

**Ministerio de Desarrollo e Inclusión Social  
Viceministerio de Políticas y Evaluación Social**

**Dirección General de Seguimiento y Evaluación-DGSE**

**Nota Metodológica para la Evaluación de Impacto del Programa  
Nacional de Alimentación Escolar *Qali Warma*<sup>1</sup>**

**Lima, 13 de Diciembre de 2013**

---

<sup>1</sup> Esta nota metodológica ha sido preparada por Stanislao Maldonado (Departamento de Economía Agrícola y de Recursos, Universidad de California en Berkeley), consultor de la Dirección General de Seguimiento y Evaluación (DGSE) del Ministerio de Desarrollo e Inclusión Social (MIDIS). El autor contó con la estrecha colaboración de Elisa Seguín, Karina Santivañez, Luis Baiocchi y Annie Chumpitaz del equipo principal de la DGSE y trabajó bajo la supervisión de Úrsula Martínez, Directora de Evaluación de la DGSE. El equipo agradece a los funcionarios del programa por la colaboración en el proceso de elaboración de esta nota y los comentarios de Indu John-Abraham, Alessandra Marini y Verónica Silva. Sarita Oré, Yasmin Villaizán y Victor Huamaní brindaron asistencia en la elaboración de esta nota. Los errores y omisiones son de exclusiva responsabilidad del autor y no comprometen a las personas anteriormente mencionadas ni a las instituciones para las que trabajan.

Correo Electrónico: [smaldonadoz@berkeley.edu](mailto:smaldonadoz@berkeley.edu)

## 1. INTRODUCCIÓN

El Programa Nacional de Alimentación Escolar *Qali Warma* fue creado en el mes de mayo de 2012 como respuesta del Gobierno del Perú a los problemas encontrados en los programas de asistencia alimentaria dirigidos a la población escolar que formaban parte del antiguo Programa Nacional de Asistencia Alimentaria (PRONAA). Las evaluaciones de proceso de dicho programa daban cuenta de una serie de deficiencias en términos de focalización y cobertura, la satisfacción de los usuarios en relación a los servicios provistos, impactos sobre nutrición, y aspectos de gestión (MIDIS 2012).

*Qali Warma* tiene como objetivo central mejorar la ingesta de alimentos, facilitando las condiciones para el aprendizaje de niñas y niños de nivel inicial y primaria de IIEE públicas. A fin de decidir la conversión del programa en una política pública, la Ley de creación del programa señala que al final de un periodo de 3 años, este será sometido a una evaluación de impacto, de modo tal que sea posible determinar su efectividad en términos del logro de los objetivos que motivaron su creación (DS N° 008-2012-MIDIS, Artículo 1). En términos de la evaluación del programa, se considera que el mismo contribuirá a mejorar las condiciones de aprendizaje mediante la mejora de los niveles de atención y memoria de corto plazo. Esta constituye la variable dependiente central del diseño de evaluación que se presenta en este documento.

El presente documento tiene como objetivo delinear los elementos básicos a tenerse en cuenta para la evaluación del impacto del programa *Qali Warma*. Dichos elementos se formulan teniendo como referencia la literatura internacional sobre la evaluación de programas de alimentación escolar y los avances recientes en las técnicas de evaluación de impacto. Así, a partir de este análisis, se propone que la Evaluación de impacto del PNAE *Qali Warma* siga un diseño de evaluación basado en un modelo de diferencias en diferencias con asignación al tratamiento a nivel de cluster, siendo este el distrito. Esta selección se realiza luego de comparar con una estrategia alternativa, basada en una regresión discontinua, y toma ventaja de un aspecto de la implementación del programa: la cobertura progresiva de escuelas no tratadas previamente por el Programa Nacional de Asistencia Alimentaria – PRONAA, en función al nivel de pobreza del distrito.

La estrategia de avance del programa consiste en atender a todas las instituciones educativas que eran beneficiarias de PRONAA a diciembre de 2012, e ir incorporando de manera progresiva a aquellas que no habían sido beneficiarias, empezando por las ubicadas en distritos del quintil I de pobreza (más pobre) en 2013, y así sucesivamente hasta llegar a la cobertura total en el año 2016.

El resto del documento se organiza del modo siguiente. La sección 2 discute aspectos vinculados con los antecedentes del programa. La sección 3 cubre aspectos conceptuales sobre la relación entre los programas de alimentos y las variables de interés en el marco de evaluación, mientras que la sección 4 presenta una breve discusión de la literatura. La sección 5 establece los objetivos y formula las preguntas de evaluación. La sección 6 discute las estrategias de evaluación mientras que la sección 7 presenta un conjunto de elementos básicos (así como algunas

extensiones) para el análisis de la evaluación. La sección 8 discute aspectos de medición de la variable central de evaluación, presenta el análisis de potencia y el cálculo de muestra, así como aspectos relacionados con el diseño muestral. El documento concluye con un conjunto de observaciones respecto de los elementos metodológicos considerados en el diseño.

Adicionalmente, el documento incluye varios apéndices que detallan elementos del diseño de evaluación. En el Apéndice I y II, se cubren aspectos generales del diseño del programa y su relación con la evaluación del mismo. En el primero de ellos, se presenta la teoría de cambio del programa, mientras que, en el segundo, se introduce la matriz de evaluación, la cual detalla las preguntas de investigación, los indicadores de impacto y las fuentes de información. En el Apéndice III, se discuten formalmente las técnicas econométricas a ser utilizadas en el contexto de la evaluación de *Qali Warma*. El Apéndice IV presenta una revisión extensa de la literatura internacional, así como algunos trabajos locales sobre el impacto de programas de alimentación escolar en donde se describen las intervenciones, las variables de impacto, las técnicas utilizadas y los principales hallazgos. El Apéndice V discute la medición de la variable de impacto de atención y memoria de corto plazo y presenta una revisión de los instrumentos utilizados en la literatura. El Apéndice VI muestra los temas de investigación cubiertos en los instrumentos de recolección de información. El Apéndice VII introduce los aspectos técnicos de la estimación del tamaño de muestra. Finalmente, el Apéndice VIII incluye mapas elaborados con el objetivo de mostrar la distribución espacial de los distritos tratados y controles en el marco de la evaluación del programa para el diseño original de la intervención. Es importante notar que, producto de los cambios implementados durante el primer año de la intervención, no es posible determinar la composición exacta de los distritos tratados y controles por lo que la discusión sobre estos aspectos es referencial.

## **2. ANTECEDENTES**

### **2.1. Información Básica del Programa**

*Qali Warma* es un programa social de alimentación dirigido a niños (as) a partir de 3 años que asisten a instituciones educativas públicas de nivel pre-escolar y primaria, a nivel nacional. Se caracteriza por tener un sistema de compras descentralizadas, operando bajo un modelo de cogestión, con representantes de la sociedad civil y la comunidad escolar, y brindando un servicio alimentario de calidad, adecuado a los hábitos de consumo locales.

Su objetivo final es mejorar la ingesta de alimentos, facilitando las condiciones para el aprendizaje de niñas y niños de nivel inicial y primaria de IIEE pública. Por ello, el programa pretende enfrentar el problema principal de inadecuada ingesta de alimentos en cantidad y calidad, mediante el aseguramiento de una provisión balanceada en cantidad y calidad a los niños, durante todos los días del año escolar.

El modelo de cogestión de *Qali Warma* establece la constitución de Comités de Compra (conformados por autoridades del gobierno local, sector salud y padres de familia) para la gestión del proceso de compra descentralizado, para lo cual se le transfieren recursos financieros.

Asimismo, se conforman los Comités de Alimentación Escolar (CAE), cuya función central es la verificación del servicio alimentario en la escuela. El servicio de *Qali Warma* tiene dos componentes: alimentario y educativo. El componente alimentario está directamente asociado con la provisión del servicio, adecuado a los hábitos de consumo local, y con los contenidos nutricionales adecuados a los grupos de edad. En esta línea, la dosificación de la alimentación se da de manera diferenciada, según quintiles de pobreza. Así, en los colegios ubicados en distritos de menores quintiles (I y II), se otorgan 2 comidas (desayuno y almuerzo), mientras que en otros distritos (quintiles III, IV y V), se otorga sólo una comida (desayuno).

Conforme a su diseño, Qali Warma tiene dos modalidades de atención: entrega de raciones preparadas y entrega de canastas de productos. En el primero de los casos, los proveedores se encargan de la preparación diaria de los desayunos, y almuerzos en los casos que corresponde, y los entregan en las instituciones educativas hasta treinta minutos antes del horario establecido para el consumo. Por su parte, bajo la segunda modalidad, los proveedores se encargan de la entrega periódica de canastas de productos no perecibles, y son los Comités de Alimentación Escolar (CAE) los que se encargan de la preparación de los alimentos en la institución educativa y su distribución a los escolares.

Por su parte, el componente educativo busca promover mejores hábitos de alimentación en los estudiantes y sus familias, a través del fortalecimiento de capacidades en equipos técnicos territoriales, miembros del Comité de Compras y del CAE.

Asimismo, la acción del programa comprende el seguimiento y monitoreo de las actividades de los CAE, a fin de asegurar una adecuada provisión de alimentos en los usuarios finales.

El ámbito de intervención del programa es universal<sup>2</sup>; sin embargo, la estrategia de intervención es progresiva cada año. Así, en el primer año (2013), se ha atendido a las IIEE beneficiarias de PRONAA a diciembre de 2012, y además, a las IIEE ubicadas en los distritos de quintil I de pobreza. Para el año 2014, se atenderá a esta misma población, y se planea llegar al resto de quintiles hacia el año 2016.

Los aspectos centrales del diseño del programa se discuten en Alcázar (2012). El marco lógico del programa asume como problema central la existencia de una inadecuada ingesta de alimentos por parte de las niñas y niños que asisten a las instituciones educativas públicas de nivel inicial y primaria, con lo cual la mejora de la ingesta, mediante un servicio de calidad, constituye el resultado específico de la intervención. El Apéndice I presenta la teoría de cambio que se desprende de este diseño, el mismo que se discutirá con más detalle más adelante.

El programa ofrece un conjunto de recetas definidas en función de su contenido nutricional. Estas se elaboran a partir de las características culturales, geográficas, ecológicas, económicas y sociales, definiéndose de este modo 8 regiones alimentarias. Por lo general, los desayunos distribuidos a los beneficiarios están conformados por bebidas semi-espesas y

---

<sup>2</sup> Brindará atención a todos los niños y niñas matriculados en educación pre-escolar y primaria, en instituciones educativas públicas a nivel nacional.

mazamorras espesas elaboradas a base de leche y cereales, cuya densidad calórica varía entre 0.6 y 0.8 Kcal/g. También incluyen alimentos sólidos como panes o galletas. En el caso de los almuerzos, estos usualmente contienen cereales, tubérculos y productos en base a carne.

## **2.2. Características del Programa y su relación con la evaluación**

El diseño de evaluación del programa debe contemplar las posibles fuentes de heterogeneidad en las respuestas de los participantes al tratamiento. Dichas fuentes son producto de diferencias institucionales, geográficas y socio-demográficas entre los individuos de la población objetivo. Estas se deben también a las características del programa y a diferencias entre los tratamientos implementados. Esta información debe ser colectada durante el recojo de la línea de base, de modo tal que sea posible estudiar el rol de estos factores para estimar, no solo el impacto del programa sino también los mecanismos que lo explican. No obstante lo anterior, la decisión de incorporar el análisis de estas heterogeneidades dentro del diseño depende también del costo en términos de recojo de información, por lo que un análisis costo-beneficio es necesario en este sentido.

En las líneas que siguen, se esboza algunas características básicas del programa y cómo estas crean retos a la estrategia de evaluación.

### ***a. El tratamiento del programa es múltiple: Qali Warma opera con dos modalidades de atención: raciones, en el caso de zonas urbanas, y productos, en el caso de zonas urbanas y rurales***

Conforme a su diseño, Qali Warma tiene dos modalidades de atención: raciones preparadas y canastas de productos<sup>3</sup>. En el primero de los casos, los proveedores se encargan de la preparación diaria de los desayunos -y almuerzos en los casos que corresponde-, y los entregan en las instituciones educativas hasta treinta minutos antes del horario establecido para el consumo. Por su parte, bajo la segunda modalidad, los proveedores se encargan de la entrega periódica de productos no perecibles, y son los Comités de Alimentación Escolar (CAE) los que se encargan de la preparación de los alimentos en la institución educativa y de su distribución a los escolares. En ambas modalidades, son los CAE los encargados de recibir las raciones / productos de los proveedores y hacerlos llegar a los estudiantes.

En ambas modalidades, se establece que se sigan las mismas recetas definidas por el programa, a fin de que se cumplan los mismos requerimientos calóricos. Sin embargo, el hecho de que sean preparados por distintos agentes (proveedores por un lado, y CAE por otro), y de que no exista una supervisión permanente a nivel de escuela, puede introducir diferencias adicionales entre ambos tratamientos.

---

<sup>3</sup> Cabe indicar que, en caso de que no sea posible adjudicar un proceso (ya sea de raciones o de productos) debido a la ausencia de proveedores que cumplan con los requisitos solicitados, el programa adjudica de manera directa a través de una canasta básica de emergencia, la cual constituye un subconjunto de las recetas inicialmente programadas para la zona.

**b. Además del componente alimentario, Qali Warma incluye un componente educativo sobre temas nutricionales, hábitos alimenticios y gestión.**

En principio, uno podría pensar que la inclusión del componente educativo puede tener sinergias importantes con el componente alimenticio, más aún si se tiene en cuenta que la unidad de análisis es la escuela<sup>4</sup>. Esta capacitación no sólo toma en cuenta la provisión de información sobre prácticas nutritivas saludables para los estudiantes, sino también incluyen capacitación para los comités de alimentación local (CAE) y para los docentes.

Si la capacitación de los CAE es crítica para la provisión del servicio alimenticio, uno debería observar una asociación positiva entre el performance desempeño del CAE y el logro alimenticio de la población objetivo. Las implicancias de incorporar a los CAE en el diseño de la evaluación podrían ser fundamentales.

En el caso de los profesores, es importante mencionar que, diferencias entre colegios en relación a con las características de los maestros, pueden tener implicancias en los efectos del programa. En un sentido más general, uno podría pensar que el impacto del programa en el logro educativo de la población objetivo depende críticamente de los docentes. En tanto que el programa sólo crea condiciones para que los estudiantes mejoren sus niveles de atención, el impacto final del mismo sobre el aprendizaje dependerá de los docentes.

**c. El componente alimentario del programa contiene distintos niveles de tratamiento en función al nivel de pobreza y a los hábitos de consumo locales**

Cuando el tratamiento difiere en intensidad, la evaluación toma en cuenta el hecho de que la respuesta de los individuos difiere en función al nivel del tratamiento al que son expuestos. En este caso, el programa contempla la provisión de desayunos y almuerzos para los potenciales beneficiarios con niveles altos de pobreza (quintiles I y II) y sólo almuerzos para aquellos con niveles bajos (quintiles III, IV y V).

Adicionalmente, el programa varía la dieta en función a las características locales y a la oferta de alimentos disponible a nivel local. Esto induce una potencial heterogeneidad en las respuestas de los individuos al programa.

**d. El diseño institucional del programa incluye la participación de la comunidad a través de los Comités de Compra, los CAE, las organizaciones de la sociedad civil y el sector privado**

El modelo de cogestión del programa descansa sobre los Comités de Compra, encargados de las compras descentralizadas de los productos o raciones, y de los Comités de Alimentación Escolar (CAE), encargados de la provisión del servicio alimentario en las escuelas<sup>5</sup>. Dado que el elemento de cogestión ya existía previamente con el PRONAA a

---

<sup>5</sup> Las funciones de los CAE son diferenciadas, dependiendo de la modalidad de atención que recibe la institución educativa a la que pertenecen. En el caso de raciones, los CAE se encargan de la recepción y

través de los CAE, es de suponer que las unidades escolares que serán expuestas al tratamiento difieren también en los niveles de organización y experiencia de las organizaciones sociales y de base sobre las cuales se construirán los CAE bajo el nuevo esquema institucional a implementarse con *Qali Warma*. Adicionalmente, entre los modelos alternativos de gestión a incorporarse en el diseño del programa, también se contempla la participación de organizaciones de la sociedad civil y el sector privado, así como el eventual involucramiento de los gobiernos locales.

Dado que -en la práctica- la implementación del programa descansa críticamente en estas organizaciones, es fundamental incluir un análisis detallado de sus características, fortalezas y debilidades, prestando particular interés a los factores que podrían explicar diferencias en relación al desempeño de estas organizaciones y por tanto inducir diferencias en términos del impacto del programa. En ese sentido, es fundamental tener mayores elementos de juicio para pensar el rol de los CAE, las organizaciones de la sociedad civil, el sector privado y los gobiernos locales en el marco de la evaluación del programa.

***e. La cobertura del programa será universal***

El programa se propone llegar a todos los niños en edad escolar (inicial y primaria) de instituciones educativas públicas. Esto constituye un reto para la evaluación del programa, en tanto que hace complicada la construcción del contrafactual (qué hubiera pasado con la población tratada en caso que no hubiese recibido el tratamiento). Por esta razón, es fundamental comprender la estrategia de implementación del programa en la medida que de esta dependerá críticamente la estrategia de evaluación.

Una alternativa consistiría en *aprovechar el hecho que la cobertura universal no es automática y sacar ventaja de las diferencias en términos de cobertura para estimar el efecto del programa*. Dicha estrategia tendrá sentido dependiendo de la velocidad en la que el programa vaya incorporando a los beneficiarios potenciales y en la manera en que dicha incorporación se realice. Este último punto es importante en tanto que pueden crearse sesgos en el proceso de selección de los beneficiarios que harían difícil la estimación precisa del impacto del programa<sup>6</sup>.

De acuerdo con el diseño del programa, se desprende que los primeros en ser atendidos serán aquellas zonas que ya contaban con el PRONAA. Esto podría ser eventualmente un reto para la evaluación en tanto que el proceso de selección de beneficiarios de este programa habría dependido mucho de factores no observables que podrían eventualmente estar correlacionados con el tratamiento de interés<sup>7</sup>. Es mucho más

---

distribución a los estudiantes. Por su parte, los CAE de productos se encargan de la recepción, almacenamiento, preparación y distribución a los estudiantes.

<sup>6</sup> Esto puede minimizarse en tanto la incorporación de los beneficiarios siga un patrón establecido mediante reglas o indicadores.

<sup>7</sup> Este punto es importante y se relaciona con el rol de las organizaciones de base en la provisión de los servicios del programa. Mayor análisis sobre este punto es necesario para el desarrollo de la estrategia de evaluación.

útil para la estrategia de evaluación si el proceso de selección se basa en un indicador observable, como es el caso de un índice de pobreza. En ese escenario, podría formularse una evaluación del efecto del programa mediante el uso de una regresión discontinua.

**f. *El programa creará externalidades a nivel de hogar***

La presencia de externalidades podría subestimar el efecto del programa si se compara individuos dentro de una misma unidad escolar tratada. Es posible también que el programa tenga un impacto directo sobre el bienestar de los hogares a los que pertenecen los individuos tratados. Este impacto debería ser contabilizado como parte de los efectos del programa. El programa podría, entre otras cosas, inducir a una re-asignación de recursos dentro del hogar o crear externalidades positivas por las mejoras en las prácticas nutricionales asociadas con el programa en los hogares de los miembros de los CAE (que, por lo general se tratarán de padres de familia).

La incorporación de estas externalidades en el diseño representará un reto para los evaluadores. Identificar todas estas fuentes posibles de externalidad requiere disponer de un marco conceptual que sirva de guía para la identificación de los efectos posibles, así como información más detallada respecto a la implementación práctica del programa.

**g. *El programa creará externalidades a nivel de la economía local***

Un elemento nuevo de *Qali Warma* en relación con el PRONAA es la compra de insumos a nivel local. Por diversas razones, esto podría tener implicancias importantes para la evaluación de impacto del programa. La compra de insumos a nivel local significará un *shock* de ingresos para los hogares de los productores locales, lo cual podría afectar la performance alimenticia de los niños en edad escolar que pertenecen a dichos hogares, en caso variara la canasta de alimentos adquiridos en el hogar. Del mismo modo, el aumento de la demanda por alimentos en las economías locales puede inducir un alza de precios, lo cual podría tener impactos negativos sobre los indicadores nutricionales de la población objetivo. Una decisión metodológica importante consistirá en evaluar si conviene incorporar estas dimensiones en la estrategia de evaluación.

**h. *El programa interactuará con programas con objetivos similares o relacionados***

Es muy probable que el programa opere en contextos en donde existen otros programas con objetivos similares o relacionados como Juntos, Cuna Más y Pensión 65. Un aspecto interesante de la evaluación de *Qali Warma* consistiría en determinar cómo el programa puede tener impactos diferenciados en función a la oferta de programas sociales a nivel local.

Es importante tener en cuenta que, aunque conceptualmente es posible identificar un conjunto posible de impactos del programa, ello no significa que la evaluación deba incorporar necesariamente todos estos impactos en el diseño por razones de costo y tiempos. La decisión dependerá de los objetivos del Programa, así como la evidencia internacional.

### 3. MARCO CONCEPTUAL Y TEORÍA DE CAMBIO

Un programa de alimentación escolar tiene, dependiendo de la naturaleza de la intervención, la capacidad de influenciar positivamente sobre un conjunto amplio de indicadores de participación escolar (matrícula, asistencia, deserción escolar y progresión, entre otros) y de logro educativo (desempeño en evaluaciones estandarizadas). Esto se da a través de su impacto sobre resultados intermedios como el estado de salud y nutrición, los procesos cognitivos de corto plazo (como la atención y memoria de corto plazo), el hambre de corto plazo y el desarrollo cognitivo, los cuales tienen un valor intrínseco más allá de su rol en términos de mejoras en el logro educativo.

Sin embargo, es importante notar que, en un programa de esta naturaleza, el efecto dependerá de una serie de factores como las características previas de la población objetivo, el tamaño de la transferencia y las condiciones de la oferta educativa. Si la transferencia, entendida en este caso como el valor nutricional / la cantidad de desayunos / almuerzos recibidos / la valoración de los desayunos / almuerzos recibidos, es muy pequeña y los niños contribuyen con el ingreso laboral mediante el trabajo infantil, es posible que no se observen cambios en términos de la asistencia a la escuela, en tanto que el valor de la transferencia no compensa el costo de oportunidad del hogar. Aún en el caso que se observe un incremento en la participación escolar, es posible que no se encuentren impactos en términos de logro educativo. Por ejemplo, si el aumento en la participación escolar no tiene como contrapartida ajustes en términos del tamaño de clase, asistencia de los maestros o un incremento en la provisión de *inputs* escolares (libros y materiales), entonces es probable que se observe un deterioro en los indicadores de performance educativa. Por esta razón, muchos de los impactos potenciales del programa son ambiguos.

En esta sección, se discute un marco de referencia que sirva para la selección de las dimensiones a evaluar así como generar hipótesis sobre la dirección de los efectos<sup>8</sup>.

#### 3.1. Impactos en Nutrición, Salud y Desarrollo Cognitivo

Los programas de alimentación escolar pueden contribuir a reducir el hambre de corto plazo, el déficit calórico, las deficiencias en el consumo de micronutrientes claves (como proteínas, hierro, zinc, vitamina A, yodo, entre otros) y el estatus nutricional (aproximado usualmente mediante el uso de indicadores antropométricos). Esto, a su vez, tiene impactos positivos sobre el desarrollo cognitivo y la mejora de los niveles de resistencia a infecciones intestinales y respiratorias, y en general, mejoras en las condiciones de salud, los cuales –bajo ciertas condiciones- pueden traducirse en una mayor participación escolar y logro educativo. Estos efectos son más importantes para el caso de la población escolar en educación inicial, en tanto –como es reconocido por la literatura- intervenciones tempranas que afectan las condiciones salud en la niñez tienen un impacto importante en el desarrollo de capacidades que

---

<sup>8</sup> Ver Bundy *et al.* (2009) y Adelman *et al.* (2008), para una discusión detallada de estos temas.

promueven el bienestar en el curso de vida de los individuos (Conti y Heckman 2012 y Currie 2009)<sup>9</sup>.

Además del impacto de estos programas sobre la población objetivo, existen impactos potenciales sobre las decisiones del hogar. En particular, la asignación de recursos dentro del hogar puede verse afectada por la existencia de esta transferencia en especie, en función al receptor de la misma. Por ejemplo, Duflo (2003) muestra que cuando la mujer es la que recibe la transferencia, existen impactos positivos en los indicadores de salud. En este caso, no es claro desde un punto de vista teórico cuál sería el impacto cuando el receptor es el niño, pero sin duda esto constituye una pregunta de interés desde el punto de vista de la evaluación.

### **3.2. Impactos en Participación Escolar y Logro Educativo**

El programa puede afectar las decisiones de los hogares en relación con la inversión en capital humano. Dado que los hogares comparan los costos y beneficios de dicha inversión, la introducción de un programa de esta naturaleza, al reducir los costos asociados a enviar a los niños a la escuela, debería aumentar la participación escolar. Esto es cierto en tanto la magnitud de las transferencia sea lo suficientemente alta como para compensar el costo de oportunidad de los hogares. Como consecuencia, la tasa de matrícula y la asistencia escolar deberían responder positivamente a la introducción del programa. Además del impacto directo del programa sobre la participación escolar, debe considerarse también la existencia de un impacto indirecto debido a cambios en los indicadores de salud y nutrición, discutidos previamente.

Una mayor participación escolar, bajo condiciones favorables desde el lado de la oferta educativa, debería estar asociado a una mejora en la performance académica. Este efecto es producto, tanto de la asistencia a clase, como consecuencia indirecta de la mejora en el aprendizaje, generada por los cambios en los niveles de atención y memoria (asociados, como se mencionó líneas arriba, con el mejoramiento de los niveles de nutrición y salud).

### **3.3. Teoría de Cambio de *Qali Warma***

De acuerdo con el marco lógico de *Qali Warma*, el objetivo de la intervención consiste en mejorar la ingesta de alimentos, en términos de calidad y cantidad, para los niños que asisten a instituciones educativas públicas de nivel inicial y primaria, de modo que se generen mejores condiciones para su aprendizaje. La teoría de cambio que se desprende de esta intervención se presenta en el Apéndice I. La provisión de un servicio de alimentación de calidad a los niños de educación pre-escolar y primaria mejora la ingesta de alimentos, lo cual a su vez tiene como efecto incrementar la asistencia escolar para finalmente mejorar los niveles de atención y memoria de corto plazo.

Es importante notar que el énfasis del programa radica en mejorar las condiciones para que los niños aprendan, entre las cuales se considera que una mejora en la ingesta cumple un rol central. A diferencia de otros programas en la región, el indicador de impacto del programa no

---

<sup>9</sup> Es importante notar que la literatura sobre este tema hace énfasis en intervenciones en los primeros años de vida, mientras que la intervención motivo de esta nota tendrá como referencia a población entre 3 y 12 años.

es una medida final de aprendizaje sino más bien un proceso intermedio relacionado con las condiciones necesarias para que este aprendizaje ocurra, como es el caso de los procesos cognitivos de atención y memoria de corto plazo. En ese sentido, el objetivo final de mejora en los indicadores de aprendizaje queda implícito en el diseño del programa, quedando como un objetivo de largo plazo al que el programa contribuirá, pero de cuyo logro no se hace explícitamente responsable.

#### **4. EVIDENCIA SOBRE EL IMPACTO DE LOS PROGRAMAS DE ALIMENTACIÓN ESCOLAR**

La evidencia local e internacional en relación con el impacto de los programas de alimentación escolar es mixta. En la mayoría de los casos, se encuentran efectos en términos de incrementos en la participación escolar, aunque ello parece no ser el caso en los países de la región para los que existen estudios al respecto. Algo similar se observa en el caso del rendimiento escolar, en donde en el mejor de los casos se encuentra evidencia parcial en favor del impacto positivo de este tipo de programas. Asimismo, la evidencia de impactos en nutrición, salud y desarrollo cognitivo sigue un patrón similar.

En esta sección, se resume brevemente los hallazgos de la literatura sobre el tema. Adelman et al (2008), Bundy et al (2009) y Jomaa et al (2011) contienen discusiones extensas sobre los resultados de la literatura. Asimismo, el programa ya dispone de revisiones de la literatura elaboradas por Alcazar (2012) y MIDIS (2012). Por esa razón, el énfasis de esta discusión estará esencialmente vinculado con aspectos de naturaleza metodológica. En el Apéndice IV, se complementan las revisiones de la literatura existentes con un conjunto de trabajos que han sido mayoritariamente publicados recientemente sobre el tema y que, a entender del autor, con diseños de investigación más rigurosos que los que predominan en esta literatura.

##### **4.1. Vista Panorámica de los Efectos de los Programas de Alimentación Escolar**

###### ***a. Evidencia sobre impactos en participación escolar***

Un número importante de estudios encuentran impactos sobre la matrícula y la asistencia, aunque la evidencia no es del todo concluyente y es muy focalizada para el caso de países del África y Asia. En el caso de países de América Latina, la evidencia es mucho más débil. Para el caso de Chile, McEwan (2012) no encuentra impactos sobre matrícula y asistencia. Adroque et al (2011) no encuentra efectos del programa de alimentación escolar en Argentina sobre la deserción escolar.

La evidencia internacional es relativamente más favorable, aunque tampoco es concluyente. Por ejemplo, Bottenheim et al (2011) encuentra evidencia débil de incrementos en la matrícula para programas de alimentación escolar en Laos. Belot y James (2011) encuentran una reducción en el número de inasistencias justificadas luego de la introducción de un programa de mejora de la calidad del menú del programa de alimentación escolar en Greenwich, Reino

Unido. Sukontamarn (2011), Meng y Ryan (2010) y Ahmed (2004) encuentran incrementos en la matrícula, aumentos de la asistencia a clases y reducciones en la probabilidad de abandonar la escuela para el Programa de Alimentación Escolar en Bagladesh, aunque con impactos diferenciados de acuerdo al género. Vermeersch y Kremer (2004) encuentran un incremento en la participación escolar de un programa de alimentación escolar asignado aleatoriamente en el norte de Kenya.

La evidencia previa para el caso peruano es escasa. Jacoby et al (1996) encuentran efectos muy modestos del programa de desayuno escolar en Huaraz, utilizando un diseño experimental aleatorio. Cueto y Chinen (2001) encuentran aumentos en la tasa de asistencia y reducciones en la deserción escolar de un programa de desayunos escolares implementado en Ayacucho, Apurímac y Huancavelica. Se desconoce la existencia de estudios recientes que hayan evaluado el impacto del PRONAA.

### ***b. Evidencia sobre impactos en salud y nutrición***

Al igual que en el caso de la asistencia escolar, la evidencia sobre impactos en salud y nutrición es mixta. En el caso de nutrición y salud, la literatura le ha prestado atención a dimensiones como el consumo de calorías, los indicadores antropométricos y el consumo de micronutrientes<sup>10</sup>. En el primer caso, son pocos los estudios que proveen evidencia respecto al consumo de calorías. Afridi (2010) encuentra un incremento en el consumo diario de nutrientes de entre 49-100% de la transferencia de alimentos del programa de alimentos escolares en la India. Ahmed (2004) encuentra que el programa de alimentación escolar de Bangladesh llevó a un incremento de calorías de alrededor del 97% de la transferencia. Jacoby (2002) no encuentra evidencia de reasignaciones de calorías provenientes de un programa de alimentación escolar en Filipinas, aunque si halla cierta evidencia de que esto ocurre en el caso de los hogares más pobres.

Es importante notar que todos los estudios anteriores se basaron en diseños no experimentales. Evidencia experimental es recogida en los estudios de Jacoby et al (1996) para Perú y Murphy et al (2003) para Kenya. En el primer caso, el incremento en el consumo de calorías es la mitad de la transferencia mientras que en el segundo dicho incremento es mayor al tamaño de la transferencia, aunque sólo para el caso de uno de tres tratamientos analizados en el estudio.

En relación con los indicadores de estado nutricional, Simeon (1998) no encuentra diferencias en ganancia de peso para un experimento en Jamaica. En otro experimento en Jamaica, Powell et al (1998) encuentra que un programa de desayunos tuvo un impacto positivo

---

<sup>10</sup> Cada una de estas dimensiones tiene ventajas y desventajas para aproximar el fenómeno de interés (Adelman et al 2008). Debido a que es posible que los alimentos entregados a los escolares sean destinados a otros miembros dentro de la escuela o en el hogar, algunos autores prefieren utilizar el consumo de calorías como indicador nutricional, aunque si el interés reside precisamente en capturar externalidades a nivel de hogar y escuela, entonces este no sería el indicador más adecuado. En el caso que el interés sea en estimar deficiencias nutricionales, algunos autores utilizan el consumo de micronutrientes como proxy para el estatus nutricional puesto que la medición del estado anterior es costosa.

sobre el peso, la talla y el índice de masa corporal de los niños que fueron expuestos al tratamiento. Van Stuijvenberg et al (1999) no encuentra diferencias significativas en el peso entre niños tratados y controles en un experimento en Sudáfrica. Neumann (2007) encuentra incrementos en talla y peso en un experimento con niños de primer grado en Kenya. Aresenault et al (2009) encuentra impactos débiles en los indicadores de talla para edad e índice de masa muscular para edad de un programa de distribución de snacks de la secretaria de educación de Bogotá, Colombia.

Respecto del impacto del consumo de micronutrientes (como hierro, zinc y vitamina A), la mayoría de los estudios evalúan la provisión de alimentos fortificados versus no fortificados. En este caso, la literatura muestra resultados más consistentes sugiriendo la importancia de la fortificación en la mejora de los niveles de micronutrientes. Jacoby et al (1996) encuentra incrementos en los niveles de proteína y consumo de hierro para el caso de un programa de alimentación en Perú. El autor encuentra además una reducción en la incidencia de anemia. Van Stuijvenberg et al (1999) encuentran aumentos en los niveles de vitamina A, hierro y una reducción en la incidencia de anemia en Sudáfrica. Murphy et al (2003) encuentran incrementos en riboflavina, vitamina A, fierro, zinc y calcio en un programa en Kenya.

### ***c. Evidencia sobre impactos en desarrollo cognitivo y logro educativo***

La evidencia sobre el impacto de los programas de alimentación escolar sobre desarrollo cognitivo es mixta. Whaley et al (2003) encuentra mejoras en el test de Raven solo para uno de los grupos de tratamientos en un experimento con niños de escuela primaria en zonas rurales de Kenya. Van Stuijvenberg et al (1999) encuentran sólo cambios en 2 (memoria de corto plazo y atención) de las 8 dimensiones de desarrollo cognitivo bajo análisis en el experimento en Sudáfrica ya mencionado. Vermeersch y Kremer (2004) no encuentran impacto sobre los scores de desarrollo cognitivo en su experimento en el norte de Kenya. Kazianga et al (2012) tampoco encuentra impactos en desarrollo cognitivo en Burkina Faso.

Acerca de los impactos en logro educativo, la evidencia internacional es menos clara. McEwan (2012) no encuentra impactos sobre los test scores de lengua y matemática para el caso del programa de alimentación escolar en Chile y Adroque et al (2011) solamente encuentra impactos para el caso de lengua en Argentina. Belot y James (2011) encuentran impactos positivos sobre los puntajes en las evaluaciones de lengua, ciencia y matemáticas para el caso del programa *Feed Me Better* en Reino Unido. Ahmed (2004) encuentra un incremento en los test scores para el caso de Bangladesh mientras que Vermeersch y Kremer (2004) encuentran un aumento de los test scores sólo para el caso de escuelas en donde los maestros tenían niveles altos de experiencia.

## **4.2. Discusión Metodológica de la Literatura**

Es muy difícil poder extraer lecciones generales a partir de la revisión de la literatura internacional debido a que los programas evaluados difieren en muchos sentidos. Los objetivos, estrategias y duración de la intervención, y el tipo de la transferencia (en términos de calidad y cantidad) son distintos entre los programas discutidos. Los instrumentos de recolección de

información, la fuente de variación explotada en la evaluación, así como el *timing* de la evaluación pueden también explicar diferencias en los resultados. Adicionalmente, las variables de impacto y los indicadores utilizados, la edad de la población bajo estudio y las características previas de las zonas de intervención y de la población objetivo, tienen consecuencias fundamentales para la interpretación de los resultados. En consecuencia, un meta-análisis de la literatura, aunque útil desde la perspectiva del evaluador, debe ser tomado con cuidado por las razones ya expuestas.

Sin embargo, es posible extraer algunas lecciones a tomar en cuenta en el diseño de la evaluación de *Qali Warma*.

- Los impactos en los indicadores de asistencia, deserción y la matrícula escolar son el resultado más importante en la mayoría de los estudios que analizan estas variables, aunque en la mayoría de los casos se trata de países de Asia y África. En el caso de países de América Latina la evidencia es más débil, salvo el trabajo de Cueto y Chinen (2000), quienes encuentran impactos positivos en el caso de un programa de desayunos escolares en la sierra del Perú.
- Los impactos sobre indicadores de salud y nutrición es mixto. Existe evidencia de un impacto en términos de consumo de nutrientes aunque los indicadores nutricionales de talla, peso y masa corporal son por lo general poco sensibles a la intervención. Existe evidencia débil también en relación con los niveles de morbilidad.
- Los impactos sobre indicadores de desarrollo cognitivo son modestos. Cuando estos existen, se tratan de programas que entregan desayunos con componentes fortificados.
- La evidencia indica que los impactos en aprendizaje son muy débiles, cuando estos existen. No hay evidencia sólida en el caso de los países de la región, salvo para el caso del test-score en lengua para el caso argentino. La mayoría de casos en donde se encontraron impactos se trata de programas que operan en países de África o Asia.

## **5. OBJETIVOS Y PREGUNTAS DE EVALUACIÓN**

### **5.1. Definiendo el Contrafactual**

La estimación del impacto del programa requiere una construcción cuidadosa del escenario contrafáctico. Dicho paso es fundamental en tanto que es preciso tener una idea acerca de qué hubiese pasado con la población en situación de tratamiento si es que no hubiese sido tratada, lo cual requiere la construcción de un grupo de control que permita representar precisamente lo anterior. Dado que es imposible observar a un mismo individuo en ambos escenarios (Holland 1986), la validez de una evaluación dependerá críticamente de la construcción de este contrafactual. Si el grupo de control seleccionado para la evaluación no

cumple el requisito anterior, entonces las diferencias observadas entre el grupo tratado y el anterior no podrán ser atribuidas exclusivamente al tratamiento<sup>11</sup>.

En el contexto de la evaluación de *Qali Warma*, la estrategia de intervención del programa y las características institucionales del mismo crean oportunidades y retos para la construcción del escenario contrafactual. En primer lugar, el programa atenderá a la misma población que era atendida previamente por el PRONAA<sup>1213</sup>. En ese sentido, y dado que la intervención que será implementada con *Qali Warma* contiene diferencias importantes en relación con la intervención que existía bajo el PRONAA, el grupo de control implícito serían los atendidos previamente por el PRONAA. En este caso, se estimaría el impacto de *Qali Warma* sobre una población previamente atendida con un programa de alimentación escolar, razón por la cual el impacto estimado es, cualquiera sea su magnitud, una medida de la performance de este programa en comparación al PRONAA. Desde un punto de vista de política, la estimación de tal efecto es de interés para los hacedores de política, en tanto que permite entender si es que los diseños institucionales y las estrategias alternativas de intervención de *Qali Warma* son superiores a las previamente existentes bajo el PRONAA.

Además de la población que era atendida por el PRONAA, *Qali Warma* planea incorporar a aquellas escuelas que no fueron previamente cubiertas por dicho programa. Estas escuelas serán incorporadas progresivamente al nuevo programa en función al nivel de pobreza del distrito en el que se encuentran ubicadas<sup>14</sup>. En este escenario, puede construirse un grupo de control que no haya sido expuesto anteriormente a un programa de alimentación escolar, pudiéndose, por tanto, estimar el impacto del programa frente a la alternativa de no intervención. En este caso, al estar compuesto el grupo de control por escuelas no previamente tratadas, el impacto estimado permitirá recuperar el efecto de *Qali Warma*. Nótese que, en este escenario, se puede recuperar el efecto del programa, mientras que en el caso anterior, el efecto estimado representa mas bien el efecto diferencial de *Qali Warma* sobre PRONAA. Claramente, las estrategias de evaluación esbozadas no son necesariamente contrapuestas en tanto que dependen del tipo de contrafactual que se pretenda construir. Nótese que ambas comparaciones son de interés y pueden ser acomodadas en un diseño de evaluación que incorpore dos grupos de control.

## 5.2. Objetivos y Preguntas de Investigación

El objetivo central de la evaluación es medir los efectos de la disponibilidad de y la participación en el programa de alimentación escolar *Qali Warma* sobre las dimensiones cognitivas de atención y memoria de corto plazo. Asimismo, se consideran un conjunto de

---

<sup>11</sup> Para una discusión conceptual sobre la construcción del contrafactual, consultar el Apéndice III.

<sup>12</sup> Específicamente, por el subcomponente pre-escolar y escolar del Programa Integral de Nutrición.

<sup>13</sup> Alrededor de 39 mil escuelas y más de 2 millones de escolares serían parte de la población de referencia en este caso, de acuerdo con la información manejada por el programa. Es importante notar que la población aquí mencionada incluye escolares de educación secundaria.

<sup>14</sup> De acuerdo a la información provista por el programa, alrededor de 20 mil escuelas serían atendidas progresivamente en los próximos años. Alrededor de 3 mil serán atendidas el primer año de la intervención, junto con las escuelas previamente atendidas por el PIN-PRONAA.

indicadores de participación escolar y logro educativo, así como sobre variables intermedias como salud, nutrición, y hambre de corto plazo.

### **A. Objetivos Centrales**

Los objetivos centrales de la evaluación se asocian directamente a la cadena causal y teoría de cambio del programa. Consideran básicamente tres dimensiones: a) la participación en el programa, entendida como el consumo efectivo de la ración provista al niño beneficiario; b) la participación escolar, aproximada por medio de la matrícula, asistencia y progresión escolar; y c) el indicador central de impacto, vinculado a las dimensiones cognitivas de atención y memoria de corto plazo. En consistencia con la discusión de nuestro marco conceptual, se formulan las siguientes preguntas de evaluación:

#### **a. *Objetivo 1: Participación en Qali Warma***

¿Cuál es el nivel de participación de estudiantes de inicial y primaria en *Qali Warma*? ¿Varían los niveles de participación en el programa en relación con las características del diseño del programa, como el nivel de tratamiento (desayunos más almuerzos versus sólo almuerzos), el tipo de CAE, la receta programada por el programa para la zona la hora de provisión, etc.? ¿Cómo difiere la participación en el programa en relación con las características de la población objetivo, como el sexo, nivel de pobreza, nivel educativo (inicial y primaria), edad, estado nutricional previo, etc.? ¿Difiere la participación en función a la existencia de programas sociales complementarios como Cuna Más, Pensión 65, Juntos, etc.?

#### **b. *Objetivo 2: Asistencia y progreso educativo en primaria***

¿Cómo afecta la disponibilidad y la participación en el programa dimensiones de participación escolar como la matrícula, la asistencia, la progresión y deserción escolar para el caso de la población beneficiaria? ¿Cómo varían estas dimensiones en función a las características del programa, de la población objetivo y la existencia de programas sociales complementarios?

#### **c. *Objetivo 3: Atención y memoria de corto plazo***

¿Cuál es el impacto del programa sobre los niveles de concentración, memoria de corto plazo y la habilidad de desarrollar tareas complejas de la población beneficiaria? ¿Cómo varían estas dimensiones en función a las características del programa, de la población objetivo y la existencia de programas sociales complementarios? En particular, ¿depende la atención en clase del momento en que los alimentos son distribuidos? ¿Difieren los impactos de acuerdo al turno (mañana y tarde) en el que los beneficiarios están matriculados?

#### **d. *Objetivo 4: Patrones de consumo e ingesta dietética***

¿Cómo afecta la disponibilidad del programa sobre el consumo de alimentos de la población objetivo dentro y fuera del hogar? ¿Cómo afecta el programa la composición de la dieta de la población objetivo en relación a componentes claves como proteínas, grasas e hidratos de carbono? ¿Existen externalidades a nivel del hogar en términos de dicha composición; en particular, sustituyen los hogares componentes de baja calidad en la dieta

producto de reasignaciones de gasto hacia alimentos de mayor calidad? ¿Cómo varían estas dimensiones en función a las características del programa, de la población objetivo y la existencia de programas sociales complementarios?

## **B. Objetivos Complementarios**

Además de los objetivos arriba indicados, de los cuales el programa es formalmente responsable, se consideran objetivos de investigación complementarios en función a la evidencia internacional y en línea con la investigación teórica sobre posibles impactos de los programas de alimentación escolar. Las preguntas de evaluación son las siguientes:

### **e. *Objetivo 5: Seguridad alimentaria***

¿Cuál es el impacto del programa sobre los niveles de hambre de corto plazo de la población beneficiaria? ¿Cómo afecta la disponibilidad del programa sobre los niveles de seguridad alimentaria de los hogares de la población beneficiaria? ¿Cómo varían estas dimensiones en función a las características del programa, de la población objetivo y la existencia de programas sociales complementarios?

### **f. *Objetivo 6: Salud infantil y nutrición***

¿Cuál es el efecto del programa sobre el estatus de salud y nutrición de la población beneficiaria (anemia, obesidad, malnutrición, etc.)? ¿Cómo varían estas dimensiones en función a las características del programa, de la población objetivo y la existencia de programas sociales complementarios?

### **g. *Objetivo 7: Logro académico***

¿Cómo afecta la disponibilidad y participación en *Qali Warma* en el logro académico de la población beneficiaria? ¿Experimentan los participantes en el programa mayores ganancias en los test estandarizados que los no participantes? ¿Cómo varían estas dimensiones en función a las características del programa, de la población objetivo y la existencia de programas sociales complementarios?

## **5.3. Matriz de Evaluación**

El Apéndice II presenta la matriz de evaluación para *Qali Warma*. Dicha matriz vincula los objetivos de la evaluación con las preguntas de investigación, los indicadores de impacto, las fuentes de información y el instrumento de medición.

Las variables de impacto centrales en la evaluación son “Asistencia y progreso educativo en primaria”, “Atención y memoria de corto plazo” e “Ingesta calórica”. En el primer caso, los indicadores de impacto cubren dimensiones asociadas a matrícula, asistencia, promoción, repitencia, deserción y retención. Dicha información se obtendrá de un módulo específico sobre información a nivel de escuela lo cual será complementado con información secundaria obtenida del Censo Escolar. En el segundo caso, se utilizará el test Woodcock-Muñoz, específicamente, las pruebas de Inversión de números, Memoria de Palabras, Cancelación de pares, y Análisis –

Síntesis de la Batería III del mencionado test. Los detalles asociados a la medición del indicador de atención y memoria de corto plazo se discuten en la sección 8.1 y en el Apéndice V.

Además de las variables de impacto centrales, se consideran en la matriz las variables de “Rendimiento Educativo”, “Nutrición y Salud” y una dimensión agregada que pretende capturar las potenciales externalidades asociadas con el programa bajo el título de “Impactos en las Escuelas, Hogares y la Comunidad”. El detalle puede estudiarse en la sección 2 de la matriz. Por otro lado, la sección 3 de la matriz analiza los factores que pueden afectar el impacto, lo cual busca aproximar la medición de efectos heterogéneos de la intervención. Entre las dimensiones consideradas, se encuentra las características de la escuela y de los docentes, variables socio-demográficas a nivel del hogar y la existencia de intervenciones complementarias en el distrito. También se consideran las características de los CAE.

Es importante anotar que la matriz de evaluación considera un conjunto amplio de dimensiones que pueden potencialmente ser evaluadas en el marco del programa. Aunque la mayoría de esas dimensiones pueden de ser sujetas a evaluación en el corto y mediano plazo, varias solo son pasibles de evaluación en el corto plazo. Asimismo, la matriz de evaluación considera indicadores que solo pueden ser estimados mediante el uso de datos secundarios. En ese sentido, la matriz considera no solo dimensiones a ser estudiadas mediante recolección de datos primarios sino también dimensiones que requieren el uso de datos administrativos e información secundaria.

#### **5.4. Hipótesis sobre el Impacto de *Qali Warma***

Para formular hipótesis respecto del impacto del programa, es necesario tener una idea precisa de la naturaleza de la intervención. Si, por ejemplo, el programa considera la distribución de alimentos fortificados, entonces es más probable observar impactos en desarrollo cognitivo y aprendizaje. Si esto no es así, entonces es muy improbable que se encuentre lo anterior.

La estructura genérica de los test de hipótesis para la evaluación de impacto del programa tiene la siguiente estructura:

$H_0$  : El programa *Qali Warma* no tiene impacto en la variable de resultado bajo análisis.

$H_A$  : El programa *Qali Warma* tiene impacto en la variable de resultado bajo análisis.

La estructura anterior asume un test de hipótesis de dos colas para testear la hipótesis nula de no efecto del programa. Las variables de resultado y los signos de los efectos esperados se presentan en el Cuadro 1. Dichos efectos esperados se formulan en función a la evidencia internacional, la naturaleza de la intervención y a las características de la población objetivo. Como se discutió líneas arriba, la literatura internacional y local sugiere que los impactos de los programas de alimentación escolar son relativamente modestos. La evidencia local encuentra cierto impacto en el caso del indicador central de la evaluación (atención y memoria de corto plazo), por lo que se espera que el impacto sea positivo, teniendo en cuenta además que la población de referencia de la evaluación estará conformada por niños de educación primaria en distritos en los dos quintiles más pobres. Al funcionar en la práctica como una transferencia en

especie al hogar, se espera también que el programa disminuya su vulnerabilidad y cause reasignaciones de recursos entre sus miembros.

Para el resto de dimensiones, es menos obvio que QW tenga un impacto positivo. Por ejemplo, es poco probable que afecte el aprendizaje medido por medio de los test de matemática y lengua. Similar situación ocurre en los indicadores nutricionales, en la medida que el programa no provee de alimentos fortificados. Dado que la evidencia internacional y local no es concluyente, dichos impactos son inciertos.

**Cuadro 1: Efectos Esperados de *Qali Warma***

Pregunta / Problema	Sub-preguntas	Efecto Esperado
<b>1. Hipótesis sobre Indicadores de Impacto:</b>		
a) Asistencia y progreso educativo en primaria	¿En qué medida QW mantiene a los niños en la escuela y facilita su progreso educativo?	+
b) Atención y memoria	¿En qué medida QW contribuye a la mejora de procesos cognitivos básicos como atención y memoria a corto plazo?	+
c) Rendimiento educativo	¿En qué medida QW promueve el aprendizaje de los niños en la escuela?	?/0
d) Nutrición y salud	¿En qué medida los niños que participan en la QW tienen dietas y condiciones nutricionales mejoradas?	?
	¿En qué medida los niños que participan en la QW han reducido la prevalencia y / o la duración de los episodios de morbilidad?	?/0
e) Impactos en las Escuelas, Hogares y la Comunidad	¿Cómo influye QW dentro del hogar, la alimentación y la distribución del trabajo relacionado?	+
	¿En qué medida QW proporciona un beneficio y sustento al hogar vulnerable?	+
	¿Cómo ha afectado QW la calidad de las IIEE?	?
	¿Cómo ha afectado la participación en QW sobre la comunidad y las IIEE?	?
<b>2. Hipótesis sobre Efectos Heterogéneos</b>		
a) Heterogeneidad de las escuelas (según ubicación geográfica, plana docente, infraestructura, inputs escolares, etc.)	¿Cuál es la influencia de las características de la escuela y la plana docente sobre la atención y memoria, y el rendimiento educativo?	?

b) Características sociales y económicas de los hogares	¿En qué medida otros factores en la escuela y el entorno familiar afectan el éxito en la escuela primaria de los niños beneficiarios del programa QW?	?
c) Diferencias en niveles de vida de los hogares	¿Cuál es el costo de oportunidad de la educación para los hogares beneficiarios de QW? ¿Cuál es el efecto incentivo de la ración de QW sobre los hogares en términos de enviar a un niño a las actividades escolares?	+
d) Otras intervenciones de redes de protección en las zonas seleccionadas	¿En qué medida han influido otras intervenciones (Juntos, Pensión 65, FONCODES, etc.) en el impacto del programa QW? ¿En qué medida otras intervenciones y actividades importantes han influido en el desempeño escolar y el entorno escolar?	+/?

Nota: + : impacto positivo, - : impacto negativo, ? : impacto incierto y 0 : no impacto.

## 6. ESTRATEGIAS DE EVALUACIÓN

Los últimos años han sido testigos de un cambio sustancial en los métodos y prácticas utilizadas para la evaluación de programas. Estos cambios se han visto reflejados en el predominio de los diseños experimentales para resolver el problema de selección (ver, por ejemplo, Banerjee y Duflo 2009). Asimismo, la literatura ha empezado a privilegiar el estudio de mecanismos (o mediadores), no siendo ya suficiente establecer el impacto del programa sino también entender las razones que explican el éxito o fracaso del mismo (ver, por ejemplo, Imai et al 2011), así como un análisis de efectos heterogéneos y las dimensiones distributivas de la intervención (Fink et al 2011, Frandsen et al 2012). Finalmente, se está prestando más atención a la robustez de los resultados de una evaluación, así como al análisis de la sensibilidad de los supuestos del diseño de la misma (ver Rosenbaum 2005 para una revisión de la literatura estadística sobre el tema). Una evaluación rigurosa del Programa *Qali Warma* implicaría la inclusión de estos aspectos en el diseño.

En la sección 5.1, se vio que, si el interés consiste en evaluar el impacto del programa frente a la alternativa de no intervención, entonces el énfasis del diseño debería estar en las zonas que no han recibido tratamiento con anterioridad. Eso tendría implicancias para la construcción de la línea de base en tanto que las observaciones provenientes de distritos no tratados previamente deberían estar sobre-representados en la línea de base. Por otro lado, si el interés es evaluar la performance de *Qali Warma* en relación al PRONAA, entonces el grupo de comparación sería –obviamente– aquellos colegios que fueron previamente expuestos al PRONAA. Como consecuencia de lo anterior, la estrategia de identificación tendrá que ajustarse a la naturaleza del grupo de control, en particular la fuente de variación a explotarse en los datos en cada uno de los casos.

Dada la naturaleza de la intervención, es posible el uso de estrategias diferenciadas en función del grupo de control de referencia. Para el caso de los colegios no intervenidos previamente por el PRONAA, la sugerencia consiste en el uso de una regresión discontinua

aprovechando el uso de los niveles de pobreza distrital como variable de asignación al tratamiento. En este caso, la comparación sería entre estudiantes y/o escuelas ubicadas en distritos alrededor de la discontinuidad inducida por la estrategia de intervención del programa. Para el caso de colegios previamente expuestos al PRONAA, una estrategia de diferencias en diferencias combinada con el método de matching podría ser tomada en cuenta, en cuyo caso la fuente de variación a utilizarse sería la exposición diferenciada al tratamiento de cohortes de edad dentro de un mismo distrito en relación a un grupo de control.

Alternativamente, se sugiere que –en el caso por problemas de implementación práctica una regresión discontinua no sea posible- se utilice un modelo similar de diferencias en diferencias con matching para el caso de aquellos colegios no tratados previamente. En este caso, se explotaría la variación entre escuelas y no entre cohortes como en el caso de colegios previamente tratados por el PRONAA.

En las líneas que siguen, se discutirá en detalle estas estrategias con el objetivo de justificar la elección en términos del contrafactual y la estrategia de identificación. En el primer caso, se opta por la comparación QW versus no intervención, por lo que la muestra de evaluación estará compuesta por estudiantes de educación primaria en escuelas previamente no intervenidas por el PRONAA en distritos pertenecientes a los dos quintiles más pobres. Asimismo, dados los problemas de implementación experimentados por el programa en sus primeros meses de intervención, se opta por descartar el diseño de regresión discontinua en tanto que este tipo de diseño es más sensible a violaciones en la regla de asignación.

Dado que el interés es facilitar la discusión de este documento entre las partes interesadas en el MIDIS y la cooperación internacional, dejaremos la discusión de los aspectos técnicos de estas estrategias para un apéndice (ver Apéndice II).

## **6.1. Evaluando el Impacto en Zonas no atendidas previamente por PRONAA**

En condiciones ideales, la asignación aleatoria del tratamiento permite recuperar el efecto causal de interés. En este caso, si fuese posible asignar aleatoriamente el programa *Qali Warma* a nivel de colegios ubicados en distritos pertenecientes al quintil I (tratados) y al quintil II (controles), una simple diferencia entre los resultados observados para los colegios tratados y no tratados sería suficiente para evaluar el impacto del programa.

### **6.1.1. Alternativa 1: Regresión Discontinua**

Por restricciones políticas e institucionales, así como la naturaleza del programa, no es posible explotar un diseño aleatorio en el contexto de *Qali Warma*. Sin embargo, bajo ciertas condiciones, es posible explotar las reglas de asignación del programa para estimar el efecto del programa para aproximarnos a un entorno experimental. En este escenario, es posible explotar un diseño de regresión discontinua tomando en cuenta la regla de asignación del programa en función del nivel de pobreza distrital. La idea básica de un diseño de regresión discontinua consiste en comparar escuelas ubicadas en distritos que tienen niveles de pobreza ubicados alrededor de la línea de corte seleccionada por el programa para decidir la asignación del

tratamiento. En nuestro caso particular, dicho valor corresponde a una tasa de pobreza distrital de 75.21% para el caso del primer periodo (2013/2014) y 27.01% para el segundo periodo (2015). Escuelas ubicadas en distritos con niveles de pobreza ligeramente más altos que la línea de corte seleccionada serían expuestas al tratamiento mientras aquellas ubicadas en distritos con niveles de pobreza ligeramente más bajos serían asignados al grupo de control.

### **a. Literatura Previa**

El uso de regresiones discontinuas es muy popular en casos en los que la asignación al programa depende de algún indicador de pobreza. La alternativa más común es el uso de índices de pobreza compuestos como los *scores* de pobreza o índices de vulnerabilidad contruidos con técnicas como componentes principales, pero también se ha utilizado la tasa de pobreza como variable de asignación, como –por ejemplo- en Van der Klauw (2008) y Ludwig y Miller (2007)<sup>15</sup>. En el contexto específico de programas de alimentación escolar, sólo estamos al tanto de los trabajos de McEwan (2012) y Schanzenbach (2009). En el primer caso, McEwan evalúa el impacto del programa nacional de alimentación escolar de Chile sobre resultados educativos mediante una regresión discontinua utilizando como variable de asignación un índice de vulnerabilidad construido a nivel de escuela. Por su parte, Schanzenbach evalúa el programa nacional de almuerzos escolares en Estados Unidos sobre la obesidad, utilizando como variable de asignación (en este caso, a nivel individual) el nivel de ingreso familiar del estudiante en relación con la línea de pobreza estatal.

Es importante notar que, a diferencia de los casos anteriores en donde la asignación al tratamiento se realiza a nivel de alumno y/o escuela, en el contexto de *Qali Warma* dicha asignación se realizará a nivel distrital. A pesar de ello, esta diferencia no tiene un impacto sustancial en términos del diseño, en tanto se pueden agregar controles a nivel de escuela en el diseño, de modo tal que se tome en cuenta heterogeneidades entre escuelas en un mismo distrito<sup>16</sup>.

### **b. Ventajas y desventajas**

Las ventajas del uso de una regresión discontinua son significativas en comparación con otros métodos de evaluación alternativos. En primer lugar, existe un consenso en la literatura en relación al alto grado de validez interna asociado a esta técnica. Buddelmeyer y Skoufias (2003), Black, Galdo y Smith (2007); y Green et al (2009) muestran que esta técnica provee estimados que se aproximan muy bien a los que se obtendrían en un diseño experimental con asignación

---

<sup>15</sup> Van der Klauw estima el impacto de un programa (Title 1) sobre los test scores de los estudiantes provenientes de escuelas primarias en Nueva York. El programa es asignado en función al nivel de pobreza de los estudiantes de las escuelas. Ludwig y Miller estudian el impacto del programa Head Start sobre la mortalidad infantil y el progreso educativo. El programa es asignado en función al nivel de pobreza del distrito (county).

<sup>16</sup> No tenemos razones para pensar que exista un sesgo importante producto de diferencias entre escuelas dentro de un mismo distrito. Dado que se trata de distritos extremadamente pobres, es razonable suponer que las características de las escuelas deben ser bastante similares por tratarse de distritos con poca población y pocas escuelas. Este problema puede ser más significativo en distritos más urbanizados.

aleatoria. En segundo lugar, autores como Lee (2008), y Lee y Lemieux (2010), una regresión discontinua es simplemente un caso particular de un diseño experimental, por lo cual este diseño permite aprovechar las ventajas de un diseño experimental sin incurrir en los costos éticos y políticos de las evaluaciones experimentales aleatorias. Finalmente, la transparencia del diseño y su alto grado de validez interna, hacen que los resultados de una evaluación mediante esta técnica sean altamente creíbles, y por tanto un instrumento poderoso para la toma de decisiones en relación con el destino del programa para los formuladores de política y la cooperación internacional.

La técnica de regresión discontinua también tiene ciertas desventajas que es preciso tomar en cuenta. En primer lugar, los resultados de la evaluación son sólo válidos a nivel local, particularmente para el caso de aquellos distritos ubicados alrededor de la discontinuidad. Los resultados son sensibles a sesgo de selección mientras más lejos se encuentren los distritos bajo análisis de la discontinuidad que define la asignación al programa. En segundo lugar, los diseños de regresión discontinua son sensibles a problemas de potencia, por lo que un número importante de observaciones alrededor de la discontinuidad son necesarias para poder estimar el efecto de interés. Finalmente, la estimación del efecto causal es sensible a la especificación econométrica utilizada. Por fortuna, salvo el primer caso, estas limitaciones pueden ser controladas en el diseño de evaluación.

### **6.1.2. Alternativa 2: Diferencias en Diferencias con Matching**

Un problema con la alternativa anterior es que puede ser sensible a problemas de implementación práctica si es que los oficiales del programa no siguen la regla de asignación en función al nivel de pobreza del distrito. Existe evidencia de problemas similares en el caso de programas como Juntos (Perova y Vakis 2009), en donde el programa no siguió el patrón de implementación originalmente establecido por problemas de naturaleza logística. Corresponderá al equipo de evaluación verificar y evaluar la magnitud de este problema en la práctica. Si el problema no es grave, todavía sería posible utilizar una regresión discontinua, aunque en este caso tendríamos un diseño tipo fuzzy, que sería equivalente al caso de un experimento con cumplimiento imperfecto (*imperfect compliance*). Sin embargo, si este problema es muy grave, debería considerarse un diseño alternativo.

En ese contexto, se propone el uso de una estrategia mixta de diferencias en diferencias con matching. La idea básica de este diseño consiste en construir un contrafactual para los distritos intervenidos bajo *Qali Warma* a partir de distritos no tratados previamente por PRONAA, cuyas características sean similares a aquellos a ser expuestos al tratamiento. Bajo un conjunto de supuestos de identificación (ver Apéndice III), esta técnica permite estimar el impacto del programa.

#### **a. Literatura Previa**

Existen varios ejemplos recientes de la aplicación del método en cuestión para el caso de programas de alimentación escolar. Adroque et al (2012) aplica una variante de esta técnica para

el caso de los programas de alimentación escolar en Argentina. Belot y James (2011) estudian el impacto de una campaña de mejora de la calidad de los alimentos provistos en la escuela en un barrio de Londres (Greenwich). Schanzenbach (2009) analiza el impacto del Programa Nacional de Almuerzos Escolares en Estados Unidos sobre la obesidad. Meng y Ryan (2010) estudian el programa "Alimento por Educación" de Bangladesh y su impacto sobre la participación escolar. Buttenheim et al (2011) analizan el impacto de programas de alimentación escolar del Programa Mundial de Alimentos (PMA) en Laos sobre la matrícula escolar e indicadores nutricionales.

En todos los casos previamente discutidos, los autores estuvieron interesados en el efecto promedio del tratamiento para los tratados, sin discutir la posibilidad de efectos distributivos importantes como consecuencia del tratamiento. Esto es importante, porque los resultados de la mayoría de estos estudios no permiten elaborar una imagen coherente de la dirección de los efectos. El uso de una estrategia que explore los impactos distributivos podría ayudar a dar una interpretación coherente a los resultados obtenidos por la literatura internacional en el tema.

Asimismo, todos los estudios mencionados se han basado en el uso de una estrategia de matching basada en el *propensity score* sin explorar la robustez de los resultados a cambios en el método de matching utilizado. Existe una amplia literatura respecto de la limitación del método de propensity scores para recuperar efectos causales (Smith y Todd 2005, Arceneaux et al 2006) y sobre las ventajas de métodos alternativos de matching como el de Abadie e Imbens (2006, 2011) en esa dirección (McKenzie et al 2010). Aunque Heckman et al (1997) encuentran que - combinado con el método de diferencias en diferencias- este método no es rechazado por el test de supuestos de identificación para la aplicación que analizan, no existen muchas aplicaciones de métodos alternativos de matching combinados con diferencias en diferencias y mucho menos una evaluación acerca de las bondades de las diversas técnicas de matching en este contexto. Por esta razón, es importante considerar métodos alternativos en el marco de la evaluación del programa.

### **b. Ventajas y desventajas**

El método de diferencias en diferencias es una estrategia de evaluación cuasi-experimental muy popular en la literatura, en razón a una serie de virtudes que se describen a continuación. En primer lugar, permite controlar por factores no observables que son fijos en el tiempo que podrían estar correlacionadas con el tratamiento. En segundo lugar, una estrategia de diferencias en diferencias, combinada con matching, permite controlar por parte de aquellos factores observables que pueden variar en el tiempo, reduciendo por tanto la heterogeneidad inicial entre tratados y controles. Finalmente, la aplicación de esta técnica en el contexto de la evaluación de *Qali Warma* sería menos sensible a problemas de implementación práctica del programa por parte del personal de campo del mismo, en tanto que las fuentes de problemas en la implementación sean identificables y por tanto incorporables al diseño de evaluación.

En su versión estándar (sin ajustes para la heterogeneidad inicial), esta técnica tiene también algunas limitaciones en tanto se basa en algunos supuestos que son relativamente

restrictivos. En primer lugar, se asume que no existen shocks temporales no observados y específicos a nivel de distrito. Asimismo, se asume que tanto tratados como controles están expuestos a los mismos shocks macro (lo que se conoce usualmente como el supuesto de tendencias comunes). Una forma de relajar los supuestos anteriores sería explotar información previa a la implementación del programa, aunque esto es sería muy complicado en el contexto de *Qali Warma*. Como ya se ha mencionado, combinado con matching, este método permite controlar por heterogeneidades en la línea de base.

## **6.2. Evaluando el Impacto en Zonas previamente atendidas por PRONAA**

En el caso de las zonas previamente atendidas por PRONAA, no es posible construir un contrafactual puesto que todas las escuelas serán atendidas en el primer año. En ese escenario, se sugiere explotar variación dentro de las escuelas producto de las reglas de asignación del programa. Dado que el programa está dirigido a estudiantes de nivel inicial y primario, es posible explotar diferencias en términos de exposición al programa para estimar el efecto causal de interés. En ese escenario, es posible utilizar un diseño de diferencias en diferencias como el ya discutido en la sección previa. Este diseño sería similar al análisis de Duflo (1999) que estudia el impacto de un programa de construcción de escuelas sobre los retornos a la educación en Indonesia.

Un problema con el diseño anterior es la potencial existencia de externalidades a nivel de escuela. Kazianga et al (2012) encuentra impactos positivos sobre el rendimiento escolar para el caso de los niños de un programa de entrega de alimentos (condicionada a un 90% de asistencia) orientado a niñas en edad escolar. Kremer et al (2009) encuentran que un programa de becas para niñas tuvo impactos positivos también sobre niños. En ese contexto, es importante tener en cuenta estas consideraciones al momento de evaluar al programa.

## **6.3. Elección Metodológica**

Como se adelantara en la introducción a esta sección, se opta por un análisis del impacto de QW versus la situación de no tratamiento, mediante el uso de un modelo de diferencias en diferencias. La justificación se basa en la robustez del método ante problemas de implementación práctica.

# **7. ANÁLISIS DE IMPACTO: ELEMENTOS BÁSICOS Y EXTENSIONES**

En esta sección, se delinear algunos criterios para el análisis de los datos. Esta discusión es importante en tanto tiene implicancias en cuanto al diseño de investigación y la medición de las variables a ser consideradas en el análisis.

## **7.1. Tipos de Análisis**

Se consideran tres tipos de análisis en el contexto de la evaluación del programa.

### ***a. Análisis descriptivo***

El objetivo de esta sección consiste en establecer el contexto básico de la evaluación así como evaluar la consistencia del diseño. En función al diseño de evaluación, la idea básica consiste en hacer comparaciones entre:

- Estudiantes provenientes de distritos con colegios sin exposición previa a programas de alimentación escolar versus distritos expuestos a *Qali Warma*.
- Participantes versus no-participantes del programa de alimentación escolar en los distritos tratados.

En el primer caso, se trata de evaluar el balance de las características observables entre distritos tratados y no tratados. En el caso del diseño de diferencias en diferencias con matching, ello consiste en evaluar la condición de soporte común (Imbens 2004).

En el segundo caso, se necesita modelar el proceso de selección en el tratamiento. Los estudiantes dentro de las escuelas tratadas se seleccionan al tratamiento en función a una serie de dimensiones observables y no observables para el evaluador. Dado que aquellos que por lo general deciden consumir los alimentos ofrecidos por el programa difieren de aquellos que no los consumen, es de esperarse que la participación en el programa sea sensible a sesgo de selección.

#### ***b. Estimación del Efecto de la Disponibilidad de Qali Warma***

La lógica de este procedimiento consiste en evaluar el impacto de la intervención para todos aquellos que *podrían* consumir los alimentos provistos por el programa. Esto, naturalmente, incluye a aquellos que consumen los alimentos como aquellos que no, lo cual tiene implicancias importantes para la interpretación de los resultados. Mientras que es posible que el programa tenga impactos importantes sobre aquellos que efectivamente consumen los alimentos, estos efectos podrían verse diluidos en un análisis de esta naturaleza si es que la proporción de niños participantes es muy baja dentro de una escuela. La ventaja de este procedimiento es que aprovecha el diseño cuasi-experimental (particularmente, en el caso del diseño de regresión discontinua) y provee estimados relevantes desde el punto de vista de política.

Además del análisis sobre toda la muestra de estudiantes provenientes de distritos QW, se debe considerar el estudio de sub-grupos claves. Ello consiste en analizar cómo el impacto del programa difiere en función a características del mismo, tales como el nivel de tratamiento (desayunos más almuerzos versus sólo almuerzos), la receta programada para la zona, la hora de provisión, etc. Incluye, además, el análisis de diferencias en relación con las características de la población objetivo como el sexo, nivel de pobreza, nivel educativo (inicial y primaria), edad, estado nutricional previo, etc. Finalmente, debe considerarse también diferencias en relación con la existencia de programas sociales complementarios como Cuna Más, Pensión 65, Juntos, etc.

#### ***c. Estimación del Efecto de la Participación en Qali Warma***

En este caso, la estimación de impacto toma en cuenta la participación efectiva de los estudiantes en el mismo. Dado que dicha participación no es aleatoria, el diseño debe tomar en

cuenta estrategias que permitan modelar el proceso de selección de los estudiantes en el tratamiento.

Al igual que en el caso anterior, se debe considerar la estimación del efecto de interés para los mismos sub-grupos claves indicados.

## 7.2. Unidad de Análisis

La evaluación tomará en cuenta al estudiante como unidad de análisis. De esta manera, la muestra estará compuesta por estudiantes de educación primaria en distritos cubiertos (o a ser cubiertos) por Qali Warma que provengan de colegios no previamente expuestos a programas de alimentación escolar. De estos niños, se obtendrá información a nivel de hogar y escuela.

## 7.3. Modelos por estimar

La especificación básica para el caso del modelo de diferencias en diferencias es la siguiente:

$$(1) \quad y_{ijt} = \alpha_j + \lambda_t + \beta(Q_{jt} \cdot P_t) + X'_{ijt} \delta + \varepsilon_{ijt};$$

donde  $y_{ijt}$  es el resultado de interés o variable de impacto (asistencia, matrícula, estado nutricional, etc.) para el estudiante  $i$  que estudia en el distrito  $j$  en el periodo  $t$ .  $\alpha_j$  y  $\lambda_t$  son respectivamente efectos fijos de distrito y periodo.  $Q_{jt} \cdot P_t$  es una variable *dummy* de interacción para observaciones tratadas después de la implementación del programa.  $X'_{ijt} \delta$  incluye características del individuo y del distrito, así como el término de error  $\varepsilon_{ijt}$ . El parámetro de interés es  $\beta$ , el cual recupera el efecto causal de interés.

Esta especificación es una generalización del método estándar de diferencias en diferencias de dos periodos y dos unidades (véase, por ejemplo, Bertrand, Duflo y Mullainathan 2004 y Hansen 2007). Los efectos fijos de tiempo dan cuenta de cambios en el tiempo de las medidas de impacto del programa. Los efectos fijos a nivel de distrito dan cuenta de características invariantes en el tiempo a nivel de escuelas y la interacción  $Q_{jt} \cdot P_t$  da cuenta de cambios en las variables de impacto en los distritos tratados después de la implementación del programa. La identificación en este escenario requiere de controlar por shocks sistemáticos en las medidas de impacto del programa de los distritos expuestos al tratamiento que estén potencialmente correlacionados con, pero no consecuencia de la implementación del programa<sup>17</sup>.

## 7.4. Efectos Heterogéneos

Para estudiar la existencia de efectos heterogéneos, se espera que la empresa seleccionada para llevar a cabo la Evaluación del PNAE Qali Warma realice un análisis empírico basado en el

---

<sup>17</sup> Formalmente, esto se conoce como el supuesto de tendencias común. En términos de contrafácticos, esto implica una estructura aditiva de los resultados potenciales de los distritos no tratados (sin tener en cuenta las covariables) de la siguiente manera:  $E(y_{ijt} / j, t) = \alpha_j + \lambda_t$ . Para una discusión, ver Angrist y Pischke (2009), chapter 6.

uso de interacciones entre la variable de tratamiento y variables que incorporen elementos relevantes como el sexo, nivel de pobreza, edad, estado nutricional previo y la existencia de programas sociales complementarios. La especificación básica sería la siguiente:

$$(2) \quad y_{ijt} = \alpha_j + \lambda_t + W_{ij} + \beta(Q_{jt} \cdot P_t \cdot W_{ij}) + \varphi_1(Q_{jt} \cdot W_{ij}) + \varphi_2(P_t \cdot W_{ij}) + \varphi_3(Q_{jt} \cdot P_t) + X'_{ijt} \delta + \varepsilon_{ijt};$$

en donde  $W_{ij}$  es una característica definida a nivel de individuo o distrito sobre la cual se evaluara la heterogeneidad del impacto.

## 7.5. Tratamiento de los Errores

Como ha sido señalado por diversos autores (Moulton 1986), hacer inferencia sin tener en cuenta la existencia de dependencia dentro de clusters puede producir una grave subestimación de los errores estándar. Esto es lo que Angrist y Pischke (2009) llaman el "problema de Moulton". Además, y especialmente relevante para la estimación de diferencias en diferencias, existe un problema potencial de correlación serial, como fue puesto de manifiesto por Bertrand, Duflo y Mullainathan (2004)<sup>18</sup>. Para hacer frente a ambos temas, en este documento sugerimos la corrección de los errores estándar para el caso del diseño de diferencias en diferencias, mediante la generalización de la matriz de covarianzas robustas de White (1980) desarrollada por Liang y Zeger (1986). Esta solución controla por cluster y heteroscedasticidad, y es válida siempre y cuando un alto número de grupos esté disponible, como es el caso en el contexto de la evaluación de *Qali Warma*<sup>19</sup>.

## 7.6. Análisis de Robustez y Sensibilidad

Además de las estimaciones de los modelos básicos, la evaluación deberá incorporar un conjunto de tests de robustez y sensibilidad a los supuestos de identificación.

Una limitación común de los métodos de matching es que descansan en el supuesto de selección sobre observables. Mientras que su combinación con diferencias en diferencias permite controlar por aquellos factores no observables que no varían en el tiempo, es probable que existan factores no observables variantes en el tiempo que podrían estar correlacionados con el tratamiento. En ese escenario, la evaluación puede considerar el uso de técnicas de sensibilidad para evaluar cómo se comporta la estrategia de matching utilizada ante el relajamiento del supuesto de selección sobre observables. DiPetre y Gangl (2004), Altonji et al (2005), Ichino et

<sup>18</sup> Según estos autores, esto se debe a las siguientes razones: a) por lo general las estimaciones se basan en series de tiempo largas, b) la variable dependiente está generalmente serialmente correlacionada en una forma positiva, y c) la variable de tratamiento cambia muy poco dentro de la unidad de tratamiento a través del tiempo.

<sup>19</sup> Para una discusión para el caso de un pequeño número de clusters, véase Angrist y Pischke (2009). Cameron, Gelbach y Miller (2007) proponen soluciones basadas en el uso de la técnica del bootstrap. En particular, el wild cluster bootstrap parece funcionar bien en un conjunto de simulaciones estudiadas por estos autores.

al (2008), Imbens (2003) y Matsa (2012) son algunas de las estrategias recientes sugeridas por la literatura.

### **7.7. Extensiones**

En la discusión anterior, el énfasis ha sido puesto en la estimación de los efectos promedio del tratamiento (o en un punto en particular de la distribución como es el caso del diseño de regresión discontinua). Dado que la evidencia internacional no es concluyente en relación con los probables impactos del programa, es preciso tener una imagen completa no solo del efecto promedio del programa sino también una evaluación de sus dimensiones distributivas, especialmente en un contexto en donde varios estudios sobre la temática encuentran efectos heterogéneos. Por esa razón, el análisis de evaluación deberá de ser complementado con un estudio del impacto distributivo del programa.

En el contexto de la evaluación de *Qali Warma*, se recomienda utilizar técnicas basadas en regresiones por cuantiles (Koenker 2005) para complementar el análisis de efectos promedios. Específicamente para el caso del diseño de diferencias en diferencias, se recomienda utilizar el estimador de “cambios en cambios” desarrollado por Athey e Imbens (2006), el cual se discute brevemente en el Apéndice III.

## **8. ASPECTOS MUESTRALES Y DE MEDICIÓN**

### **8.1. Medición de la Variable Dependiente**

La variable de impacto de interés en el contexto de la evaluación de impacto de *Qali Warma* es una medida que aproxima los procesos cognitivos básicos de atención y memoria. Dicha medición requiere el uso de pruebas especializadas algunas de las cuales han sido previamente utilizadas en la literatura de evaluación de impacto de programas de alimentación escolar. En el Apéndice V se presenta una revisión de los estudios relevantes en la literatura mencionada.

En el contexto de *Qali Warma*, se utilizarán pruebas especializadas de atención, memoria de corto plazo y razonamiento por lo que es preciso seleccionar entre los instrumentos disponibles algún test que contenga los sub-tests necesarios para medir dichas habilidades. En base a la revisión de Guerrero y León (2013), se sugiere que los sub-tests sean equivalentes a la prueba de Cancelación, Retención de dígitos y Matrices Progresivas de Raven Escala Coloreada. La Batería III del Woodcock Muñoz precisamente contiene pruebas equivalentes a estos sub-tests. La ventaja de esta aproximación se basa en que estos tres sub-test permitirán cubrir todos los procesos cognitivos descritos. Asimismo, los tres tipos de sub-tests son comúnmente usados en la literatura sobre evaluación de impacto de programas de alimentación escolar para medir procesos cognitivos y han sido usadas previamente en el Perú en evaluaciones similares. En ese sentido, se buscará evitar comprometer el diseño de evaluación trabajando en pruebas que ya han mostrado su utilidad en otros contextos. Asimismo, esta aproximación permitirá comparar

los resultados que se obtengan en el caso de *Qali Warma* con la literatura internacional sobre alimentación escolar.

## 8.2. Temas de Investigación

Los temas de investigación cubiertos en los instrumentos de recolección de información se presentan en el Apéndice VI.

## 8.3. Análisis de Poder y Tamaño de Muestra<sup>20</sup>

En relación al análisis de poder para la evaluación de impacto de *Qali Warma*, se ha determinado que el tamaño de muestra necesario para estimar un efecto en caso que este existiera asciende a 3,464 observaciones. Dicha estimación toma como punto de partida un diseño de evaluación basado en un modelo de diferencias en diferencias a nivel de cluster en donde el nivel de intervención es el distrito. En particular, se utiliza la propuesta metodológica para el cálculo de muestra recientemente desarrollada por Teerenstra et al. (2012). Los detalles estadísticos del procedimiento de cálculo de muestra se detallan en el Apéndice VII y se discuten en extenso en Maldonado (2013a y 2013b). La fórmula básica es la siguiente:

$$(3) J = (t_{(1-k)} + t_{\alpha})^2 * \frac{1}{P(1-P)} \frac{\sigma^2}{n * ES^2} * [1 + (n-1)\rho] * [2(1-r)],$$

en donde ES es el tamaño del efecto esperado por la intervención medido en la unidad de referencia correspondiente. n es el número de observaciones por clúster,  $\rho$  es el coeficiente de correlación a nivel de cluster y P es la proporción de tratados en la muestra.  $\sigma^2$  es la varianza de la variable de resultado (atención y memoria de corto plazo) y J es el número de clústeres. El parámetro r cumple un rol importante en el diseño y es definido del modo siguiente:

$$(4) r = \frac{n\rho}{1+(n-1)\rho} \rho_c + \frac{1-\rho}{1+(n-1)\rho} \rho_s,$$

en donde  $\rho_c$  y  $\rho_s$  son respectivamente los indicadores de autocorrelación a nivel de cluster y a nivel de individuo.

El indicador de referencia para la evaluación de impacto es la medida de atención y memoria. Dado que no se disponía de información primaria y secundaria relevante para los indicadores de atención y memoria de corto plazo, los cálculos se basaron en información referencial de variables de resultado similares como es el caso de rendimiento escolar. Los parámetros seleccionados para el ejercicio de cálculo de muestra son los siguientes:

- Efecto tamaño: 0.20 desviaciones estándar tanto para el indicador de atención como para el de memoria de corto plazo
- Nivel de potencia: 90%
- Nivel correlación a nivel de clúster (distrito): 0.17

<sup>20</sup> Esta sección se basa en Maldonado (2013a, 2013b).

Los parámetros anteriores se basan en información colectada a partir de la revisión de la literatura y en estimaciones propias en base la Evaluación Censal de Estudiantes 2011. Para el caso del efecto tamaño, se asume un efecto modesto en relación a los estándares usuales en la literatura.

Como sugieren autores como Cohen (1988), se considera un efecto de 0.2 desviaciones estándar como un efecto pequeño. Este efecto es consistente con los estimados de un meta-análisis de la literatura sobre el impacto de los programas de alimentación escolar sobre atención y memoria encargado por el MIDIS (Guerrero y León 2013). Por otro lado, se utiliza un nivel de potencia del 90%, en línea con la literatura educativa internacional (Schochet 2005). Este parámetro es conservador y mayor al estándar usual de 80% que se utiliza en la literatura económica. Finalmente, el nivel de correlación intracluster (CCI) fue estimado a partir de la información de la ECE 2011 a nivel distrital tomando como referencia la base de 334,557 estudiantes indicada líneas arriba.

Además de los parámetros indicados, es necesario computar el coeficiente de correlación  $r$ , el mismo que depende de  $\rho_c$  y  $\rho_s$  (los indicadores de autocorrelación a nivel de cluster y a nivel de individuo respectivamente). Para el análisis que sigue, se asumen los siguientes valores para dichos parámetros:

- Coeficiente de autocorrelación a nivel de clúster: 0.30
- Coeficiente de autocorrelación a nivel de individuo: 0.45

El cálculo de los coeficientes de autocorrelación requiere de datos longitudinales a nivel de estudiantes y distritos para el caso de los indicadores de atención y memoria de corto plazo. Como ya se mencionó, dichos datos no están disponibles. Una alternativa consistiría en el uso del Estudio Longitudinal realizado por el Ministerio de Educación para el caso del indicador de rendimiento. Este estudio solo considera estudiantes de la región Lima. Sin embargo, dado que no se dispone de una fundamentación adecuada acerca de la validez de utilizar el rendimiento educativo como un proxy de procesos cognitivos como la atención y la memoria de corto plazo, se opta en este documento por utilizar parámetros más conservadores de los que se obtienen de esta encuesta (alrededor de 0.7 en términos de autocorrelación individual). Este valor es bastante alto en relación a los estándares internacionales<sup>21</sup>.

Finalmente, el cálculo del tamaño de muestra para un diseño a nivel de cluster requiere fijar un parámetro adicional: el número de observaciones por cluster<sup>22</sup>. Como referencia, utilizamos el siguiente valor:

- Número de observaciones por cluster: 10.

---

<sup>21</sup> De acuerdo con McKenzie (2102), el valor de este parámetro para países en desarrollo como la India es de alrededor de 0.5.

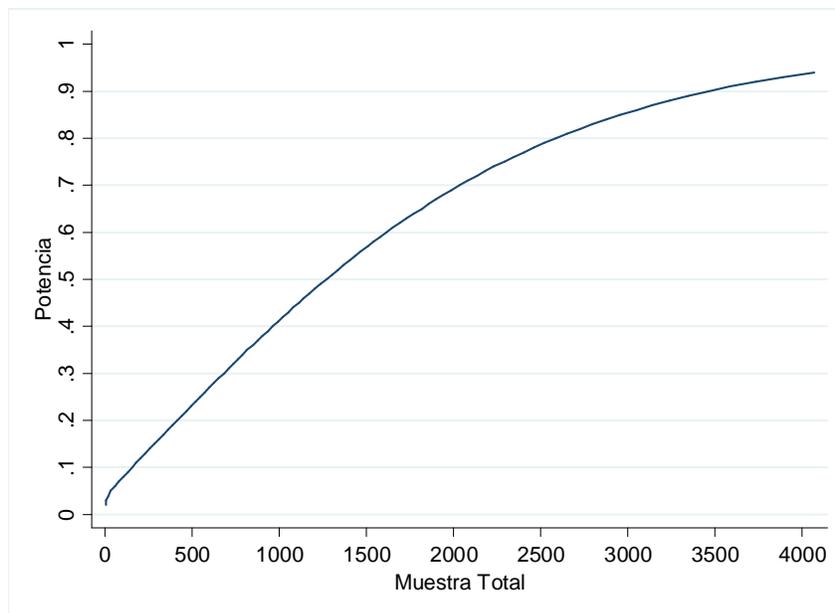
<sup>22</sup> En general, el investigador tiene dos opciones: fijar el número de observaciones a nivel de cluster y estimar el número de clusters o, alternativamente, fijar el número de clusters y estimar el número de observaciones por cluster. En este trabajo, se opta por la primera vía.

Este valor se basa en el número promedio de hogares por cluster que se utilizan en las encuestas de hogares levantadas bajo el programa Encuestas de Medición de Niveles de Vida del Banco Mundial. Grosh y Muñoz (1996) reportan tamaños de cluster entre 10 y 16 hogares para el caso peruano.

Como es conocido, el incremento de observaciones por cluster tiene menos impacto en el cálculo de poder que el incremento del número de clusters. En ese sentido, es más eficiente aumentar el número de distritos que el número de estudiantes en cada distrito para alcanzar un nivel de poder determinado, minimizando el tamaño de muestra total. Sin embargo, un número alto de distritos tienen implicancias sobre el costo total, especialmente si se tratan de distritos alejados y de difícil acceso como es el caso de *Qali Warma*, en tanto incrementan el costo por unidad. Así, mientras que incluir más distritos reduce la muestra total, puede venir asociado a un mayor costo por unidad. Por otro lado, si se reduce el número de distritos será necesario aumentar el número de observaciones por distrito para alcanzar el nivel de poder inicial lo cual llevará a un incremento de la muestra total, pero el hecho de que las observaciones provengan de menos distritos puede hacer más eficiente el proceso de recolección de datos y reducir el costo unitario.

A partir de los supuestos sobre los parámetros para el cálculo del tamaño de muestra, se realiza la estimación del modelo básico. Los resultados se resumen en la Figura 1.

**Figura 1: Potencia como Función del Tamaño de Muestra**



La Figura 1 presenta la curva de potencia para los parámetros anteriores. A un nivel de potencia de 90%, se estima que se necesitarían **3,464 observaciones** para tener la habilidad de estimar un efecto no menor a 0.2 desviaciones estándar. A partir de dicha figura, es posible determinar cómo se afectan nuestras conclusiones si es que alteramos los niveles de potencia.

Así por ejemplo, si modificamos el nivel de potencia a 80%, se requerirían ahora solo 2,588 observaciones.

Una característica del programa consiste en que el tratamiento difiere en función a la zona de intervención. En las zonas urbanas, el programa utiliza la opción de entregas de raciones mientras que en las rurales se distribuyen productos alimenticios a los Comités de Alimentación Escolar (CAE) de modo que estos se encarguen de la preparación de los alimentos. Debido a problemas de implementación en el terreno, el programa consideró una tercera alternativa denominada “canasta de productos”, en donde un sub-conjunto de los productos disponibles bajo la segunda modalidad es distribuido en aquellas zonas en donde los procesos de licitación del programa no funcionaron. Dado que el diseño original de la evaluación solo presta atención a los dos quintiles más pobres, en la práctica implicaban la comparación entre el grupo de control y un grupo de tratamiento conformado esencialmente por distritos en donde la intervención es vía distribución de productos. En ese escenario, la introducción de la “canasta de productos” puede tener implicancias en el diseño original de la evaluación. En particular, el análisis de poder de la evaluación puede verse comprometido puesto que el diseño ahora considera en la práctica un tratamiento múltiple en donde existen dos niveles de intervención además del grupo de control original.

El problema de comparaciones múltiples ocurre cuando se considera un conjunto de hipótesis respecto a significancia estadística de una intervención de manera simultánea. En el contexto de *Qali Warma*, la comparación no sería solamente entre distritos tratados y controles, sino también entre distritos tratados que reciben productos y aquellos que reciben la canasta básica. En este escenario, el número de comparaciones posibles serían 3 en vez del caso simple en donde solo una comparación (tratados versus controles) es posible. Por esta razón, el test de hipótesis que subyace a la estimación original del tamaño de muestra podría ser incorrecto pues asume una probabilidad de cometer un Error Tipo I de 5% cuando en realidad, si el conjunto de test de hipótesis posibles es considerado conjuntamente, el Error Tipo I es sustancialmente mayor. Dado que controlar por este problema implica reducir el Error Tipo I para cada test individual, esto automáticamente implica un incremento del Error Tipo II, lo cual deriva finalmente en una reducción del nivel de poder en relación al considerado originalmente en el caso de *Qali Warma*<sup>23</sup>.

La solución más común en la literatura consiste en utilizar la corrección sugerida por Bonferroni. La fórmula correspondiente se detalla en el Apéndice VI. El parámetro adicional es el siguiente:

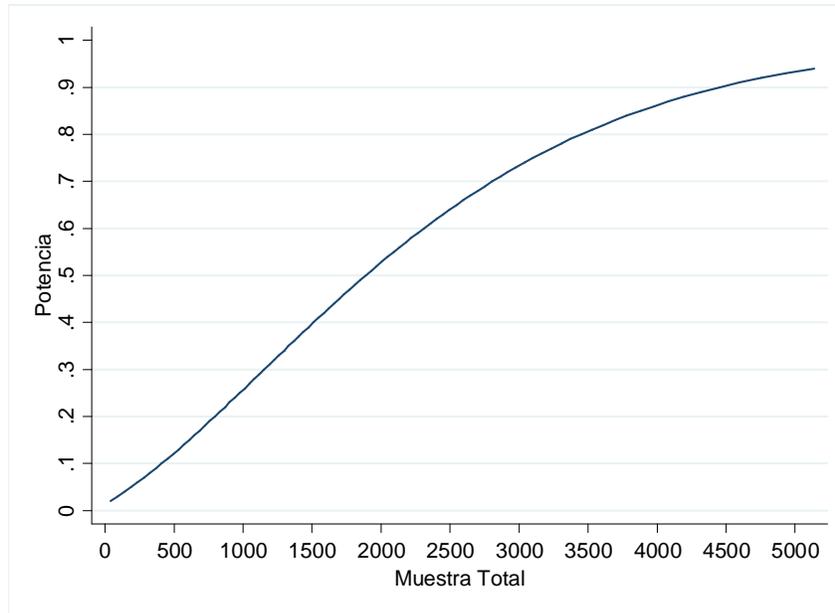
- Número de test de hipótesis considerados para la comparación múltiple: 3.

Los resultados del ajuste son los siguientes:

### **Figura 2: Potencia como Función del Tamaño de Muestra bajo la Corrección de Bonferroni**

---

<sup>23</sup> Es importante anotar que no existe consenso universal respecto a la utilización de este tipo de ajuste en la literatura. Al respecto, ver Gelman et al (2012) para una discusión de porque no siempre tiene sentido preocuparse por comparaciones múltiples.



Se estima que serán necesarias **4,452 observaciones** para tener la capacidad de detectar un efecto de 0.2 desviaciones estándar con un nivel de potencia de 90%, para un diseño de diferencias en diferencias con asignación al tratamiento a nivel de clúster y ajuste por comparación múltiple, usando la corrección de Bonferroni. Así, la comparación múltiple hace referencia a dos grupos de tratamiento (modalidad Producto y Canasta Básica) y un grupo de control.

Maldonado (2013a y 2013b) presenta un análisis de poder más detallado. En particular, discute la sensibilidad de los resultados básicos aquí presentados, ante cambio en los supuestos básicos del ejercicio.

#### 8.4. Diseño Muestral

El cálculo de muestra anterior tomó como referencia un diseño de cluster a nivel distrito, con unidad de análisis a nivel de estudiante. Ciertamente, tomar el distrito como unidad de referencia introduce dificultades al diseño muestral. Así, por ejemplo, esto implicaría dificultades en términos de la elaboración de listado, una de las primeras tareas en el proceso de recojo de información. Sin embargo, el trabajo se reduce sustancialmente si dentro de cada distrito se selecciona aleatoriamente un centro poblado y dentro de dicho centro poblado uno o dos conglomerados definidos por el Instituto Nacional de Estadística e Informática (INEI). Los encuestadores actualizarían el listado solamente de los conglomerados elegidos, lo cual simplificaría el trabajo de campo.

De esta manera, la unidad primaria de muestreo (UPM) sería el centro poblado. Dentro de cada UPM, se seleccionará uno o dos conglomerados que constituirían la unidad secundaria de muestreo (USM). Cada USM está constituido por entre 100 y 120 viviendas si tomamos como referencia el diseño de la Encuesta Nacional de Hogares (ENAHOG). Finalmente, una vez identificados los hogares con niños en edad de atender primaria en el listado elaborado por el

encuestador, se procedería a seleccionar aleatoriamente la muestra final entre las viviendas con niños en edad escolar. Estas viviendas constituyen la unidad terciaria de muestreo (UTM).

Es importante notar que esta evaluación no toma la escuela como unidad de referencia para la recolección de información. Una parte de la literatura sobre programas de alimentación escolar típicamente procede a partir de información recolectada a nivel de escuela. Ello no permite evaluar el impacto del programa sobre matrícula pues la muestra estaría obviamente compuesta por estudiantes matriculados. Por esa razón, se opta por muestrear población en edad escolar en las zonas de referencia de la intervención. Ello tiene implicancias sobre el diseño de evaluación y el tamaño de muestra, como se discute en Maldonado (2013a). Así, por ejemplo, cuando se estima el tamaño de muestra tomando como referencia la escuela en vez del distrito, se obtiene que se necesitarían alrededor de 700 escuelas y 21,000 estudiantes para estimar un efecto mínimo detectable de 0.2 desviaciones estándar en el indicador de referencia. La razón de esta cifra elevada radica en la existencia de un nivel de correlación a nivel de cluster de 0.45 cuando se toma la escuela como referencia en vez del distrito. En ese sentido, existe un trade-off entre costos unitarios y costo totales del levantamiento de información pues aunque el diseño seleccionado puede tener altos costos unitarios por encuesta tiene un costo total menor.

### **8.5. Distribución Geográfica de Distritos Tratados y Controles**

De acuerdo con el diseño original de la intervención, el programa atenderá los primeros dos años a estudiantes de educación básica en las escuelas previamente atendidas por el PRONAA y en escuelas no atendidas previamente ubicadas en distritos con altos niveles de pobreza pertenecientes a los quintiles 1 y 2. Dado que el diseño de evaluación excluye a las escuelas previamente atendidas por el PRONAA, el marco muestral de la evaluación consistía originalmente en distritos del quintil más pobre (que conforman el grupo de tratamiento) y del segundo quintil (que conforman el grupo de control). Como se mencionó líneas arriba, el programa modificó durante su primer año de intervención, lo cual hace imposible determinar a la fecha la composición de los distritos tratados y controles para la evaluación.

El Apéndice VIII muestra la distribución espacial de los distritos tratados (Quintil I) y los distritos controles (Quintil II) para el diseño original de la intervención.

## **9. OBSERVACIONES FINALES**

En este documento, se ha presentado los elementos centrales de la evaluación de impacto del Programa Nacional de Alimentación Escolar *Qali Warma*. La discusión se ha concentrado en aspectos conceptuales vinculados con los problemas de evaluación y las preguntas de investigación del diseño. También se ha propuesto diversas estrategias para la evaluación del impacto del programa y discutido diversos criterios para el análisis de la información.

La variable de impacto central de la evaluación es una medida de procesos cognitivos básicos de atención y memoria de corto plazo. Se considera además el impacto sobre la ingesta

calórica, la participación en el programa y la asistencia y progreso educativo dentro del núcleo central del diseño de la evaluación. Complementariamente, el programa estudia dimensiones asociadas al rendimiento educativo, nutrición y salud, así como la potencial generación de externalidades a nivel del hogar y la escuela.

En función a las características del programa, se propone un diseño de diferencias en diferencias combinado con alguna variante del método de matching. Dicha estrategia será complementada con un diseño de cambios en cambios con el objeto de evaluar los impactos distributivos del programa. Se considera también el uso de técnicas para evaluar la sensibilidad de los resultados a relajamientos al supuesto de selección sobre observables asociado a la técnica de matching. En ese sentido, se considera que el diseño sugerido en este documento tiene un nivel de rigurosidad que va en línea con la práctica internacional, y que, en algunos aspectos, introduce innovaciones a las aproximaciones existentes para la evaluación de programas de alimentación escolar.

La variable central de la evaluación es una medida de atención y memoria de corto plazo. Esto implica el uso de un instrumento especializado para capturar esas dimensiones. En un apéndice a este documento, se ha cubierto con algún nivel de detalle los aspectos asociados a la medición de la variable dependiente del estudio. Guerrero y León (2013) ofrecen un tratamiento más completo al respecto.

La unidad de análisis de la evaluación es el estudiante. Dado que la implementación del programa se basa en el nivel de pobreza del distrito donde el estudiante se encuentra, se opta por un diseño que considera el distrito como el cluster de análisis. El análisis de potencia considera esta dimensión, calculando un tamaño de muestra de 3,464 observaciones en el caso del diseño estándar y 4,452 si se considera el ajuste por comparaciones múltiples. Para reducir los problemas prácticos que esto implicaría en términos del trabajo de campo, se propone seleccionar aleatoriamente un centro poblado por distrito. Este constituiría la UPM. Bajo la misma lógica se procedería con la selección de la USM (el conglomerado) y la UTM (la vivienda).

Se considera que el diseño propuesto en este documento contiene aspectos novedosos que permitirán mejorar la comprensión del impacto de los programas de alimentación escolar en el Perú y la región. Es, a nuestro entender, el primer diseño de evaluación en considerar explícitamente las dimensiones distributivas para el caso de los programas de alimentación escolar. Asimismo, se espera que estas innovaciones puedan servir de referencia a otras evaluaciones bajo responsabilidad del MIDIS.

## 10. BIBLIOGRAFIA

Abadie, Alberto y Guido Imbens (2006). "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica*, 74 (1), 235-267.

Abadie, Alberto y Guido Imbens (2010), "Bias-Corrected Matching Estimators for Average Treatment Effects," *Journal of Business and Economic Statistics*, 29(1), 1-11.

Adelman, Sarah; Daniel Gilligan, y Kim Lehrer (2008), "How Effective are Food for Education Programs? A Critical Assessment of the Evidence from Developing Countries," *Food Policy Review* 9, IFPRI, Washington DC.

Adroque, Cecilia y Maria Orlicki (2011), "Do in-School Feeding Programs Have Impact on Academic Performance and Dropouts? The Case of Public Argentine Schools," mimeo, Universidad de San Andres, Buenos Aires.

Ahmed, Akhter (2004), "Impact of Feeding Children in School: Evidence from Bangladesh," mimeo, IFPRI.

Alcazar, Lorena (2012), "Diseño del Programa Nacional de Alimentación Escolar *Qali Warma*," GTZ, Lima.

Angrist, Joshua y Jorn-Steffen Pischke (2009), *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press.

Arceneaux, Kevin; A. S. Gerber y D. P. Green (2006), "Comparing Experimental and Matching Methods Using a Large-Scale Voter Mobilization Experiment," *Political Analysis*, 14, 37-62.

Athey, Susan y Guido Imbens (2006), "Identification and Inference in Nonlinear Differences-in-Differences Models," *Econometrica*, Vol. 74, pp.431-497.

Banerjee, Abhijit y Esther Duflo (2009), "The Experimental Approach to Development Economics," *Annual Review of Economics*, 1, 151-178.

Belot, Michele y Jonathan James (2011), "Healthy School Meals and Educational Outcomes," *Journal of Health Economics*, 30, 489-504.

Bertrand, Marianne; Esther Duflo y Sendhil Mullainathan (2004), "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics*, 119, 249-275.

Black, Dan; Jose Galdo y Jeffrey Smith (2007), "Evaluating the Regression Discontinuity Design Using Experimental Data," mimeo, Chicago University.

Blundell, Richard y Monica Costa Dias (2009), "Alternative Approaches to Evaluation in Empirical Microeconomics," *Journal of Human Resources*, 44 (3), 565-640.

Buddelmeyer, Hielke y Enmanuel Skoufias (2003), "An Evaluation of the Performance of Regression Discontinuity Design on PROGRESA," IZA Working Paper 827, Bonn.

Bundy, Donald; Carmen Burbano, Margaret Grosh, Aulo Gelli, Matthew Jukes, y Lesley Drake (2008). *Rethinking School Feeding. Social Safety Nets, Child Development, and the Education Sector*. Directions in Development. World Bank.

Buttenheim, Alison; Harold Alderman y Jed Friedman (2011), "Impact Evaluation of School Feeding Programmes in Lao People's Democratic Republic," *Journal of Development Effectiveness*, 3(4), 520-542.

Caliendo, Marco y Sabine Kopeining (2008), "Some Practical Guidance for the Implementation of Propensity Score Matching," *Journal of Economic Surveys*, 22, 31-72.

Cameron, Colin; Jonah B. Gelbach, y Douglas L. Miller (2011), "Robust Inference with Multiway Clustering," *Journal of Business and Economic Statistics*, 29, 238-249.

Conti, Gabriella y James Heckman (2012), "The Economics of Child Well-Being," IZA Working Paper 6930.

Dewey, Kathryn y Seth Adu-Afarwuah (2008), "Systematic Review of the Efficacy and Effectiveness of Complementary Feedings Interventions in Developing Countries," *Maternal and Child Nutrition*, 4, 24-85.

Diamond, Alexis y Jasjeet Sekhon (2012), "Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies," Mimeo, Department of Political Science, UC Berkeley.

Duflo, Esther (2001), "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment," *American Economic Review*, 91(4), 795-813.

Frandsen, Brigham; Markus Frölich y Blaise Melly (2012), "Quantile Treatment Effects in the Regression Discontinuity Design," *Journal of Econometrics*, 168(2), 382-395.

Fink, Gunter; Margaret McConnell y Sebastian Vollmer (2011), "Testing for Heterogeneous Treatment Effects in Experimental Data: False Discovery Risks and Correction Procedures," mimeo, Harvard School of Public Health.

Gertler, Paul; Sebastian Martinez, Patrick Premand, Laura Rawlings, y Christel Vermeersch (2010). *Impact Evaluation in Practice*. World Bank.

Glewwe, Paul (2005), "The Impact of Child Health and Nutrition on Education in Developing Countries: Theory, Econometric Issues, and Recent Empirical Evidence," *Food and Nutrition Bulletin*, 26(2), S235-S250.

Green, Donald; T. Leong, H. Kern, A. Gerber y C. Larimer (2009). "Testing the Accuracy of the Regression Discontinuity Analysis using Experimental Benchmarks," *Political Analysis*, 17, 400-417.

Hansen, Christian (2007), "Generalized Least Squares Inference in Panel and Multilevel Models with Serial Correlation and Fixed Effects," *Journal of Econometrics*, 140, 670-694.

Heckman, James; Hidehiko Ichimura, y Petra Todd (1998a), "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies*, 65 (2), 261-294.

Heckman, James; H. Ichimura, J. Smith y P. Todd (1998b), "Characterizing Selection Bias Using Experimental Data," *Econometrica*, 66, 1017-1098.

Heckman, James; Robert LaLonde y Jeffrey Smith (1999), "The Economics and Econometrics of Active Labor Market Programs," *Handbook of Labor Economics*, Vol. 3. Elsevier Science.

Holland, P. W. (1986), "Statistics and Causal Inference," *Journal of the American Statistical Association*, 81, 945-970.

Imai, Kosuke; Luke Keele, Dustin Tingley, y Teppei Yamamoto (2011), "Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies," *American Political Science Review*, 105(4) 765-789.

Imbens, Guido (2004), "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review," *Review of Economics and Statistics*, 86, 4-29.

Imbens, Guido y Jeffrey Wooldridge (2009), "Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature*, March 2009.

Imbens, Guido y Karthik Kalyanaraman (2012), "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," *Review of Economic Studies*, 79, 933-959.

Jomaa, Lamis; Elaine McDonnell y Claudia Probart (2011), "School Feeding Programs in Developing Countries: Impacts on Children's Health and Educational Outcomes," *Nutrition Reviews*, 69(2), 83-98.

Kazianga, Harounan; Damien de Walque y Harold Alderman (2012), "Educational and Child Labour Impacts of Two Food-for-Education Schemes: Evidence from a Randomised Trial in Rural Burkina Faso," *Journal of African Economies*, 21(5), 723-760.

Koenker, Roger (2005), *Quantile Regression*, Cambridge University Press.

Kremer, Michael; Edward Miguel y Rebecca Thornton (2009), "Incentives to Learn," *Review of Economics and Statistics*, 91(3), 437-456.

Lee, David (2008), "Randomized Experiments from Non-Random Selection in U.S. House Elections," *Journal of Econometrics*, 142, 675-697.

Lee, David y Thomas Lemieux (2010), "Regression Discontinuity Design in Economics," *Journal of Economic Literature*, 48(2), 281-355.

Liang, Kung-Yee y Scott Zeger (1986), "Longitudinal Data Analysis using Generalized Linear Models," *Biometrika*, Vol. 73, 13-22.

Ludwig, Jens y Douglas Miller (2007), "Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design," *Quarterly Journal of Economics* 122, 159-208.

Mathematica Policy Research (1999), "Universal-Free School Breakfast Program Evaluation Design Project," USDA Food and Nutrition Service.

McCrary, Justin (2008), "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test," *Journal of Econometrics*, 142(2), 698-714.

McEwan, Patrick (2013), "The Impact of Chile's School Feeding Program on Education Outcomes," *Economics of Education Review*, 32, 122-139.

Meng, Xin y Jim Ryan (2010), "Does a Food for Education Program Affect School Outcomes? The Bangladesh Case," *Journal of Population Economics*, 23, 415-447.

Ministerio de Inclusión y Desarrollo Social (2012), "Consolidado de las Recomendaciones a PRONAA en base a la Evidencia Científica Nacional e Internacional," Informe Técnico 006-2012 DGSYE-VMPEM-MIDIS, Lima.

Moulton, B.R. (1990), "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units," *Review of Economics and Statistics*, 72, 334-38.

Murphy, Suzanne; Constance Gewa, Li-Jung Liang, Monika Grillenberger, Nimrod Bwibo y Charlotte Nuemann (2003), "School Snacks Containing Animal Source Foods Improve Dietary Quality for Children in Rural Kenya," *Journal of Nutrition*, 3950S-3956S.

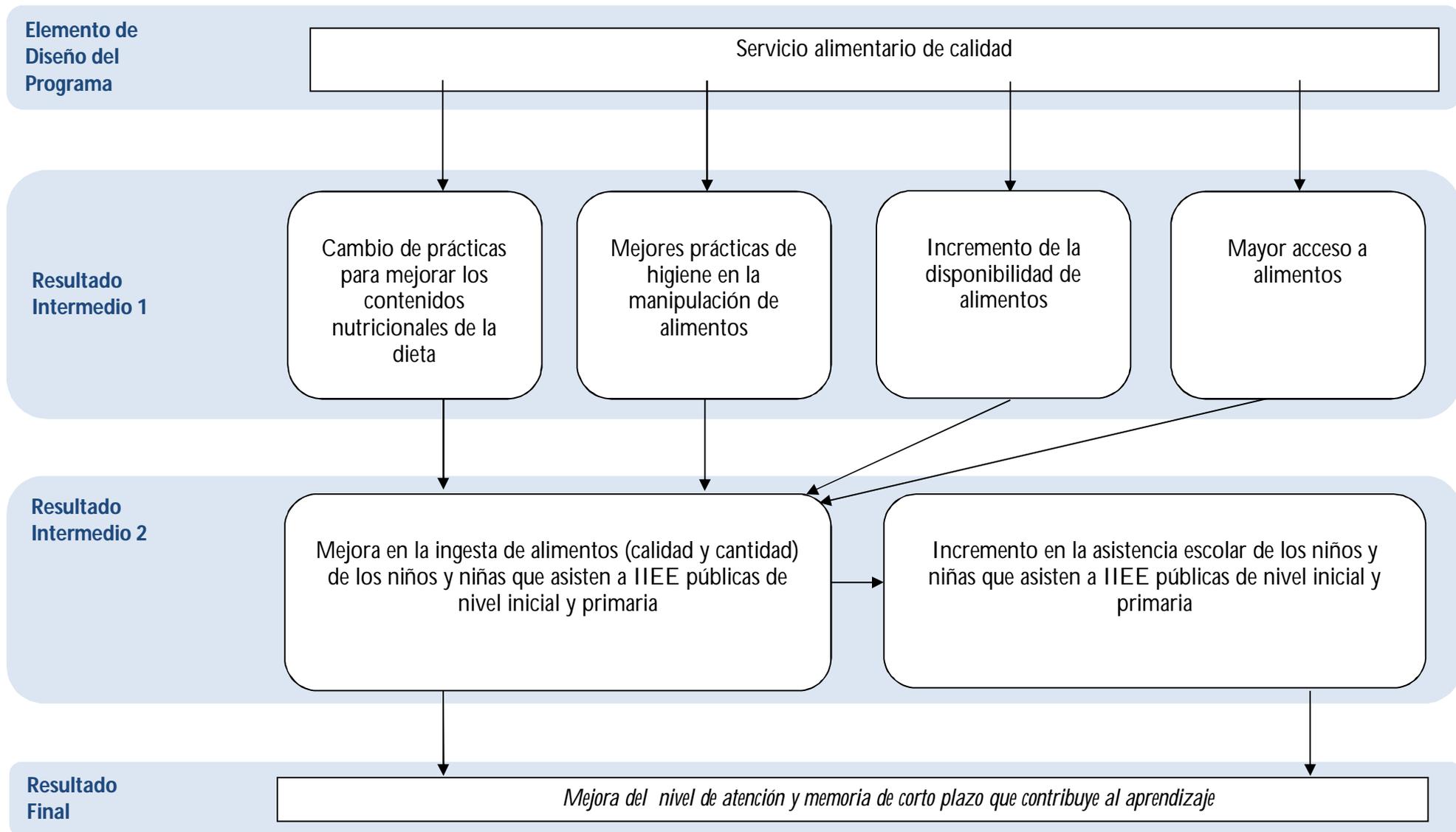
Perova, Elizaveta y Renos Vakis (2009), "Welfare Impacts of the 'Juntos' Program in Peru: Evidence from a Non-Experimental Evaluation," mimeo, World Bank.

Porter, Jack (2003), "Estimation in the Regression Discontinuity Model," mimeo, Harvard University.

Ravallion, Martin (2008), "Evaluating Anti-poverty Programs". *Handbook of Development Economics*, Volume 4. Elsevier Science.

- Rosenbaum, Paul (2005), "Sensitivity Analysis in Observational Studies," En: Brian Everitt y David Howell (Editores), *Encyclopedia of Statistics in Behavioral Science*. John Willey and Sons.
- Schanzenbach, Diane (2009), "Do School Lunches Contribute to Childhood Obesity?," *Journal of Human Resources*, 44(3), 684-709.
- Shadish, William; Thomas Cook y Donald Campbell (2002), *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin Company.
- Smith, Jeffrey y Petra Todd (2005), "Does Matching Overcome LaLonde's Critique of Non-experimental Methods?" *Journal of Econometrics*, 125, 305-353.
- Stuart, Elizabeth (2010), "Matching Methods for Causal Inference: A Review and a Look Forward," *Statistical Science*, 25, 1-21.
- Van der Klauw, Wilbert (2008), "Breaking the Link between Poverty and Low Student Achievement: An Evaluation of Title I," *Journal of Econometrics*, 142(2), 731-756.
- Van Stuijvenberg, Elizabeth; Jane Kvalsvig, Mieke Faber, Marita Krueger, Diane Kenoyer y Spinnler Benade (1999), "Effect of Iron-, Iodine-, and  $\beta$ -carotene Fortified Biscuits on the Micronutrient Status of Primary School Children: A Randomized Controlled Trial," *American Journal of Clinical Nutrition*, 69, 497-503.
- Vermeersch, Christel y Micheal Kremer (2004), "School Meals, Educational Achievement and School Competition: Evidence from a Randomized Evaluation," mimeo, World Bank.
- Whaley, Shannon; Marian Sigman, Charlotte Neuman, Nimrod Bwibo, Donald Guthrie, Robert Weiss, Susan Alber y Suzanne Murphy (2003), "The Impact of Dietary Intervention on the Cognitive Development of Kenyan School Children," *Journal of Nutrition*, 3965S-3971S.
- White, Halbert (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, Vol. 48, 817-838.

# APÉNDICE I: TEORÍA DE CAMBIO DE QALI WARMA



**APÉNDICE II:  
MATRIZ DE EVALUACIÓN**

Pregunta / Problema	Sub-preguntas	Principales Indicadores	Fuentes de Información	Instrumento
<b>1. Resultados e impactos: Mejor nivel de atención y memoria de corto plazo en los niños y niñas que asisten a los niveles de inicial y primaria de las instituciones educativas públicas.</b>				
<b>2. Resultados e impactos en la escuela primaria, mayores beneficios sociales y otras áreas que incluyen:</b>				
a) Asistencia y progreso educativo en primaria.	¿En qué medida QW mantiene a los niños en la escuela y facilita su progreso educativo?	<p><i>Comparación entre QW y no QW:</i></p> <ul style="list-style-type: none"> <li>- Tasas de matrícula en nivel primaria.</li> <li>- Tasas de asistencia en nivel primaria.</li> <li>- Tasas de promoción en nivel primaria.</li> <li>- Tasas de repitencia en nivel primaria.</li> <li>- Tasas de deserción en nivel primaria.</li> <li>- Tasa de retención (Matricula sexto sobre primer grado).</li> </ul>	<p>Análisis de datos secundarios (Censo Escolar).</p> <p>Encuestas cuantitativas (escuelas).</p>	Módulo sobre Escuela.
b) Atención y memoria de corto plazo.	¿En qué medida QW contribuye a la mejora de procesos cognitivos básicos como atención y memoria a corto plazo?	<p><i>Comparación entre QW y no QW:</i></p> <ul style="list-style-type: none"> <li>- Indicador de memoria de corto plazo en la prueba seleccionada.</li> <li>- Indicador de atención en la prueba seleccionada.</li> </ul>	Encuestas cuantitativas.	Módulo sobre Atención y Memoria.

c) Rendimiento educativo.	¿En qué medida QW promueve el aprendizaje de los niños en la escuela?	<p><i>Comparación entre QW y no QW:</i></p> <ul style="list-style-type: none"> <li>- Indicador de rendimiento escolar en lengua.</li> <li>- Indicador de rendimiento escolar en matemática.</li> </ul>	Análisis de datos secundarios (Evaluación Censal).	N.A.
d) Nutrición y salud.	¿En qué medida los niños que participan en la QW tienen dietas y condiciones nutricionales mejoradas?	<p><i>Comparación entre QW y no QW:</i></p> <ul style="list-style-type: none"> <li>- La ingesta dietética de niño (24 horas del día).</li> <li>- Índice de masa corporal de los niños en edad escolar primaria.</li> <li>- Indicadores antropométricos (talla y peso).</li> <li>- Porcentaje de niños en primaria con déficit calórico.</li> <li>- Micro y macronutrientes en la ración (proteínas, hierro).</li> <li>- Contenido de calorías (Kcal) en la ración.</li> </ul>	<p>Encuesta cuantitativa (niños, hogares).</p> <p>Encuesta antropométrica (niños en edad escolar).</p>	Módulo de Alimentos y Módulo de Salud.
	¿En qué medida los niños que participan en la QW han reducido la prevalencia y / o la duración de los episodios de morbilidad?	<p><i>Comparación entre QW y no QW:</i></p> <ul style="list-style-type: none"> <li>- La prevalencia de la morbilidad.</li> <li>- La duración media de los episodios de enfermedad.</li> <li>- Niveles de hemoglobina.</li> </ul>	<p>Encuesta cuantitativa.</p> <p>Análisis de datos secundarios (registros escolares).</p>	Módulo de Alimentos y Módulo de Salud.
	¿Cómo influye QW dentro del hogar, la alimentación y la distribución del trabajo relacionado?	<p><i>Comparación entre QW y no QW:</i></p> <ul style="list-style-type: none"> <li>- Cantidad y calidad de la ración.</li> <li>- Valor de comidas sustituido por la ración.</li> <li>- Valor de aperitivos sustituido por la ración.</li> <li>- Grado en que la sustitución de la comida proporciona más alimento para sus hermanos menores / otros miembros del hogar.</li> </ul>	<p>Encuesta cuantitativa (hogares).</p> <p>Encuesta cuantitativa (estudiantes).</p>	Encuesta de Hogar.

e) Impactos en las Escuelas, Hogares y la Comunidad	¿En qué medida QW proporciona un beneficio y sustento al hogar vulnerable?	-Situación socio económica del hogar. - Los patrones de los ingresos familiares gasto de los hogares. - La participación de los niños en edad escolar en el trabajo infantil.	Encuesta cuantitativa (hogares).	Módulo de Hogar.
	¿Cómo ha afectado QW la calidad de las IIEEs?	- Efectos de congestionamiento en el aula. - Asignación de horas de trabajo de los docentes. - Ratio Número de alumnos/Número de aulas.	Encuesta cuantitativa (escuela).	Módulo de Escuelas y Comunal.
	¿Cómo ha afectado la participación en QW sobre la comunidad y las IIEEs?	- Actividades de la escuela. - Nivel y tipo de participación de los padres en la escuela.	Encuesta cuantitativa (escuela).	Módulo de Escuelas y Comunal.
<b>3. Factores que afectan el impacto, incluyendo:</b>				
a) Heterogeneidad de las escuelas (según ubicación geográfica, plana docente, infraestructura, inputs escolares, etc.)	¿Cuál es la influencia de las características de la escuela y la plana docente sobre la atención y memoria, y el rendimiento educativo?	<i>Comparación de los diferentes tipos de escuela con QW:</i> - Calidad de la escuela (ratio profesor/alumno). - Calidad de los maestros (asistencia de maestros, tipo de titulación de docentes, etc.) - Características de la infraestructura educativa en relación a la calidad de la escuela y a la intervención de QW (ratio alumnos/aulas).	Encuesta cuantitativa (escuela), Censo Escolar y MINEDU.	Encuesta de Escuelas.

<p>b) Características sociales y económicas de los hogares.</p>	<p>¿En qué medida otros factores en la escuela y el entorno familiar afectan el éxito en la escuela primaria de los niños beneficiarios del programa QW ?</p>	<p><i>Comparación entre QW y no QW:</i></p> <ul style="list-style-type: none"> <li>- Condiciones de vida de los Hogares.</li> <li>- La capacidad de los padres para proporcionar otras formas de apoyo educativo (clases particulares, etc.).</li> </ul>	<p>Encuesta cuantitativa (hogares).</p> <p>Informantes clave (encuestas).</p>	<p>Módulo de Hogar.</p>
<p>c) Diferencias en niveles de vida de los hogares.</p>	<p>¿Cuál es el costo de oportunidad de la educación para los hogares beneficiarios de QW?</p> <p>¿Cuál es el efecto incentivo de la ración de QW sobre los hogares en términos de enviar a un niño a las actividades escolares?</p>	<p><i>Dentro de las comunidades con QW, comparación de los hogares y los niños participantes y no participantes en QW:</i></p> <ul style="list-style-type: none"> <li>- Valor del trabajo de los niños frente a la educación (por edad y sexo).</li> <li>- Valor de la escuela y la enseñanza-aprendizaje en el aula como incentivo para que los niños vengan o los padres envíen a sus hijos a la escuela.</li> </ul>	<p>Encuesta cuantitativa (hogar).</p>	<p>Encuesta de Hogares.</p>
<p>d) Otras intervenciones de redes de protección en las zonas seleccionadas.</p>	<p>¿En qué medida han influido otras intervenciones (Juntos, Pensión 65, FONCODES, etc.) en el impacto del programa QW?</p> <p>¿En qué medida otras intervenciones y actividades importantes han influido en el desempeño escolar y el entorno escolar?</p>	<ul style="list-style-type: none"> <li>- Número de hogares que se benefician de otras intervenciones.</li> <li>- Percepción los beneficios reales de otras intervenciones.</li> </ul>	<p>Encuesta cuantitativa (escuelas).</p> <p>Entrevistas (profesores y hogar).</p>	<p>Encuesta de Hogares.</p>
<p><b>4. Los cambios que se podrían hacer en la estrategia y el diseño del programa.</b></p>				

a) CAE	<p>¿Cuáles son los factores clave que influyen en el involucramiento de los hogares en los CAE?</p> <p>¿Qué asociaciones entre las ONG, organizaciones sociales de base y el sector privado son necesarias para el apoyo y soporte institucionales a los CAE?</p> <p>¿Cómo puede mejorar QW a través de los CAE la relación escuela-comunidad para mejorar la educación?</p>	- Opiniones y percepciones de los involucrados (oficiales de QW, miembros de CAE y Comités de Compras, ONG, Organizaciones Sociales de Base, etc).	Encuesta cuantitativa (hogar, escuela).  Entrevistas a informantes clave.	Encuesta de Escuelas.
b) Ministerio de Educación (vía UGELs y otras unidades administrativas)	¿Qué acciones o aportes podrían ser adoptadas por QW para cooperar con MINEDU hacer frente a los obstáculos a la educación de calidad?	- Capacitación docente.  - Otros por determinarse.	Entrevistas a informantes clave (Gobierno, los asociados)	Entrevistas.
c) Qali Warma	<p>¿Qué tan eficiente y oportuna es la entrega de la ración de QW?</p> <p>¿Cuál es la percepción de la población beneficiaria sobre QW?</p>	<p>- Porcentaje de raciones entregadas a los CAE en forma oportuna</p> <p>- Duración promedio del servicio de atención de QW durante la jornada escolar</p> <p>- Aceptación de la ración por la población beneficiaria</p> <p>- Percepción de QW por la población beneficiaria, padres de familia y comunidad escolar</p>	Encuesta cuantitativa (escuelas).  Análisis de datos secundarios (logística del programa).	Encuesta de Escuelas.

Fuente: Elaboración propia.

## APÉNDICE III: ESTIMACIÓN DE EFECTOS DE TRATAMIENTO EN EL CASO DE *QALI WARMA*

### 1. Introducción

El propósito de este apéndice consiste en presentar un breve resumen de la literatura de estimación de los efectos de tratamiento con las técnicas sugeridas en las secciones 6 y 7. Un tratamiento técnico de los tópicos a ser cubiertos en este breve apéndice puede encontrarse en Heckman, La Londe y Smith (1999); Heckman, Ichimura y Todd (1997, 1998); Imbens (2004); Blundell y Costa-Dias (2009); Wooldridge y Imbens (2009); Gertler et al (2011); Ravallion (2008) y Shadish, Campbell y Cook (2002).

### 2. Enfoque de Resultados Potenciales

El objetivo fundamental de esta literatura consiste en determinar el efecto de un tratamiento sobre un resultado a nivel individual. En el contexto de este estudio, los resultados considerados son la participación educativa y logro educativo, mientras que el tratamiento vendría a ser la presencia (o ausencia) de *Qali Warma*. El problema fundamental con el que se enfrenta un análisis de esta naturaleza radica en el hecho de que no es posible observar a una misma persona bajo los dos estados (es decir, recibiendo y no recibiendo el programa en un mismo momento del tiempo). De ello se desprende que encontrar un grupo de control adecuado es fundamental a fin de hacer posible la comparación entre aquellos que recibieron el tratamiento y los que no, sobre todo porque los que recibieron el tratamiento (en nuestro caso, *Qali Warma*) pueden diferir de aquellos que no lo recibieron en muchas más dimensiones que exclusivamente el tratamiento. De esta manera, la estimación del impacto del tratamiento puede verse afectado por la presencia de sesgo de selección.

Existen diversas técnicas en la literatura diseñadas con el propósito de estimar el efecto del tratamiento en entornos no-experimentales. Dado que la asignación al tratamiento no es aleatoria, el objetivo de estos métodos consiste en usar la información disponible de los tratados y no tratados a fin de hacer posible la comparación entre ambos grupos. En particular, discutiremos las técnicas de regresiones discontinuas, diferencias y diferencias, y emparejamiento a ser utilizadas en el marco de una evaluación.

La idea básica de estos métodos consiste en estimar el impacto de un tratamiento siguiendo el enfoque de resultados potenciales de Rubin<sup>24</sup>. Bajo este enfoque, cada individuo tiene dos respuestas potenciales a un tratamiento. Sea  $Y_{1i}$  el resultado (en nuestro caso, ser expuesto o participación en *Qali Warma*) si es que el individuo  $i$  es expuesto al tratamiento, y sea  $Y_{0i}$  el resultado en caso el mismo individuos no sea tratado. Sea asimismo  $T_i = 1$  el caso en el que el individuo  $i$  es tratado, o sea si recibe *Qali Warma*, y  $T_i = 0$  cualquier otro caso. Bajo estas condiciones, el parámetro de interés consiste en estimar el impacto promedio de recibir *Qali Warma*, el cual no es otra cosa mas que la diferencia entre la situación con tratamiento (es decir, recibir *Qali Warma*) y la situación sin

---

<sup>24</sup> Para una discusión respecto al enfoque de Rubin, véase Holland (1986).

tratamiento, dado el tratamiento. En la literatura, dicho parámetro es conocido como el efecto promedio del tratamiento sobre los tratados o ATET por sus siglas en ingles. Formalmente:

$$(A3.1) \quad \tau_{ATET} = E[Y_{1i} - Y_{0i} / T = 1] = E[Y_{1i} / T = 1] - E[Y_{0i} / T = 1]$$

Dado que es imposible observar a un mismo individuo  $i$  en el escenario contra-factual en un mismo momento del tiempo,  $ATET$  no está identificado a partir de los datos. En caso la probabilidad de ser receptor de *Qali Warma* siguiese un proceso aleatorio, la estimación de  $ATET$  sería posible utilizando simplemente el efecto promedio para el caso del grupo de control, esto es, aquellos que no recibieron el tratamiento. Como se discutió en la sección correspondiente, este no parece ser el caso.

Ante la imposibilidad de aproximarnos plenamente a un entorno experimental, supuestos de identificación son necesarios. Una posible solución consiste en considerar que las características observables de los individuos influyen sobre el tratamiento y el resultado de interés, asumiendo asimismo que dichas características observables no están afectadas por el tratamiento. En esas condiciones, el modelo queda identificado si es que condicionamos sobre el vector de características observables  $X$  debido a que dentro de los subgrupos definidos por  $X$  ser parte del grupo de control no estaría correlacionado con el resultado en el escenario contra-factual en caso se hubiera sido parte del grupo de tratados. Este supuesto es conocido en la literatura como “independencia condicional”, “selección sobre observables” o “asignación ignorable al tratamiento”. Formalmente:

$$(A3.2) \quad (Y_{1i}, Y_{0i}) \perp T / X = x; \quad \forall x \in X; \quad X \subseteq \Omega$$

En esas condiciones,  $E[Y_{0i} / T = 1, X] = E[Y_{0i} / T = 0, X]$  estando por tanto  $ATET$  identificado. De esta manera  $ATET$  sería:

$$(A3.3) \quad ATET = E[Y_{1i} / T = 1] - E_{X/S=1} [E[Y_{0i} / T = 0, X = x] / T = 1]$$

Un factor importante a considerar en el análisis es que bajo este supuesto no tomamos en cuenta el rol que pueden jugar las variables no observables en el análisis. Adicionalmente, un supuesto importante a fin de identificar el impacto del tratamiento radica en asumir que el tratamiento no tiene un impacto sobre el conjunto de variables explicativas observables  $X$  pues ello configuraría un problema de endogeneidad.

### 3. Diferencias en Diferencias

Para la evaluación de impacto de *Qali Warma* se considera un diseño de diferencias en diferencias. En este caso, se requiere disponer de información de línea de base y de por lo menos de un seguimiento para recuperar el efecto causal de interés. La intuición básica del método consiste en comparar al grupo de tratamiento y control, antes y después de la intervención. Asumiendo que las tendencias entre el grupo de tratamiento y control son las mismas, esta aproximación permite recuperar el efecto de tratamiento.

En este escenario, es preciso incorporar la dimensión temporal en el análisis. El tratamiento puede ser escrito del modo siguiente:

$$(A3.4) T_i = \begin{cases} 1 & \text{si } t_{i1} = 1 \\ 0 & \text{en otro caso.} \end{cases}$$

Bajo el supuesto de tendencias comunes y de no selección en base a shocks transitorios, es posible escribir la ecuación de interés de la evaluación del modo siguiente:

$$(A3.5) y_{it} = \alpha + \beta T_{it} + \varepsilon_{it};$$

donde  $E[\varepsilon_{it}/T_i, t] = E[p_i/T_i] + q_t$ , siendo  $p_i$  un efecto fijo individual no observable para el analista y  $q_t$  un shock macro agregado. Bajo los supuestos anteriores podemos escribir la siguiente expresión:

$$(A3.6) E[y_{it}/T_i, t] = \begin{cases} \alpha + E[\beta/T_i = 1] + E[p_i/T_i = 1] + q_t & \text{if } T_{it} = 1, t = 1 \\ \alpha + E[p_i/T_i] + q_t. & \end{cases}$$

De lo anterior se desprende que la constante y los términos de error pueden ser eliminados mediante el uso secuencial de diferencias del modo siguiente:

$$(A3.7) \beta^{DID} = \{E[y_{it}/T_i = 1, t = 1] - E[y_{it}/T_i = 1, t = 0]\} \\ - \{E[y_{it}/T_i = 0, t = 1] - E[y_{it}/T_i = 0, t = 0]\}.$$

Esta expresión puede ser calculada a partir del análogo muestral correspondiente. En este escenario, es posible estimar *ATE*.

Es importante notar que aquí hemos considerado el caso simple de dos periodos y dos unidades. La ecuación (1) en el texto es simplemente una generalización para más de un periodo y más de una unidad tratada.

#### 4. Matching

En el caso del diseño anterior, es posible que heterogeneidades a nivel de individuos puedan introducir sesgos en la estimación del efecto de tratamiento. En ese contexto, la combinación de DID con matching puede contribuir sustancialmente con la reducción de dichos sesgos.

A fin de utilizar matching, un supuesto adicional es necesario a fin de asegurar la consistencia de nuestros estimadores. Necesitamos que para cada unidad tratada exista una unidad de control con similares características  $X$ . Este supuesto, conocido como la condición de soporte común, puede ser expresado de la siguiente forma:

$$(A3.8) 0 < P(T = 1/X_1 = x_1) < 1; \quad \forall x_1 \in X_1 \subset \square$$

De acuerdo con este supuesto, si todos los individuos con un patrón similar de variables explicativas eligen el tratamiento, entonces no deberían existir observaciones sobre individuos similares que decidan no aceptar el tratamiento (Abadie, Drukker, Leber e Imbens 2004).

Como ya se mencionó, diversas estrategias de emparejamiento han sido sugeridas en la literatura. De acuerdo con Imbens (2004), estas pueden ser agrupadas en: a) emparejamiento de covariados, b) emparejamiento basados en el *propensity score* y, c) combinación de algunas de las variantes entre los dos grupos anteriores con otras estrategias no paramétricas<sup>25</sup>. En las líneas que siguen discutiremos brevemente las dos estrategias complementarias que utilizaremos en este estudio.

#### 4.1. Emparejamiento por Propensity Score

Empecemos con las estrategias de emparejamiento basadas en *propensity score*, nombre con el cual se conoce a la probabilidad condicional al tratamiento,  $p(x_i) = P[T = 1/X_i = x_i]$ . La idea básica detrás de esta literatura consiste en obtener un estimado de *ATE* que no este sujeto a la “maldición de la dimensionalidad”; esto es, una baja densidad por celdas cuando se utiliza alguna estrategia no paramétrica para una alta dimensionalidad de covariados  $X$ . Este problema se expresa en un alto nivel de imprecisión en los estimados. De acuerdo con Rosenbaum y Rubin (1983), solo es necesario comparar observaciones con el mismo *propensity score*, el cual tiene la virtud de reducir el problema de estimación que implicaría estimar *ATE* con varias dimensiones a una estimación de una sola dimensión. De esta manera, la estimación de *ATE* procedería como sigue:

$$(A3.9) \quad ATE = E[Y_i/T = 1] - \int_{p(x_i)/S=1} E[Y_{0i}/T = 0, p(X_i) = p(x_i)]/T = 1$$

Existen diversas formas de estimar *ATE* mediante el *propensity score*. En particular, en este estudio consideraremos el método del vecino más cercano, el de Kernel, el de radio y el estratificado. La diferencia básica entre estas aproximaciones radica en la forma en que se pesan las observaciones que son asignadas a los grupos de tratados y de control.

Discutamos brevemente cada uno de los métodos mencionados. Sea  $C$  el conjunto de unidades de control. Sea  $C(i)$  el conjunto de unidades de control emparejadas a la unidad tratada  $i$  con un valor estimado de *propensity score* de  $p(x_i)$ . El método de emparejamiento de vecino más cercano establece que  $C(i) = \min_j \|p(x_i) - p(x_j)\|$ . Asimismo, en el caso del método de radio, todas las unidades de control con valores de *propensity score* que caen dentro de un radio  $r$  determinado desde  $x_i$  son emparejadas a la unidad tratada  $i$ ; esto es,  $C(i) = \{p(x_j) \mid \|p(x_i) - p(x_j)\| < r\}$ . Denotemos ahora al número de unidades de control emparejadas con la unidad tratada  $i$  como  $N_i^C$  y definamos los pesos  $w_{ij}$  como  $w_{ij} = 1/N_i^C$  en caso  $j \in C(i)$  o  $w_{ij} = 0$  en otros casos. Dado lo anterior, los estimadores de emparejamiento de vecino más cercano y de radio pueden escribirse -después de algunas manipulaciones- de la forma siguiente:

$$(A3.10) \quad ATE^{VMCR} = \frac{1}{N^T} \sum_{i \in T} Y_i^T - \frac{1}{N^T} \sum_{j \in C} w_j Y_j^C$$

<sup>25</sup> La clasificación de Imbens (2004) incluye además modelos no paramétricos y métodos bayesianos, los cuales no serán discutidos ni utilizados en este estudio.

El estimador de emparejamiento por el método de Kernel tiene la siguiente forma:

$$(A3.11) \quad ATET^K = \frac{1}{N^T} \sum_{i \in T} \left\{ Y_i^T - \frac{\sum_{j \in C} Y_j^C G\left(\frac{p(x_j) - p(x_i)}{h_n}\right)}{\sum_{k \in C} G\left(\frac{p(x_k) - p(x_i)}{h_n}\right)} \right\}$$

En donde  $G(\cdot)$  es la función Kernel y  $h_n$  es el parámetro de ancho de banda.

Finalmente, el estimador de emparejamiento estratificado se basa en el procedimiento de estratificación utilizado en el computo del *propensity score*. En dicho procedimiento, se computan bloques o estratos dentro de los cuales la asignación al tratamiento se considera aleatoria y los covariados están balanceados. El estimador de emparejamiento es entonces:

$$(A3.12) \quad ATET^S = \sum_{q=1}^Q \left( \frac{\sum_{i \in I(q)} Y_j^T}{N_q^T} - \frac{\sum_{j \in I(q)} Y_j^C}{N_q^C} \right) \frac{\sum_{i \in I(q)} T_i}{\sum_{\forall i} T_i}$$

Donde  $q$  cataloga a los estratos definidos sobre intervalos del *propensity score*;  $I(q)$  es el conjunto de unidades en el estrato  $q$ ;  $N_q^T$  y  $N_q^C$  definen respectivamente el número de unidades tratadas y de control dentro de cada estrato  $q$ , y  $Q$  es el número de estratos. El peso de cada estrato viene dado por la proporción de unidades tratadas dentro de cada uno de ellos.

Antes de culminar esta sección, convendría discutir brevemente las virtudes y limitaciones de los algoritmos presentados para el cómputo del estimador de emparejamiento mediante el *propensity score*. Como es reconocido por Heckman, Ichimura y Todd (1997), la elección del algoritmo puede ser importante cuando el tamaño muestral es pequeño. Así, la elección entre algoritmos implicará siempre una disyuntiva entre sesgo y varianza. Va más allá de los objetivos de este apéndice una discusión detallada respecto a las diferencias entre las diferentes alternativas discutidas. El lector interesado puede encontrar una discusión al respecto en Smith (2000), y Caliendo y Kopeining (2008).

## 4.2. Emparejamiento por Covariados

Discutamos ahora brevemente la propuesta de Abadie e Imbens (2006, 2011). Estos autores han sugerido un estimando alternativo denominado efecto promedio condicional del tratamiento sobre los tratados o *CATET* por sus siglas en ingles. Formalmente:

$$(A3.13) \quad CATET = E_{i:T_i=1}[Y_{1i} - Y_{0i} / X_i]$$

Este estimando se computa mediante técnicas de emparejamiento. Dicho estimando es propuesto por Abadie e Imbens (2006, 2011) en un contexto de discusión sobre las propiedades de los estimadores de emparejamiento para muestras grandes. De acuerdo con estos autores, los métodos de emparejamiento tradicionales trabajan con estimadores que incluyen un termino de

sesgo condicional que no permite que el estimador sea consistente a  $\sqrt{N}$ . Asimismo, estos métodos suelen computarse sin reemplazo; esto es, cada unidad del grupo de control es emparejada con una y solo una tratada. Abadie e Imbens (2002, 2006) proponen un estimador de emparejamiento simple con reemplazo para un número fijo de emparejamientos, el cual tiene la virtud de producir un menor sesgo aunque una mayor varianza, y sugieren un estimador corregido de sesgo que es consistente a la tasa de  $\sqrt{N}$ .

Introduzcamos algunos elementos previos antes de discutir con algún nivel de detalle los estimadores mencionados. Sea  $\|x\|_v = (x'Vx)^{1/2}$  el vector norma con una matriz de pesos positiva definida  $V$  y  $\|z-x\|_v$  la distancia entre los vectores  $x$  y  $v$ . Sea, además,  $d_M(i)$  la distancia desde los covariados  $X_i$  para la unidad  $i$  al  $M$ -ésimo emparejamiento más cercano con el tratamiento contrario, donde  $d_M(i)$  es un número real. Denominemos por  $J_M(i)$  al conjunto de índices para los emparejamientos para la unidad  $i$  que está al menos tan cerca como  $M$ -ésimo emparejamiento. Finalmente, sea  $K_M(i)$  la suma de pesos de la unidad  $i$  como un emparejamiento para las otras unidades y  $K_M'$  la suma de los pesos al cuadrado en los emparejamientos. Formalmente:

$$(A3.14) \quad \begin{aligned} K_M(i) &= \sum_{i=1}^N 1\{i \in J_M(i)\} \cdot \frac{1}{\#J_M(i)} \\ K_M'(i) &= \sum_{i=1}^N 1\{i \in J_M(i)\} \cdot \left( \frac{1}{\#J_M(i)} \right)^2 \end{aligned}$$

Dado lo anterior, definamos ahora los estimadores sugeridos por Abadie e Imbens (2006, 2011). En primer lugar, consideremos el *estimador simple* definido como:

$$(A1.11) \quad ATET_{AI}^S = \frac{1}{N_T} \sum_{i=1}^N [T_i - (1-T_i)K_M(i)]Y_i$$

Un problema con el estimador simple de emparejamiento es que es susceptible de sesgos en muestras finitas cuando el emparejamiento no es exacto. A fin de dar cuenta de esta limitación, los autores sugieren un *estimador corregido de sesgo* que ajusta las diferencias en el proceso de emparejamiento por las diferencias en sus valores de covariados. Dicho ajuste se basa en la estimación de la siguiente función de regresión:  $\hat{\mu}_T(x) = E[Y(T)/X=x]$ , la cual es aproximada mediante funciones lineales estimadas vía mínimos cuadrados sobre la muestra emparejada.

Para el cálculo del estimador corregido por sesgo y ajustado heterocedasticidad, tanto el estándar como el que utiliza la métrica de Mahalanobis, referimos al lector interesado directamente a Abadie e Imbens (2006) para más detalles al respecto.

## 5. Efectos Distributivos

A fin de evaluar impactos distributivos, la empresa seleccionada deberá complementar el análisis anterior mediante el uso de la técnica de regresiones cuantílicas. Dado el diseño de evaluación, ello implicara el uso de una técnica de diferencias en diferencias cuantílicas. La ventaja de esta aproximación es que permite evaluar el impacto de la intervención en diferentes partes de la distribución de la variable de impacto y no solamente en un momento de esta (la media). La aproximación sugerida es la generalización del modelo de diferencias en diferencias sugerida por Athey e Imbens (2006) conocida como el modelo de cambios en cambios (CIC).

Esta aproximación tiene la ventaja de estimar toda la distribución hipotética del efecto del tratamiento en los tratados en un entorno no paramétrico. El modelo básico en (1) se extiende utilizando la siguiente especificación para los resultados en ausencia de la intervención:

$$(A3.15) Y_i^0 = h(u_i, t_i),$$

donde  $h(\cdot)$  es creciente en las características no observables  $u$  y  $t$  representa los efectos fijos de tiempo. A distribución de no observables puede cambiar a través de los distritos pero no a través de tiempo. Bajo un conjunto de supuestos, el modelo CIC es identificado y la distribución de  $Y_{11}^0$  es:

$$(A3.16) F_{Y^0,11}(y) = F_{Y,10}(F_{Y,00}^{-1}(F_{Y,01}(y))),$$

donde  $F_{Y,00}^{-1}(\cdot)$  es la inversa de la función de distribución. El efecto promedio del tratamiento puede ser escrito del modo siguiente:

$$(A3.17) \begin{aligned} \beta^{CIC} &= E[Y_{11}^1 - Y_{11}^0] = E[Y_{11}^1] - E[\kappa^{CIC}(Y_{10})] \\ &= E[Y_{11}^1] - E[F_{Y,01}^{-1}(F_{Y,00}(Y_{10}))], \end{aligned}$$

donde  $\kappa^{CIC}(\cdot)$  es una transformación que permite a la distribución contra-fáctica del segundo periodo  $Y_{11}^0$  ser igualada a la distribución de  $\kappa^{CIC}(Y_{10})$ .  $\beta^{CIC}$  puede ser estimada utilizando la distribuciones empíricas y las medias muestrales.

## APENDICE IV:

### ESTUDIOS SELECCIONADOS SOBRE EL IMPACTO DE PROGRAMAS DE ALIMENTACION ESCOLAR

Estudio (I)	Descripción de la intervención (II)	Tratamiento y Variables de Interés (III)	Población de Referencia (IV)	Técnicas Utilizadas y Datos (V)	Principales Hallazgos (VI)
McEwan (2012), Programa de Alimentación Escolar-Chile	<ul style="list-style-type: none"> <li>Se estableció estándares mínimos para las comidas: 1) contenido nutricional y calorías, 2) frecuencias mínimas y máximas de ciertos alimentos, 3) umbrales de calidad de los ingredientes y 4) condiciones mínimas de las condiciones de operación e infraestructura del servicio.</li> <li>Las calorías de las comidas varían de acuerdo a la escuela: 1) desayuno y almuerzo (o almuerzo y snack) con 700 kcal/día para las escuelas inicial, 2) las escuelas primaria reciben una de las tres opciones: desayunos de 250kcal/día, desayunos/almuerzos o almuerzos/snacks de 700kcal/día, o desayuno/almuerzo o almuerzo/snack de 1000 kcal/día y 3) las escuelas secundaria reciben desayunos de 350 kcal/día o almuerzos de 650 kcal/día.</li> </ul>	Impacto del contenido calórico sobre indicadores de matrícula, repetición, asistencia y test scores.	<ul style="list-style-type: none"> <li>Alumnos de educación primaria que estudian en escuelas consideradas vulnerables (públicas o privadas).</li> </ul>	Regresión discontinua utilizando como variable de asignación un índice de "vulnerabilidad". Se utiliza diversas líneas de corte para asignar niveles distintos de calorías en los alimentos. Se utilizan datos administrativos.	<ul style="list-style-type: none"> <li>No se observa un incremento en la matrícula ni la asistencia.</li> <li>No se encontró efecto en la edad de matrícula ni en la repetición del primer grado.</li> <li>No efecto en los test scores de matemática y lenguaje.</li> </ul>
Kazianga et al (2012), programa de alimentación escolar en Sahel, Burkina Faso.	<ul style="list-style-type: none"> <li>El programa de provisión de alimentos provee desayunos y/o almuerzos escolares posiblemente.</li> <li>El programa de entrega de alimentos para el hogar provee a las niñas de 10kg de harina de cereal cada mes condicionado a tener 90%</li> </ul>	Comparación de dos alternativas, un esquema (A1) de provisión de alimentos en la escuela, y otro (A2) de entrega condicionada de	<ul style="list-style-type: none"> <li>Ambos programas fueron para niños en etapa escolar (6 a 15 años).</li> <li>Mientras que el programa de provisión de alimentos en la</li> </ul>	Experimento aleatorio. Asignación a nivel de escuelas (16 escuelas al esquema A1, 16 a A2 y 14 al grupo de control).	<ul style="list-style-type: none"> <li>Aumento en alrededor 5 puntos porcentuales en la matrícula escolar en ambos programas. Incremento en 8.4-9.6% en test de aritmética en ambos programas.</li> <li>No efecto sobre desarrollo</li> </ul>

Estudio (I)	Descripción de la intervención (II)	Tratamiento y Variables de Interés (III)	Población de Referencia (IV)	Técnicas Utilizadas y Datos (V)	Principales Hallazgos (VI)
	de asistencia.	alimentos para el hogar. Se evalúa el impacto de ambas alternativas sobre matrícula, asistencia, performance académica y desarrollo cognitivo.	escuela no discrimina entre géneros, el programa de entrega de alimentos para el hogar solo va dirigido a las niñas.		cognitivo. • Aumento en asistencia en 0.7-0.9 días en ambos programas. • No efecto sustancial sobre el trabajo infantil para los niños previamente matriculados.
Adroque et al (2011), programas de alimentación escolar, Argentina.	<ul style="list-style-type: none"> <li>• El programa en sí hace entrega de fondos a las provincias para que fortalezcan las comidas entregadas en los colegios. Así, hay una gran variedad de situaciones.</li> <li>• El programa no tiene un marco legal en el que especifique los objetivos en términos de micronutrientes ni calorías que deban otorgarse, los menús son monótonos y proveen un contenido nutricional similar al recibido en casa de modo que no complementan los déficits del hogar, y principalmente se trata de almuerzos (pocas veces desayuno).</li> </ul>	Impacto de programas de alimentación escolar sobre el rendimiento académico y la deserción escolar.	<ul style="list-style-type: none"> <li>• Las escuelas (públicas) que reciben el programa son aquellas que atienden a niños de estatus socioeconómico bajo.</li> <li>• El estudio se centra en estudiantes de tercer año de educación primaria o educación básica general.</li> </ul>	Modelo estándar de diferencias en diferencias utilizando información de encuestas de hogares y datos administrativos. Se explota variación en el tiempo y espacio de la participación en programas de alimentación escolar.	<ul style="list-style-type: none"> <li>• No efecto sobre la deserción escolar.</li> <li>• No efecto sobre el rendimiento en los test de matemática. Incremento de 3.5% en los resultados del test de lengua.</li> </ul>
Belot y James (2011), campaña "Feed me Better" en Greenwich, Reino Unido.	<ul style="list-style-type: none"> <li>• El programa consistió en modificar los menús escolares con el mismo presupuesto. Se reemplazó las comidas chatarras (escasez de nutrientes básicos como hierro y vitamina C) por comidas nutritivas.</li> </ul>	Impacto de cambios en los estándares nutricionales de los alimentos provistos en las escuelas sobre rendimiento escolar, participación en el programa de alimentos	<ul style="list-style-type: none"> <li>• El estudio se limita al caso de estudiantes de educación primaria.</li> </ul>	Método estándar de diferencias en diferencias. Se utilizan como grupo de control 5 barrios de Londres con características similares a Greenwich. Los datos a nivel individual y de escuela provienen de registros administrativos.	<ul style="list-style-type: none"> <li>• Incremento en test scores de los alumnos de tercer a sexto grado: 3% en matemáticas, 6% en inglés y 8% en ciencias.</li> <li>• Reducción de 14% en las inasistencias autorizadas (usualmente asociadas a problemas de salud). No</li> </ul>

Estudio (I)	Descripción de la intervención (II)	Tratamiento y Variables de Interés (III)	Población de Referencia (IV)	Técnicas Utilizadas y Datos (V)	Principales Hallazgos (VI)
		escolares y asistencia.			efecto sobre las asistencias no autorizadas. • No evidencia de incremento en la participación en el programa de alimentos.
Butteheim et al (2011), programas de alimentación escolar administrados por el PMA en 3 provincias del norte de Laos.	<ul style="list-style-type: none"> <li>• El programa de alimentación escolar en la escuela provee un snack diario hecho de maíz y soya.</li> <li>• El programa de raciones para llevar a la casa provee a las niñas de conservas de pescado y arroz.</li> </ul>	Comparación de programas de alimentación escolar en la escuela con programas de raciones para llevar a casa. Se evalúa impacto sobre matrícula y resultados nutricionales.	<ul style="list-style-type: none"> <li>• El estudio analiza el caso de estudiantes en edad de 3 a 14 años de zonas rurales.</li> </ul>	Método de diferencias en diferencias combinado con propensity score matching. Los datos provienen de una muestra longitudinal de 4,500 hogares con niños en edad escolar.	<ul style="list-style-type: none"> <li>• Evidencia débil pero no concluyente del impacto de los programas sobre la matrícula.</li> <li>• Alguna evidencia de incremento de matrícula en edad temprana, pero no concluyente en áreas en donde ambos tratamientos estuvieron disponibles.</li> <li>• Evidencia mixta pero inconsistente sobre indicadores nutricionales (peso para edad, talla para edad y anemia).</li> </ul>
Schanzenbach (2009), programa nacional de almuerzos escolares en Estados Unidos.	<ul style="list-style-type: none"> <li>• Los programas de alimentación brindan tanto almuerzos como desayunos escolares.</li> </ul>	Impacto del programa alimentario en indicadores de masa corporal y obesidad.	<ul style="list-style-type: none"> <li>• El estudio analiza niños desde que entran a inicial hasta que alcanzan el octavo grado.</li> </ul>	Dos metodologías: análisis de cohortes via modelos de panel de datos y regresión discontinua. Para la última hace uso de la variable de asignación del programa: ingreso familiar (línea de pobreza). Los datos provienen de encuestas longitudinales a aproximadamente 15 000 niños	<ul style="list-style-type: none"> <li>• Los almuerzos escolares incrementan en los participantes del programa tanto el peso como la probabilidad de ser calificados como obesos.</li> <li>• Los almuerzos escolares solo explican una pequeña parte de la tasa de obesidad</li> </ul>

Estudio (I)	Descripción de la intervención (II)	Tratamiento y Variables de Interés (III)	Población de Referencia (IV)	Técnicas Utilizadas y Datos (V)	Principales Hallazgos (VI)
				en etapa escolar.	general.
Alderman et al (2008), programa de alimentación escolar en el norte de Uganda.	<ul style="list-style-type: none"> <li>• El programa de comida por educación provee a de un snack fortificado a mitad de mañana y almuerzos escolares. El snack consiste en una papilla enriquecida con soya y maíz, azúcar y agua. El almuerzo consiste en frijoles con arroz o harina de maíz; además, incluye aceite vegetal y sal. Entre ambos, ofrecen 1049 kcal/día, 32.6gm de proteína y 24.9 gm de grasa.</li> <li>• El programa de raciones para llevar a casa otorga alimentos que son casi sustitutos de las comidas disponibles en el hogar (granos, aceite, frijoles, maíz fortificado).</li> </ul>	Comparación del impacto de dos programas, uno de alimentación escolar y otro de raciones para llevar a casa. Se examinan cómo los outcomes (tasa de matrícula, asistencia, edad en que se inicia la etapa escolar, continuidad con la educación secundaria y tasa de repitencia) son afectados por el timing de los almuerzos y la colocación de incentivos a los niños frente a los padres.	<ul style="list-style-type: none"> <li>• La población estudiada es la de alumnos que cursan la educación primaria.</li> </ul>	Análisis prospectivo con un experimento aleatorio. La data proviene de encuestas.	<ul style="list-style-type: none"> <li>• Efecto positivo en las tasas de asistencia, la magnitud varía dependiendo del grado y género analizado (entre 9 y 30 puntos porcentuales).</li> <li>• Solo el programa de alimentación escolar (SFP) reduce la edad de inicio escolar en el caso de las chicas.</li> <li>• Ninguno de los dos programas afecta la progresión a los estudios de educación secundaria.</li> </ul>
Ahmed et al (2006), programa "Alimentos por Educación" en Bangladesh.	<ul style="list-style-type: none"> <li>• El programa brinda raciones de alimentos a los hogares de 15-20 kg de trigo o 12-16 kg de arroz mensual dependiendo del número de niños que fuesen enviados a la escuela primaria, quienes debían asistir al menos al 85% de las clases.</li> </ul>	Impacto del programa de alimentación sobre indicadores de logro estudiantil (test scores en pruebas de matemáticas, conciencia ambiental, inglés y Bangla) a través del tamaño de clase.	<ul style="list-style-type: none"> <li>• Se analiza el efecto del programa en niños en edad de educación primaria (6 a 10 años) pertenecientes a hogares pobres.</li> </ul>	Análisis multivariado (modelo Tobit). La data proviene de encuestas a nivel de escuelas y hogares.	<ul style="list-style-type: none"> <li>• Efectos negativos en los test scores de los alumnos que no recibieron el programa.</li> <li>• Efecto positivo pero estadísticamente igual a cero en los test scores de matemáticas y bangla de los alumnos que recibieron el programa.</li> </ul>

Estudio (I)	Descripción de la intervención (II)	Tratamiento y Variables de Interés (III)	Población de Referencia (IV)	Técnicas Utilizadas y Datos (V)	Principales Hallazgos (VI)
Vermeersch et al (2004), programa de desayunos escolares en Kenya.	<ul style="list-style-type: none"> <li>El programa provee desayunos escolares. Este consiste en un potaje hecho de azúcar, aceite de maíz y agua. Cada 500 ml brinda 422 calorías. Este desayuno resulta más nutritivo al que reciben los niños típicamente en sus hogares.</li> </ul>	Impacto del programa de desayunos escolares en los indicadores de participación escolar, indicadores de talla y peso, habilidades cognitivas, aprendizaje y tasa de ausentismo del profesor.	<ul style="list-style-type: none"> <li>Se evalúa el impacto en los niños del preescolar (4 a 6 años). Este estudio se concentra en escuelas preescolares informales de áreas rurales.</li> </ul>	Experimento aleatorio. 25 escuelas elegidas aleatoriamente de un total de 50 recibieron el programa.	<ul style="list-style-type: none"> <li>Efecto en la participación escolar tanto en aquellos alumnos que ya habían participado anteriormente como en aquellos que no hubieran participado en ausencia del programa. El efecto fue de 30%.</li> <li>No efecto en las habilidades cognitivas. Solo hubo efecto positivo en aquellos colegios con profesores bastante experimentados previamente al programa.</li> <li>Efecto positivo en el peso de los niños, pero no hubo efecto en los indicadores de peso y talla en las niñas ni en el de talla en niños.</li> <li>No efecto en la tasa de ausentismo de los profesores.</li> </ul>
Timothy et al. (2010), programas de alimentación escolar (WFP) en Kenya.	<ul style="list-style-type: none"> <li>El programa provee almuerzos escolares cuyos ingredientes varían dependiendo de la disponibilidad de alimentos; sin embargo, el patrón consiste en 150 gr de cereal, 40 gr de legumbres y 5 gr de aceite vegetal. Esta ración provee alrededor de 700 kcal/día y 13.5 de proteínas.</li> <li>En algunas escuelas, los alumnos de educación inicial recibieron potajes</li> </ul>	Impacto del programa en indicadores de salud, nutrición y educación (test scores y la proporción de alumnos con un desempeño por encima del promedio).	<ul style="list-style-type: none"> <li>La población analizada es la de alumnos cursando la educación primaria o inicial.</li> </ul>	Comparación de indicadores obtenidos de un muestreo aleatorio en el grupo de control y tratamiento. La data es administrativa.	<ul style="list-style-type: none"> <li>Se encuentra un efecto positivo del programa en el indicador de aprendizaje: 11 puntos de diferencia en los scores entre tratados y controles.</li> <li>La proporción de alumnos con desempeño por encima del promedio aumentó en 7</li> </ul>

Estudio (I)	Descripción de la intervención (II)	Tratamiento y Variables de Interés (III)	Población de Referencia (IV)	Técnicas Utilizadas y Datos (V)	Principales Hallazgos (VI)
	fortalecidos con micronutrientes.				puntos porcentuales.
Nielsen et al. (2010), programas de alimentación escolar (WFP) en Cambodia.	<ul style="list-style-type: none"> <li>Hubieron dos programas: 1) programa de almuerzo escolar y 2) programa de raciones para llevar a casa.</li> </ul>	Impacto de los programas en indicadores de salud y educación.	<ul style="list-style-type: none"> <li>Se analizan los resultados de los alumnos de educación primaria.</li> </ul>	Metodologías mixtas, entre ellas la de panel de datos. La data proviene de encuestas a nivel de hogares y escuelas, y entrevistas.	<ul style="list-style-type: none"> <li>Efecto positivo en la tasa de matrícula (2-2.5%). El efecto es mayor para el caso de las niñas.</li> <li>Reduce la tasa de deserción escolar entre 1.8-2.7%.</li> <li>No hubo efecto significativamente distinto de cero en las pruebas estandarizadas de matemáticas y lenguaje (escritura y comprensión).</li> </ul>
Ahmed et al. (2002), programa de comida por educación en Bangladesh.	<ul style="list-style-type: none"> <li>El programa brinda raciones de alimentos a los hogares de 15-20 kg de trigo o 12-16 kg de arroz mensual dependiendo del número de niños que fuesen enviados a la escuela primaria, quienes debían asistir al menos al 85% de las clases.</li> </ul>	Impacto del programa en indicadores educativos de matrícula, asistencia, deserción escolar y calidad educativa.	<ul style="list-style-type: none"> <li>Se analiza el efecto del programa en niños en edad de educación primaria (6 a 10 años) pertenecientes a hogares pobres.</li> </ul>	Uso de diversos métodos cuantitativos y cualitativos. La data proviene de encuestas realizadas a nivel de hogares y escuelas.	<ul style="list-style-type: none"> <li>Efecto positivo en la tasa de matrícula escolar y asistencia. El efecto en la matrícula fue mayor para las niñas que los niños.</li> <li>Disminución en la tasa de deserción escolar.</li> <li>No resultados concluyentes en los indicadores de calidad educativa.</li> </ul>
Afridi (2010), programa de alimentación en la India.	<ul style="list-style-type: none"> <li>El programa ofrece comidas gratuitas que consisten en 100 gr de trigo o arroz cocido dependiendo de la zona a los alumnos de educación primaria de las escuelas públicas.</li> </ul>	Impacto del programa de alimentación sobre el consumo diario de nutrientes (calorías, carbohidratos, proteínas, calcio y	<ul style="list-style-type: none"> <li>La muestra para el estudio estuvo conformada por alumnos de educación primaria de las escuelas públicas de</li> </ul>	Estos datos son recogidos mediante encuestas aleatorias en las que se recoge información sobre lo consumido y las actividades llevadas a cabo.	<ul style="list-style-type: none"> <li>El consumo diario de nutrientes se incrementa en 49-100% de las transferencias.</li> </ul>

Estudio (I)	Descripción de la intervención (II)	Tratamiento y Variables de Interés (III)	Población de Referencia (IV)	Técnicas Utilizadas y Datos (V)	Principales Hallazgos (VI)
		hierro).	zonas rurales (mayor pobreza).		
Hinrichs (2010), programa nacional de almuerzos escolares en Estados Unidos.	<ul style="list-style-type: none"> <li>El programa hace entrega de dinero a los gobiernos federales. A los colegios beneficiarios (estos deciden si participar o no) se les otorga este dinero y ayudas en mercancías y se les pide cumplir con determinados requerimientos en el almuerzo que ofrecen (tres categorías de almuerzo con diferentes requerimientos).</li> </ul>	Impacto del programa de almuerzo escolar en indicadores como la altura, índice de masa corporal y años de educación alcanzados.	<ul style="list-style-type: none"> <li>La población analizada es la de alumnos en el rango de edad de 5 a 17 años.</li> </ul>	El método empleado es el de variables instrumentales. La data utilizada es de tipo administrativo, de encuestas de hogares y del censo.	<ul style="list-style-type: none"> <li>No se encuentran efectos de largo plazo en cuanto a salud se refiere (indicador de talla).</li> <li>No efecto en el indicador de peso.</li> <li>Un incremento en la exposición al programa de 10% está relacionado con un incremento de 0.365 en los años de educación alcanzados en las mujeres y de un año en el caso de los varones. Este efecto encontrado es impreciso y no robusto.</li> </ul>
Meng y Ryan (2010), programa "Alimentos por Educación" en Bangladesh.	<ul style="list-style-type: none"> <li>El programa brinda raciones de alimentos a los hogares de 15-20 kg de trigo o 12-16 kg de arroz mensual dependiendo del número de niños que fuesen enviados a la escuela primaria, quienes debían asistir al menos al 85% de las clases.</li> </ul>	Impacto del programa alimentario en los resultados de participación escolar y años de educación alcanzados.	<ul style="list-style-type: none"> <li>Se analiza el efecto del programa en niños en edad de educación primaria (6 a 10 años) pertenecientes a hogares pobres.</li> </ul>	La metodología empleada es la de propensity score matching combinado con diferencias en diferencias. Se hace de uso de la encuesta de hogares.	<ul style="list-style-type: none"> <li>Efecto entre 15-26% en la tasa de participación escolar.</li> <li>Efecto positivo del programa en el indicador de años de educación alcanzados, entre 0.7-1.5 años más condicional a la participación escolar.</li> </ul>
Cueto y Chinen (2001), programa de desayunos	<ul style="list-style-type: none"> <li>El desayuno escolar brindado por el programa consiste en una bebida similar a la leche pero sin lactosa y 6 galletas pequeñas. Este brinda 600kcal/día, 22.5 gramos de</li> </ul>	Impacto del programa de desayunos escolares en indicadores de peso y talla, tasa de	<ul style="list-style-type: none"> <li>Se analizan los efectos del programa en alumnos de cuarto grado (590) de</li> </ul>	Se aplicaron varias metodologías dependiendo del output analizado: estadísticas descriptivas, análisis de	<ul style="list-style-type: none"> <li>Se encontró efecto positivo en los indicadores de memoria de corto plazo, hemoglobina, tasa de</li> </ul>

Estudio (I)	Descripción de la intervención (II)	Tratamiento y Variables de Interés (III)	Población de Referencia (IV)	Técnicas Utilizadas y Datos (V)	Principales Hallazgos (VI)
escolares en Perú	proteínas y 20 gramos de grasa.	matrícula, deserción escolar, asistencia, aptitudes cognoscitivas y rendimiento en pruebas estandarizadas de matemáticas y comprensión lectora.	primaria de escuelas estatales del ámbito rural de zonas altoandinas (mayor pobreza).	covarianzas y modelo jerárquico lineal. La data utilizada es recogida directamente para los fines de la evaluación y también de tipo administrativo.	deserción escolar, y asistencia. • No hubo efecto en los indicadores de talla, masa corporal ni rendimiento en las pruebas estandarizadas.
Van Stuijvenberg et al. (1998), Programa de Alimentación Escolar-Sudáfrica.	• El programa escolar provisionó galletas fortificadas con hierro, yodo y b-caroteno.	Impacto del programa de alimentación escolar sobre indicadores antropométricos, de función cognitiva y morbilidad.	• Niños de entre 6 y 11 años de escuelas primarias de comunidades rurales de bajo nivel económico.	Experimento aleatorio. 115 niños fueron asignados al grupo de tratamiento, y 113, al grupo de control.	• No se encontró efecto en ningún indicador de peso y talla. • Efecto positivo en la prueba de retención de dígitos. • Reducción en los indicadores de morbilidad.
Chandler et al. (1995), Desayunos Escolares – Jamaica.	• El programa otorgó desayunos escolares a los alumnos a ser tratados y una naranja (placebo) a los alumnos del grupo de control.	Impacto de los desayunos otorgados a los alumnos sobre los scores de 4 test cognitivos: fluidez verbal, retención de dígitos, búsqueda visual y un test de velocidad de procesamiento de información.	• Niños de tercer y cuarto grado de primaria de escuelas rurales.	Experimento aleatorio. De un grupo de 197 (97 niños malnutridos y 100 con buen estado nutricional) alumnos de 4 escuelas primaria se eligió aleatoriamente quiénes recibirían el tratamiento y quiénes no.	• Solo hubo efecto positivo en la prueba de fluidez verbal. • Se encontró que existe una interacción muy significativa entre los grupos nutricionales y el tratamiento.
Simeon (1998), Programa de Alimentación Escolar –	• En el primer estudio una clase recibió un almuerzo escolar estándar mientras que otras dos sirvieron como control. • Se testeó a los alumnos durante varias	Se analizan 2 programas de alimentación escolar. El primero examina el	• El primer estudio tiene como población de referencia a niños de 12-13 años de	Experimento aleatorio.	• En el primer estudio se encontró efectos positivos en los indicadores de asistencia y de desempeño en los test de

Estudio (I)	Descripción de la intervención (II)	Tratamiento y Variables de Interés (III)	Población de Referencia (IV)	Técnicas Utilizadas y Datos (V)	Principales Hallazgos (VI)
Jamaica.	mañanas habiendo recibido el desayuno escolar y no.	<p>efecto de los almuerzos escolares en indicadores de logro académico, asistencia y crecimiento.</p> <p>El segundo estudio analiza el impacto de la omisión del desayuno en indicadores cognitivos.</p>	<p>escuelas rurales pobres.</p> <ul style="list-style-type: none"> <li>• La población de referencia del segundo estudio son los niños de 9-10 años.</li> </ul>		<p>matemáticas. No hubo efecto en la talla.</p> <ul style="list-style-type: none"> <li>• En el segundo estudio, se concluye que el alivio del hambre es un mecanismo a través del cual los programas de alimentación escolar mejoran el rendimiento académico.</li> </ul>
Whaley et al. (2003), Programa de Alimentación Escolar – Kenia.	<ul style="list-style-type: none"> <li>• Hubieron tres tipos de tratamientos los cuales suplementaban el almuerzo tradicional "githeri" (compuesto por maíz y frijoles) con carne, leche o energía (aceites adicionales). Estos tres suplementos ofrecen aproximadamente 240 kcal.</li> <li>• El grupo de control no recibió ningún alimento ni fuente de energía.</li> </ul>	Este estudio analiza el impacto de tres dietas diferentes sobre indicadores de desarrollo cognitivo.	<ul style="list-style-type: none"> <li>• La población de referencia son los niños de primer grado de primaria de escuelas rurales pobres.</li> </ul>	<p>Experimento aleatorio. 12 escuelas (con 555 estudiantes en total) recibieron aleatoriamente una de las cuatro intervenciones alimentarias. La selección de las escuelas estuvo basada en su tamaño y localización.</p> <p>Tras la aleatorización, 134 niños fueron designados a recibir carne, 114 a recibir leche, 148 a recibir otra fuente de energía y 129 al grupo de control.</p> <p>Además, se hizo uso de un modelo lineal jerárquico con efectos aleatorios para examinar los efectos del tratamiento a través del tiempo.</p>	<ul style="list-style-type: none"> <li>• Se encontró que los niños que recibieron comida suplemental con carne tuvieron un mejor desempeño en la prueba de Matrices Progresivas de Raven.</li> <li>• Los niños con alimentación suplementaria de leche y otras fuentes de energía se desarrollaron mejor que el grupo control en pruebas de aritmética.</li> <li>• No efecto en ninguno de los grupos en los test de comprensión lectora.</li> </ul>

Estudio (I)	Descripción de la intervención (II)	Tratamiento y Variables de Interés (III)	Población de Referencia (IV)	Técnicas Utilizadas y Datos (V)	Principales Hallazgos (VI)
Simeon et al. (1989) – Jamaica.	<ul style="list-style-type: none"> <li>El tratamiento consistió en otorgar una cena estándar (pollo, pan y papas fritas: 940 kcal aprox.) durante la noche. Además, a la mañana siguiente se proveyó un desayuno estándar con 590 kcal aprox. o una taza de té. El desayuno se otorgó durante la primera visita y solo una taza con té en la segunda, o viceversa.</li> </ul>	<p>En el estudio se examina el efecto de la omisión del desayuno sobre indicadores cognitivos en tres grupos: niños con buena nutrición, desnutridos y otro grupo que previamente ha tenido desnutrición severa.</p>	<ul style="list-style-type: none"> <li>Niños en el rango de edad de 9 y 10,5 años.</li> </ul>	<p>Experimento aleatorio. Los niños para cada grupo fueron identificados usando el registro de hospitales y de escuelas primaria ubicadas en áreas pobres.</p> <p>La data proviene de encuestas realizadas por los autores.</p>	<ul style="list-style-type: none"> <li>En general, se encuentra que las funciones cognitivas son vulnerables a la omisión de desayuno.</li> <li>Se encontró que el grupo con desnutrición tuvieron menor desempeño en las pruebas de fluidez y codificación.</li> <li>El grupo control mostró mejores resultados cuando no tuvieron desayuno mientras que en los otros dos grupos combinados no hubo efecto.</li> <li>No hubo efectos significativos en los test de MFFT hard-items ni en el de HCI.</li> </ul>
Mahoney et al. (2005) – EE.UU.	<ul style="list-style-type: none"> <li>El grupo tratado recibe un tipo de desayuno: avena instantánea o cereales preparados. Ambos ofrecen aproximadamente la misma cantidad de energía; sin embargo, difieren en la composición de los macronutrientes, características del proceso, efectos en la digestión y metabolismo, y niveles de glucosa.</li> </ul>	<p>Este estudio analiza el efecto de la composición del desayuno sobre indicadores de desarrollo cognitivo.</p>	<ul style="list-style-type: none"> <li>El primer experimento tuvo como población objetivo a niños de entre 9 y 11 años de edad mientras que el segundo estuvo enfocado en niños de entre 6 y 8 años.</li> <li>Los niños, para ambos experimentos, son de clase media de escuelas católicas privadas.</li> </ul>	<p>Se realizaron dos experimentos, cada uno realizó estudios sobre 30 niños (15 niños y 15 niñas).</p> <p>La data es de fuente primaria.</p>	<ul style="list-style-type: none"> <li>Se encuentra que el desayuno favorece el desempeño cognitivo, particularmente en tareas que requieren de procesamiento visual complejo.</li> <li>Se encuentran efectos diferenciales dependiendo del tipo de desayuno suministrado.</li> <li>Tanto niñas y niños (9-11 años) mejoraron su memoria espacial, y las niñas</li> </ul>

Estudio (I)	Descripción de la intervención (II)	Tratamiento y Variables de Interés (III)	Población de Referencia (IV)	Técnicas Utilizadas y Datos (V)	Principales Hallazgos (VI)
					<p>mejoraron su memoria de corto plazo tras consumir avena.</p> <ul style="list-style-type: none"> <li>• Los niños y niñas (6-8 años) mostraron mejores resultados en test de memoria espacial y atención; las niñas mejoraron en pruebas de memoria de corto plazo.</li> </ul>

## **APÉNDICE V: MEDICIÓN DE LA VARIABLE DE ATENCIÓN Y MEMORIA DE CORTO PLAZO**

La habilidad cognitiva, en particular los aspectos relacionados con dimensiones de atención y memoria, han sido tratados de manera recurrente en la literatura de evaluación de programas de alimentación. Esto se debe, en gran medida, al papel esencial que juegan estos aspectos en la adquisición de nuevos conocimientos. En línea con lo mencionado, la literatura sobre esta temática ha puesto especial atención en el desarrollo e implementación de una serie de instrumentos para medir las habilidades cognitivas, tanto de corto como de largo plazo.

En lo que respecta a las habilidades cognitivas de largo plazo, un aspecto que ha tomado preponderancia en la literatura es la memoria de largo plazo. Esta ha sido recogida mediante pruebas de fluidez verbal que exigen al niño recordar palabras que han aprendido durante toda su vida. Una prueba bastante conocida al respecto es el *Peabody Picture Vocabulary Test*, el cual consiste en presentar al niño una serie de imágenes de diferente complejidad que debe proceder a nombrar (Whaley et al, 2003); otra prueba que mide este aspecto consiste en pedir al niño mencionar todas las palabras que conozca sobre un tema específico (Chandler et al, 1995).

Por otra parte, dentro de las habilidades cognitivas de corto plazo, reciben especial atención dentro de la literatura las pruebas y/o instrumentos que brindan actividades o tareas que requieran de memoria y velocidad de procesamiento de información, así como de cierto nivel de atención puesto para su adecuada realización.

Entre los instrumentos utilizados para recoger indicadores de atención y vigilancia se encuentran el Test de Hagen, el test de cancelación, el test de búsqueda visual de ciertos patrones y otros test de atención auditiva que consisten en identificar una determinada combinación de palabras de una serie de estas (Simeon, 1998; Mahoney et al, 2005; Dickie et al, 1982; Chandler et al, 1995). Por otro lado, con la finalidad de recoger indicadores de velocidad de procesamiento de información, se han utilizado pruebas como la de verificación de oraciones en la que se le presenta al niño una serie de proposiciones sobre las cuales debe responder si son verdaderas o falsas (Dickie et al, 1982), y pruebas que ofrecen pares de cifras para comparar en magnitud (Chandler et al, 1995) pues ambas exigen que el niño razone rápidamente.

Respecto a las pruebas de memoria de corto plazo, resaltan las pruebas de retención de dígitos hacia adelante y hacia atrás que miden específicamente la memoria auditiva. Además, están las pruebas de codificación, de memoria espacial y de percepción visual como the Rey Complex Figure Copy and Recall Test que consiste en mostrar una figura y pedir que se la dibuje según lo que se recuerde de ella con el mayor detalle posible tanto en ese momento como tras varios minutos después (Van Stuijvenberg et al, 1998; Simeon, 1998; Simeon et al, 1989; Mahoney et al, 2005; Chandler et al, 1995). También resaltan las pruebas de memoria verbal que consisten en otorgar una historia escrita para luego pedir que sea narrada con el máximo detalle que se recuerde (Mahoney et al, 2005).

Además de estas pruebas, existen otros instrumentos que permiten medir otras habilidades cognitivas en los niños. Entre estos, la prueba de matrices de Raven que busca medir la habilidad del

niño para organizar los detalles de percepción, y para razonar por analogías y comparación de formas (Whaley et al, 2003; Vermeersch et al, 2004); pruebas de aritmética (para resolver mentalmente o no) cuya complejidad varía de acuerdo a la edad del público objetivo (Whaley et al, 2003; Simeon, 1998; Dickie et al, 1982); pruebas de comprensión oral (Simeon, 1998); pruebas de emparejamiento de figuras similares que permiten medir la impulsividad y la eficiencia para resolver problemas (Simeon, 1998); entre otros.

## Instrumentos sobre Atención y Memoria utilizados en Programas de Alimentación Escolar

Título	autor(es) / año	Pruebas Utilizadas	Descripción de las Pruebas		Nivel educativo/edad de niños evaluados
School Meals, Educational Achievement and School Competition: Evidence from a Randomized Evaluation	Vermeersch et al. (2004)	<i>Oral curriculum testing</i> (10 min)	Evaluó la fluidez verbal, la habilidad para reconocer similitudes entre figuras, habilidad en la Matriz de Raven, habilidades de patrones de construcción, y memoria usando la técnica de los dígitos (hacia adelante).		preescolar (4-6 años)
		Test cognitivo oral (10 min)			
		<i>Written curriculum testing</i> (40 min)	Estuvo compuesta por dos secciones: 1) escribir el alfabeto, escribir los números del 1 al 20; 2) rellenar las letras que faltan en palabras, anote el nombre de un objeto representado en un cuadro, responder a preguntas sobre el contenido de una frase y realizar operaciones aritméticas básicas.		
Effect of iron-, iodine-, and b-carotene-fortified biscuits on the micronutrient status of primary school children: a randomized controlled trial	Van Stuijvenberg et al. (1998)	Copia de dígitos, contar letras, cancelar letras, leer números, contar hacia atrás.	Los test fueron realizados fuera de clase. Los test estuvieron diseñados para obtener la velocidad de procesamiento y capacidad de llevar a cabo actividades relacionadas con las habilidades requeridas por la escuela que requerían de memoria.	Se midió el tiempo que tomo al niño completar la tarea	niños en edad de educación primaria (segundo a cuarto grado)

		Pruebas de fluidez verbal, escribir cruces, retención de dígitos hacia adelante y hacia atrás.		se consideró el número de tareas completadas en determinado tiempo	
The Impact of Dietary Intervention on the Cognitive Development of Kenyan School Children	Whaley et al. (2003)	The Raven's Colored Progressive Matrices	Toda la prueba tiene un tiempo promedio de 30 minutos	Se le presentó al niño una matriz de símbolos y tuvo que completar la matriz seleccionando el símbolo correcto faltante de un grupo de símbolos. Este test mide la habilidad del niño para organizar los detalles de percepción y para razonar por analogías y comparación de formas.	Niños de primer grado
		The verbal meaning Test (Peabody Picture Vocabulary test)		Se le presentó al niño cuatro imágenes y se le pidió nombrarlo. En total habían 36 objetos de diferentes complejidades: palabras comunes y otras palabras que requerían de cierto conocimiento de conceptos abstractos.	
		Test de aritmética (adaptado del test revisado de Wechsler Intelligence Scale for Children)		El test constaba de 19 ejercicios. Los primeros ejercicios eran de suma y resta, y los siguientes de división, multiplicación y decimales.	
		Ejercicios mentales de aritmética	Subtests del Wechsler Intelligence Scale for Children revisado	-	
		Retención de dígitos		Busca medir la memoria auditiva de corto plazo	
		Ejercicios de codificación		Busca medir la memoria visual de corto plazo	
		Fluidez verbal	Obtenidos de The Clinical Evaluation of Language Functions	Busca medir la memoria de largo plazo	
		Test de comprensión oral		-	

School feeding in Jamaica: a review of its evaluation	Simeon (1998)	Match de figuras familiares	Test reconocidos por ser sensibles a los efectos de corto plazo de la privación de comida.	Mide la impulsividad y la eficiencia para resolver problemas. El puntaje de eficiencia fue obtenido tras sumar los z-scores del número de errores y el tiempo de latencia antes de responder. Un puntaje menor es señal de mayor eficiencia, por ejemplo, menos errores y tiempo de respuesta mayor.	Niños de 9 y 10 años
		Hagen's Central Incidental Task		Mide la memoria visual de corto plazo y el nivel de atención	
		The Peabody Picture Vocabulary test	Utilizado para medir el coeficiente intelectual del niño.		
School feeding in Jamaica: a review of its evaluation	Simeon et al. (1989)	Prueba de aritmético	Subtests del Wechsler Intelligence Scale for Children revisado. Estos son elegidos por ser afectados por el nivel de atención puesto y la distracción.	El puntaje fue el número de respuestas correctas.	Niños de 9 y 10 años
		Retención de dígitos (hacia adelante y hacia tras)		la tarea consiste en sustituir símbolos por números tan rápido como sea posible. El puntaje es el número de símbolos correctos escritos en un tiempo determinado. Mide la memoria visual de corto plazo	
		Ejercicios de codificación			

Effect of breakfast composition on cognitive processes in elementary school	Mahoney et	Test de memoria espacial	Se crearon tres mapas (para controlar por la exposición previa) Cada mapa consistió en 24 países de los 4 continentes. Los nombres para los países se eligieron de 3 categorías: naturaleza, animales, colores. Durante el ejercicio, los nombres de los países aparecieron en la pantalla uno a la vez. Los participantes pasaban de a uno en uno los nombres mediante un teclado. Una vez que el participante paso por todos ellos, se le daba 8 minutos para que los estudie. Pasado este tiempo, el niño recibe un mapa en blanco y rellena los espacios vacíos de los paisas tanto como su memoria se lo permita.	Niños en edad de educación primaria
		Retención de dígitos	Se les hizo escuchar grupos de números y repetirlos. Tras haberlo repetido correctamente, el siguiente set de números aumenta en uno. Si el participante no repite el número correctamente, se le da otra oportunidad con otro conjunto de la misma cantidad de números. Dos intentos fallidos y el test termina. Las repeticiones eran para hacerlas hacia adelante y hacia atrás (a la inversa del orden de mención).	
		Test de percepción visual: The Rey Complex Figure Copy and Recall Test	El niño recibe una de las tres figuras de igual complejidad y una hoja en blanco. Se les pide copiar la figura de manera tan exacta como sea posible. Se les vuelve a pedir que dibujen la figura de su memoria en una hoja en blanco pasados 20 y 50 minutos.	
		Prueba de atención visual	Los niños miran una serie de letras a través de una pantalla a una velocidad de una por segundo durante diez minutos. Se le dio al niño buscar una combinación en especial (por ejemplo, "x" inmediatamente después de "B"), al encontrarla debía presionar la barra espaciadora del teclado.	

children	al. (2005)	Prueba de atención auditiva	Los niños escuchaban palabras a una velocidad de una por segundo mediante auriculares conectados a una PC. Se les pidió que tras escuchar determinada combinación de palabras (por ejemplo, "ratón" seguido por "casa") presionaran la barra espaciadora.	Niños en edad de educación primaria
		Prueba de memoria verbal	La prueba consistía en dos historias de un párrafo cada una que contenían entre 15 y 20 líneas. Los participantes tenían 5 minutos para leer las historias tantas veces como quisiesen, luego debían relatar la historia con el mayor detalle posible.	
Breakfast and performance in schoolchildren	Dickie et al. (1982)	Test de cancelación (3 min.)	Diseñado para medir el grado de atención y vigilancia. Consistía en 4 páginas de letras compuestas por letras aleatorias, se les pedía a los niños tachar los casos en que apareciesen dos letras idénticas juntas. El puntaje estaba en función del número de aciertos (expresado como porcentaje del total de casos presentados en la hoja) y la velocidad.	niños en etapa escolar
		Test de sumas (5 min.)	Se les pedía los niños que resuelvan ejercicios de sumas de 5 números de 2 dígitos cada uno. La hoja de ejercicios contenía 70 ejercicios de sumas.	
		Test de memoria y búsqueda (2 min.)	Se le pide buscar entre una serie de líneas de 20 letras aleatorias cada una un conjunto de letras ( 4 o 6 letras)	
		Test de verificación de oraciones	Se les presenta a los niños una lista de proposiciones que describen como están ubicadas dos símbolos: + y *. La tarea asignada al niño es la de señalar si estas proposiciones son verdaderas o falsas. Este test estaba compuesto por 5 secciones, cada una con un ordenamiento aleatorio de los símbolos de 64 items.	

School Breakfast Improves Verbal Fluency in undernourished Jamaican Children	Chandler et al. (1995)	Test de búsqueda visual (45 segundos por página)	Este test mide la velocidad de procesar información visual, la cual refleja la capacidad de atención puesta. Consiste en tener páginas de 8 líneas con 30 letras aleatorias del alfabeto. En el extremo izquierdo se ponen las letras que deben ser encontradas en cada oración. En las primeras 3 hojas de ejercicios se pide encontrar una letra, en las 3 siguientes se pide encontrar 2. El puntaje obtenido es el número de aciertos menos el número de errores (sin marcar o marcados incorrectamente)	niños de segundo a cuarto grado
		Retención de dígitos	Consiste en que el niño intente recordar dígitos cada vez de mayor tamaño. Se le otorga un punto por cada intento correcto	
		Test de Fluidez Verbal	Es una prueba para la memoria de largo plazo. El niño tenía que mencionar tantas palabras como recordase de un determinado ítem: animales o cosas para comer. Un punto por cada palabra	
		Test de velocidad de proceso de información	Se le presento al niño dos dígitos de los cuales debía comparar cuál de ellos era mayor. Cada línea de la hoja estuvo compuesta por 8 pares a comparar	

## **APÉNDICE VI: TEMAS DE INVESTIGACION DE INSTRUMENTOS DE RECOLECCION**

### **1. Cuestionario de Hogar**

#### **1.1. Características de los miembros del hogar**

- Identificación individual
- Número de orden de los miembros del hogar
- Identificación de los miembros de hogar
- Sexo
- Identidad de rasgo
- Miembros en ausencia en el hogar
- Relación de parentesco
- Fecha de nacimiento
- Edad

#### **1.2. Educación**

- Alcance del informante
- Identificación individual
- Nivel de alfabetismo
- Asistencia a un Centro de Enseñanza
- Razón por la que no asiste a un Centro de Enseñanza
- Año o grado de estudios y nivel alcanzado
- Edad de Iniciativa a la educación primaria
- Matriculado en algún centro o programa de educación básica en este año
- Razón por la que no asiste a un centro o programa de educación básica
- Edad del matriculado en algún centro o programa de educación básica por última vez
- Asistencia a un centro o programa de educación básica el mes pasado
- Razón por la que no asistió a un centro o programa de educación básica el mes pasado
- Tiempo que le toma ir al centro educativo
- Incentivo para que asista a un centro o programa de educación básica este año
- Resultados que obtuvo en sus grados de educación
- Frecuencia de desaprobación de año o grado de estudio
- Razón por la que no aprobó de año o grado de estudio en la educación primaria

#### **1.3. Empleo**

- Identificación individual
- Tenencia de trabajo en actividad agrícola dentro del hogar el mes pasado
- Disposición de trabajo en actividad agrícola
- Tenencia de trabajo en actividad agrícola fuera del hogar
- Actividad principal el mes pasado
- Total Meses o Días Trabajados en el último año
- Máximo y Mínimo de Horas Trabajadas en días laborales
- Total Horas Trabajadas en días laborales
- Máximo y Mínimo de Horas Trabajadas en no días laborales
- Total Horas Trabajadas en no días laborales

- Trabaja ambos días no laborales
- Horario de trabajo
- Categoría Ocupacional

#### **1.4. Actividades no Agrícolas**

- Tenencia de algún negocio propio en el hogar el último año
- Tipo de negocio
- Trabajadores encargados de las operaciones y finanzas del negocio
- Duración del negocio en el año pasado o mes pasado
- Miembros del hogar en el negocio
- Total ingreso bruto del negocio el mes pasado

#### **1.5. Ingresos**

- Ingreso bruto de miembros del hogar el mes pasado
- Contribución y aporte del ingreso al hogar
- Total de otros ingresos (Juntos, Pensión 65, otros)
- Total ingreso por transferencia procedente del exterior
- Ingreso por alquiler de tierras agrícolas

#### **1.6. Activos**

- Tenencia de artefactos y/o vehículos del hogar, ganado, bienes inmobiliarios y de granja
- Número de artefactos y/o vehículos del hogar, ganado, bienes inmobiliarios y de granja
- Valor de artefactos y/o vehículos del hogar, ganado, bienes inmobiliarios y de granja

## **2. Cuestionario de Escuela**

### **2.1. Características de la escuela**

- Tipo de escuela
- Distancia de la escuela a la plaza principal
- Niños en el área de influencia de la escuela
- Grados que se enseñan en la escuela
- Escuela atendida por Qali Warma (QW)
- Inicio de recepción de alimentos de QW
- Días que no se recibieron alimentos en el Semestre I y II , 2013
- Periodo de tiempo en que la escuela no tuvo alimentos en el semestre anterior
- Inscripción de la escuela
- Asistencia a la escuela de los beneficiarios

### **2.2. Infraestructura**

- Ambientes en la escuela
- Pupitres y bancas para los estudiantes
- Salón de clase con material predominante: paredes, pisos y techos
- Ambiente donde se almacenan alimentos con material predominante: paredes, pisos y techos

- Ambiente donde se producen alimentos con material predominante: paredes, pisos y techos
- Ambiente donde se distribuyen alimentos con material predominante: paredes, pisos y techos
- Escuela con biblioteca, electricidad y/o fuentes de agua potable
- Escuela con letrinas para profesores, niños y niñas
- Abastecimiento de agua en la escuela
- Tipo de servicio higiénico en la escuela
- Mantenimiento de la servicio higiénico
- Servicios higiénicos con lavaderos o lavatorios

### **2.3. Mobiliario**

- Suficientes pupitres y bancas para los estudiantes
- Adecuado suministro de enseñanza en la escuela
- Proporción de libros de texto para todos los estudiantes
- Tipo, cantidad y estado del mobiliario
- Cantidad de estudiantes con falta de mobiliario

### **2.4. Actividades complementarias**

- Orientación por la escuela el año anterior

### **2.5. Hambre**

- Observación de cambios por el director en el comportamiento en clase, el aprendizaje y la salud de los estudiantes desde que el programa QW comenzó
- La escuela tiene Comités de Alimentación Escolar (CAE)
- Frecuencia de reunión del CAE

### **2.6. Hábitos**

- Cargo en el colegio
- Tiempo, lugar y supervisión en la comida del alumnado
- Tienda, cafetería y/o máquinas expendedoras de alimentos o bebidas en el colegio
- Programas subvencionados o de suscripción para el alumnado
- Tiendas donde el alumnado puede comprar golosinas y/o frutas
- Normativa o política nutricional en la municipalidad o región sobre la alimentación del
  - Alumnado
  - Concepto de la política alimentaria
  - Tiempo que rige la política alimentaria en el colegio
  - Comidas o bebidas prohibidas por la norma

### **2.7. Opinión del profesor**

- Cambios observados en el comportamiento y el aprendizaje en clase de los estudiantes desde que comenzó el programa QW
- Formas de cambio

### **3. Cuestionario de Salud**

#### **3.1. Percepción de la salud y morbilidad**

- Estado general de salud del individuo
- Padecimiento de enfermedades y tipo de ellas

#### **3.2. Accidentalidad y agresión**

- Accidente de cualquier tipo en el último año
- Tipo del accidente y número de veces sufrido
- Lugar del accidente
- Lugar de atención del accidente
- Daños producidos por el accidente
- Agresión física y lugar de la agresión

#### **3.3. Restricción de la actividad**

- Restricción de la actividad por dolores o síntomas de enfermedad
- Días de permanencia en cama por motivos de salud
- Dolores o síntomas que obligaron a quedarse en cama

#### **3.4. Tipos de seguro de salud**

- Tipo de seguro al cual está afiliado

#### **3.5. Atención primaria**

- Tiempo desde que acudió por última vez a una consulta médica
- Número de veces que acudió en el último mes
- Lugar de atención de la última consulta
- Motivo de la consulta
- Tiempo que pasó entre la cita y la consulta con el médico
- Tipo de los profesionales consultados en el último mes
- Tuvo fiebre o tos en las últimas dos semanas
- Lugar de atención de la fiebre o tos
- Tuvo diarrea
- Consumo de líquidos cuando estuvo con diarrea
- Consumo de otras sustancias cuando estuvo con diarrea
- Lugar de atención de la diarrea

#### **3.6. Peso y talla**

- Identificación individual
- Fecha de nacimiento
- Talla
- Peso
- Fecha de peso y talla
- Resultados
- Talla y peso dentro del rango

#### **3.7. Anemia**

- Identificación individual

- Nivel de hemoglobina
- Resultados

#### **4. Cuestionario de Alimentos**

##### **4.1. Consumo de alimentos**

- Consumo de alimentos y bebidas en la mañana después de levantarse después del día anterior
- Olvido del consumo alimentos y bebidas en la mañana después de levantarse después del día anterior
- Horario del consumo
- Descripción detallada de los ingredientes y porción de los alimentos y bebidas consumidos.
- Nivel de lo que se consumió el día anterior
- Consumo de desayuno en la escuela o en otro lugar
- Consumo de alimentos en la escuela el día anterior
- Suficiencia de las raciones –en cantidad- consumidas en el desayuno y almuerzo
- Acción realizada con el alimento no consumido

##### **4.2. Alimentos en la escuela**

- Valoración del día anterior a la encuesta
- Horario en que se levantó
- Actividad realizada durante cada una de sus comidas
- Tuvo desayuno
- Dónde, con quién y qué hacía mientras desayunaba
- Qué consumió en el desayuno: alimento y cantidad
- Comió o bebió algo en la mañana en el colegio
- Qué consumió en la media mañana y a la hora de comer en el colegio
- Tuvo almuerzo
- Dónde, con quién y qué hacía mientras almorzabas
- Qué consumió en el almuerzo: alimento y cantidad
- Qué consumió en la merienda, en la cena y después de cenar
- Tuvo merienda
- Dónde, con quién y qué hacía mientras merendaba
- Qué consumió de merienda: alimento y cantidad
- Tuvo cena
- Dónde, con quién y qué hacía mientras cenaba
- Qué consumió de cena: alimento y cantidad

##### **4.3. Opinión de niños**

- Razón por la que no consume todo el desayuno escolar
- Apreciación sobre el desayuno
- Apreciación sobre el almuerzo

#### **5. Cuestionario de Comunidad**

- Servicios que ofrece la comunidad
- Organismo que financió la obra relacionada al servicio

- Conservación y si funciona el servicio
- Actividad económica más importante de las personas en la comunidad
- Segunda actividad económica más importante de las personas en la comunidad
- Camino transitable en la comunidad
- Hogares con electricidad o generadores
- Instituciones de ahorro y crédito en la comunidad
- Mercado permanente en la comunidad
- Principal medio de transporte público
- Principal medio de riego
- Principales recursos naturales
- Recursos turísticos: tipo, distancia y tipo de acceso

## APÉNDICE VII: ANÁLISIS DE POTENCIA Y CALCULO DE MUESTRA

### 1. Estimación del Efecto Mínimo Detectable

Para el cálculo de muestra se utiliza la propuesta metodológica desarrollada por Teerenstra et al. (2012) para el caso del modelo de diferencias en diferencias con asignación de tratamiento a nivel de cluster. En este modelo de diferencias en diferencias, el efecto de tratamiento es estimado del modo siguiente:

$$(A7.1) \hat{\beta}_{DD} = (\bar{Y}_1^T - \bar{Y}_1^C) - (\bar{Y}_0^T - \bar{Y}_0^C)$$

Un paso fundamental en el cálculo del tamaño de muestra es la correcta especificación de la varianza. Teerenstra et. al. (2012) muestran que la varianza para un diseño de diferencias en diferencias con asignación al tratamiento a nivel de clusters es la siguiente:

$$(A7.2) \text{Var}(\hat{\beta}_{DD}) = 2(1-r) \cdot [1 + (n-1)\rho] \cdot \left( \frac{1}{P} + \frac{1}{1-P} \right) \cdot \frac{\sigma^2}{Jn},$$

donde n es el número de observaciones por clúster,  $\rho$  es el coeficiente de correlación a nivel de cluster y P es la proporción de tratados en la muestra.  $\sigma^2$  es la varianza de la variable de resultado (atención y memoria de corto plazo) y J es el número de clústeres. El parámetro r cumple un rol importante en el diseño y es definido del modo siguiente:

$$(A7.3) r = \frac{n\rho}{1 + (n-1)\rho} \rho_c + \frac{1-\rho}{1 + (n-1)\rho} \rho_s,$$

en donde  $\rho_c$  y  $\rho_s$  son respectivamente los indicadores de autocorrelación a nivel de cluster y a nivel de individuo.

A partir de (A7.2), es posible derivar una expresión para estimar el efecto mínimo detectable, el nivel de potencia o el tamaño de muestra. La fórmula general para el efecto mínimo detectable (EMD) es la siguiente<sup>26</sup>:

$$(A7.4) \text{EMD}_{DD\_cluster} = \underbrace{(t_{(1-k)} + t_\alpha)}_A * \underbrace{\sqrt{\frac{1}{P(1-P)}} \sqrt{\frac{\sigma^2}{nJ}}}_{B} * \underbrace{\sqrt{1 + (n-1)\rho}}_B * \underbrace{\sqrt{2(1-r)}}_C,$$

en donde el EMD es expresado en términos de la fórmula del EMD para el caso de un diseño aleatorio simple con asignación a nivel individual (A), el efecto diseño que permite ajustar la formula en A por la similitud existente en individuos provenientes del mismo clúster (B), y el factor asociado al coeficiente de correlación r (C), el cual captura el rol conjunto del CCI y los coeficientes de autocorrelación a nivel individual y a nivel de cluster.

La ventaja de la formula anterior es que permite interpretar el EMD para el caso de un diseño de diferencias en diferencias a nivel de clúster en términos de expresiones ya conocidas en la literatura, con la excepción del factor asociado al coeficiente de correlación. Esto simplifica el

<sup>26</sup> Nótese que esta formulación no es parte del artículo original. Es una derivación del autor de esta nota a partir de la discusión presentada por los autores del artículo.

proceso de computo del tamaño de muestra puesto que solo es necesario computar  $r$  para estimar la expresión  $C$  en la formula anterior, y multiplicarlo con las expresiones usualmente calculadas en la literatura ( $A$  y  $B$ ) mediante las rutinas disponibles en cualquier software estándar.

A partir de (A7.4), es posible derivar una fórmula para el cálculo del tamaño de muestra (definido en términos de clusters y asumiendo un número fijo de observaciones por cluster):

$$(A7.5) \quad J = (t_{(1-k)} + t_{\alpha})^2 * \frac{1}{P(1-P)} \frac{\sigma^2}{n * ES^2} * [1 + (n-1)\rho] * [2(1-r)],$$

en donde  $ES$  es el tamaño del efecto esperado por la intervención medido en la unidad de referencia correspondiente. A fin de simplificar la lectura, y debido a la ausencia de información de línea de base para los indicadores de referencia, la formula puede expresarse en términos de desviaciones estándar a partir de una simple normalización que surge de la división de (A7.5) por la desviación estándar de la variable de resultado de interés. Esta normalización es sumamente útil para hacer comparable la intervención bajo análisis con la literatura internacional sobre el impacto de la alimentación escolar.

En suma, el proceso de cálculo de muestra para un diseño de diferencias en diferencias con asignación al tratamiento a nivel de clúster consiste en el desarrollo de las siguientes estimaciones:

Paso 1: Calculo del tamaño de muestra para un diseño de asignación aleatoria simple con información a ser analizada durante el seguimiento.

Paso 2: Calculo del efecto diseño mediante el CCI y obtención del tamaño de muestra para un diseño de asignación aleatoria a nivel de cluster.

Paso 3: Calculo del coeficiente de correlación  $r$  y obtención del tamaño de muestra para un diseño de diferencias en diferencias (o comparación de línea de base y seguimiento) con asignación a nivel de cluster.

## 2. Ajustes por Comparaciones Múltiples

El problema de comparaciones múltiples ocurre cuando se considera un conjunto de hipótesis respecto a significancia estadística de una intervención de manera simultánea. En el contexto de *Qali Warma*, la comparación no sería solamente entre distritos tratados y controles, sino también entre distritos tratados que reciben productos y aquellos que reciben la canasta básica. En este escenario, el número de comparaciones posibles serían 3 en vez del caso simple en donde solo una comparación (tratados versus controles) es posible. Por esta razón, el test de hipótesis que subyace a la estimación original del tamaño de muestra podría ser incorrecto pues asume una probabilidad de cometer un Error Tipo I de 5% cuando en realidad, si el conjunto de test de hipótesis posibles es considerado conjuntamente, el Error Tipo I es sustancialmente mayor. Dado que controlar por este problema implica reducir el Error Tipo I para cada test individual, esto automáticamente implica un incremento del Error Tipo II, lo cual deriva finalmente en una reducción del nivel de poder en relación al considerado originalmente<sup>27</sup>.

---

<sup>27</sup> Es importante anotar que no existe consenso universal respecto a la utilización de este tipo de ajuste en la literatura. Al respecto, ver Gelman et al (2012) para una discusión de porque no siempre tiene sentido preocuparse por comparaciones múltiples.

La aproximación más común para la corrección del problema de comparaciones múltiples consiste en el uso de la corrección de Bonferroni. Esta aproximación, parte de la definición de la siguiente desigualdad:

$$(A7.6) \Pr\{\min(p_j : 1 \leq j \leq m) \leq \alpha/m\} \leq \alpha;$$

Donde  $p_1, \dots, p_m$  son los p-values observados y  $\alpha$  es el p-value crítico. Nótese que  $m$  es el número de comparaciones posibles, por lo que la desigualdad anterior define un nivel de Error Tipo I para la familia de tests derivados de la comparación entre los diferentes niveles de tratamiento. Esto implica que, para mantener un nivel de Error Tipo I para toda la familia de hipótesis, el nivel de Error Tipo I de cada hipótesis individual sea ajustado a un nivel  $\alpha/m$ .

En relación a nivel de poder, la corrección de Bonferroni implica que el p-value asociado al Error Tipo I en la estimación del Efecto Mínimo Detectable para una comparación individual sería:

$$(A7.7) p_i = \alpha/m$$

Este p-value ajustado es el que se usa en el ejercicio de cálculo de muestra.

A fin de ajustar por la existencia de comparaciones múltiples, la corrección de Bonferroni se implementa del modo siguiente:

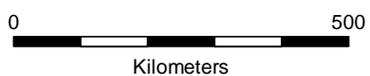
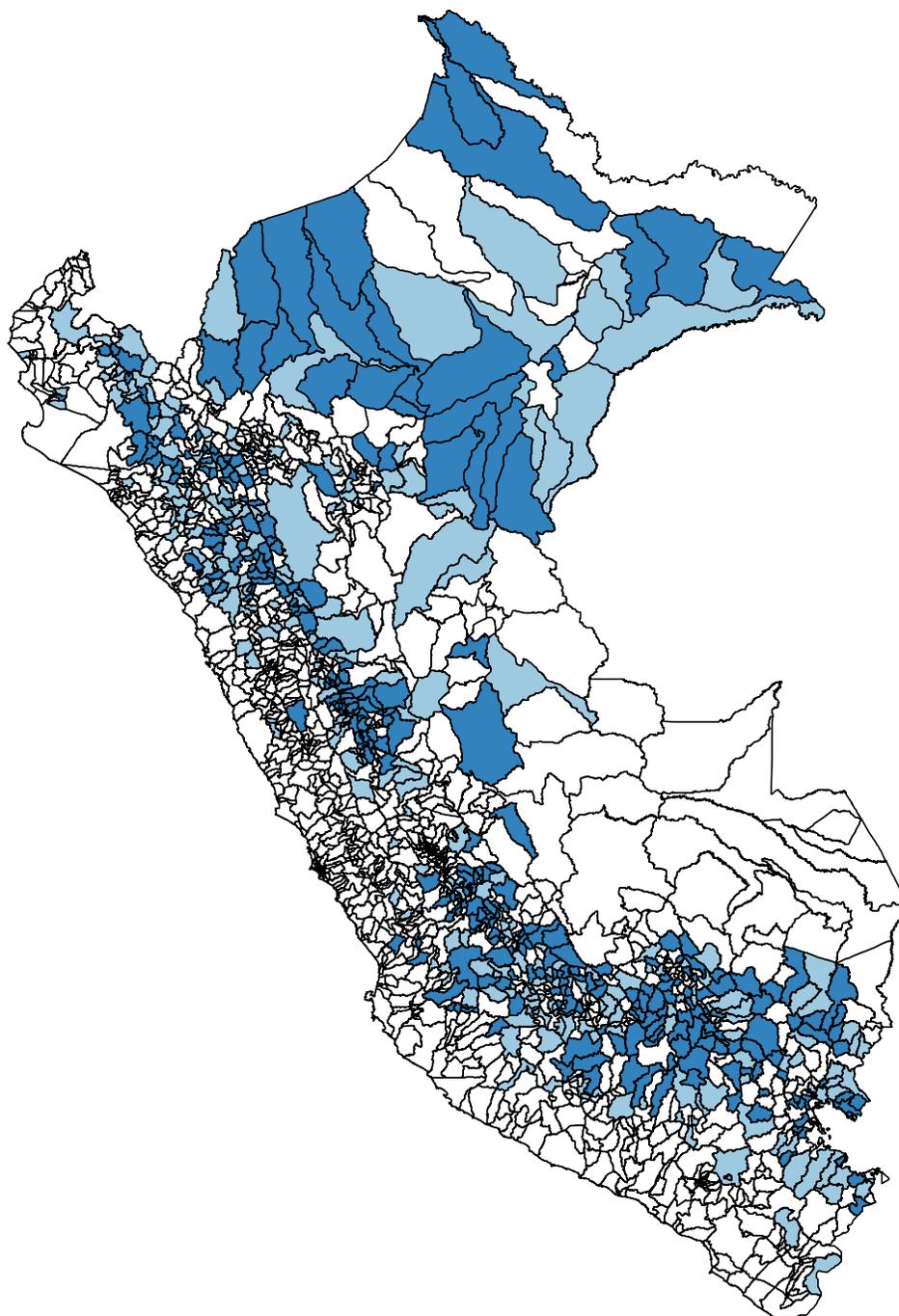
$$(A7.8) EMD_{DD\_cluster}^{Bonf} = \underbrace{(t_{(1-k)} + t_{(\alpha/m)})}_{A} * \underbrace{\sqrt{\frac{1}{P(1-P)}} \sqrt{\frac{\sigma^2}{nJ}}}_{B} * \underbrace{\sqrt{1+(n-1)\rho}}_{B} * \underbrace{\sqrt{2(1-r)}}_{C},$$

en donde  $t_{(\alpha/m)}$  es el t-estadístico para el Error Tipo I ajustado. A partir de (A7.8), es posible derivar una fórmula para el cálculo del tamaño de muestra (definido en términos de clusters y asumiendo un número fijo de observaciones por cluster):

$$(A7.9) J_{Bonf} = (t_{(1-k)} + t_{\alpha/m})^2 * \frac{1}{P(1-P)} \frac{\sigma^2}{n * ES^2} * [1 + (n-1)\rho] * [2(1-r)],$$

en donde  $ES$  es el tamaño del efecto esperado por la intervención medido en la unidad de referencia correspondiente.

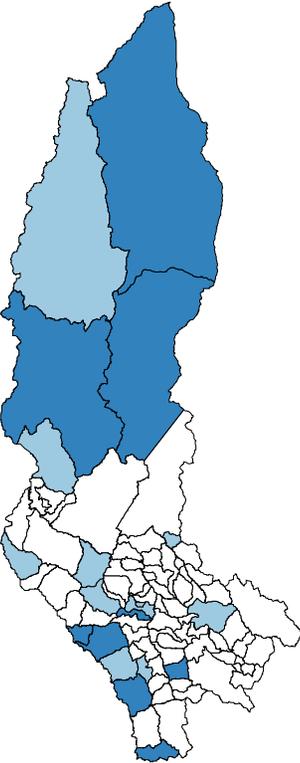
**APÉNDICE VIII:  
DISTRITOS TRATADOS Y CONTROLES**



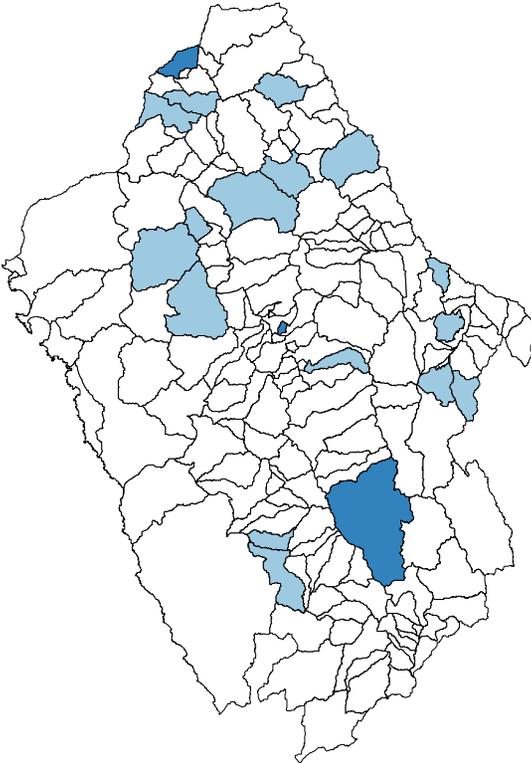
Nota: Los distritos de color azul constituyen el grupo de tratamiento para el periodo 2013/2014. Los distritos de color celeste forman parte del grupo de control y serán cubiertos por el programa en el 2015.

**MAPAS REGIONALES**

**Mapa 1: Amazonas**

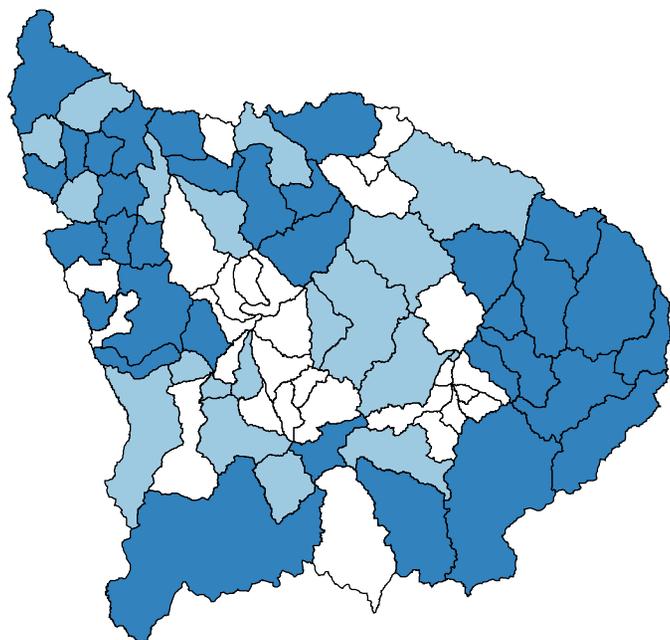


**Mapa 2: Ancash**

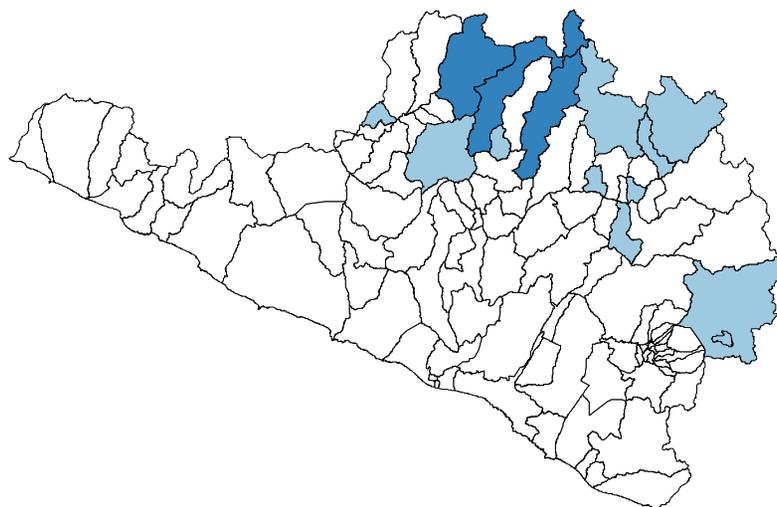


Nota: Los distritos tratados se representan con el color azul mientras que los controles con color celeste.

**Mapa 3: Apurímac**

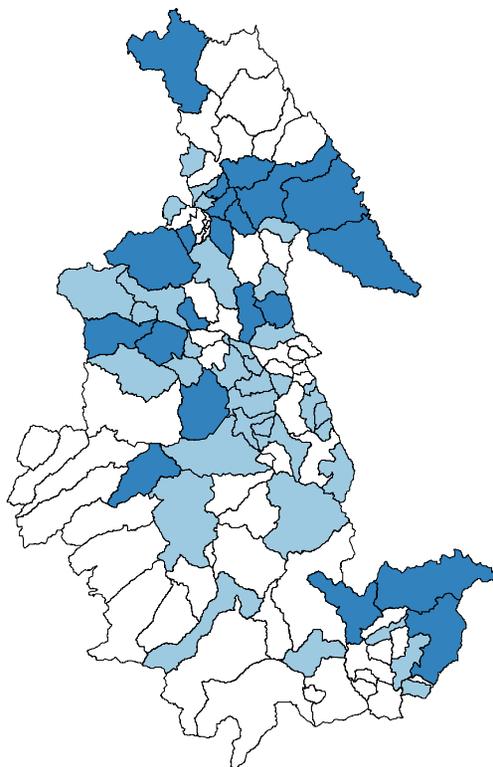


**Mapa 4: Arequipa**

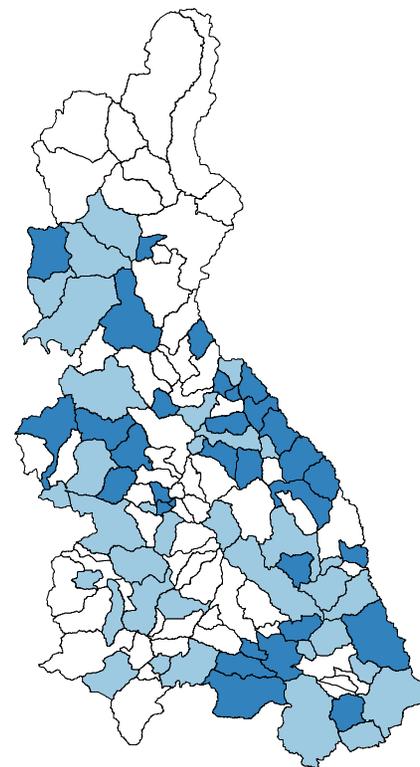


Nota: Los distritos tratados se representan con el color azul mientras que los controles con color celeste.

**Mapa 5: Ayacucho**

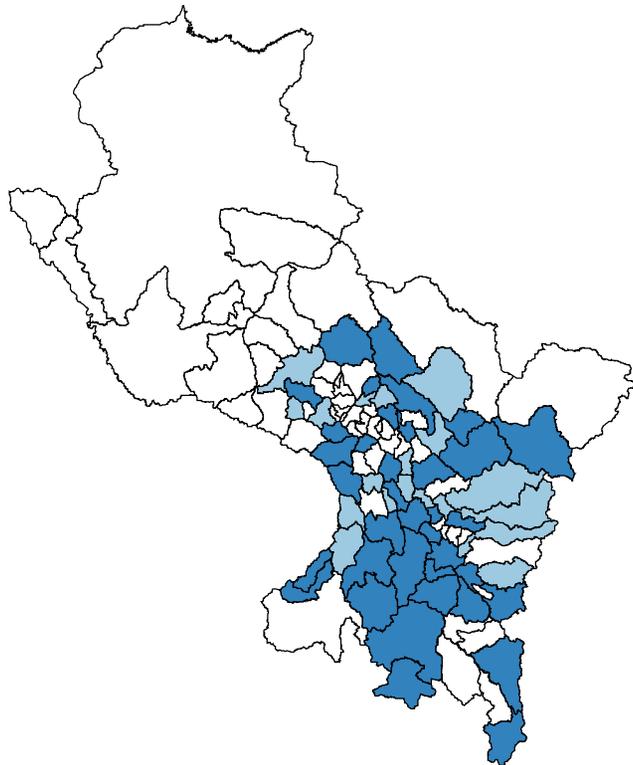


**Mapa 6: Cajamarca**

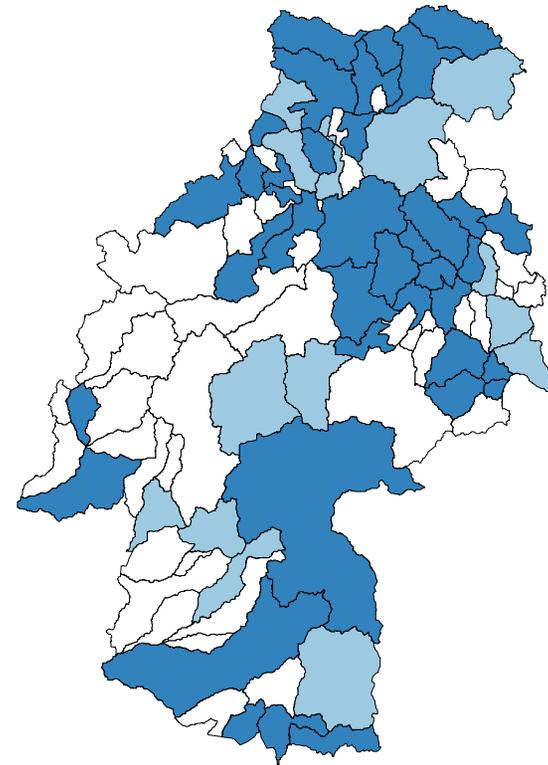


Nota: Los distritos tratados se representan con el color azul mientras que los controles con color celeste.

**Mapa 7: Cusco**

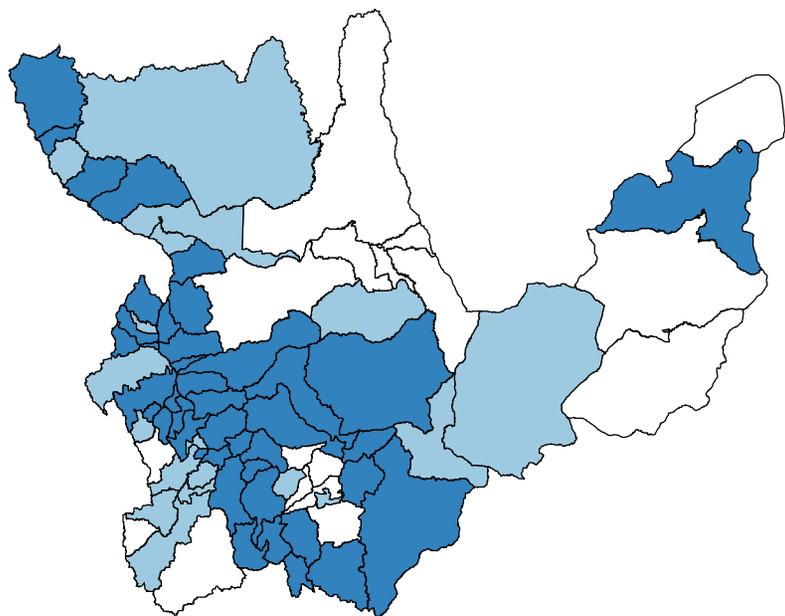


**Mapa 8: Huancavelica**



Nota: Los distritos tratados se representan con el color azul mientras que los controles con color celeste.

**Mapa 9: Huánuco**

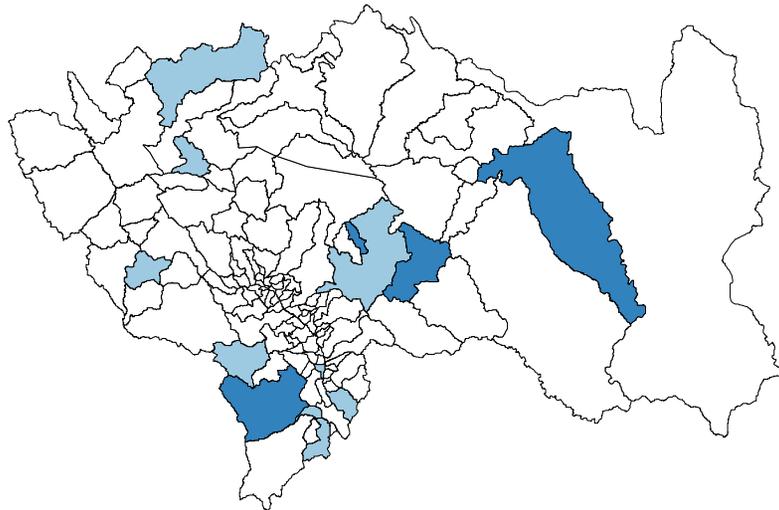


**Mapa 10: Ica**

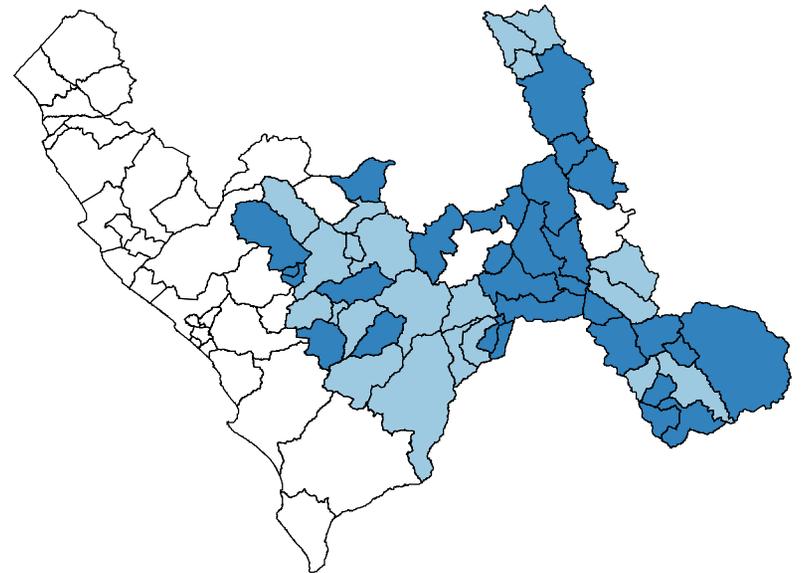


Nota: Los distritos tratados se representan con el color azul mientras que los controles con color celeste.

**Mapa 11: Junín**

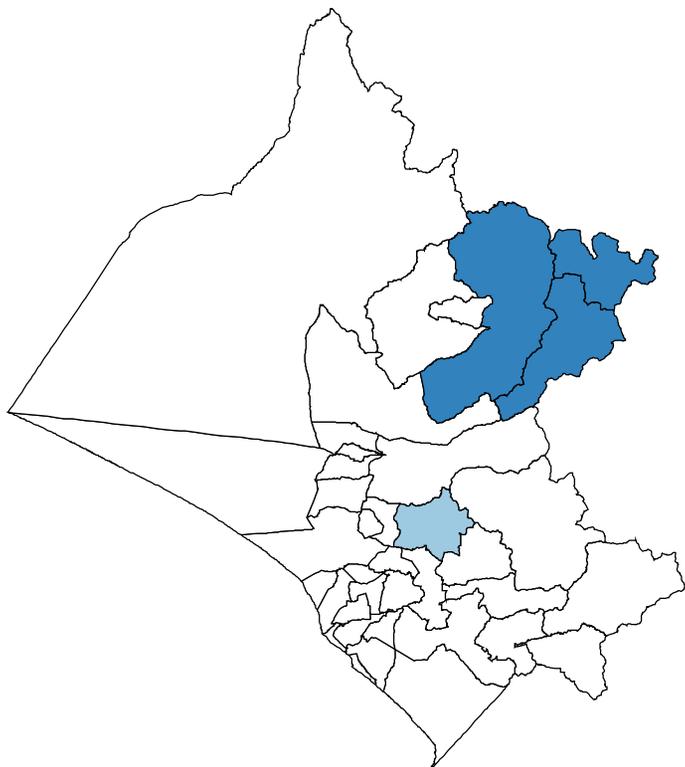


**Mapa 12: La Libertad**



Nota: Los distritos tratados se representan con el color azul mientras que los controles con color celeste.

**Mapa 13: Lambayeque**

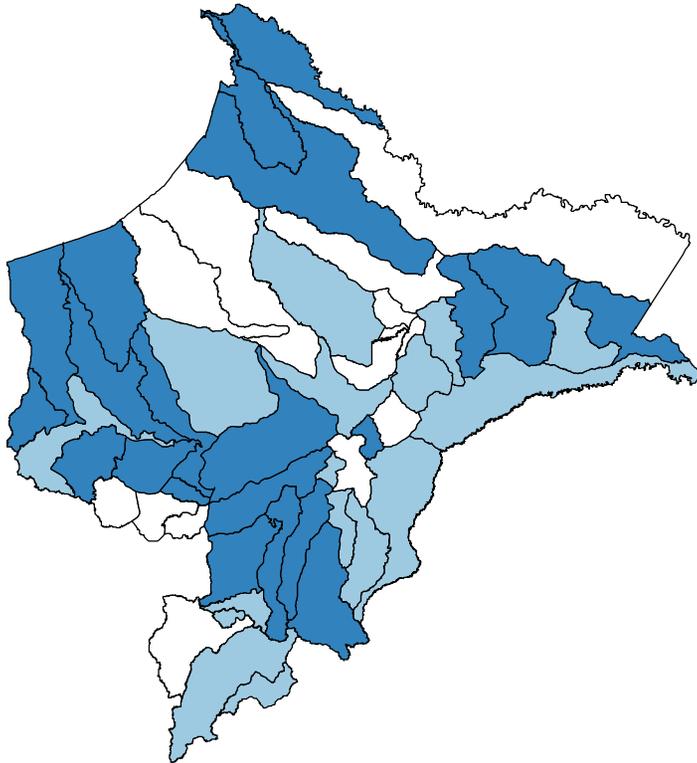


**Mapa 14: Lima (incluyendo Callao)**

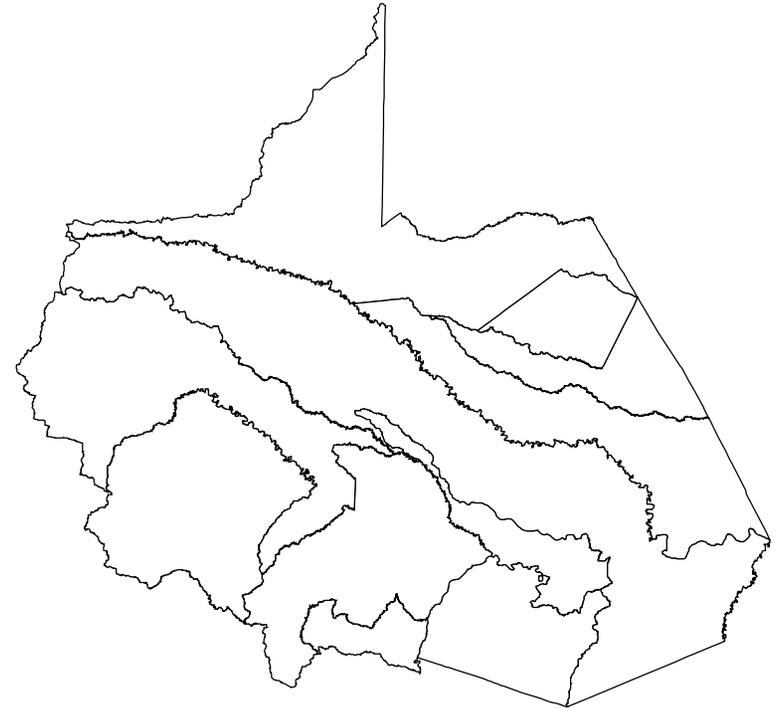


Nota: Los distritos tratados se representan con el color azul mientras que los controles con color celeste.

**Mapa 15: Loreto**

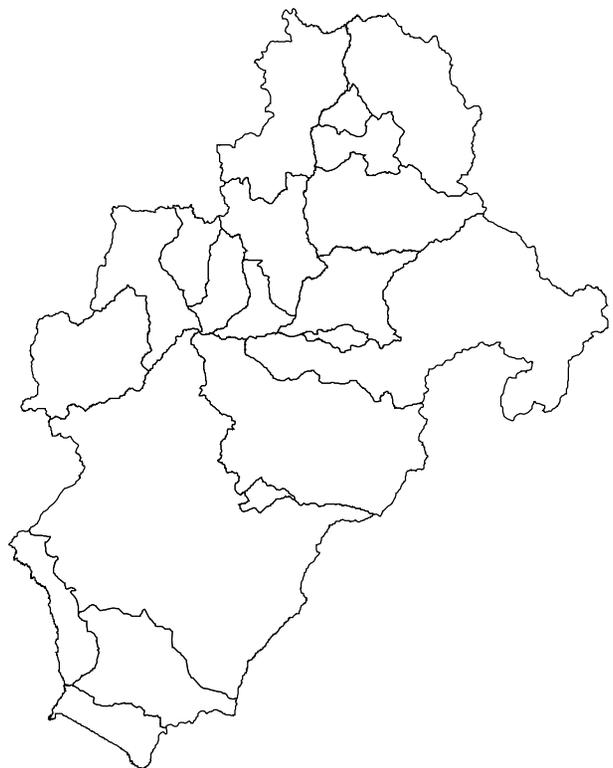


**Mapa 16: Madre de Dios**

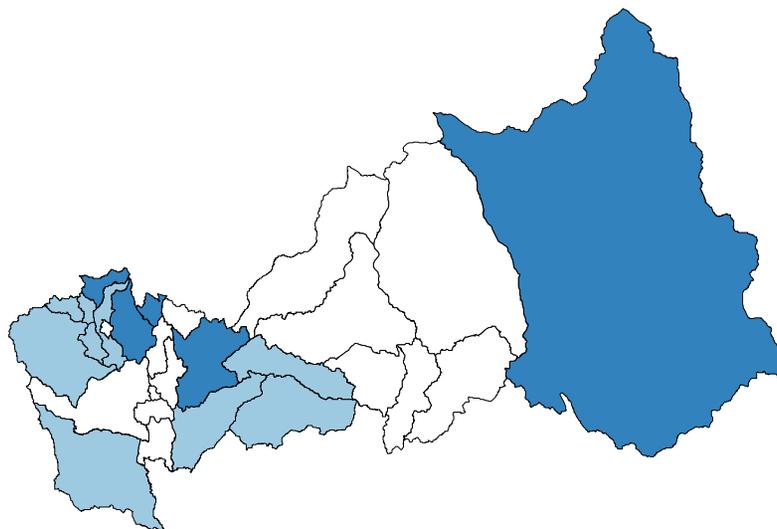


Nota: Los distritos tratados se representan con el color azul mientras que los controles con color celeste.

**Mapa 17: Moquegua**

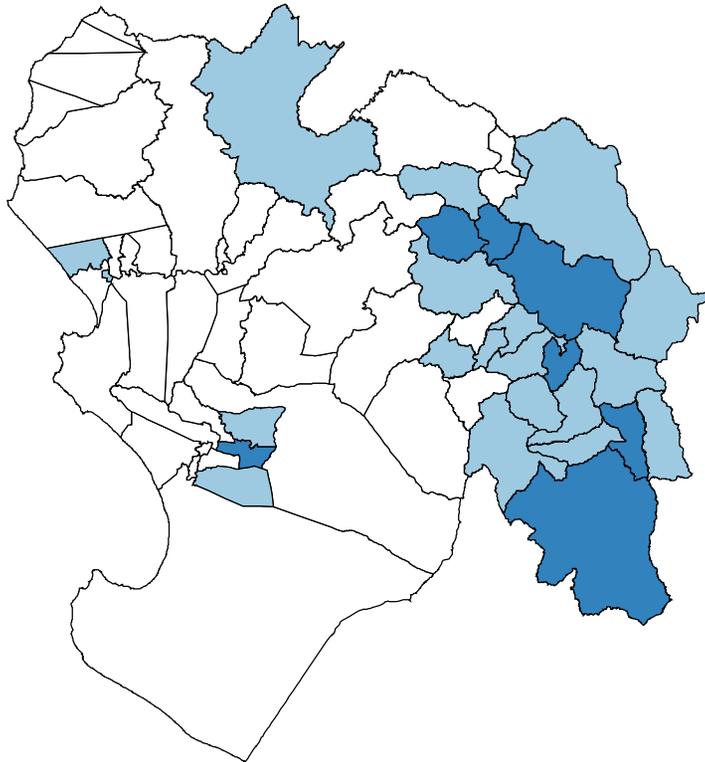


**Mapa 18: Pasco**

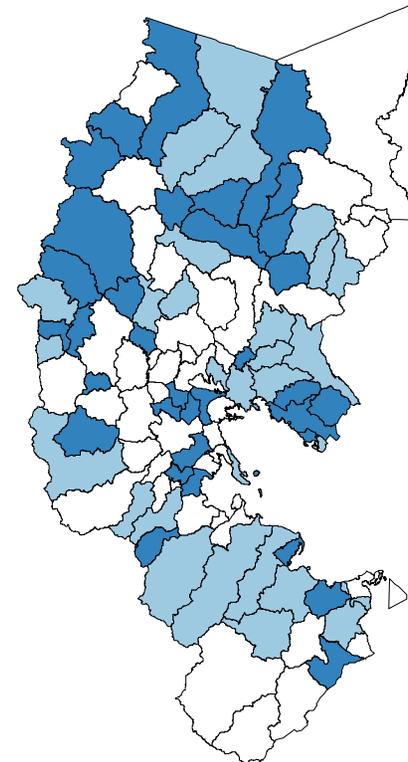


Nota: Los distritos tratados se representan con el color azul mientras que los controles con color celeste.

**Mapa 19: Piura**

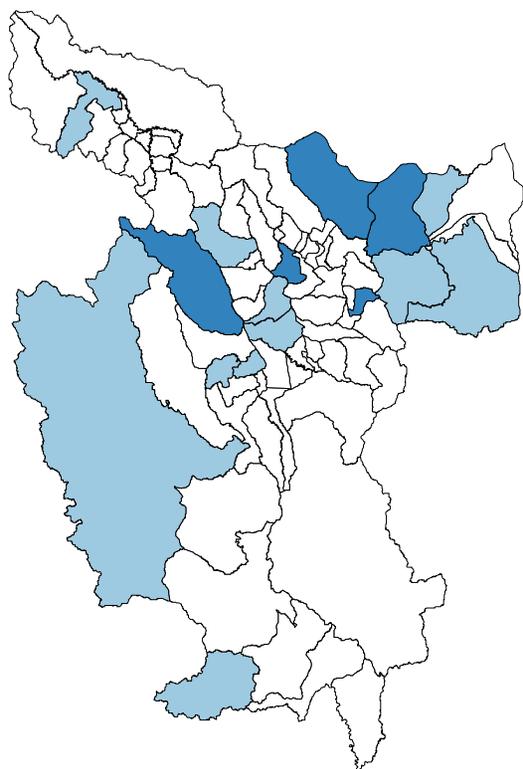


**Mapa 20: Puno**

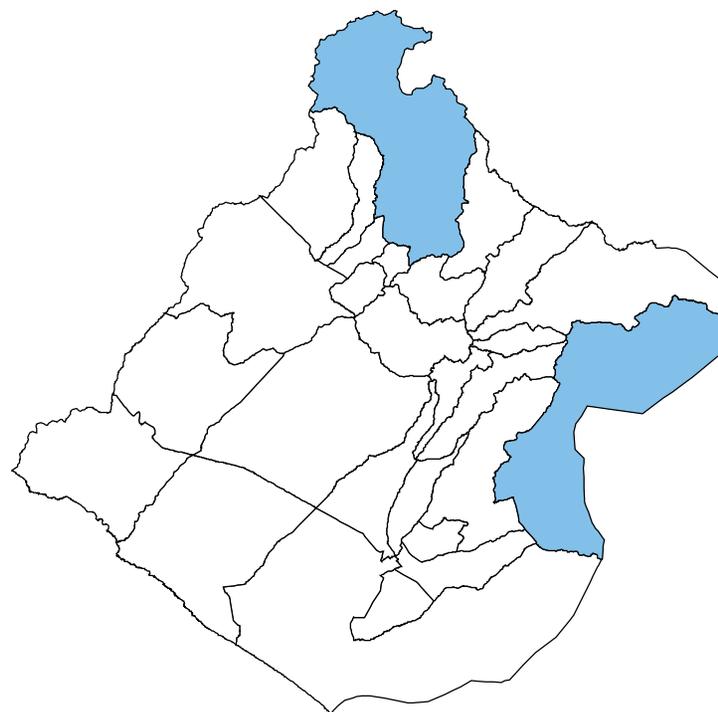


Nota: Los distritos tratados se representan con el color azul mientras que los controles con color celeste.

**Mapa 21: San Martín**



**Mapa 22: Tacna**



Nota: Los distritos tratados se representan con el color azul mientras que los controles con color celeste.

**Mapa 23: Tumbes**



**Mapa 24: Ucayali**



Nota: Los distritos tratados se representan con el color azul mientras que los controles con color celeste.