



2023 EANNA

ENCUESTA DE ACTIVIDADES DE NIÑOS, NIÑAS Y ADOLESCENTES

Informe Diseño Muestral EANNA 2023

Instituto Nacional de Estadísticas – Ministerio de Desarrollo Social y Familia

Febrero 2025

Informe revisado y editado por el Ministerio de Desarrollo Social y Familia a partir de Informe de Diseño Muestral de la Encuesta de Actividades de Niños, Niñas y Adolescentes 2023, entregado por el Instituto Nacional de Estadísticas (INE), en el marco del convenio *“Programa de trabajo conjunto para la elaboración del diseño para la Encuesta de Actividades de Niños, Niñas y Adolescentes (EANNA 2023) y la Encuesta de Bienestar Social (EBS 2023)”*, correspondiente a la entrega N° 9 del 10 de diciembre de 2024.

La actualización de este informe en febrero de 2025, respecto con la versión publicada en junio de 2023, incorpora los capítulos V y VI, que detallan el diseño del factor de expansión y el cálculo y evaluación de las estimaciones de la encuesta, respectivamente.

ÍNDICE

I.	ANTECEDENTES	5
II.	CARACTERÍSTICAS GENERALES DEL DISEÑO MUESTRAL DE EANNA	6
II.1.	Objetivo	6
II.2.	Población objetivo	6
II.3.	Unidad de información.....	6
II.4.	Marco muestral	7
II.5.	Dominios de estudio	7
II.6.	Objetivo de precisión y tamaño de muestra	8
II.7.	Estrategia muestral	8
III.	CÁLCULO Y DISTRIBUCIÓN DEL TAMAÑO MUESTRAL.....	9
III.1.	Etapas para la estimación del tamaño muestral para EANNA 2023	10
III.2.	Parámetros utilizados para el cálculo del tamaño muestral.....	14
III.3.	Metodología de cálculo del tamaño muestral utilizando muestras complejas	15
III.4.	Resumen proceso de simulación.....	17
III.5.	Escenario elegido	21
III.6.	Tamaño muestral propuesto.....	22
III.7.	Errores esperados	24
IV.	SELECCIÓN DE UNIDADES MUESTRALES	25
V.	DISEÑO FACTORES DE EXPANSIÓN.....	27
V.1.	Factor de expansión de primera fase.....	29
V.2.	Factor de expansión de segunda fase	29
V.2.1.	Ponderador de selección de NNA como inverso de la probabilidad de selección 29	
V.2.2.	Producto entre los ponderadores de primera y segunda fase.....	30
V.2.3.	Ajuste por elegibilidad desconocida.....	30
V.2.4.	Ajuste por no elegibilidad.....	31
V.2.5.	Ajuste por no respuesta	32
V.3.	Suavizamiento	34
V.3.1.	Método de contracción a la media (CM).....	34
V.3.2.	Método R-K.....	35
V.3.3.	Método mixto.....	35
V.4.	Calibración de factores de expansión	36
V.5.	Redondeo Probabilístico	39
VI.	CÁLCULO Y EVALUACIÓN DE ESTIMACIONES.....	39
VI.1.	Estimación de la varianza compleja	39
VI.2.	Variables que identifican el diseño muestral complejo	41
VI.3.	Estadísticos asociados a los parámetros de interés	41

VII. ANEXOS.....	43
VII.1. Anexo N°1: Códigos de Disposición Final de Casos EANNA 2023	43
VIII. REFERENCIAS	44

ÍNDICE DE TABLAS

Tabla 1. Resumen de parámetros utilizados y tamaños y errores resultantes según escenario, EANNA 2023.....	20
Tabla 2: Estimación de tamaños muestrales según proyecciones de población al 15 de junio de 2023 con marco Casen 2022 actualizado.....	21
Tabla 3: Tamaños y errores finales a nivel nacional, regional, urbano-rural, EANNA 2023	23
Tabla 4: Tamaños y errores EANNA 2012 y propuesta EANNA 2023. Nacional Urbano - Rural	24
Tabla 5: Errores esperados asociados al parámetro de interés según tamaño muestral objetivo ..	24
Tabla 6: Distribución del número de NNA por región y área	26
Tabla 7. Distribución viviendas según elegibilidad.....	32
Tabla 8. Variables utilizadas en el modelo de no respuesta EANNA 2023.....	32
Tabla 9. Veintiles no respuesta EANNA 2023.....	34
Tabla 10. Proyecciones de población de 5 a 17 años al 30 de julio de 2023 según región y área....	37
Tabla 11. Distribución de los cambios de estrato en la muestra lograda, EANNA 2023.	37
Tabla 12. Distribución de la muestra y tamaño poblacional expandido según nacional – área y región, EANNA 2023	38
Tabla 13. Estimación del trabajo infantil con sus respectivos errores muestrales a nivel nacional, área y regional, EANNA 2023	42

ÍNDICE DE CUADROS

Cuadro 1. Parámetros utilizados para el cálculo de tamaño muestral actual	14
Cuadro 2: Parámetros comunes utilizados en escenarios EANNA 2023.....	19

ÍNDICE DE FIGURAS

Figura 1. Diagrama de procesos de desarrollo de factor de expansión EANNA 2023	28
Figura 2. Ordenamiento de factores de expansión atípicos y no atípicos	35

I. ANTECEDENTES

La Encuesta de Actividades de Niños, Niñas y Adolescentes, EANNA 2023, es un proyecto a cargo del Ministerio de Desarrollo Social y Familia (MDSF), que cuenta con la colaboración técnica del Ministerio del Trabajo y Previsión Social, el Ministerio de la Mujer y la Equidad de Género (MMEG), la Organización Internacional del Trabajo (OIT) y el Fondo de las Naciones Unidas para la Infancia (UNICEF). Las cuatro instituciones componen la Mesa Técnica de EANNA, la cual ha asesorado los procesos de diseño, construcción, recolección y análisis de la encuesta.

El objetivo de la EANNA es estimar la magnitud y las principales características del trabajo infantil (TI), así como las principales actividades que realizan los niños, niñas y adolescentes en el país, permitiendo actualizar el diagnóstico proporcionado por EANNA 2012.

La encuesta constituye, además, un instrumento importante para el diseño, implementación, seguimiento y evaluación de políticas y programas sociales que se ejecutan en el país, especialmente en materias relacionadas con el trabajo infantil y las principales actividades que realizan los niños, niñas y adolescentes.

El diseño muestral y el proceso de recolección de la encuesta estuvo a cargo del Instituto Nacional de Estadísticas (INE), para lo cual el MDSF e INE convinieron un programa de trabajo conjunto para la elaboración del diseño muestral de EANNA 2023.

EANNA es una encuesta bifásica de la Encuesta de Caracterización Socioeconómica Nacional (Casen). Esta última se desarrolla generalmente, de manera bianual con el objeto de dar continuidad a la serie de información sobre la situación socioeconómica de los hogares y la población del país. Lo anterior significa que, desde la base de los hogares encuestados en Casen 2022, se llevó a cabo la selección de personas (niños, niñas y adolescentes) que fueron encuestados en EANNA 2023.

El presente informe es parte de los productos entregables del convenio INE – MDSF y presenta el diseño metodológico elaborado para EANNA con el fin de responder a los objetivos propios de la encuesta y, además, el proceso de cálculo de los factores de expansión.

La metodología del diseño muestral de la EANNA 2023, describe el conjunto de opciones metodológicas y de procedimientos estadísticos adoptados a lo largo del proceso de planificación, que son considerados los más apropiados para seleccionar una muestra probabilística, estadísticamente representativa de los niños, niñas y adolescentes (NNA) de Chile y que permita estimar la tasa de trabajo infantil en cada región del territorio nacional, así como también, en las áreas geográficas urbanas y rurales a nivel nacional y para el país en su conjunto.

Este informe contempla los siguientes capítulos:

- Características generales del diseño muestral de EANNA
- Cálculo y distribución del tamaño muestral
- Selección de unidades muestrales
- Diseño de factores de expansión
- Estimación de parámetros de interés y su respectivo cálculo de la varianza bajo muestreo complejo.

II. CARACTERÍSTICAS GENERALES DEL DISEÑO MUESTRAL DE EANNA

II.1. Objetivo

Estimar la magnitud y las principales características del trabajo infantil (TI), así como las principales actividades que realizan los niños, niñas y adolescentes en el país, actualizando el diagnóstico proporcionado por EANNA 2012.

II.2. Población objetivo

La población objetivo es el universo de niños, niñas y adolescentes (NNA) de 5 a 17 años cumplidos a la fecha de realización de la encuesta, que habitan el territorio nacional y son representados por la Encuesta de Caracterización Socioeconómica Nacional (Casen) 2022.

II.3. Unidad de información

La EANNA recoge información a través de la entrevista directa a niños, niñas y adolescentes de 5 a 17 años, permitiendo identificar a los niños en trabajo infantil, conocer sus características y analizar las condiciones en que desarrollan su actividad. Al mismo tiempo, mediante entrevistas a las personas cuidadoras de estos NNA, se obtiene información sobre sus hogares y se recoge la percepción de la persona cuidadora sobre el trabajo infantil.

II.4. Marco muestral

Los NNA entrevistados en EANNA 2023 son una muestra que proviene de los NNA que integran los hogares efectivamente entrevistados por la Encuesta Casen 2022. De esta manera, ambas encuestas comparten el mismo marco muestral basado en el Censo de Población y Viviendas de 2017 y actualizado al año 2020.

El marco muestral de viviendas utilizado en Casen 2022 fue elaborado a partir del MMV 2020 y está constituido por unidades primarias de muestreo (UPM) que corresponden a áreas geográficas homogéneas, en términos del número de viviendas particulares que las conforman, excluyendo viviendas de temporada y colectivas.

Las UPM presentan una estratificación geográfica, dada por la división político-administrativa, según la conformación región-provincia-comuna y desagregada, además según la división censal, clasificándolas en áreas urbanas y rurales. A esta estratificación se suma la clasificación socioeconómica (con niveles bajo, medio y alto) construida a partir de un modelo que considera variables del Censo 2017¹.

A través de las UPM se accede a las viviendas y hogares, cuyos integrantes son principalmente el objeto final de estudio. Para la selección de viviendas o unidades secundarias de muestreo (USM) de Casen 2022 se consideró la actualización de las UPM seleccionadas en primera etapa donde, 60% fueron actualizadas a través de la verificación de direcciones en gabinete, a partir de los listados obtenidos de Precenso 2016 y utilizando imágenes satelitales, Certificados de Recepción Final de conjuntos habitacionales, servidores de mapas, entre otros. En el 40% restante la actualización corresponde a la enumeración en terreno donde, a partir de los listados de Precenso, se recorren las UPM verificando y/o añadiendo direcciones de nuevas edificaciones.

II.5. Dominios de estudio

EANNA 2023 se encuentra diseñada para obtener indicadores confiables con errores de muestreo aceptables para los niveles nacional, nacional urbano, nacional rural y regional.

¹ Documentos de trabajo: Estratificación socioeconómica del marco muestral de viviendas 2017 MMV 2017. <https://www.ine.cl/inicio/documentos-de-trabajo/documento/estratificaci%C3%B3n-socioecon%C3%B3mica-del-marco-muestral-de-viviendas-2017>.

II.6. Objetivo de precisión y tamaño de muestra

Considerando un levantamiento presencial, con un tamaño objetivo ajustado de 18.069 NNA, se espera obtener estimaciones del parámetro de interés con un error absoluto nacional de 0,6% y, un error relativo, de 8,5%. Con este tamaño los errores absolutos regionales no superan 3,4% y, los errores relativos, 51,1%. Pese a que existen errores relativos elevados, a partir del error estándar se aprecia que los valores a nivel regional cumplen con el estándar de calidad que utiliza el INE para la publicación de parámetros de interés de sus estudios.

Dados los tamaños muestrales objetivos a nivel nacional y regional, para la distribución de los tamaños a nivel de estratos (región-área-tramo etario) se establece un mínimo de 10 NNA, especialmente, en áreas rurales.

II.7. Estrategia muestral

El diseño muestral de EANNA 2023 implica la obtención de una muestra probabilística bifásica. La primera fase está dada por el diseño muestral de Casen 2022, que corresponde a una muestra probabilística, estratificada y bietápica, siendo los estratos conformados por la combinación comuna-área y nivel socioeconómico². La segunda fase se compone de una muestra probabilística estratificada, siendo los estratos muestrales conformados por región-área-tramo etario³.

² El nivel socioeconómico se incorpora para algunas combinaciones de comuna-área.

³ Se consideran 4 tramos etarios: 5 a 8 años, 9 a 11 años, 12 a 14 años y 15 a 17 años.

III. CÁLCULO Y DISTRIBUCIÓN DEL TAMAÑO MUESTRAL

Entre enero y marzo de 2023, el INE tuvo la misión de desarrollar la estrategia de muestreo de EANNA 2023 que consistió en proponer y elaborar, de acuerdo con los requerimientos planteados por el MDSF, simulaciones de tamaños muestrales con los términos específicos de errores muestrales aceptables y distribución muestral, optimizando la selección de las unidades muestrales, de acuerdo con criterios de factibilidad y de presupuesto.

Durante esta etapa, previa al desarrollo de la estrategia de trabajo de campo, se estableció como hito inicial del convenio INE-MDSF obtener un tamaño muestral óptimo para la estimación del principal parámetro de interés, con márgenes de precisión que permitan que EANNA 2023 tenga representatividad a nivel nacional, nacional urbano, nacional rural, y regional. Para ello, se simularon distintos escenarios o alternativas de tamaño muestral y distribución, basados en la unificación de criterios estadísticos solicitados por MDSF y discutidos en reuniones de trabajo.

Las simulaciones se realizaron considerando los parámetros del levantamiento de EANNA 2012, Casen en Pandemia 2020, Piloto Casen Cambio de Marco y Casen 2022. Se formularon diferentes alternativas de distribución de la muestra objetivo que fueron presentadas en reuniones de trabajo conformadas por los equipos de la División de Observatorio Social del Ministerio y el Subdepartamento de Diseño de Marcos y Muestras del INE. Finalmente, se estableció un escenario que se ajusta a los requerimientos establecidos por el MDSF y considera el presupuesto acordado.

El escenario elegido utiliza como marco de referencia los 33.989 NNA que fueron resultado del levantamiento de Casen 2022⁴. El error absoluto propuesto para cumplir con el estándar de calidad y el orden de la distribución regional de la población es de 0,6% a nivel nacional. Con errores absolutos entre 1,5% y 3,4% a nivel regional, se obtiene un tamaño bajo muestreo aleatorio simple (M.A.S.) de 8.072 NNA a nivel nacional.

Para obtener el efecto de diseño se contempla el efecto de diseño de EANNA 2012, el efecto de diseño utilizado para Casen 2022 y el efecto de diseño resultante del cálculo de la tasa de pobreza en hogares con NNA del Piloto Cambio de Marco Casen 2022 donde, el efecto de diseño propuesto va a corresponder al máximo entre estos tres valores, con excepción de las regiones de Coquimbo, Metropolitana, Ñuble, Los Ríos, Aysén y Magallanes. A nivel regional se establece que los efectos de diseño utilizados fluctúen entre 1,5 y 3,0.

⁴ Corresponde a la última actualización realizada por el MDSF con fecha 11 de abril de 2023 y puesto a disposición del INE el 20 de abril de 2023.

Luego, el tamaño objetivo ajustado por efecto de diseño es de 18.188 NNA que, al aplicar el ajuste por finitud, considerando como universo las proyecciones de población al 15 de junio de 2023⁵, entrega un tamaño de 18.066 NNA a nivel nacional. La distribución de los tamaños por estrato de muestreo (región - área - tramo etario), se realiza en forma proporcional a las proyecciones de población a junio de 2023 y se establece un mínimo de 10 NNA por cada uno, aumentando la muestra a 18.069 NNA con un error absoluto es de 0,6% y un error relativo de 8,5%. Si bien a nivel regional se observan errores relativos elevados, a partir del error estándar se aprecia que estos valores cumplen con el estándar de calidad.

Finalmente, para obtener el tamaño con sobremuestreo, se utiliza una tasa de no logro a nivel regional donde se selecciona el máximo entre: la tasa de no logro de Casen 2022, ponderada por la proporción de la TNL de EANNA 2012 respecto de Casen 2011 y el máximo de la TNL entre Casen 2017 y Casen 2022, que se pondera con EANNA 2012, que equivale a 60% y, Casen a 40%. Este procedimiento dio lugar a un tamaño con sobremuestreo de 25.611 NNA.

III.1. Etapas para la estimación del tamaño muestral para EANNA 2023

El cálculo del tamaño muestral para EANNA 2023 inicia con la obtención de parámetros asociados a la última aplicación de la encuesta (EANNA 2012) en los niveles de estimación definidos para el nuevo estudio (EANNA 2023). Así, a nivel regional se obtiene la tasa de trabajo infantil junto con sus estadísticos asociados (error estándar y efecto del diseño). Luego, se prevé una serie de procedimientos que corresponden al paso desde un diseño bajo muestreo aleatorio simple a un diseño complejo, seguido de otros dos ajustes, determinados por la finitud de la población y la tasa de no logro, obteniendo así el tamaño con sobremuestreo.

En los párrafos siguientes se describe, en detalle, cada uno de los siete pasos realizados para definir el tamaño muestral con el total de NNA a encuestar, en principio, a nivel de región, luego a nivel de estrato de muestreo (región - área - tramo etario). Obtenidos los tamaños muestrales para el escenario elegido se determina un tamaño de muestreo por estrato de manera tal que existan, al menos, 10 NNA a encuestar por estrato.

El cálculo del tamaño para una encuesta que presenta un cambio en su diseño muestral debe considerar la obtención de un tamaño bajo M.A.S. para luego, aplicar los correspondientes ajustes según el efecto del nuevo diseño. A continuación, se detallan los pasos a seguir:

⁵ Estas proyecciones consideran el 15 de junio de 2023 como fecha de referencia; fecha que se encuentra dentro del periodo estipulado para el levantamiento.

Paso 1: Estimación de parámetros

Se obtienen estimaciones, a nivel regional, de la tasa de trabajo infantil (TI) junto con el error estándar y el efecto del diseño asociado a este indicador, usando los resultados del trabajo de campo de EANNA 2012. A partir de estos parámetros se obtiene la cuasivarianza poblacional $S(r)_r^2$ según la ecuación (1):

$$S(ti)_r^2 = \frac{n \cdot \bar{m}}{Def f(ti)_r} \cdot SE(ti)_r^2 \quad (1)$$

Donde:

- n : Número de conglomerados logrados en la región r .
- \bar{m} : Número promedio de NNA por conglomerado en la región r .
- $n \cdot \bar{m} = m_r$: Número de NNA logrados en la región r .
- $Def f(ti)_r$: Efecto del diseño asociado a la tasa de trabajo infantil (ti) en la región r .
- $SE(ti)$: Error estándar de la estimación de la tasa de trabajo infantil (ti) en la región r .

El efecto del diseño asociado al parámetro de interés se define como:

$$Def f(ti) = \sigma^2(ti)_{MC} / \sigma^2(ti)_{MAS} \quad (2)$$

Donde:

- $\sigma^2(ti)_{MC}$: Cálculo de la varianza bajo un diseño complejo.
- $\sigma^2(ti)_{MAS}$: Cálculo de la varianza bajo un muestreo aleatorio simple.

En el proceso de obtención del efecto de diseño de una encuesta, es posible que se presenten valores menores que 1 o mayores que lo esperado desde el diseño muestral o respecto de lo observado en diferentes niveles de estimación. En los casos de encuestas que hacen uso de diseños muestrales complejos y que la estimación del efecto del diseño se encuentra por debajo del valor 1, puede deberse a que los algoritmos de cálculo de la varianza se basan en aproximaciones que, en general, utilizan el método de conglomerado último, cuyo efecto puede traducirse en una subestimación de la varianza. Luego, al comparar esta estimación subestimada con la varianza bajo diseño aleatorio simple, esta última resulta mayor, obteniéndose un $Def f$ menor que 1. Ante esta situación, y bajo el supuesto que una muestra polietápica tiene un mayor error de muestreo que bajo muestreo aleatorio simple, se justifica que se considere un valor mínimo de 1 para el efecto de diseño utilizado en el cálculo de tamaños muestrales.

Por otra parte, en los casos en que el investigador considere que los efectos de diseño estimados exceden de manera importante lo esperado, puede tomar la decisión de establecer cotas superiores para su uso en el cálculo de tamaños, bajo el supuesto que los altos valores observados se deben a factores particulares del diseño y de la muestra recolectada y, considerando además, que la

inferencia se realiza sobre una submuestra particular de un gran número de muestras posibles obtenidas desde el levantamiento de la encuesta.

Paso 2: Cálculo de Tamaño bajo M.A.S.

Se obtienen los errores absolutos “objetivo” por región (d_{0r}) y luego, los errores relativos (e_{0r}) según una aproximación del cuantil de una distribución t de Student con nivel de confianza de 95% que, para efectos de la simulación, tiene un valor correspondiente a $t=2$. Luego, junto con la cuasivarianza calculada en la ecuación (1) se obtiene un tamaño muestral inicial a nivel regional (m_{0r}) considerando un muestreo aleatorio simple monoetápico.

$$m_{0r} = \frac{t^v_{1-\alpha/2} \cdot S(ti)_r^2}{d_{0r}^2} \quad (3)$$

Paso 3: Ajuste por efecto del diseño

Dado que el diseño muestral definido para EANNA 2023 es un diseño complejo, el tamaño calculado bajo M.A.S. se debe ponderar por un valor que representa el efecto provocado en la varianza al muestrear bajo un diseño estratificado en lugar de un M.A.S.

Así, el tamaño asociado al diseño complejo se obtiene desde la ecuación (4):

$$m_{1r} = m_{0r} \cdot Deff(ti)_r \quad (4)$$

Donde:

- m_{0r} : Es el número de NNA a encuestar en la región r bajo un muestreo aleatorio simple.
- m_{1r} : Es el número de NNA a encuestar en la región r considerando un diseño complejo.
- $Deff(ti)_r$: Es el efecto de diseño en la región r obtenido desde la estimación de alguna variable vinculada a la tasa de trabajo infantil.

Generalmente, los valores de $Deff(ti)_r$ son acotados al rango (1 – 5) basado en el supuesto que, por un lado, para un mismo tamaño muestral, una muestra polietápica tiene un mayor error de muestreo que bajo un muestreo aleatorio simple, lo que implica que $Deff$ no puede ser menor que 1 y, por otro, dado que la evidencia empírica en estudios de este tipo apunta a que la ganancia en precisión, producto de corregir el tamaño muestral por un efecto del diseño mayor a 5, es marginal respecto al incremento de los costos de levantamiento del nuevo tamaño muestral.

Paso 4: Corrección por población finita

El tamaño m_{1r} a su vez, es ajustado por un ponderador que da cuenta de la finitud de la población. Su forma de cálculo se observa en la ecuación (5):

$$m_{2r} = \frac{t^v_{1-\alpha/2} \cdot S(r)_r^2 \cdot Deff(ti)_r}{d_{0r}^2 + t^v_{1-\alpha/2} \cdot Deff(ti)_r \cdot S(ti)_r^2 / M_r} = \frac{m_{1r}}{1 + m_{1r} / M_r} \quad (5)$$

Donde:

m_{2r} : Es el total de NNA a encuestar en la región r considerando un diseño complejo y asumiendo finitud de la población.

M_r : Es el total de NNA en la región r , según marco de selección.

Finalmente, este tamaño es el que determina la muestra objetivo regional, que se debe distribuir en cada uno de los estratos de muestreo definidos.

Paso 5: Distribución por estrato de muestreo: región – área – tramo etario

Una vez determinado el tamaño muestral objetivo m_{2r} a nivel regional, se distribuye este valor, en forma proporcional al marco de selección, en cada uno de los estratos que las conforman, primero a nivel de área y luego a nivel de tramo etario.

Es importante mencionar que, el tamaño inicial calculado en cada estrato de muestreo puede ajustarse en la medida que los estratos no cumplan con tamaños mínimos definidos, por lo que, los tamaños regionales calculados en el paso 4 eventualmente podrían aumentar, caso en el cual, se deben recalcular los errores muestrales propuestos inicialmente.

Paso 6: Ajuste por no logro

Una vez definido el total de unidades a encuestar para estimar la TI, se debe tener en consideración que, en el trabajo de campo es posible no lograr algunas unidades por diversas razones, tales como: rechazos, moradores ausentes, entre otras. Por este motivo se hace necesario ajustar el tamaño muestral definido en el paso 5 por un ponderador que da cuenta de las posibles pérdidas de unidades muestrales.

Para calcular este ponderador, se utiliza información relacionada con el logro de la encuesta en levantamientos anteriores de la misma encuesta o de operaciones estadísticas de similar temática o método de recolección. El nivel de desagregación idóneo es a nivel de estrato de muestreo. Obtenidos los tamaños con sobremuestreo a este nivel, los tamaños a nivel regional son calculados como la agregación de los tamaños de los estratos que las conforman.

Para calcular el ponderador de ajuste (tasa de no logro) se utiliza el máximo entre la tasa de Casen 2022 ponderada por la proporción de la tasa de no logro (TNL) de EANNA 2011 respecto de Casen 2011 y el máximo de la TNL entre Casen 2017 y 2022 ponderando, un 60% EANNA y, un 40% Casen. Esta tasa se encuentra a nivel regional y se aplica a cada estrato de muestreo de EANNA.

Luego, el tamaño con sobremuestreo a nivel de estrato m_{3em} es calculado como se observa en la ecuación (6):

$$m_{3em} = \frac{m_{2em}}{1 - tnr_{em}} \quad (6)$$

Donde:

m_{3em} : Número total de NNA a encuestar en el estrato de muestreo em .

tnr_{em} : Tasa de no respuesta aplicada en el estrato de muestreo em .

Paso 7: Estimación de la tasa de no respuesta a niveles agregados

Dado que el tamaño de muestra regional con sobremuestreo m_{3r} , es calculado como la suma o agregación de estratos de muestreo em , del total de NNA de cada una de las regiones, se puede despejar la tasa de no logro y estimar la tasa regional a partir de la ecuación (7):

$$tnr_r = \frac{m_{3r} - m_{2r}}{m_{3r}} \quad (7)$$

De forma análoga se estima la tasa de no logro a nivel nacional.

III.2. Parámetros utilizados para el cálculo del tamaño muestral

Para el cálculo del tamaño muestral se consideran los elementos presentados en el Cuadro 1:

Cuadro 1. Parámetros utilizados para el cálculo de tamaño muestral actual

Parámetro	Descripción
Parámetro asociado	Tasa de trabajo infantil (TI) en la frontera de producción (porcentaje de NNA en situación de TI)
Estimador asociado	Estimador de razón de trabajo infantil (TI) en la frontera de producción (porcentaje de NNA en situación de TI) según resultados obtenidos de EANNA 2012: $ti = \frac{\text{Número de NNA en TI}}{\text{Número total de NNA}}$
Tamaño de la población objetivo	Población estimada de niños, niñas y adolescentes (NNA) y de los hogares que los contienen por región, área y tramos etario (5 a 8 años, 9 a 11 años, 12 a 14 años y 15 a 17 años) a partir de las proyecciones de población oficiales para el periodo de recolección de la encuesta (15 de junio de 2023).
Variable de diseño	$X_j = \begin{cases} 0, & \text{si el NNA no está en situación de TI} \\ 1, & \text{si el NNA está en situación de TI} \end{cases}$

Parámetro	Descripción
Niveles de estimación	Nacional Nacional urbano, Nacional rural Regional
Errores de muestreo	Errores absolutos de acuerdo con el estándar para la evaluación de la calidad de las estimaciones en encuestas de hogares (INE, 2020), cuya definición de aceptación se establece mediante el siguiente umbral: $\text{Máximo ee tolerable} = \begin{cases} \sqrt[3]{r^2/9} & ; r < 0,5 \\ \sqrt[3]{(1-r)^2/9} & ; r \geq 0,5 \end{cases}$
Cuasivarianza	EANNA 2012
Efecto del diseño	Se utiliza el máximo entre: 1. El efecto obtenido en EANNA 2012 2. El efecto obtenido para la tasa de pobreza en viviendas con NNA en Casen Piloto cambio de marco (CM). 3. El efecto utilizado para el cálculo del tamaño muestral de Casen 2022. Luego, se define un mínimo de 1,5 y un máximo de 3,0
Tasa de sobremuestreo	Obtenida como el máximo entre la tasa de Casen 2022, ponderada por la proporción de la TNL de EANNA 2012 respecto de Casen 2011 y el máximo de la TNL entre Casen 2017 y 2022, ponderando 60% con la TNL de EANNA 2012 y 40% para Casen.

Elaborado por el Instituto Nacional de Estadísticas.

III.3. Metodología de cálculo del tamaño muestral utilizando muestras complejas

La población objetivo, los niveles de estimación deseados y la cobertura geográfica de la encuesta EANNA 2023, hacen necesaria la confección de un diseño eficiente para la selección de una muestra de NNA, en términos de ejecución presupuestaria e implementación operativa.

El muestreo aleatorio simple de viviendas necesitaría, en principio, de un marco de NNA con sus respectivas direcciones de domicilios, generado por un censo de población y viviendas que, por un lado, resultaría muy costoso de mantener actualizado y, por otro, los NNA seleccionados de un marco de este tipo, se encontrarían tan dispersos geográficamente que, los costos asociados al levantamiento excederían la asignación presupuestaria, por tanto, el muestreo aleatorio simple iría en desmedro de la eficiencia deseada.

Por otra parte, en los diseños de muestra complejos, las unidades que finalmente componen la muestra (en este caso NNA) se determinan después de varias etapas; en consecuencia, no es necesario mantener un marco de NNA actualizado, lo cual reduce significativamente los costos.

La conformación de estratos de muestreo, la conglomeración de unidades muestrales y la definición de las probabilidades de inclusión, caracterizan los diseños de muestra complejos (Heeringa, West, & Berglund, 2010, pág. 13). El diseño de Casen presenta elementos que lo complejizan, específicamente, por el hecho que sea probabilístico, estratificado y bietápico.

Mientras, algunos de estos elementos tienden a reducir la varianza, como la estratificación, otros tienden a aumentarla, como la conglomeración. Las distintas etapas de muestreo tienden a aumentar la varianza mientras que, la disminución del tamaño de los conglomerados tiende a disminuirla. También es necesario considerar otros efectos tales como la distribución de la prevalencia de la pobreza en las unidades secundarias de muestreo o como la correlación de las unidades secundarias al interior de los conglomerados, respecto a la característica de interés, llamada correlación intraclásica, que afecta la varianza y es el principal factor en la medición del efecto del diseño o efecto de no considerar un muestreo aleatorio simple (Kish, 1965).

El cociente entre la estimación de la varianza bajo un diseño complejo versus la estimación bajo un diseño aleatorio simple se denomina efecto del diseño y, en términos de tamaño muestral, equivale a cuánto se debe aumentar la muestra bajo diseño complejo para alcanzar el mismo nivel de precisión que se obtendría si aplicáramos un muestreo aleatorio simple.

Luego, para determinar el tamaño de la muestra compleja se incorpora este factor de ajuste a partir del tamaño de una muestra aleatoria simple, lo que permite aproximarse al número de viviendas necesarias para que el diseño complejo proporcione la misma varianza (Kish, 1965).

En este contexto, la metodología de cálculo del tamaño muestral utilizada en EANNA 2023 se inicia con la determinación de un tamaño muestral bajo un muestreo aleatorio simple, fijando previamente el error absoluto a alcanzar con base en los resultados de encuestas anteriores. Posteriormente, se realizan tres ajustes secuenciales: el primero, que da cuenta del efecto del diseño; el segundo, que corresponde a un ajuste por población finita y; el tercero, referido al ajuste por la tasa de no logro.

La obtención de tamaños muestrales para EANNA 2023 comienza con la determinación del tamaño bajo muestreo aleatorio simple y, a partir de este tamaño se obtiene un error de estimación. Luego, utilizando efectos que dan cuenta del diseño complejo, se calculan tamaños necesarios que permitan obtener estimaciones con el grado de precisión calculado en el paso anterior. Los efectos de diseño utilizados están basados en: los resultados del levantamiento del Piloto Cambio de Marco de Casen, los utilizados en el diseño de Casen 2022 y los efectos obtenidos en el levantamiento de EANNA 2012.

Además, la corrección por finitud de la población es estimada a partir de la población completa sobre la que se desea realizar inferencias (el universo de estudio) y por esta razón se utilizan las proyecciones de población.

De este modo, para EANNA 2023 se obtuvieron tamaños de muestra regionales, asegurando representatividad a nivel nacional, nacional urbano, nacional rural y regional.

Posteriormente, se realiza un procedimiento mediante el cual se distribuye el tamaño de muestra de cada región por el estrato de muestreo que incluye el área y tramo etario donde, se busca, por un lado, resguardar coherencia entre los tamaños muestrales y aquellos de la población (según el marco muestral utilizado) en términos de proporcionalidad y, por otro, que dicha distribución no difiriera mayormente de la distribución de tamaños de las proyecciones de población al 15 de junio de 2023.

Bajo esta distribución, los tamaños con sobremuestreo son calculados a partir de una tasa de no logro que considera el máximo entre la tasa de no logro de Casen 2022, ponderada por la proporción de la TNL de EANNA 2012 respecto de Casen 2011 y el máximo de la TNL entre Casen 2017 y 2022, ponderando 60% para EANNA y 40% para Casen. Finalmente, para que la muestra sea eficiente, en términos de la cantidad de NNA a seleccionar, la fracción de muestreo respecto del marco debe considerar el establecimiento de cotas (especialmente superiores) que, idealmente no debieran superar 70% a nivel regional. Sin embargo, se hace necesario también la revisión de las fracciones de muestreo a niveles más desagregados, es decir, desagregación por área (urbana – rural) y por estratos de muestreo (región - área - tramo etario).

III.4. Resumen proceso de simulación

En este apartado se presentan las simulaciones de tamaños muestrales derivadas de distintos escenarios que fueron trabajados inicialmente, hasta llegar a la propuesta elegida. La estructura del capítulo mantiene la cronología de las simulaciones realizadas, así como también, la metodología empleada en ellas.

Como base para el desarrollo de todos los escenarios trabajados se presentan los resultados del trabajo de campo de EANNA 2012, desde donde se obtiene la estimación de la tasa de trabajo infantil junto con sus estadísticos asociados, a saber, el error estándar, el coeficiente de variación y el efecto de diseño, tanto a nivel nacional como regional.

A partir de estos resultados, fueron analizados seis escenarios que se describen a continuación:

- Escenario MDSF: Este escenario fue formulado por el MDSF y utiliza como parámetros base los resultados de EANNA 2012, sin embargo, no considera el efecto que provoca el nuevo diseño muestral de la encuesta, así como tampoco la corrección por población finita. Además, en la obtención de los errores de muestreo no se consideran las orientaciones del estándar de calidad que utiliza el INE.
- Escenario n°0: Utiliza como marco muestral los NNA de Casen en Pandemia al no contar, al momento de la simulación, con los resultados de Casen 2022 y, a su vez, presenta información relativa a los parámetros de EANNA 2012 desde donde se obtuvo una estimación de la tasa de trabajo infantil nacional de 6,6%, asociada a un error absoluto de 0,8% y a un error relativo de 11,8%.
- Escenario n°1: Esta simulación considera como marco muestral la estimación de los NNA de Casen en Pandemia 2020 y se utilizan los totales de NNA logrados en Casen en Pandemia 2020 para el cálculo de la fracción de muestreo.
- Escenario n°1.1: Mantiene los errores y efectos del diseño utilizados en el escenario n°1 y la modificación se encuentra en el insumo utilizado para calcular la corrección por finitud ya que en este caso se utilizan las proyecciones de población al 15 de enero de 2023.
- Escenario n°2: Utiliza como insumo para la corrección por población finita las proyecciones de población al 15 de junio de 2023 y los resultados del levantamiento de Casen 2022 para calcular las fracciones de muestreo. Además, se realizan ajustes a los efectos del diseño definidos en los escenarios anteriores para establecer un mínimo de 1,5 y un máximo de 3,0. Además, este escenario presenta un ajuste en la distribución urbano-rural debido a que no todos los estratos poseen suficientes NNA para seleccionar.
- Escenario n°2.1: Este escenario presenta una diferencia con respecto al escenario n° 2 y se refiere a una actualización del marco muestral a utilizar para el cálculo de las fracciones de muestreo. La fecha de actualización por parte de MDSF corresponde al 16 de marzo de 2023, que fue recibida por el INE el 22 de marzo de 2023.
- Escenario n°2.2: Este escenario es el seleccionado. Utiliza los mismos parámetros que el escenario n°2, sin embargo, las fracciones de muestreo son obtenidas a partir del marco actualizado por MDSF al 11 de abril de 2023 y recibido por el INE, el 20 de abril de 2023.

Obtenidos los tamaños muestrales objetivos del escenario elegido (escenario n°2.2) se realiza una comparación entre su distribución y la distribución de las proyecciones de población a nivel regional, con la finalidad de mostrar que no existen mayores diferencias con dichos referentes.

El Cuadro 2 presenta los parámetros comunes utilizados en los escenarios considerados.

Cuadro 2: Parámetros comunes utilizados en escenarios EANNA 2023

Parámetro	Descripción
Parámetro a estimar	Tasa de trabajo infantil (TI)
Variable de diseño	$X_j = \begin{cases} 0, & \text{si el NNA no está en situación de TI} \\ 1, & \text{si el NNA está en situación de TI} \end{cases}$
Estimador asociado	Estimador de razón: $t_i = \frac{\text{Número de NNA que trabajan}}{\text{Número total de NNA}}$

Elaborado por el Instituto Nacional de Estadísticas.

La Tabla 1 presenta un resumen de los parámetros utilizados en cada una de las simulaciones, así como también, los tamaños y errores muestrales resultantes, por orden cronológico, comenzando con el escenario propuesto por el MDSF, para seguir con los resultados del trabajo de campo de EANNA 2012 (escenario n°0) y luego, presentar el resto de los escenarios.

Tabla 1. Resumen de parámetros utilizados y tamaños y errores resultantes según escenario, EANNA 2023

Nro.	N° NNA según marco utilizado	Niveles de estimación	Error Absoluto	Error Relativo	Efectos de Diseño	Tamaño Muestral Objetivo	Tamaño Muestral Objetivo (ajustado) ⁶	Tasa de no logro	Tamaño con sobremuestreo
MDSF		Nacional		8,5%			18.470	TNR EANNA 2012 corregida por la variación entre la TNR de Casen 2011 y el máximo entre la TNR Casen 2017 y la TNR Cambio Marco	21.782
		Regional		18,2% - 45,5%					
0	32.499	Nacional	0,8%	11,8%	2,5	6.830			
		Regional	1,4% - 6,3%	24,2% - 61,2%	0,7 - 3,4				
1	3.435.717	Nacional	0,6%	8,5%	2,1	15.938	15.940	TNR EANNA 2012 corregida por la variación entre la TNR de Casen 2011 y el máximo entre la TNR Casen 2017 y la TNR Cambio Marco	19.661
		Regional	1,4% - 3,5%	21,0% - 51,0%	1,0 - 3,0				
1.1	3.290.782	Nacional	0,6%	8,5%	2,1	15.921	15.922	TNR EANNA 2012 corregida por la variación entre la TNR de Casen 2011 y el máximo entre la TNR Casen 2017 y la TNR Cambio Marco	19.625
		Regional	1,4% - 3,5%	21,0% - 51,0%	1,0 - 3,0				
2 - 2.1	3.287.283	Nacional	0,6%	8,5%	2,1	18.066	18.069	Máximo entre la TNL de Casen 2022 ponderada por la proporción de la TNL de EANNA 2012 respecto de Casen 2011 y el máximo de la TNL entre Casen 2017 y 2022 ponderando un 60% EANNA y un 40% Casen.	23.489
		Regional	1,4% - 3,4%	21,0% - 51,1%	1,0 - 3,0				
2.2	3.287.283	Nacional	0,6%	8,5%	2,1	18.066	18.069	Máximo entre la TNL de Casen 2022 ponderada por la proporción de la TNL de EANNA 2012 respecto de Casen 2011 y el máximo de la TNL entre Casen 2017 y 2022 ponderando un 60% EANNA y un 40% Casen.	25.611
		Regional	1,4% - 3,4%	21,0% - 51,1%	1,0 - 3,0				

Elaborado por el Instituto Nacional de Estadísticas

⁶ Los errores absolutos y relativos son calculados con el tamaño objetivo ajustado.

III.5. Escenario elegido

Tabla 2: Estimación de tamaños muestrales según proyecciones de población al 15 de junio de 2023 con marco Casen 2022 actualizado⁷

Región	NNA Casen 2022	Proyección NNA 15/06/23	Tasa TI	Efecto Diseño Ajustado	Tamaño ajustado efecto diseño	Error absoluto	Error Relativo	Tamaño ajustado por finitud	Tamaño Objetivo Ajustado	Error Absoluto ajustado	Error Relativo tamaño ajustado	Tamaño con Sobre-muestreo
Nivel País	33.989	3.287.283	6,6%	2,1	18.188	0,6%	8,5%	18.066	18.069	0,6%	8,5%	25.611
Urbano	27.622	2.926.900	6,1%	2,1	16.187	0,5%	8,2%	16.079	15.743	0,5%	8,3%	21.828
Rural	6.367	360.383	9,7%	2,1	2.001	1,9%	19,4%	1.987	2.326	1,7%	18,0%	3.783
Arica y Parinacota	1.490	46.845	4,9%	1,8	652	2,3%	46,1%	644	644	2,3%	46,3%	1.041
Tarapacá	1.795	76.333	4,5%	1,7	717	2,0%	44,7%	711	711	2,0%	44,9%	1.066
Antofagasta	1.667	125.710	5,5%	1,7	774	2,1%	37,2%	770	770	2,1%	37,3%	1.239
Atacama	1.701	61.560	5,2%	1,7	672	2,2%	42,3%	665	665	2,2%	42,5%	912
Coquimbo	1.420	154.893	5,3%	1,9	908	2,1%	38,9%	903	903	2,1%	39,0%	1.362
Valparaíso	3.303	323.958	5,1%	2,1	2.098	1,4%	27,5%	2.085	2.085	1,4%	27,6%	2.879
Metropolitana	6.336	1.306.947	6,7%	3,0	3.798	1,4%	21,0%	3.787	3.787	1,4%	21,0%	5.555
O'Higgins	2.213	173.978	8,7%	1,8	999	2,4%	27,6%	994	995	2,4%	27,6%	1.456
Maule	2.186	197.757	4,8%	1,7	1.221	1,5%	31,5%	1.214	1.215	1,5%	31,6%	1.701
Ñuble	1.312	86.134	10,3%	2,4	775	3,4%	33,1%	769	769	3,4%	33,2%	927
Biobío	3.234	282.035	4,4%	2,6	1.916	1,5%	34,3%	1.904	1.905	1,5%	34,4%	2.302
La Araucanía	2.244	179.839	11,6%	2,1	1.029	2,9%	25,1%	1.024	1.024	2,9%	25,2%	1.464
Los Ríos	1.852	68.634	11,9%	1,5	693	3,0%	25,1%	687	687	3,0%	25,2%	1.127
Los Lagos	1.716	153.757	8,0%	2,6	887	2,9%	36,3%	882	882	2,9%	36,4%	1.293
Aysén	732	20.263	7,7%	1,5	504	2,9%	37,6%	492	492	2,9%	38,0%	613
Magallanes	788	28.640	3,0%	1,5	545	1,5%	50,6%	535	535	1,5%	51,1%	674

Elaborado por el Instituto Nacional de Estadísticas.

⁷ Marco actualizado con fecha 11 de abril de 2023 y puesto a disposición del INE el día 20 de abril de 2023.

Este escenario se elabora con base en los parámetros obtenidos del trabajo de campo de EANNA 2012 (escenario n°0). La diferencia con los escenarios anteriores se encuentra en que el marco de referencia corresponde a 33.989 NNA que fueron resultado del levantamiento de Casen 2022. El error absoluto propuesto es ajustado respecto al escenario n°1 para cumplir con el estándar de calidad y el orden de la distribución regional de la población. Con errores absolutos entre 1,4% y 3,4% a nivel regional se obtiene un tamaño bajo M.A.S. de 8.072 NNA a nivel nacional.

Para obtener el efecto de diseño se contempla el efecto de diseño de EANNA 2012, el efecto de diseño utilizado para Casen 2022 y el efecto de diseño resultante del cálculo de la tasa de pobreza en hogares con NNA del Piloto Cambio de Marco Casen 2022 donde, el efecto de diseño propuesto va a corresponder al máximo entre estos tres valores, con excepción de las regiones de Coquimbo, Metropolitana, Ñuble, Los Ríos, Aysén y Magallanes. A nivel regional se establece que los efectos de diseño utilizados fluctúen entre 1,5 y 3,0.

Luego, el tamaño objetivo ajustado por efecto de diseño es de 18.188 NNA que, al aplicar el ajuste por finitud, considerando como universo las proyecciones de población al 15 de junio de 2023, resulta un tamaño de 18.066 NNA a nivel nacional.

La distribución de los tamaños por estrato de muestreo (región-área-tramo etario) se realiza en forma proporcional a las proyecciones de población a junio de 2023 y se establece un mínimo de 10 NNA por cada uno, aumentando la muestra a 18.069 NNA con un error absoluto de 0,6% y un error relativo de 8,5%. Si bien a nivel regional se observan errores relativos elevados, a partir del error estándar se aprecia que estos valores cumplen con el estándar de calidad.

Finalmente, para obtener el tamaño con sobremuestreo, se utiliza una tasa de no logro a nivel regional donde se selecciona el máximo entre: la tasa de no logro de Casen 2022, ponderada por la proporción de la TNL de EANNA 2012 respecto de Casen 2011 y el máximo de la TNL entre Casen 2017 y Casen 2022, que se pondera con EANNA 2012 que equivale a 60% y Casen a 40%. Este procedimiento dio lugar a un tamaño con sobremuestreo de 25.611 NNA.

III.6. Tamaño muestral propuesto

El tamaño calculado a nivel regional se debe distribuir por área (urbana – rural) y, al interior de estas, por los tramos etarios definidos según una distribución conocida. Para esto, se cuenta con la distribución resultante de Casen 2022 o bien, con las proyecciones de población estimadas al 15 de junio de 2023 (fecha que coincide con el periodo de levantamiento de la encuesta).

El tamaño objetivo ajustado del escenario elegido (n°2.2) definido para obtener estimaciones de la tasa de trabajo infantil con suficiente grado de precisión a nivel regional, alcanza 18.069 NNA a nivel nacional, cuyos errores absolutos y relativos asociados son de 0,6% y 8,5%, respectivamente. Aplicando las tasas de no logro a nivel regional con ciertos ajustes (con el objetivo de que el marco muestral cuente con suficientes unidades para seleccionar la muestra) se obtiene el tamaño con sobremuestreo que alcanza 25.611 NNA.

La Tabla 3 muestra los tamaños objetivo ajustados junto con sus errores muestrales asociados y el tamaño con sobremuestreo a nivel nacional y regional, desagregados por área urbano-rural.

Tabla 3: Tamaños y errores finales a nivel nacional, regional, urbano-rural, EANNA 2023

Región	Error Absoluto	Error Relativo	Tamaño objetivo			Tamaño con sobremuestreo		
			Total	Urbano	Rural	Total	Urbano	Rural
Nivel País	0,6%	8,5%	18.069	15.743	2.326	25.611	21.828	3.783
Arica y Parinacota	2,3%	46,3%	644	595	49	1.041	932	109
Tarapacá	2,0%	44,9%	711	671	40	1.066	1.000	66
Antofagasta	2,1%	37,3%	770	730	40	1.239	1.149	90
Atacama	2,2%	42,5%	665	617	48	912	847	65
Coquimbo	2,1%	39,0%	903	760	143	1.362	1.088	274
Valparaíso	1,4%	27,6%	2.085	1.906	179	2.879	2.474	405
Metropolitana	1,4%	21,0%	3.787	3.627	160	5.555	5.177	378
O'Higgins	2,4%	27,6%	995	779	216	1.456	1.103	353
Maule	1,5%	31,6%	1.215	940	275	1.701	1.317	384
Ñuble	3,4%	33,2%	769	567	202	927	683	244
Biobío	1,5%	34,4%	1.905	1.703	202	2.302	2.058	244
La Araucanía	2,9%	25,2%	1.024	743	281	1.464	1.062	402
Los Ríos	3,0%	25,2%	687	504	183	1.127	808	319
Los Lagos	2,9%	36,4%	882	685	197	1.293	983	310
Aysén	2,9%	38,0%	492	425	67	613	529	84
Magallanes	1,5%	51,1%	535	491	44	674	618	56

Elaborado por el Instituto Nacional de Estadísticas

Luego, a partir de las desagregaciones de tamaños por estrato de muestreo (región – área – tramo etario) y de los resultados del trabajo de campo de EANNA 2012 se obtienen, por agregación, tamaños y errores asociados para los niveles nacional urbano y nacional rural que se muestra en la Tabla 4.

Tabla 4: Tamaños y errores EANNA 2012 y propuesta EANNA 2023. Nacional Urbano - Rural

Región	Resultados EANNA 2012				Propuesta EANNA 2023			
	NNA logrados	Tasa TI	Error Absoluto	Error Relativo	Tamaño Objetivo	Error Absoluto	Error Relativo	Tamaño con Sobremuestreo
Nivel País	9.978	6,6%	0,8%	11,8%	18.069	0,6%	8,5%	25.611
Urbano	7.859	6,1%	0,8%	13,1%	15.743	0,5%	8,2%	21.828
Rural	2.119	9,7%	1,6%	16,5%	2.326	1,9%	19,4%	3.783

Elaborado por el Instituto Nacional de Estadísticas.

A nivel nacional urbano, con un tamaño objetivo ajustado de 15.743 NNA se obtiene una estimación para la tasa de TI con un error absoluto de 0,5% y un error relativo de 8,2%. Por otra parte, a nivel nacional rural, con un tamaño objetivo de 2.326 NNA se obtiene una estimación para la tasa de TI con un error absoluto de 1,9% y un error relativo de 19,4%.

III.7. Errores esperados

La Tabla 5 presenta los errores esperados, asociados a la estimación del parámetro de interés para un tamaño muestral objetivo de 18.069 NNA⁸.

Tabla 5: Errores esperados asociados al parámetro de interés según tamaño muestral objetivo

Región	Tasa TI EANNA 2012	Error Absoluto Propuesto EANNA 2023	Error Relativo Propuesto EANNA 2023	Tamaño ajustado por finitud EANNA 2023	Tamaño objetivo ajustado EANNA 2023	Tamaño con sobremuestreo EANNA 2023
Nacional	6,6%	0,6%	8,5%	18.066	18.069	25.611
Arica y Parinacota	4,9%	2,3	46,1	644	644	1.041
Tarapacá	4,5%	2,0	44,7	711	711	1.066
Antofagasta	5,5%	2,1	37,2	770	770	1.239
Atacama	5,2%	2,2	42,3	665	665	912
Coquimbo	5,3%	2,1	38,9	903	903	1.362
Valparaíso	5,1%	1,4	27,5	2.085	2.085	2.879
Metropolitana	6,7%	1,4	21,0	3.787	3.787	5.555
O'Higgins	8,7%	2,4	27,6	994	995	1.456
Maule	4,8%	1,5	31,5	1.214	1.215	1.701
Ñuble	10,3%	3,4	33,1	769	769	927
Biobío	4,4%	1,5	34,3	1.904	1.905	2.302
La Araucanía	11,6%	2,9	25,1	1.024	1.024	1.464
Los Ríos	11,9%	3,0	25,1	687	687	1.127
Los Lagos	8,0%	2,9	36,3	882	882	1.293

⁸ La estimación de los errores se obtiene a partir del tamaño objetivo y no del tamaño con sobremuestreo, pues dadas las pérdidas naturales de unidades muestrales, se espera obtener un tamaño de 18.069 NNA.

Región	Tasa TI EANNA 2012	Error Absoluto Propuesto EANNA 2023	Error Relativo Propuesto EANNA 2023	Tamaño ajustado por finitud EANNA 2023	Tamaño objetivo ajustado EANNA 2023	Tamaño con sobremuestreo EANNA 2023
Aysén	7,7%	2,9	37,6	492	492	613
Magallanes	3,0%	1,5	50,6	535	535	674

Elaborado por el Instituto Nacional de Estadísticas.

IV. SELECCIÓN DE UNIDADES MUESTRALES

La selección de unidades muestrales se realiza desde el marco muestral actualizado al 11 de abril de 2023, conformado por NNA logrados en Casen 2022, en forma sistemática y con igual probabilidad al interior de cada estrato de muestreo (región - área - tramo etario) contemplando aquellos registros de NNA entre 5 y 17 años cumplidos a junio de 2023.

La selección de NNA es implementada en el *software* R, utilizando el paquete “Sampling”, fijando una semilla aleatoria, a fin de que pueda ser replicable en cualquier momento.

Sean M_{em} el número de NNA encontrado en el estrato y m_{em}^{sel} el número de NNA a seleccionar. Para la selección de m_{em}^{sel} NNA, el *software* ejecuta los siguientes pasos:

Paso 1

En primera instancia se ordenan geográficamente y en forma ascendente todos los NNA según estrato de muestreo, comuna, sexo e identificador de la UPM⁹.

Paso 2

Al interior de cada estrato de muestreo se calcula el período (k) que corresponde a:

$$k = M_{em} / m_{em}^{sel}$$

Notar que “ k ” puede ser un número real, no entero (puede tener decimales).

Paso 3

Luego se determina el arranque “ A ” o primera selección, que corresponde a una semilla aleatoria propia para la encuesta.

⁹ En la sintaxis enviada a MDSF la variable “correlativo” replica este ordenamiento.

Paso 4

Posteriormente se suma sucesivamente el período " k " al arranque " A " para obtener distintos valores, los que dan origen a la selección de unidades de la siguiente forma: " A ", " $A + k$ ", " $A + 2k$ ", " $A + 3k$ ", ..., " $A + (m_{em}^{sel} - 1)k$ ".

El primer NNA seleccionado(a) es " A " y es un número entero, el segundo es el redondeo de " $A + k$ ", el tercero es el redondeo de " $A + 2k$ " y así sucesivamente, hasta la m_{em}^{sel} selección, dada por el redondeo de " $A + (m_{em}^{sel} - 1)k$ ".

Luego, la probabilidad de inclusión del j –ésimo NNA dentro del estrato em está dada por:

$$P_{re}(j) = \frac{m_{em}^{sel}}{M_{em}} \tag{1}$$

Donde:

m_{em}^{sel} : Número de NNA seleccionados en el estrato em

M_{em} : Número actualizado de NNA en el marco de selección en el estrato em .

La Tabla 6 presenta la distribución del total de NNA seleccionados, según región y área (Urbano - Rural) obtenida a partir de la distribución del marco de NNA logrados en Casen 2022.

Tabla 6: Distribución del número de NNA por región y área

Región	Área	Total NNA	
		Marco Casen 2022	Selección EANNA 2023
Total País		33.989	25.611
Arica y Parinacota	Urbano	1.294	932
	Rural	196	109
Tarapacá	Urbano	1.659	1.000
	Rural	136	66
Antofagasta	Urbano	1.519	1.149
	Rural	148	90
Atacama	Urbano	1.531	847
	Rural	170	65
Coquimbo	Urbano	1.104	1.088
	Rural	316	274
Valparaíso	Urbano	2.696	2.474
	Rural	607	405
Metropolitana	Urbano	5.812	5.177
	Rural	524	378
O'Higgins	Urbano	1.626	1.103
	Rural	587	353
Maule	Urbano	1.605	1.317
	Rural	581	384
Ñuble	Urbano	874	683
	Rural	438	244

Región	Área	Total NNA	
		Marco Casen 2022	Selección EANNA 2023
Biobío	Urbano	2.584	2.058
	Rural	650	244
La Araucanía	Urbano	1.431	1.062
	Rural	813	402
Los Ríos	Urbano	1.308	808
	Rural	544	319
Los Lagos	Urbano	1.268	983
	Rural	448	310
Aysén	Urbano	606	529
	Rural	126	84
Magallanes	Urbano	705	618
	Rural	83	56

Elaborado por el Instituto Nacional de Estadísticas.

V. DISEÑO FACTORES DE EXPANSIÓN

Los análisis y estudios basados en encuestas de hogares con diseño muestral probabilístico utilizan un ponderador para la estimación de parámetros de interés con el objetivo que estos reflejen las características del diseño muestral y, además, tengan validez sobre la población de estudio. Este ponderador guarda relación con las probabilidades de selección de las distintas unidades de muestreo y da cuenta del número de personas de la población que representa cada participante en dicha encuesta. Este ponderador es conocido como factor de expansión.

En EANNA 2023 se desarrollan factores para cada NNA presente en la muestra y así representar el universo desde el cual se seleccionaron las unidades. Su construcción se asocia a un diseño muestral bifásico. Esto quiere decir que, en la primera fase, las unidades muestrales son seleccionadas desde los hogares logrados en Casen 2022 y, en la segunda fase, a partir de un diseño estratificado, se seleccionan NNA en forma sistemática y con igual probabilidad al interior de cada estrato de muestreo.

En detalle, la metodología de cálculo de los factores de expansión consiste en la obtención secuencial de una serie de ponderadores que se inicia con el factor de no respuesta de las viviendas logradas en Casen 2022 y que corresponde a la primera fase del diseño. Luego, se obtiene el factor de segunda fase que refleja las probabilidades de selección del NNA presente en la muestra de EANNA.

A continuación, se obtiene el factor de selección de NNA a partir del producto entre el factor de primera y segunda fase para luego, aplicar una serie de ajustes según elegibilidad y no respuesta, que tiene como objetivo reducir el sesgo asociado a la falla en conseguir la cooperación de las

personas seleccionadas en la muestra. Dadas las múltiples etapas para la obtención del factor, es necesario realizar un suavizamiento que permite acotar la variabilidad producida para finalmente, aplicar la calibración con el fin de cuadrar las estimaciones poblacionales obtenidas de la encuesta con las proyecciones poblacionales desarrolladas por el INE y que son utilizadas en todas las encuestas de hogares.

Para realizar la calibración se utiliza la técnica *Raking*, que será descrita en los próximos apartados, que permite controlar las estimaciones por diversas variables simultáneamente, manteniendo el diseño muestral y disminuyendo el sesgo que se produce por errores no muestrales, provocados por la no respuesta y/o la imperfección de los marcos de selección, mejorando además la precisión de las estimaciones. Para realizar la calibración, se utiliza la información auxiliar proveniente de las Proyecciones de Población según sexo y tramo etario y para distintos niveles de desagregación.

La Figura 1 muestra el flujo del proceso del cálculo de los factores de expansión para EANNA 2023. Luego, a lo largo del capítulo se detalla el proceso de construcción de los factores de expansión donde cada apartado hace referencia a las rutinas correspondientes a cada ponderador mencionado.

Figura 1. Diagrama de procesos de desarrollo de factor de expansión EANNA 2023



Elaborado por el Instituto Nacional de Estadísticas¹⁰.

¹⁰ VPO corresponde a Viviendas Particulares Ocupadas.

V.1. Factor de expansión de primera fase

En Casen 2022 las unidades muestrales se obtienen desde el MMV 2020 a partir de un diseño probabilístico, estratificado y bietápico, donde los estratos muestrales están conformados por la combinación Comuna-Área-NSE, donde aplique, y los conglomerados son seleccionados en forma sistemática con probabilidad proporcional al tamaño, según el número de viviendas que los conforman. Por su parte, las viviendas o unidades secundarias de muestreo son seleccionadas en forma aleatoria y con igual probabilidad al interior de cada UPM elegida.

El cálculo del factor de expansión de primera fase de EANNA considera el diseño muestral de Casen 2022 donde, al factor de selección se le aplica un ajuste por elegibilidad y un ajuste por no respuesta redistribuyendo los pesos de las viviendas que no respondieron Casen entre las que sí lo hicieron.

Para EANNA 2023 el factor de primera fase corresponde al ponderador ajustado por no respuesta de Casen 2022 (w_{hij}^{NR}) obtenido según la ecuación (9):

$$F1_{em,k} = w_{hij}^{NR} = \hat{R}_{g,R} \cdot w'_{hij} \quad (9)$$

Donde,

- em, k : Índice del estrato de muestreo de EANNA (región - área - tramo etario) y de la persona seleccionada, respectivamente.
- w_{hij}^{NR} : Ponderador de selección de viviendas de Casen 2022 ajustado por no respuesta.

V.2. Factor de expansión de segunda fase

Dado que el diseño muestral de EANNA es bifásico, luego de obtener el ponderador de no respuesta desde Casen 2022 (que corresponde a la primera fase), se deben calcular los ponderadores de selección de las unidades participantes en EANNA además de una serie de ajustes propios de la encuesta y que corresponden a la segunda fase. Estos se detallan a continuación:

V.2.1. Ponderador de selección de NNA como inverso de la probabilidad de selección

Este ponderador corresponde al inverso de la probabilidad de selección del NNA para la muestra de EANNA según el estrato de muestreo (región - área - tramo etario). Este ponderador se interpreta como el número de personas entre 5 y 17 años, en la población, que representa el NNA seleccionado(a) en la muestra. Se debe considerar que, a diferencia de los ponderadores de personas desarrollados en Casen, el diseño muestral de la EANNA permite que NNA que residen en

una misma vivienda puedan tener distintos ponderadores, debido a que pertenecen a tramos etarios distintos dados por su edad cumplida a junio de 2023.

El ponderador de selección del NNA es calculado como el inverso de la probabilidad de selección del NNA k perteneciente al estrato em . Esto se observa en la ecuación (10):

$$F2_{em,k} = \frac{M_{em}}{m_{em,sel}} \quad (10)$$

Donde,

em, k : Índice del estrato de muestreo (región - área - tramo etario) y del NNA k seleccionado(a).

$F2_{em,k}$: Ponderador de selección de la persona k en el estrato em .

M_{em} : Total de personas elegibles en el estrato de muestreo, según el levantamiento de Casen 2022.

$m_{em,sel}$: Total de personas seleccionadas en EANNA 2023 en el estrato em .

V.2.2. Producto entre los ponderadores de primera y segunda fase

Este ajuste da cuenta de la selección de personas considerando ambas fases de muestreo, es decir, contempla la selección de la k -ésima persona para EANNA, dado que fue lograda en Casen 2022. Este factor corresponde al factor de selección y se expresa según la ecuación (11):

$$w_{em,k} = F1_{em,k} * F2_{em,k} \quad (21)$$

Donde,

$w_{em,k}$: Ponderador de selección de personas considerando la primera y segunda fase.

V.2.3. Ajuste por elegibilidad desconocida

La muestra seleccionada incluye, en principio, solo personas elegibles, considerando que el marco de selección de EANNA es construido a partir de las viviendas logradas en Casen 2022. Sin embargo, durante el trabajo de campo existen situaciones como la imposibilidad de acceder a la vivienda de la persona seleccionada por razones climáticas o desastres naturales, por tratarse de áreas peligrosas o por no localizar la dirección, entre otros motivos que llevan a que el NNA seleccionado(a) sea clasificado como de elegibilidad desconocida.

En los casos donde la persona seleccionada es clasificada con elegibilidad desconocida sus pesos son redistribuidos proporcionalmente entre las personas que resultaron con elegibilidad conocida (elegibles y no elegibles) al interior del estrato de muestreo.

La proporción de los pesos a redistribuir $\hat{R}_{em,known}$ se calcula como la razón entre dos estimaciones realizadas con base en el ponderador de selección de personas.

El numerador corresponde a la estimación del total de personas en la población y el denominador corresponde a la estimación del total de personas con elegibilidad conocida. Esto se observa en la ecuación (12):

$$\hat{R}_{em,known} = \frac{\sum_{\epsilon \in \Omega_{em}} \sum_{\epsilon \in \theta_k} w_{em,k}}{\sum_{\epsilon \in \Omega_{em}} \sum_{\epsilon \in \theta_{k,known}} w_{em,k}} \quad (12)$$

Donde,

- θ_k : : Conjunto de personas k seleccionadas.
- $\theta_{k,known}$: Conjunto de personas k seleccionadas y clasificadas como elegibilidad conocida.
- Ω_{em} : Conjunto de estratos de muestreo em .
- $w_{em,k}$: Ponderador de selección de la persona k en el estrato em .

La razón antes descrita es la misma para todas las unidades muestrales pertenecientes a un mismo estrato de muestreo. Luego, el ponderador de selección del NNA, corregido por elegibilidad desconocida, queda definido según la ecuación (13):

$$w_{em,k}^{known} = \hat{R}_{em,known} \cdot w_{em,k} \quad (33)$$

Donde $w_{em,k}$ es el ponderador de selección de la persona k perteneciente al estrato em .

V.2.4. Ajuste por no elegibilidad

Durante el levantamiento puede suceder que, al momento de visitar la vivienda de la persona seleccionada, no sea posible ubicar a un informante elegible o, que la vivienda se encuentre desocupada, sea de temporada o en construcción, o bien, se trate de una oficina de gobierno, hospital, hogar de ancianos, etc., que corresponden a edificaciones que no contienen a la población de interés. Estas situaciones llevan a que el NNA seleccionado(a) sea clasificado como no elegible y, a partir de esta etapa no se considera para fines analíticos.

A los NNA clasificados como no elegibles se les asigna un valor blanco (“missing”) en el ponderador ajustado por elegibilidad desconocida. La Tabla 7 presenta la distribución de las personas seleccionadas de acuerdo con su estado de elegibilidad según el trabajo de campo de EANNA 2023.

Tabla 7. Distribución viviendas según elegibilidad

Elegibilidad	Frecuencia	Porcentaje
Total	25.611	100,0%
Elegible	22.286	87,0%
Elegibilidad desconocida	546	2,1%
No elegible	2.779	10,9%

Elaborado por el Instituto Nacional de Estadísticas.

V.2.5. Ajuste por no respuesta

En esta etapa solo son consideradas las NNA elegibles, pues son los que cumplen con las características necesarias para participar en la encuesta. Sin embargo, es posible que los padres o tutores o el NNA seleccionado(a) no deseen participar o simplemente no puedan ser contactados. En EANNA 2023, de las 22.286 personas que resultaron elegibles, 17.659 fueron entrevistadas lo que corresponde a una tasa de respuesta de 79,2%¹¹.

Utilizar un ponderador de selección sin tomar en cuenta el fenómeno de la no respuesta a la unidad, significa obtener estimaciones representativas solo de aquellas unidades que participan en la encuesta y, por lo tanto, una parte de la población no sería cubierta.

Por otro lado, si existe alguna relación entre la no respuesta y la variable de interés es posible que las estimaciones resulten sesgadas, motivo por el cual, se hace necesario aplicar un ajuste al factor a partir de un modelo que reproduzca el comportamiento de la no respuesta.

El método elegido para el ajuste es el propensity score donde se modela la probabilidad de respuesta a partir de variables predictoras obtenidas desde Casen 2022 y de la hoja de ruta de EANNA 2023. Una vez obtenida la probabilidad de respuesta, las unidades se ordenan de menor a mayor y se agrupan en veintiles (20 grupos de igual tamaño).

En la Tabla 8 se presentan las variables utilizadas en el modelo y la fuente desde donde se obtienen.

Tabla 8. Variables utilizadas en el modelo de no respuesta EANNA 2023

Variable	Fuente
Estado físico del edificio o vivienda	Hoja de ruta EANNA 2023
Modo de contacto (presencial o telefónico)	Hoja de ruta EANNA 2023
Cantidad de encuestas realizadas por la persona encuestadora en tramos	Hoja de ruta EANNA 2023
Región	Casen 2022

¹¹ En esta etapa de cálculo del factor de expansión el peso de las personas que resultaron de elegibilidad desconocida ha sido distribuido entre todas las personas elegibles y no elegibles y, por lo tanto, ya no se consideran para el análisis. En ese sentido, la tasa de respuesta obtenida (79,2%) corresponde al total de personas que respondieron la encuesta sobre el total de personas que resultaron elegibles y es diferente a la reportada en los indicadores de desempeño del trabajo de campo (77,3%).

Variable	Fuente
Edad NNA en tramos ([5-8], [9-11], [12-14], [15-17])	Casen 2022
Presencia de rayados o daños a casas o vehículos en el entorno de la vivienda	Casen 2022
Presencia de consumo de drogas o alcohol en el entorno de la vivienda	Casen 2022
Presencia de personas peleando o amenazándose en el entorno de la vivienda	Casen 2022
Nivel educacional NNA (educación especial, no asiste, educación tradicional)	Casen 2022
Sistema previsional de salud del NNA	Casen 2022
Condición de tenencia de la vivienda	Casen 2022
Tipo de hogar	Casen 2022
Vivienda a menos de 20 cuadras o 2,5 km de equipamiento comunitario	Casen 2022
Sexo del NNA	Casen 2022
Quintil autónomo del jefe/a de hogar al que pertenece el NNA	Casen 2022

Elaborado por el Instituto Nacional de Estadísticas.

Con el modelo de predicción se espera que al interior de las 20 celdas de ajuste las unidades sean similares en sus características y así, aquellas entrevistadas puedan representar a las no entrevistadas. Bajo este supuesto, los pesos asociados a las personas contactadas elegibles que no respondieron deben ser redistribuidos entre las personas que sí lo hicieron. Luego, por veintil se calcula la razón de ajuste. Esto se observa en la ecuación (14) que utiliza el ponderador ajustado por no elegibilidad.

En esta ecuación, en el numerador se incluye la estimación del total de personas elegibles y, en el denominador, la estimación del total de personas elegibles que responden.

$$\hat{R}_{v,R} = \frac{\sum_{\epsilon \in \Phi_v} \sum_{\epsilon \in \theta_v} w_{em,k,eleg}}{\sum_{\epsilon \in \Phi_v} \sum_{\epsilon \in \theta_{v,R}} w_{em,k,resp}} \quad (14)$$

Donde,

- v : : Índice del grupo para la corrección de no respuesta (veintil).
- Φ_v : : Conjunto de personas asignadas al veintil v .
- θ_v : : Conjunto de personas pertenecientes al veintil v elegibles.
- $\theta_{v,R}$: : Conjunto de personas pertenecientes al veintil v que respondieron.
- $w_{em,k,eleg}$: Estimación del total de personas elegibles.
- $w_{em,k,resp}$: Estimación del total de personas que responde.

Una vez estimados los factores de ajuste, el ponderador de selección corregido por no respuesta $w_{em,k}^{NR}$ se expresa según la ecuación (15):

$$w_{em,k}^{NR} = \hat{R}_{v,R} \cdot w_{em,k,eleg} \quad (15)$$

Este ponderador se asigna a las 17.659 NNA elegibles entrevistadas y a los NNA elegibles no entrevistados se les asigna un valor blanco (“missing”).

La Tabla 9 presenta la distribución de los veintiles y la probabilidad de respuesta de cada uno de ellos.

Tabla 9. Veintiles no respuesta EANNA 2023

Veintil	NNA elegibles	NNA que no responden	NNA que responden	Promedio probabilidad de respuesta
Total país	22.286	4.627	17.659	
1	1.115	857	258	0,090
2	1.115	376	739	0,346
3	1.115	334	781	0,394
4	1.115	277	838	0,424
5	1.115	280	835	0,448
6	1.115	237	878	0,468
7	1.114	223	891	0,487
8	1.114	239	875	0,504
9	1.114	218	896	0,521
10	1.114	213	901	0,538
11	1.114	193	921	0,555
12	1.114	178	936	0,573
13	1.114	176	938	0,591
14	1.114	170	944	0,610
15	1.114	165	949	0,632
16	1.114	141	973	0,659
17	1.114	117	997	0,697
18	1.114	105	1.009	0,738
19	1.114	74	1.040	0,779
20	1.114	54	1.060	0,839

Elaborado por el Instituto Nacional de Estadísticas.

V.3. Suavizamiento

El suavizamiento de los factores de expansión tiene por objetivo acotar la variabilidad de valores atípicos debido a los múltiples ajustes que se deben realizar. Para ello se aplica una metodología mixta que, por un lado, los contrae a la media (CM) y, por otro, aplica un aumento proporcional a los factores no atípicos y una disminución proporcional a los atípicos (método R-K).

V.3.1. Método de contracción a la media (CM)

Este método tiene por objetivo que la distribución de los factores sea más homogénea y para ello contrae los factores de expansión, acercándolos al valor promedio de éstos. El factor suavizado o contraído a la media (o promedio) viene dado por:

$$w_i^S = \beta \cdot w_i + (1 - \beta) \cdot \bar{w}; \quad 0 \leq \beta \leq 1 \tag{16}$$

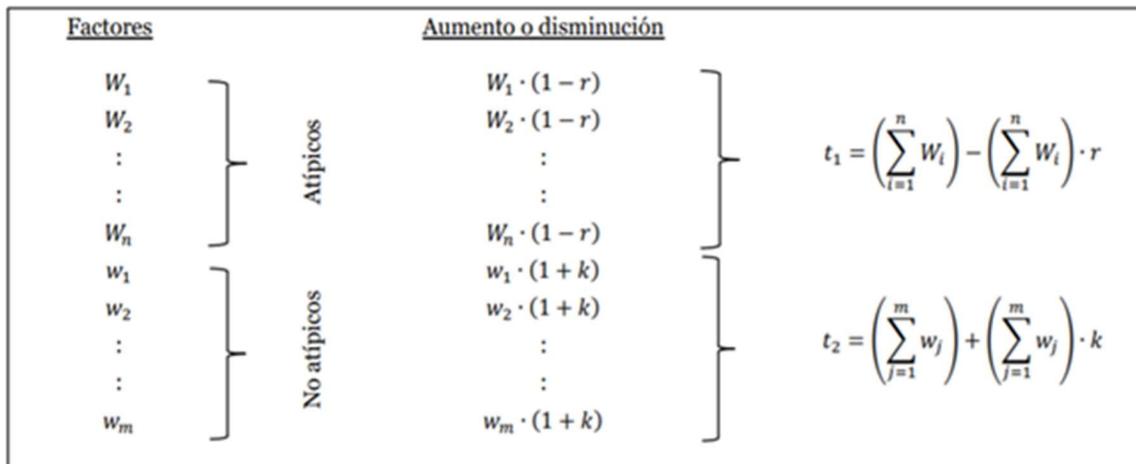
Donde,

- β : Es un parámetro real que permite controlar la intensidad de la contracción.
- w_i : : Factor ajustado por no respuesta.
- \bar{w} : : Media de los factores a nivel de estrato (región - área - tramo etario) y de valores atípicos y no atípicos.

V.3.2. Método R-K

Este método da un tratamiento diferenciado a los conjuntos de datos atípicos o no atípicos. Por un lado, genera un aumento proporcional en los factores no atípicos y, por otro lado, produce una disminución proporcional en los factores atípicos. Esto significa que el método R-K disminuye en una proporción R los valores atípicos y aumenta en una proporción K los valores no atípicos como se indica en la figura 2.

Figura 2. Ordenamiento de factores de expansión atípicos y no atípicos



Fuente: Instituto Nacional de Estadísticas.

V.3.3. Método mixto

El método mixto combina los métodos CM y R-K y consiste en separar los factores de expansión en un subconjunto de factores no atípicos y otro de factores atípicos. Para los factores no atípicos se aplica el método de contracción a la media considerando $\beta = 0,8$ mientras que, para los factores atípicos se utiliza $\beta = 0,2$ con el fin de que, en caso de tener un único valor atípico al interior del estrato, el método CM no tenga efecto sobre ese valor y se conserve el valor del factor atípico.

Luego, se aplica el método R-K donde se unen las distribuciones tanto de factores atípicos como en un continuo que se puede lograr con valores de λ cercanos a la unidad. En este caso se utiliza $\lambda =$

0,99 para conservar una pequeña distancia entre el mayor de los factores no atípicos y el menor de los factores atípicos, después del suavizamiento.

V.4. Calibración de factores de expansión

La calibración de los factores de expansión sirve para ajustar los factores a los totales poblacionales dados por las proyecciones de población. En el caso de EANNA 2023 las proyecciones de población utilizadas son al 30 de julio de 2023, fecha que corresponde a la mitad del periodo de levantamiento de la encuesta.

Para la calibración se utiliza el método *Raking* que permite incorporar información auxiliar y tiene tres propósitos, el primero de ellos es que la encuesta, a partir de los pesos calibrados, sea capaz de reproducir totales poblacionales conocidos, el segundo objetivo es alcanzar mejoras en la precisión de los parámetros de interés por medio de la obtención de estimaciones con un menor error de muestreo y finalmente, se pretende obtener una mayor precisión en las estimaciones (Alvarado, Pizarro & Guarda, 2020).

Las variables marginales de calibración utilizadas corresponden a las siguientes:

- Población total por región y área.
- Población total por región y sexo.
- Población total por tramo etario ([5-8], [9-11], [12-14], [15-17]) y sexo.

Dado que la población objetivo del estudio es la que reside en viviendas particulares es necesario realizar un ajuste a las proyecciones de población. Para esto se considera la proporción de población residente en estas viviendas sobre el total de población en el país según los resultados del Censo de Población y Vivienda 2017. El ajuste se realiza a nivel de las marginales de calibración, lo que da como resultado diferentes stocks poblacionales para cada marginal. Por este motivo, es necesario realizar un ajuste adicional que toma como referencia la marginal de región-sexo y los stocks poblacionales de la marginal región-área y tramo etario-sexo son ajustados para que las tres marginales cuadren y sea posible realizar la calibración con el método *Raking*.

La Tabla 10 muestra las proyecciones de población de personas entre 5 a 17 años, según región y área, y aquellas ajustadas a población residente en viviendas particulares utilizadas en EANNA 2023.

Tabla 10. Proyecciones de población de 5 a 17 años al 30 de julio de 2023 según región y área

Región	Estimación de población total			Estimación de población residente en viviendas particulares		
	Urbano	Rural	Total	Urbano	Rural	Total
Total	2.926.311	359.935	3.286.246	2.913.667	357.033	3.270.700
Arica y Parinacota	43.291	3.522	46.813	43.008	3.415	46.423
Tarapacá	74.162	2.215	76.377	73.818	2.168	75.986
Antofagasta	123.600	1.984	125.584	122.900	1.936	124.836
Atacama	57.056	4.457	61.513	56.818	4.444	61.262
Coquimbo	130.595	24.330	154.925	129.942	24.275	154.217
Valparaíso	296.315	27.517	323.832	295.004	27.362	322.366
Metropolitana	1.252.011	55.025	1.307.036	1.248.822	54.792	1.303.614
O'Higgins	136.176	37.749	173.925	135.918	37.686	173.604
Maule	153.024	44.750	197.774	152.282	44.677	196.959
Ñuble	63.467	22.578	86.045	63.072	22.543	85.615
Biobío	251.902	29.869	281.771	250.795	29.667	280.462
La Araucanía	130.560	49.171	179.731	129.467	48.684	178.151
Los Ríos	50.359	18.202	68.561	49.579	17.560	67.139
Los Lagos	119.277	34.260	153.537	118.153	33.794	151.947
Aysén	17.582	2.640	20.222	17.405	2.570	19.975
Magallanes	26.934	1.666	28.600	26.684	1.460	28.144

Elaborado por el Instituto Nacional de Estadísticas.

Adicionalmente, en el proceso de calibración se incorpora información del trabajo de campo respecto a cambios en los estratos de muestreo definidos originalmente en el diseño. Estos cambios responden, en su mayoría, a diferencias en el tramo etario, principalmente debido a que la edad bajo la cual se realizó la estratificación en el diseño muestral corresponde a una edad estimada a partir de los datos levantados en Casen 2022. El propósito de este ajuste es que cada NNA de la muestra se asocie a la región, área y tramo etario en el que fue encuestado(a) y no en el definido para el diseño de la encuesta, logrando con esto, obtener estimaciones acordes a la realidad actual y evitar posibles sesgos asociados a la desactualización (Jang, Sukasih, Lin, Kang, & Cohen, 2009). En la Tabla 11 se presenta la distribución de los NNA logrados y de aquellos que experimentaron cambios de estrato a nivel regional, así como la proporción que estos representan sobre el total.

Tabla 11. Distribución de los cambios de estrato en la muestra lograda, EANNA 2023.

Región	NNA con cambio de estrato	NNA logrados	Proporción cambio de estrato
Nacional	531	17.659	3,0%
Arica y Parinacota	24	676	3,6%
Tarapacá	31	753	4,1%
Antofagasta	29	720	4,0%
Atacama	31	634	4,9%
Coquimbo	36	878	4,1%
Valparaíso	46	1.902	2,4%

Región	NNA con cambio de estrato	NNA logrados	Proporción cambio de estrato
Metropolitana	100	3.509	2,8%
O'Higgins	26	1.015	2,6%
Maule	33	1.212	2,7%
Ñuble	30	746	4,0%
Biobío	59	1.951	3,0%
La Araucanía	20	1.028	1,9%
Los Ríos	22	705	3,1%
Los Lagos	29	878	3,3%
Aysén	8	490	1,6%
Magallanes	7	562	1,2%

Elaborado por el Instituto Nacional de Estadísticas.

El resultado del proceso de calibración se presenta en la Tabla 12 donde se observa que, luego del ajuste, los factores de expansión alcanzan un total de 3.270.700 personas.

Tabla 12. Distribución de la muestra y tamaño poblacional expandido según nacional – área y región, EANNA 2023

Región	Tamaño de muestra lograda	Tamaño expandido a población entre 5 y 17 años
Total	17.659	3.270.700
Urbano	15.022	2.913.667
Rural	2.637	353.033
Arica y Parinacota	676	46.423
Tarapacá	753	75.986
Antofagasta	720	124.836
Atacama	634	61.262
Coquimbo	878	154.217
Valparaíso	1.902	322.366
Metropolitana	3.509	1.303.614
O'Higgins	1.015	173.604
Maule	1.212	196.959
Ñuble	746	85.615
Biobío	1.951	280.462
La Araucanía	1.028	178.151
Los Ríos	705	67.139
Los Lagos	878	151.947
Aysén	490	19.975
Magallanes	562	28.144

Elaborado por el Instituto Nacional de Estadísticas.

V.5. Redondeo Probabilístico

Finalmente, como los resultados son referidos a personas, para obtener un número entero en cada uno de los factores de expansión, se aplica un redondeo de forma probabilística con el objetivo de que los nuevos pesos calibrados no produzcan un sesgo en las estimaciones de interés.

El redondeo probabilístico distribuye los excesos de la parte decimal en forma aleatoria con base en una distribución de probabilidades entre los factores, a partir de semillas de aleatorización diferentes para cada factor que permiten replicar el procedimiento y alcanzar el stock poblacional a nivel nacional; a diferencia del redondeo determinístico que aproxima el valor de los factores al entero más cercano.

Por último, cabe señalar que, en el caso del factor regional, para el redondeo son utilizadas las marginales de calibración región-área, región-sexo y tramo etario-sexo, sin embargo, esto no garantiza que se alcance el stock poblacional para cada una de las categorías de las marginales y solo es posible aproximarse al total nacional como fue señalado previamente.

VI. CÁLCULO Y EVALUACIÓN DE ESTIMACIONES

Este capítulo tiene como objetivo, por un lado, presentar los métodos de estimación de varianza en diseños muestrales complejos que, en el caso de EANNA corresponde a un diseño bifásico elaborado a partir del diseño muestral de Casen 2022 y; por otro, la obtención de estimaciones de parámetros de interés que, para efectos del estudio considera la variable trabajo infantil, junto con sus errores muestrales, para los dominios definidos.

VI.1. Estimación de la varianza compleja

Los métodos exactos para estimar varianza en diseños muestrales estándar, cuando son aplicables, representan la mejor alternativa para obtener estimaciones precisas. Sin embargo, los diseños empleados en encuestas sociales, como es el caso de EANNA, presentan mayor complejidad, al momento de obtener la estimación de parámetros de interés y sus respectivas varianzas, que los diseños aleatorio simple o estratificados.

Un diseño muestral complejo se origina cuando las unidades muestrales son seleccionadas en más de una etapa de selección. En el caso de EANNA, la selección de NNA se obtuvo a partir de las

unidades logradas en Casen 2022¹², la cual ya cuenta con un diseño complejo. Este hecho añade un grado adicional de complejidad al cálculo de las estimaciones de parámetros de interés, debido a la variabilidad adicional introducida por el submuestreo de segunda fase (Gutiérrez, 2015).

En diseños muestrales estratificados, la estimación de la varianza total se obtiene mediante la suma de las varianzas estimadas en cada estrato, y la varianza en cada estrato se estima a partir de la variabilidad de la característica de interés entre los conglomerados. En este punto cabe señalar que, EANNA cuenta con un diseño estratificado, sin embargo, los conglomerados considerados en EANNA no corresponden a una etapa de la selección, sino que se obtienen de la primera fase presente en Casen 2022.

Una aproximación para la estimación de varianza comúnmente utilizada en encuestas complejas, de muestreo polietápico por conglomerados, corresponde al método del conglomerado último (EVCU) o por reemplazo (WR por sus siglas en inglés). Este método permite capturar cierto grado de covarianza por medio de la correlación intraclásica de las unidades al interior de los conglomerados, aun cuando las estimaciones se basen en muestras sin traslape. También es posible la captura de covarianza a través de la correlación intraclásica de las unidades al interior de los estratos o dominios comunes (INE, 2022).

Estas estimaciones con EVCU, pueden efectuarse indistintamente por linealización mediante el desarrollo de Series de Taylor o con una técnica de replicación. Con este enfoque, las UPM se tratan como si se hubieran seleccionado con reemplazo entre los estratos de la primera etapa. En tal caso, cada elemento o individuo solo tiene que identificarse por el estrato de la primera etapa y por la UPM (dentro del estrato) de la que se haya seleccionado.

Para efectos de la estimación de varianzas no se necesita información sobre las etapas de muestreo por debajo del nivel de las UPM. De este modo la descripción del plan de muestreo, necesario para la estimación de varianzas con diseños complejos, se simplifica de forma que se asemeja a un muestreo por conglomerados estratificado en una etapa, esto es, una muestra estratificada de conglomerados finales completamente enumerados. Este enfoque por conglomerados finales produce una buena aproximación para la estimación de la varianza, siempre que sea razonable partir de la hipótesis de una primera etapa con reemplazo.

Así, cuando el plan de muestreo se describe como EVCU, solo se necesitan tres variables del diseño de la encuesta para la estimación de varianzas:

1. La variable de ponderación de la muestra o factor de expansión redondeado designado “*expr*” cuyos totales expanden territorialmente a nivel de región - área.

¹² Casen 2022 presenta un diseño muestral probabilístico, estratificado y bietápico.

2. La variable de estratificación utilizada para el diseño muestral de la encuesta y que corresponde a la combinación de las variables región – área – tramo etario del NNA obtenidas desde Casen 2022¹³, designada “estrato”.
3. La variable UPM (o pseudo-conglomerado) designada como “varunit” obtenida desde Casen 2022.

El empleo de EVCU para la estimación de varianzas considerando el diseño complejo puede dar lugar a una ligera sobreestimación. Sin embargo, al realizar el análisis en general, no hay problemas en aceptar un cierto grado de sobreestimación por la relativa simplicidad de la aproximación mediante EVCU. No obstante, hay que tener en cuenta que la sobreestimación puede ser apreciable si hay varios estratos en los que el muestreo en la primera etapa se hace sin reemplazo y con grandes fracciones de muestreo. En este caso, puede ser preferible optar por un software que cuente con la opción de incorporar los factores de corrección por población finita *cpf* o corrección por finitud para la primera etapa.

VI.2. Variables que identifican el diseño muestral complejo

En la muestra de EANNA 2023, para identificar los estratos definidos en el diseño, se construye la variable “estrato” con la combinatoria de las variables región, área y tramo de edad, considerando la información previa al levantamiento. En relación con los conglomerados, como se mencionó en párrafos previos, la selección de personas se realiza directamente al interior de cada estrato; sin embargo, el diseño de la primera fase (Casen 2022) sí contempla la utilización de conglomerados, por lo que, al momento de realizar las estimaciones, es necesario considerarlos en la definición del diseño complejo a través de la variable “varunit” obtenida desde Casen 2022. Según lo anterior la muestra lograda de personas se distribuye en 128 estratos muestrales conteniendo 8.038 conglomerados, los cuales se encuentran traslapados, pues la unidad de conglomeración se encuentra anidada en más de un estrato de segunda fase.

VI.3. Estadísticos asociados a los parámetros de interés

A partir de la muestra lograda en EANNA 2023, se calculan los estadísticos asociados al parámetro de interés definido por el Ministerio, el cual corresponde a la variable:

¹³ Si bien, durante el levantamiento se identificaron cambios de estrato, para el cálculo de las estimaciones se utiliza la estratificación definida en el diseño muestral original. Esta decisión responde a la recomendación de utilizar el diseño inicial cuando existe asociación entre la variable de interés y las variables de estratificación (Mitani, Mercaldo, Haneuse, & Schildcrout, 2021)

- Trabajo infantil (ti) correspondiente al porcentaje de niños, niñas y adolescentes distribuidos en las categorías: “0. Si NNA no está en situación de trabajo infantil” y “1. Si NNA está en situación de trabajo infantil”.

La Tabla 13 presenta la estimación del trabajo infantil junto con el respectivo cálculo de la varianza representada por los correspondientes errores absolutos y relativos a nivel nacional, por área y regional.

Tabla 13. Estimación del trabajo infantil con sus respectivos errores muestrales a nivel nacional, área y regional, EANNA 2023

Región	Trabajo Infantil		
	Estimación	Error absoluto	Error relativo
Nacional	15,5%	0,6%	4,0%
Urbano	14,7%	0,7%	4,6%
Rural	22,1%	1,7%	7,9%
Arica y Parinacota	12,7%	2,6%	20,6%
Tarapacá	14,2%	2,6%	18,5%
Antofagasta	6,4%	1,8%	27,9%
Atacama	11,8%	2,7%	22,6%
Coquimbo	16,8%	2,6%	15,6%
Valparaíso	15,7%	1,7%	11,0%
Metropolitana	14,4%	1,2%	8,5%
O'Higgins	14,4%	2,2%	15,0%
Maule	17,3%	2,1%	12,3%
Ñuble	14,5%	2,6%	17,6%
Biobío	17,0%	1,7%	10,0%
La Araucanía	26,0%	2,7%	10,5%
Los Ríos	19,6%	3,1%	15,7%
Los Lagos	17,4%	2,5%	14,2%
Aysén	13,2%	3,1%	23,8%
Magallanes	11,9%	2,7%	22,4%

Elaborado por el Instituto Nacional de Estadísticas.

VII. ANEXOS

VII.1. Anexo N°1: Códigos de Disposición Final de Casos EANNA 2023

Glosa código de disposición final	Total	Porcentaje
Total	25.611	100,0%
Personas entrevistadas	17.659	
11. Entrevistada completa	17.426	68,0%
12. Entrevistada parcial	233	0,9%
Personas elegibles, no entrevistadas	4.627	
21. Rechazo a nivel hogar	2.165	8,5%
22. Rechazo informante	31	0,1%
31. Vivienda ocupada, sin moradores presentes	794	3,1%
32. Se impide acceso a la vivienda	47	0,2%
33. Informante no ubicable, se contacta a otra persona	125	0,5%
34. Informante ausente en período de levantamiento de la encuesta	153	0,6%
41. Informante impedido física y/o mentalmente para contestar	210	0,8%
42. Informante no habla idioma español	2	0,0%
43. Muerte del informante o algún familiar	14	0,1%
45. Información falseada	15	0,1%
49. Otra razón de no entrevista	1.053	4,1%
71. Marca tono, pero sin respuesta	10	0,0%
72. Responde buzón de voz	3	0,0%
74. Se encuentra apagado o fuera del área de cobertura	2	0,0%
75. Número telefónico no existe	3	0,0%
Personas con elegibilidad desconocida	546	
51. No se envió a terreno	33	0,1%
52. No fue posible localizar la dirección	336	1,3%
53. Inaccesibilidad por razones climáticas o desastres naturales	29	0,1%
54. Área peligrosa	37	0,1%
59. Otra razón de elegibilidad desconocida	32	0,1%
91. Informante/hogar ya no reside en la vivienda	79	0,3%
Personas no elegibles	2.779	
61. Establecimiento, oficina de gobierno u otra organización	15	0,1%
62. Institución (por ejemplo: hospital, cárcel, asilo de ancianos, etc.)	2	0,0%
64. Vivienda en demolición, incendiada, destruida o erradicada	14	0,1%
65. Vivienda desocupada (por ejemplo: en venta, arriendo, etc.)	200	0,8%
66. Vivienda de veraneo o de uso temporal	23	0,1%
67. Edificación no calificable (por ejemplo: en construcción, en reparaciones, etc.)	16	0,1%
68. No existe informante elegible	998	3,9%
69. Otra razón para ser considerada no elegible	1.511	5,9%

Elaborado por el Instituto Nacional de Estadísticas

VIII. REFERENCIAS

- Alvarado, Pizarro, & Guarda (2020). *La nueva metodología de calibración de la Encuesta Nacional de Empleo: método de calibración Raking*. Instituto Nacional de Estadística de Chile.
- Gutiérrez, A. (2015). *Estrategias de Muestreo, Diseño de Encuestas y Estimación de Parámetros*. Bogotá.
- Heeringa, S. G., West, B. T., & Berglund, P. A. (2010). *Applied Survey Data Analysis*. USA: Chapman & Hall / CRC Press.
- INE. (marzo de 2020). *Fundamentos del Estándar para la evaluación de la calidad de las estimaciones en encuestas de hogares*. Obtenido de INE Chile: <https://www.ine.gob.cl/inicio/documentos-de-trabajo/documento/fundamentos-del-est%C3%A1ndar-para-la-evaluaci%C3%B3n-de-la-calidad-de-las-estimaciones-en-encuestas-de-hogares>
- INE. (2022). *Separata técnica. Metodología de estimación anual Encuesta Nacional de Empleo (ENE)*. Obtenido de <https://www.ine.gob.cl/docs/default-source/ocupacion-y-desocupacion/publicaciones-y-anuarios/separatas/tematicas/estimacion-anual-2022.pdf>
- Jang, D., Sukasih, A., Lin, X., Kang, K., & Cohen, S. (2009). Effects of Misclassification of Race/Ethnicity Categories in Sampling Stratification on Survey Estimates. *Proceedings of the American Statistical Association, survey methods section*.
- Kish, L. (1965). *Survey Sampling*. Nueva York: John Wiley & Sons.
- MDSF. (2013). *Encuesta de Actividades de Niños, Niñas y Adolescentes 2012: Metodología del Diseño Muestral y Factores de Expansión*. .
- Mitani, A., Mercado, N., Haneuse, S., & Schildcrout, J. (2021). Survey design and analysis considerations when utilizing misclassified sampling strata. *BMC Medical Research Methodology*.