



PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
FACULTAD DE CIENCIAS SOCIALES
INSTITUTO DE SOCIOLOGÍA

FACTORES DE EXPANSIÓN MÁS COMPLEJOS

¿MÁS EFICIENTES?

APLICACIÓN PARA EL CASO DE BARÓMETRO DE LA FELICIDAD 2014

POR

CRISTINA MARCHANT ARAYA

Actividad Final de Grado presentada al Instituto de Sociología de la Pontificia Universidad Católica de Chile, para optar al grado académico de Magíster en Diseño y Análisis de Encuestas Sociales

Profesor Guía:

Carolina Casas-Cordero, PhD.

Diciembre, 2018

Santiago, Chile

Cristina Marchant Araya cemarcha@uc.cl

Abstract

Objetivo: el presente trabajo tiene como principal objetivo observar el desempeño de distintos factores de expansión que varían en su construcción para la estimación de una variable de bienestar subjetivo de la encuesta chilena Barómetro de la Felicidad 2014.

Métodos: para la encuesta Barómetro 2014 se construyeron factores de expansión con variaciones en el cálculo de los ponderadores base y sus ajustes de no respuesta, para luego compararse en torno a: un proxy de sesgo relativo respecto a una estimación de bienestar subjetivo en la Encuesta Casen 2013, medición de Efecto Diseño en las estimaciones, y Error Cuadrático Medio para conjugar el sesgo y la varianza, dando cuenta del tradeoff clásico que la literatura ha señalado respecto a la utilización de factores de expansión.

Resultados: los factores que incluyeron ajuste de no respuesta explícito obtuvieron mejores resultados en la reducción del sesgo, que aquellos que no lo incluyeron. A su vez, la varianza de las estimaciones no aumentó de manera sustantiva cuando se utilizaron factores de expansión con ajuste de no respuesta. Con ello, dichos factores obtuvieron el mejor desempeño entre la reducción del sesgo y el aumento de la varianza en las estimaciones para la variable de bienestar subjetivo en Barómetro 2014.

Conclusión: cuando se realizan ajustes de no respuesta explícitos en función de variables que se encuentren relacionadas tanto con la participación en la encuesta, como con la variable de interés, se obtienen factores de expansión más eficientes, en tanto contribuyen a reducir el sesgo de las estimaciones, sin aumentar la varianza.

Contenido

1	Introducción.....	5
2	Antecedentes de la Literatura	8
	A. Problemas de No respuesta y Fallas de cobertura.....	8
	B. Métodos de ponderación	9
	C. Variables auxiliares para ajuste de ponderadores	11
	D. Indicadores de desempeño para comparar ponderadores	13
3	Métodos	14
	A. Tercer Barómetro de la Felicidad	14
	i. Medición de bienestar subjetivo	16
	B. Construcción de factores de expansión alternativos	18
	C. Desempeño de factores de expansión	22
4	Resultados.....	25
	A. ¿Qué tan adecuadas son las variables auxiliares disponibles para los ajustes de No Respuesta?	25
	B. ¿Qué tan distintos son los ponderadores desarrollados?.....	27
	C. Impacto de los ponderadores en la precisión de las estimaciones	28
	D. Impacto de los ponderadores en la acuracidad de las estimaciones.....	29
	E. Impacto de los ponderadores en el Error Cuadrático Medio de las estimaciones .	31

5	Discusión	33
6	Bibliografía.....	36
7	Anexos	42
	i. Estructura de los datos de Encuesta Barómetro 2014.....	46
	ii. Ponderador base de viviendas y personas	49
	iii. Ajustes explícitos de no respuesta.....	52
	iv. Ponderador de personas.....	55
	v. Calibración	56

1 Introducción

Los factores de expansión se utilizan comúnmente para asignarle a los individuos de una muestra un “peso” de tal manera que las estimaciones ponderadas de la encuesta representen de la mejor manera posible a la población de objetivo (Kalton & Flores-Cervantes, 2003). Este trabajo tiene como principal objetivo estudiar la performance de diversos factores de expansión en la estimación de una variable de bienestar subjetivo en la encuesta Barómetro de la Felicidad 2014.

En la práctica, la construcción de dichos factores puede variar, pero en general contempla tres etapas; ponderación por probabilidades de selección, ajustes de no respuesta y ajuste de calibración a totales poblacionales externos (Kalton & Flores-Cervantes, 2003; Lohr, 2009; Groves, 2006; Little & Vartivarian, 2003, Valliant, Dever, & Kreuter, 2013). En la primera etapa, se genera un ponderador base que refleja las probabilidades de selección de los individuos de la muestra. En la segunda, se efectúan ajustes para compensar por la no respuesta de los individuos inicialmente seleccionados. Y en la tercera, se realiza una corrección por fallas de cobertura, para que (1) la muestra se asemeje a la población en variables auxiliares relevantes, y (2) mejorar la precisión en estimaciones de subpoblaciones de interés.

Ahora bien, las investigaciones que utilizan encuestas como modo de recolección de datos no siempre elaboran factores de expansión que contemplen los tres componentes mencionados. Como ejemplo de ello, la Tabla 1 agrupa encuestas según tres estrategias distintas utilizadas para calcular los factores de expansión en una encuesta.

La Estrategia A consiste en construir los ponderadores base ignorando el diseño muestral complejo efectivamente empleado, asumiendo que la muestra se seleccionó bajo muestreo aleatorio simple (M.A.S). A continuación, se construye un solo factor de ajuste que corrige, en forma implícita, tanto la no respuesta como la falla en la cobertura utilizando totales poblacionales externos.

La Estrategia B consiste en construir ponderadores base que reflejen el diseño muestral complejo efectivamente empleado y, a continuación, construir un solo factor de ajuste que

corrige, en forma implícita, tanto la no respuesta como la falla en la cobertura utilizando totales poblacionales externos.

Finalmente, la Estrategia C construye ponderadores base que reflejan el diseño muestral complejo empleado y, a continuación, construye factores de ajuste explícitos para la no respuesta y la falla en la cobertura.

La literatura y las buenas prácticas actuales recomiendan el uso de estrategia de tipo C – donde se reflejen de cerca las probabilidades de selección de la muestra, y que utilicen, en la medida de lo posible, ajustes explícitos para las correcciones estadísticas empleadas en la corrección de las fallas de representación de las unidades entrevistadas. Estrategias más sencillas (o ingenuas) como las A y B, todavía se encuentran en la práctica debido a falta de información auxiliar para el desarrollo de modelos de cálculo/corrección más sofisticados, deseo de no afectar comparabilidad histórica de los cálculos, o excesiva ingenuidad para simplificación de los cálculos.

Independiente de las razones, las distintas estrategias empleadas en la construcción de factores de expansión de una encuesta con diseño muestral complejo tienen efecto en las propiedades de los ponderadores y en las estimaciones que se desarrollan a partir de ellos.

Considerando todo lo anterior es que surgen las preguntas: ¿Cuáles son las implicancias de calcular ajustes de expansión complejos? ¿Cómo se comportan factores de expansión más y menos complejos en su elaboración respecto a la reducción de sesgo y aumento de varianza?

Este trabajo muestra, en particular, el desempeño relativo de distintas estrategias de cálculo de factores de expansión para la encuesta Barómetro de la Felicidad 2014, que tiene un diseño muestral y tamaño que es similar a varias otras encuestas que se levantan en Chile. La variable de interés del estudio corresponde a una medición de bienestar subjetivo, temática abordada en políticas públicas, y por tanto, encontrándose presente en otras encuestas de relevancia del país, como la encuesta Casen, la Encuesta Nacional de Calidad de Vida y Salud y la encuesta del Centro de Estudios Públicos.

Así, el objetivo de este trabajo es contribuir a la discusión de las decisiones metodológicas prácticas en el contexto de las encuestas que buscan producir inferencia a la población chilena, y en particular, de la temática del bienestar subjetivo en el país.

Tabla 1. Ejemplos de encuestas chilenas e internacionales que utilizan distintas estrategias para el cálculo de ponderadores

Estrategia de cálculo de ponderadores	Características de los ponderadores			Encuesta en Chile	Encuesta internacionales
	Ponderador base	Ajuste NR* y Cobertura			
Estrategia A	Calculado asumiendo MAS (cuando el diseño muestral es complejo)	Ajuste realiza	NR: no	ENDC 2016, ENMA 2016, Barómetro de la Felicidad 2013.	Latinobarómetro 2016, Climate Change in the American Mind 2018, Public Perceptions of the Health Consequences of Global Warming 2014, A Baseline Human Rights Survey 2013, Asian Barometer 2016.
Estrategia B	Calculado asumiendo el diseño muestral complejo	Ajuste realiza	NR: no	Encuesta Nacional Bicentenario 2017, ENTV 2017, ENDH 2015, ENPE 2015, NENE 2018, NESI 2017.	EQLS 2016, European Working Conditions Survey 2015, EHS 2016
Estrategia C	Calculado asumiendo el diseño muestral complejo	Ajuste realiza	NR: sí realiza	Encuesta Casen 2015, ENS 2017, ENCAVI 2017, Encuesta de Ocupación y Desocupación en el Gran Santiago 2018, EPS 2015, ENPV 2017, ENVAE 2014, ENDPG 2016, ENPC 2017, ENUSC 2017.	ACS 2017, The Nationwide Food Consumption Survey, AHS 2015, ATUS 2018, NCVS 2016, LFS 2013, Adult Education Survey 2016, EU Statistics on Income and Living Conditions 2013

Nota: en la Tabla 8 de Anexos se encuentran las referencias de metodología para cada una de las encuestas referenciadas.
Ajuste NR: Ajuste de No respuesta.

2 Antecedentes de la Literatura

A. Problemas de No respuesta y Fallas de cobertura

Las encuestas actualmente enfrentan dos grandes problemas para realizar inferencia; el fenómeno de la no respuesta y las fallas de cobertura.

Por una parte, la teoría del muestreo que fundamenta las muestras en encuestas supone que todos los elementos inicialmente seleccionados contestan, sin embargo, este supuesto en la actualidad y desde hace décadas es poco realista (Groves, 2006; Little & Vartivarian, 2003; Little & Rubin, 2014). De hecho, en general, las tasas de respuesta en las encuestas no superan el 80% (algunos ejemplos de ello en la Tabla 2).

Tabla 2. Tasas de respuesta en encuestas en Chile e internacionales

Encuestas en Chile	Tasa de respuesta %	Encuestas Internacionales	Tasa de respuesta %
Casen 2017	75,5	EQLS 2016	34,0
ENS 2017	66,2	European Working Conditions Survey 2015	42,5
ENCAVI 2016	78,7	ATUS 2018	45,6
ENDPG 2016	64,0	Adult Education Survey 2016	51,7

Por otro lado, los marcos muestrales que se utilizan para seleccionar las muestras, y otros procedimientos para seleccionar a entrevistados pueden tener problemas, y por tanto, la muestra resultante no cubre perfectamente a la población objetivo.

Por ambas razones, la información de los entrevistados que está en las bases de datos de las encuestas puede no representar adecuadamente a la población de inferencia del estudio. Para hacer frente a este problema hay distintas estrategias.

Mientras la imputación es ampliamente utilizada para sobrellevar el problema de la no respuesta al ítem, es decir, cuando no se responde alguna pregunta del cuestionario (ver más sobre el tópico en Royston, 2004; Andridge & Little, 2010; Little & Rubin, 2014), el modelamiento y el desarrollo de ponderadores -por otro lado-, son los mecanismos más

utilizados en estadísticas oficiales para hacer frente a la no respuesta a la unidad (que corresponde a cuando un individuo o vivienda rechaza participar en la encuesta) y a los problemas de cobertura en encuestas.

A continuación nos enfocaremos en los métodos de ponderación.

B. Métodos de ponderación

La construcción de ponderadores contempla dos grandes componentes; el cálculo de los ponderadores base, y ajustes estadísticos que permitan corregir potenciales sesgos de no respuesta y de cobertura de los datos recolectados (Bethlehem, 2002; Carlson & Williams, 2001; Kalton & Flores-Cervantes, 2003)

En la primera fase, a cada elemento de la muestra se genera un ponderador base. El cálculo del ponderador base busca reflejar las características del diseño muestral complejo implementado, tales como la estratificación, la conglomeración, y las probabilidades de selección de las unidades de la muestra. El ponderador base corresponde al inverso de la probabilidad de selección calculadas (Asparouhov, 2005; Kalton & Flores-Cervantes, 2003; Lohr, 2009).

La segunda etapa en la construcción de los factores corresponde a ajustes por no respuesta a la unidad. El ajuste de no respuesta a través de ponderadores busca redistribuir los ponderadores de los elementos seleccionados que no respondieron entre elementos seleccionados que sí respondieron de características similares. Dicho de otra forma, a los que sí responden se les aumenta el “peso” para que representen a los individuos que no respondieron (Groves & Peytcheva, 2008; Kalton & Flores-Cervantes, 2003). Métodos utilizados para el cálculo de ajustes de no respuesta incluyen el método de celdas y los modelos de propensión a responder. (Para más detalles sobre los métodos ver Groves & Peytcheva, 2008; Kalton & Flores-Cervantes, 2003; Deville & Särndal, 1992; Carlson & Williams, 2001; Valliant, Dever, & Kreuter, 2013).

El ajuste de cobertura o calibración busca lograr que la muestra expandida se ajuste a valores conocidos de la población y mejorar la precisión de las estimaciones en la encuesta

(para las subpoblaciones de interés para las cuales se tiene la información de totales externos disponibles) (Kalton & Flores-Cervantes, 2003). A su vez, el ajuste de cobertura expande los respondientes a la población objetivo, a diferencia del ajuste explícito de no respuesta, donde se expande a la muestra seleccionada. Además, la calibración también se conoce como un tipo de ajuste implícito de no respuesta, donde al ponderar para representar a la población objetivo en base a características sociodemográficas, se estará ajustando implícitamente por aquellos individuos que no respondieron la encuesta en base a sus características.

VARIABLES sociodemográficas usualmente utilizadas en la calibración incluyen estimaciones de población por edad y sexo derivadas de modelos de proyección de la población desarrollados por las oficinas nacionales de estadísticas en base datos de Censos de población y estadísticas vitales. Además, dependiendo de la temática de la encuesta -y de la disponibilidad de datos- se pueden encontrar ajustes por raza o etnia, situación ocupacional, nivel educacional, entre otros (Valliant, 1993; Lee, 2006).

En particular, encuestas relacionadas con la medición de bienestar subjetivo han utilizado en el proceso de calibración variables sociodemográficas que se vinculan con el fenómeno y son relevantes para el análisis de los resultados, respondiendo a la necesidad de ajustar los datos a las subpoblaciones para las cuales se tiene planeado publicar estimaciones, tales como la edad, el sexo, nivel educacional y región de residencia (EQLS, 2011; NORC at the University of Chicago; World Values Survey, 2017; Latinobarómetro Org; Dirección de Estudios Sociales UC - Instituto de la Felicidad de Coca-Cola, 2015).

C. Variables auxiliares para ajuste de ponderadores

Según la literatura, las variables que son más efectivas en reducir los potenciales sesgos de cobertura y no respuesta son aquellas asociadas directamente a las unidades seleccionadas. En encuestas de individuos, por ejemplo, las variables individuales tendrían mayor impacto, seguidas por las variables a nivel de hogar y finalmente variables del entorno del hogar. Estas variables deben estar disponibles para quienes responden y quienes no responden la encuesta. Y deben, además, estar relacionadas con el fenómeno bajo estudio y con la probabilidad de participar en la encuesta.

Los investigadores no cuentan con dichas variables para respondentes y no respondentes. Y por ello, se deben utilizar auxiliares para realizar los ajustes de no respuesta que se encuentren disponibles para todos los individuos. Ahora, la elección de dichas variables no es trivial, pues hacer una “buena” elección reducirá el sesgo de no respuesta, y no afectará de sobremanera la variabilidad de los datos (Kreuter, et al., 2010; Kalton & Flores-Cervantes, 2003; Little & Vartivarian, 2005).

Dichas las variables deberán relacionarse con; a) la probabilidad de responder o participar de la encuesta, y b) la variable de interés del estudio. Por ello, la decisión de cuales variables incluir en los ajustes dependerá del “trade-off” que se de entre la reducción del sesgo y el aumento de la varianza (Kalton & Flores-Cervantes, 2003; Kreuter, et al., 2010).

En ese contexto, existe amplia literatura que aborda cómo variables del contexto social donde se aplica el estudio, del diseño de éste, y las características de las unidades de muestreo y encuestadores influyen en la no respuesta en las encuestas (ver más detalles en Groves, 2006, Groves & Couper, 1998; Groves, Cialdini, & Couper, 1992; Lipps & Pollien, 2011; Bethlehem, 2002).

El uso de variables auxiliares que no cumplan con los requisitos mencionados anteriormente pueden presentar problemas. Por ejemplo, Kreuter y otros mostraron que cuando existen débiles asociaciones entre las variables incluidas en los ajustes de no respuesta, y las variables de participación o de resultado de la encuesta, los ponderadores construidos en base a dichos ajustes no modifican sustantivamente las estimaciones. En ese

sentido, se destacó que cuando los predictores de las variables resultado de los estudios también son predictores de la respuesta, estos deberían incluirse en los ajustes de no respuesta (Kreuter, y otros, 2010). En línea con ese resultado, Groves y otros investigadores en el año 2007 en su estudio de la National Survey of Family Growth encontraron que cuando se realizan ajustes de no respuesta con variables que tienen niveles de correlación modestos con la variable de respuesta y variables resultado, los efectos en las estimaciones ponderadas también eran modestos (Groves, Wagner, & Peytcheva, 2007).

Por otra parte, Wun en su estudio de la Medical Expenditure Panel Survey (MEPS) mostró que cuando se realizan ajustes de no respuesta con variables vinculadas a la variable de resultado (en este caso, salud), el sesgo y la varianza de las estimaciones ponderadas disminuye, obteniendo factores de expansión con mejor performance que aquellos donde se excluyen dichas variables (Wun & Ezzati-Rice, 2007).

En concordancia con lo anterior, otros investigadores mostraron que se observaban efectos en las estimaciones ponderadas solo cuando las variables utilizadas para ajustar por no respuesta estaban correlacionadas con las variables resultado y la probabilidad de responder. Sin embargo, los cambios en las estimaciones en general son pequeñas (Kreuter, Lemay, & Casas-Cordero, 2007).

La National Survey of Family Growth, que incorpora la medición de bienestar subjetivo, utiliza variables como tipo de contacto con el hogar, zona de residencia, tipo de viviendas, raza y lenguaje de ocupantes del hogar para realizar ajustes de no respuesta (Lepkowski, Mosher, & Davis, 2006).

D. Indicadores de desempeño para comparar ponderadores

La literatura ha planteado que la idoneidad de los factores de expansión se vería reflejado en un positivo balance entre la reducción del sesgo y el incremento en la varianza. Estudios que han investigado respecto a ese tradeoff han utilizado como indicadores de performance de los factores; el sesgo y sesgo relativo (Little & Vartivarian, 2005), varianza y Efecto Diseño de los factores de expansión (Carlson & Williams, 2001), varianza en la estimación (Kreuter, y otros, 2010), y error cuadrático medio (Little & Vartivarian, 2005; Wun & Ezzati-Rice, 2007)

Por ejemplo, Kreuter et al. (2010) compararon el desempeño de factores de expansión utilizando los cambios en los promedios estimados de variables de interés y cambios en las varianzas de las estimaciones cuando se incorporan nuevas variables a los ajustes de no respuesta de los factores (Kreuter, y otros, 2010).

Wun y Ezzati-Rice (2007) utilizaron el Error Cuadrático Medio para comparar factores de expansión que incluyen y no incluyen ajustes de no respuesta explícitos, al igual que Little y Vartivarian (2005). Casas-Cordero (2010), además, utiliza el Factor de Inflación de la Varianza de Kish para evaluar la performance de diversos factores de expansión que varían en las variables utilizadas para realizar el ajuste de no respuesta.

3 Métodos

Para responder a los objetivos del trabajo se construyeron diversos factores de expansión para la encuesta Tercer Barómetro de la Felicidad 2014, donde se variará la metodología de cálculo del ponderador base y la de los ajustes de no respuesta, para luego comparar el desempeño de dichos factores en la estimación de una variable de interés de bienestar subjetivo de la encuesta Tercer Barómetro de la Felicidad.

De este modo, este trabajo intenta en primer lugar responder a la pregunta si factores de expansión que contemplen ajustes explícitos por no respuesta que incluyan diversos tipos de variables son más eficientes en reducir el sesgo y la varianza en comparación con factores que omiten o las probabilidades de selección, o las probabilidades de participar. En segundo lugar, se busca analizar los efectos de realizar ajustes de no respuesta en base a variables con un mayor acceso (que se pueden obtener fácilmente desde fuentes administrativas disponibles a todo público) y menor acceso (que se deben construir a través de un procesamiento avanzado de bases de datos más complejas), con tal de entregar algunas orientaciones prácticas a investigadores que se desempeñan en la materia.

A. Tercer Barómetro de la Felicidad

El Tercer Barómetro de la Felicidad 2014 es una encuesta cara a cara aplicada por encuestador, de 2.267 casos. La población objetivo del estudio son las personas de 18 a 90 años de edad residentes de las zonas urbanas de comunas con más de 25.000 habitantes de Chile. Su diseño muestral fue probabilístico, estratificado y multietápico, seleccionando en una primera etapa 94 unidades primaria de muestreo (UPM) que corresponde a partes urbanas de las comunas¹. En la segunda etapa las unidades secundarias de muestreo (USM) que son las manzanas censales. En la tercera etapa las unidades terciarias de muestreo

¹ De las 94 comunas seleccionadas, 63 de ellas fueron incluidas con una probabilidad igual a 1 (denominadas comunas con Inclusión forzosa, pues tenían más de 100.000 habitantes y/o eran capitales regionales).

(UTM) son las viviendas. Finalmente, las personas de 18 años y más corresponden a las últimas unidades de muestreo (UUM)².

El diseño muestral consideró una muestra de 2.800 casos, que consideró una sobremuestra del 30% para hacer frente a la no respuesta esperada en este tipo de encuestas. Así, se obtuvo una tasa de respuesta³ de 82,2% con 2.267 entrevistas completas. La tasa de rechazo⁴ correspondió a un 7,7%, habiendo 213 encuestas rechazadas en total. En total, se logró contacto con 2.480 casos, obteniendo una tasa de contacto⁵ del 92,2%.

Para enfrentar dichos problemas de no respuesta y cobertura se construyó un ponderador original que consideró un ponderador base y ajustes de calibración. El ponderador base⁶ para todas las personas entrevistadas (2.267) se construyó asumiendo muestreo aleatorio simple de casos sin reemplazo, asignando a cada elemento en la muestra el ponderador base igual a 1. El ponderador base fue posteriormente calibrado utilizando el método raking a los parámetros poblacionales de sexo, edad y tramos educacionales que se presentan en la Tabla 3.

El ponderador calibrado fue finalmente estandarizado para tener una media de 1 y desviación estándar de 0,45. La suma del ponderador estandarizado corresponde al tamaño

² Para mayores detalles respecto a la estructura de los datos de la encuesta Barómetro 2014, consultar sección en Anexos.

³ Corresponde a tasa AAPOR RR1, que considera el número de entrevistas completas partido por el total de casos elegibles seleccionados (viviendas particulares ocupadas). Para una aproximación más conservadora, esta tasa asume que aquellos casos con elegibilidad desconocida son elegibles (N = 129).

⁴ Corresponde a tasa AAPOR REF1, que considera la proporción de rechazos (se considera rechazo hogar y persona) del total de casos elegibles (viviendas particulares ocupadas). Para una aproximación más conservadora, esta tasa asume que aquellos casos con elegibilidad desconocida son elegibles (N = 129).

⁵ Corresponde a tasa AAPOR COOP1, que considera la proporción de casos en donde se llegó a algún tipo de contacto (ya sea encuesta completa o rechazo), del total de casos elegibles de la muestra. Para una aproximación más conservadora, esta tasa asume que aquellos casos con elegibilidad desconocida son elegibles.

⁶ Acceso a la sintaxis de cálculo del ponderador fue provisto por DESUC, entidad responsable del levantamiento de la encuesta y cálculo de los ponderadores.

muestral de 2.267 casos. Las estadísticas descriptivas del ponderador se pueden ver en la Tabla 4.

Tabla 3. Parámetros poblacionales utilizados como insumo para la calibración del ponderador Barómetro 2014

Variable	Atributo	%
Sexo	Hombre	0,49
	Mujer	0,51
Macrozona	Norte	0,12
	Centro	0,21
	Sur	0,26
	Región Metropolitana	0,41
Edad	18 a 29 años	0,26
	30 a 45 años	0,31
	46 a 65 años	0,31
	66 años o más	0,12
Educación del encuestado	Superior completa	0,31
	Escolar completa	0,31
	Escolar incompleta	0,38

Nota: proporciones para sexo, macrozona y edad se obtuvieron de Proyecciones de población de INE para el año 2013 ajustadas a población urbana. Proporción de educación del encuestado se obtuvo de NENE (oct-nov-dic 2013) ajustada a población urbana.

Información provista desde sintaxis de cálculo del ponderador de DESUC.

Tabla 4. Estadísticas descriptivas del ponderador original de Barómetro 2014

Promedio	Desviación estándar	Mínimo	Máximo	Suma	N muestra
1,0	0,45	0,3	3,2	2.267	2.267

Fuente: Estimación propia a partir de base de datos del estudio Barómetro 2014.sav

i. Medición de bienestar subjetivo

La variable de interés del estudio corresponde a la medición de bienestar subjetivo, que tiene exactamente el mismo enunciado, y alternativas de respuesta en las encuestas Barómetro 2014 y Casen 2013 (encuesta a utilizar como parámetro en tanto corresponde a la encuesta de mayor envergadura en Chile): *¿Cuán satisfecho está usted con su vida en este momento? Por favor, use esta tarjeta en que 1 significa que está “completamente*

insatisfecho” y 10 significa que usted está “completamente satisfecho”. ¿Dónde se ubica usted? De este modo, se realizaron estimaciones del promedio de dicha variable, tratando como casos perdidos los No sabe y No responde⁷.

Ahora, en la encuesta Casen la pregunta referida al bienestar subjetivo es respondida solo por los adultos que se encuentran presentes al momento de la entrevista. Debido a ello, el Ministerio de Desarrollo Social construyó un factor de expansión particular para la pregunta por bienestar subjetivo, que pretende corregir posibles sesgos de selección en las respuestas de los adultos que se encontraban presentes al momento de la entrevista⁸.

Para efectos de los análisis, se estimó el promedio de la satisfacción con la vida solo para la población comparable a la encuesta de Barómetro 2014, es decir, personas de 18 a 90 años de las 94 comunas incluidas en la encuesta Barómetro (donde no se observaron diferencias en la distribución de la muestra completa Casen, y la submuestra analítica). La distribución de dicha variable en conjunto con la distribución en Barómetro se puede ver en la Tabla 5.

⁷ Dado que el porcentaje de casos perdidos por No sabe – No responde era menor al 3%, se optó por dejar estos casos fuera del análisis, siguiendo práctica de Peng y Dong (2013).

⁸ El ponderador de la pregunta de satisfacción "Expr_r20" de Casen 2013 utiliza un ajuste de no respuesta para corregir por la probabilidad de "no estar presente" durante la entrevista Casen. El factor de ajuste se realizó utilizando el método del propensity score, que ajustó un modelo de la propensión a estar presente en la entrevista en base a 11 variables disponibles en la encuesta Casen tanto para los que estaba presentes como lo no presentes. En particular, se utilizaron; si la encuesta se realizó o no durante el fin de semana, si la persona tiene alguna condición permanente de discapacidad, si la persona es mayor de 15 años, la región de residencia, la zona de residencia, las categorías de ocupación de la persona, asistencia a un establecimiento educacional, si es jefe de hogar o pareja del jefe de hogar, el estado civil, tramos de edad de 10 años, años de escolaridad e ingresos totales per cápita del hogar. Para más detalles, ver documento http://observatorio.ministeriodesarrollosocial.gob.cl/casenmultidimensional/casen/docs/Metodologia_factor_expansion_orientacion_sexual_Identidad_genero.pdf, que sigue un procedimiento análogo.

Tabla 5. Distribución y estadísticas descriptivas de variable de bienestar subjetivo en Encuesta Casen 2013 base completa, subpoblación equivalente a Barómetro 2014, y Barómetro 2014

	Casen 2013 base completa %	Casen 2013 subpoblación Barómetro 2014 %	Barómetro 2014
Completamente insatisfecho	1,3	1,4	0,9
2	1,0	1,0	0,7
3	1,4	1,5	1,9
4	3,7	3,7	4,5
5	12,0	11,8	10,6
6	11,0	10,6	10,4
7	15,4	15,0	16,2
8	17,5	17,6	21,0
9	11,4	11,6	15,2
Completamente satisfecho	25,4	25,8	18,6
Media	7,52	7,53	7,5
Desviación estándar	2,1	2,1	2,0
N	87.006	67.748	2.267

Nota: estimaciones Casen ponderadas utilizando factor de expansión Expr_r20 para pregunta sobre bienestar subjetivo y estimaciones Barómetro 2014 ponderadas con factor original POND_B.

B. Construcción de factores de expansión alternativos

Para este trabajo se construyeron ocho factores de expansión que varían según la estrategia de cálculo utilizado para la construcción de ponderador base, ajustes de no respuesta, y ajustes de calibración. La Tabla 6 presenta las características generales de cada uno.

Los ponderadores *W1* y *W2* utilizan la estrategia A planteada anteriormente, donde ignoran el diseño muestral complejo de la encuesta, asumiendo igual probabilidad de selección para los elementos en la muestra. El ponderador *W1* asume $prob=1$, donde su suma corresponde al tamaño de la muestra (2.267 casos), mientras que el ponderador *W2* asume $prob=n/N$, expandiendo a la población objetivo (10.499.821 individuos).

Los ponderadores *W3* a *W8* se construyeron reflejando las características del diseño muestral complejo efectivamente empleado, que considera distintas probabilidades de selección, estratificación y conglomeración.

En los ponderadores *W1*, *W2* y *W3* se realizó un ajuste implícito de no respuesta (donde *W3* tuvo una estrategia B de ponderación), mientras que los ponderador *W4* a *W8* se realizó un ajuste explícito a partir de variables comunales y de zonas censales provenientes de distintas fuentes (administrativas del Ministerio de Desarrollo Social y Censo 2012).

Con dichas variaciones se trata de evaluar si, tal como se señala en la literatura, usar variables más cercanas al fenómeno de interés tiene mayor poder predictivo. Y en ese sentido, se realizará la comparación entre las agrupaciones de las mismas variables a nivel de zonas censal vs comuna. Por otro lado, también interesa observar si las variables disponibles comunales (MDS) pueden ser sustitutos razonables de las variables creadas del Censo, ya que estas últimas no se encuentran de manera tan accesible pues implica el procesamiento avanzado y construcción a partir de una base de datos más compleja.

En específico, el ponderador *W4* usa variables auxiliares de nivel comunal disponibles a partir de datos administrativos públicos (ver MDS 2014).

El ponderador *W5* agrega variables auxiliares de nivel comunal de elaboración propia, creadas a partir de datos del Censo 2012.

El ponderador *W6* usa variables auxiliares de nivel de zona censal, de elaboración propia, utilizando los datos del Censo 2012.

El ponderador *W7* usa variables auxiliares de nivel de comuna y zona censal, de elaboración propia a partir de datos del Censo 2012. Las variables a nivel de comuna fueron generadas cuidando de no crear autocorrelaciones de los datos entre variables creadas a nivel de comuna y zona censal de agregación anidados a partir de la misma fuente de datos. Para ello, se siguió la metodología planteada por Suzuki et al, 2012⁹.

⁹ La metodología consiste en realizar las estimaciones de las diversas variables de interés al nivel más alto, excluyendo las estimaciones para dichas variables en los niveles más bajos que la componen. En este caso, para realizar los cálculos de variables a nivel de comuna, se excluyeron las zonas censales seleccionadas en la muestra de dichas comunas.

Por último, en el ponderador *W8* se combinan las variables provenientes de Censo y Ministerio de Desarrollo Social tanto a nivel de comuna como de zona censal, para con ello evaluar la contribución adicional de todas las variables auxiliares.

Todos los ponderadores construidos incorporaron ajustes de calibración a totales poblaciones ajustados para sexo, tramos etarios (4), macrozona (4) y nivel educacional (3). Para revisar en detalle el procedimiento seguido en cada una de estas etapas revisar secciones i, ii, iii y iv de Anexos.

Tabla 6. Factores de expansión contruidos

Nombre Factor de expansión	Descripción			Fórmula
	Ponderador base (wb)	Ajuste de NR (Aj_NR)	Ajustes de calibración (Aj_CAL)	
w1: UNO_CAL	Igual a 1 para todos los elementos	No	Raking a totales de sexo, edad, macrozona y nivel educacional	$1 * Aj_{calib}$
w2: MAS_CAL	Igual a N/n asumiendo MAS para todos.	No	Raking a totales de sexo, edad, macrozona y nivel educacional	$\left(\frac{N}{n}\right) * Aj_{calib}$
w3: PSEL_CAL	Según diseño complejo empleado	No	Raking a totales de sexo, edad, macrozona y nivel educacional	$wb_{personas} * Aj_{calib}$
w4:NR_COM_MDS	Según diseño complejo empleado	Sí, ajuste de celdas en función a variables de MDS 2012-2013	Raking a totales de sexo, edad, macrozona y nivel educacional	$wb * Aj_{NR_mds} * Aj_{calib}$
w5:NR_COM	Según diseño complejo empleado	Sí, ajuste de celdas en función a variables a nivel de comuna de MDS 2012-2013 y Censo 2012	Raking a totales de sexo, edad, macrozona y nivel educacional	$wb * Aj_{NR_com} * Aj_{calib}$
w6:NR_ZON	Según diseño complejo empleado	Sí, ajuste de celdas en función a variables a nivel de zona censal de Censo 2012	Raking a totales de sexo, edad, macrozona y nivel educacional	$wb * Aj_{NR_zon} * Aj_{calib}$
w7:NR_FULL_CENSO	Según diseño complejo empleado	Sí, ajuste de celdas en función a variables a nivel de comuna y zona censal de Censo 2012	Raking a totales de sexo, edad, macrozona y nivel educacional	$wb * Aj_{NR_censo} * Aj_{calib}$
w8:NR_FULL	Según diseño complejo empleado	Sí, ajuste de celdas en función a variables a nivel de comuna y zona censal de MDS 2012-2013 y Censo 2012	Raking a totales de sexo, edad, macrozona y nivel educacional	$wb * Aj_{NR_full} * Aj_{calib}$

C. Desempeño de factores de expansión

El modo de evaluar el desempeño de los ponderadores se realiza en torno al equilibrio buscado entre la reducción del sesgo y la variabilidad de las estimaciones (Kreuter, et al., 2010; Little & Vartivarian, 2005).

Para el primer aspecto es necesario establecer un gold standard respecto al cual se puedan comparar las distintas estimaciones ponderadas bajo evaluación. En este caso se medirá el sesgo relativo, donde se utilizará como aproximación al valor verdadero (gold standard¹⁰) la satisfacción con la vida estimada a partir de la encuesta Casen 2013, Así las estimaciones con menor valor de sesgo relativo indicarían un mejor desempeño de los ponderadores asociados a dicha medida para la reducción del sesgo (Kreuter, et al., 2010; Little & Vartivarian, 2005; Wun & Ezzati-Rice, 2007). Siguiendo a Kreuter et al (2010), el sesgo relativo se expresará en:

$$r\widehat{Bias} = \left| \left(\frac{\widehat{X}_{factorBarómetro} - \bar{X}_{Casen}}{\bar{X}_{Casen}} \right) * 100 \right|$$

Donde,

$\widehat{X}_{factorBarómetro}$: Promedio estimado de variable de bienestar subjetivo utilizando los factores de expansión W1 a W8 con datos de la encuesta Barómetro 2014

\bar{X}_{Casen} : Promedio estimado de variable de bienestar subjetivo en Casen 2013 para población comparable de Barómetro 2014.

En segunda instancia, se utilizará el Efecto Diseño (Carlson & Williams, 2001) de las estimaciones para la encuesta Barómetro de la Felicidad 2014 realizadas con los distintos ponderadores W1 a W8. El efecto diseño estima la varianza en la estimación al utilizar el

¹⁰ Se señala que corresponde a un proxy de gold standard en tanto no existen estimaciones inequívocas sobre la variable, por lo cual no corresponde a un “valor verdadero” como se entiende tradicionalmente al gold standard.

factor de expansión específico, en vez de haber asumido un diseño muestral basado en el muestreo aleatorio simple (Lohr, 2009; Kish, 1965), como se observa en la siguiente fórmula. Los efectos diseño de las estimaciones realizadas a partir de la encuesta Barómetro 2014 se calcularon en el programa estadístico Stata asumiendo el diseño muestral complejo, utilizando como variable de estrato a las regiones, y como variable de conglomerado a las manzanas¹¹.

$$\widehat{DEFF} = \frac{\widehat{Var}_{estimación\ factorx}}{\widehat{Var}_{estimación\ MAS}}$$

Donde,

$\widehat{Var}_{estimación\ factorx}$: Varianza estimada para variable de bienestar subjetivo utilizando el factor de expansión específico en Barómetro 2014

$\widehat{Var}_{estimación\ MAS}$: Varianza estimada para variable de bienestar subjetivo bajo supuesto de Muestreo Aleatorio Simple

En tercer lugar, se utilizará la medida del Error Cuadrático Medio, donde se conjuga la medida de sesgo y la medida de varianza, para de este modo evaluar el trade-off que entre la reducción del sesgo y el aumento de la varianza (Little & Vartivarian, 2005; Kreuter, y otros, 2010), expresado en:

$$\widehat{EMC} = \left(\widehat{X}_{factorx} - \bar{X}_{Casen} \right)^2 + \widehat{Var}_{factorx_barómetro}$$

Los tres indicadores se calcularán para el total y las siguientes subpoblaciones de interés relacionadas con el bienestar subjetivo; sexo (2), tramos etarios (4), macrozona (4) y nivel educacional del encuestado (4) y situación ocupacional del encuestado (3).

¹¹ Siguiendo la práctica de Casen de utilizar el último conglomerado para reportar el diseño muestral complejo en una encuesta.

De esta manera, se compararán estos tres indicadores entre los factores de expansión con ajuste de no respuesta explícitos, y el resto de los factores, y adicionalmente, se analizará el comportamiento entre los factores con ajuste explícito de no respuesta, para evaluar las implicancias de la utilización de distintos sets de variables para realizar el ajuste.

4 Resultados

A. ¿Qué tan adecuadas son las variables auxiliares disponibles para los ajustes de No Respuesta?

El Gráfico 1 muestra las correlaciones entre las variables utilizadas para realizar los ajustes de no respuesta, tanto con la respuesta en la encuesta (disponible para todos los elementos de la muestra), como con la variable de bienestar subjetivo (disponible solo para respondentes).

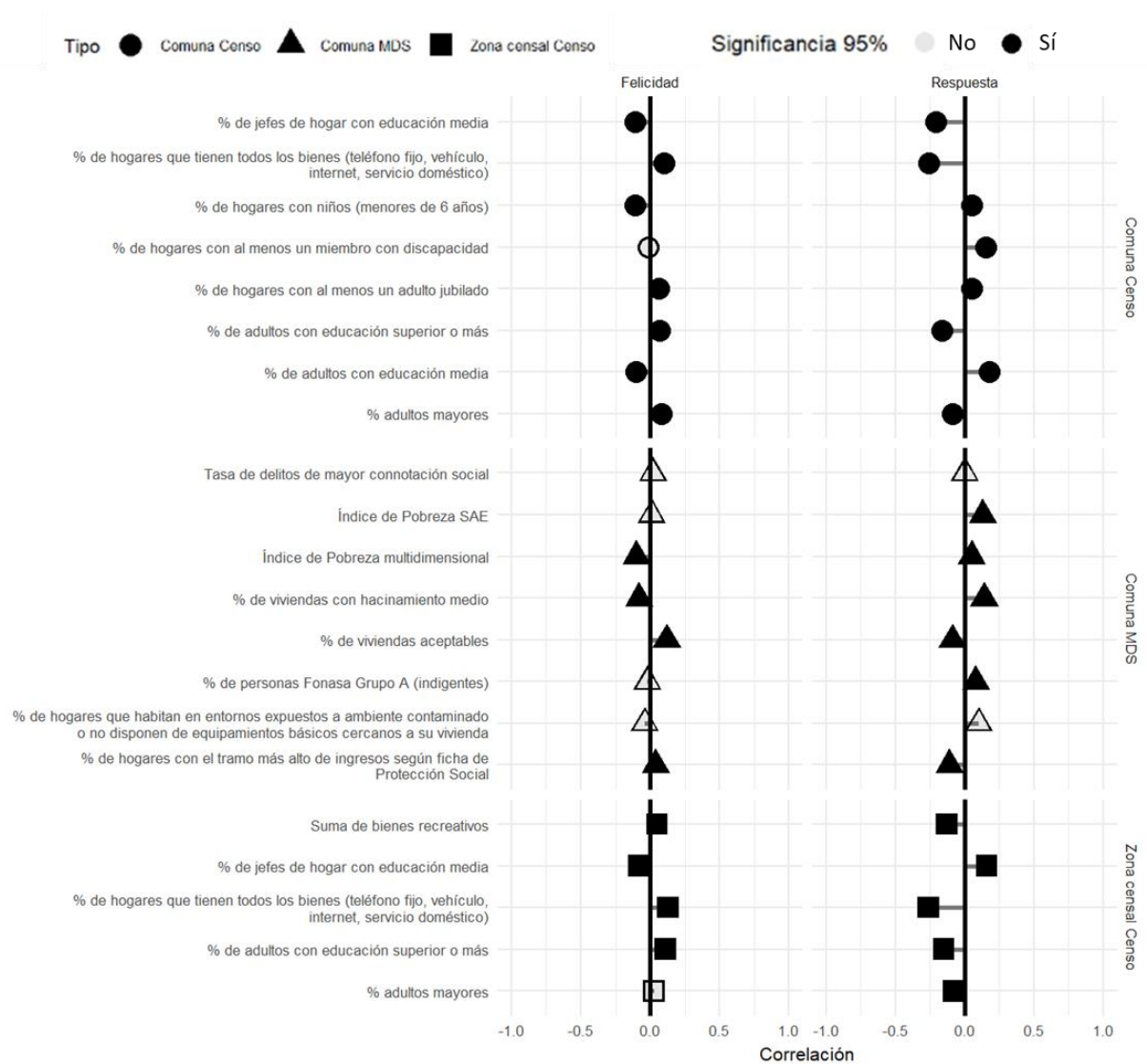
Como se puede ver, las variables tienen asociaciones estadísticamente significativas al 95% de nivel de confianza con la respuesta en la encuesta entre el orden del 0,05 y 0,26, siendo las más grandes en magnitud el porcentaje de hogares que tienen todos los bienes preguntados en Censo 2012¹² tanto a nivel de comuna como de zona censal (ambos con un $r=0,26$), seguido del porcentaje de jefes de hogar con educación media completa a nivel de comuna calculado a partir de Censo 2012 ($r = 0,21$).

En el caso del bienestar subjetivo, las correlaciones son más bajas (entre 0,04 y 0,13), pero aun estadísticamente significativas. Destaca nuevamente el porcentaje de hogares a nivel de zona que tienen todos los bienes preguntados en Censo 2012 ($r = 0,13$), seguido por el porcentaje de viviendas en la categoría de aceptables a nivel de comuna a partir de información de MDS ($r = 0,12$).

Dichas correlaciones se encontraron dentro de los rangos hallados por otros investigadores, como Kreuter (2007), Casas-Cordero (2010) y Kreuter et al. (2010)., con correlaciones entre 0,20 y 0,50. Para ver en detalle las correlaciones analizadas, consultar Tabla 18 y Tabla 19 en Anexos.

¹² Bienes consultados en Censo 2012 corresponden a vehículo, teléfono fijo, internet en el hogar y servicio doméstico).

Gráfico 1. Correlación bivariada entre variables de ajuste de no respuesta, y variable de respuesta y bienestar subjetivo



B. ¿Qué tan distintos son los ponderadores desarrollados?

La Tabla 7 muestra las estadísticas descriptivas de los factores de expansión construidos, donde se evidencian las diferencias entre el ponderador original de la encuesta, con el resto de los factores elaborados. En particular, se puede ver que de éstos el de menor variabilidad (DE = 2.061) corresponde al factor *MAS_CAL*, que ignora el diseño muestral complejo y no considera ajustes explícitos de no respuesta. En el otro extremo, el factor con la mayor variabilidad es el que incluye ajustes de no respuesta en base a variables comunales.

Tabla 7. Estadísticas descriptivas de factores de expansión

Factor de expansión	Media	Mínimo	Máximo	Desviación estándar	Suma
W1: UNO_CAL	1	0,3	3,2	0,45	2.267
W2: MAS_CAL	4.631	1.320	13.791	2.062	10.499.821
W3: PSEL_CAL	4.197	712	24.115	2.541	10.499.821
W4: NR_COM_MDS	4.631	517	43.066	3.643	10.499.821
W5: NR_COM	4.631	468	45.949	3.728	10.499.821
W6: NR_ZON	4.631	454	42.750	3.567	10.499.821
W7: NR_FULL_CENSO	4.631	441	37.711	3.503	10.499.821
W8: NR_FULL	4.631	407	38.819	3.600	10.499.821

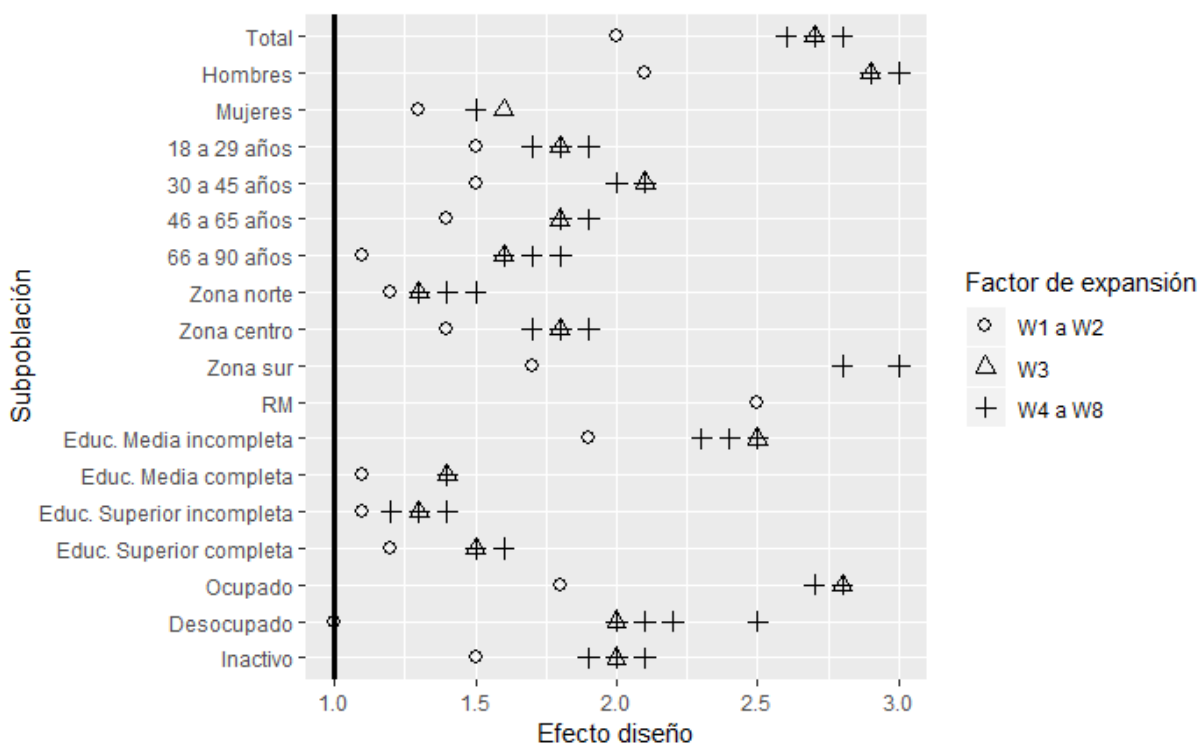
C. Impacto de los ponderadores en la precisión de las estimaciones

En el Gráfico 2 se muestran los indicadores de Efecto Diseño (DEFF) para las estimaciones de bienestar subjetivo en distintas subpoblaciones. Como es de esperar, los Efecto Diseño más cercanos a 1 se encuentran en los factores de expansión que asumen un diseño muestral base simplificado (MAS) y no incorporan ajustes de no respuesta, lo que corresponde al modelo más simplificado y poco realista de ponderación (factores $W1:UNO_CAL$ y $W2:MAS_CAL$).

El Gráfico 2 también muestra que no se observan diferencias importantes entre los DEFFs calculados para los ponderadores con ajuste de NR ($W4$ a $W8$) y el ponderador que no tiene ajuste NR pero sí usa como base las probabilidades efectivas de selección de la muestra ($W3:PSEL_CAL$). Dicho de otra forma, el mayor impacto en los DEFFs viene dado por la incorporación de las probabilidades de selección efectivas en el cómputo del ponderador, y no por la incorporación del ajuste de no respuesta (cualquiera de ellos). Para ver los estimadores puntuales, consultar Tabla 22 en sección de Anexos.

A su vez, también destaca que en general el Efecto Diseño entre los distintos factores de expansión con ajuste explícito de no respuesta son equivalentes, no mostrando grandes variaciones entre los que incorporan más variables en el ajuste.

Gráfico 2. Efecto Diseño de tipos de factores de expansión contruidos para promedio de variable de interés estimado para total y subpoblaciones de sexo, edad, macrozona, nivel educacional y situación ocupacional del encuestado



D. Impacto de los ponderadores en la acuracidad de las estimaciones

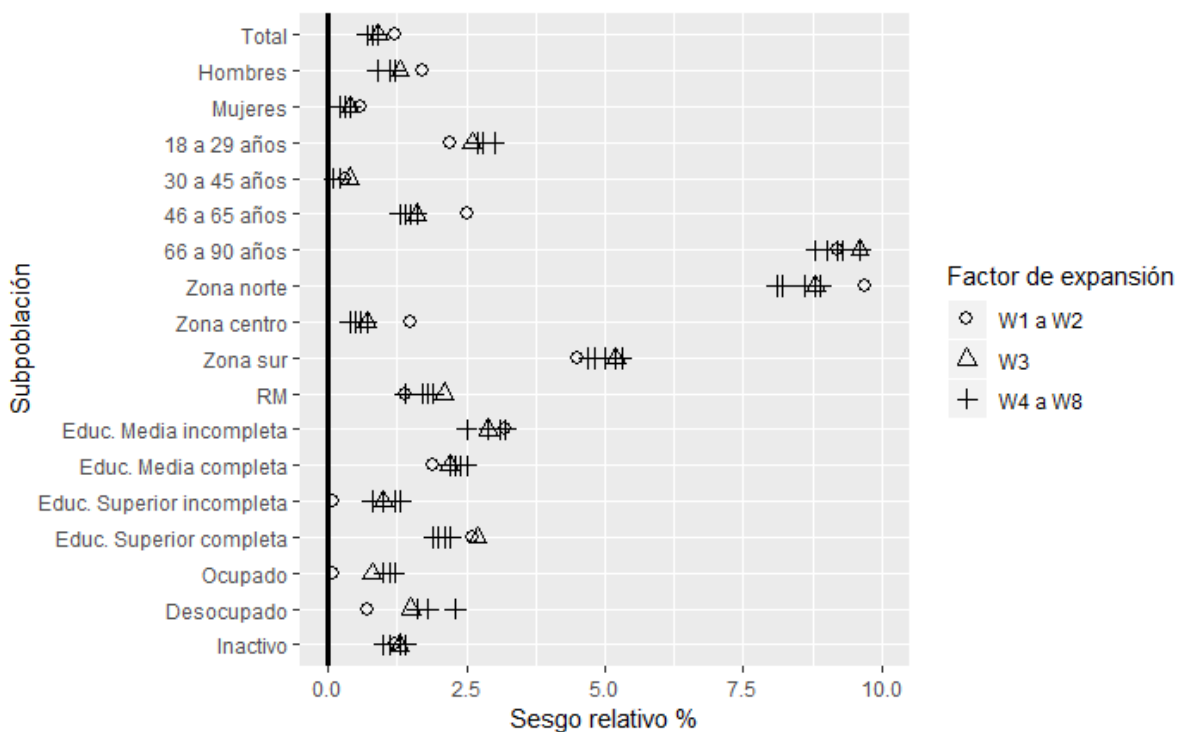
El Gráfico 3 muestra el indicador de sesgo relativo de las estimaciones del promedio de la variable de interés, respecto a dicho promedio para la Encuesta Casen 2013 para distintas subpoblaciones.

Los resultados muestran que los factores de expansión que incluyen ajustes explícitos de no respuesta (*W4 a W8*) tienen un menor o igual sesgo relativo en 12 de las 18 estimaciones realizadas, si se comparan con aquellos factores de expansión que no incluían el ajuste (*W1 a W3*). Además, en general, aquellas estimaciones con mayor sesgo relativo, este no era

notoriamente mayor que el factor de expansión ajustado por no respuesta que mostró el menor sesgo relativo el indicador.

A su vez, en las columnas 6 a 9 de la Tabla 23 de la sección de Anexos se muestra el comportamiento de los factores de expansión con distintos ajustes de no respuesta. Aquí se aprecia que los menores indicadores de sesgo relativo los obtuvo el factor de expansión FULL, donde se utilizaron las 13 variables auxiliares evaluadas. Seguido por él, se encontró el factor que utilizó solo variables a nivel de las zonas censales (*NR_ZON*), y aquel que utilizó un set de variables provenientes del Censo (*NR_FULL_CENSO*).

Gráfico 3. Sesgo relativo de factores de expansión construidos para promedio de satisfacción con la vida para total y subpoblaciones de sexo, edad, macrozona, nivel educacional y situación ocupacional del entrevistado



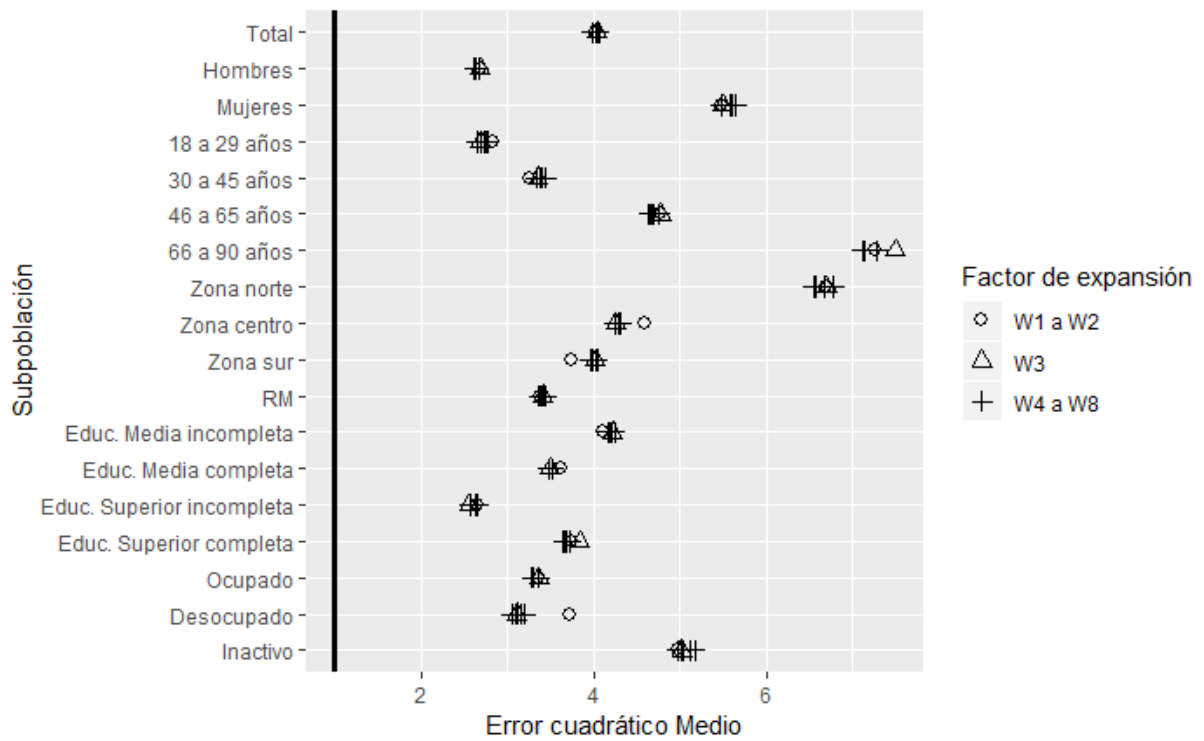
E. Impacto de los ponderadores en el Error Cuadrático Medio de las estimaciones

En el Gráfico 4 se pueden ver las estimaciones del Error Cuadrático Medio para el promedio de la variable de interés según las mismas subpoblaciones presentadas anteriormente, mostrando de esta manera un balance entre el sesgo y la varianza de la estimación.

El principal resultado que se evidencia radica en que en general los factores de expansión con ajustes explícitos de no respuesta ($w4$ a $w8$) obtuvieron los menores errores cuadráticos medios en 14 de las 18 estimaciones realizadas para la variable de bienestar subjetivo, mientras que el factor de expansión ingenuo obtuvo el error cuadrático medio menor en cinco estimaciones ($w1$ a $w2$), y el factor con ajuste implícito de no respuesta solo en una ($w3$).

En específico, dentro de los factores de expansión con ajuste explícitos de no respuesta, en la Tabla 24 de la sección de Anexos se ve que el de mejor performance corresponde al que utilizó un modelo global para el ajuste de no respuesta (incorporando variables de comunas y zonas, de fuentes administrativas y a partir del Censo), donde se obtuvo el menor Error Cuadrático Medio en seis estimaciones. Lo sigue el factor de expansión con ajuste mediante variables de comunas -con cuatro estimaciones con el Error Cuadrático Medio menor- y seguido de ello, el factor con un ajuste a partir de variables administrativas, donde en tres estimaciones se obtuvo el Error Cuadrático Medio más bajo.

Gráfico 4. Error Cuadrático Medio de factores de expansión construidos para promedio de variable de interés estimado para total y subpoblaciones de sexo, edad, macrozona, nivel educacional y situación ocupacional del encuestado



5 Discusión

Los resultados expuestos en la sección anterior entregan pistas claves respecto a la importancia e implicancias que tiene la realización de factores de expansión que involucren ajustes explícitos de no respuesta para las estimaciones que se realicen con los datos de las encuestas.

En primer lugar, la mayoría de los factores de expansión con ajuste explícito de no respuesta- y en la mayoría de las subpoblaciones- obtuvieron menores sesgos que aquellos factores que no incluían el ajuste explícito, lo cual se encuentra en línea por lo planteado por otras investigaciones (Wun & Ezzati-Rice, 2007; Little & Vartivarian, 2003; Bethlehem, 2002; Kalton & Flores-Cervantes, 2003; Lepkowski, Mosher, & Davis, 2006). Este resultado se debe poner en contexto en tanto las correlaciones entre las variables utilizadas en el ajuste, y la variable de respuesta y bienestar subjetivo se encontraron dentro del rango empírico hallado por otros investigadores, como Kreuter, que encuentra correlaciones cercanas a de hasta 0,20 (2007), Casas-Cordero con correlaciones que alcanzan el 0,4 (2010), aunque en algunos casos pudieron encontrarse correlaciones más altas, siendo cercanas a 0,5 (Kreuter, y otros, 2010).

En segundo lugar, y respecto al Efecto Diseño en las estimaciones, naturalmente el factor de expansión más simplificado tuvo menores Efecto Diseño. Ahora bien, los resultados también mostraron que la utilización de factores de expansión más complejos en su construcción en variadas estimaciones no aumentó de manera drástica el Efecto Diseño. De hecho, el mayor aumento en los DEFFs se ve explicado por la inclusión de ponderador base asumiendo un diseño complejo, y no por los ajustes explícitos de no respuesta en los factores. Esto indica, por ende, que las estimaciones realizadas con dichos factores no aumentaron en demasía su varianza. Este resultado mostraría que, tal como se menciona en la literatura, al realizar ajustes de no respuesta basados en variables que se escogieron por su vinculación con la participación en la encuesta y con la variable de interés, permite construir factores de expansión que no incrementen demasiado la varianza (Kreuter, y otros, 2010; Wun & Ezzati-Rice, 2007; Little & Vartivarian, 2003).

En línea con lo anterior, los resultados vinculados al Error Cuadrático Medio permitieron mostrar que los factores de expansión con ajuste explícito de no respuesta obtuvieron en general la mejor performance considerando el trade-off entre reducción de sesgo y aumento de la varianza, entregando indicadores más certeros en comparación con otros métodos de expansión más simplificados, lo que se condice con lo planteado por la literatura.

Ahora, al comparar los ajustes de no respuesta con diversos sets de variables, a pesar de que el factor que utilizó el set completo de variables obtuvo la mejor performance en general, el resto de los factores con ajuste explícito no se distanciaron notoriamente ni en la reducción del sesgo, ni en menores Efecto Diseño, ni en Error Cuadrático Medio más pequeños. En base a ello, es que a pesar de que la utilización de variables tanto de comunas como de sectores más cercanos a la vivienda (por ejemplo zonas censales), contribuyan a construir factores de expansión más eficientes, los resultados no permiten establecer que un tipo de variable sea notoriamente más deseable que otra para realizar los ajustes de no respuesta, ni tampoco la necesidad de incluir todas las variables para obtener resultados satisfactorios.

En ese sentido, los resultados muestran que para el caso específico de Barómetro 2014 y de la variable de estudio, realizar un ajuste de no respuesta explícito- independiente de las variables que se utilicen y la fuente de la que provengan- contribuirá a lograr mejores estimaciones que un factor de expansión que realice un ajuste implícito.

Esto es importante en un contexto donde los estudios no siempre consideran la elaboración de factores de expansión complejos, lo cual puede deberse principalmente a falta de información para la realización de los ajustes necesarios, especialmente en el caso de la no respuesta, donde se necesita que las variables sobre las cuales se realizará el ajuste se encuentren disponibles tanto para respondentes, como para no respondentes. También algunas encuestas no han modificado la metodología de construcción de sus ponderadores, pues viene establecida desde variadas versiones anteriores, y modificarla implicaría la pérdida de comparabilidad con dichas versiones (por ejemplo, la Nueva Encuesta Nacional de Empleo). Mientras que otras encuestas, en cambio, en un escenario similar sí han hecho modificaciones, donde se incorporan ajustes explícitos de no respuesta. Ejemplo de ello son la encuesta Casen y la de Ocupación y Desocupación en el Gran Santiago.

En base a lo anterior, se releva la importancia de que investigadores que utilicen encuestas tengan fácil acceso a información que pueda ser relevante para realizar ajustes de no respuesta, e idóneamente, a diversos niveles de proximidad con las viviendas, tales como zonas censales, barrios o manzanas.

Ahora bien, los resultados también mostraron que la realización de ajustes sin variables censales dieron paso igualmente a factores de expansión con buena performance, y considerando ello, investigadores que no cuenten con acceso a dichas variables pueden realizar ajustes eficientes de igual manera.

Considerando todo ello, es que este trabajo viene a reforzar la idea ya consensuada en la literatura de la importancia de realizar ajustes explícitos de no respuesta que puedan afrontar en parte los posibles sesgos derivados de la no respuesta. En esta misma línea, los resultados plantean la importancia de nuevas investigaciones que se centren en la búsqueda de variables adecuadas para realizar dichos ajustes y el mecanismo que explica su adecuación en cada caso particular, considerando las ventajas que se derivarían de realizarlos con variables con altos niveles de correlación con la respuesta y variables de interés del estudio.

Finalmente, se debe considerar también que los análisis precedentes se realizaron con información disponible del Censo 2012, por lo cual para replicar estos análisis con investigaciones más reciente es ampliamente recomendable que se utilicen los datos del Censo 2017 y/o información proveniente de fuentes administrativas más actualizadas.

Además, se debe destacar que el presente trabajo se centró en una encuesta focalizada en felicidad y bienestar subjetivo, por lo tanto, los resultados y conclusiones aquí expuestos responden a dicha variable. Dado eso, es importante relevar la importancia de realizar estudios particulares para cada encuesta en la que se quieran aplicar factores de expansión que ajusten explícitamente por la no respuesta, pues las variables que en este estudio resultaron útiles para realizar los ajustes, no necesariamente lo serán para otros estudios.

6 Bibliografía

- Andridge, R. R., & Little, R. J. (2010). A review of hot deck imputation for survey non-response. *International statistical review*, 78(1), 40-64.
- Asparouhov, T. (2005). Sampling weights in latent variable modeling. *Structural equation modeling*, 12(3), 411-434.
- Atrostic, B. K., Bates, N., & Silberstein, A. (2001). Nonresponse in US government household surveys: consistent measures, recent trends, and new insights. *Journal of Official Statistics*, 17(2), 2009.
- Bethlehem, J. G. (2002). *Weighting nonresponse adjustments based on auxiliary information*.
- Brick, J. (2013). Unit Nonresponse and Weighting Adjustments:. *Journal of Official Statistics*, 29(3), 329–353.
- Carlson, B. L., & Williams, S. (2001). A comparison of two methods to adjust weights for non-response: propensity modeling and weighting class adjustments. *Proceedings of the annual meeting of the American Statistical Association*.
- Casas-Cordero, C. (2010). *Neighborhood characteristics and participation in household surveys*. University of Maryland, College Park.
- Castellano, R., & Punzo, G. (2012). The impact of territorial factors on the total non-response error in the European Union-Survey on Income and Living Condition (EU-SILC). *Statistica*, 72(2), 211.
- Curtin, R., Presser, S., & Singer, E. (2005). Changes in telephone survey nonresponse over the past quarter century. *Public opinion quarterly*, 69(1), 87-98.
- DeMaio, T. J. (1980). Refusals: Who, where and why. . *Public Opinion Quarterly*, 44(2), 223-233.

- Deville, J. C., & Särndal, C. E. (1992). Calibration estimators in survey sampling. *Journal of the American statistical Association*, 87(418), 376-382.
- Diener, E., Suh, E. M., Lucas, R. E., & Smith, H. L. (1999). Subjective well-being: Three decades of progress. *Psychological bulletin*, 125(2), 276.
- Dirección de Estudios Sociales UC - Instituto de la Felicidad de Coca-Cola. (2015). *Tercer Barómetro de la Felicidad. El bienestar subjetivo de los chilenos: la importancia de nuestros vínculos*. (M. Browne, & P. Velasco, Edits.) Santiago.
- Dolan, P., Peasgood, T., & White, M. (2008). Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being. *Journal of Economic Psychology*, 29, 94-122.
- Dong, Y., & Peng, C. Y. (2013). Principled missing data methods for researchers. *SpringerPlus*, 2(1), 222.
- Easterlin, R. A. (2003). Explaining happiness. *Proceedings of the National Academy of Sciences*, 100(19), 11176-11183.
- EQLS. (2011). *3rd European Quality of Life Survey Weighting Report*.
- Fordyce, M. W. (2005). A review of research on the happiness measures: A sixty second index of happiness and mental health. En *Citation classics from social indicators research* (págs. 373-399).
- Fowler, J. H., & Christakis, N. A. (2008). Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study. *Bmj*, 337, a2338.
- Goyder, J., Lock, J., & McNair, T. (1992). Urbanization effects on survey nonresponse: a test within and across cities. *Quality and Quantity*, 26(1), 39-48.

- Graham, C., & Felton, A. (2005). *Does inequality matter to individual welfare? An initial exploration based on happiness surveys from Latin America*. Washington, DC: The Brookings Institution.
- Groves, R. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, 70(5), 646-675.
- Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public opinion quarterly*, 646-675.
- Groves, R. M., & Couper, M. P. (2012). *Nonresponse in household interview surveys*. John Wiley & Sons.
- Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: a meta-analysis. *Public opinion quarterly*, 72(2), 167-189.
- Groves, R. M., Wagner, J., & Peytcheva, E. (2007). Use of interviewer judgments about attributes of selected respondents in post-survey adjustment for unit nonresponse: An illustration with the National Survey of Family Growth. *Proc. Surv. Res. Meth. Sect. Am. Statist. Ass*, 3428-3431.
- Groves, R., & Couper, M. (1998). *Nonresponse in Household Interview Surveys*.
- Groves, R., Cialdini, R., & Couper, M. (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly*, 56, 475-495.
- House, J. S., & Wolf, S. (1978). Effects of urban residence on interpersonal trust and helping behavior. *Journal of Personality and Social psychology*, 36(9), 1029.
- Hox, J., De Leeuw, E., Couper, M., Groves, R., De Heer, W., Kuusela, V., . . . Martin, J. (2002). The influence of interviewers' attitude and behavior on household survey nonresponse: An international comparison. *Survey nonresponse*, 103-120.

- Instituto Nacional de Estadísticas. (2013). *Demográficas y Vitales*. Obtenido de Productos estadísticos:
http://historico.ine.cl/canales/chile_estadistico/familias/demograficas_vitales.php
- Kahneman, D., & Krueger, A. B. (2006). Developments in the Measurement of Subjective Well-Being. *The journal of economic perspectives*, 20(1), 3-24.
- Kalton, G., & Flores-Cervantes, I. (2003). Weighting Methods. *Journal of Official Statistics*, 19(2), 81-97.
- Kish, L. (1965). *Survey Sampling Principles*. New York: Marcel Dekker, Inc.
- Kreuter, F., Lemay, M., & Casas-Cordero, C. (2007). Using proxy measures of survey outcomes in post-survey adjustments: Examples from the European Social Survey (ESS). *Proc. Surv. Res. Meth. Sect. Am. Statist. Ass*, 3142-3149.
- Kreuter, F., Olson, K., Wagner, J., Yan, T., Ezzati-Rice, T., Casas-Cordero, C. L., . . . Raghunathan, T. (2010). Using proxy measures and other correlates of survey outcomes to adjust for non-response: examples from multiple surveys. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 173(2), 389-407.
- Latinobarómetro Org. (s.f.). *Latinobarómetro Database*. Recuperado el Octubre de 2017, de <http://www.latinobarometro.org/latContents.jsp>
- Lee, S. (2006). Propensity score adjustment as a weighting scheme for volunteer panel web surveys. *Journal of official statistics*, 22(2), 329.
- Lepkowski, J., Mosher, W., & Davis, K. (2006). National Survey of Family Growth, Cycle 6: Sample design, weighting, imputation, and variance estimation . *Vital Health Stat* , 2(142).
- Lipps, O., & Pollien, A. (2011). Effects of interviewer experience on components of nonresponse in the European Social Survey. *Field Methods*, 23(2), 156-172.

- Little, R. J. (1986). Survey nonresponse adjustments for estimates of means. *International Statistical Review/Revue Internationale de Statistique*, 139-157.
- Little, R. J., & Rubin, D. B. (2014). *Statistical analysis with missing data*. John Wiley & Sons.
- Little, R. J., & Vartivarian, S. (2003). On weighting the rates in non-response weights. *Statistics in medicine*, 22(9), 1589-1599.
- Little, R. J., & Vartivarian, S. (2005). Does weighting for nonresponse increase the variance of survey means? *Survey Methodology* 31(2), 161.
- Little, R. J., Lewitzky, S., Heeringa, S., Lepkowski, J., & Kessler, R. C. (1997). Assessment of weighting methodology for the National Comorbidity Survey. *American Journal of Epidemiology*, 146(5), 439-449.
- Lohr, S. (2009). *Sampling: design and analysis*. Nelson Education.
- Lyubomirsky, S., Sheldon, K. M., & Schkade, D. (2005). Pursuing happiness: The architecture of sustainable change. *Review of general psychology*, 9(2), 111.
- NORC at the University of Chicago. (s.f.). *The General Social Survey* . Recuperado el Octubre de 2017, de <http://gss.norc.org>
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41-55.
- Royston, P. (2004). Multiple imputation of missing values. *Stata journal*, 4(3), 227-41.
- Stoop, I. A. (2005). *The hunt for the last respondent: Nonresponse in sample surveys*. Sociaal en Cultureel Planbu.
- Suzuki, E., Yamamoto, E., Takao, S., Kawachi, I., & Subramanian, S. V. (2012). Suzuki, E., Yamamoto, E., Takao, S., Kawachi, I., & Subramanian, S. V. *PloS one*, 7(12), e51717.

- Valliant, R. (1993). Poststratification and conditional variance estimation. *Journal of the American Statistical Association*, 88(421), 89-96.
- Valliant, R., Dever, J. A., & Kreuter, F. (2013). *Practical tools for designing and weighting survey samples*. New York: Springer.
- Van Loon, A. J., Tijhuis, M., Picavet, H. S., Surtees, P. G., & Ormel, J. (2003). Survey non-response in the Netherlands: effects on prevalence estimates and associations. *Annals of epidemiology*, 13(2), 105-110.
- World Values Survey. (2017). *Banco de datos - MANAGING WEIGHTS AND POPULATION WEIGHTS WITHIN THE WVS*. Recuperado el Octubre de 2017, de <http://www.jdsurvey.net/jds/jdsurveyActualidad.jsp?Idioma=I&SeccionTexto=0405>
- Wun, L., & Ezzati-Rice, T. (2007). Assessment of the impact of health variables on nonresponse adjustment in the medical expenditure panel survey (MEPS). *Proc. Surv. Res. Meth. Sect. Am. Statist. Ass*, 2857-2864.

7 Anexos

Tabla 8. Referencias de encuestas chilenas e internacionales

Encuesta	Referencia de metodología
Encuesta Nacional de Derechos Ciudadanos - ENDC 2016	Informe final en: http://www.coordinacionoirs.gob.cl/wp-content/uploads/2017/07/ENDC-Informe-Final.pdf
Encuesta Nacional de Medio Ambiente - ENMA 2016	Informe final en: http://www.facso.uchile.cl/documentos/descarga-informe-completo-3era-encuesta-nacional-de-medio-ambiente-y-cambio-climatico_132429_0_0314.pdf .
Barómetro de la Felicidad 2014	Libro basado en informe final en: http://www.duna.cl/media/2015/10/Libro-Barometro-de-la-Felicidad.pdf
Encuesta Centro de Estudios Públicos – CEP 2017	Documento web de metodología en: https://www.cepchile.cl/cep/site/artic/20180517/asocfile/20180517155926/analisis_metodologico_encuestacep_pablo_marshall.pdf
Encuesta Nacional Bicentenario 2017	Documento web de metodología en: https://encuestabicentenario.uc.cl/wp-content/uploads/2017/10/Metodologia_EncuestaBicentenario_2017.pdf
Encuesta Nacional de Televisión – ENTV 2017	Informe final en: https://www.cntv.cl/cntv/site/artic/20180502/asocfile/20180502113330/ix_entv_final.pdf
Encuesta Nacional de Derechos Humanos – ENDH 2015	Informe final en: https://www.indh.cl/bb/wp-content/uploads/2017/01/Informe.pdf
Encuesta Nacional de Drogas en Población Escolar – ENPE 2015	Informe final en: http://www.senda.gob.cl/wp-content/uploads/media/estudios/PE/Informe%20Nacional%20Final%20ENPE%202015.pdf
Nueva Encuesta Nacional de Empleo – NENE 2018	Metodología en: http://historico.ine.cl/canales/chile_estadistico/mercado_del_trabajo/empleo/metodologia/pdf/031110/manual_metodologico031110.pdf
Nueva Encuesta Suplementaria de Ingresos –	Metodología en:

NESI 2017	http://historico.ine.cl/canales/chile_estadistico/mercado_del_trabajo/nene/nesi/archivos/antecedentes_metodologicos/manual_conceptual_y_metodologico_nesi.pdf
Encuesta Casen 2015	Metodología en: http://observatorio.ministeriodesarrollosocial.gob.cl/casen-multidimensional/casen/docs/Metodologia_de_Disen_Muestral_Casen_2015.pdf
Encuesta Nacional de Salud – ENS 2017	Presentación en: http://www.minsal.cl/wp-content/uploads/2017/11/ENS-2016-17_PRIMEROS-RESULTADOS.pdf
Encuesta Nacional de Calidad de Vida y Salud – ENCAVI 2017	Presentación en: http://www.minsal.cl/wp-content/uploads/2017/02/PRESENTACION-ENCAVI-2016-11.02.2017.pdf
Encuesta de Ocupación y Desocupación en el Gran Santiago 2018	Metodología en: http://www.microdatos.cl/Documentos/Encuestas/Ocupacion/1/Disen%CC%83o_muestral_VF.pdf
Encuesta de Protección Social – EPS 2015	Informe de factores de expansión: https://www.previsionsocial.gob.cl/sps/download/estudios-previsionales/encuesta-de-proteccion-social/documentos-eps/documentos-eps-2016/informe-metodologico-factoresexpansio%CC%81n-eps-2015.pdf
Encuesta Nacional de Polivictimización – ENP 2017	Presentación en: http://www.seguridadpublica.gov.cl/media/2018/02/PPT-Primera-Encuesta-de-Polivictimizacion.pdf
Encuesta Nacional de Violencia en el Ámbito Escolar – ENVAE 2014	Presentación en: http://www.seguridadpublica.gov.cl/media/2016/11/Presentaci%C3%B3n-ENVAE.pdf
Encuesta Nacional de Drogas en Población General – ENDPG 2016	Informe final en: http://www.senda.gob.cl/wp-content/uploads/2017/12/InformeENPG2016.pdf
Encuesta Nacional de Participación Cultural – ENDPG 2017	Informe de factores de expansión: http://www.cultura.gob.cl/wp-content/uploads/2018/03/informe-metodologico-factores-de-expansion.pdf
Encuesta Nacional Urbana de Seguridad Ciudadana. – ENUSC 2017	Metodología en: file:///C:/Users/crism/Downloads/documento-metodol%C3%B3gico-de-dise%C3%B1o-muestral-xiv-enusc-2017.pdf

American Community Survey – ACS 2017	Metodología en: https://www2.census.gov/programs-surveys/acs/methodology/design_and_methodology/acs_design_methodology_ch11_2014.pdf
Latinobarómetro 2016	Documento web en: http://www.latinobarometro.org/latContents.jsp
Climate Change in the American Mind 2018	Documento web en: http://climatecommunication.yale.edu/wp-content/uploads/2018/04/Climate-Change-American-Mind-March-2018-1.pdf
Public Perceptions of the Health Consequences of Global Warming 2014	Documento web en: http://climatecommunication.yale.edu/wp-content/uploads/2014/12/Global-Warming-Public-Health-October_2014_FINAL-1.pdf
A Baseline Human Rights Survey 2013	Documento web en: http://www.mv.undp.org/content/maldives/en/home/library/democratic_governance/SurveyReport.html
Eurobarometer 2018	Metodología en: https://surveyinsights.org/wp-content/uploads/2013/01/Methodological_Background.pdf
European Quality of life Survey – EQLS 2016	Metodología en: https://www.eurofound.europa.eu/sites/default/files/ef_publication/field_ef_document/ef1733en.pdf
European Working Conditions Survey	Metodología en: https://www.lu.lv/materiali/biblioteka/es/pilnieteksti/nodarbinatiba/Fourth%20European%20Working%20Conditions%20Survey.pdf
European Health Interview Survey (EHIS) 2016	Metodología en: https://ec.europa.eu/eurostat/cache/metadata/en/hlth_det_esms.htm
American Housing Survey – AHS 2015	Metodología en: https://www2.census.gov/programs-surveys/ahs/2015/2015%20AHS%20National%20Sample%20Design,%20Weighting,%20and%20Error%20Estimation.pdf

American Time Use Survey – ATUS 2018	<p>Metodología en:</p> <p>https://www.bls.gov/tus/atususersguide.pdf#nameddest=CHAPTER%207:%20WEIGHTS,%20LINKING,%20AND%20ESTIMATION</p>
National Crime Victimization Survey – NCVS 2016	<p>Metodología en:</p> <p>https://www.bjs.gov/index.cfm?ty=dcdetail&iid=245#Methodology</p>
Asian Barometer 2016	<p>Metodología en:</p> <p>http://www.asianbarometer.org/survey/survey-methods</p>
Labour Force Survey in the EU 2013	<p>Metodología en:</p> <p>https://ec.europa.eu/eurostat/documents/3888793/5858653/KS-TC-14-004-EN.PDF/365e7792-7724-4bf4-9b2e-c8df216376f3</p>
Adult Education Survey 2016	<p>Metodología en:</p> <p>https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/714753/Adult_Education_Survey_2016_technical_report.pdf</p>
EU Statistics on Income and Living Conditions 2013	<p>Metodología en:</p> <p>https://ec.europa.eu/eurostat/statistics-explained/index.php/EU_statistics_on_income_and_living_conditions_(EU-SILC)_methodology_%E2%80%93_data_quality#Weighting</p>

i. Estructura de los datos de Encuesta Barómetro 2014

En la Tabla 9 se presenta la estructura de los datos de la encuesta a nivel de comunas (94 seleccionadas), zonas censales, que agrupan geográficamente a un grupo de manzanas (551), y manzanas, que corresponden a unidades territoriales que dan forma a los centros poblados determinados por límites naturales (560). En promedio se asignaron 29,7 viviendas por comuna (con un mínimo de 5 y un máximo de 110), 5,1 viviendas para zonas censales (con un mínimo de 5 y un máximo de 10), y en manzanas las viviendas asignadas fueron 5 para todas las manzanas. Respecto a encuestadores, en promedio se asignaron 2,4 por comuna (con un mínimo de 1 y un máximo de 8), y 1,1 en zonas censales y manzanas (ambas con un mínimo de 1 y un máximo de 3).

Esta estructura inicial muestra que la distinción entre zona censal y manzana no es tan relevante en este caso, ya que para 98% de las zonas censales solo hay una manzana por zona.

En cuanto a encuestas completas, en promedio hubieron 24,1 por comuna (con un mínimo de 0 y un máximo de 99), 4,5 en zonas censales y manzanas (con máximos de 10 y 5 respectivamente). Por otro lado, hubo un promedio de 2,3 encuestas rechazadas en comunas (con un máximo de 20), y de 0,4 a nivel zonas censales y comunas (con un máximo de 5 para ambos casos). Respecto al no contacto, hubo 1,8 casos de no contacto en promedio por comuna (con un máximo de 14), y de 0,3 en zonas censales y manzanas (con un máximo de 6 y 5 casos de no contacto respectivamente).

Tabla 9. Estructura de los datos a nivel de comunas, zonas censales y manzanas

	Totales	Comunas (n: 94)			Zonas censales (n: 551)			Manzanas (n: 560)		
		\bar{X}	Min.	Máx.	\bar{X}	Min.	Máx.	\bar{X}	Min.	Máx.
Muestra seleccionada	2.800	29,7	5	110	5,1	5	10	5	5	5
Encuestas completas (110)	2.267	24,1	0	99	4,5	0	10	4,5	0	5
Encuestas rechazadas (211)	212	2,3	0	20	0,4	0	5	0,4	0	5
No contacto (224, 225, 236)	165	1,8	0	14	0,3	0	6	0,3	0	5
Otros	156	1,6	5	20	0,2	5	20	0,3	5	20
Asignación encuestadores	107	2,4	1	8	1,1	1	3	1,1	1	3

Nota: Códigos entre paréntesis denotan código AAPOR correspondiente, los cuales se pueden revisar en detalle en la Tabla 11 de la sección de Anexos.
 \bar{X} = Promedio, DE = Desviación estándar, Min= Mínimo, Max=Máximo.

Al analizar los resultados a nivel de encuestador, la Tabla 10 muestra que el promedio de encuestas completas por encuestador (2.267 en total) fueron 21,1 casos, con un mínimo de 0 y un máximo de 145, encontrándose la moda en 5 encuestas completas por encuestador. Por otro lado, hubo 1,9 rechazos por encuestador en promedio, con un mínimo de 0 y un máximo de 25 (de los 212 casos de rechazo en total). En cuanto al no contacto, en promedio se dieron 1,5 de estos casos por encuestador, con un máximo de 0 y un máximo de 16 (de los 165 casos de no contacto en total).

Tabla 10. Encuestas completas, rechazadas y no contacto por encuestador

Encuestas por encuestador	Total	Mínimo	Promedio	Máximo	Desviación estándar
Completas	2.267	0	21,1	145	26,1
Rechazadas	212	0	1,9	25	4,5
No contacto	165	0	1,5	16	2,8

Tabla 11. Códigos AAPOR de Disposición Final de Casos

Entr.					
Entr.	110	Entrevista completa	No Entr. Elegibilidad desc.	311	No se envió a terreno
	120	Entrevista parcial		317	Área peligrosa o de difícil acceso
No entrevistados (NE), Elegibles	211	Se rechazó entrevista	No entrevistados, No elegibles	318	No fue posible localizar la dirección
	2111	Informante rechazó entrevista (nivel hogar)		390	Otra razón de elegibilidad desconocida
	2112	Persona Kish rechazó entrevista (nivel persona)		410	Fuera de muestra
	212	Se interrumpió la entrevista		451	Empresa, oficina de gobierno y otra organización
	223	Se impidió acceso a la vivienda		452	Instituciones (hospital, cárcel, asilo de anciano, etc.)
	224	Vivienda ocupada sin moradores presentes		453	Dormitorio colectivo (militar de trabajo, internados, etc.)
	225	Informante no ubicable o no puede atender		454	Vivienda en demolición, incendiada, destruida o erradicada
	231	Muerte del informante		461	Vivienda particular desocupada
	232	Informante impedido físico/mental para contestar		462	Vivienda de veraneo o de uso temporal
	233	Problemas de idioma		463	Otra razón no elegible
	236	Otra razón elegible (correspondiendo en este caso a encuestas rechazadas en la etapa de supervisión)			

ii. Ponderador base de viviendas y personas

Para la construcción del ponderador base de viviendas se utilizó la información de la selección de la muestra a partir del script del programa R, que a su vez utilizó el paquete *sampling* para seleccionar a las unidades.

En primer lugar, se debe reconstruir la probabilidad de selección de la comuna¹³. Las comunas fueron seleccionadas en base a dos criterios; aquellas de inclusión forzosa (IF) (con más de 100.000 habitantes y/o son capitales regionales) que tienen una probabilidad de selección igual a 1, y las de no inclusión forzosa (NO-IF), donde su probabilidad de selección era proporcional a la cantidad de viviendas de la comuna. Esto queda representado por:

$$P_h(i) = \begin{cases} 1 & \text{si } i \in h; \text{ es IF} \\ \frac{U_{hi}}{U_h} & \text{si } i \in h; \text{ es no IF} \end{cases}$$

Donde,

- h : Representa el índice de los estratos de muestreo de primera etapa, en este caso, las 15 regiones de Chile.
- i : Representa el índice de la UPM, correspondiente a una parte urbana de una comuna.
- U_{hi} : Representa el total de viviendas que posee la UPM i , del estrato h , según el Marco de muestreo, sin considerar las manzanas de 7 o menos viviendas.
- U_h : Representa el total de viviendas que posee el estrato h (región), según el Marco de muestreo, sin considerar las manzanas de 7 o menos viviendas.

De esta manera, el ponderador base de comunas se constituye a partir del inverso de la probabilidad de selección de la comuna:

¹³ A pesar de que en documentos oficiales de la encuesta no se establece que se seleccionan las partes urbanas de las comunas, esto se puede extraer del código de selección de la muestra, donde se excluyen las partes rurales de las comunas del marco muestral.

$$w_{comuna} = \frac{1}{P_h(i)}$$

En segundo lugar, se incorpora la probabilidad de selección de las manzanas, las cuales fueron seleccionados con una probabilidad proporcional a la cantidad de viviendas de la manzana, lo cual queda expresado en:

$$P_{hi}(j|i) = \frac{n_{hi} M_{hij}}{U_{hi}}$$

Donde,

- j : Representa el índice de la USM, correspondiente a una manzana censal.
- n_{hi} : Representa el total de USM a seleccionar en la UPM i , del estrato h .
- U_{hi} : Representa el total de viviendas que posee la UPM i , del estrato h , según el Marco de muestreo, sin considerar las manzanas de 7 o menos viviendas.
- M_{hij} : Representa el total de viviendas que posee la USM j , en la UPM i , del estrato h , según el Marco de muestreo.

Así, el ponderador de selección de la manzana queda constituido como el inverso de su probabilidad de selección:

$$w_{manzana} = \frac{1}{P_{hi}(j)}$$

En tercer lugar, se involucra la probabilidad de selección de la vivienda en la manzana. Esto se realizó con una selección aleatoria con salto sistemático, donde cada vivienda tenía una igual probabilidad de selección.

$$P_{hij}(k|j) = \frac{m_{hij}}{M_{hij}^*}$$

Donde,

- k : Representa el índice de la UTM, correspondiente a una vivienda.
- m_{hij} : Total de viviendas a seleccionar en la USM j , de la UPM i , del estrato h . Esto se encuentra por diseño definido en “5”, donde en cada manzana se seleccionaron 5 viviendas a participar.
- M_{hij}^* : Total de viviendas particulares elegibles (viviendas residenciales ocupadas) que posee la USM j , en la UPM i , del estrato h , de acuerdo a lo registrado en el proceso de empadronamiento.

Así, el ponderador de viviendas corresponde al inverso de la probabilidad de selección de la vivienda en la manzana:

$$w_{vivienda} = \frac{1}{P_{hij}(k)}$$

Ahora bien, para la construcción del ponderador base a nivel de viviendas se debe dar cuenta del carácter condicional de las probabilidades de selección. Es decir, es necesario considerar que la probabilidad de selección de una vivienda, también viene determinada por la probabilidad condicional de que la manzana donde se encuentra haya sido seleccionada, y a su vez, que la comuna donde se encuentra la manzana también haya sido seleccionada. En base a ello, el **ponderador base de viviendas** queda expresado en:

$$w_{b_{viviendas}} = w_{comuna} * w_{manzana} * w_{vivienda}$$

Luego de obtener el ponderador base de viviendas la muestra seleccionada representa a la población de viviendas objetivo. Este ponderador será el punto de partida para realizar los ajustes explícitos de no respuesta.

Adicionalmente, para efectos de este estudio, no se consideraron ajustes de elegibilidad como se recomienda en Valliant et al, 2013 en función de simplificar el análisis. Se debe considerar que dicha simplificación no afecta los resultados, pues dichos ajustes se habrían realizado a todos los factores de expansión por igual, y por lo tanto, no habrían tenido

efectos sobre los resultados de las comparaciones de los factores alternativos que son foco del estudio.

iii. Ajustes explícitos de no respuesta

Para realizar el ajuste explícito de no respuesta se utilizó la técnica de “propensity stratification”, método introducido por Rosenbaum y Rubin, bajo el supuesto de que nos encontramos en un escenario *Missing At Random* (1983).

En ese método, se crean celdas de ajustes a partir de la predicción de responder o no en la encuesta en base a una serie de variables auxiliares, que en el caso ideal debiesen estar correlacionadas con la probabilidad de responder y la variable resultado del estudio.

En particular, se realiza un modelo de regresión logística que tiene como variable dependiente 1 = responde, y 0 = no responde para cada una de las viviendas seleccionadas (Valliant, Dever, & Kreuter, 2013). Luego de ello, los pasos a seguir corresponden a;

- a. Calcular la probabilidad de responder para cada una de las unidades de la muestra en función de modelo de regresión utilizado.
- b. Ordenar los casos de mayor a menor en función de la probabilidad predicha.
- c. Formar celdas que agrupen a un mismo número de elementos de la muestra, que incluirán a respondientes y no respondientes.

De esta manera, se construyeron 20 celdas de ajuste a partir de la predicción de responder, y en cada una de ellas, se calculó una razón de ajuste definido por:

$$\hat{R}_{g,Resp} = \frac{\sum_{k \in \varphi_{g,eleg}} w b_{viviendas}}{\sum_{k \in \varphi_{g,eleg,Resp}} w b_{viviendas}}$$

Donde,

$wb_{viviendas}$: Ponderador de viviendas, asociado a la vivienda k , de la manzana j , seleccionada al interior de la comuna i en el estrato h .

g : Índice de las 20 clases (ventiles) creadas para el ajuste de no respuesta.

$\varphi_{g,eleg}$: Conjunto de viviendas elegibles¹⁴ pertenecientes al ventíl g .

$\varphi_{g,eleg,Resp}$: Conjunto de viviendas elegibles que respondieron la encuesta, pertenecientes al ventíl g .

De esta manera, se puede obtener el **ponderador base de viviendas ajustado por no respuesta**.

$$wb_{viviendas}^{NR} = \hat{R}_{g,Resp} * wb_{viviendas}$$

En específico, para seleccionar las variables a incluir en el modelo de ajuste de no respuesta se estudió la correlación bivariada entre los indicadores de respuesta (1 = responde, 0 = no responde), cooperación (1 = coopera, 0 = no coopera) y contacto (1 = hubo contacto, 0 = no hubo contacto) y diversas variables auxiliares disponibles para respondentes y no respondentes que pudieran relacionarse con la participación en la encuesta y con el bienestar subjetivo. Dichas variables muestran características a dos niveles; (1) nivel de comuna¹⁵ (obtenidas a través de datos administrativos del Ministerio de Desarrollo Social o construidas a partir de la base de datos del Censo 2012) y zonas censales¹⁶ (construidas a

¹⁴ Para efectos de este estudio, se asumirá que todas las viviendas seleccionadas en el estudio eran elegibles. Esto se fundamenta en que de los 2.800 casos seleccionados en la muestra, solo 40 correspondieron a viviendas no elegibles, por lo cual la tasa de elegibilidad corresponde al 0,99.

¹⁵ De las 94 comunas de la muestra del estudio, en 17 de ellas no se encontraban disponibles algunos indicadores a nivel de comuna. Para subsanar ello, se utilizó la metodología de para la estimación de la Pobreza SAE del Ministerio de Desarrollo Social, donde se identifican clusters de comunas que son similares respecto a una serie de indicadores. De esta forma, en aquellas comunas donde no se tuviera el dato de una variable, se les imputó el promedio de dicha variable entre las comunas de su mismo cluster que estuvieran disponibles en la muestra de la encuesta.

¹⁶ En 4 de las 551 zonas censales construidas a partir de la muestra del Barómetro 2014 no se pudieron encontrar las variables construidas a partir de la base de datos del Censo, pues correspondían a áreas RAU sin geocódigo que permitiera identificarlos en el Censo. En esos casos, se les imputó el valor promedio de dicha variable en la comuna donde pertenecía la zona censal, pues correspondía al nivel geográfico más cercano con datos disponibles.

partir de las variables del Censo 2012). Las variables utilizadas a ambos niveles se muestran en la Tabla 14 y Tabla 15 del Anexo. De dichas variables, se escogieron aquellas que tuvieran una correlación significativa al menos con un 95% nivel de confianza con la tasa de respuesta, o de contacto, o de cooperación, para luego ser incluidas en los modelos de regresión logística utilizados para crear las clases de ajustes de no respuesta (ver resultados de relaciones bivariadas en la Tabla 16 y Tabla 17 del Anexo).

Adicionalmente, en las Tabla 18 y Tabla 19 del Anexo se pueden ver las relaciones bivariadas de dichas variables con la pregunta sobre bienestar subjetivo.

Con el objeto de responder a la pregunta de investigación sobre el tipo de variables auxiliares que son más útiles para reducir el sesgo y la varianza en los ajustes de no respuesta, se realizaron cinco modelos de regresión logística diversos para la predicción de la probabilidad de responder. En el primero, se incorporaron solo variables a nivel de comuna de fuentes administrativas del Ministerio de Desarrollo Social y del Censo. En el segundo, solo variables a nivel de zona censal provenientes del Censo de población 2012. En el tercero, variables de fuentes administrativas del Ministerio de Desarrollo Social a nivel de comuna. En el cuarto, variables provenientes del Censo de población 2012. Finalmente, el quinto modelo se construyó incorporando todas las variables disponibles.

Los resultados de los modelos de regresión logística mostraron que en general variables vinculadas a los niveles educacionales de adultos y jefes de hogar de comunas y zonas censales, estadísticas de pobreza de la comuna, proporciones de hogares con discapacitados, niños y adultos mayores se vincularon con la respuesta total de la encuesta a nivel de viviendas. Para más detalles de los resultados de los modelos de regresión consultar la Tabla 20 del Anexo.

A partir de los resultados anteriores, se retuvieron las variables que en los modelos de regresión logística resultaron significativas para predecir la respuesta a nivel de vivienda. De este modo, y a partir de las variables significativas de los cinco modelos de regresión, se realizaron cinco ajustes de no respuesta alternativos, todos mediante la metodología de

“propensity stratification” para crear celdas de ajuste de la no respuesta¹⁷ (Valliant, Dever, & Kreuter, 2013).

iv. Ponderador de personas

Ahora, teniendo el ponderador de viviendas ajustado por no respuesta, es necesario pasar a representar a la población objetivo a nivel de individuos. Para eso, debe incorporar el ponderador de selección de personas, considerando la probabilidad de selección del encuestado en la vivienda (la cual se encontrará disponible solo para aquellas viviendas que decidieron participar de la encuesta).

En el Barómetro de la Felicidad 2014, la estrategia de selección del encuestado se basó en el método Kish, donde a cada persona elegible de la vivienda (personas entre 18-90 años) se le otorga una igual probabilidad de selección. Así, la probabilidad de selección queda en:

$$P_{hijk}(l|k) = \frac{1}{p_k}$$

Donde,

p_k : Total de personas elegibles en la vivienda k .

Así, el ponderador de selección de personas quedará definido como el inverso de probabilidad de selección de la persona:

$$w_{persona} = \frac{1}{P_{hijk}(l|k)}$$

Finalmente, para conseguir el **ponderador base de personas** (w_b), que permitirá expandir a la población de individuos, se tiene que:

$$wb_{personas} = wb_{viviendas} * w_{persona}$$

¹⁷ En la Tabla 21 del Anexo se muestran los promedios de respuesta observados y promedios de probabilidad predicha de respuesta por ventila para cada uno de los ajustes de no respuesta realizados.

Como una manera de evaluar el funcionamiento de los factores de expansión se utiliza la suma de los ponderadores base, para verificar que su suma se aproxime a la población de viviendas y personas que quiere representar. Este resultado se puede ver en la Tabla 12 y Tabla 13 del Anexo.

$$wb_{personas}^{NR} = wb_{viviendas}^{NR} * wb_{personas}$$

v. Calibración

Para la calibración de la encuesta Barómetro de la Felicidad 2014 se utilizó el paquete *survey* en el programa estadístico R, utilizando el método de raking¹⁸. Para ello, se consideraron los totales poblacionales a partir de las estimaciones de población por comuna, sexo y edad simple de INE para años 2002 y 2020 (Instituto Nacional de Estadísticas, 2013).

Sin embargo, al no encontrarse los totales poblacionales desagregados por ruralidad, para estimar la población objetivo de la encuesta (personas entre 18-90 años de las partes urbanas de las comunas con más de 25.000 habitantes), se estimó el porcentaje de población urbana en cada una de las 94 comunas en la muestra utilizando datos de la Encuesta Casen 2015. Los totales de población empleados para la calibración vienen dados por la fórmula:

$$TotU_p = Tot_p * PctU_p$$

¹⁸ Ver más detalles del paquete en <https://cran.r-project.org/web/packages/survey/survey.pdf>.

Donde,

Tot_p : Total de personas en la subpoblación p

$PctU_p$: Porcentaje de personas residentes en áreas urbanas en la subpoblación p estimado a partir de Casen 2015

$TotU_p$: Total de personas residentes en áreas urbanas en la subpoblación p .

p : Índice asociado a las 94 comunas en la muestra de Barómetro 2014.

De esta manera, se obtuvieron los totales poblacionales ajustados para sexo, tramos etarios (4), macrozona (4) y nivel educacional (3), los cuales corresponden a los mismos utilizados para la calibración original de la encuesta (ver Tabla 13 del Anexo), con lo que se construyó el Ajuste de calibración Aj_{calib} .

Tabla 12. Suma de ponderadores base de selección de viviendas sin y con ajustes explícito de No Respuesta

		Sin ajuste de No Respuesta	Con ajuste explícito de No Respuesta				
Región	Marco muestral de Viviendas urbanas de 141 comunas con más de 25000 habitantes según Marco manzanas urbanas 2008 INE	PSEL_CAL (N:2800)	NR_COM_MDS (N:2267)	NR_COM (N:2267)	NR_ZON (N:2267)	NR_FULL_CENSO (N:2267)	NR_FULL (N:2267)
I	67.871	59.398	63.616	58.978	55.645.	58.073	56.090
II	116.743	109.411	105.297	112.513	116.299	124.478	103.401
III	53.102	37.251	39.713	42.666	42.212	46.962	44.013
IV	142.847	139.401	117.695	127.615	118.500	117.178	131.657
V	441.090	421.418	407.113	407.265	410.601	412.051	386.370
VI	133.141	130.987	140.752	126.551	136.926	136.771	124.371
VII	179.608	174.918	185.031	197.437	189.271	180.000	201.407
VIII	401.922	387.257	433.773	405.321	444.815	451.021	405.949
IX	137.891	135.323	133.298	137.685	140.873	136.317	135.541
X	120.854	114.734	119.298	124.879	99.554	96.022	109.749
XI	19.782	126.56	12.517	12.054	11.995	12.160	12.226
XII	36.163	33.099	34.057	35.039	39.559	36.981	39.458
RM	1.779.731	1.698.461	1.659.254	1.675.763	1.646.648	1.650.553	1.711.080
XIV	54.088	51.872	49.198	46.527	51.406	45.152	48.442
XV	52.865	47.463	51.642	42.215	48.244	48.525	42.833
Total	3.737.698	3.553.649	3.552.253	3.552.508	3.552.546	3.552.245	3.552.587

Tabla 13. Suma de ponderadores base de selección de personas sin y con ajustes explícito de No Respuesta

		Sin ajuste de No Respuesta	Con ajuste explícito de No Respuesta				
Región	Marco muestral de Personas con 18+ años en las partes urbanas de las 141 comunas con más de 25000 habitantes según población estimada por comuna 2002-2020 INE y ajustada por ruralidad a partir de proyecciones Casen 2015	PSEL_CAL (N:2800)	NR_COM_MDS (N:2267)	NR_COM (N:2267)	NR_ZON (N:2267)	NR_FULL_CENSO (N:2267)	NR_FULL (N:2267)
I	204.505	159.394	207.288	196.344	184.230	192.581	183.498
II	387.790	355.146	397.632	424.776	441.532	476.197	390.009
III	149.609	155.100	165.351	177.644	176.141	197.231	184.692
IV	414.704	326.083	394.491	421.913	393.762	392.586	441.030
V	1.096.146	1.177.514	1.395.862	1.390.913	1.398.891	1.408.588	1.314.387
VI	371.510	323.408	408.567	369.068	399.438	396.967	359.350
VII	428.261	435.278	527.940	561.983	546.029	520.013	576.433
VIII	1.133.779	1.029.338	1.221.268	1.137.037	1.245.533	1.262.428	1.131.971
IX	364.017	365.596	450.276	465.338	475.456	462.171	459.103
X	338.237	244.169	358.703	377.390	302.481	291.519	332.938
XI	56.913	35.437	43.808	42.189	42.016	42.534	43.056
XII	94.747	94.332	102.171	105.118	118.914	110.299	118.747
RM	5.156.558	4.625.879	5.817.676	5.837.024	5.783.308	5.798.844	5.975.725
XIV	153.622	101.721	131.586	124.439	136.109	121.275	128.781
XV	149.421	87.543	115.845	94.698	108.664	109.835	96.250
Total	10.499.821	9.515.937	11.738.464	11.725.873	11.752.504	11.783.068	11.735.970

Nota: la diferencia de personas entre los factores de expansión con ajuste explícito de no respuesta y la población estimada de personas puede explicarse a que los hogares contactados en la encuesta podrían tener una mayor cantidad de residentes, y por tanto, son más posibles de contactar. Esta diferencia se corrige luego de los ajustes de calibración poblacionales.

Tabla 14. Variables utilizadas para análisis bivariados a nivel de comuna

Dimensión	Variable	Fuente
Educación	% de adultos con educación superior o más	Estimación propia de Censo 2012
	% de jefes de hogar con educación superior o más	Estimación propia de Censo 2012
	% de adultos con educación superior incompleta	Estimación propia de Censo 2012
	% de jefes de hogar con educación superior incompleta	Estimación propia de Censo 2012
	% de adultos con educación media	Estimación propia de Censo 2012
	% de jefes de hogar con educación media	Estimación propia de Censo 2012
	% de hogares con personas mayores de 18 años sin nivel educacional establecido por ley	MDS (Casen 2015)
Pobreza	Índice de Pobreza multidimensional	MDS (Casen 2013)
	Índice de Pobreza SAE	MDS (Casen 2013)
	% de hogares que tienen todos los bienes (teléfono fijo, vehículo, internet, servicio doméstico)	Estimación propia de Censo 2012
	% de hogares con déficit a servicios básicos	MDS (Casen 2015)
	% de hogares con el tramo más alto de ingresos según ficha de Protección Social	MDS (Casen 2015)
	% de hogares con el tramo más bajo de ingresos según ficha de Protección Social	MDS (Casen 2015)
Seguridad	Tasa de delitos de mayor connotación social	Subsecretaría de Prevención del Delito 2012
Ocupación	% hogares con al menos un adulto dedicado a quehaceres del hogar	Estimación propia de Censo 2012
	% de hogares con al menos un adulto desempleado	Estimación propia de Censo 2012
	% de hogares con al menos un adulto jubilado	Estimación propia de Censo 2012
Sociales	% de hogares con al menos un miembro con discapacidad	Estimación propia de Censo 2012
	% de hogares que habitan en entornos expuestos a ambiente contaminado o no disponen de equipamientos básicos cercanos a su vivienda	MDS (Casen 2015)
	% de personas de 5 años o más que utilizan internet, al menos una vez a la semana o más	MDS (Casen 2015)

	% de hogares donde al menos un mayor de 15 años ha sufrido discriminación en los últimos 12 meses	MDS (Casen 2015)
Demográficos	% adultos mayores	Estimación propia de Censo 2012
	% de mujeres	Estimación propia de Censo 2012
	% de hogares con niños (menores de 6 años)	Estimación propia de Censo 2012
Vivienda y entorno	% de hogares con saneamiento deficitario	MDS (Casen 2015)
	% de viviendas aceptables	MDS (Casen 2015)
	% de viviendas recuperables	MDS (Casen 2015)
	% de viviendas irrecuperables	MDS (Casen 2015)
	% de viviendas con hacinamiento medio	MDS (Casen 2015)
	% de viviendas con hacinamiento crítico	MDS (Casen 2015)
Salud	% de personas Fonasa Grupo A (indigentes)	Registros Fonasa 2010

Tabla 15. Variables utilizadas para análisis bivariados a nivel de zonas censales

Dimensión	Variable	Fuente
Educación	% de adultos con educación superior o más	Estimación propia de Censo 2012
	% de jefes de hogar con educación superior o más	Estimación propia de Censo 2012
	% de adultos con educación superior incompleta	Estimación propia de Censo 2012
	% de jefes de hogar con educación superior incompleta	Estimación propia de Censo 2012
	% de adultos con educación media	Estimación propia de Censo 2012
	% de jefes de hogar con educación media	Estimación propia de Censo 2012
Pobreza	% de hogares que tienen todos los bienes (teléfono fijo, vehículo, internet, servicio doméstico)	Estimación propia de Censo 2012
Ocupación	% hogares con al menos un adulto dedicado a quehaceres del hogar	Estimación propia de Censo 2012
	% de hogares con al menos un adulto desempleado	Estimación propia de Censo 2012
	% de hogares con al menos un adulto jubilado	Estimación propia de Censo 2012
Sociales	% de hogares con al menos un miembro con discapacidad	Estimación propia de Censo 2012
Demográficos	% adultos mayores	Estimación propia de Censo 2012
	% de mujeres	Estimación propia de Censo 2012
	% de hogares con niños (menores de 6 años)	Estimación propia de Censo 2012
Vivienda y entorno	Suma de escombros en la zona	Estimación propia de Precenso 2011
	Suma de bienes recreativos en la zona	Estimación propia de Precenso 2011

Tabla 16. Correlaciones bivariadas significativas de variables a nivel de comuna con Tasa de Respuesta (RR1), Tasa de Contacto (CON1) y Tasa de Cooperación (COOP1)

Dimensión	Variable	RR1	CON1	COOP1
Educación	% de adultos con educación superior o más	-0,21		-0,39
	% de jefes de hogar con educación superior o más	-0,26		-0,42
	% de adultos con educación media	0,33		0,37
	% de jefes de hogar con educación media	-0,35		0,37
	% de hogares con personas mayores de 18 años sin nivel educacional establecido por ley			0,29
Pobreza	Índice de Pobreza multidimensional			0,23
	Índice de Pobreza SAE	0,29		0,29
	% de hogares que tienen todos los bienes (teléfono fijo, vehículo, internet, servicio doméstico)	-0,38	-0,21	-0,47
	% de hogares con el tramo más alto de ingresos según ficha de Protección Social			-0,32
Seguridad	Tasa de delitos de mayor connotación social			-0,21
Ocupación	% hogares con al menos un adulto dedicado a quehaceres del hogar			0,31
	% de hogares con al menos un adulto desempleado	0,24		0,35
	% de hogares con al menos un adulto jubilado	0,21		
Sociales	% de hogares con al menos un miembro con discapacidad	0,29		0,36
	% de hogares que habitan en entornos expuestos a ambiente contaminado o no disponen de equipamientos básicos cercanos a su vivienda			0,36
	% de personas de 5 años o más que utilizan internet, al menos una vez a la semana o más			-0,36
Demográficos	% adultos mayores			-0,22
	% de hogares con niños (menores de 6 años)			0,25
Vivienda y entorno	% de hogares con saneamiento deficitario			0,27
	% de viviendas aceptables			-0,23
	% de viviendas con hacinamiento medio			0,32
Salud	% de personas Fonasa Grupo A (indigentes)			0,26
Nota: se presentan los coeficientes estadísticamente significativos a un 95% de nivel de confianza.				
N: 94 comunas.				

Tabla 17. Correlaciones bivariadas significativas de variables a nivel de zonas censales con Tasa de Respuesta (RR1), Tasa de Contacto (CON1) y Tasa de Cooperación (COOP1)

Dimensión	Variable	RR1	CON1	COOP1
Educación	% de adultos con educación superior o más	-0,18	-0,11	-0,13
	% de jefes de hogar con educación superior o más	-0,21	-0,13	-0,16
	% de adultos con educación superior incompleta	0,15	0,11	0,1
	% de adultos con educación media	0,21	0,12	0,17
	% de jefes de hogar con educación media	0,19	0,10	0,17
Pobreza	% de hogares que tienen todos los bienes (teléfono fijo, vehículo, internet, servicio doméstico)	-0,31	-0,19	-0,27
Ocupación	% hogares con al menos un adulto dedicado a quehaceres del hogar	0,14		0,12
	% de hogares con al menos un adulto desempleado	0,17	0,09	0,14
Sociales	% de hogares con al menos un miembro con discapacidad	0,14	0,09	0,310
Demográficos	% adultos mayores	-0,10		-0,12
	% de hogares con niños (menores de 6 años)	0,09		0,12
Vivienda y entorno	Suma de bienes recreativos en la zona	-0,16	-0,17	
Nota: se presentan los coeficientes estadísticamente significativos a un 95% de nivel de confianza.				
N: 551 zonas censales.				

Tabla 18. Correlaciones bivariadas de variables a nivel de comuna con participación en la encuesta y pregunta sobre bienestar subjetivo

Dimensión	Variable	Correlación con participación en la encuesta	Correlación con pregunta sobre bienestar subjetivo
Educación	% de adultos con educación superior o más	-0,16*	0,07*
	% de adultos con educación media	0,18*	-0,10*
	% de jefes de hogar con educación media	-0,21*	-0,11*
Pobreza	Índice de Pobreza multidimensional	0,05*	-0,10*
	Índice de Pobreza SAE	0,13*	0,01
	% de hogares que tienen todos los bienes (teléfono fijo, vehículo, internet, servicio doméstico)	-0,26*	0,10*
	% de hogares con el tramo más alto de ingresos según ficha de Protección Social	-0,11*	0,04*
Seguridad	Tasa de delitos de mayor connotación social	0,0	0,02
Ocupación	% de hogares con al menos un adulto jubilado	0,05*	0,06*
Sociales	% de hogares con al menos un miembro con discapacidad	0,15*	-0,01
	% de hogares que habitan en entornos expuestos a ambiente contaminado o no disponen de equipamientos básicos cercanos a su vivienda	0,1	-0,04
Demográficos	% adultos mayores	-0,09*	0,08*
	% de hogares con niños (menores de 6 años)	0,05*	-0,11*
Vivienda y entorno	% de viviendas aceptables	-0,09*	0,12*
	% de viviendas con hacinamiento medio	0,14*	-0,08*
Salud	% de personas Fonasa Grupo A (indigentes)	0,08*	-0,02
Nota: * denotan coeficientes estadísticamente significativos a un 95% de nivel de confianza.			
N: 2.267 casos que participaron en Barómetro 2014.			

Tabla 19. Correlaciones bivariadas de variables a nivel de zona censal con participación en la encuesta y pregunta sobre bienestar subjetivo

Dimensión	Variable	Correlación con participación en la encuesta	Correlación con pregunta sobre bienestar subjetivo
Educación	% de adultos con educación superior o más	-0,15*	0,11*
	% de jefes de hogar con educación media	0,16*	-0,08*
Pobreza	% de hogares que tienen todos los bienes (teléfono fijo, vehículo, internet, servicio doméstico)	-0,26*	0,13*
Demográficos	% adultos mayores	-0,08*	0,03
Vivienda y entorno	Suma de bienes recreativos en la zona	-0,13*	0,05*
Nota: * denotan coeficientes estadísticamente significativos a un 95% de nivel de confianza.			
N: 2.267 casos que participaron en Barómetro 2014.			

Tabla 20. Odds ratio de modelos de regresión logística para la Respuesta en la encuesta (1) y la No respuesta en la encuesta (0), según variables a nivel de comuna y zonas censales provenientes de datos administrativos del Ministerio de Desarrollo Social (MDS) y Censo 2012.

Dimensión	Variable	Modelo comunas	Modelo zonas censales	Modelo datos administrativos (MDS)	Modelo datos de Censo 2012	Modelo global
Educación	% de adultos con educación superior o más	1,13* (0,05)	1,03* (0,02)		Comuna 1,14* (0,05)	Comuna 1,05*** (0,01)
					Zona censal 1,01 (0,02)	Zona censal 1,00 (0,003)
	% de jefes de hogar con educación superior o más	0,92 (0,07)	0,98 (0,02)		Comuna 0,88 (0,07)	
					Zona censal 0,99 (0,02)	
	% de adultos con educación superior incompleta		0,98 (0,03)		Zona censal 0,94 (0,03)	
	% de adultos con educación media	1,23*** (0,05)	1,01 (0,02)		Comuna 0,97 (0,04)	Comuna 1,10* (0,03)
					Zona censal 1,05 (0,03)	
% de jefes de hogar con educación media	0,81** (0,06)	0,98 (0,02)		Comuna 1,06 (0,05)	Comuna 0,93 (0,04)	
				Zona censal 0,05* (0,02)		
% de hogares con personas mayores de 18 años sin nivel educacional establecido por ley	0,99 (3,44)			0,99 (0,01)		
Pobreza	Índice de Pobreza multidimensional	0,85*** (0,02)		0,89*** (0,01)		Comuna 0,85*** (0,02)
	Índice de Pobreza SAE	1,01 (0,01)		1,04* (0,01)		
	% de hogares que tienen todos los bienes (teléfono fijo, vehículo, internet, servicio	0,96* (0,01)	0,96*** (0,01)		Comuna 0,98	Comuna 0,97

	doméstico)				(0,02)	(0,01)
					Zona censal 0,98 (0,01)	Zona censal 0,99 (0,008)
	% de hogares con el tramo más alto de ingresos según ficha de Protección Social	1,04** (0,01)		1,03*** (0,009)		Comuna 1,03** (0,009)
Seguridad	Tasa de delitos de mayor connotación social en comuna	1,00*** (0,00007)		1,00 (0,00005)		Comuna 1,00*** (0,0005)
Ocupación	% hogares con al menos un adulto dedicado a quehaceres del hogar	0,97 (0,02)	1,01 (0,01)		Comuna 0,96 (0,02)	
					Zona censal 1,03 (0,01)	
	% de hogares con al menos un adulto desempleado	1,02 (0,07)	1,05 (0,03)		Comuna 1,09 (0,07)	
					Zona censal 0,97 (0,04)	
% de hogares con al menos un adulto jubilado	1,13* (0,05)			Comuna 0,99 (0,05)	Comuna 1,16*** (0,04)	
Sociales	% de hogares con al menos un miembro con discapacidad	1,10*** (0,02)	1,01 (0,01)		Comuna 1,08** (0,02)	Comuna 1,10*** (0,02)
					Zona censal 0,98 (0,01)	
	% de hogares que habitan en entornos expuestos a ambiente contaminado o no disponen de equipamientos básicos cercanos a su vivienda	1,15*** (0,02)		1,11*** (0,01)		Comuna 1,14*** (0,01)
	% de personas de 5 años o más que utilizan internet, al menos una vez a la semana o más	0,98 (0,02)		0,99 (0,01)		
Demográficos	% adultos mayores	0,68** (0,1)	0,94*** (0,02)		Comuna 0,88 (0,10)	Comuna 0,72*** (0,09)
					Zona censal 0,96	Zona censal 0,96***

					(0,02)	(0,01)
	% de hogares con niños (menores de 6 años)	0,97 (0,04)	0,99 (0,02)		Comuna 0,91* (0,04)	
					Zona censal 1,00 (0,02)	
Vivienda y entorno	% de hogares con saneamiento deficitario	0,98 (0,01)		0,99 (0,009)		
	% de viviendas aceptables	0,91*** (0,02)		0,94*** (0,01)		Comuna 0,91*** (0,01)
	% de viviendas con hacinamiento medio	0,96 (0,03)		1,07** (0,02)		
	Suma de bienes recreativos		0,99*** (0,001)		Zona censal 0,99*** (0,001)	Zona censal 0,99*** (0,001)
Salud	% de personas Fonasa Grupo A (indigentes)	1,06** (0,01)		1,03 (0,01)		Comuna 1,04* (0,01)
Intercepto		2,54	1,24	5,79	1,95	2,39
N		2.800	2.780	2.800	2.780	2.800
AIC		2441	2492	2502	2434	2348

Nota: errores estándar entre paréntesis.

Niveles de significancia: *** (menor a 0.001), ** (menor a 0.01), * (menor a 0.05).

En modelo de Censo 2012 y Global, se anuncia el nivel de la variable que resulta significativo, comuna o zona censal.

Para variables construidas a partir del Censo 2012 que se encuentran disponibles a niveles de comuna y zona censal, las estimaciones a nivel de comuna se realizaron excluyendo a las zonas censales al interior de ellas.

Tabla 21. Promedio de respuesta observada y Promedio de predicción de probabilidad de responder por ventiles construidos para los ajustes de no respuesta de los factores de expansión alternativos construidos.

Ventil	NR_COM_MDS		NR_COM		NR_ZON		NR_FULL_CENSO		NR_FULL	
	Promedio de Respuesta observada	Promedio de Predicción Probabilidad de responder	Promedio de Respuesta observada	Promedio de Predicción Probabilidad de responder	Promedio de Respuesta observada	Promedio de Predicción Probabilidad de responder	Promedio de Respuesta observada	Promedio de Predicción Probabilidad de responder	Promedio de Respuesta observada	Promedio de Predicción Probabilidad de responder
1	0.39	0.4	0.3	0.37	0.46	0.42	0.45	0.53	0.36	0.4
2	0.65	0.65	0.62	0.61	0.65	0.69	0.7	0.66	0.61	0.58
3	0.68	0.71	0.65	0.68	0.7	0.75	0.69	0.73	0.65	0.65
4	0.64	0.74	0.74	0.72	0.84	0.78	0.79	0.76	0.86	0.7
5	0.88	0.78	0.74	0.74	0.76	0.78	0.91	0.78	0.61	0.73
6	0.71	0.78	0.85	0.77	0.81	0.81	0.86	0.79	0.75	0.76
7	0.8	0.8	0.82	0.79	0.79	0.82	0.86	0.81	0.86	0.79
8	0.86	0.81	0.85	0.81	0.71	0.82	0.78	0.81	0.76	0.81
9	0.83	0.83	0.76	0.82	0.87	0.83	0.89	0.82	0.81	0.83
10	0.81	0.84	0.88	0.82	0.8	0.84	0.75	0.83	0.83	0.85
11	0.84	0.85	0.89	0.85	0.8	0.84	0.81	0.84	0.74	0.86
12	0.85	0.86	0.9	0.87	0.87	0.85	0.8	0.84	0.94	0.87
13	0.92	0.86	0.82	0.88	0.93	0.85	0.9	0.85	0.86	0.88
14	0.92	0.87	0.79	0.89	0.88	0.86	0.82	0.86	0.95	0.89
15	0.95	0.88	0.88	0.89	0.87	0.86	0.85	0.86	0.94	0.9
16	0.9	0.89	0.93	0.91	0.91	0.87	0.79	0.87	0.92	0.91
17	0.85	0.9	0.92	0.93	0.94	0.87	0.84	0.88	0.88	0.93
18	0.88	0.92	0.91	0.94	0.92	0.88	0.91	0.88	0.89	0.94
19	0.95	0.93	0.97	0.95	0.85	0.88	0.95	0.89	0.99	0.95
20	0.92	0.95	0.98	0.98	0.86	0.89	0.85	0.91	0.98	0.98

Tabla 22. Efecto Diseño de factores de expansión contruidos para promedio de variable de interés estimado para total y subpoblaciones de sexo, edad, macrozona, nivel educacional y situación ocupacional del encuestado

		PondB = 1 Ajuste NR = no Ajuste CAL = sí	PondB = N/n Ajuste NR = no Ajuste CAL = sí	PondB = sí Ajuste NR = no Ajuste CAL = sí	PondB = sí Ajuste NR = sí Ajuste CAL = sí				
Subpoblación		W1:UNO_CAL	W2:MAS_CAL	W3:PSEL_CAL	W4:NR_COM_ MDS	W5:NR_ COM	W6:NR_ ZON	W7:NR_FULL_ CENSO	W8:NR_FULL
1	Total	2,0	2,0	2,7	2,7	2,7	2,8	2,7	2,6
2	Hombres	2,1	2,1	2,9	3,0	3,1	3,1	3,0	2,9
	Mujeres	1,3	1,3	1,6	1,5	1,5	1,5	1,5	1,5
3	18 a 29 años	1,5	1,5	1,8	1,8	1,7	1,9	1,9	1,8
	30 a 45 años	1,5	1,5	2,1	2,1	2,1	2,1	2,0	2,1
	46 a 65 años	1,4	1,4	1,8	1,8	1,9	1,8	1,8	1,8
	66 a 90 años	1,1	1,1	1,6	1,7	1,8	1,7	1,7	1,6
4	Zona norte	1,2	1,2	1,3	1,3	1,4	1,5	1,4	1,3
	Zona centro	1,4	1,4	1,8	1,7	1,9	1,9	1,7	1,8
	Zona sur	1,7	1,7	3,1	2,8	2,8	3,0	3,0	2,8
	RM	2,5	2,5	3,1	3,2	3,3	3,3	3,3	3,2
5	Educ. Media incompleta	1,9	1,9	2,5	2,3	2,4	2,5	2,4	2,3
	Educ. Media completa	1,1	1,1	1,4	1,4	1,4	1,4	1,4	1,4
	Educ. Superior incompleta	1,1	1,1	1,3	1,3	1,2	1,4	1,4	1,3
	Educ. Superior completa	1,2	1,2	1,5	1,5	1,6	1,6	1,6	1,5
6	Ocupado	1,8	1,8	2,8	2,7	2,8	2,8	2,7	2,7
	Desocupado	1,0	1,0	2,0	2,2	2,5	2,1	2,0	2,0
	Inactivo	1,5	1,5	2,0	2,0	2,0	2,1	2,0	1,9

Tabla 23. Sesgo relativo de factores de expansión contruidos para promedio de satisfacción con la vida para total y subpoblaciones de sexo, edad, macrozona, nivel educacional y situación ocupacional del entrevistado

Sub-población	\bar{X}_{Casen}^{19}	PondB = 1	PondB = N/n	PondB = sí	PondB = sí				
		Ajuste NR = no Ajuste CAL = sí	Ajuste NR = no Ajuste CAL = sí	Ajuste NR = no Ajuste CAL = sí	Ajuste NR = sí Ajuste CAL = sí	Ajuste NR = sí Ajuste CAL = sí	Ajuste NR = sí Ajuste CAL = sí	Ajuste NR = sí Ajuste CAL = sí	Ajuste NR = sí Ajuste CAL = sí
		W1:UNO_CAL	W2:MAS_CAL	W3:PSEL_CAL	W4:NR_COM_MDS	W5:NR_COM	W6:NR_ZON	W7:NR_FULL_CENSO	W8:NR_FULL
1 Total	7,5	1,2%	1,2%	0,9%	0,8%	0,9%	0,8%	0,8%	0,7%
2	Hombres	7,6	1,7%	1,7%	1,3%	1,1%	1,2%	1,1%	0,9%
	Mujeres	7,4	0,6%	0,6%	0,4%	0,4%	0,4%	0,2%	0,2%
3	18 a 29 años	7,6	2,2%	2,2%	2,6%	2,8%	2,7%	2,8%	2,7%
	30 a 45 años	7,7	0,3%	0,3%	0,4%	0,1%	0,2%	0,1%	0,2%
	46 a 65 años	7,4	2,5%	2,5%	1,6%	1,5%	1,6%	1,3%	1,3%
	66 a 90 años	7,5	9,2%	9,2%	9,6%	9,2%	9,0%	9,6%	8,8%
4	Zona norte	7,7	9,7%	9,7%	8,8%	8,8%	8,9%	8,2%	8,6%
	Zona centro	7,6	1,5%	1,5%	0,7%	0,7%	0,6%	0,6%	0,5%
	Zona sur	7,4	4,5%	4,5%	5,2%	4,8%	4,7%	5,3%	5,0%
	RM	7,5	1,4%	1,4%	2,1%	1,7%	1,7%	1,9%	1,4%
5	Educ.Medía incompleta	7,2	3,2%	3,2%	2,9%	2,9%	3,2%	3,1%	2,5%
	Educ.Medía completa	7,5	1,9%	1,9%	2,2%	2,2%	2,2%	2,4%	2,3%
	Educ.Superior incompleta	7,8	0,1%	0,1%	1,0%	1,0%	1,0%	1,2%	0,8%
	Educ. Superior completa	8,1	2,6%	2,6%	2,7%	2,2%	2,0%	1,9%	2,1%
6	Ocupado	7,6	0,1%	0,1%	0,8%	1,0%	1,0%	1,2%	1,0%

¹⁹ Estimación en base a población de 18-90 años, de 94 comunas, y utilizando ponderador Exp_r20 construido especialmente para indicador de satisfacción con la vida, y no el ponderador exp_R que utilizan la mayoría de las estimaciones de otras variables en Casen.

	Desocupado	7,1	0,7%	0,7%	1,5%	2,3%	1,8%	1,6%	1,8%	1,6%
	Inactivo	7,3	1,2%	1,2%	1,3%	1,3%	1,4%	1,3%	1,1%	1,0%

Nota: se sombrea celdas que destacan indicadores de Sesgo Relativo más bajos en la estimación para población específica.

Tabla 24. Error Cuadrático Medio de factores de expansión contruidos para promedio de variable de interés estimado para total y subpoblaciones de sexo, edad, macrozona, nivel educacional y situación ocupacional del encuestado

		PondB = 1 Ajuste NR = no Ajuste CAL = sí	PondB = N/n Ajuste NR = no Ajuste CAL = sí	PondB = sí Ajuste NR = no Ajuste CAL = sí	PondB = sí Ajuste NR = sí Ajuste CAL = sí				
Subpoblación		W1:UNO_CAL	W2:MAS_CAL	W3:PSEL_CAL	W4:NR_COM_ MDS	W5:NR_ COM	W6:NR_ ZON	W7:NR_FULL_ CENSO	W8:NR_FULL
1	Total	4,03	4,03	4,04	4,03	4,04	4,06	4,03	3,99
2	Hombres	2,69	2,69	2,69	2,62	2,63	2,67	2,64	2,63
	Mujeres	5,48	5,48	5,50	5,64	5,64	5,61	5,59	5,48
3	18 a 29 años	2,83	2,83	2,69	2,70	2,73	2,77	2,75	2,65
	30 a 45 años	3,25	3,25	3,36	3,44	3,41	3,41	3,39	3,34
	46 a 65 años	4,77	4,77	4,78	4,65	4,68	4,69	4,64	4,76
	66 a 90 años	7,27	7,27	7,50	7,14	7,11	7,14	7,11	7,29
4	Zona norte	6,70	6,70	6,70	6,55	6,78	6,68	6,68	6,57
	Zona centro	4,60	4,60	4,26	4,26	4,29	4,31	4,31	4,25
	Zona sur	3,75	3,75	4,02	4,04	4,04	4,02	3,96	3,98
	RM	3,39	3,39	3,42	3,42	3,39	3,44	3,41	3,37
5	Educ. Media incompleta	4,11	4,11	4,22	4,21	4,24	4,24	4,19	4,16
	Educ. Media completa	3,63	3,63	3,50	3,52	3,48	3,52	3,49	3,52
	Educ. Superior incompleta	2,66	2,66	2,56	2,64	2,65	2,66	2,65	2,57
	Educ. Superior completa	3,75	3,75	3,84	3,68	3,65	3,67	3,73	3,65
6	Ocupado	3,36	3,36	3,37	3,31	3,29	3,37	3,37	3,30
	Desocupado	3,72	3,72	3,11	3,10	3,06	3,20	3,15	3,12
	Inactivo	4,97	4,97	5,02	5,12	5,17	5,03	4,98	5,02

Nota: se somborean celdas que destacan indicadores de Error Cuadrático Medio más bajos en la estimación para población específica.

