de interés (peso-para-la-edad o talla-para-la-edad o peso-para-la-talla) para el niño i, a la edad e. Nótese que el subíndice e denota la edad cronológica del niño, la cual demarca el 'antes' y 'después' de ingresar al programa.
- $TRAT_i$, es una variable dicótoma que indica si el niño pertenece al grupo tratamiento
- $POST_e$, es una variable dicótoma que indica si el niño está en el periodo posterior al tratamiento, definido por si su edad es mayor a la edad de ingreso al programa.
- $TRAT_i \times POST_e$ captura el efecto del tratamiento.
- D_t es una variable dicótoma para cada año de ingreso (efecto fijo por año de ingreso), tomando como categoría base el año $t = 2015$, y
- B_b es una variable dicótoma para cada barrio, tomando como categoría base el barrio Obligado $b =$ Obligado (efecto fijo por barrio).

Al incluir efectos fijos por año (D_t) y barrio B_b, se está controlando por factores no observados que pueden afectar la variable dependiente y estar correlacionados con el tratamiento, mejorando la precisión de la estimación del efecto del tratamiento e incrementando la validez de las estimaciones obtenidas mediante diferencias en diferencias, evitando que estas se confundan con variables omitidas. Además, permite controlar por los pequeños desbalances observados en términos de barrios y cohortes de nacimiento entre GT y GC. De todas formas, también se estimaron modelos sin controles.

Para asegurar la fiabilidad de los resultados obtenidos, se realizaron varios ejercicios de robustez. En primer lugar, se estimó la misma regresión (4) considerando que el tiempo "antes y después" está determinado por la edad del niño al *egreso* del programa (nuevamente, observada para el GT e imputada para el GC). Se implementaron también dos pruebas placebo con el fin de evaluar si los efectos observados en el análisis principal también aparecían en escenarios sin tratamiento real. En una prueba, el GT ficticio fue el grupo de niños nacidos en el primer semestre de cada año, y el grupo de control estuvo formado por los niños nacidos en el segundo semestre de cada año. Asimismo, otro ejercicio placebo consistió en tomar como GT a todas las niñas frente a los niños que eran el grupo de control. Los resultados sugieren que los efectos encontrados con el verdadero grupo tratado son consistentes y confiables.

De cualquier manera, aun cuando la estrategia de adaptar los datos del programa 1.000 Dias al formato del modelo canónico DD 2x2 es un aporte metodológico novedoso cuyas estimaciones parecen ser robustas, deben reconocerse una limitación significativa, y es que para el estimador DD 2x2 se promedian todos los z-scores posteriores al ingreso de los niños al programa. Esto implica que se ignoran dos cuestiones importantes: (1) si el efecto del programa varía con el tiempo de permanencia en el programa, y (2) si el efecto del programa es heterogéneo dependiendo de la edad a la que el niño ingresó al programa. Para poder incorporar estas cuestiones, se adoptó una segunda estrategia metodológica, implementando la propuesta de Callaway y Sant'Anna (2021).

7. Segunda Estrategia Metodológica: Diferencias en diferencias reconociendo múltiples periodos de tiempo y tratamiento escalonado. La metodología de Callaway y Sant'Anna (2021)

El modelo de DD canónico de 2x2 resulta atractivo por su sencillez y por su correspondencia con un test de diferencia de medias en un cuadro de doble entrada. El problema surge, como en nuestro caso, cuando los datos no encajan en ese esquema simplificado, lo que es más común de lo que se cree (Goodman-Bacon, 2021, p. 255). En esos casos, suele estimarse una regresión con efectos fijos por unidad y tiempo, cuyo estimador se le llama "Estimador de efectos fijos bidireccionales de diferencias en diferencias" (Two-way fixed effects DD estimator, TWFE). Goodman-Bacon (2021) muestra que dicho estimador puede descomponerse como un promedio ponderado de múltiples comparaciones 2x2, incluidas aquellas entre grupos tratados en distintos momentos. Ello implica que, cuando existe variación en el tiempo de adopción del tratamiento, la interpretación causal del TWFE requiere supuestos fuertes, en particular que los efectos promedio del tratamiento sean constantes a lo largo del tiempo; en caso contrario, las estimaciones resultantes pueden estar sesgadas.

Cabe señalar, sin embargo, que la estrategia de adaptar el modelo de2x2, no fue el TWFE, ya que utilizamos la edad cronológica del niño al ingreso del programa como variable temporal e imputamos la edad de ingreso del grupo de control, encuadrando los datos en un esquema 2x2. En cualquier caso, dado que el programa en estudio tiene en la práctica

más de dos periodos de tiempo y variación en el momento del tratamiento, la estrategia econométrica recomendable es una como la metodología desarrollada por Callaway y Sant'Anna (2021) (en adelante, C&S). La metodología no se basa directamente en una ecuación de regresión tradicional como el modelo canónico de 2x2 o el TWFE, sino que se trata de un enfoque no paramétrico para estimar los efectos del tratamiento con múltiples períodos y adopción escalonada del tratamiento.

Esta metodología, permite estimar un modelo DD que identifica una familia de Efectos Promedio del Tratamiento (ATT) en lugar de uno solo, al que denominan efecto promedio del tratamiento grupo-tiempo. ATT(g,t) es el efecto promedio del tratamiento para el grupo g en el tiempo t, donde el grupo se define por el momento en que entró en tratamiento. Estos estimadores posibilitan, además de obtener un efecto promedio para todos los grupos tratados, analizar: (a) cómo varía el efecto del tratamiento según el tiempo de exposición al programa, (b) cómo varía el efecto promedio del tratamiento entre los grupos tratados (que entraron en diferentes momentos), y (c) cómo evoluciona el efecto acumulado del tratamiento.

En este caso, los grupos g se definieron en función del bimestre de vida en el que el niño ingresó al programa. La variable t también se define en función bimestre de vida en el que se midió al niño. Así, por ejemplo, tenemos el grupo de niños que ingresaron al programa en su primer bimestre de vida (independientemente del año), el grupo que ingresó en el segundo bimestre de vida, y así sucesivamente. A su vez, el impacto del programa se mide para cada uno de estos grupos g en diferentes momentos t de sus vidas, en su primer bimestre, segundo bimestre, y así sucesivamente, según lo permita la disponibilidad de datos. Todos los modelos estimados incluyen variables de control, a saber: sexo y edad del niño, barrios, año de nacimiento y año de cada medición antropométrica.

Asimismo, para que las estimaciones sean válidas, como en el modelo canónico, se debe cumplir el supuesto de tendencias paralelas condicionales. Sin embargo, este supuesto implica dos alternativas, según el grupo de control utilizado: el grupo nunca tratado o el aún no tratado. El análisis principal se realizó tomando como grupo de control los nunca tratados, pero se realizó una prueba de robustez con los aún no tratados.

También se implementaron diferentes tipos de estimandores de DD propuestos por Callaway y Sant'Anna, a saber:

1. DRIMP**,** Implementa un estimador DiD doblemente robusto mejorado basado en la probabilidad inversa de inclinación y mínimos cuadrados ponderados (Sant'Anna y Zhao, 2020).
2. DRIPW: Implementar un estimador DiD doblemente robusto basado en IPW y MCO. (Sant'Anna y Zhao, 2020; Callaway y Sant'Anna, 2021).
3. OR: Implementar un estimador DiD de regresión de resultados basado en MCO (Heckman et al., 1997; Callaway y Sant'Anna, 2021).
4. IPW: utiliza ponderaciones de probabilidad inversas (de recibir el tratamiento dadas sus covariables) para corregir las diferencias en las covariables entre los grupos de tratamiento y control (Abadie, 2005; Callaway y Sant'Anna, 2021).

El análisis principal se basó en el método DRIMP. Al igual que con los modelos 2x2, implementamos las mismas pruebas de placebo (niños nacidos en el primer semestre de cada año vs los nacidos en el segundo semestre, y niñas vs niños) dentro de este modelo

y los resultados tampoco fueron significativos, lo que sugiere que los resultados encontrados con el experimento real son confiables.

8. Adecuación de los datos del programa 1000 Días para la implementación de la metodología de C&S

Como se mencionó las mediciones de los niños se agruparon en bimestres de vida. La selección del uso de edad en bimestres y no otra definición temporal, como podrían ser semestres o meses, estuvo basada en encontrar un equilibrio entre tener suficiente desagregación temporal para poder analizar los efectos de un programa que se espera que tenga impacto en muy breves periodos de tiempo en la edad de los infantes, y el mínimo de agregación necesario para poder contar con suficiente cantidad de unidades de tratamiento y control en varios grupos (g, t) que permitan realizar las estimaciones.

Algo adicional para clarificar consiste en que para muchos niños se tiene más de una medición antropométrica dentro del mismo bimestre. Entonces, fue necesario decidir cuál medición utilizar. Los resultados principales de las estimaciones se realizaron tomando la primera medición de cada bimestre e ignorando las demás, pero dentro de los múltiples ejercicios de robustez realizados, se realizaron las mismas estimaciones tomando la última medición de cada bimestre.

Por otra parte, cabe señalar que dentro de las principales desventajas del uso de bimestres es que se pierden mediciones del mismo niño, y que los niños con mediciones previas al ingreso al programa, pero dentro del mismo bimestre, también se pierden, ya que aparecen como "siempre tratados" (cuando en realidad no lo están). Por lo que, hubo una depuración adicional de los datos en relación con la metodología 2x2, lo que redujo aún más los tamaños efectivos del GT y del GC, ya que fue necesario descartar a los niños cuya única medición pre-tratamiento coincidía con el mismo bimestre de ingreso al programa.

A continuación, se presenta la Tabla 13 con los resultados de la depuración de datos -manteniendo la primera medición de cada bimestre- utilizada para la estimación con el método de C&S.

Tabla 13: Cantidad y porcentaje de niños del GT de cada índice antropométrico que ingresaron al programa 1000 Días en cada bimestre de vida

Bimestre	Peso-para-la-edad		Talla-para-la-edad		Peso-para-la-talla	
de vida	Casos	%	Casos	%	Casos	%
1	0	0%	0	0%	0	0%
2	58	25%	24	21%	24	22%
3	50	22%	18	16%	16	15%
4	31	14%	16	14%	16	15%
5	19	8%	12	11%	12	11%
6	14	6%	6	5%	6	6%
7	9	4%	3	3%	3	3%
8	11	5%	9	8%	8	7%
9	12	5%	6	5%	6	6%
10	10	4%	6	5%	6	6%
11	4	2%	3	3%	2	2%
12	2	1%	2	2%	1	1%
13	0	0%	0	0%	0	0%
14	3	1%	3	3%	3	3%
15	2	1%	2	2%	2	2%
16	2	1%	1	1%	1	1%
17	0	0%	0	0%	0	0%
18	1	0%	0	0%	0	0%
19	1	0%	1	1%	1	1%
Total	**229**	**100%**	**112**	**100%**	**107**	**100%**

Fuente: Elaboración propia utilizando los datos de BAMIDEI

8.1. Tests de tendencias paralelas con la metodología de C&S

Como se ha mencionado, en un esquema de DD con múltiples periodos y distintos momentos de adopción del tratamiento, uno de los supuestos centrales para la identificación de efectos causales es el de tendencias paralelas. Este supuesto establece que, en ausencia del tratamiento, la evolución promedio de la variable resultado para el GT habría sido la misma que la del GC.

Sin embargo, para aplicar este supuesto, en la metodología C&S, se puede tomar como grupo de comparación a los "nunca tratado" (NT) o a los "aún no tratado" (ANT). Si bien el análisis principal se realizó considerando a los nunca tratados como grupo de comparación, también se replicaron los análisis como prueba de robustez utilizando el grupo de aún no tratados.

En la Tabla 14 se presentan los resultados del test de tendencias paralelas para los z-score de peso-para-la-edad, talla-para-la-edad y peso-para-la-talla, utilizando como grupo de comparación a los “Nunca tratados” y el método de estimación DRIMP. Cabe destacar que las estimaciones con DRIMP y DRIPW son exactamente iguales, lo cual sugiere que el modelo está bien especificado y no presenta problemas de colapso en las probabilidades de tratamiento (overlap). En otras palabras, cuando no hay errores de especificación ni problemas de multicolinealidad o de solapamiento, ambos métodos *doubly robust* (DRIMP y DRIPW) suelen conducir al mismo estimador.

Tabla 14: Test de tendencia paralela. Grupo de control: Nunca tratados. Método DRIMP. Base C&S - manteniendo la primera medición de cada bimestre

Indicador	Chi2	p-valor	Test de hip. (95%)
Peso-para-la-edad	0.106	0.745	NRHo
Talla-para-la-edad*	1940.264	0.0000	Rho
Peso-para-la-talla	270.350	0.000	Rho

Fuente: Elaboración propia utilizando los datos de BAMIDEI. * En talla-para-la-edad hubo que excluir a las variables dummy de los barrios para que se pudiera realizar la estimación.

La hipótesis nula en este test es que no hay violaciones a la tendencia paralela entre GC y GT (condicionado a las variables de control) previa al tratamiento. A partir de la tabla, se advierte que únicamente no se rechaza la hipótesis nula en la variable peso-para-la-edad; en cambio, sí se rechaza en talla-para-la-edad y en peso-para-la-talla. Si bien presentaremos los resultados del ATT promedio para los tres indicadores antropométricos, el resultado de este test indica que el único ATT estimado confiable es el de peso-para-la-edad.

Como ejercicio adicional de robustez se realizó el análisis con los aún-no-tratados como grupo de control. En este caso, al utilizar a los ANT como GC, el test de tendencias paralelas se rechaza para el indicador peso-para-la-edad, algo que no ocurre cuando se utilizan los NT como GC. Esto permite observar lo señalado por Goodman-Bacon (2021) de por qué la comparación tratados con aun-no-tratados -que es tácita en el modelo TWFE en vez de explícita como en C&S- puede ser problemática: precisamente porque estos grupos de unidades pueden teenr tendencias pretratamiento diferentes. A su vez, C&S enfatizan que el supuesto de tendencias paralelas cuando el GC son los ANT es más exigente que cuando se usan los NT, y que esto es una desventaja. Así, queda claro que los NT son el GC que debe utilizarse en este caso.

En cuanto al indicador de talla-para-la-edad no se rechaza el test de tendencias paralelas cuando se usan los ANT como GC, a diferencia del escenario con los NT como GC, donde sí se rechazaba. Sin embargo, el efecto si bien positivo, es no significativo. Finalmente, en el indicador de peso-para-la-talla, cuando se usan los ATN como GC, se rechaza el test de tendencias paralelas, al igual que ocurre cuando se usan los NT como GC.

9. Resultados de las estimaciones por ambas estrategias metodológicas

En esta sección se reporta y analiza el impacto del programa, sobre los z-score de: peso-para-la-edad, talla-para-la-edad y peso-para-la-talla, a través de las dos metodologías mencionadas. En primer lugar, se reporta la Tabla 14 correspondiente al Modelo Canónico 2×2 (sin controles adicionales), en el cual se estima el efecto promedio del tratamiento en cada una de las tres medidas. La Tabla 15 presenta los resultados de las estimaciones incluyendo además variables de control. En estas tablas se observa que, en el caso de la variable “Peso-para-la-edad” y “Peso-para-la-talla”, el coeficiente del efecto de tratamiento es positivo y significativo a un nivel del 1% y 5% respectivamente. Para “Talla-para-la-edad” si bien el resultado es positivo, no es significativo, y además se incumple el test de tendencias paralelas.

Tabla 14: Comparación de medias. Modelo Canónico 2x2

z-scores	GT		P-valor	GC		P-valor	Dif. En Dif.	P-valor
	Antes	Después		Antes	Después			
Peso-para-la-edad	-0.890	-0.249	0.000	0.228	0.137	0.556	0.732***	0.001
Talla-para-la-edad	-0.831	-0.641	0.080	-0.731	-0.675	0.708	0.133	0.535
Peso-para-la-talla	-0.001	0.461	0.000	0.682	0.652	0.857	0.492**	0.023

Fuente: Elaboración propia utilizando los datos de BAMIDEI

Tabla 15: Resultados Modelo Canónico 2x2 incluyendo variables de control

	Peso-para-la-edad	Talla-para-la-edad	Peso-para-la-talla
Tratado	-1.079***	-0.006	-0.709***
	(0.226)	(0.221)	-0.182
Post	-0.139	-0.0134	0.158
	(0.212)	(0.194)	-0.204
Tratado*Post	0.735***	0.0913	0.473**
	(0.219)	(0.216)	-0.209
Edad	0.00379	0.00782	-0.0142**
	-0.004	-0.006	-0.00602
Sexo	0.0191	0.115	0.0507
	(0.120)	(0.137)	-0.118
Sarmiento	-0.0365	0.242	-0.312
	(0.219)	(0.283)	-0.248
Don Alfonso	0.199	0.219	-0.174
	(0.253)	(0.293)	-0.244
Barrufaldi	0.514***	0.474**	0.068
	(0.190)	(0.231)	-0.239
Mariló/San Ambrosio/Trujui	0.231	0.214	-0.0327
	(0.200)	(0.263)	-0.217
San Miguel Centro	0.303	-0.0429	0.317
	(0.234)	(0.303)	-0.365
Mitre	0.255	0.488**	-0.0643
	(0.246)	(0.241)	-0.27
Santa Brígida	0.234	0.0731	0.0615
	(0.232)	(0.280)	-0.253
Año 2016	-0.166	-0.354	-0.46
	(0.394)	(0.708)	-0.302
Año 2017	0.381	-0.299	-0.202
	(0.386)	(0.669)	-0.267
Año 2018	0.242	0.216	-0.298
	(0.400)	(0.658)	-0.264
Año 2019	0.342	-0.0475	-0.202
	(0.397)	(0.661)	-0.258
Constant	-0.283	-1.194	0.957***
	(0.455)	(0.754)	-0.349
Casos	693	508	497
P-valor	0	0.1027	0.0032
R-cuadrado	0.039	0.016	0.018
R-cuadrado ajus.	0.0352	0.0101	0.0114
Root MSE	2.204	2.235	2.329

Error estándar robusto entre paréntesis. *** p<0.01, ** p<0.05, * p<0.1.
Fuente: Elaboración propia utilizando los datos de BAMIDEI

En segundo lugar, se presenta la Tabla 16, en la cual se confrontan los resultados obtenidos de los efectos promedio del tratamiento (ATT), con la metodología el modelo DD 2×2 (con controles) y los estimados mediante la metodología C&S (2021) utilizando el método DRIMP. Según la tabla, para "Peso-para-la-edad" el ATT es positivo y significativo en ambos enfoques; sin embargo, en el modelo 2×2 el coeficiente es de 0.735, mientras que en la metodología C&S se estima en 1.618. En el caso de "Peso-para-la-talla", el coeficiente obtenido mediante la metodología 2×2 resulta positivo y significativo, en tanto que el de la metodología C&S no alcanza significancia estadística. Finalmente, para "Talla-para-la-edad" el coeficiente es positivo, pero no resulta estadísticamente significativo en ninguna de las dos metodologías.

Tabla 16: Comparación de resultados: ATT 2x2 vs. C&S

	Peso-para-la-edad		Talla-para-la-edad		Peso-para-la-talla	
	DD 2x2 (con controles)	C&S-NT-DRIMP	DD 2x2 (con controles)	C&S-NT-DRIMP	DD 2x2 (con controles)	C&S-NT-DRIMP
Coef.	0.735***	1.618***	0.0913	0.867	0.473**	-0.142
Error estándar	(0.219)	0.4253	(0.216)	0.787	(0.209)	0.615
Observaciones	693	582	508	451	497	441

*** p<0.01, ** p<0.05, * p<0.1

Fuente: Elaboración propia utilizando los datos de BAMIDEI

Estos resultados revelan diferencias entre las dos estrategias metodológicas y los resultados son complementarios. En primer lugar, se observa que tanto en el modelo 2x2 como en la metodología C&S, se detecta un impacto positivo y significativo del programa en el indicador de peso-para-la-edad. Cierto es que hay una diferencia importante en la magnitud del efecto estimado: el modelo 2x2 sugiere que el programa 1.000 Dias incrementaría el z-score de peso-para-la talla en 0.735 DS, en tanto que de acuerdo con la metodología C&S, esto sería de 1.618 DS, más del doble. Parece razonable atribuir esta diferencia de magnitud al hecho de que el modelo 2x2 promedia todas las mediciones pre-ingreso de los niños, lo que produce estimaciones de impacto que se tienden a diluir, debido a la heterogeneidad en las edades y en las características de las unidades. En cambio, la metodología C&S permite obtener estimaciones de ATT más precisas al comparar unidades tratadas y de control (GT y GC) de manera más rigurosa, a través de emparejamientos que respetan la variación en el momento de ingreso y en las covariables relevantes. Más allá de la diferencia de magnitud, ambas estimaciones indican de manera consistente que el programa 1000 Días podría estar contribuyendo significativamente a sacar a los niños de condiciones de desnutrición.

En segundo lugar, las estimaciones por ambos modelos también coinciden en no encontrar un impacto en el indicador de talla-para-la-edad. Esto puede deberse a la reducción en el tamaño de los grupos de análisis para estas variables, lo que limita la capacidad estadística para detectar efectos significativos.

Asimismo, las estimaciones por ambos modelos difieren en el indicador de peso-para-la-talla. Con el modelo 2x2 se encuentra un impacto positivo (y válido, ya que se cumple el supuesto de tendencias paralelas), en tanto que con la metodología de C&S, no se encuentra impacto (y además tampoco se cumple el supuesto de tendencias paralelas). Ante la menor cantidad de niños con mediciones de talla, es posible que el modelo 2x2

pueda encontrar efecto ya que simplemente promedia las mediciones pre- y post-tratamiento, mientras que con C&S se buscan emparejamientos con un nivel de detalle mucho más exigente, reduciendo la cantidad de observaciones. De este modo, este es sin dudas un resultado menos robusto.

Por último, cabe señalar que, uno de los supuestos del modelos de C&S, es que el tratamiento "no se apaga" (irreversibilidad del tratamiento). Por lo que como ejercicio de robustez, se realizaron las estimaciones, ignorando las mediciones post-egreso y post-abandono. En ellas se observa que los resultados se mantienen. En particular, los coeficientes de peso-para-la-edad son muy similares y mantienen el mismo nivel de significancia estadística.

En síntesis, los hallazgos principales de la evaluación de impacto, son que es posible afirmar que el programa 1000 Días tiene un impacto positivo y significativo en el peso-para-la-talla de los niños, que estaría entre 0.7 y 1.6 DS de la mediana de la población de referencia, lo cual es una magnitud muy importante. Se encuentra además algún indicio de impacto positivo en el indicador de peso-para-la-talla. A continuación, se analizan los efectos desagregados que permite estimar la metodología de C&S para el indicador para el cual se encontró impacto con la misma, el de peso-para-la-edad.

9.1. Efectos desagregados de C&S para peso-para-la-edad

Como se mencionó anteriormente, la metodología de C&S permite obtener un *conjunto* de efectos ATT, que buscan dar respuesta a diferentes interrogantes de interés.

9.1.1. Efectos promedio según el momento en que se inició el tratamiento (Efectos específicos de grupo)

En esta sección se explora la heterogeneidad en los efectos promedio del tratamiento en función del momento de inicio de la intervención. El análisis busca responder si el impacto varía significativamente según el timing de la intervención y, en particular, identificar cuáles son los grupos que presentan los mayores efectos. Esta aproximación resulta fundamental para comprender la dinámica de la política y para determinar en qué momentos la intervención resulta más o menos efectiva.

La Tabla 17 reporta los coeficientes y errores estándar de los efectos promedio específicos para los grupos de ingreso para los cuales este efecto específico pudo ser estimado. Como se muestra en la tabla, no se presenta estimación del promedio de los efectos grupales. Eso se debe, a que para varias combinaciones de mayores valores de (g, t) (es decir, ingresos y mediciones más tarde en la vida) no hay suficiente cantidad de niños que puedan emparejarse entre el GT y el GC (satisfaciendo las variables de control). Al no tener estimaciones para varias de las combinaciones (g, t), no se reporta ese promedio. Sin embargo, aun cuando limitados, los resultados desagregados que se obtienen son sugerentes.

Tabla 17: Efectos específicos del grupo

Grupo	ATT
GAverage	(omitted)
G2	1.803*** (0.569)
G3	1.542** (0.744)
G5	0.836 (1.042)

*** p<0.01, ** p<0.05, * p<0.1. Errores estándar entre paréntesis.

Fuente: Elaboración propia utilizando los datos de BAMIDEI

Los resultados muestran que los niños que ingresaron en el segundo bimestre de vida tienen un impacto positivo, significativo y aún mayor que el ATT general estimado. A su vez, es mayor que el ATT que se observa sobre los niños que ingresaron en el tercer bimestre de vida, quienes de todas formas también exhiben un ATT muy alto. En contraste, el grupo que ingresa en el quinto bimestre (G5) presenta un efecto de magnitud mucho menor y un error estándar relativamente elevado, lo que indica una mayor incertidumbre en esta estimación. Es importante señalar que no se dispone de observaciones para niños que hayan ingresado en el primer bimestre, lo cual impide evaluar el potencial impacto de iniciar el tratamiento aún más temprano.

En síntesis, los resultados indican que, dentro del rango de bimestres con suficiente información (2,3 y 5), el programa tiene un impacto positivo y robusto en el peso-para-la-edad y que éste es mayor para el grupo que ingresa en bimestres de vida más tempranos. Este hallazgo respalda la hipótesis de que las intervenciones, cuanto más tempranas son, más pueden contribuir a mejorar el estado nutricional y de salud en la primera infancia, (Cunha y Heckman, 2007; Conti y Heckman, 2012; OMS, 2009).

9.1.2. Efectos promedio según la duración de la exposición al tratamiento (Estudio de Evento)

Por otra parte, la metodología de C&S, también permite realizar el análisis dinámico de los efectos del tratamiento, mediante un estudio de eventos que permite evaluar la evolución del impacto a lo largo del tiempo de exposición. Mediante la estimación de coeficientes para distintos períodos pre y post intervención, se pretende identificar la trayectoria del efecto del programa 1000 Días en el z-score de peso-para-la-edad. La metodología de estudio de eventos resulta especialmente adecuada para detectar si existe un retraso en la manifestación del impacto o si, por el contrario, el efecto se concentra en determinados intervalos de tiempo tras el ingreso al tratamiento.

En la Tabla 18 se reportan los efectos estimados. El coeficiente "Pre_avg" es el promedio de los efectos en los periodos previos al tratamiento y, al no ser estadísticamente significativo, sugiere que las tendencias pretratamiento son paralelas. Por otro lado, el "Post_avg" muestra un efecto positivo, lo que indica que, en promedio, el tratamiento tiene un impacto positivo en la variable peso-para-la-edad una vez que se inicia la intervención. Asimismo, no se observa un impacto inmediato y robusto en los periodos iniciales; en cambio, el efecto aparece en el 4° bimestre posterior al ingreso al programa,

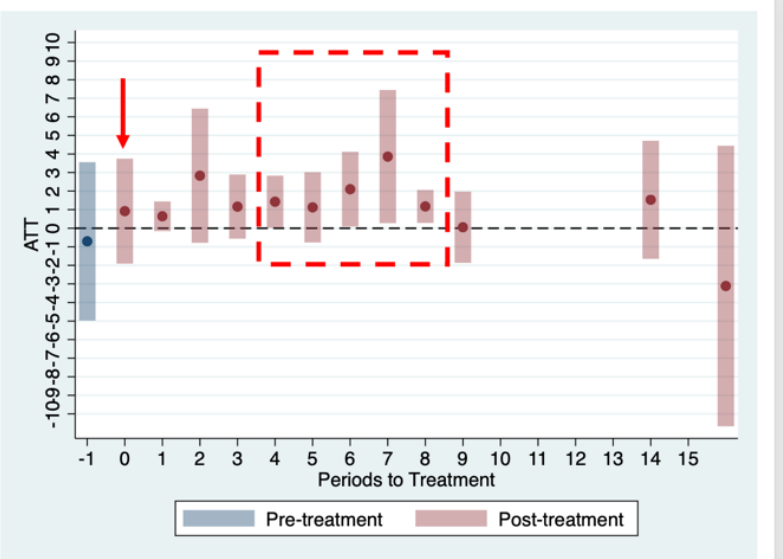
alcanzando su mayor magnitud entre el 6° y el 7° bimestre, y mostrando también un efecto alto y significativo 8 bimestres después del ingreso al programa, para luego disminuir y perder significancia en periodos posteriores. Estos coeficientes se representan también en el Gráfico 4.

Tabla 18: Resultados de Estudio de Eventos: Efectos dinámicos

Periodos pre y post ingreso al programa	**ATT**
Pre_avg	-0.707 (2.176)
Post_avg	1.143** (0.495)
Tm1	-0.707 (2.176)
Tp0	0.919 (1.442)
Tp1	0.643 (0.407)
Tp2	2.828 (1.844)
Tp3	1.162 (0.882)
Tp4	1.426** (0.716)
Tp5	1.124 (0.964)
Tp6	2.102** (1.026)
Tp7	3.858** (1.830)
Tp8	1.174** (0.454)
Tp9	0.054 (0.979)
Tp14	1.531 (1.622)
Tp16	-3.113 (3.855)

*** p<0.01, ** p<0.05, * p<0.1. Errores estándar entre paréntesis.
Fuente: Elaboración propia utilizando los datos de BAMIDEI

Gráfico 4: Efectos dinámicos del tratamiento



Fuente: Elaboración propia utilizando los datos de BAMIDEI

9.1.3. Efectos acumulativo del programa hasta diferentes edades

En este apartado se presenta el efecto acumulativo promedio de la política entre todos los grupos hasta el período $\tilde{t}$, analizado en términos de "Período Calendario". No obstante, es importante aclarar que, en este análisis, los efectos de tiempo calendario se refieren a la edad cronológica de los niños. Es decir, los coeficientes reflejan el impacto acumulativo de la intervención en función de la edad a la que se observan los resultados.

La Tabla 19 presenta los efectos acumulados estimados para distintos bimestres de vida. Resultan positivos y significativos los efectos acumulados de todos los niños que ingresaron antes del 6to, 8vo, 9no y 10mo bimestre de vida. Se destaca el coeficiente particularmente alto asociado al bimestre 9. Es decir, los niños de aproximadamente un año y medio que han ingresado al tratamiento antes de esa edad, presentan un aumento sustancial y significativo en su puntaje z de peso-para-la-edad.

Tabla 19: Efectos promedio acumulados de haber ingresado al programa antes de cada bimestre de edad

Bimestres después de ingreso	ATT (acumulado)
CAverage	1.404* 0.770
T3	6.095 6.505
T4	0.698 0.474
T5	-0.116 0.585
T6	1.375** 0.584
T7	2.828 1.844
T8	1.428** 0.718
T9	3.489** 1.472
T10	1.174** 0.454
T11	(omitted)
T14	0.054 0.979
T16	1.531 1.622
T18	-3.113 3.855

*** p<0.01, ** p<0.05, * p<0.1

Fuente: Elaboración propia utilizando los datos de BAMIDEI

En resumen, los resultados indican que el programa 1000 Días tiene un impacto significativo y de gran magnitud en el peso-para-la-edad de los niños, de entre 0.7 y 1.6 desvíos estándar, dependiendo del método de estimación. También hay evidencia, aunque menos robusta, de efecto positivo y significativo en la relación de peso-para-la-altura. Se considera que las limitaciones en la cantidad de datos de talla ha sido el principal obstáculo para identificar resultados en los indicadores que involucran esta medición.

A su vez, el análisis revela que el impacto sobre el peso-para-la-edad se concentra principalmente en aquellos niños que ingresan en el segundo y tercer bimestre de vida, y que cuanto antes ingresan, mayor es el efecto. Se encuentra también que el efecto no es inmediato, sino que se manifiesta entre el cuarto y el octavo bimestre luego del ingreso, con picos destacados en el sexto y séptimo bimestre. Por último, se evidencian efectos acumulados significativo de todos los niños que ingresaron antes de los 12, 16, 18 y 20 meses, con un impacto particularmente alto en su puntaje z de peso-para-la-edad en aquellos que ingresaron antes del año y medio año.

Estos resultados no son aislados. La evidencia empírica reciente muestra que, en otros contextos subnacionales de Argentina, los programas de visitas domiciliarias también han logrado impactos positivos en el desarrollo infantil. En particular, la evaluación del programa implementado en la provincia de Santiago del Estero por Tuñón et al. (2024) identificó mejoras significativas en dimensiones cognitivas, motrices y socioemocionales. Aunque dicho estudio no evaluó indicadores antropométricos, sus hallazgos refuerzan la hipótesis de que las intervenciones basadas en el acompañamiento familiar y la proximidad territorial pueden generar mejoras sustantivas y sostenibles en el bienestar infantil.

A su vez, los resultados aquí presentados contrastan con los hallazgos de evaluaciones de otros programas de protección social implementados en Argentina. Un estudio reciente realizado por Cherñavsky y Gherardi (2023) sobre la Asignación Universal por Hijo (AUH) no encontró efectos significativos sobre los indicadores antropométricos de los niños beneficiarios, específicamente en el peso-para-la-edad y la talla-para-la-edad. Este contraste resulta ilustrativo: mientras la AUH opera mediante transferencias monetarias no condicionadas, actuando de manera indirecta sobre los determinantes de la nutrición infantil, el Programa 1000 Días combina la provisión de bienes y servicios con un acompañamiento familiar personalizado y sostenido. Por lo tanto, la evidencia acumulada en estas investigaciones refuerza la hipótesis de que las intervenciones integrales, territoriales y basadas en el trabajo directo con las familias son más efectivas para incidir en el estado nutricional y el desarrollo infantil en contextos de alta vulnerabilidad.

11. Conclusiones

A lo largo del trabajo se mostró por medio de dos metodologías de diferencias en diferencias que el Programa 1000 Días-San Miguel tiene un impacto positivo en el z-score de peso-para-la-edad. La rigurosa depuración y construcción de la base de datos (BAMIDEI) ha permitido conformar grupos de tratamiento y control que, a pesar de presentar algunas diferencias en términos de sexo, cohorte de nacimiento y procedencia por barrio, resultan adecuadamente balanceados para el análisis de impacto. Cabe resaltar que, debido a la naturaleza no aleatorizada de la asignación al programa, los grupos presentan ciertas discrepancias iniciales, las cuales se han considerado inherentes al proceso de focalización basado en criterios de vulnerabilidad. En consecuencia, la implementación de técnicas cuasi-experimentales, como el enfoque de Diferencias en Diferencias, se justifica plenamente para aislar el efecto causal del programa, mitigando el sesgo de selección que podría derivarse de dichas diferencias estructurales. Los pequeños desbalances de los grupos son controlados incluyendo las variables de control correspondientes.

Los resultados obtenidos mediante la aplicación del modelo canónico de Diferencias en Diferencias (DiD 2×2) y la metodología no paramétrica propuesta por Callaway y Sant'Anna (2021) son altamente consistentes, pese a que la magnitud de los efectos difiere entre ambos enfoques. En particular, se destaca que, aunque ambos métodos confirman la existencia de un efecto positivo del tratamiento, la estimación del ATT derivada del enfoque de Callaway y Sant'Anna muestra un coeficiente de mayor magnitud (1.6 DS vs. 0.73 DS en el método 2x2), lo que sugiere que el impacto del programa se revela de forma más marcada cuando se captura la heterogeneidad en los tiempos de exposición. Esta metodología ha permitido saber, además, que el impacto es mayor para los niños que ingresan más temprano al programa, que el efecto se hace visible entre el cuarto y el octavo bimestre luego del ingreso y que se evidencian efectos

acumulados significativo de todos los niños que ingresaron antes de los 12, 16, 18 y 20 meses, con un impacto particularmente alto en su puntaje z de peso-para-la-edad en aquellos que ingresaron antes del año y medio. Asimismo, los ejercicios de robustez, que incluyeron tests placebos, refuerzan la validez interna del ATT estimado para peso-para-la-edad.

Con el modelo de 2x2 se encontró un efecto positivo y significativo sobre el indicador de peso-para-la-talla. Sin embargo, este resultado no goza de la robustez del resultado referido al peso-para-la-edad. Este conjunto de resultados metodológicos, al tiempo que aporta una doble aproximación al impacto del programa, evidencia además la conveniencia de adoptar estrategias alternativas en contextos donde la heterogeneidad temporal y la no aleatorización son elementos predominantes.

Desde una perspectiva de política pública, los resultados obtenidos tienen importantes implicancias, ya que la mejora en los indicadores antropométricos no solo repercute en el desarrollo físico y nutricional de la primera infancia, sino que también se alinea con los objetivos internacionales de erradicación del hambre y reducción de la pobreza. En este sentido, el impacto positivo observado respalda la continuidad y posible ampliación de intervenciones focalizadas en la primera infancia, aportando evidencia empírica de gran relevancia tanto para la política local como para la agenda de desarrollo global.

Además, los resultados obtenidos en esta investigación encuentran respaldo en evidencia reciente sobre intervenciones basadas en visitas domiciliarias en contextos subnacionales de Argentina. En particular, el estudio de Tuñón et al. (2024) sobre un programa implementado en la provincia de Santiago del Estero identificó efectos positivos y significativos en distintas dimensiones del desarrollo infantil, reforzando la hipótesis de que las políticas de acompañamiento familiar y proximidad territorial son eficaces para promover el bienestar infantil. Asimismo, los hallazgos de la investigación contrastan con los resultados de la evaluación de la Asignación Universal por Hijo (AUH), realizada por Cherñavsky y Gherardi (2023) y de Giannasi y Ramos (2024), que no encontró efectos sobre indicadores antropométricos. Este contraste resulta ilustrativo, en tanto destaca que los programas integrales que combinan provisión de servicios, acompañamiento personalizado y trabajo territorial muestran un mayor potencial para incidir directamente en la nutrición y el desarrollo infantil, en comparación con las transferencias monetarias generalizadas. De este modo, los hallazgos de esta tesis fortalecen la evidencia empírica sobre la efectividad de los programas focalizados e integrales para reducir las desigualdades en la primera infancia.

Finalmente, los hallazgos encontrados no solo demuestran la efectividad del Programa 1000 Días-San Miguel en mejorar indicadores críticos de desarrollo infantil, sino que también constituye una contribución metodológica al abordar los desafíos inherentes a la evaluación de programas no aleatorizados y con tiempos de tratamiento heterogéneos. La doble aproximación metodológica adoptada y los exhaustivos análisis de robustez permiten afirmar, con un alto grado de confiabilidad, que el efecto positivo estimado sobre el z-score de peso-para-la-edad es atribuible al impacto del programa, lo que representa un avance significativo en la evaluación de políticas públicas en contextos de vulnerabilidad en Argentina.

12. Bibliografia

Abadie, A. (2005). Semiparametric difference-in-differences estimators. The Review of Economic Studies, 72(1), 1-19.

Allison, P. D. (2001). *Missing Data*. Thousand Oaks, CA: Sage Publications.

Almond, D., Currie, J., & Duque, V. (2018). Childhood circumstances and adult outcomes: Act II. Journal of Economic Literature, 56(4), 1360-1446.

Angrist, J. D., & Pischke, J. S. (2009). Mostly harmless econometrics: An empiricist's companion. Princeton University Press.

Angrist, J., & Pischke, J. S. (2015). Mastering 'Metrics: The Path from Cause to Effect. Princeton University Press.

Athey, S., & Imbens, G. W. (2006). Identification and inference in nonlinear difference-in-differences models. Econometrica, 74(2), 431-497.

Azur, M. J., Stuart, E. A., Frangakis, C., & Leaf, P. J. (2011). Multiple Imputation by Chained Equations: What is it and how does it work? *International Journal of Methods in Psychiatric Research, 20*(1), 40–49.

Banerjee, A. V., & Duflo, E. (2011). Poor economics: A radical rethinking of the way to fight global poverty. Public Affairs.

Beretta, L., & Santaniello, A. (2016). Nearest neighbor imputation algorithms: a critical evaluation. *BMC Medical Informatics and Decision Making, 16*(1), 74.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? The Quarterly Journal of Economics, 119(1), 249-275.

Blasco, J., & Casado, D. (2009). Evaluación de impacto. Ivàlua: Institut Català d'Avaluació de Polítiques Públiques.

Blasco, J., & Casado, D. (Diciembre de 2009). Guía práctica 5: evaluación de impacto. Colección Ivàlua de guías prácticas. Catalunya, España: Ivàlua.

Bloom, H. S. (1984). "Accounting for No-Shows in Experimental Evaluation Designs." Evaluation Review, 8(2), 225-246.

CAF (2019). 3 aspectos imprescindibles para hacer una evaluación de impacto rigurosa-Recuperado de: https://www.caf.com/es/actualidad/noticias/2019/01/3-aspectos-imprescindibles-para-hacer-una-evaluacion-de-impacto-rigurosa/

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. Journal of Econometrics, 225(2), 200-230.

Card, D., & Krueger, A. B. (1994). Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. American Economic Review, 84(4), 772-793.

Cunha, F., y Heckman, JJ (2007). La tecnología de la formación de habilidades. *American Economic Review* , 97(2), 31–47. https://doi.org/10.1257/aer .

DiTraglia (2024). Lecture Notes on Treatment Effects. University of Oxford. Recuperado de: https://www.treatment-effects.com/treatment-effects.pdf

Efron, B., & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall/CR

FAO, FIDA, OPS, UNICEF y PMA. 2023. *latín América y el Caribe - Panorama regional de Seguridad alimentaria y nutrición 2023: estadísticas y tendencias.*

GARGANTA, S. & GASPARINI, L. The impact of a social program on labor informality: The case of AUH in Argentina. Journal of Development Economics, 115, pp. 99-110, 2015.

GARGANTA, S.; GASPARINI, L. & MARCHIONNI, M. Cash transfers and female labor force participation: the case of AUH in Argentina. IZA Journal of Labor Policy, 6(10), pp. 1-22, 2017a.

GARGANTA, S.; GASPARINI, L.; MARCHIONNI, M. & TAPPAT., M. The effect of cash transfers on fertility: evidence from Argentina. Population Research and Policy Review, vol. 36, n. 1, pp. 1-24, 2017b.

Gertler, P. J., Heckman, J. J., Pinto, R., Zanolini, A., Vermeersch, C., Walker, S., Chang, S. M., & Grantham-McGregor, S. (2014). Labor market returns to an early childhood stimulation intervention in Jamaica. Science, 344(6187), 998-1001.

Gonzalez, MS & Santos, ME (2023a), "Mil Días: Un Programa para la Primera Infancia Vulnerable en Argentina: Focalización, factores de riesgo de deserción y correlatos del tiempo hasta la graduación". *Niño: Cuidado, Salud y Desarrollo* (en prensa). Documento de la investigación: 10.1111/cch.13030

González, Maria Sol & Santos, Maria Emma (2023b), " Ciudades sustentables, inversiones inteligentes: una caracterización de "Mil Días-San Miguel", un programa para la primera infancia vulnerable en Argentina ". *Sustentabilidad* . https://doi.org/10.3390/su151612205

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. Journal of Econometrics, 225(2), 254-277.

Hausmann, R. (2016). METHODS OF PROGRAM EVALUATION: AN ANALYTICAL REVIEW AND IMPLEMENTATION STRATEGIES. Nutrition Economics, 205-230.

Heckman, J. Desarrollo humano y desarrollo de la primera infancia. Contribución especial. En Informe sobre Desarrollo Humano 2014. Mantener el progreso humano: reducir las vulnerabilidades y aumentar la resiliencia; Naciones Unidas: Nueva York, NY, EE.UU., 2014.

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. Science, 312(5782), 1900-1902.

Heckman, J. J., Ichimura, H., & Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. The Review of Economic Studies, 64(4), 605-654.

Heckman, JJ Formación de habilidades y economía de la inversión en niños desfavorecidos. Science 2006, 312, 1900–1902.

Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. Journal of Economic Literature, 47(1), 5-86.

Kleinke , K. (2017). Imputación múltiple bajo supuestos de patrones faltantes violados. Métodos de información en medicina, 56(06), 491-496.

Lechner, M. (2011). The estimation of causal effects by difference-in-difference methods. Foundations and Trends in Econometrics, 4(3), 165-224.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical Analysis with Missing Data* (2nd ed.). Wiley.

Molina, I. (2018). [Referencia específica a la publicación de Molina sobre MICE].

Morris, TP, White, IR y Royston, P. (2014). Ajuste de la imputación múltiple mediante correspondencia de medias predictiva y extracciones de residuos locales. Metodología de investigación médica de BMC, 14(1), 1-13.

Organización Mundial de la Salud (OMS), (2007). WHO Child Growth Standards: STATA igrowup package. Geneva: WHO. Disponible en: https://www.who.int/tools/child-growth-standards/software

Raghunathan T.E., Solenberger P.W., Van Hoewyk J. (2002) IVEware: Imputation and Variance Estimation Software User Guide, Ann Arbor, MI, University of Michigan; http://www.isr.umich.edu/src/smp/ive/ *[19 May 2008]*. [Google Scholar]

Ríos Avila (ND). CSDID Version. Recuperado de: 1.6https://friosavila.github.io/playingwithstata/main_csdid.html

Rios-Avila, F., Callaway, B., & Sant'Anna, P. H. C. (2021, August). *csdid: Difference-in-Differences with Multiple Time Periods in Stata*. Presentación en la Stata Conference. Recuperado de https://www.stata.com/meeting/us21/slides/US21_SantAnna.pdf

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons.

Rüttenauer, T., & Aksoy, O. (2024, February 1). When Can We Use Two-Way Fixed-Effects (TWFE): A Comparison of TWFE and Novel Dynamic Difference-in-Differences Estimators. https://doi.org/10.31235/osf.io/cpvzf

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. Chapman & Hall.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*(2), 147–177.

StataCorp. (2021). Stata Multiple-Imputation Reference Manual. Release 17. College Station, TX: StataCorp LLC.

StataCorp. (2023). *Stata Statistical Software: Release 16*. StataCorp LLC.

Van Buuren , S. y Groothuis-Oudshoorn , K. (2011). ratones: imputación multivariada mediante ecuaciones encadenadas en R. Journal of Statistical Software, 45(3), 1-67.

Van Buuren, S. (2018). *Flexible Imputation of Missing Data* (2nd ed.). CRC Press.

Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). MICE: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software, 45*(3), 1-67.

Vink , G., Frank, LE, Pannekoek, J. y van Buuren , S. (2014). Se revisó la imputación predictiva de coincidencia de medias de datos semicontinuos. estadística Neerlandesa , 68(1), 61-74.

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. Statistics in Medicine, 30(4), 377-399.

White, IR, Royston, P. y Wood, AM (2011). Imputación múltiple mediante ecuaciones encadenadas: problemas y orientación para la práctica. Statistics in Medicine, 30(4), 377-399.

Wooldridge, J. M. (2002). Econometric analysis of cross section and panel data. MIT Press.