

LEVANTAMIENTO VII Ronda DE LA ENCUESTA DE PROTECCIÓN SOCIAL

Informe Metodológico sobre Muestreo,
Atrición, Factores de Expansión, Errores
Muestrales y Cálculo de Varianzas



Contenido

1. Consideraciones previas.....	9
1.1 Efectos de la pandemia de COVID 19 en la aplicación de la EPS.....	9
2. Muestreo en EPS VII Ronda.....	12
2.1. Configuración de la muestra de la EPS VII Ronda.....	12
2.2. Encuesta Presencial	13
2.3. Efectos de la pandemia en la EPS VII Ronda.....	17
2.4. Cobertura telefónica.....	18
2.5. Encuesta de Re-Entrevista	19
2.5.1. Marco muestral.....	19
2.5.2. Tamaño muestral	20
2.5.3. Selección de la muestra	21
2.5.3.1. Observaciones de exclusión forzosa	21
2.5.3.2. Observaciones de inclusión forzosa	21
2.5.3.3. Sobredimensionamiento de la muestra.....	22
2.5.4. Tamaño de la muestra	22
2.5.5. Selección de la muestra de Vivos.....	23
2.5.6. Selección de la muestra de Impedidos	26
2.5.7. Comportamiento de la muestra de Re-Entrevista.....	27
2.6. Encuesta de Continuidad	27
2.6.1. Marco muestral y tamaño muestral	27
2.6.2. Selección de la muestra	28
2.6.3. Tamaño de la muestra	29
2.6.4. Comportamiento de la muestra de Continuidad.....	33
3. Atrición en la EPS.....	34
3.1. Razones de no respuesta.....	36
3.2. Atrición por olas.....	38
3.3. Comportamiento de variables seleccionadas a través de las olas	41
3.4. Atrición en la encuesta de Re-Entrevista.....	48
3.5. Conclusiones sobre atrición.....	51
4. Factores de expansión de corte transversal	52
4.1. Consideraciones generales sobre los factores de expansión en la EPS VII Ronda	52

4.2.	Factores de expansión de corte transversal de la encuesta Presencial	54
4.2.1.	Configuración de la muestra de la encuesta Presencial	54
4.2.2.	Ajustes por atrición y no respuesta	56
4.2.3.	Post-estratificación de los factores de corte transversal	69
4.2.4.	Factores de expansión finales de corte transversal	73
4.3.	Factores de expansión de corte transversal de la encuesta de Re-entrevista	75
4.3.1.	Configuración de la muestra	75
4.3.2.	Ajustes por cobertura telefónica	80
4.3.3.	Ajustes por atrición y no respuesta	82
4.3.4.	Post-estratificación de los factores de expansión de corte transversal	83
4.3.5.	Factores de expansión finales de corte transversal	85
4.4.	Factores de expansión de corte transversal de la encuesta de Continuidad	85
4.4.1.	Configuración de la muestra de la encuesta de Continuidad	85
4.4.2.	Ajustes por selección de la muestra	87
4.4.3.	Ajustes por cobertura telefónica	89
4.4.4.	Ajustes por atrición y no respuesta	91
4.4.5.	Post-estratificación	92
4.4.6.	Factores de expansión finales de corte transversal	95
5.	Factores de expansión longitudinales	97
5.1.	Factores de expansión longitudinales de la encuesta Presencial	97
5.1.1.	Panel 2002-2004-2006-2009-2015-2019	97
5.1.2.	Panel 2004-2006-2009-2015-2019	100
5.1.3.	Panel 2015-2019	102
5.2.	Factores de expansión longitudinales para la encuesta Presencial y de Re-entrevista	103
5.3.	Factores de expansión longitudinales de la encuesta de Continuidad	105
5.3.1.	Panel 2002-2004-2006-2009-2015-2019	105
5.3.2.	Panel 2004-2006-2009-2015-2019	107
5.3.3.	Panel 2015-2019	108
5.4.	Ajustes adicionales a los factores de expansión	110
6.	Estimación de varianza y errores muestrales	111
6.1.	Errores muestrales en la encuesta Presencial y de Continuidad	113
6.2.	Errores muestrales en la encuesta de Re-Entrevista	117

6.3. Coeficiente de variación	120
6.4. Intervalos de confianza.....	121
6.5. Efecto diseño	121
Referencias	123

Índice de Tablas

Tabla 1. Encuesta Presencial: tipos de cuestionario aplicados	14
Tabla 2. Encuesta Presencial: distribución de la muestra por olas en que participó el entrevistado	14
Tabla 3. Encuesta Presencial: distribución de la muestra nominal, contactada y entrevistada.....	17
Tabla 4. Disponibilidad de número telefónico de contacto en la muestra de EPS VII Ronda.....	18
Tabla 5. EPS VII Ronda: disponibilidad de números de teléfono según fuente y total.....	19
Tabla 6. Encuestas logradas en la encuesta Presencial, por región y tipo de cuestionario aplicado	20
Tabla 7. <i>Clusters</i> por cada sub-muestra representada en la muestra final después de considerar cobertura telefónica – Muestra de Vivos para Re-entrevista.....	24
Tabla 8. Encuesta de Re-Entrevista: distribución de la muestra nominal, contactados y respuesta	27
Tabla 9. Disposición de casos según participación en la encuesta Presencial	28
Tabla 10. Cobertura telefónica en encuesta de Continuidad	30
Tabla 11. Encuesta de Continuidad: Distribución de la muestra nominal, contactados y respuesta.....	34
Tabla 12. Origen de las observaciones disponibles en la EPS VII Ronda.....	35
Tabla 13. Encuesta Presencial: comportamiento de la muestra.....	56
Tabla 14. Ejemplo de resultado de estimación de la probabilidad de ser visitado en la encuesta Presencial, Afiliados 2002	59
Tabla 15. Ejemplo de distribución de probabilidad estimada de ser visitado en la encuesta Presencial, sub-muestra Afiliados 2002.....	61
Tabla 16. Ejemplo de resultado de estimación de la probabilidad de Responder + Rechazar Ser Visitado en la encuesta Presencial, Afiliados 2002	61
Tabla 17. Ejemplo de distribución de probabilidad estimada de Responder + Rechazar Ser Visitado en la encuesta Presencial, sub-muestra Afiliados 2002.....	64
Tabla 18. Ejemplo de resultado de estimación de la probabilidad de Responder Responder + Rechazar, Ser Visitado en la encuesta Presencial, Afiliados 2002	64
Tabla 19. Ejemplo de distribución de probabilidad estimada de Responder (Responder + Rechazar, Ser Visitado) en la encuesta Presencial, sub-muestra Afiliados 2002.....	66
Tabla 20. Ejemplo de distribución de probabilidad incondicional estimada de Responder la encuesta Presencial, sub-muestra Afiliados 2002	67
Tabla 21. Ejemplo de agrupación por ventiles de la probabilidad incondicional estimada de no respuesta de la encuesta Presencial, sub-muestra Afiliados 2002.....	68
Tabla 22. Población de 23 años y más a Noviembre de 2019, según región, sexo y grupo etario.....	70
Tabla 23. Muestra de entrevistados Vivos de la encuesta Presencial, según región, sexo y grupo etario ..	71
Tabla 24. Muestra expandida de entrevistados Vivos de la encuesta Presencial, según región, sexo y grupo etario	72
Tabla 25. Coeficientes de ajuste de post-estratificación para entrevistados Vivos de la encuesta Presencial, según región, sexo y grupo etario	73
Tabla 26. Comportamiento de los factores de expansión de corte transversal de la encuesta Presencial, y de las poblaciones representadas por cada uno, según sub-muestras y muestra total.....	73
Tabla 27. Encuesta de Re-Entrevista a Vivos: comportamiento de la muestra	79
Tabla 28. Encuesta de Re-Entrevista a Impedidos: comportamiento de la muestra.....	80
Tabla 29. Estimación de la probabilidad de tener cobertura telefónica en la encuesta	81

Tabla 30. Distribución de la probabilidad estimada de tener teléfono en la encuesta de Re-Entrevista, sub-muestra Vivos	82
Tabla 31. Proyección de población de 23 años y más a Julio de 2020, según macrozona,	83
Tabla 32. Muestra de entrevistados Vivos de la encuesta de Re-Entrevista, según macrozona,	83
Tabla 33. Muestra expandida de entrevistados Vivos de la encuesta de Re-Entrevista, según macrozona, sexo y grupo etario	84
Tabla 34. Coeficientes de ajuste de post-estratificación para entrevistados Vivos de la encuesta de Re-Entrevista, según macrozona, sexo y edad	84
Tabla 35. Comportamiento de los factores de expansión de corte transversal de la encuesta de Re-Entrevista, y de las poblaciones representadas por cada uno	85
Tabla 36. Encuesta de Continuidad: Comportamiento de la muestra	87
Tabla 37. Estimación de la probabilidad de tener cobertura telefónica en la encuesta de Continuidad	90
Tabla 38. Distribución de la probabilidad estimada de cobertura telefónica en la encuesta de Continuidad	91
Tabla 39. Proyección de población de 23 años y más a Agosto de 2020, según macrozona,	93
Tabla 40. Muestra de entrevistados Vivos de la encuesta de Continuidad, según macrozona,	93
Tabla 41. Muestra expandida de entrevistados Vivos de la encuesta de Re-Entrevista, según macrozona, sexo y grupo etario	94
Tabla 42. Coeficientes de ajuste de post-estratificación para entrevistados Vivos de la encuesta de Continuidad, según macrozona, sexo y edad	94
Tabla 43. Comportamiento de los factores de expansión de corte transversal de la encuesta de Re-Entrevista, y de las poblaciones representadas por cada uno	95
Tabla 44. Estimación de la probabilidad de participar en el panel 2002-2004-2006-2009-2015-2019, encuesta Presencial	98
Tabla 45. Panel 2002-2004-2006-2009-2015-2019, encuesta Presencial: Comportamiento de la muestra y los factores de expansión	99
Tabla 46. Proyección de población de 35 años y más a Noviembre de 2019, según macrozona, sexo y grupo etario	100
Tabla 47. Coeficientes de ajuste de post-estratificación: Muestra Panel 2002 2004 2006 2009 2015 2019, encuesta Presencial, según macrozona, sexo y grupo etario	100
Tabla 48. Panel 2004-2006-2009-2015-2019, encuesta Presencial: Comportamiento de la muestra y los factores de expansión	101
Tabla 49. Proyección de población de 34 años y más a Noviembre de 2019, según macrozona, sexo y grupo etario	101
Tabla 50. Coeficientes de ajuste de post-estratificación: Muestra Panel 2004 2006 2009 2015 2019, encuesta Presencial, según macrozona, sexo y grupo etario	102
Tabla 51. Panel 2015-2019, encuesta Presencial: Comportamiento de la muestra y los factores de expansión	102
Tabla 52. Proyección de población de 23 años y más a Noviembre de 2019, según macrozona, sexo y grupo etario	103
Tabla 53. Coeficientes de ajuste de post-estratificación: Muestra Panel 2015 2019, encuesta Presencial, según macrozona, sexo y grupo etario	103

Tabla 54. Panel encuesta Presencial y encuesta de Re-Entrevista: Comportamiento de la muestra y los factores de expansión	104
Tabla 55. Proyección de población de 23 años y más a Julio de 2020, según macrozona, sexo y grupo etario	104
Tabla 56. Coeficientes de ajuste de post-estratificación: Panel encuesta Presencial y encuesta de Re-Entrevista, según macrozona, sexo y grupo etario	105
Tabla 57. Panel 2002-2004-2006-2009-2015-2019, encuesta de Continuidad: Comportamiento de la muestra y los factores de expansión.....	106
Tabla 58. Proyección de población de 35 años y más a Agosto de 2020, según macrozona, sexo y grupo etario	106
Tabla 59. Coeficientes de ajuste de post-estratificación: Muestra Panel 2002 2004 2006 2009 2015 2019, encuesta de Continuidad, según macrozona, sexo y grupo etario .	107
Tabla 60. Panel 2004-2006-2009-2015-2019, encuesta de Continuidad: Comportamiento de la muestra y los factores de expansión	107
Tabla 61. Proyección de población de 34 años y más a Agosto de 2020, según macrozona, sexo y grupo etario	108
Tabla 62. Coeficientes de ajuste de post-estratificación: Muestra Panel 2004 2006 2009 2015 2019, encuesta de Continuidad, según macrozona, sexo y grupo etario	108
Tabla 63. Panel 2015-2019, encuesta de Continuidad: Comportamiento de la muestra y los factores de expansión	109
Tabla 64. Proyección de población de 23 años y más a Agosto de 2020, según macrozona, sexo y grupo etario	109
Tabla 65. Coeficientes de ajuste de post-estratificación: Muestra Panel 2015 2019, encuesta de Continuidad, según macrozona, sexo y grupo etario.....	110

Índice de Figuras

Figura 1. EPS VII Ronda: temporalidad del trabajo de campo	11
Figura 2. Ajuste a los factores de expansión originales de la muestra de Vivos.....	26
Figura 3. Composición de la muestra EPS VII Ronda según última participación en la encuesta.....	36
Figura 4. Atrición por olas en la encuesta Presencial, en unidades observacionales y porcentajes	38
Figura 5. Atrición acumulada por ola hasta la EPS VII Ronda.	40
Figura 6. Distribución de respuestas por región y por ola	41
Figura 7. Distribución de sexo de entrevistados entre olas	43
Figura 8. Distribución de afiliados al sistema de pensiones por olas.....	44
Figura 9. Distribución edad según participación en encuestas de EPS VII Ronda	45
Figura 10. Grupos etarios por ola	47
Figura 11. Distribución del estado civil entre olas	48
Figura 12. Contacto por ola en encuesta de Re-Entrevista.....	49
Figura 13. Tasas de respuesta y rechazo a la encuesta de Re-Entrevista	49
Figura 14. Atrición por tramo etario en la encuesta de Re-Entrevista	50
Figura 15. Situación de no respuesta por tramo etario en la encuesta de Re-Entrevista	51
Figura 16. Atrición y no respuesta por sexo en la encuesta de Re-Entrevista.....	52
Figura 17. Configuración de la muestra de la encuesta Presencial de la EPS VII Ronda.....	54
Figura 18. Configuración de la muestra de la encuesta de Re-Entrevista de la EPS VII Ronda	76
Figura 19. Configuración de la muestra de la encuesta de Continuidad de la EPS VII Ronda	86

1. Consideraciones previas

En este informe se presentan cinco aspectos íntimamente relacionados con la capacidad de una encuesta para representar el fenómeno que está midiendo. Las encuestas, como tales, suelen apuntar a capturar información de calidad, pero sólo para un grupo pequeño de participantes. Frente a esta situación, el rol del muestreo y cómo luego se recomponen los comportamientos poblacionales a partir de la muestra, son claves para obtener resultados plausibles.

La VII ronda EPS tuvo por objetivo realizar el seguimiento a las personas que conformaban el panel del marco muestral de la EPS (es decir, personas que respondieron o contestaron la encuesta en alguna ronda de la EPS. No se agregó nueva muestra, ya que se buscaba la recuperación y fidelización del panel de entrevistados.

La EPS es una encuesta longitudinal de tipo panel fijo más nacimientos, esto significa que en un momento del tiempo se realizó una selección de personas representativas a nivel nacional, a las que se agregan muestras de casos nuevos (o nacimientos) para actualizar el panel cada cierta cantidad de años, también conocidos como “refrescos”.

Dado que la última actualización del marco muestral de la EPS se realizó en el año 2015, incorporando al panel quienes habían cumplido 18 años entre 2004 y 2015, la actual ronda, al dedicarse sólo a recopilar los casos panel, corresponde a una encuesta con un panel “envejecido” representativo de la población adulta del país con una muestra de personas de 22 años y más¹.

1.1 Efectos de la pandemia de COVID 19 en la aplicación de la EPS

La VII Ronda de la EPS, originalmente prevista para desarrollarse hacia finales de 2019 y durante el primer semestre de 2020, tenía un objetivo de 18.000 encuestas efectivas logradas a partir de una muestra panel de 24.882 observaciones disponibles. Estas observaciones correspondían a participantes de panel de las seis olas previas, tratándose en algunos casos de personas que respondieron en varias de ellas, y en otros casos, de personas que participaron sólo en alguna ola en particular.

Sin embargo, la llegada de la pandemia de coronavirus y su impacto en la salud del país, obligaron a terminar anticipadamente el trabajo de campo presencial, atendiendo a las indicaciones sanitarias correspondientes. Esto implicó que el trabajo de campo de la encuesta presencial (en adelante EPS

¹ Dado que por causas coyunturales la VII ronda EPS consistió en la ejecución de 3 encuestas, principalmente, durante el 2020, la elaboración de los factores de expansión de corte transversal, se ajustó a la población de 23 años y más. Se sumaron 5 años a la edad de la muestra considerando que el refresco de la EPS VI ronda fue seleccionado en el año 2015 y la mayor parte del terreno de la VII ronda se ejecutó 5 años después.

presencial) se cerrara, preliminarmente, en marzo de 2020 con un total de 8.656 encuestas logradas (vivos, fallecidos e impedidos). Ésta se desarrolló, entonces entre el 14 de diciembre de 2019 y 22 de marzo de 2020.

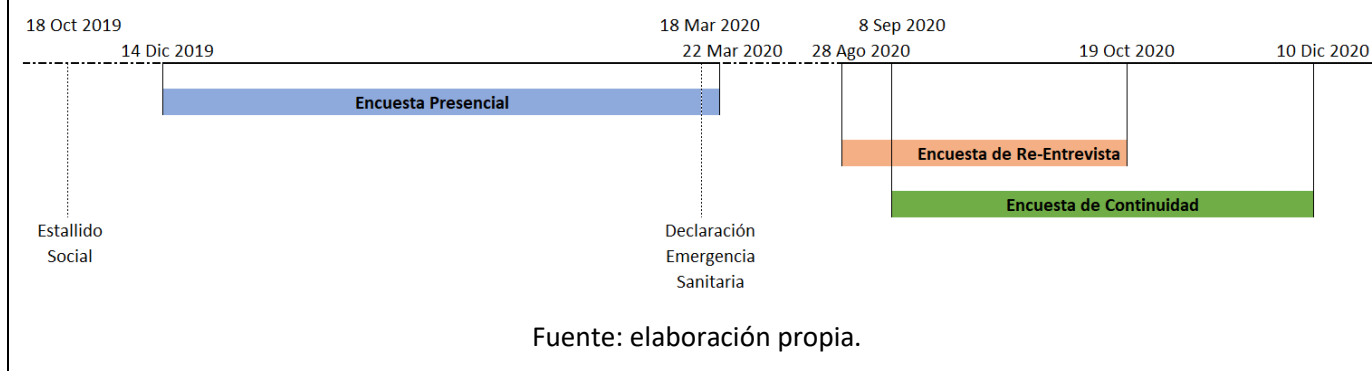
La crisis sanitaria levantó un gran desafío, ya que para darle continuidad a la EPS se debía proceder con una nueva estrategia para completar el objetivo de la encuesta en cuanto a número de observaciones y contenido informativo. Para ello, se adjudicó de forma directa una extensión del proyecto original en el que se contemplaron dos encuestas telefónicas:

- Encuesta de “Continuidad”: que pretendía conseguir un mínimo de 5.000 entrevistas a partir de todos los entrevistados que no respondieron la encuesta presencial, para recabar información similar a la obtenida en la encuesta en terreno (encuesta de continuación del levantamiento presencial, llegando aquellas personas a las que no habían respondido la encuesta presencial).
- Encuesta de “Re-entrevista”: que pretendió conseguir información de un mínimo de 1.800 entrevistados que participaron de la EPS Presencial, para actualizar la información de las historias laborales de los panelistas y recopilar datos sobre diversos aspectos relacionados con el impacto de la pandemia en el corto plazo.

De esta forma, la VII ronda está conformada por tres sub-muestras que responden a distintas encuestas (Presencial, de Re-Entrevista y de Continuidad) usando distintas modalidades, pero persiguiendo un mismo objetivo: caracterizar el comportamiento de la población chilena en términos laborales y previsionales.

La **Figura 1** muestra las fechas en que se desarrolló el trabajo de campo de cada encuesta que compone la VII Ronda de la EPS, junto a otros hitos importantes. Como puede observarse, la información fue relevada en tres momentos distintos del tiempo, marcados por un antes y un después de la declaración de emergencia a raíz de la pandemia de coronavirus. El trabajo de campo del proyecto inicial debió ser retrasado algunas semanas debido a los efectos del estallido social comenzado el 18 de octubre de 2019. Y el advenimiento de la pandemia implicó el cierre del trabajo de campo presencial y el replanteo del proyecto cambiando su modalidad y alcance originales. Así, la encuesta presencial se desarrolló entre el 14 de diciembre de 2019 y el 22 de marzo de 2020, mientras que la encuesta de Re-Entrevista, realizada a quienes habían respondido la encuesta Presencial, tomó lugar entre el 28 de agosto y el 19 de octubre de 2020. Por su parte, el trabajo de campo de la encuesta de Continuidad, dirigida a quienes no habían respondido la encuesta Presencial, se desarrolló entre el 8 de septiembre y el 10 de diciembre de 2020.

Figura 1. EPS VII Ronda: temporalidad del trabajo de campo



A continuación, se estudia el comportamiento de los entrevistados y su participación y respuesta a la encuesta. Los participantes de las encuestas de panel no necesariamente participan habitualmente en el estudio. Algunos de ellos lo hacen una vez y no vuelven a ser parte del estudio por diversas razones (por ejemplo, no se los puede ubicar, rechazan la encuesta en rondas posteriores, fallecen). Otros participan intermitentemente del estudio, y en cada oportunidad pueden tener diferentes motivaciones para autoseleccionarse como entrevistados que responden o rechazan la encuesta. Cualquiera sea la razón de no participación, estas situaciones constituyen una fuente de potencial sesgo en los resultados de las encuestas en general, que se tratan en el análisis de atrición y no respuesta de la encuesta.

Para corregir los sesgos emergentes por atrición y no respuesta, se deben realizar un conjunto de ajustes a los factores de expansión. Éstos representan la importancia de la contribución informativa de cada observación a la muestra completa, y es lo que permite utilizar los resultados de la encuesta para hacer inferencia sobre el comportamiento poblacional. Los factores de expansión se obtienen como el inverso de la probabilidad de haber sido elegido para constituir parte de la muestra, por lo que los problemas de cobertura del marco muestral o de la ejecución del trabajo en terreno, pérdidas de observaciones por atrición y no respuesta y, finalmente, cambios en la población del momento de la encuesta respecto a la población muestreada originalmente, deben ser ajustados a través de estos factores para lograr mantener la capacidad de inferencia de la encuesta relevada. En este informe, se detallan, paso a paso, todos los ajustes que se realizaron a los factores de expansión, tanto de corte transversal como longitudinales, para los tres componentes de la EPS VII Ronda.

Dado que se está haciendo una encuesta para inferir el comportamiento de la población, y teniendo en cuenta el diseño complejo de la EPS, es esperable que los resultados observados tengan errores respecto al verdadero comportamiento de la población. Por ello, se explica el comportamiento de los errores muestrales, el efecto diseño sobre las varianzas, y los métodos de cálculo de errores y varianzas para las distintas encuestas que constituyen la EPS VII Ronda y que permiten realizar inferencia.

Dada la complejidad de la estructura de la EPS VII Ronda, se abordarán las tres partes de la encuesta en forma conjunta, resaltándose las particularidades técnicas de cada una de ellas.

2. Muestreo en EPS VII Ronda

2.1. Configuración de la muestra de la EPS VII Ronda

La muestra de la EPS se ha realizado siguiendo una metodología de “panel fijo + nacimientos”, es decir, en cada ocasión que se agregó muestra, se hizo en razón de la actualización del marco muestral (o para extenderlo, como sucedió en 2004, para representar la población total del país de 18 o más años de edad). Así, en la primera ronda (2002), sólo se incluyó población afiliada al sistema de pensiones representativa del perfil demográfico del país para mayores de 18 años. En el 2004, se incorporaron dos muestras: una muestra correspondió a la actualización del marco muestral de la EPS I Ronda incorporando a quienes habían cumplido 18 años en el período 2002-2004. La otra muestra se incorporó para extender el marco muestral para que la encuesta tuviera representatividad nacional para mayores de 18 años, hayan o no estado afiliados al sistema de pensiones. Por lo tanto, esta segunda muestra incorporó a mayores de 18 años al 2004 que no estaban afiliados al sistema. En el año 2012, se decidió incorporar una muestra para actualizar el marco muestral; sin embargo, la nueva muestra fue descartada por razones de calidad estadística. Finalmente, en 2015 se incluyó una nueva actualización del marco muestral incorporando a quienes habían cumplido 18 años entre 2004 y 2015. En la presente ronda, no se incluyó refresco ni actualización del marco muestral y sólo se buscó fortalecer el vínculo con el panel.

La selección de la muestra de 2002 se realizó a través de un proceso en dos etapas para lograr obtener un total de 20.000 observaciones: en primer término, se seleccionaron *clusters* geográficos (comunidades de inclusión forzosa más comunidades de inclusión probabilística), mientras que en la segunda etapa se seleccionó a los entrevistados según su condición de afiliado al sistema, sexo y edad. De esta manera, se configuró una muestra representativa de estos aspectos de la población de cada *cluster*, obteniéndose idéntica probabilidad de selección para cada entrevistado dentro del *cluster* condicional en estas características². Las incorporaciones de muestra en 2004 y 2015 siguieron el mismo diseño, manteniendo la estructura de *clusters* y selección de unidades primarias de muestreo.

La VII Ronda de la EPS, originalmente prevista para desarrollarse hacia finales de 2019 y durante el primer semestre de 2020, tenía un objetivo de 18.000 encuestas efectivas logradas a partir de una muestra panel de 24.882 observaciones disponibles.

Si bien el marco muestral y los detalles de selección original de cada muestra de panel que conforma la EPS, se pueden encontrar en el Informe Técnico de la EPS VI Ronda, y se omiten aquí en razón del público conocimiento de esta información, vale la pena resumir los puntos más salientes:

- Según el documento “Diseño muestral y factores de expansión VI Ronda EPS” de la SPS – Centro UC Encuestas y Estudios Longitudinales de 2016 (en adelante, “Diseño 2015”), las cuatro muestras

² Véase Lohr (2010), capítulo 8.

que están incluidas en la EPS fueron realizadas tomando como distribución de la población de Chile las correspondientes a la obtenidas en las CASEN 2000 y 2003 (Diseño 2015, pp. 9).

- Las cuatro muestras de la EPS corresponden a:
 1. Afiliados 2002 – EPS I Ronda
 2. Nuevos Afiliados 2004 – EPS II Ronda
 3. No afiliados 2004 – EPS II Ronda
 4. Nuevos mayores de 18 años y más 2015 – EPS VI Ronda³
- A estas muestras se le fueron deduciendo los fallecidos entrevistados en cada ronda y otras depuraciones como entrevistados viviendo en el extranjero, lo que implicó ajustes proporcionales al tamaño en cada ronda para dar cuenta de la reducción del marco muestral.
- Luego de estos ajustes, llegamos al gran total de 24.882 observaciones disponibles para la VII Ronda de la EPS.
- Dado que el conjunto de *clusters* de la EPS ha permanecido inalterado a través de las actualizaciones muestrales, se toma esa información para determinar la importancia relativa de cada comuna en el muestreo.⁴ La selección de las comunas y los entrevistados dan origen al denominado “factor base”, es decir, el factor de expansión primigenio para el trabajo de muestreo de la EPS VII Ronda.

Estas 24.882 observaciones correspondían a participantes de panel de las seis olas previas, tratándose en algunos casos de personas que respondieron en varias de ellas, y en otros casos, de personas que participaron sólo en alguna ola en particular. El objetivo de la EPS VII Ronda no era sólo el levantamiento, sino también la recuperación y fidelización del panel de entrevistados.

2.2. Encuesta Presencial

La EPS VII Ronda fue concebida originalmente como una encuesta presencial, en que se contacta cara a cara a cada entrevistado. El trabajo de campo de esta encuesta comenzó el 14 de diciembre de 2019. Sin

³ Dado que los “nacimientos” corresponden a todos los que se incorporaron a la población en edad de trabajar entre 2004 (última fecha de incorporación de nueva muestra a la EPS) y 2015, los nacimientos incluyen a personas con edades entre 18 y 32 años cumplidos a 2015.

⁴ Nótese que, en la primera fase de la selección de comunas, tal como se explica en el documento de Diseño 2015, se empleó información poblacional basada en la CASEN 2000 que era la mejor información disponible en el momento (el Censo se realizó en 2002). Adicional a esto, en una segunda etapa, se pasa del “número de habitantes por comuna” al “número de afiliados por comuna”, ajustando de esta manera los cálculos de probabilidades de selección al marco muestral de la encuesta (es decir, al listado de individuos proveniente de los registros administrativos que cumplían restricciones de edad y afiliación/no afiliación al sistema de pensiones). Para la ampliación del marco muestral realizado en 2004, se recurrió a información poblacional proyectada a partir de la CASEN 2003, mientras que, para la actualización del marco muestral de 2015, se empleó información proporcionada por el Registro Civil.

embargo, la extensión de la pandemia de coronavirus y su impacto en la salud del país, obligaron a terminar anticipadamente el trabajo de campo, atendiendo a las indicaciones sanitarias correspondientes. Esto implicó que el trabajo de campo de la EPS se cerrara el 22 de marzo de 2020 con un total de 8.655 encuestas logradas. Estas encuestas se levantaron, según la situación del entrevistado, en uno de los tres cuestionarios disponibles: entrevistados vivos, fallecidos e impedidos, tal como se detalla en la **Tabla 1**.

Tabla 1. Encuesta Presencial: tipos de cuestionario aplicados

Cuestionario	Frecuencia	Porcentaje	Porcentaje Acumulado
Vivos	7.800	90,12	90,12
Fallecidos	516	5,96	96,08
Impedidos	339	3,92	100
Total	8.655	100	

Fuente: Elaboración propia

La situación general de la muestra, dependiendo de las olas en las que cada entrevistado del panel ha participado, se presenta en la **Tabla 2**. Como se puede observar, el comportamiento es relativamente homogéneo entre quienes fueron visitados (hayan respondido o no la encuesta) y los entrevistados de panel que no alcanzaron a ser recorridos.

Tabla 2. Encuesta Presencial: distribución de la muestra por olas en que participó el entrevistado

	Respondió Encuesta Presencial	No respondió, pero fue visitado	No fue visitado	Total	Respondió Encuesta Presencial (%)	No respondió, pero fue visitado (%)	No fue visitado (%)
2002	144	243	412	799	18,0%	30,4%	51,6%
2002 2004	97	174	260	531	18,3%	32,8%	49,0%
2002 2004 2006	178	242	344	764	23,3%	31,7%	45,0%
2002 2004 2006 2009	362	408	658	1.428	25,4%	28,6%	46,1%
2002 2004 2006 2009 2012	449	374	692	1.515	29,6%	24,7%	45,7%
2002 2004 2006 2009 2012 2015	2.136	646	1.626	4.408	48,5%	14,7%	36,9%
2002 2004 2006 2009 2015	638	291	645	1.574	40,5%	18,5%	41,0%
2002 2004 2006 2012	50	67	92	209	23,9%	32,1%	44,0%
2002 2004 2006 2012 2015	112	61	111	284	39,4%	21,5%	39,1%
2002 2004 2006 2015	132	77	116	325	40,6%	23,7%	35,7%
2002 2004 2009	36	60	76	172	20,9%	34,9%	44,2%
2002 2004 2009 2012	17	18	37	72	23,6%	25,0%	51,4%
2002 2004 2009 2012 2015	64	34	63	161	39,8%	21,1%	39,1%
2002 2004 2009 2015	46	28	62	136	33,8%	20,6%	45,6%

2002 2004 2015	59	47	58	164	36,0%	28,7%	35,4%
2002 2006	53	125	154	332	16,0%	37,7%	46,4%
2002 2006 2009	66	98	128	292	22,6%	33,6%	43,8%
2002 2006 2009 2012	74	56	78	208	35,6%	26,9%	37,5%
2002 2006 2009 2012 2015	159	99	130	388	41,0%	25,5%	33,5%
2002 2006 2009 2015	84	58	77	219	38,4%	26,5%	35,2%
2002 2006 2012	9	22	26	57	15,8%	38,6%	45,6%
2002 2006 2012 2015	29	9	24	62	46,8%	14,5%	38,7%
2002 2006 2015	31	25	37	93	33,3%	26,9%	39,8%
2002 2009	29	49	71	149	19,5%	32,9%	47,7%
2002 2009 2012	8	16	17	41	19,5%	39,0%	41,5%
2002 2009 2012 2015	26	8	21	55	47,3%	14,5%	38,2%
2002 2009 2015	30	11	36	77	39,0%	14,3%	46,8%
2002 2015	36	26	55	117	30,8%	22,2%	47,0%
2004	26	53	107	186	14,0%	28,5%	57,5%
2004 2006	37	70	124	231	16,0%	30,3%	53,7%
2004 2006 2009	84	115	244	443	19,0%	26,0%	55,1%
2004 2006 2009 2012	129	93	184	406	31,8%	22,9%	45,3%
2004 2006 2009 2012 2015	625	182	487	1.294	48,3%	14,1%	37,6%
2004 2006 2009 2015	173	70	236	479	36,1%	14,6%	49,3%
2004 2006 2012	9	15	27	51	17,6%	29,4%	52,9%
2004 2006 2012 2015	36	11	29	76	47,4%	14,5%	38,2%
2004 2006 2015	29	23	30	82	36,1%	27,7%	36,1%
2004 2009	15	21	31	67	22,4%	31,3%	46,3%
2004 2009 2012	2	6	17	25	8,0%	24,0%	68,0%
2004 2009 2012 2015	21	6	15	42	50,0%	14,3%	35,7%
2004 2009 2015	19	8	9	36	52,8%	22,2%	25,0%
2004 2012	3	4	6	13	23,1%	30,8%	46,2%
2004 2015	15	12	14	41	36,6%	29,3%	34,1%
2012 2015	4	2	6	12	33,3%	16,7%	50,0%
2015	2.274	1.551	2.941	6.766	33,6%	22,9%	43,5%
Total	8.655	5.614	10.613	24.882	34,8%	22,6%	42,7%

Nota: Levantamiento al 22 de marzo de 2020 dada crisis sanitaria del país.

Fuente: Elaboración propia.

En la **Tabla 3** se muestra el comportamiento de la muestra de la encuesta Presencial, según el panel del que provienen sus entrevistados. Cada columna de la tabla identifica el panel de origen (momento y razón de la incorporación de la muestra). Por ejemplo, en la columna “Afiliados 2002”, podemos observar que, al inicio del trabajo de campo, se tenían disponibles 14.632 entrevistados que habían sido incorporados en la

primera ola de la EPS, y que podían, potencialmente, responder la encuesta de VII Ronda.⁵ Esto constituye lo que se denomina “muestra nominal” de la encuesta, y el concepto se extiende a cada sub-muestra de la EPS y a la muestra total. Por su parte, en la EPS II Ronda, se incorporaron dos grupos de nuevos entrevistados que aparecen en las columnas “Nuevos Afiliados 2004” y “No Afiliados 2004”, con muestras nominales de 918 y 2.554 entrevistados, respectivamente. Finalmente, en la columna “Nuevos 2015”, que fueron incorporados en la EPS VI Ronda representando a toda la nueva población en edad de trabajar que cumplió 18 años entre 2004 y 2015, se informa una muestra nominal de 6.778 entrevistados.

De este conjunto de puntos de contacto disponibles, se visitaron 8.464 entrevistados. Los otros 6.168 potenciales entrevistados no llegaron a ser visitados para la realización de la encuesta Presencial debido a la interrupción de estudio causado por la pandemia.

De éstos 8.464 entrevistados que se visitaron, un total de 5.669 entrevistados tuvieron una respuesta definitiva a la encuesta, ya sea que aceptaran o rechazaran llevarla a cabo. Los 2.795 casos restantes no quedaron en una situación intermedia en que se debían realizar nuevas gestiones para conseguir contacto, respuesta o rechazo a la encuesta.

Finalmente, de los 5.669 entrevistados para los que se obtuvo una respuesta a su participación en la encuesta, 5.154 decidieron responder la encuesta completa. Es decir, 515 entrevistados rechazaron la realización de la encuesta presencial.

Teniendo en cuenta las cuatro sub-muestras que conforman la EPS VII Ronda, se lograron 8.655 encuestas (vivos, fallecidos e impedidos) que implicaron la visita a 14.187 puntos de contacto.

La población representada por cada sub-muestra, en la parte final de la **Tabla 3**, corresponde a la cantidad de personas que fueron muestreadas originalmente. Ese es el resumen del marco muestral de cada sub-muestra. El factor de expansión base (o de selección) promedio es un indicador de la contribución promedio de información de cada observación de la encuesta.⁶

⁵ Nótese que, si bien la muestra original de la EPS 2002 tenía 17.246 entrevistados, a través de las olas, se fueron dando de baja observaciones por razones de fallecimiento, impedimento u otras de fuerza mayor como residencia en el extranjero.

⁶ El factor de expansión suele ser interpretado como la cantidad de personas similares que se están relevando en la encuesta al entrevistar a una de ellas. Sin embargo, esa interpretación procede cuando el factor de expansión es construido bajo un enfoque frecuentista, mientras que, en el presente contexto, el factor de expansión es probabilístico, ya que representa el inverso de la probabilidad de selección de cada individuo en la muestra. Por lo tanto, la interpretación no es directa.

Tabla 3. Encuesta Presencial: distribución de la muestra nominal, contactada y entrevistada

Tipo de Muestra	Afiliados 2002	Nuevos Afiliados 2004	No Afiliados 2004	Nuevos 2015	Total
Muestra Nominal	14.632	918	2.554	6.778	24.882
Visitados	8.464	476	1.427	3.820	14.187
Responde + Rechaza Visitado	5.669	271	1.018	2.417	9.375
Responde Responde + Rechaza, Visitado	5.154	253	970	2.278	8.655
Población representada originalmente	8.074.667	527.914	3.994.879	3.250.561	15.848.021
Factor de expansión base promedio	1566,680	2086,617	4118,432	1426,936	1831,083

Fuente: Elaboración propia.

2.3. Efectos de la pandemia en la EPS VII Ronda

Como se mencionó anteriormente, el advenimiento de la pandemia de COVID-19 implicó una terminación temprana del estudio en terreno, habiendo transcurrido aproximadamente la mitad del tiempo destinado al mismo.

Después de un largo análisis y consultas, la SPS planteó la continuación del estudio, pero con cambios en la modalidad y en el enfoque, aprovechando las particulares circunstancias que se estaban viviendo. Por una parte, la modalidad pasó de entrevistas cara a cara, a entrevistas telefónicas. Por otra parte, se buscó obtener información que permitiera dimensionar los efectos de la pandemia sobre la población representada por la EPS. Esta estrategia permitiría contar con información antes-después de la pandemia sobre varios aspectos relevantes del comportamiento de los entrevistados, como evolución de la historia laboral, comportamiento previsional a través del tiempo, formalidad en el empleo, salud de los entrevistados, entre otros. También se podría analizar el impacto de la pandemia en el mercado laboral, al mismo tiempo que los alcances de algunas de las medidas implementadas por el gobierno en la contingencia.

Para llevar adelante estas tareas y cerrar el trabajo analítico de la encuesta presencial, la SPS propuso una extensión del proyecto original en el que se contemplan dos encuestas telefónicas:

- **Encuesta de Continuidad:** esta es una encuesta que pretendía conseguir 5.000 entrevistas a partir de todos los entrevistados que no respondieron la encuesta presencial, para recabar información similar a la obtenida en esta última.
- **Encuesta de Re-Entrevista:** esta encuesta pretendía recolectar información de 1.800 entrevistados que respondieron la encuesta Presencial entre los meses de diciembre de 2019 y marzo de 2020, y que informaría sobre diversos aspectos relacionados con el impacto de la pandemia en el corto plazo.

Por lo tanto, la EPS VII Ronda queda finalmente conformada por tres encuestas:

1. **Encuesta Presencial:** 8.655 casos obtenidos de forma presencial entre los meses de diciembre de 2019 y marzo de 2020.
2. **Encuesta de Continuidad:** encuesta telefónica realizada a quienes no respondieron la encuesta Presencial, con un total de 5.296 entrevistas completas.
3. **Encuesta de Re-Entrevista:** encuesta telefónica realizada a quienes respondieron la encuesta Presencial, con un total de 2.225 entrevistas completas.

2.4. Cobertura telefónica

La realización de encuestas telefónicas presentan un conjunto de particularidades que las diferencian de manera importante del resto de los enfoques de levantamiento.⁷ El primero y más evidente es que se requiere que el entrevistado tenga un número de teléfono (fijo o móvil) donde pueda ser contactado.

Si bien la base de datos de panel, a través de las olas, ha ido recolectando información de contacto telefónico, la información más confiable sólo estaba disponible para la muestra que respondió la encuesta presencial de la EPS VII Ronda y, en algunos casos, para aquellos que fueron visitados durante esa encuesta, pero para quienes no se pudo concretar la entrevista para la realización de la encuesta. La situación inicial de cobertura telefónica para la muestra completa se presenta en la **Tabla 4**.

Tabla 4. Disponibilidad de número telefónico de contacto en la muestra de EPS VII Ronda

	Respondió Encuesta Presencial	No respondió	Total	Respondió Encuesta Presencial (%)	No respondió (%)	Total (%)
No tiene teléfono registrado	1.786	5.813	7.599	23,5%	76,5%	30,5%
Teléfono de contacto registrado	6.852	10.364	17.217	39,8%	60,2%	69,2%
Sólo correo electrónico	17	50	67	25,4%	74,6%	0,3%
Total	8.655	16.227	24.882	34,8%	65,2%	

Fuente: elaboración propia en base a datos de la SPS.

IPSOS Chile utilizó esta información, enriqueciéndola con los números telefónicos levantados en terreno (muchos de ellos validados al ser usados en la supervisión telefónica) y con el poblamiento de información de contacto a través de una empresa externa, que contó con todos los resguardos de confidencialidad de información de los entrevistado. El resultado de esta combinación de fuentes mejoró sustancialmente la

⁷ Por ejemplo, véase Lavrakas (2008), Lohr (2008) y Steeh (2008).

cantidad de números telefónicos que se pudieron usar para contactar a los entrevistados de panel de la EPS VII Ronda.

En la **Tabla 5** se muestra la disponibilidad de números de contacto telefónico resultante de las tres fuentes: (i) la información contenida en el panel (histórica), (ii) la información actualizada a partir de la entrevista lograda en la EPS VII Ronda Presencial, y (iii) la información resultante del poblamiento realizado por IPSOS Chile.

Tabla 5. EPS VII Ronda: disponibilidad de números de teléfono según fuente y total

Situación de los entrevistados de panel		¿Tiene, al menos, un número de teléfono de contacto?			
		Si	No	Total	
					Si (%) No (%)
Respondió	Vivos	7,763	37	7,8	99.5% 0.5%
Encuesta	Fallecidos	479	37	516	92.8% 7.2%
Presencial	Impedidos	337	2	339	99.4% 0.6%
No Respondió Encuesta Presencial		14,901	1,326	16,227	91.8% 8.2%
Total		23,48	1,402	24,882	94.4% 5.6%

Fuente: Elaboración propia.

En la **Tabla 5** se observa que un 94,4% de las observaciones cuenta con, al menos, un número de contacto. Entre aquellas personas que no respondieron la encuesta presencial, el 91,8% tiene algún teléfono de contacto. La cifra es más alta para quienes respondieron la encuesta presencial, llegando casi al 100%.

2.5. Encuesta de Re-Entrevista

Para el caso particular de la encuesta de Re-Entrevista, dispusimos de 8.579 teléfonos antes de reducir la muestra para contabilizar fallecimientos detectados en la encuesta Presencial. Los entrevistados fallecidos, tal como en rondas anteriores de la EPS, fueron excluidos de la muestra, ya sea que hubiesen sido personas para las que se respondió un cuestionario, o que se contara con información sobre su fallecimiento.⁸

2.5.1. Marco muestral

La encuesta de Re-Entrevista de la EPS VII Ronda recabó información sobre diversos aspectos de los entrevistados y su hogar, entre los que se pueden mencionar: situación laboral y sus cambios por la pandemia, ingresos, salud, subsidios recibidos por la pandemia, entre otros; ahondando en los efectos de corto plazo de la pandemia sobreveniente. Esta encuesta fue diseñada (en especial, su instrumento) pensando en el formato de encuesta telefónica.

⁸ CDF=231. Véase la explicación a la **Tabla 3**.

El objetivo de esta encuesta fue conseguir 1.800 entrevistas a partir de todos los entrevistados en la encuesta Presencial. Esto implica que el marco muestral de la encuesta de Re-entrevista corresponde a todos los participantes de rondas anteriores de la EPS que hubiesen respondido la encuesta Presencial de la VII Ronda, y que dispusieran de un número de teléfono en el que se pudiera localizar al entrevistado de panel.

La muestra incluye entrevistados vivos (al momento del levantamiento), cuestionarios administrados a familiares sobrevivientes de entrevistados fallecidos, y cuestionarios aplicados a cuidadores y curadores de personas impedidas (temporal o permanentemente) de responder la encuesta. La **Tabla 6** muestra la distribución regional de los tres tipos de cuestionarios respondidos en la encuesta Presencial y que constituyen el marco muestral de la encuesta de Re-entrevista.

Tabla 6. Encuestas logradas en la encuesta Presencial, por región y tipo de cuestionario aplicado

Región	Encuestas aplicadas			Total encuestas
	Vivos	Fallecidos	Impedidos	
Tarapacá	87	2	5	94
Antofagasta	243	24	34	301
Atacama	217	10	1	228
Coquimbo	280	12	7	299
Valparaíso	598	41	30	669
Libertador B. O'Higgins	319	26	9	354
Maule	279	21	18	318
Biobío	1.081	98	37	1.216
Araucanía	209	9	7	225
Los Lagos	593	40	42	675
Aysén	35	0	1	36
Magallanes	101	9	1	111
Metropolitana	3.382	196	123	3.701
Los Ríos	187	14	16	217
Arica Y Parinacota	85	5	2	92
Ñuble	104	9	6	119
Total De Entrevistas	7.800	516	339	8.655

Fuente: Elaboración propia.

2.5.2. Tamaño muestral

Típicamente, el tamaño muestral se determina a partir de la pregunta de investigación (variable de interés) y las características de la población bajo estudio. Sin embargo, en este caso, la cantidad de encuestas a recabar se fijó *a priori* en 1.800 teniendo en cuenta varios factores como la restricción presupuestaria, el

periodo de tiempo para realizar el trabajo de campo, las tasas de éxito en encuestas telefónicas, y las particulares circunstancias que se vivían en el país en esos momentos.

2.5.3. Selección de la muestra

A partir de la información general sobre la selección de la muestra original de la EPS, se consideró que para la encuesta de Re-entrevista, tanto el marco muestral como la muestra seleccionada (participantes de la encuesta Presencial) estaban dadas. Así, se disponía de 8.655 observaciones que podían ser abordadas para obtener el objetivo de levantamiento. Sin embargo, un 0,9% de las observaciones no tenía teléfono de contacto que se pudiera emplear para llegar a ellas. Por lo tanto, se dispuso de 8.579 observaciones con cobertura telefónica. Sobre esta muestra se aplicaron los criterios de elegibilidad para la encuesta de Re-entrevista.⁹

Para determinar la muestra requerida para obtener exitosamente 1.800 encuestas completas, se tuvieron en cuenta tres aspectos adicionales instruidos por la SPS que determinan la elegibilidad de los entrevistados y, por ende, el tamaño final de la misma: (i) observaciones de exclusión forzosa, (ii) observaciones de inclusión forzosa, y (iii) determinación del sobredimensionamiento de la muestra debido a la no respuesta esperada.

2.5.3.1. Observaciones de exclusión forzosa

Por indicación de la SPS, se excluyó forzosamente el conjunto de observaciones correspondiente a entrevistados a quienes se aplicó el cuestionario de Fallecidos en la encuesta Presencial. En la muestra disponible con número de teléfono, los “fallecidos” representan un 5.6% de las observaciones (479 casos sobre 8.579 posibles). La exclusión de estas observaciones implica que la muestra disponible es de 8.100 casos

2.5.3.2. Observaciones de inclusión forzosa

También por indicación de la SPS, se estableció que se debía contactar y encuestar a todos los entrevistados a quienes se les aplicó un cuestionario de Impedidos en la ronda presencial.

Esto implica que la muestra incluye forzosamente los 337 casos de “Impedidos” para los que se ha conseguido un teléfono de contacto (99.4% del total de impedidos). Sin embargo, esto genera un desbalance de la muestra, ya que estos entrevistados tienen una participación mayor en el objetivo de

⁹ Nótese que se hace referencia a la cobertura telefónica antes de la determinación de la elegibilidad de las observaciones incluidas, porque la recopilación de números telefónicos de contacto (“poblamiento de la base de contacto”) se realizó sin tener en cuenta esta elegibilidad. De hecho, el poblamiento se hizo para las 24.882 observaciones que eran parte del marco muestral disponible para la EPS VII Ronda.

levantamiento de esta encuesta al que tienen en el marco muestral (encuesta Presencial) debido a su inclusión forzosa.

Para solucionar esta situación, se propuso a la SPS mantener la proporción de casos de Impedidos según los resultados del trabajo de campo de la encuesta Presencial para ser incluidos dentro del objetivo de 1.800 casos exitosos. Es decir, los “impedidos” con número de contacto (337 casos) representan el 4.2% de los 8.100 casos totales elegibles disponibles.

2.5.3.3. Sobredimensionamiento de la muestra

Como en la generalidad de las encuestas de panel, aquí también se tuvieron en cuenta los problemas de atrición, no respuesta y otros impedimentos que resultan en la imposibilidad de levantar las encuestas. Dado que se trata de una muestra sin posibilidad de reemplazos, se estableció un tamaño de la muestra seleccionada que excediera el tamaño objetivo del estudio, teniendo en cuenta las tasas de respuesta esperada para este tipo de encuesta.

2.5.4. Tamaño de la muestra

Teniendo en cuenta los tres aspectos mencionados anteriormente sobre elegibilidad de los entrevistados, se propuso la siguiente estrategia de distribución de encuestas según hubiesen respondido cuestionario de Vivos o Impedidos en la etapa presencial del presente estudio:

- Para mantener el balance entre Vivos e Impedidos, tal como surgieron de la encuesta Presencial, se propuso que las 1.800 encuestas a lograr se distribuyeran de la siguiente manera: el 95.8% de las mismas se administraría a Vivos, mientras que el 4.2% restante sería cubierto con cuestionarios realizados a Impedidos. Es decir, se planteó como objetivo de levantamiento **1.725 encuestas a Vivos y 75 encuestas a Impedidos**.¹⁰
- Respecto a los Impedidos, se aplicaría el criterio propuesto anteriormente. De los 337 casos disponibles de inclusión forzosa, dado que 75 serían parte del objetivo de la encuesta normal, **el resto (262 casos) se relevarían por encima del objetivo de las 1.800 encuestas**.
- Por lo tanto, **el sobredimensionamiento de muestra afectaría sólo a los entrevistados con el status de “vivos”** en la encuesta Presencial.

¹⁰ Participación de cada tipo de encuesta en el total:

$$\begin{aligned} \text{Encuestas a Vivos} &= \frac{7.763}{8.100} \cdot 1.800 = 1.725 \text{ encuestas} \\ \text{Encuestas a Impedidos} &= \frac{337}{8.100} \cdot 1.800 = 75 \text{ encuestas} \end{aligned}$$

Para determinar el sobredimensionamiento de muestra, se calculó la tasa de respuesta esperada para la encuesta de Re-Entrevista que depende de la probabilidad de tener cobertura telefónica, ser contactado telefónicamente, responder al contacto, y completar la encuesta exitosamente:

$$TR_{Re-Entrevista} = \Pr(CT|TF) \cdot \Pr(RE|CT) \cdot \Pr(CE|RE) = 31.2\%$$

donde TR es la tasa de respuesta, CT corresponde a “Logro de contacto telefónico”, TF corresponde a “Entrevistado en la encuesta Presencial que dispone de número telefónico para ser contactado”, RE corresponde a “Acepta responder la encuesta”, y CE corresponde a “Completa la respuesta de la encuesta”. Nótese que TF representa la probabilidad de tener disponible un teléfono de contacto, es decir, el alcance de la cobertura de la encuesta en la modalidad telefónica.¹¹

La tasa esperada de respuesta permitió determinar que el tamaño de la muestra sobredimensionada para Vivos era de:

$$N_{Re-Entrevista} = \frac{\text{observaciones objetivo}}{\text{tasa de respuesta}} = \frac{1.725 \text{ obs}}{31,2\%} = 5.529 \text{ observaciones de Vivos}$$

Es decir, para lograr adecuadamente el objetivo de 1.725 encuestas de Vivos en estas condiciones, se requería una muestra nominal de 5.529 observaciones de entrevistados que respondieron este tipo de cuestionario en la etapa Presencial.

2.5.5. Selección de la muestra de Vivos

La muestra se seleccionó siguiendo los mismos criterios de estratificación por *clusters* geográficos y características de los entrevistados (condición de afiliado, edad, sexo) usados en el diseño original para mantener la representatividad y comparabilidad de la muestra. A esto se sumó la necesidad de utilizar la muestra como un panel respecto a los entrevistados que respondieron la encuesta Presencial para realizar análisis antes-después de la pandemia.

El algoritmo de selección muestral fue el siguiente:

1. Para cada *cluster* (~comuna) de cada sub-muestra,¹² se sumaron los factores de expansión base de los individuos de la sub-muestra particular incluidos dentro de los 7.763 casos disponibles¹³. Es decir, para la comuna c dentro de la sub-muestra

$$m \in \{\text{Afiliados 2002, Nuevos Afiliados 2004, No Afiliados 2004, Nuevos 2016}\}$$

y entrevistados i , se tiene:

¹¹ Como se explica más adelante, en el ajuste a los factores de expansión, cada una de estas probabilidades se estimó utilizando modelos de *propensity score* basados en regresiones logísticas (o, simplemente, modelos *logit*).

¹² Sub-muestras de la EPS VII Ronda: Afiliados 2002, Nuevos Afiliados 2004, No Afiliados 2004, Nacimientos 2015.

¹³ Estos casos corresponden a los entrevistados Vivos de la encuesta Presencial para quienes se disponía de un número de teléfono de contacto.

$$FE_{m,c} = \sum_{i=1}^{I_{m,c}} FE_{m,c,i}$$

2. Luego, a nivel comunal para cada sub-muestra, se calculó el peso original de la comuna en la sub-muestra, $FE_{m,c}^{original}$, versus la suma de los factores de expansión, $FE_{m,c}$. Esta proporción es mayor a 1,¹⁴ y actúa como un factor de ajuste (expansión) para los factores de expansión originales de las observaciones disponibles. Por lo tanto,

$$FC_{m,c} = \frac{FE_{m,c}^{original}}{FE_{m,c}} > 1$$

3. Para cada sub-muestra, se calcularon los factores de expansión de cada observación disponible en la comuna ajustados por el factor comunal encontrado en el paso anterior, $FC_{m,c}$. Dado que el problema de cobertura telefónica dejó algunas comunas fuera de la muestra, se realizó un ajuste provisorio de factores de expansión por cobertura telefónica para el muestreo. En la **Tabla 7** se muestra la cantidad de comunas en esta situación.

Esto implica que, si se suman los factores de expansión de cada comuna (o de toda la sub-muestra), se obtiene un tamaño menor a la población representada. Para lograr que las observaciones con cobertura telefónica representaran a la población correspondiente a cada sub-muestra, se realizó un ajuste provisorio manteniendo las proporciones por tamaño de estrato efectivamente representado.¹⁵

Tabla 7. Clusters por cada sub-muestra representada en la muestra final después de considerar cobertura telefónica – Muestra de Vivos para Re-entrevista

	Afiliados 2002	Nuevos Afiliados 2004	No Afiliados 2004	Nuevos 2015
Clusters con observaciones	109	84	105	106
Clusters totales	113	113	113	114
Clusters sin observaciones	4	29	8	8

Fuente: elaboración propia.

¹⁴ Esto es así porque en ninguna comuna se logró la meta establecida en los objetivos de levantamiento de la encuesta Presencial debido a la interrupción del estudio.

¹⁵ En el tratamiento de factores de expansión definitivo, se realizó el correspondiente ajuste por cobertura en base a la estimación de la probabilidad de tener algún teléfono de contacto disponible. Por su parte, el ajuste provisorio consistió en tomar la proporción representada por cada observación y expandirla proporcionalmente al 100% del tamaño poblacional representado en cada estrato. Esta estrategia es similar a la que se ha usado en la EPS en olas anteriores para ajustar la disminución del tamaño del panel debido a fallecimientos. Si bien permite mantener las proporciones, tiene alguna deficiencia en términos de información que no se incorpora en el entendimiento del fenómeno de cobertura. Dado que, en general, hay acuerdo en usar la mayor cantidad de información posible para realizar los ajustes de factores de expansión final (y esto se realiza, según se explica, más adelante), no tiene mayores implicancias en términos del ajuste necesario para el muestreo.

4. A los factores del paso anterior, se los ajustó por el tamaño de la muestra relativo al marco muestral disponible, es decir, se calculó la proporción en que debían amplificarse los factores de expansión de cada observación debido a que el tamaño muestral seleccionado es de 5.529 encuestas:

$$PF = \frac{7,764 \text{ casos disponibles}}{5,529 \text{ observaciones objetivo}} = 1.404232$$

Esta proporción, PF , se aplica a cada uno de los factores de expansión de la muestra.¹⁶ El cálculo sigue la forma:

$$\widehat{FE}_{m,c,l} = \widehat{FE}_{m,c,l} \cdot PF \quad \forall m, c, l$$

5. A continuación, se realizó un muestreo aleatorio simple al interior de cada grupo de observaciones formado por el cruce de sub-muestra de origen y estrato (*cluster*). El procedimiento que se llevó a cabo fue el siguiente:

- Se generó un número aleatorio uniforme para todos los casos reunidos en el marco muestral (es decir, las 7.764 observaciones).¹⁷
- Se ordenaron las observaciones según sub-muestra de origen, estrato (*cluster*) y número aleatorio.
- Dentro de cada grupo formado por el cruce de la sub-muestra de origen y el estrato, se acumularon los factores de expansión que se obtuvieron anteriormente con las correcciones de los factores base, $\widehat{FE}_{m,c,l}$.
- Se seleccionó la muestra aleatoriamente dentro de cada celda conformada por la sub-muestra de origen y el estrato, de tal forma que los factores de expansión base ajustados representaran, al menos, la población de esa comuna en esa sub-muestra, es decir, que cumplieran la siguiente condición:

$$\sum_{i=1}^{I_{m,c}} \widehat{FE}_{m,c,l} \geq FE_{m,c}^{original} \quad \forall m, c$$

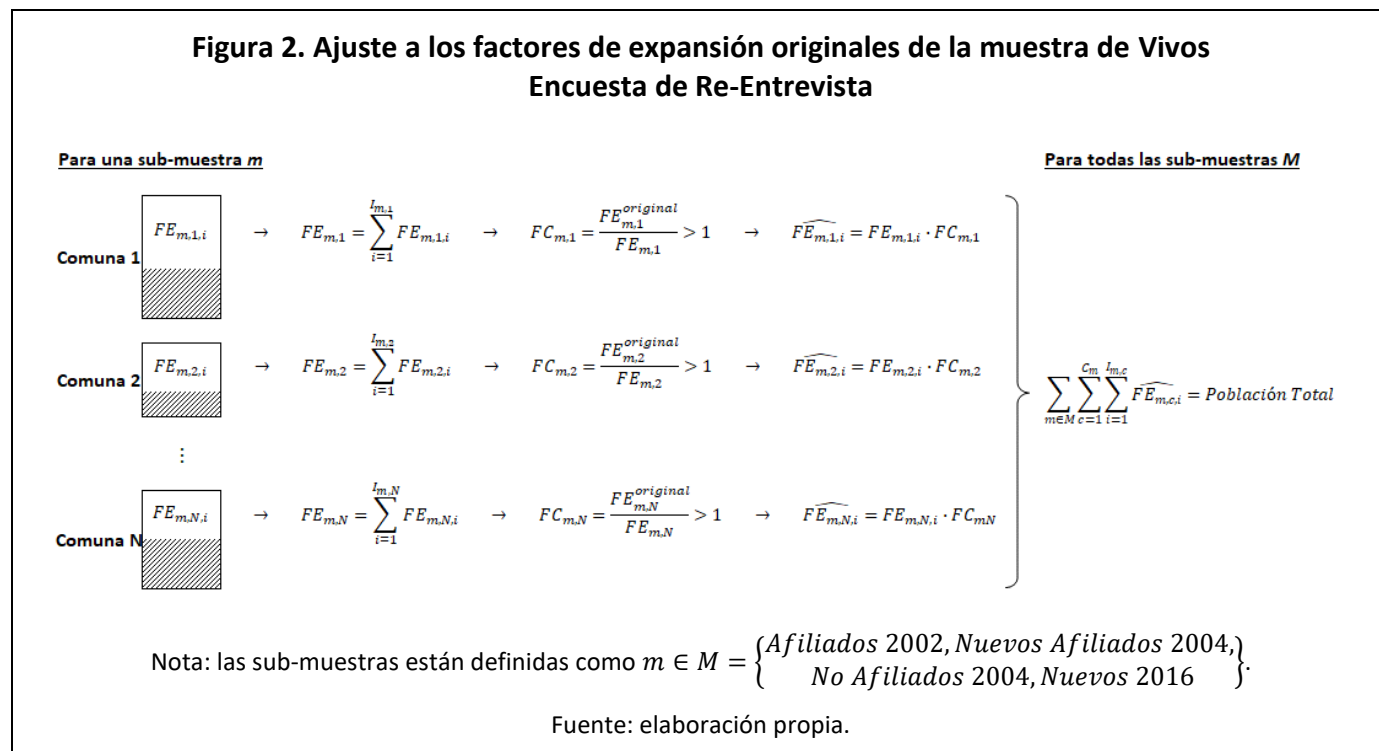
- Finalmente, al conjunto de observaciones seleccionadas se las ordenó por el número aleatorio asignado, de tal manera que ese fuera el orden en que se abordaría el contacto

¹⁶ La proporción ha sido redondeada al sexto decimal. En doble precisión (redondeo al décimo quinto decimal), el número obtenido es 1.404232230059685. Este número no se recalcula en el software (es decir, se lo copia literalmente en el programa para realizar los cálculos) para evitar que posibles cambios en la capacidad de operar con números de punto flotante de los computadores donde se pudiera emplear el software de creación de las muestras, generen resultados diferentes. En el muestreo original se había incluido una observación no elegible que provenía de la muestra de la EPS VI Ronda. Esta observación fue eliminada posteriormente al constatar su no elegibilidad. Por ello, en los cálculos de muestreo, esa observación apareció y generó la diferencia entre 7.764 (usado en los cálculos) y 7.763 observaciones efectivas de la encuesta.

¹⁷ Ver nota al pie anterior.

con la muestra. Esta estrategia fue pensada para permitir avanzar de manera más o menos balanceada a través de diferentes características de la muestra resultante (por ejemplo, sub-muestra de origen, sexo, grupos etarios, condición de afiliado, estrato, etc.).

Con este procedimiento, aplicado a cada comuna de cada sub-muestra, se obtuvo un conjunto de 5.569 observaciones que se “recorrieron” en su totalidad.¹⁸ Estos ajustes a los factores de expansión base se representan, esquemáticamente, en la **Figura 2**.



2.5.6. Selección de la muestra de Impedidos

Dado que la muestra de entrevistados para los que se respondió un cuestionario de “impedidos” en la encuesta presencial fue incluido forzosamente, no se requirió selección. Sin embargo, para asegurar un avance equilibrado de la muestra, se la ordenó a partir de un número aleatorio generado para cada unidad muestral.

¹⁸ Nótese que se requerían, al menos, 5.529 observaciones para asegurar el levantamiento comprometido para esta encuesta. El exceso de observaciones seleccionadas se debió a un efecto similar al “redondeo”: la suma de las ponderaciones de las observaciones seleccionadas no coincidía exactamente con la población comunal representada, por lo que se usó el criterio de que “al menos, debe representar a la comuna”, obteniéndose, en la mayoría de los casos, una representación en exceso. Esta situación no afectó la representatividad de la muestra, ya que se realizó en el ajuste de factores de expansión definitivo, un ajuste proporcional al tamaño de cada comuna (*cluster*).

2.5.7. Comportamiento de la muestra de Re-Entrevista

En la **Tabla 8** se muestra el comportamiento de esta muestra en el trabajo de campo. De los 8.137 casos disponibles de la encuesta Presencial (8.100 de Vivos y 337 de Impedidos), 8.100 tenían un número de teléfono de contacto registrado en las distintas fuentes de información empleadas, según se detalló anteriormente. Así, la muestra nominal ascendió a 5.906, incluyendo los 337 Impedidos más la selección de 5.569 Vivos. De estos casos, 5.874 fueron tratados en la encuesta (se intentó contacto). De ellos, 2.659 fueron efectivamente contactados, accediendo a responder la encuesta un total de 2.251. De estos, 26 abandonaron la encuesta antes de finalizarla, por lo que éstas fueron descartadas. El resultado final fue el logro de 2.225 encuestas, de la cuales se 136 se aplicaron a Impedidos, 7 a Fallecidos, y las 2.082 restantes, a Vivos.

Tabla 8. Encuesta de Re-Entrevista: distribución de la muestra nominal, contactados y respuesta

Tipo de Muestra	Afiliados 2002	Nuevos Afiliados 2004	No Afiliados 2004	Nuevos 2015	Total
Muestra Total	4.750	247	869	2.271	8.137
Muestra con Cobertura Telefónica	4718	247	864	2271	8.100
Muestra Nominal	3.442	197	668	1.599	5.906
Muestra Efectiva	3.419	197	660	1.598	5.874
Contactado	1.575	94	308	682	2.659
Responde Contactado	1.322	74	266	589	2.251
Responde Completa Responde, Contactado	1.306	74	263	582	2.225
Población representada originalmente	8.074.667	527.914	3.994.879	3.250.561	15.848.021
Factor expansión base promedio	6596,950	7133,973	16439,831	5672,881	7496,699

Fuente: Elaboración propia.

2.6. Encuesta de Continuidad

2.6.1. Marco muestral y tamaño muestral

La encuesta de Continuidad de la EPS VII Ronda recabó información similar a la correspondiente a la encuesta presencial, aunque con un menor nivel de detalle, en especial en lo que se refiere a la historia laboral, y con la adaptación correspondiente del cuestionario presencial al formato telefónico. El cuestionario se reformateó para que las entrevistas tuvieran una duración promedio de 25 minutos, dependiendo de la complejidad de la información aportada por el entrevistado.

El objetivo de esta encuesta se fijó en 5.000 entrevistas a partir de todos los entrevistados que no respondieron la encuesta presencial. Esto implica que el marco muestral de la encuesta de Continuidad corresponde a todas las personas que no respondieron la encuesta Presencial, es decir, 16.227 observaciones disponibles.

La muestra incluye observaciones que fueron visitadas (“recorridas” con distintos tipos de resultados, según se observa en la 0) y otras a las que no se visitó debido a la interrupción del estudio en su modalidad presencial. Todas las observaciones visitadas en la etapa presencial quedaron registradas con algún CDF distinto de “Entrevista Completa”, mientras que el resto de las observaciones se consideraron simplemente “no recorridas”. La Tabla 9 muestra la situación de las 16.227 observaciones que conforman el marco muestral según ha sido definido por la contraparte.

Tabla 9. Disposición de casos según participación en la encuesta Presencial

Disposición final de casos (CDF)	Frecuencia	Porcentaje	Porcentaje Acumulado
Anuladas por supervisión	82	0.5	0.5
Concertó cita con entrevistado	262	1.6	2.1
Se rechazó entrevista	720	4.4	6.6
Entrevistado no disponible	505	3.1	9.7
Se impidió acceso a la vivienda	31	0.2	9.9
Vivienda ocupada sin moradores presente	1,158	7.1	17.0
Informante no ubicable o no puede atender	454	2.8	19.8
Muerte del informante	90	0.6	20.3
Problema de idioma	1	0.0	20.4
Otra razón (elegible)	51	0.3	20.7
Informante impedido Físico/Mental para contestar temporalmente	15	0.1	20.8
Informante impedido Físico/Mental para contestar permanentemente	16	0.1	20.9
Área peligrosa o de difícil acceso	33	0.2	21.1
No fue posible localizar la dirección	591	3.6	24.7
Cambio de domicilio conocido	146	0.9	25.6
Cambio de domicilio desconocido	1,151	7.1	32.7
Otra razón (elegibilidad desconocida)	28	0.2	32.9
Empresa, oficina de gobierno u otra organización	21	0.1	33.0
Instituciones (hospital, cárcel, asilo, etc.)	10	0.1	33.1
Dormitorio colectivo (militar, de trabajo, etc.)	1	0.0	33.1
Vivienda en demolición, incendiada, destruida, etc.	17	0.1	33.2
Vivienda particular desocupada	34	0.2	33.4
Vivienda de veraneo o de uso temporal	4	0.0	33.4
Otra razón no elegible (con comentario)	193	1.2	34.6
No recorridos	10,613	65.4	100.0
TOTAL	16,227	100.0	

Fuente: Elaboración propia.

2.6.2. Selección de la muestra

Para la encuesta de Continuidad, se considera que el marco muestral y la muestra seleccionada (residual del trabajo de campo de la encuesta Presencial) están dadas. Sin embargo, se determinó el tamaño de la sub-muestra que se debía recorrer para conseguir las 5.000 encuestas objeto de esta parte del estudio.

2.6.3. Tamaño de la muestra

Teniendo en cuenta las diversas vertientes de la literatura sobre encuestas sin reemplazo y, en especial, en encuestas de tipo panel, se tuvo en cuenta el problema de no respuesta que puede deberse a diversas razones: atrición, imposibilidad de contacto, rechazo, etc. A esto, se sumó el efecto del cambio de la modalidad en la realización de esta encuesta, al pasar a ser telefónica, una estrategia que ha demostrado tener tasas de respuesta bastante inferiores a las encuestas cara a cara. Para solucionar el problema, se analizó el comportamiento que podría tener la tasa de respuesta, para luego calcular un tamaño muestral sobredimensionado que se haga cargo de la no respuesta debido a las distintas situaciones.¹⁹

Al igual que en la encuesta de Re-Entrevista, para estimar la tasa de respuesta esperada, se tuvieron en cuenta la probabilidad de disponer de teléfono de contacto, la probabilidad de ser contactado telefónicamente, la probabilidad de responder la encuesta, aunque sea parcialmente, y la probabilidad de completar exitosamente la encuesta de Continuidad. Con esta información, se estimó la siguiente tasa de respuesta:

$$TR_{Continuidad} = \Pr(CT|NE) \cdot \Pr(RE|CT) \cdot \Pr(CE|RE) = 26,9\%$$

donde TR es la tasa de respuesta esperada para esta encuesta, CT corresponde a “Logro de contacto telefónico”, NE corresponde a “No entrevistado en la encuesta presencial que dispone de número telefónico para ser contactado”, RE corresponde a “Acepta responder la encuesta”, y CE corresponde a “Completa la respuesta de la encuesta”. Nótese que NE representa la probabilidad de tener disponible un teléfono de contacto, es decir, el alcance de la cobertura de la encuesta en la modalidad telefónica.

Por lo tanto, el tamaño de la muestra sobredimensionada necesaria para obtener los 5.000 casos esperados de la encuesta de Continuidad asciende a:

$$N_{Continuidad} = \frac{\text{observaciones objetivo}}{\text{tasa de respuesta esperada}} = \frac{5.000 \text{ obs}}{26,9\%} = 18.588 \text{ observaciones}$$

Es decir, para lograr adecuadamente el objetivo de 5.000 encuestas en estas condiciones, requerimos una muestra nominal de 18.588 observaciones. Sin embargo, el conjunto de entrevistados no recorridos sólo asciende a 16.227 observaciones, de las cuáles sólo el 91.8% poseía un número de teléfono para intentar un contacto, basado en las primeras fuentes de información, que denominamos Poblamiento 1 (ver sección 2.4).

Debido a esto, Ipsos implementó dos estrategias adicionales para mejorar la información de cobertura telefónica para la encuesta de Continuidad y asegurar el éxito del estudio:

¹⁹ Véase Dabalen et al. (2016), pp. 23. Nótese que la sugerencia fue realizar sobredimensionamiento de la muestra, prefiriendo la configuración de la tasa de respuesta bajo los escenarios algo más pesimistas, de tal manera que luego fuera sencillo ajustar el sobre-muestreo con los factores de expansión. La situación inversa (sucesivos re-muestreos para completar la muestra objetivo) genera pérdida de propiedades estadísticas a la muestra conseguida ya que se trata de sub-muestras obtenidas en sucesivas rondas de muestreo.

- Se hizo un segundo relevamiento de números de teléfono, contratado a una empresa externa que recolecta información de fuentes alternativas de datos. A esto se denominó Poblamiento 2.
- Se realizaron visitas en terreno, siguiendo la estrategia usada en otras encuestas, como la CASEN 2020, a aquellos entrevistados con los que no se lograba contacto telefónico (por ejemplo, teléfono equivocado, buzón de voz, bloqueo de teléfono) y otros que no tenían teléfono de contacto. De esta manera, se pudo conseguir un conjunto de número de teléfonos altamente fiables y con alta probabilidad de contacto. A esto se denominó Poblamiento 3.

Esto mejoró la información de contacto en dos aspectos: permitió contar con más números alternativos de contacto para los entrevistados que ya tenían esta información, pero con los que no había éxito, e incrementó la cobertura. En la **Tabla 10**, se muestra la disponibilidad de teléfonos de contacto después de estas acciones, llegando a una cobertura de 95,6% de la muestra disponible.

Tabla 10. Cobertura telefónica en encuesta de Continuidad

		Poblamiento 1	Poblamiento 2	Poblamiento 3	Total	Porcentaje
Afiliados 2002	Si	8.492	6.919	1.907	8.961	94,5%
	No	986	2.559	7.571	517	5,5%
Nuevos Afiliados 2004	Si	624	553	109	647	97,3%
	No	41	112	556	18	2,7%
No Afiliados 2004	Si	1.324	903	316	1.416	89,4%
	No	260	681	1.268	168	10,6%
Nuevos 2015	Si	4.461	2.869	1.221	4.482	99,6%
	No	39	1.631	3.279	18	0,4%
Total	Si	14.901	11.244	3.553	15.506	
	No	1.326	4.983	12.674	721	

Fuente: elaboración propia.

Dado que, aún con las múltiples fuentes de información sobre teléfonos, la cantidad de muestra disponibles era menor a la requerida, se propuso “recorrer” el conjunto total de unidades muestrales disponibles. Además, se implementaron medidas para fortalecer el resultado del trabajo de campo para mejorar las tasas de contacto, respuesta y respuesta completa de encuestas esperadas antes del trabajo de campo.

Una medida final, que sólo apuntaba a balancear el avance de la muestra en el trabajo de campo, fue aleatorizar el orden de abordaje de las encuestas, tal como se hizo en el caso Impedidos en la encuesta de Re-Entrevista. Dado que se debía recorrer la muestra completa, esto no generaba otro impacto que un avance balanceado que permitía un mejor monitoreo de la calidad de las respuestas en el levantamiento.

RECUADRO 1: Tasas de contacto, respuesta y respuesta completa en entrevistas telefónicas

Las estimaciones de la tasa de respuesta esperada que realizó Ipsos, en especial, las referidas a tasas de contacto, respuesta y respuesta completa esperadas, obedecen a múltiples casos reportados y estudiados en la literatura²⁰ y al juicio experto de los miembros del equipo de Ipsos, con amplia experiencia en este tipo de encuestas, tal como consta en la descripción de la experiencia que es parte de la propuesta técnica de este estudio.

En la siguiente tabla, se resumen algunas de las experiencias observadas en la aplicación de encuestas telefónicas con números de teléfono celular, que es el medio de contacto predominante en la muestra disponible para la VII Ronda de la EPS.

Encuesta	Entidad que la realiza / País	Fecha/ Periodo	Tamaño de la muestra	Encuestas completas logradas	Tasa de Contacto	Tasa de Respuesta	Tasa de Respuesta Incompleta	Tasa de Rechazo
Encuesta del Sistema de Vigilancia de Factores de Riesgo asociados a Enfermedades No Transmisibles (SIVFRENT) (*1)	SIVFRENT de la Comunidad de Madrid / España	Enero/1999- Mayo/2000	2456	1391	68.2%	83.0%	1.4%	15.6%
Asthma Call-back Survey (*2)	Centers for Disease Control and Prevention (CDC) / Estados Unidos	2012/2013	15275	5005	75.5%	43.4%		38.6%
New York Adult Tobacco Survey (*3)	NY Tobacco Control Program / Estados Unidos	Mayo/ Junio 2015	34959	2736	53.2%	14.7%		85.3%

²⁰ Véase, por ejemplo, Díaz de Rada (2019), AAPOR (2010, 2014), Dabalen et al. (2016), Kennedy y Hartig (2019), CEPAL (2020a, b), INE Chile (2020), entre otros.

Encuesta Experimental CEP (*4)	Centro de Estudios Públicos y DATAVOZ / Chile	Octubre/ Noviembre 2018	372	121	50.0%	11.0%	2.0%	23.0%
Encuesta Telefónica de Ocupación y Empleo (ETOE) (*5)	INEGI / México	Abril/2020	14294	9220	98.7%	65.3%		34.7%
Encuesta Telefónica de Ocupación y Empleo (ETOE) (*6)	INEGI / México	Mayo/ 2020	14294	8515	96.7%	61.6%		38.4%
Encuesta Telefónica de Ocupación y Empleo (ETOE) (*7)	INEGI / México	Junio/2020	14294	7611	96.4%	55.3%		44.7%
Encuesta Telefónica sobre Confianza del Consumidor (ETCO) (*8)	INEGI / México	Abril/2020	7000	1135	30.6%	53.0%		47.0%
Encuesta Telefónica sobre Confianza del Consumidor (ETCO) (*9)	INEGI / México	Mayo/ 2020	12109	2143	31.1%	57.0%		43.0%
Encuesta Telefónica sobre Confianza del Consumidor (ETCO) (*10)	INEGI / México	Junio/2020	15381	2605	29.2%	57.9%		42.1%
Encuesta Telefónica sobre Confianza	INEGI / México	Julio/2020	N/D	2796	N/D	62.8%		32.5%

**del Consumidor
(ETCO) (*11)**

Encuesta Telefónica sobre COVID-19 y Mercado Laboral (ECOVID-ML) (*12)	INEGI / México	Abril/2020	28619	5593	29.3%	66.7%	2.9%	29.1%
---	-------------------	------------	-------	------	-------	-------	------	-------

Referencias:

- (*1) Galán, I., F. Rodríguez-Artalejo y B. Zorrilla (2004), "Comparación entre encuestas telefónicas y encuestas «cara a cara» domiciliarias en la estimación de hábitos de salud y prácticas preventivas," Gaceta Sanitaria, Vol. 18, Suplemento 2.
- (*2) Qin, X., C. M. Bailey y H. S. Zahran (2019), "Comparison response patterns on landline and cell phone in a call back survey: effects of demographic characteristics and lag days," Survey Methods: Insights from the Field.
- (*3) Brown, E. M., L. T. Olson, M. C. Farrelly, J. M. Nonnemaker, H. Battles y J. Hampton (2018), "Comparing Response Rates, Costs, and Tobacco-Related Outcomes Across Phone, Mail, and Online Surveys," Survey Practice, Vol. 11, Issue 2.
- (*4) González, R., P. Valenzuela, J. Fábrega, C. Isla y G. Castillo (2019), "INFORME TÉCNICO: Explorando modos mixtos en la encuesta CEP," Centro de Estudios Públicos y DATAVOZ.
- (*5) INEGI (2020), "Nota técnica: Alcance metodológico de la Encuesta Telefónica de Ocupación y Empleo 2020 (ETOE)", Abril.
- (*6) INEGI (2020), "Nota técnica: Alcance metodológico de la Encuesta Telefónica de Ocupación y Empleo 2020 (ETOE)", Mayo.
- (*7) INEGI (2020), "Nota técnica: Alcance metodológico de la Encuesta Telefónica de Ocupación y Empleo 2020 (ETOE)", Junio.
- (*8) INEGI (2020), "Nota Técnica: Encuesta Telefónica sobre Confianza del Consumidor (ETCO)," Abril.
- (*9) INEGI (2020), "Nota Técnica: Encuesta Telefónica sobre Confianza del Consumidor (ETCO)," Mayo.
- (*10) INEGI (2020), "Nota Técnica: Encuesta Telefónica sobre Confianza del Consumidor (ETCO)," Junio.
- (*11) INEGI (2020), "Nota Técnica: Encuesta Telefónica sobre Confianza del Consumidor (ETCO)," Julio.
- (*12) INEGI (2020), "Complementariedad de la Encuesta Telefónica de Ocupación y Empleo (ETOE) y la Encuesta Telefónica sobre COVID-19 y Mercado Laboral (ECOVID-ML)," Nota Técnica.

Fuente: Elaboración propia.

2.6.4. Comportamiento de la muestra de Continuidad

En la **Tabla 11**, se muestra el comportamiento de esta muestra en el trabajo de campo. Su interpretación es similar a la **Tabla 3**. De los 16.227 casos disponibles para recorrer, 10.695 no habían recibido ningún tratamiento (intento de contacto) en la encuesta Presencial, 4.812 habían sido sujetos de, al menos, un intento de contacto, sin que fuera posible llegar a una conclusión respecto a la respuesta o rechazo de la encuesta Presencial (es decir, tenían un CDF distinto de entrevista completa), y 720 entrevistados habían rechazado la encuesta Presencial.

La cobertura telefónica lograda con los tres poblamientos llegó a los 15.506 casos, de los que se recorrieron efectivamente 15.209. En total, se logró contactar a 6.696 entrevistados, de los que accedieron a responder la encuesta 5.325, y la completaron exitosamente 5.296. De estas encuestas, 5.031 correspondieron a cuestionarios de Vivos, a los cuáles se les realizaron los ajustes de factores de expansión, según se detalla más adelante, para dejarlos disponibles para análisis estadístico. También se obtuvieron 103 cuestionarios completos acerca de entrevistados fallecidos y 162, de impedidos.

Tabla 11. Encuesta de Continuidad: Distribución de la muestra nominal, contactados y respuesta

Tipo de Muestra	Afiliados 2002	Nuevos Afiliados 2004	No Afiliados 2004	Nuevos 2015	Total
Muestra Nominal	9.478	665	1.584	4.500	16.227
No visitados en Presencial	6.168	442	1.127	2.958	10.695
No revisitados en Presencial	2.795	205	409	1.403	4.812
Rechazos en Presencial	515	18	48	139	720
Muestra con Cobertura Telefónica	8.961	647	1.416	4.482	15.506
Muestra Efectiva	8.730	636	1.379	4.464	15.209
Contactado	3.848	282	598	1.968	6.696
Responde Contactado	2.952	222	488	1.663	5.325
Responde Completa Responde, Contactado	2.932	221	486	1.657	5.296
Población representada originalmente	8.074.667	527.914	3.994.879	3.250.561	15.848.021
Factor de expansión base promedio	2753,979	2388,751	8219,916	1961,715	2992,451

Fuente: Elaboración propia.

3. Atrición en la EPS

De los 24.882 casos que podían ser abordados durante el estudio (potenciales entrevistados), se esperaba llegar a 18.000 encuestas efectivas atendiendo a problemas de atrición de diversa índole, y a un nivel de esfuerzo razonable en el trabajo de campo.

Al momento de interrumpirse el estudio, el campo había alcanzado a más de 14.187 entrevistados potenciales (direcciones recorridas), con 8.655 cuestionarios conseguidos (7.800 a entrevistados vivos, 516 a entrevistados fallecidos y 339 a entrevistados impedidos). Para ello, al momento de paralización del estudio, se habían realizado 20.680 visitas.

La siguiente tabla muestra la antigüedad de la participación, es decir, la última vez que participaron según el panel al que pertenecen:

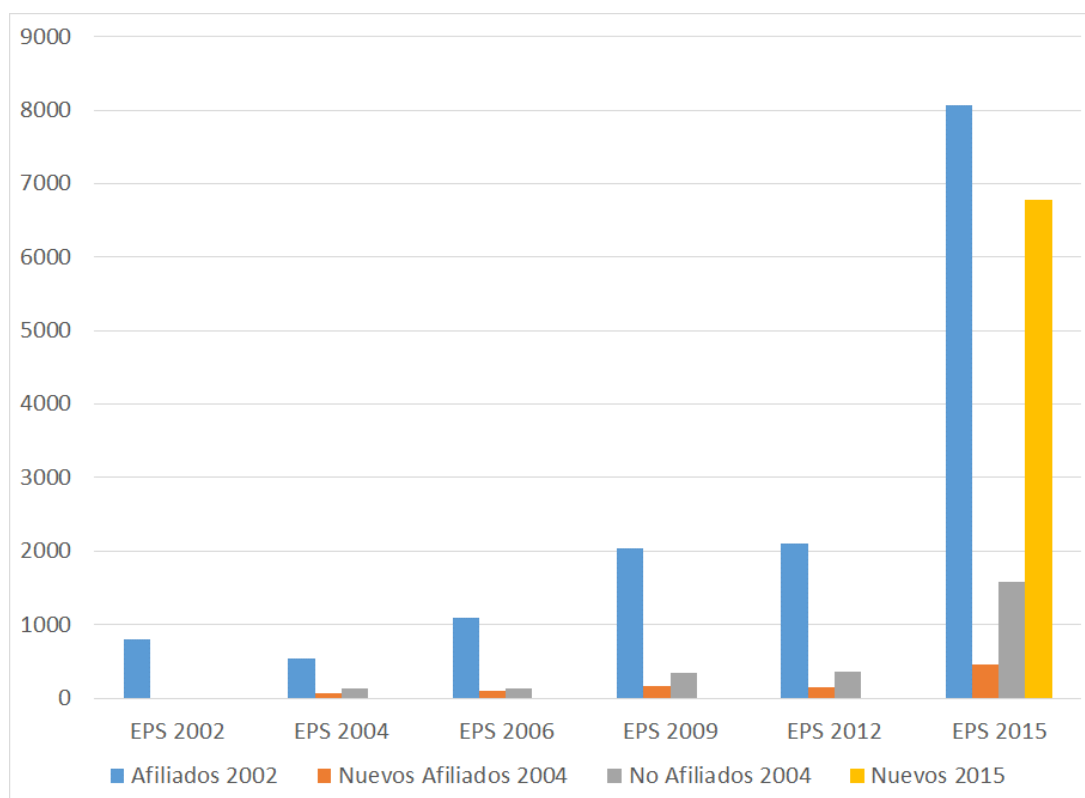
Tabla 12. Origen de las observaciones disponibles en la EPS VII Ronda

	Afiliados 2002	Nuevos Afiliados 2004	No Afiliados 2004	Nuevos 2015	TOTAL
EPS 2002	800				800
EPS 2004	533	61	125		719
EPS 2006	1,093	92	138		1,323
EPS 2009	2,041	159	351		2,551
EPS 2012	2,102	142	353		2,597
EPS 2015	8,063	464	1,587	6,778	16,892
TOTAL	14,632	918	2,554	6,778	24,882

Fuente: elaboración propia.

Para facilitar la comparación, se muestra en la **Figura 3**, cuál es la composición de la muestra de la EPS VII Ronda de 24.882 entrevistados de panel disponibles, dependiendo de la última vez que participaron en la EPS.

Figura 3. Composición de la muestra EPS VII Ronda según última participación en la encuesta



Fuente: elaboración propia.

Sin embargo, la extensión de la pandemia de coronavirus y su impacto en la salud del país, obligaron a terminar anticipadamente el trabajo de campo, como se explicó en la sección anterior.

Dado que estas dos encuestas se aplicaron sobre muestras mutuamente excluyentes, y a la vez, partieron de la misma muestra principal, conviene también revisar el comportamiento agregado de las dos muestras versus toda la distribución.

Por su parte, la encuesta de Re-Entrevista es una sub-muestra de la encuesta Presencial, heredando las características de esta última.

3.1. Razones de no respuesta

De acuerdo a Daniels y Hogan (2008), para entender el comportamiento de una muestra, es necesario diferenciar los distintos tipos de no respuesta que nos podemos encontrar. En particular, ellos diferencian, al menos, cinco tipos de pérdida de observaciones muestrales:

1. visitas perdidas, ya sea aleatoriamente o por razones relacionadas con la respuesta.

2. la salida de un participante del estudio ya sea que lo decida el mismo participante o el investigador que lleva a cabo el estudio.
3. pérdida del entrevistado por imposibilidad de su seguimiento posterior por razones desconocidas o no reportadas.
4. fallecimiento o impedimento que no le permita continuar participando del panel (no atribuibles al estudio, en este caso).
5. pérdida por diseño, en que sólo un subgrupo de individuos es seleccionado para su seguimiento en el tiempo.

Un caso particular de pérdida de información es cuando el entrevistados participa intermitentemente en el estudio longitudinal, generando brechas informativas en su historia de participación. Pero, ya sea que los entrevistados perdidos a lo largo del estudio longitudinal sean permanentes o intermitentes, hay varias estrategias para comprender los patrones de comportamiento y modelar las probabilidades de participación para realizar los ajustes correspondientes.²¹

Sin embargo, sin importar las razones particulares de no respuesta que se observen, ni el patrón de la no participación que muestre un individuo particular, para hablar de aleatoriedad de la muestra obtenida, debemos diferenciar tres casos largamente discutidos en la literatura (Wood et al., 2019):

1. Pérdida aleatoria (“*Missing at random*” o MAR): definida por Rubin (1976), este tipo de atrición implica que la pérdida de la observación no depende de variables no observadas, pero puede deberse (explicarse) por la información contenida en variables observadas. Por ejemplo, si los hombres son más o menos propensos a participar y continuar participando en el panel (atrición o no respuesta total), o si las mujeres son menos propensas a revelar su ingreso en la encuesta que los hombres (no respuesta parcial), podemos decir que el mecanismo de pérdida de información que observamos es MAR.
2. Pérdida totalmente aleatoria (“*Missing completely at random*” o MCAR): en este caso, la pérdida de información no está relacionada con información observable o no observable. En particular, las razones son enteramente ajenas a la encuesta en sí. Por lo tanto, la información faltante (total o parcial) es simplemente un subconjunto de la información total que sí se logró recolectar.
3. Pérdida no aleatoria (“*Missing not at random*” o MNAR): este tipo de pérdida de información es estudiada por Little y Rubin (2014), quienes aseguran que cuando la información no observada que

²¹ Véase, por ejemplo, Albert (2000) y Albert et al. (2002), que construyen modelos con una variable aleatoria multinomial especificada para cada momento del tiempo, en que clasifican el comportamiento como “todavía participando”, “pérdida intermitente” y “pérdida permanente”. También Lin et al. (2004) tratan el problema de intermitencia en la participación con el uso del enfoque de patrón mixto latente. También se puede modelar la no respuesta con el enfoque de pseudo-verosimilitud, como lo proponen Troxel et al. (1998), usando información de pérdida del mismo periodo condicional sólo en la no respuesta potencial para esa ola de la encuesta. Finalmente, Rotnitzky et al. (2001) realiza una revisión del tratamiento que se puede aplicar a no respuestas por causas múltiples.

tiene el entrevistado es el factor que determina su participación en la encuesta, entonces se debe modelar la pérdida de información (por ejemplo, se pueden utilizar imputaciones múltiples) para reemplazar esta información perdida y evitar sesgos en los resultados. Sin embargo, se debe tener conocimiento (o información) previa sobre la observación, algo que está disponible para todos los casos abordados en la EPS VII Ronda, al tratarse solamente de observaciones de panel (sin refresco).

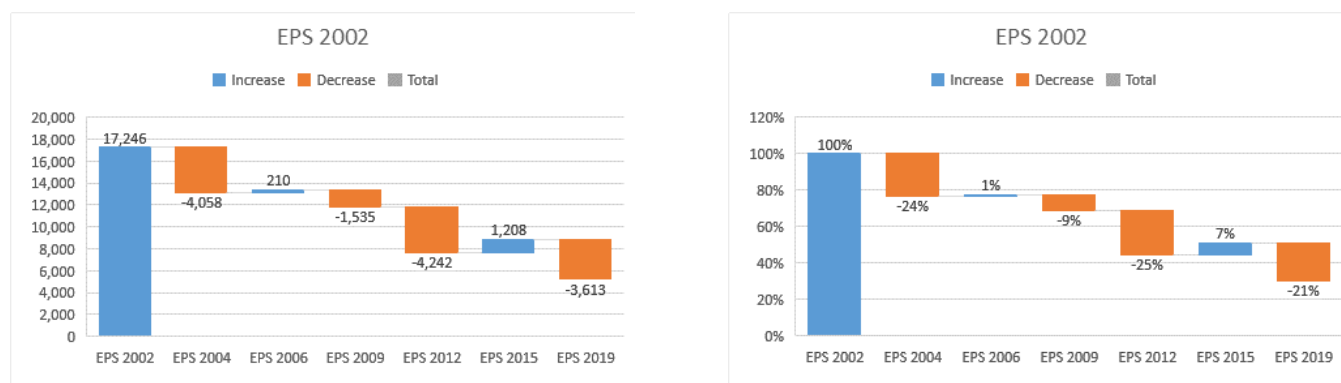
Cada una de estas formas de pérdida de información requieren un tratamiento. En particular, es posible que aun con tratamiento, en el caso de MNAR, no sea posible controlar los sesgos que se pueden presentar por la autoselección. En general, se espera que las pérdidas sean de tipo MAR y MCAR.

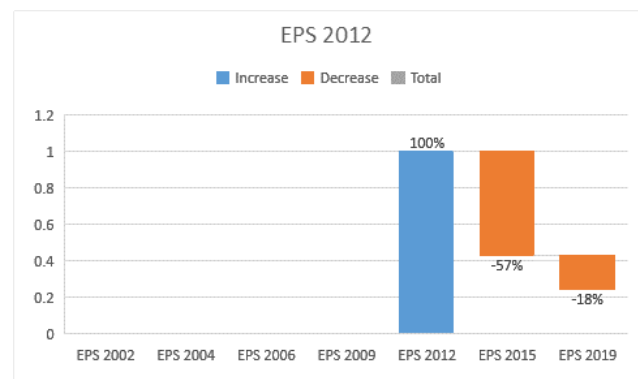
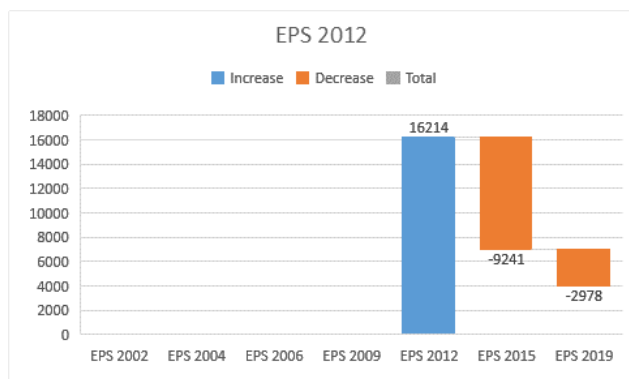
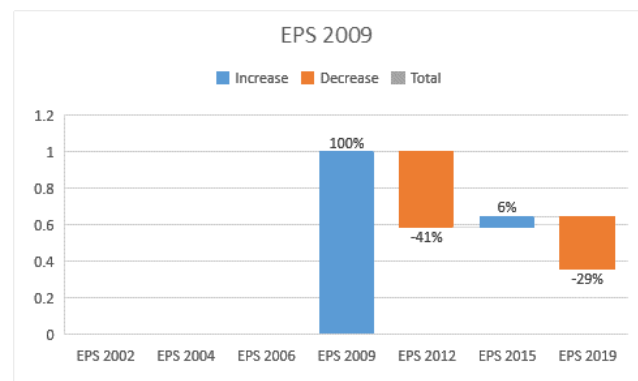
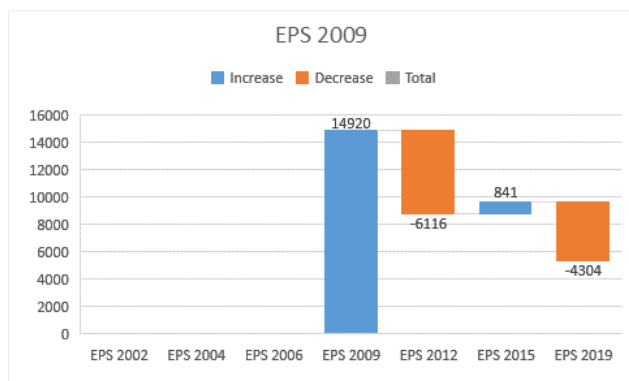
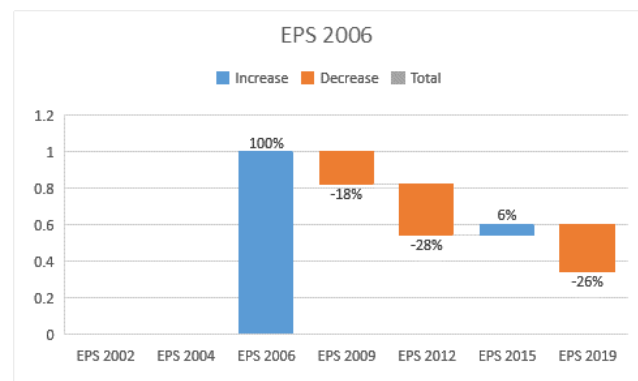
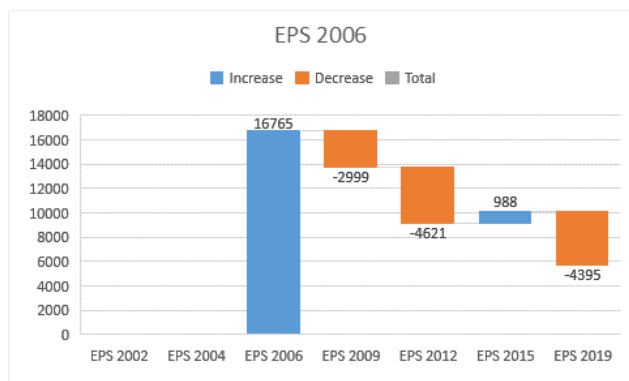
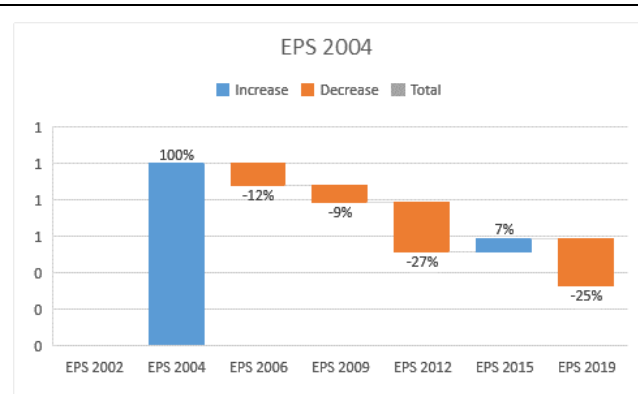
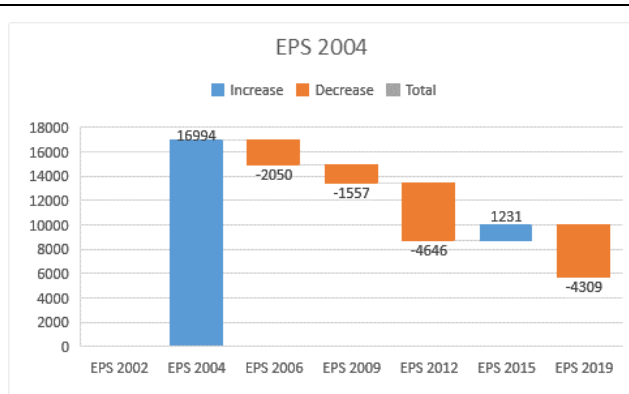
A continuación, se analiza el comportamiento de las muestras según un conjunto de variables que suelen determinar el comportamiento de autoselección de los participantes, y que se emplean en los modelos para ajuste de no respuesta.

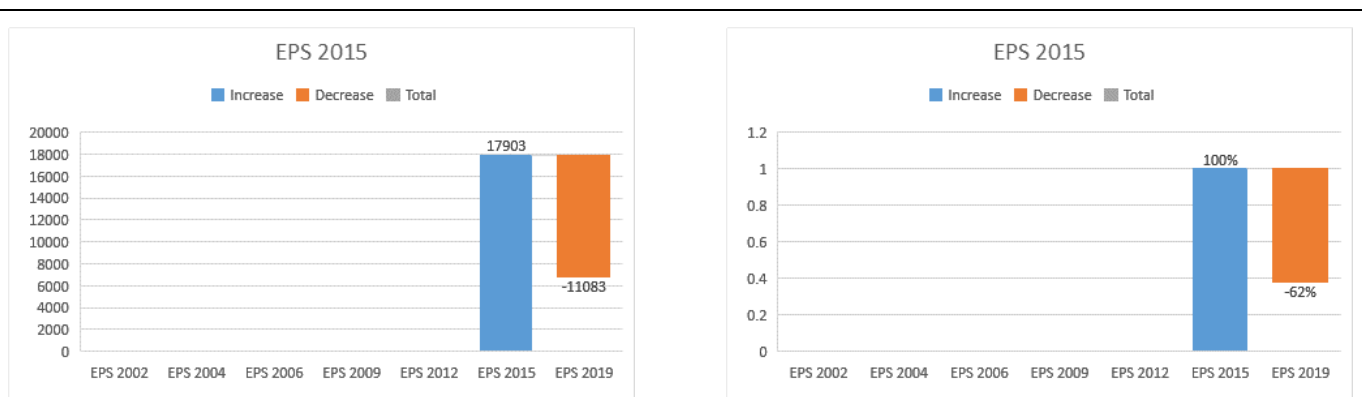
3.2. Atrición por olas

La atrición histórica que ha tenido la encuesta se muestra en la **Figura 4**. Como puede observarse en las sucesivas olas, el comportamiento de atrición en la EPS VII Ronda no se muestra especialmente distinto a las ediciones anteriores. La EPS VII Ronda muestra niveles de atrición de entre 20% y 30% en las olas anteriores a 2012, con un repunte en esta ola por tamaño muestral menor (en realidad, en 2015 se observó la pérdida importante de observaciones de la encuesta 2012). En el caso de la atrición entre la EPS VI Ronda y la EPS VII Ronda, el impacto mayor se debe a que la atrición de la ola 2015 de la EPS fue cubierta por observaciones de refresco, obedeciendo a decisiones de diseño.

Figura 4. Atrición por olas en la encuesta Presencial, en unidades observacionales y porcentajes





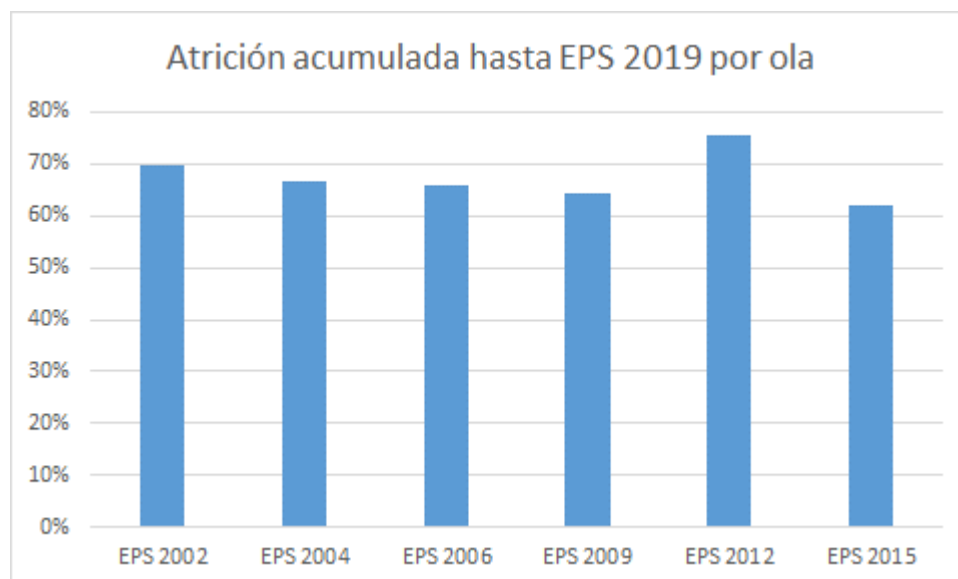


Nota: para cada una de las olas anteriores a las EPS VII Ronda, la atrición ha sido calculada sobre el total de casos efectivamente logrados, lo que implica que los niveles de atrición están subestimados. Por el momento, no se cuenta con información sobre cuáles fueron los objetivos de levantamiento en cada una de las olas anteriores para los paneles.

Fuente: elaboración propia.

Si se revisa la acumulación de atrición por ola previo a la EPS VII Ronda, se observa que los niveles de atrición son similares entre olas, y que la composición de la muestra recolectada no tiene sub/sobre-representación de los participantes de ninguna ola en particular, tal como surge de la **Figura 5**.²²

Figura 5. Atrición acumulada por ola hasta la EPS VII Ronda.



Fuente: elaboración propia.

3.3. Comportamiento de variables seleccionadas a través de las olas

A continuación, ampliamos el análisis para observar si la atrición observada en la EPS VII Ronda obedeció a algún patrón sistemático (MNAR). referencia de la encuesta, que han sido observadas a través de las olas y que se reportan habitualmente.

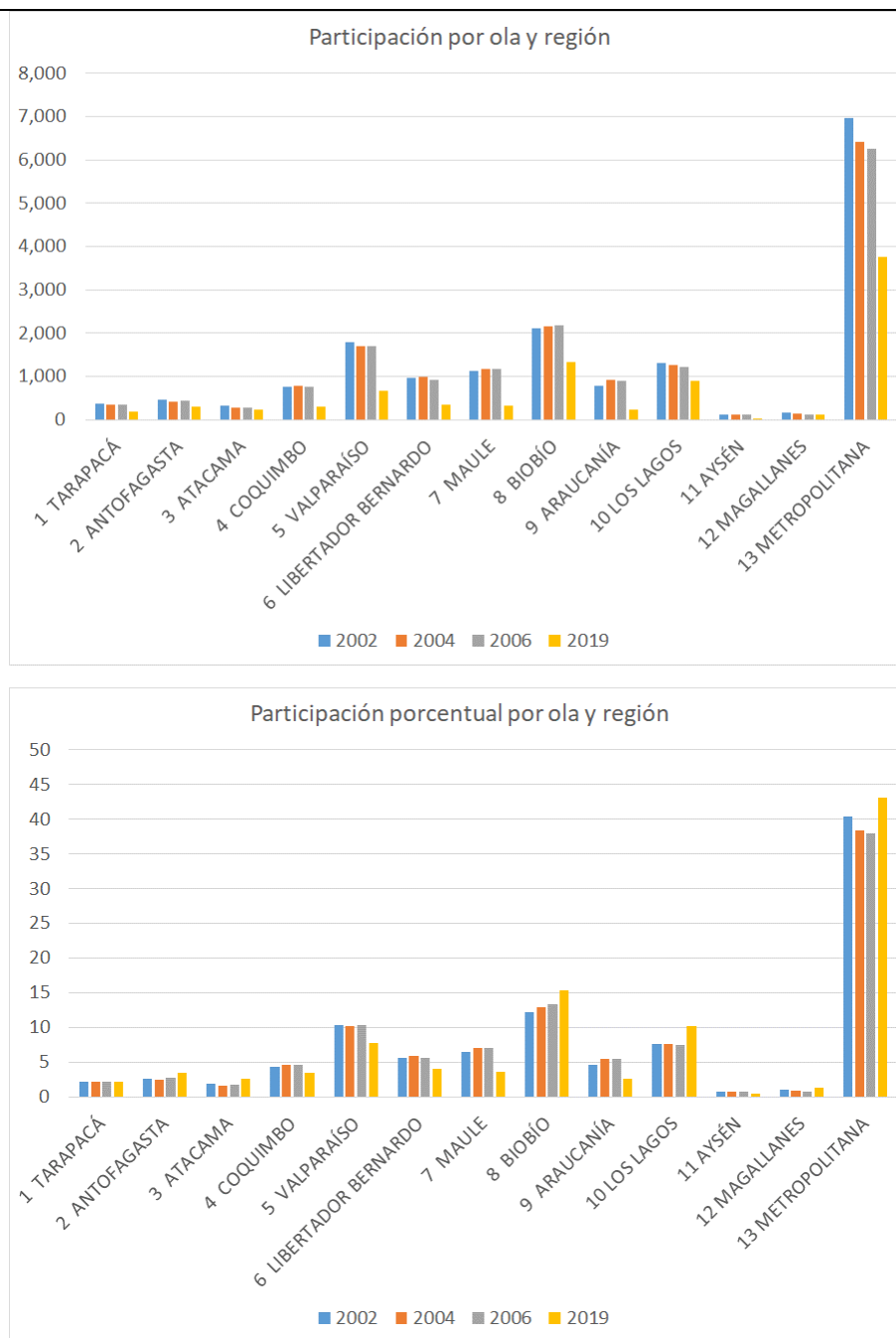
Se incluyen en este conjunto de variables el sexo y edad del entrevistado (también representado a través de tramos etarios), región en la que habita, la calidad de afiliado al sistema previsional y el estado civil.

Las comparaciones que se realizan a continuación se abstraen de la representación a nivel país del fenómeno, ya que deseamos saber cuál es el desempeño de la muestra como tal, y no los agregados a nivel país que se obtendrían aplicando factores de expansión.

Si bien la encuesta no ha sido diseñada para obtener inferencia a nivel de región, las características del mercado laboral y la alfabetización financiera varían de manera importante entre regiones, y es bueno que la encuesta cuente con una configuración regional adecuada o, al menos, comparable con versiones anteriores de la misma. En la **Figura 6**, se muestra la distribución regional de casos levantados en cada encuesta en términos porcentuales respecto al total de encuestas recabadas en la respectiva ola.

Como se puede observar, en términos absolutos, dado el menor tamaño de la encuesta de 2019 (13.951 observaciones totales), todas las regiones poseen menos casos que los reportados en olas anteriores. La más afectada por esta merma en la información es la Región Metropolitana. Sin embargo, cuando revisamos el desempeño relativo de las distintas muestras en términos porcentuales, en el segundo panel de la **Figura 6**, observamos que tenemos dos situaciones: regiones en que el estudio ha recogido relativamente más observaciones que en ocasiones anteriores, como en las regiones de Antofagasta, Atacama, Biobío, Los Lagos, Magallanes y RM.

Figura 6. Distribución de respuestas por región y por ola



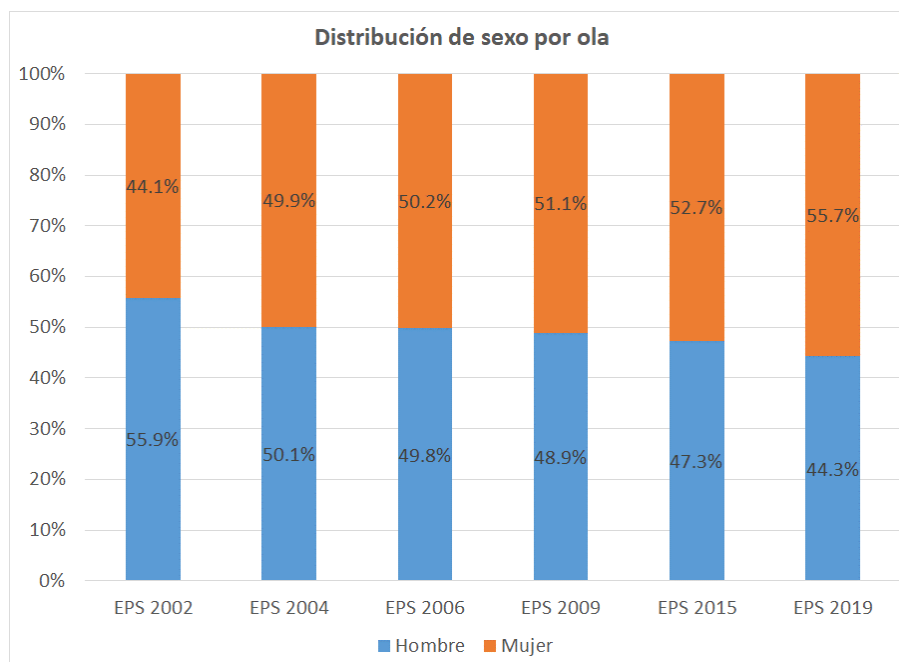
Nota: el reporte incluye las olas 2002, 2004, 2006 y 2019.

Fuente: elaboración propia.

La segunda variable que observamos es la composición de género entre olas de la encuesta, como se muestra en la **Figura 7**. Con el avance de las olas de la Encuesta, se observa un desplazamiento de la composición de la muestra de los hombres hacia las mujeres. Sin embargo, este cambio es paulatino y no aparece como un comportamiento anómalo particular de la EPS VII Ronda. De hecho, desde la EPS II Ronda,

ola a ola, la composición ha ido cambiando en favor de la mujer aproximadamente un 2% por ola. Esto también se cumple para la EPS VII Ronda respecto a la EPS VI Ronda, en que ha llegado a un 3%.

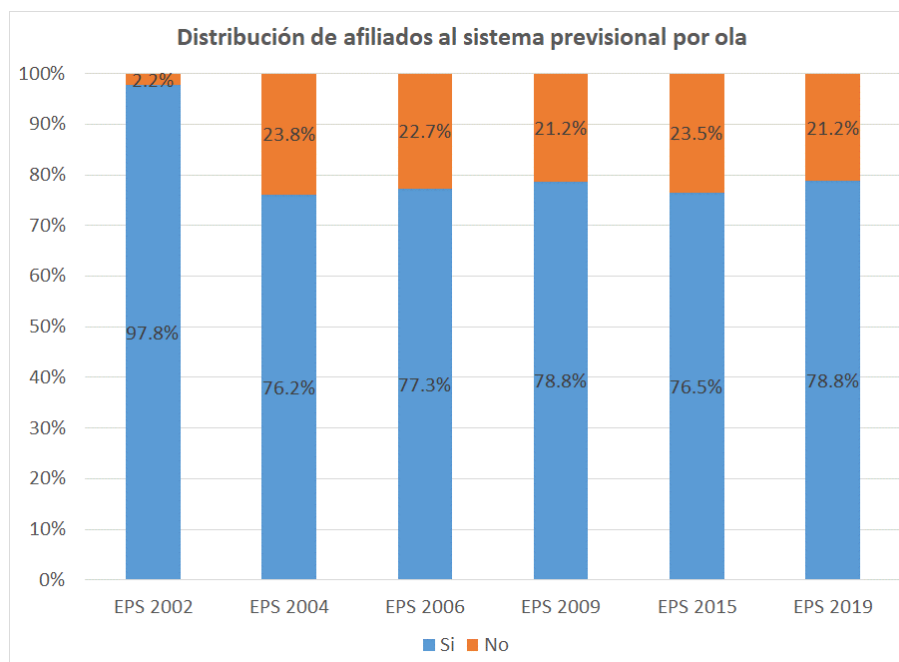
Figura 7. Distribución de sexo de entrevistados entre olas



Fuente: elaboración propia.

La siguiente variable de interés es la distribución de los afiliados al sistema de pensiones. Esta es una variable relevante, ya que la encuesta fue originalmente pensada para estudiar el fenómeno previsional en Chile, y la condición de aportante al sistema previsional era un determinante de participación en la primera ola. En ese sentido, se puede observar que, en 2002, los participantes declaran estar afiliado al sistema previsional casi en su totalidad, mientras que, a partir de 2004, el porcentaje de participantes (afiliados) al sistema previsional según el auto-reporte, ronda el 77%. Esta cifra también es cercana a la observada en la EPS VII Ronda. Por lo tanto, no se observa una sub- o sobre-representación de los afiliados al sistema previsional para la ola en discusión. Véase la **Figura 8** para una comparación gráfica.

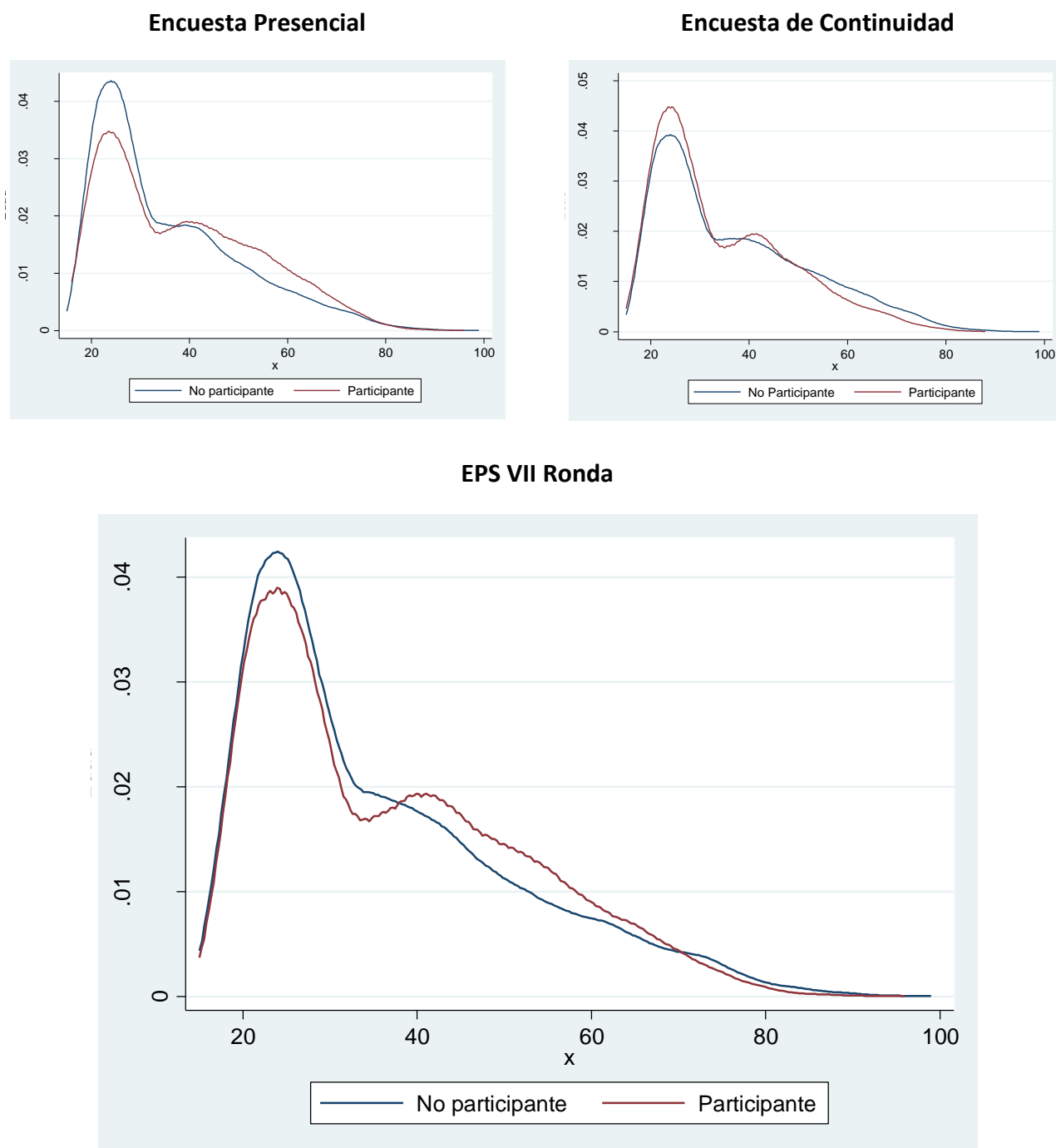
Figura 8. Distribución de afiliados al sistema de pensiones por olas



Fuente: elaboración propia.

La cuarta variable analizada corresponde a la edad. Esta variable es clave para entender, junto al sexo, cómo se comporta demográficamente la muestra. Los perfiles de edad según la participación o no en cada encuesta, y el perfil para la suma de ambas, se muestra en la **Figura 9**. Los perfiles no difieren de forma aparente entre participantes y no participantes. De hecho, mientras los más jóvenes estaban más representados en la encuesta Presencial entre los no participantes, su participación subió en la encuesta de Continuidad, compensando en el agregado, la participación por edades. En el agregado, se evidencia una menor participación de jóvenes, en favor de participantes de entre 40 y 55 años de edad.

Figura 9. Distribución edad según participación en encuestas de EPS VII Ronda



Fuente: elaboración propia.

Una visión alternativa, es el comportamiento de los grupos etarios que habitualmente son utilizados en el análisis para evitar las imperfecciones propias del reporte de la edad y para generar masa crítica en los distintos tramos.

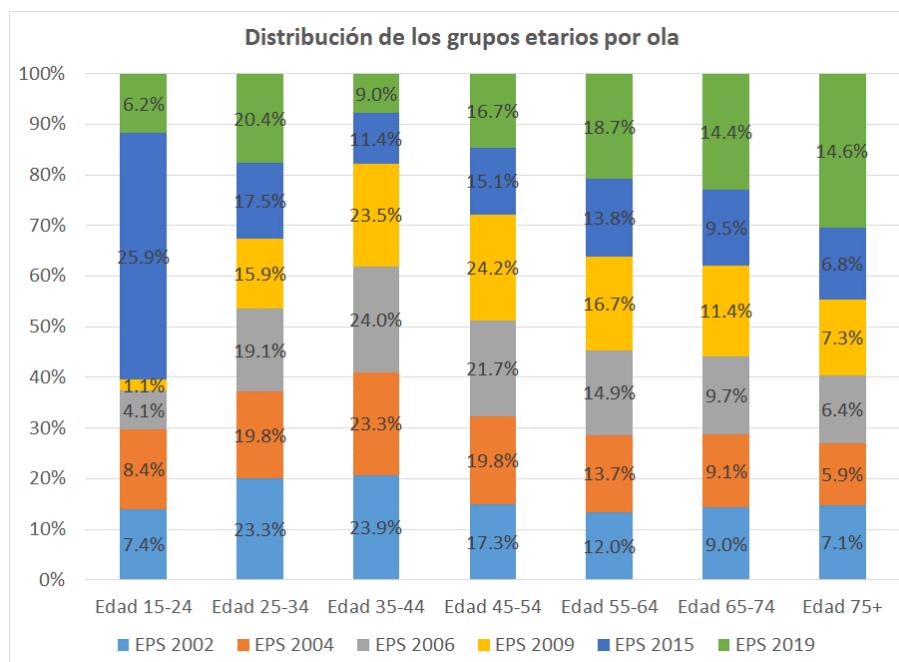
Así, en la **Figura 10**, podemos observar que tenemos una menor concentración de casos en el grupo de 35 a 44 años de edad, que tiene una doble explicación: por un lado, es parte del efecto de envejecimiento del panel (que hoy ya tiene 18 años de antigüedad), por lo que esa escasez se asocia al grupo de jóvenes que comenzó en el panel en las primeras olas. Esta explicación también puede apoyarse en el crecimiento paulatino de los otros tramos etarios para la EPS VII Ronda.

La segunda explicación viene por el lado de la introducción del refresco en 2015 que produjo un quiebre en el perfil etario de la encuesta, generando una bimodalidad de la distribución de edades. Y eso se ve reflejado en un grupo de jóvenes adultos más acotado, flanqueado por jóvenes y adultos maduros.

Sin embargo, este comportamiento parece ser característico de la distribución actual del panel si es que tenemos en cuenta el comportamiento de la muestra en 2015, cuando se produce la gran contracción en este tramo etario que alimenta al mismo grupo en 2019. Nótese al respecto, la caída desde 23,5% para el grupo etario en 2009, a 11,4% en 2015, ola en la que también debemos sumar el efecto de la fallida incorporación de refresco joven en la EPS V Ronda (no representada en este gráfico).

Si bien este comportamiento no lleva a pensar en la edad como un mecanismo importante de autoselección, posiciona a la variable como un desafío para la continuación del estudio, y que se nota un “extrangulamiento” de la muestra entre el grupo de jóvenes y los adultos maduros.

Figura 10. Grupos etarios por ola

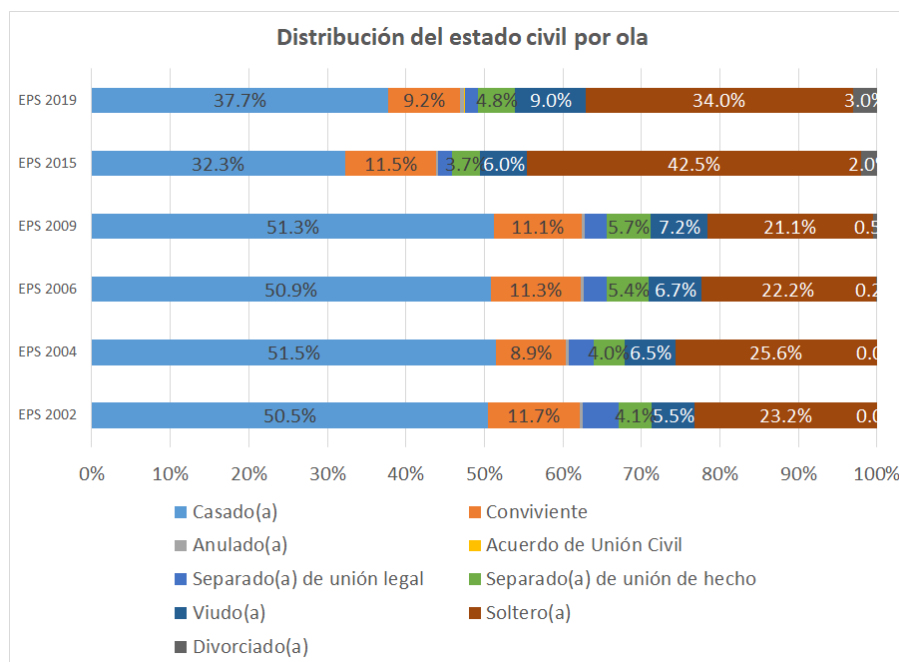


Fuente: elaboración propia.

Finalmente, analizamos el comportamiento de otra variable socio-demográfica como es el estado civil. Esta variable se supone no influenciada por la crisis sanitaria, aun cuando estudios realizados a personas en situación de aislamiento han arrojado resultados que van, en general, en contra de la continuidad de la convivencia.

Los resultados son similares a olas anteriores, aunque vale la pena resaltar la variación en los casos de personas que declaran estar casadas y solteras. Dada la introducción del refresco en el 2015, el panel sufrió un brusco cambio en la composición de estas dos categorías por la entrada de jóvenes entre los que domina la soltería. Pero la EPS VII Ronda comienza a evidenciar el paso del tiempo en el comportamiento del estado civil, incorporando más personas a la relación de pareja en detrimento de la soltería. Este es un comportamiento esperable que habla de la buena representatividad que tiene la muestra para medir este fenómeno observado en otras encuestas, y también a nivel censal. Véase la **Figura 11**.

Figura 11. Distribución del estado civil entre olas



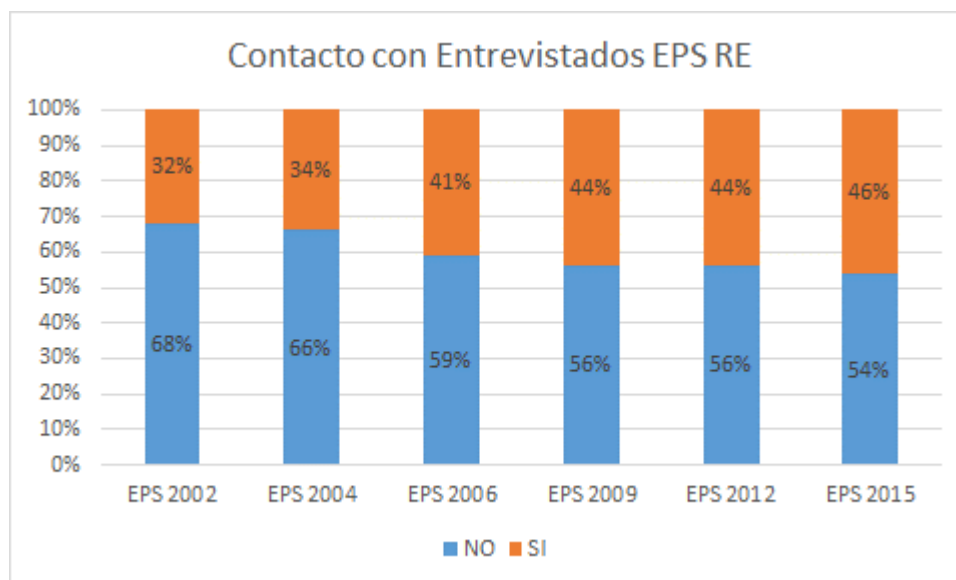
Fuente: elaboración propia.

3.4. Atrición en la encuesta de Re-Entrevista

La encuesta de Re-Entrevista se realizó en condiciones muy diferentes a la encuesta Presencial. En esta encuesta, el principal desafío consistió en contactar a los entrevistados ya que se cambió la forma de comunicación. A favor de la encuesta jugó el hecho de que los entrevistados habían participado en la encuesta Presencial 6 meses antes.

La siguiente figura muestra las tasas de contacto por ola en la encuesta de Re-Entrevista:

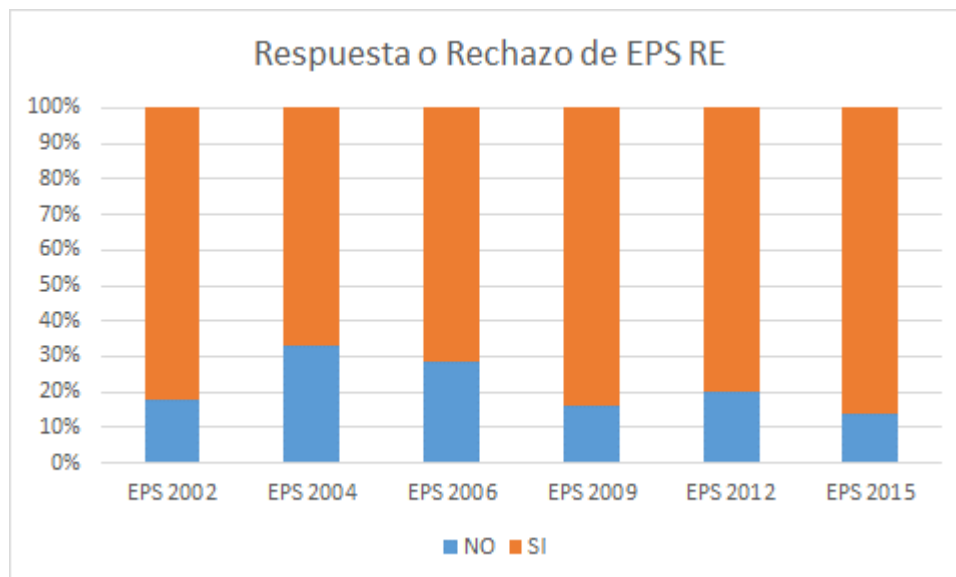
Figura 12. Contacto por ola en encuesta de Re-Entrevista



Fuente: elaboración propia.

Como puede observarse, hay una disminución constante de la atrición al tratarse de olas más jóvenes. A continuación, está representada la tasa de respuesta/rechazo al estudio en la **Figura 13**.

Figura 13. Tasas de respuesta y rechazo a la encuesta de Re-Entrevista

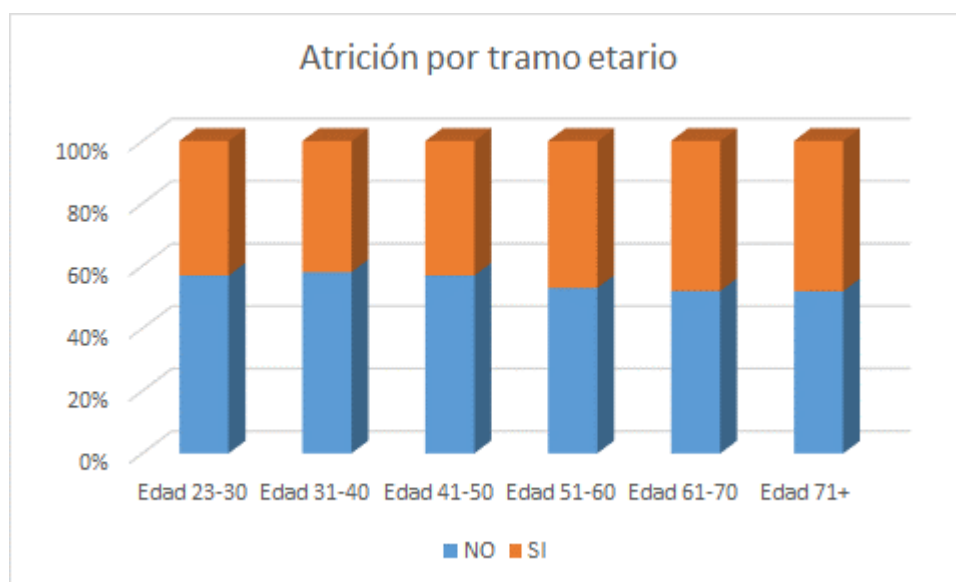


Fuente: elaboración propia.

Como sucedía con la encuesta Presencial, las personas que comenzaron su historia en 2004 se han mostrado más reacias a continuar involucradas en el estudio.

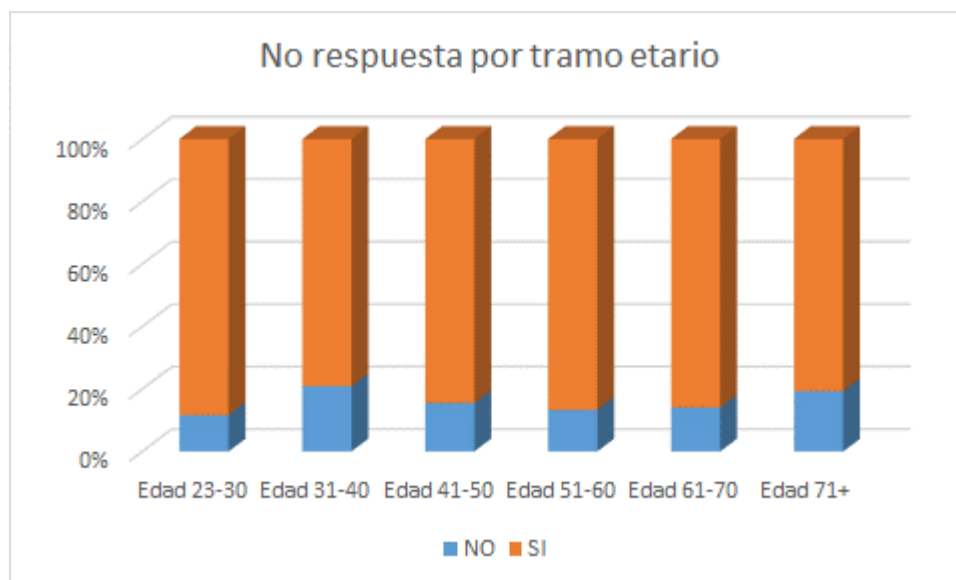
A continuación, se muestran la atrición y no respuesta por tramo etario:

Figura 14. Atrición por tramo etario en la encuesta de Re-Entrevista



Fuente: elaboración propia.

Figura 15. Situación de no respuesta por tramo etario en la encuesta de Re-Entrevista



Fuente: elaboración propia.

Contrario a lo que se suele suponer, los más jóvenes se mostraron más dispuestos a responder una vez contactados. Sin embargo, domina entre ellos la dificultad de lograr contacto, al menos en los grupos hasta 40 años.

Respecto a la atrición y no respuesta por sexo, los hombres parecen ser los más reticentes a participar, según surge de la **Figura 16**.

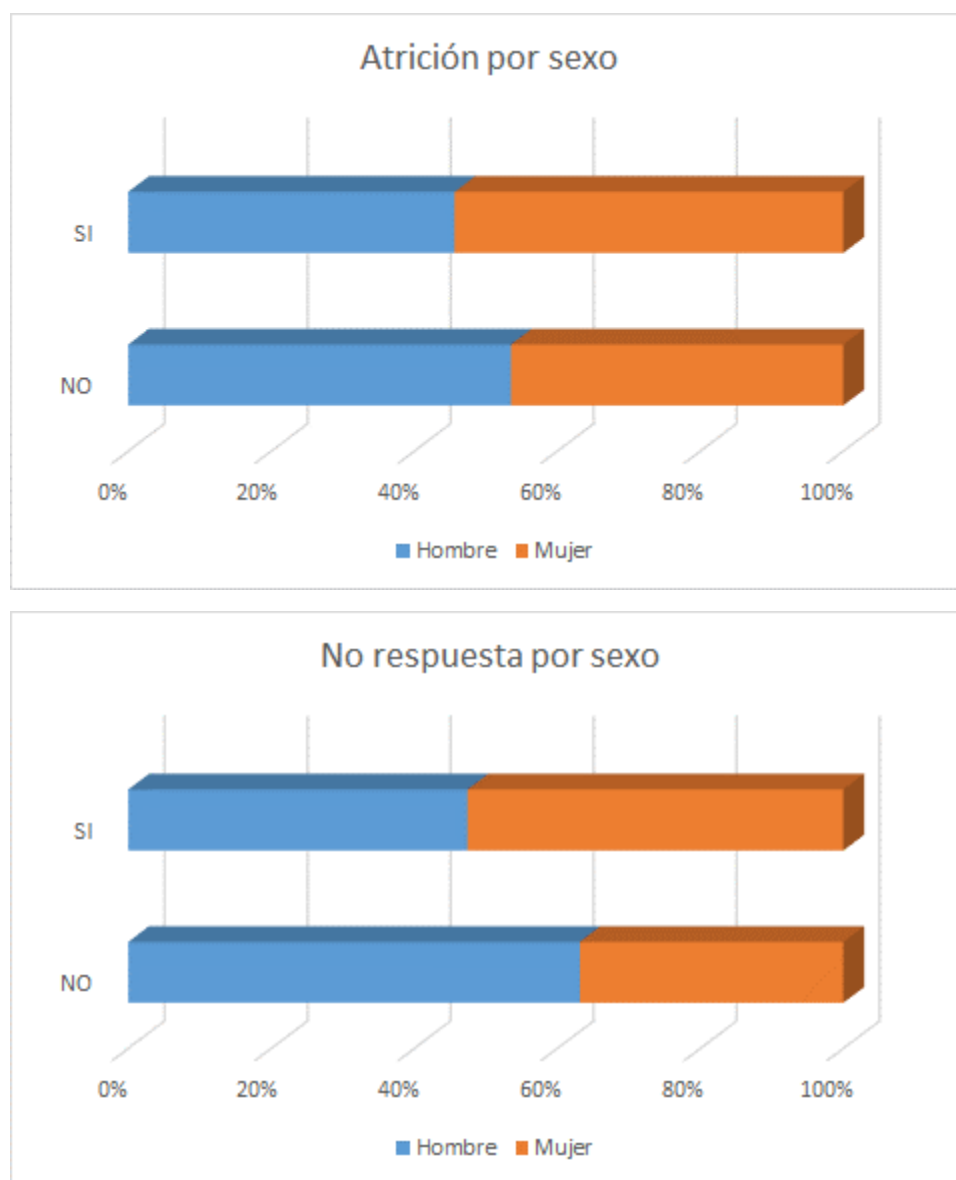
3.5. Conclusiones sobre atrición

Si bien la EPS VII Ronda tuvo por delante muchos obstáculos en su desarrollo, los resultados muestran que la atrición y no respuesta no son atribuibles a situaciones particulares que se puedan identificar y que sesgarían los resultados obtenidos en los tres componentes del estudio.

Por el contrario, y sumando a esto los resultados que se ven en la sección siguiente sobre factores de expansión, las típicas variables que en los estudios ayudan a entender la aparición de sesgos MNAR en los levantamientos, no son importantes en describir el comportamiento en ninguna de las tres encuestas de la presente ola.

Eso permite trabajar el ajuste de factores partiendo de una base no sesgada, que sigue representando a la población adulta de Chile en los temas relacionados con mercado laboral y previsión.

Figura 16. Atrición y no respuesta por sexo en la encuesta de Re-Entrevista



Fuente: elaboración propia.

4. Factores de expansión de corte transversal

4.1. Consideraciones generales sobre los factores de expansión en la EPS VII Ronda

Dadas las características de las encuestas realizadas, la no respuesta a la encuesta por diversas razones es habitual. Esto implica que los factores de expansión de diseño (o base) no sean adecuados para ponderar

a los participantes finales de la encuesta y representar adecuadamente a la población. Esas razones dependen de la encuesta de que se trate (Presencial, de Re-entrevista o de Continuidad), metodología empleada, cobertura, y características de la muestra relevada. Por lo tanto, los factores de expansión de las encuestas finalmente obtenidas deben ser ajustados para que representen adecuadamente la población, tanto transversal como longitudinalmente, evitando que se produzcan sesgos sistemáticos en la información obtenida.

Si bien es cierto que existe consenso sobre la necesidad de corregir estos errores, la discusión comienza muy tempranamente con el dilema acerca de qué tipo de ponderación o factores de expansión emplear en las estimaciones. En este sentido, la bondad del uso de factores de expansión basados en diseño (los más aproximados a los factores de selección o factores base) *versus* el uso de factores basados en modelos, como los que se han obtenido para los tres componentes de la EPS VII Ronda, y que son descriptos en las secciones siguientes, suele ser parte de la discusión, y la respuesta depende del tipo de análisis que se esté realizando y el alcance que se le quiera dar a la inferencia.

Aun cuando es mayoritaria la opinión de que los segundos son los más adecuados, los factores basados en diseño pueden funcionar en algunas circunstancias particulares, por ejemplo, cuando no hay información suficiente para estimar modelos de corrección del comportamiento de la muestra, o por pérdida repentina de representatividad de la muestra.²³

Sin embargo, lo habitual es realizar correcciones a los factores base en función del conocimiento que se tenga de la elegibilidad, cobertura y no respuesta de los entrevistados seleccionados para la encuesta. En este sentido, lo más relevante es el conjunto de información de que se disponga para realizar estimaciones de ajustes a los factores de expansión, bajo el enfoque de ajuste basado en modelos.²⁴

A continuación, se detallan los ajustes a los factores de expansión que se realizaron en los tres componentes de la EPS VII Ronda. Estos ajustes implicaron no sólo el control de problemas de no respuesta (para las tres encuestas) y cobertura (Re-Entrevista y Continuidad) sino que también se debieron ajustar los factores de expansión obtenidos para acercar los totales a la población del país que se quería representar. Esto último se realizó a través de post-estratificación.

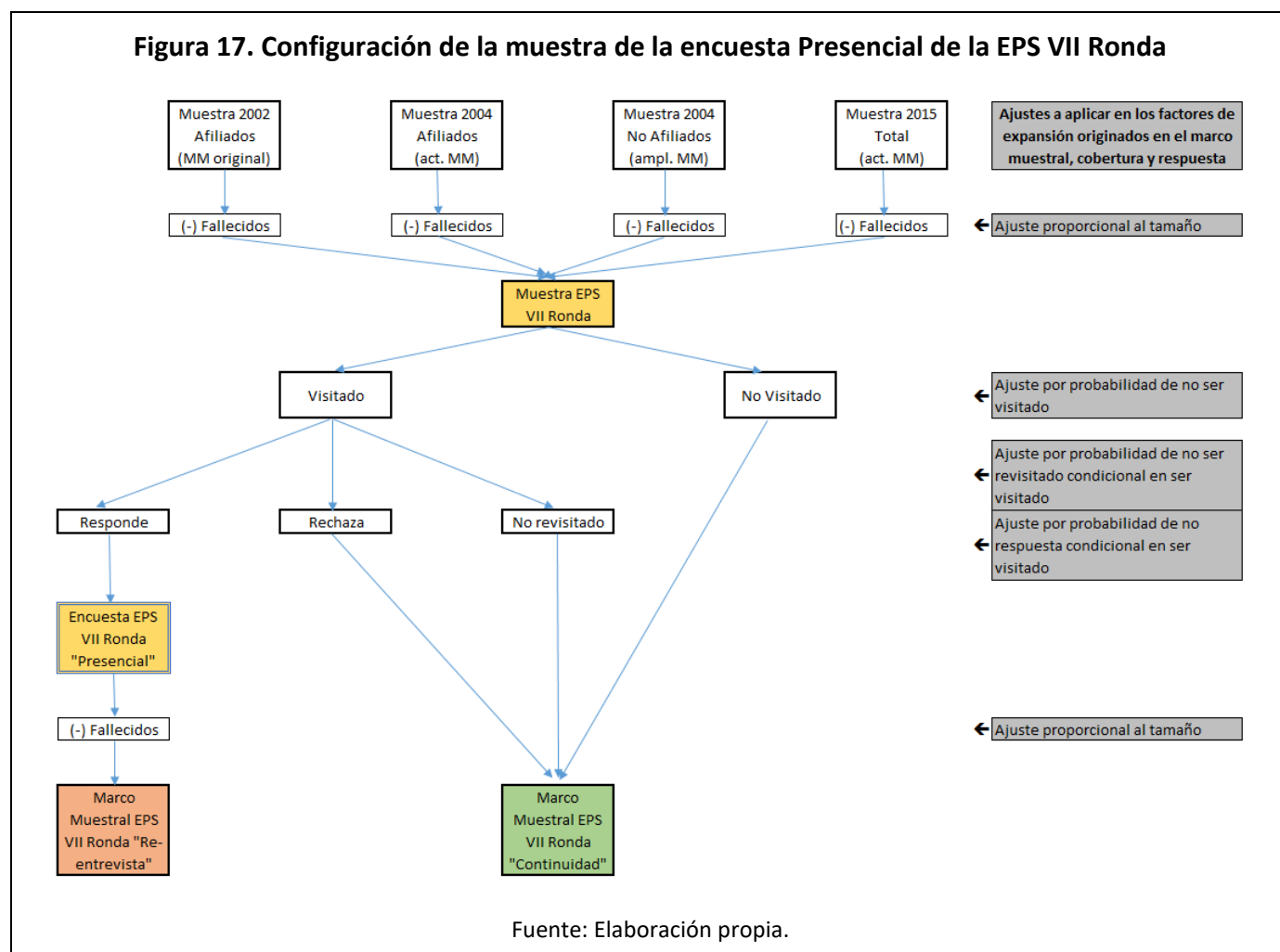
²³ Llama la atención en este ámbito, el uso de factores de diseño que realizó el INE de Chile para su Encuesta Nacional de Empleo en el trimestre móvil Enero-Febrero-Marzo de 2020, en que debió recurrir a los factores de diseño y ajustes del trimestre móvil Diciembre-Enero-Febrero para corregir los desvíos que presentó el relevamiento en el advenimiento de la pandemia de coronavirus y la imposibilidad de levantar la encuesta en terreno. Véase INE Chile (2020).

²⁴ Véase Lohr (2010).

4.2. Factores de expansión de corte transversal de la encuesta Presencial

4.2.1. Configuración de la muestra de la encuesta Presencial

En la **Figura 17**, se esquematiza la estructura y comportamiento de la muestra de la encuesta Presencial, incluyendo la forma en que se originan los marcos muestrales para las encuestas de Continuidad y Re-Entrevista que se desprenden de la encuesta Presencial.



En la parte central de la **Figura 17**, se muestra la configuración de la muestra producto del desarrollo de la encuesta presencial. Los cuadros sombreados con color gris a la derecha del esquema indican los ajustes a los factores de expansión que se hacen necesarios para incorporar posibles sesgos debido a las situaciones representadas en el esquema.

En primer término, la muestra de la EPS VII Ronda está constituida por cuatro sub-muestras, como fue mencionado anteriormente. A cada una de estas sub-muestras, se le deben deducir los fallecidos. Esto requiere que cada sub-muestra de la EPS VII Ronda sea ajustada proporcionalmente para incorporar el

tamaño de las observaciones excluidas debido al fallecimiento dentro de cada *cluster*. De esta manera, las observaciones de “vivos” ex ante representan a toda la población. Estos factores ajustados configuran, en definitiva, los factores de expansión base para esta muestra, $FE_{Presencial}^{Base}$.

En la **Tabla 13** se muestra el comportamiento de la muestra, la representatividad de los factores en cada etapa, los ajustes a tamaños muestrales sobrevivientes, y el total final de encuestados por cada una de la sub-muestras de la encuesta Presencial. Esta tabla será referida profusamente en las explicaciones siguientes.

Además de presentar el comportamiento de la muestra completa, la **Tabla 13** utiliza el mismo esquema de desagregación para cada sub-muestra de la EPS, y también para los entrevistados que respondieron cuestionarios de Vivos, Fallecidos e Impedidos. Nótese que, en los últimos dos casos, el entrevistado es alguien vinculado²⁵ con la persona que forma parte del panel y que, por razones obvias, no puede responder por sí mismo la encuesta. También vale la pena notar que se detectaron fallecidos e impedidos que no respondieron la encuesta Presencial: estos casos corresponden a los códigos de disposición final (“CDF”) 231 para fallecidos, y 237 y 238, para impedidos temporales y permanentes, respectivamente. Esta información fue recolectada durante el trabajo de campo y, si bien no son encuestas logradas, es información relevante al momento de estimar ajustes a los factores de expansión (véase la sección 4.2.2).

²⁵ Durante el trabajo de campo, los encuestadores debían asegurarse, según los protocolos establecidos en este estudio, de que la persona que respondía por fallecidos o impedidos tenía conocimiento suficiente de la situación de estos.

Tabla 13. Encuesta Presencial: comportamiento de la muestra

Tipo de Muestra	Afiliados 2002	Nuevos Afiliados 2004	No Afiliados 2004	Nuevos 2015	Total
Muestra Nominal	14.632	918	2.554	6.778	24.882
Visitados	8.464	476	1.427	3.820	14.187
Responde + Rechaza Visitado	5.669	271	1.018	2.417	9.375
Responde Responde + Rechaza, Visitado	5.154	253	970	2.278	8.655
Vivos	13.886	907	2.371	6.742	23.906
Visitados	7.718	465	1.244	3.784	13.211
Responde + Rechaza Visitado	5.020	263	853	2.384	8.520
Responde Responde + Rechaza, Visitado	4.505	245	805	2.245	7.800
Fallecidos	477	8	113	8	606
Visitados	477	8	113	8	606
*Sin encuesta (CDF 231)	75	2	12	1	90
Responde + Rechaza Visitado	402	6	101	7	516
Responde Responde + Rechaza, Visitado	402	6	101	7	516
Impedidos	269	3	70	28	370
Visitados	269	3	70	28	370
**Sin encuesta, temporales (CDF 237)	11	0	3	1	15
***Sin encuesta, permanentes (CDF 238)	11	1	3	1	16
Responde + Rechaza Visitado	247	2	64	26	339
Responde Responde + Rechaza, Visitado	247	2	64	26	339
Total Encuestas	5.154	253	970	2.278	8.655
Población representada originalmente	8.074.667	527.914	3.994.879	3.250.561	15.848.021
Factor de expansión base promedio	1566,680	2086,617	4118,432	1426,936	1831,083
Población representada con ajuste por NR	8.072.384	524.405	4.021.555	3.250.389	15.868.733
Factor de expansión ajustado por NR promedio	1791,872	2140,428	4995,720	1447,835	2034,453
Población representada post-estratificada	6.085.856	537.055	2.993.179	3.801.727	13.417.817
Factor de expansión post-estratificado promedio	1350,911	2192,060	3718,235	1693,420	1720,233

Fuente: elaboración propia.

4.2.2. Ajustes por atrición y no respuesta

Para realizar los ajustes a los factores de expansión finales, se trató cada sub-muestra por separado, y luego se las reunió nuevamente para los ajustes finales de post-estratificación. El procedimiento seguido fue el siguiente:

PASO 1: Se tomaron los factores de expansión base provenientes de la ola en que se hizo la selección muestral respectiva. Esos factores base se encuentran ajustados por la probabilidad inicial de no

respuesta (es decir, no son exactamente los factores de diseño, sino que aquellos resultantes del ajuste proporcional de la muestra original al resultado de campo de la primera ronda de levantamiento de cada sub-muestra, sin que medien otros ajustes). A modo de ejemplo, en la EPS I Ronda se seleccionaron 30.000 observaciones, con un objetivo de levantamiento de 20.000 encuestas, de las que se recolectaron 17.246 encuestas. Sin embargo, de éstas sólo correspondieron a Vivos un total de 16.032 encuestas que, desde entonces, conforman el denominado Panel Afiliados 2002.

PASO 2: A los factores de expansión base se les realizó un ajuste proporcional al tamaño por pérdida de muestra²⁶ entre la ola original, y las observaciones disponibles dentro del conjunto de contactos entregados por la SPS. Estos 24.882 posibles entrevistados son el resultado de la depuración de la base original de contactos conformada a través de las olas, deducidos los fallecidos y otras pérdidas de muestra (por ejemplo, mudanzas al exterior del país del entrevistado). Se identifica a ésta como la muestra nominal de la EPS VII Ronda. Siguiendo con el ejemplo de la sub-muestra Afiliados 2002, según se reporta en la **Tabla 13**, de 16.032 casos originales, tuvimos a disposición 14.632 casos nominales.

PASO 3: Se realizó un segundo ajuste proporcional al tamaño para lograr que los entrevistados que respondieron el cuestionario de Vivos representen el universo adecuadamente. Este ajuste obedece a lo que se denomina muestra efectiva, es decir, el conjunto de potenciales entrevistados Vivos, después de deducir de la base de contactos aquellos casos en que se conoce fehacientemente que los entrevistados están fallecidos o impedidos, en virtud del resultado del trabajo de campo. Para el ejemplo de los Afiliados 2002, esa muestra nominal se redujo de 14.632 a una muestra efectiva de 13.886 potenciales entrevistados vivos a incluir en el ajuste por no respuesta.²⁷

PASO 4: Cuando se llevó a cabo el levantamiento de la encuesta Presencial, 10.695 personas no alcanzaron a ser visitados. Por lo tanto, no se tiene ninguna información sobre ellos a excepción de la proveniente de la última ola en que los entrevistados respondieron la encuesta. Si bien se estima que estos hogares no diferían de manera particular de los hogares que sí fueron visitados, se realizó un ajuste a los factores de expansión de Visitados y No Visitados que da cuenta de los

²⁶ Los ajustes proporcionales se emplean cuando no existe una interpretación razonable de que la pérdida de muestra está relacionada con la encuesta, su contenido, o su ejecución. Para la EPS, entonces, es razonable asumir que un fallecimiento o impedimento temporal o permanente no está relacionado con la encuesta, por lo que la realización de un ajuste proporcional no distorsiona las probabilidades de cada individuo de participar en la encuesta. Más aun, con ese ajuste se logra redistribuir el peso que tienen las unidades muestrales perdidas entre el resto de los entrevistados participantes, sin distorsionar los tamaños de *cluster* ni su probabilidad de selección. Nótese que este razonamiento no es general: si, por ejemplo, la encuesta fuese sobre prevalencia de una enfermedad mortal, la causa del fallecimiento sería informativo respecto de la encuesta, por lo que se debería incorporar esa información en los factores de expansión ajustados.

²⁷ Los ajustes de los pasos 2 y 3 corresponden a atrición en sentido propio, ya que no existe (hasta donde es posible conocer) que exista autoselección. En los pasos siguientes, el ajuste es por atrición por no respuesta, o simplemente, ajuste por no respuesta.

posibles diferenciales en la probabilidad de haber sido o no visitado durante el trabajo de campo de la encuesta Presencial.²⁸

El ajuste se realizó mediante la estimación de un modelo de propensión a ser visitado/no visitado basado en regresión logística para obtener la probabilidad de ser visitado, $Pr(Visitado)$. Se probaron diversas especificaciones de los modelos explicativos de la situación de los entrevistados y su probabilidad de ser visitados. El poder explicativo de todos ellos resultó bajo si observamos el Pseudo R^2 propuesto por McFadden. De todos modos, esto también es indicio de que la información con que habitualmente se cuenta para llevar adelante el muestreo y el trabajo de campo, no influyó en el desarrollo del trabajo de campo ni en la probabilidad de ser visitados.

En la **Tabla 14** (y sólo a modo ilustrativo, ya que se generan muchas tablas de este tipo a lo largo del proceso de ajuste de factores), se muestra el resultado de la estimación de los efectos marginales del modelo finalmente elegido para el ajuste de la probabilidad de ser visitado, en base a 13.886 casos posibles para la muestra efectiva de Afiliados 2002:

²⁸ Este aspecto del comportamiento de la muestra, en conjunto con las características de los entrevistados que no pudieron ser revisitados para conseguir respuesta o rechazo condicional en haber sido visitados al menos una vez, son analizados el “Informe sobre la factibilidad técnica de combinación de muestras: efectos de la irrupción de COVID-19 sobre el levantamiento de la EPS VII Ronda”.

Tabla 14. Ejemplo de resultado de estimación de la probabilidad de ser visitado en la encuesta Presencial, Afiliados 2002

	Modelo 1	Modelo 2	Modelo 3	Modelo 4
Fue visitado en la encuesta Presencial				
Edad 45-54	0.0839 (0.0634)	0.0892 (0.0632)	0.101 (0.0635)	0.107 (0.0609)
Edad 55-64	0.0986 (0.0642)	0.105 (0.0641)	0.122 (0.0646)	0.124* (0.0620)
Edad 65-74	0.114 (0.0719)	0.122 (0.0716)	0.162* (0.0728)	0.160* (0.0698)
Edad 75 o más	-0.245** (0.0863)	-0.229** (0.0863)	-0.177* (0.0888)	-0.167* (0.0854)
I. Tarapacá	-0.0674 (0.127)	-0.0630 (0.126)	-0.0640 (0.127)	
II. Antofagasta	0.707*** (0.131)	0.691*** (0.130)	0.670*** (0.130)	
III. Atacama	0.705*** (0.146)	0.683*** (0.145)	0.664*** (0.145)	
IV. Coquimbo	-1.087*** (0.0910)	-1.091*** (0.0904)	-1.094*** (0.0905)	
V. Valparaíso	-0.744*** (0.0635)	-0.725*** (0.0630)	-0.732*** (0.0631)	
VI. O'Higgins	-0.627*** (0.0803)	-0.614*** (0.0796)	-0.609*** (0.0799)	
VII. Maule	-1.160*** (0.0788)	-1.139*** (0.0786)	-1.128*** (0.0789)	
VIII. Biobío	0.126* (0.0616)	0.136* (0.0615)	0.141* (0.0616)	
IX. Araucanía	-1.282*** (0.0966)	-1.264*** (0.0967)	-1.257*** (0.0966)	
X. Los Lagos	0.858*** (0.0811)	0.843*** (0.0815)	0.846*** (0.0817)	
XI. Aysén	-1.003*** (0.199)	-1.007*** (0.195)	-0.988*** (0.195)	
XII. Magallanes	-0.261 (0.170)	-0.281 (0.169)	-0.274 (0.168)	
Hombre	-0.185*** (0.0393)	-0.191*** (0.0393)	-0.191*** (0.0395)	-0.186*** (0.0382)
2002 2004	0.209 (0.131)			
2002 2004 2006	0.305* (0.121)			
2002 2004 2006 2009	0.390*** (0.105)			
2002 2004 2006 ~2012	0.519*** (0.103)			
2002 2004 2006 ~2015	0.988*** (0.0922)			
2002 2004 2006 ~2015	0.730*** (0.103)			
2002 2004 2006 2012	0.485** (0.178)			
2002 2004 2006 ~2015	0.757*** (0.159)			
2002 2004 2006 2015	0.856*** (0.157)			
2002 2004 2009	0.414* (0.198)			
2002 2004 2009 2012	0.306 (0.308)			
2002 2004 2009 ~2015	0.728*** (0.195)			
2002 2004 2009 2015	0.581** (0.213)			
2002 2004 2015	0.751***			

2002 2006	(0.211)			
	0.268			
2002 2006 2009	(0.156)			
	0.415**			
2002 2006 2009 2012	(0.158)			
	0.753***			
2002 2006 2009 ~2015	(0.196)			
	0.981***			
2002 2006 2009 2015	(0.152)			
	0.980***			
2002 2006 2012	(0.176)			
	0.336			
2002 2006 2012 2015	(0.301)			
	0.790*			
2002 2006 2015	(0.335)			
	0.713**			
2002 2009	(0.246)			
	0.333			
2002 2009 2012	(0.208)			
	0.756*			
2002 2009 2012 2015	(0.333)			
	0.702*			
2002 2009 2015	(0.314)			
	0.494			
2002 2015	(0.281)			
	0.149			
Afiliado: Sí	(0.234)			
	0.114	0.109	0.114	0.0925
Pensionado: Sí	(0.0877)	(0.0877)	(0.0879)	(0.0850)
	0.0226	0.0167	0.0270	0.0416
EPS 2004	(0.0405)	(0.0404)	(0.0414)	(0.0400)
	0.209	0.186	0.130	
EPS 2006	(0.131)	(0.131)	(0.128)	
	0.293**	0.295**	0.275**	
EPS 2009	(0.111)	(0.110)	(0.106)	
	0.389***	0.373***	0.343***	
EPS 2012	(0.0988)	(0.0981)	(0.0946)	
	0.525***	0.496***	0.404***	
EPS 2015	(0.0983)	(0.0979)	(0.0943)	
	0.874***	0.839***	0.736***	
Educación Básica	(0.0889)	(0.0884)	(0.0850)	
	0.335*	0.412**		
Educación Media	(0.169)	(0.159)		
	0.434*	0.509**		
Educación Terciaria	(0.171)	(0.160)		
	0.393*	0.471**		
Educación Universi~a	(0.179)	(0.168)		
	0.141	0.203		
Norte	(0.179)	(0.168)		
	-0.106			
Centro	(0.0627)			
	-0.473***			
Sur	(0.0429)			
	-0.199***			
Constant	(0.0570)			
	-0.173	-0.174	-0.531**	-0.545**
	(0.103)	(0.103)	(0.198)	(0.188)

Observations	13886	13886	13886	13886
Pseudo R-squared	0.065	0.062	0.064	0.023

Standard errors in parentheses

* p<0.05, ** p<0.01, *** p<0.001

Fuente: elaboración propia.

A continuación, en la **Tabla 15**, se puede observar la distribución de las probabilidades de ser visitado según se obtienen de este modelo:

Tabla 15. Ejemplo de distribución de probabilidad estimada de ser visitado en la encuesta Presencial, sub-muestra Afiliados 2002

Pr (d_Visitado)				
	Percentiles	Smallest		
1%	.2225619	.1342433		
5%	.2871078	.1490785		
10%	.3407283	.1556023	Obs	13,886
25%	.4406999	.1556023	Sum of Wgt.	13,886
50%	.5608495		Mean	.5530053
		Largest	Std. Dev.	.1499006
75%	.6744148	.8591787		
90%	.7298673	.8591787	Variance	.0224702
95%	.7894773	.8591787	Skewness	-.2538642
99%	.8398272	.8591787	Kurtosis	2.371052

Fuente: elaboración propia.

PASO 5: Dentro del grupo de entrevistados de panel que fueron visitados, se debe ajustar por la probabilidad conjunta de responder o rechazar *versus* no ser revisitado, condicional en ser visitados al menos una vez. El ajuste se realizó mediante la estimación de un modelo de propensión de (responder + rechazar)/no ser revisitado basado en regresión logística para obtener las probabilidades de (responder + rechazar)/no ser revisitado, condicional en haber sido visitado al menos una vez durante el trabajo de campo de la encuesta Presencial, es decir, $Pr(\text{Responder} \cup \text{Rechazar} | \text{Visitado})$. Se probaron varias especificaciones del modelo, aunque todas con bajo *Pseudo R*². Para esta estimación, se disponía de 7.714 observaciones que cumplieran con la condición dentro de la muestra de Afiliados 2002.²⁹ A continuación, en la **Tabla 16**, y sólo como ejemplo, se muestra el resultado de la estimación de los efectos marginales del modelo elegido para la sub-muestra de Afiliados 2002:

Tabla 16. Ejemplo de resultado de estimación de la probabilidad de Responder + Rechazar | Ser Visitado en la encuesta Presencial, Afiliados 2002

	Modelo 1	Modelo 2	Modelo 3	Modelo 4
Responde o rechaza la encuesta Presencial				
Edad 45-54	0.267** (0.0876)	0.272** (0.0874)	0.274** (0.0866)	0.269** (0.0872)
Edad 55-64	0.410*** (0.0896)	0.428*** (0.0892)	0.468*** (0.0881)	0.412*** (0.0893)
Edad 65-74	0.595*** (0.103)	0.615*** (0.103)	0.679*** (0.100)	0.596*** (0.102)
Edad 75 o más	0.449***	0.483***	0.610***	0.436***

²⁹ En la **Tabla 13**, se puede observar la cantidad de observaciones para cada una de las otras sub-muestras de la EPS VII Ronda. Las explicaciones dadas en este procedimiento paso-a-paso se aplicaron de forma idéntica a las demás sub-muestras.

	(0.128)	(0.127)	(0.123)	(0.126)
I. Tarapacá	-0.818***	-0.812***	-0.823***	
	(0.171)	(0.167)	(0.166)	
II. Antofagasta	-0.330*	-0.341*	-0.362*	
	(0.147)	(0.148)	(0.147)	
III. Atacama	-0.284	-0.353*	-0.311*	
	(0.150)	(0.150)	(0.152)	
IV. Coquimbo	0.419*	0.436**	0.460**	
	(0.169)	(0.167)	(0.167)	
V. Valparaíso	-0.260**	-0.213*	-0.228*	
	(0.0991)	(0.0973)	(0.0968)	
VI. O'Higgins	-0.261*	-0.219	-0.156	
	(0.127)	(0.125)	(0.124)	
VII. Maule	-0.412**	-0.343*	-0.305*	
	(0.139)	(0.137)	(0.136)	
VIII. Biobío	0.0228	0.0540	0.0837	
	(0.0802)	(0.0798)	(0.0797)	
IX. Araucanía	-0.219	-0.161	-0.130	
	(0.170)	(0.168)	(0.168)	
X. Los Lagos	0.124	0.123	0.172	
	(0.0936)	(0.0933)	(0.0925)	
XI. Aysén	-0.537	-0.596	-0.557	
	(0.348)	(0.346)	(0.342)	
XII. Magallanes	1.705***	1.692***	1.707***	
	(0.420)	(0.424)	(0.419)	
Hombre	-0.299***	-0.313***	-0.281***	-0.299***
	(0.0554)	(0.0551)	(0.0546)	(0.0551)
Educación Básica	-0.237	-0.191		-0.232
	(0.282)	(0.276)		(0.283)
Educación Media	-0.419	-0.392		-0.416
	(0.282)	(0.276)		(0.282)
Educación Terciaria	-0.697*	-0.681*		-0.698*
	(0.291)	(0.285)		(0.291)
Educación Universi~a	-0.675*	-0.681*		-0.685*
	(0.290)	(0.285)		(0.291)
2002	-1.628***			-1.594***
	(0.132)			(0.130)
2002 2004	-1.597***			-1.595***
	(0.153)			(0.151)
2002 2004 2006	-1.405***			-1.381***
	(0.125)			(0.125)
2002 2004 2006 2009	-1.351***			-1.309***
	(0.0994)			(0.0986)
2002 2004 2006 ~2012	-1.086***			-1.075***
	(0.0966)			(0.0958)
2002 2004 2006 ~2015	-0.447***			-0.430***
	(0.0957)			(0.0954)
2002 2004 2006 2012	-1.385***			-1.386***
	(0.229)			(0.228)
2002 2004 2006 ~2015	-0.577**			-0.582**
	(0.202)			(0.202)
2002 2004 2006 2015	-0.625***			-0.601***
	(0.176)			(0.176)
2002 2004 2009	-1.318***			-1.287***
	(0.234)			(0.230)
2002 2004 2009 2012	-1.468***			-1.506***
	(0.392)			(0.388)
2002 2004 2009 ~2015	-0.227			-0.244
	(0.256)			(0.261)
2002 2004 2009 2015	-0.821**			-0.729**
	(0.283)			(0.279)
2002 2004 2015	-0.883***			-0.852***
	(0.246)			(0.237)
2002 2006	-1.850***			-1.829***
	(0.187)			(0.186)
2002 2006 2009	-1.368***			-1.352***
	(0.186)			(0.183)
2002 2006 2009 2012	-0.825***			-0.818***
	(0.218)			(0.220)
2002 2006 2009 ~2015	-0.656***			-0.693***

2002 2006 2009 2015	(0.161)			(0.158)
	-0.534**			-0.509*
2002 2006 2012	(0.200)			(0.199)
	-2.580***			-2.575***
2002 2006 2012 2015	(0.467)			(0.470)
	-0.0946			-0.122
2002 2006 2015	(0.458)			(0.453)
	-1.008***			-0.986***
2002 2009	(0.300)			(0.299)
	-1.448***			-1.359***
2002 2009 2012	(0.242)			(0.241)
	-1.702***			-1.697***
2002 2009 2012 2015	(0.443)			(0.443)
	0.163			0.217
2002 2009 2015	(0.503)			(0.506)
	-0.0379			-0.0492
2002 2015	(0.429)			(0.430)
	-0.562			-0.538
Afiliado: Sí	(0.324)			(0.326)
	-0.0550	-0.0683	-0.112	-0.0363
Pensionado: Sí	(0.126)	(0.125)	(0.124)	(0.124)
	0.0708	0.0647	-0.00547	0.0706
EPS 2004	(0.0576)	(0.0574)	(0.0559)	(0.0572)
		0.0317	0.0430	
EPS 2006		(0.187)	(0.185)	
		0.0946	0.0683	
EPS 2009		(0.153)	(0.153)	
		0.269	0.272*	
EPS 2012		(0.138)	(0.138)	
		0.466***	0.505***	
EPS 2015		(0.139)	(0.138)	
		1.356***	1.401***	
Norte		(0.126)	(0.126)	-0.321***
				(0.0893)
Centro				-0.153*
				(0.0624)
Sur				0.110
				(0.0828)
Constant	1.629***	-0.0312	-0.494***	1.612***
	(0.297)	(0.308)	(0.146)	(0.297)
Observations	7718	7718	7718	7718
Pseudo R-squared	0.095	0.085	0.081	0.088

Standard errors in parentheses
 * p<0.05, ** p<0.01, *** p<0.001

Fuente: elaboración propia.

A continuación, en la **Tabla 17**, se puede observar la distribución de las probabilidades predichas:

Tabla 17. Ejemplo de distribución de probabilidad estimada de Responder + Rechazar | Ser Visitado en la encuesta Presencial, sub-muestra Afiliados 2002

Pr(d_RespRech)				
	Percentiles	Smallest		
1%	.2686871	.0981138		
5%	.3450892	.1195174		
10%	.39924	.1291336	Obs	13,886
25%	.4962177	.1406801	Sum of Wgt.	13,886
50%	.6517504		Mean	.6310215
		Largest	Std. Dev.	.1665926
75%	.7773371	.9729031		
90%	.8339633	.9729582	Variance	.0277531
95%	.8585035	.9773587	Skewness	-.3245164
99%	.8953995	.9778542	Kurtosis	2.090429

Fuente: elaboración propia.

PASO 6: A su vez, se sabe que aquellos que responden o rechazan definitivamente el cuestionario, típicamente difieren entre sí, por lo que se debe realizar un ajuste a los factores de expansión siguiendo la lógica de los ajustes por no respuesta. Para ello, se estima un modelo de propensión de respuesta/rechazo basado en regresión logística para obtener la probabilidad de responder/rechazar, condicional en haber sido visitado y no ser parte del grupo de los “no revisitados”, es decir, $Pr(Responder|Responder \cup Rechazar, Visitado)$. Nuevamente, se probaron varias especificaciones, y a continuación se muestra como ejemplo, la estimación de los efectos marginales para la sub-muestra Afiliados 2002, basado en 5.020 observaciones que cumplían la condición, y de la cuáles, 4.505 respondieron efectivamente la encuesta:

Tabla 18. Ejemplo de resultado de estimación de la probabilidad de Responder | Responder + Rechazar, Ser Visitado en la encuesta Presencial, Afiliados 2002

	Modelo 1	Modelo 2	Modelo 3	Modelo 4
Responde en la ron~				
Edad 45-54	-0.0424 (0.182)	-0.0588 (0.182)		-0.0423 (0.182)
Edad 55-64	0.106 (0.183)	0.0577 (0.183)		0.111 (0.183)
Edad 65-74	0.166 (0.199)	0.102 (0.200)		0.167 (0.199)
Edad 75 o más	0.00687 (0.222)	-0.126 (0.227)		0.0101 (0.223)
I. Tarapacá	-0.643* (0.312)	-0.652* (0.309)	-0.653* (0.307)	
II. Antofagasta	-0.166 (0.264)	-0.148 (0.264)	-0.133 (0.265)	
III. Atacama	-0.000907 (0.338)	-0.0861 (0.336)	-0.0955 (0.335)	
IV. Coquimbo	1.039* (0.427)	1.000* (0.429)	1.008* (0.429)	

V. Valparaíso	-0.0585 (0.194)	-0.0440 (0.194)	-0.0504 (0.194)	
VI. O'Higgins	-0.690*** (0.200)	-0.762*** (0.203)	-0.758*** (0.202)	
VII. Maule	-0.208 (0.249)	-0.248 (0.251)	-0.243 (0.250)	
VIII. Biobío	0.195 (0.162)	0.152 (0.164)	0.155 (0.164)	
IX. Araucanía	0.527 (0.420)	0.498 (0.427)	0.485 (0.425)	
X. Los Lagos	-0.152 (0.158)	-0.216 (0.160)	-0.227 (0.160)	
XI. Aysén	-1.262* (0.521)	-1.291* (0.519)	-1.293* (0.516)	
XII. Magallanes	0.242 (0.444)	0.248 (0.442)	0.255 (0.443)	
Hombre	-0.297** (0.105)	-0.330** (0.104)	-0.327** (0.104)	-0.282** (0.104)
EPS 2004	0.154 (0.319)	0.175 (0.320)	0.186 (0.320)	0.121 (0.319)
EPS 2006	0.225 (0.274)	0.304 (0.272)	0.298 (0.272)	0.192 (0.272)
EPS 2009	0.529* (0.252)	0.557* (0.252)	0.564* (0.251)	0.521* (0.250)
EPS 2012	0.967*** (0.266)	0.946*** (0.264)	0.947*** (0.263)	0.943*** (0.264)
EPS 2015	1.544*** (0.232)	1.505*** (0.230)	1.512*** (0.229)	1.511*** (0.229)
Sí	-0.225 (0.204)	-0.168 (0.210)	-0.250 (0.179)	-0.211 (0.204)
Sí	-0.0935 (0.107)	0.00200 (0.110)	0.00965 (0.109)	-0.111 (0.106)
Educación Básica		0.0847 (0.489)	0.108 (0.491)	
Educación Media		-0.165 (0.485)	-0.144 (0.486)	
Educación Terciaria		-0.497 (0.501)	-0.482 (0.502)	
Educación Universi~a		-0.635 (0.500)	-0.606 (0.502)	
Norte				-0.0715 (0.174)
Centro				-0.0804 (0.118)
Sur				-0.0620 (0.145)
Constant	1.228*** (0.278)	1.465** (0.548)	1.440** (0.527)	1.252*** (0.277)
Observations	5020	5020	5020	5020
Pseudo R-squared	0.058	0.065	0.064	0.047

Standard errors in parentheses

* p<0.05, ** p<0.01, *** p<0.001

Fuente: elaboración propia.

En la **Tabla 19**, se muestran los principales estadísticos de la distribución de la probabilidad de respuesta predicha a partir del modelo anterior:

Tabla 19. Ejemplo de distribución de probabilidad estimada de Responder | (Responder + Rechazar, Ser Visitado) en la encuesta Presencial, sub-muestra Afiliados 2002

Pr (d_Responde)				
	Percentiles	Smallest		
1%	.6191812	.340484		
5%	.7093183	.3790597		
10%	.7562068	.395265	Obs	13,886
25%	.8300663	.412545	Sum of Wgt.	13,886
50%	.9024892		Mean	.8748303
		Largest	Std. Dev.	.0823655
75%	.9362084	.9845881		
90%	.950608	.9845881	Variance	.0067841
95%	.9592175	.9845881	Skewness	-1.385778
99%	.977272	.9845881	Kurtosis	5.11484

Fuente: elaboración propia.

PASO 7: A continuación, se obtuvo la probabilidad incondicional de responder la encuesta Presencial, es decir,

$$\begin{aligned}
 PR(\text{Respuesta}) &= \Pr(\text{Responde} | \text{Responde} \cup \text{Rechaza}, \text{Visitado}) \\
 &\cdot \Pr(\text{Responde} \cup \text{Rechaza} | \text{Visitado}) \cdot \Pr(\text{Visitado})
 \end{aligned}$$

La siguiente tabla muestra la distribución de la probabilidad predicha para la sub-muestra Afiliados 2002:

Tabla 20. Ejemplo de distribución de probabilidad incondicional estimada de Responder la encuesta Presencial, sub-muestra Afiliados 2002

Prop_Score				
	Percentiles	Smallest		
1%	.0547531	.0192231		
5%	.0971252	.0195988		
10%	.1263831	.0231015	Obs	13,886
25%	.1960502	.0236543	Sum of Wgt.	13,886
50%	.3034412		Mean	.321907
		Largest	Std. Dev.	.1548501
75%	.4439037	.7254544		
90%	.5437365	.7254544	Variance	.0239785
95%	.5866237	.7254544	Skewness	.2783207
99%	.6539441	.7332186	Kurtosis	2.102759

Fuente: elaboración propia.

PASO 8: una vez que se ha estimado la probabilidad de responder la encuesta, se debe realizar el ajuste por no respuesta. Si bien existen múltiples enfoques para llevarlo a cabo, y la literatura es prolífera en estos tópicos³⁰, se optó por revisar la factibilidad de dos de ellos: (i) enfoque de ajuste por celdas, y (ii) enfoque de ajuste directo.

En el enfoque de ajuste por celdas, se debe decidir cómo agrupar las observaciones según algún criterio. Los criterios pueden ir desde características poblacionales, hasta métodos sofisticados de *clusterización*, *cóputas* y *random forest classification*. En el caso de la EPS VII Ronda, se recurrió a un método bastante empleado en la práctica y que consiste en dividir la distribución de la probabilidad de no respuesta predicha en tramos idénticos. Para ello, se utilizaron tres alternativas: separación por deciles (10 partes iguales), ventiles (20 partes iguales) y percentiles (100 partes iguales) de la distribución del *propensity score* estimado. Dentro de cada separación, se toma un promedio simple de las probabilidades y se utiliza ese factor de ajuste para las observaciones dentro del *cluster*.

A continuación, se puede ver un ejemplo del comportamiento de la distribución del *propensity score* cuando se lo agrupa en ventiles, para la sub-muestra Afiliados 2002:

³⁰ Véase Buuren (2017), Lohr (2010), Valliant et al. (2018), Valliant y Dever (2018), entre otros, que resumen varias de las alternativas disponibles.

Tabla 21. Ejemplo de agrupación por ventiles de la probabilidad incondicional estimada de no respuesta de la encuesta Presencial, sub-muestra Afiliados 2002

	Ventil	Minimo	Media	Maximo
1.	.0970543	.0192231	.0719907	.0970161
2.	.1263803	.0971252	.1117898	.1263739
3.	.1522628	.1263831	.1397951	.1521632
4.	.1742548	.1522628	.1634776	.1742337
5.	.1960465	.1742865	.1846147	.1960354
6.	.2169945	.1960502	.206258	.2169945
7.	.2370722	.2171598	.2266315	.2370395
8.	.2593128	.2371122	.2481117	.2593128
9.	.2810855	.2593369	.2703054	.2810847
10.	.3034412	.2810901	.2914118	.303416
11.	.3279516	.3034664	.3154961	.3278967
12.	.3529578	.3279613	.340919	.3529467
13.	.3810632	.3530025	.3667524	.380977
14.	.4134789	.3811338	.3973396	.4133576
15.	.4439037	.4134923	.429172	.4439037
16.	.4800726	.4441019	.4622467	.4800726
17.	.5127254	.4800967	.49709	.5127254
18.	.5437365	.5128428	.5293393	.5437365
19.	.5866237	.5437818	.5649701	.5866237
20.	.7332186	.5868964	.6253268	.7332186

Fuente: elaboración propia.

El enfoque de ajuste directo, por su parte, simplemente implica que el *propensity score* obtenido se utiliza como denominador de los factores base de cada observación, “inflándolos” para que representen a la población completa.

Para determinar cuál de los métodos es más adecuado, se tuvieron en cuenta los efectos de los distintos ponderadores sobre el tamaño poblacional representado, y cómo éste conversaba con el tamaño poblacional original. También se tuvo en cuenta el rango intercuartílico, varianza y valores extremos resultantes de la aplicación de cada uno de estos ponderadores de probabilidad de responder la encuesta.

Finalmente, se utilizó alguno de estos factores para ajustar por no respuesta el factor de expansión. En el caso de la encuesta Presencial, las poblaciones representadas después de cada ajuste por no respuesta, y para la muestra total, se pueden ver en la sección final de la **Tabla 13**, en comparación el factor base.

Después de aplicar este procedimiento, que implicó ajustes sucesivos por muestra nominal, efectiva y no respuesta en sus distintos niveles, se logró representar algo más de 15,8 millones de personas mayores de 23 años que habitan Chile.³¹

4.2.3. Post-estratificación de los factores de corte transversal

Aun cuando los factores de expansión han sido ajustados por atrición y no respuesta, no logran representar a la población contingente del momento de la encuesta. Por lo tanto, se realiza un último ajuste para empatar la población representada con la inferencia que se puede realizar a partir de la muestra. Para hacer este procedimiento, se deben calibrar los factores de expansión.³²

Los factores de expansión de la EPS VII Ronda se han ajustado empleando post-estratificación por celdas poblacionales, como se hizo en las olas anteriores. En el procedimiento de post-estratificación, se determinan celdas de ajuste para las cuales es posible fijar (multidimensionalmente) la cantidad de personas que se deberían representar. Si bien el método es simple (sólo requiere el cálculo de proporciones), cuando las celdas se determinan con muchas variables o categorías, o cuando una determinada configuración deja una celda con pocas observaciones encuestadas, los ponderadores pueden “explotar”, es decir, puede aumentar o disminuir bruscamente causando que los factores de expansión finales muestren exacerbaciones que pueden requerir *trimming* (es decir, “recortes” teniendo en cuenta máximos y mínimos de la distribución, varianzas, etc.) o construcción de pseudo-estratos y/o pseudo-clusters.

En el caso de esta ola de la EPS, se aplicó la post-estratificación por sexo, grupo etario y región como es tradición en la encuesta. Los totales poblacionales se tomaron de las proyecciones censales del INE, disponibles públicamente. La fecha que se tomó como referencia para la encuesta Presencial fue el mes de noviembre de 2019, periodo anterior al inicio del levantamiento. Los totales poblacionales para los que se post-estratificó se muestran en la **Tabla 22**.

Nótese que las proyecciones de población entregadas por el INE, se refieren a junio de cada año. Por lo tanto, la información empleada en la post-estratificación de la EPS y mostrada en la **Tabla 22**, corresponde a una interpolación entre junio de 2019 y noviembre del mismo año. La interpolación se realizó asumiendo que

³¹ Nótese que la EPS VII Ronda representa a personas con 23 o más años de edad, que son los más jóvenes incorporados a la muestra de panel, es decir, aquellos que en 2015 tenían 18 años de edad, cuando se realizó la última incorporación de muestra.

³² La calibración puede tomar muchas formas, entre las que se sugiere particularmente la calibración iterativa por márgenes/totales poblacionales (también conocida como *raking*). Este tipo de calibración ha demostrado buenos comportamientos en contextos de encuestas complejas como la EPS, y tiene algunas ventajas técnicas que han sido discutidas en la literatura especializada. Sin embargo, en esta edición de la EPS, y siguiendo los términos contractuales, se ha utilizado el procedimiento de post-estratificación que es un tipo de calibración por celdas específicas (en lugar de los marginales o los totales).

la tasa de crecimiento entre 2019 y 2020 fue una tasa efectiva, por lo que la tasa de crecimiento del periodo se puede modelar como:³³

$$\bar{r} = \frac{1}{T} \ln \left(\frac{N_T}{N_0} \right)$$

Donde T es la cantidad de periodos de duración sobre el que se mide la tasa de crecimiento, N_T es la población al final del periodo, y N_0 es la población al inicio. Para el caso de la EPS, $T = 1$ año, por lo que la expresión anterior se reduce a:

$$\bar{r} = \ln \left(\frac{N_{2020}}{N_{2019}} \right)$$

Por su parte, empleando esta tasa de crecimiento, se puede obtener la población en el sub-periodo t mediante la siguiente expresión:

$$N(t) = N_0 \exp \left(t \cdot \left[\frac{\bar{r} \cdot T}{k} \right] \right)$$

Donde k es la cantidad de subperiodos dentro de T y t es el sub-periodo para el que se quiere interpolar. Po lo tanto, dado que el mes de noviembre es el quinto mes después de junio de cada año, la expresión anterior queda como:

$$N(5) = N_{2019} \exp \left(\frac{5 \cdot \bar{r}}{12} \right)$$

Tabla 22. Población de 23 años y más a Noviembre de 2019, según región, sexo y grupo etario

Región	Hombres				Mujeres				TOTAL
	23-40	41-55	55+	TOTAL	23-40	41-55	55+	TOTAL	Pobl.
I / XV	96.131	57.448	51.688	205.267	90.912	56.438	59.250	206.600	411.867
II	109.037	66.379	51.251	226.667	105.325	63.811	57.993	227.129	453.796
III	44.355	30.421	29.980	104.756	42.255	28.793	33.083	104.131	208.887
IV	109.424	78.685	84.439	272.548	110.710	83.054	99.291	293.055	565.603
V	262.862	178.998	215.238	657.098	255.529	188.405	267.926	711.860	1.368.958
VI	129.123	102.951	110.154	342.228	125.319	101.713	121.351	348.383	690.611
VII	141.305	112.497	125.441	379.243	145.292	116.336	142.327	403.955	783.198
VIII / XVI	273.609	215.323	232.193	721.125	284.483	228.187	277.305	789.975	1.511.100
IX	126.293	98.658	108.938	333.889	132.561	101.556	126.573	360.690	694.579
X / XIV	168.778	136.999	134.479	440.256	170.600	133.601	153.600	457.801	898.057
XI / XII	40.831	28.938	30.072	99.841	39.735	27.742	31.462	98.939	198.780
RM	1.271.104	763.509	709.879	2.744.492	1.194.406	779.797	913.685	2.887.888	5.632.380
Total	2.772.852	1.870.806	1.883.752	6.527.410	2.697.127	1.909.433	2.283.846	6.890.406	13.417.816

Fuente: elaboración propia en base a proyecciones INE a Junio de 2019. Interpolación según Sadinle (2008).

³³ Véase Sadinle (2008).

En la **Tabla 23**, se puede observar la cantidad de observaciones disponibles en cada celda de post-estratificación para los cuestionarios de Vivos de la encuesta Presencial:

Tabla 23. Muestra de entrevistados Vivos de la encuesta Presencial, según región, sexo y grupo etario

<i>Región donde vive/vivía el entrevistado</i>	<i>Hombres</i>				<i>Mujeres</i>				<i>TOTAL</i>
	23-40	41-55	55+	TOTAL	23-40	41-55	55+	TOTAL	Pobl.
I / XV	38	12	23	73	37	22	35	94	167
II	46	26	43	115	52	31	49	132	247
III	39	24	32	95	44	25	44	113	208
IV	34	28	52	114	45	40	76	161	275
V	46	60	106	212	46	88	160	294	506
VI	91	36	60	187	100	47	75	222	409
VII	42	36	58	136	56	36	59	151	287
VIII / XVI	164	120	267	551	217	147	259	623	1.174
IX	35	22	40	97	44	25	57	126	223
X / XIV	122	94	143	359	136	132	159	427	786
XI / XII	15	13	24	52	29	21	38	88	140
RM	467	316	589	1.372	657	520	829	2.006	3.378
Total	1.139	787	1.437	3.363	1.463	1.134	1.840	4.437	7.800

Fuente: elaboración propia.

A organización de celdas (grupos etarios y agrupación de regiones) obedecen a la condición de que las celdas tuvieran, al menos, 10 observaciones disponibles. Esto implicó que se agruparan algunas regiones que administrativamente habían sido una sola región hasta hace algunos años. Nótese el caso de la Región de los Ríos que se suma a la Región de los Lagos (regiones X y XIV), o las regiones más extremas que también se agrupan. Vale la pena destacar que, al menos para los casos de I + XV regiones, VII + XVI, y X + XIV regiones, se encontraban como una unidad administrativa al momento de hacer los muestreos. Y en el caso de los nuevos entrevistados incorporados en la EPS VI Ronda, el muestreo siguió la agrupación (estratos y clusters) originales de 2002, por lo que la agrupación de las regiones es algo natural en la encuesta. En el caso de las comunas del sur, XI + XII, la agrupación se debió a la baja densidad poblacional de tales regiones, una práctica habitual en las estadísticas chilenas.

Por su parte, la muestra, después de los diversos ajustes explicados en las secciones anteriores, representaba la distribución de población informada en la **Tabla 24**.

Tabla 24. Muestra expandida de entrevistados Vivos de la encuesta Presencial, según región, sexo y grupo etario

Región donde vive/vivía el entrevistado	Hombres				Mujeres				TOTAL
	23-40	41-55	55+	TOTAL	23-40	41-55	55+	TOTAL	Pobl.
I / XV	87.687	72.100	68.904	228.691	59.887	64.977	87.440	212.304	440.995
II	158.581	68.329	93.187	320.097	88.845	76.214	112.968	278.027	598.123
III	34.334	34.000	36.794	105.128	33.593	26.950	62.634	123.177	228.305
IV	84.862	80.469	114.242	279.572	79.492	87.929	154.046	321.466	601.038
V	140.430	181.764	308.752	630.947	100.323	228.276	428.114	756.714	1.387.660
VI	220.225	91.118	113.129	424.473	182.670	116.304	240.049	539.023	963.496
VII	108.195	124.837	199.004	432.035	130.186	114.494	211.541	456.221	888.256
VIII / XVI	212.836	206.305	457.370	876.511	284.392	248.204	501.549	1.034.145	1.910.656
IX	146.116	111.391	178.445	435.953	116.906	127.517	227.929	472.352	908.305
X / XIV	202.185	137.465	189.904	529.553	147.893	171.016	226.459	545.368	1.074.921
XI / XII	59.821	28.918	37.532	126.272	42.135	41.335	76.330	159.801	286.073
RM	930.821	778.144	1.164.653	2.873.619	978.649	1.137.032	1.591.604	3.707.285	6.580.904
Total	2.386.092	1.914.842	2.961.916	7.262.849	2.244.971	2.440.249	3.920.663	8.605.883	15.868.732

Fuente: elaboración propia.

Al final del proceso, la realización del ajuste por post-estratificación es sencillo: se toman, celda a celda, las proporciones entre la población estimada por estrato, y la población representada por la muestra. Esto entrega un factor que multiplica a los factores de expansión de la muestra. Si la población representada en la muestra es más grande que la población real, el ajuste de post-estratificación será menor a 1, y al multiplicarlo por los factores de expansión de los entrevistados de la celda, les disminuirá el peso relativo en la muestra. Si la población representada es mayor a la ajustada por no respuesta, entonces el factor de post-estratificación será mayor que 1 e incrementará la representatividad de los individuos de la celda. En la **Tabla 25**, se puede ver el factor de post-estratificación para cada celda.

Tabla 25. Coeficientes de ajuste de post-estratificación para entrevistados Vivos de la encuesta Presencial, según región, sexo y grupo etario

Región donde vive/vivía el entrevistado	Hombres			Mujeres		
	23-40	41-55	55+	23-40	41-55	55+
I / XV	1,096301	0,796781	0,750145	1,518069	0,868580	0,677609
II	0,687580	0,971463	0,549981	1,185498	0,837256	0,513360
III	1,291866	0,894735	0,814806	1,257849	1,068392	0,528196
IV	1,289442	0,977833	0,739127	1,392720	0,944560	0,644556
V	1,871841	0,984780	0,697122	2,547058	0,825338	0,625828
VI	0,586322	1,129861	0,973703	0,686042	0,874542	0,505525
VII	1,306024	0,901154	0,630345	1,116031	1,016092	0,672811
VIII / XVI	1,285537	1,043710	0,507670	1,000319	0,919353	0,552898
IX	0,864335	0,885688	0,610484	1,133910	0,796410	0,555318
X / XIV	0,834772	0,996611	0,708142	1,153536	0,781219	0,678269
XI / XII	0,682548	1,000676	0,801234	0,943032	0,671143	0,412182
RM	1,365573	0,981192	0,609520	1,220464	0,685818	0,574065

Fuente: elaboración propia.

4.2.4. Factores de expansión finales de corte transversal

Después de aplicar los factores de post-estratificación de la **Tabla 25** a cada entrevistado, obtenemos los factores finales corregidos de corte transversal. A continuación, en la **Tabla 26**, se muestran estos factores para cada sub-muestra y para la muestra total, junto a las poblaciones representadas en cada *cluster*.

Tabla 26. Comportamiento de los factores de expansión de corte transversal de la encuesta Presencial, y de las poblaciones representadas por cada uno, según sub-muestras y muestra total

	Obs	Media	Desv. Std.	Min	Max
Afiliados 2002					
Población por cluster					
Base	4.505	99.633	58.111	3.140	283.247
Nominal	4.505	99.633	58.111	3.140	283.247
Efectiva	4.505	99.633	58.111	3.140	283.247
Ajustada por No Respuesta	4.505	112.119	62.454	1.058	278.090
Post-estratificada	4.505	180.141	96.715	3.615	409.881
Factores de Expansión					
Base	4.505	522,8	174,0	363,5	2.091,3
EPS I Ronda	4.505	457,5	155,8	301,0	1.837,0
Ajustados por Muestra Nominal	4.505	538,0	181,5	369,5	2.115,6

Ajustados por Muestra Efectiva	4.505	571,8	189,4	380,6	2.218,8
Ajustados por No Respuesta	4.505	1.791,9	1.282,3	662,2	14.624,9
Post-estratificados	4.505	1.350,9	1.164,6	370,3	12.478,9
Nuevos Afiliados 2004					
Población por cluster					
Base	245	5.877	5.018	282	45.633
Nominal	245	5.877	5.018	282	45.633
Efectiva	245	5.877	5.018	282	45.633
Ajustada por No Respuesta	245	7.615	5.880	418	31.771
Post-estratificada	245	170.453	95.188	2.385	409.881
Factores de Expansión					
Base	245	566,8	604,1	200,9	6.519,0
EPS II Ronda	245	617,3	287,8	307,3	2.853,7
Ajustados por Muestra Nominal	245	567,0	602,6	200,9	6.519,0
Ajustados por Muestra Efectiva	245	573,7	602,2	210,4	6.519,0
Ajustados por No Respuesta	245	2.140,4	2.677,1	342,4	18.866,4
Post-estratificados	245	2.192,1	3.113,3	220,7	21.692,7
No Afiliados 2004					
Población por cluster					
Base	805	46.568	22.340	4.409	99.044
Nominal	805	46.568	22.340	4.409	99.044
Efectiva	805	46.568	22.340	4.409	99.044
Ajustada por No Respuesta	805	55.626	32.327	3.124	155.563
Post-estratificada	805	168.936	91.963	6.982	409.881
Factores de Expansión					
Base	805	1.513,7	314,0	1.102,3	3.982,7
EPS II Ronda	805	867,2	365,9	392,7	3.250,6
Ajustados por Muestra Nominal	805	1.573,6	324,1	1.102,3	4.779,2
Ajustados por Muestra Efectiva	805	1.719,9	399,0	1.102,3	5.974,0
Ajustados por No Respuesta	805	4.995,7	4.069,7	1.913,0	49.310,4
Post-estratificados	805	3.718,2	4.452,5	983,3	67.337,0
Nuevos 2015					
Población por cluster					
Base	2.245	38.920	25.062	2.980	117.213
Nominal	2.245	38.920	25.062	2.980	117.213
Efectiva	2.245	38.920	25.062	2.980	117.213
Ajustada por No Respuesta	2.245	46.790	27.518	1.766	114.828
Post-estratificada	2.245	181.606	100.172	2.385	409.881
Factores de Expansión					
Base	2.245	475,2	111,3	262,8	1.490,0

EPS VI Ronda	2.245	498,8	156,1	216,8	1.561,2
Ajustados por Muestra Nominal	2.245	475,2	111,3	262,8	1.490,0
Ajustados por Muestra Efectiva	2.245	478,0	112,1	262,8	1.490,0
Ajustados por No Respuesta	2.245	1.447,8	584,4	615,4	3.911,5
Post-estratificados	2.245	1.693,4	835,2	644,4	6.587,7
Muestra Completa					
Población por cluster					
Ajustada por No Respuesta	7.800	84.203	60.835	418	278.090
Post-Estratificado	7.800	179.102	97.262	2.385	409.881
Factores de Expansión					
Ajustados por No Respuesta	7.800	2.034,5	2.004,5	342,4	49.310,4
Post-Estratificados	7.800	1.720,2	1.956,8	220,7	67.337,0
Fuente: elaboración propia.					

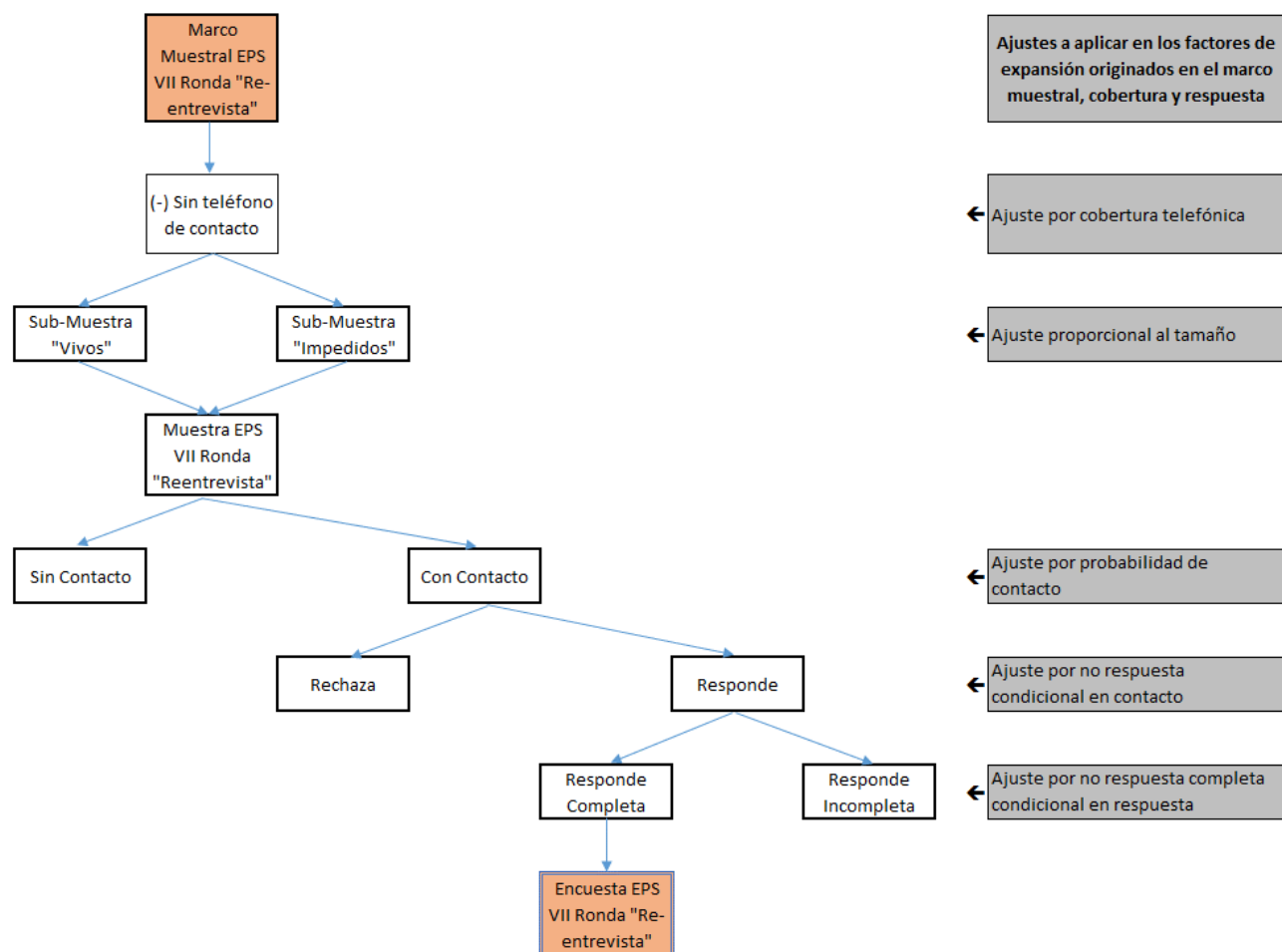
Nótese que cada uno de estos pasos se aplicó a cada sub-muestra detallada en la **Tabla 13**, a la que se remitió anteriormente. También vale la pena observar no sólo el comportamiento de la sub-muestra Afiliados 2002, que se ha usado como ejemplo para corroborar estos números, sino que también las otras tres sub-muestras y la muestra total de la encuesta Presencial. En resumen, los factores permitieron representar la población de Chile a noviembre de 2019 según lo mostrado en la última sección de la **Tabla 13**.

4.3. Factores de expansión de corte transversal de la encuesta de Re-entrevista

4.3.1. Configuración de la muestra

A continuación, se analizan los factores de expansión de *cross-section* de la encuesta de Re-Entrevista. La configuración de esta encuesta se muestra en la **Figura 18**, en donde, sobre el lado derecho, en cuadros sombreados con color gris, se observan los ajustes realizados según los diferentes aspectos que se deben corregir por la modalidad de realización de la misma.

Figura 18. Configuración de la muestra de la encuesta de Re-Entrevista de la EPS VII Ronda



Fuente: elaboración propia.

El procedimiento aplicado para el ajuste de los factores de expansión es similar a lo ya explicado para la encuesta Presencial. Sin embargo, hay algunos aspectos que destacan:

- **Factor de expansión base:** el factor de expansión base para la encuesta de Re-Entrevista, $FE_{Re-Entrevista}^{Base}$, corresponde al factor ajustado final de la encuesta Presencial, con el ajuste proporcional por tamaño por pérdida de la unidades correspondientes a fallecidos. Este factor final es el factor post-estratificado de la encuesta Presencial, que ya incorpora todos los ajustes para hacer de la muestra de la encuesta Presencial, una muestra representativa de la población del momento de la realización de la selección muestral de esta encuesta.
- **Corrección por cobertura telefónica:** dado que esta encuesta se aplicó a través de llamados telefónicos, se debe controlar por el sesgo de cobertura que se discutió en el diseño muestral (sección 2.4). Para realizar el ajuste por cobertura, se estima un modelo de propensión a tener un

número de teléfono de contacto basado en regresión logística para obtener la probabilidad de tener un teléfono de contacto registrado, $\Pr(\text{Tener teléfono})$. Por lo tanto, el factor de expansión base de la encuesta de Re-Entrevista (para cada una de las observaciones), ajustado por cobertura,³⁴ está dado por:

$$\widehat{FE}_{Re-Entrevista}^{Base} = \frac{FE_{Re-Entrevista}^{Base}}{\Pr(\text{Tener teléfono})}$$

Ajuste por no respuesta: los entrevistados con número de teléfono no necesariamente responderán la encuesta. Diversas son las razones de no respuesta, y en esta encuesta se ajusta por:

- la probabilidad de ser contactado, dado que se tiene un número de teléfono de potencial contacto, $\Pr(\text{Contacto})$
- la probabilidad de responder la encuesta, dado que fue contactado, $\Pr(\text{Responder}|\text{Contacto})$
- la probabilidad de responder completa la encuesta,³⁵ dado que se comenzó a responder la misma, $\Pr(\text{Responder completa}|\text{Responder}, \text{Contacto})$

Así, el factor de expansión ajustado por no respuesta para esta encuesta estará dado por:

$$\begin{aligned} \widehat{FE}_{Re-Entrevista}^{Ajustado} \\ = \widehat{FE}_{Re-entrevista}^{Base} \cdot \Pr(\text{Contacto})^{-1} \cdot \Pr(\text{Responder}|\text{Contacto})^{-1} \\ \cdot \Pr(\text{Respuesta completa}|\text{Responder}, \text{Contacto})^{-1} \end{aligned}$$

- **Sub-muestras longitudinales:** en la encuesta de Re-Entrevista se abandona el esquema de muestras paralelas que se aplicó en encuesta Presencial, y que ha sido usado en las anteriores olas de la EPS. Dos buenas razones conducen a esta decisión: en primer lugar, no habría cantidad suficiente de observaciones para muchas celdas del muestreo original. Si bien esto se puede solucionar estadísticamente con pseudo-estratos y pseudo-clusters, estaríamos agregando mayores distorsiones y un cambio a la estrategia de muestreo original de esta encuesta. La segunda razón va en línea con esta última idea: la encuesta de Re-Entrevista fue concebida como un panel corto para análisis antes-después de la pandemia por COVID-19. Por lo tanto, no es necesario mantener las muestras separadas para lograr el cálculo de eventuales factores longitudinales que nos remonten

³⁴ Nótese que tener/no tener un teléfono de contacto no altera el marco muestral, sino que obliga a un ajuste por cobertura telefónica. Esto se debe a que el marco muestral se realizó con todos los entrevistados que no respondieron la encuesta Presencial, y para cada uno de ellos, se buscó un medio de contacto telefónico. Por lo tanto, el marco no fue condicionado en la posibilidad de localización telefónica. Véase Cobben y Bethlehem (2005) y Cobben (2009).

³⁵ En las encuestas telefónicas, es bastante habitual encontrar situaciones en las que, inicialmente se acepta realizar la encuesta, pero ésta es abandonada antes de ser finalizada. Esto suele ser menos habitual en las entrevistas cara a cara, en la que la presencia física tiende a impedir un abrupto quiebre de la entrevista. Véase el Recuadro 1, en que se muestran que las tasas de abandono pueden atentar contra el éxito de una encuesta, aun cuando la tasa de respuesta inicial sea alta.

a olas anteriores de la EPS. Por lo tanto, a los fines del tratamiento estadístico de los ajustes de factores, se considera que la muestra de la encuesta de Re-Entrevista es única.

- **Sub-muestras de corte transversal:** El siguiente aspecto a tener en cuenta es que la muestra contiene dos sub-muestras de *cross-section*: (i) la muestra de “impedidos”, de inclusión forzosa, que sobre-representa al grupo en el levantamiento; y (ii) la muestra de “vivos” que es una muestra probabilística obtenida a partir del marco muestral de la encuesta Presencial. Los factores de expansión ajustados se aplicarán a la muestra de Vivos, que es la muestra que tiene un marco y diseño muestral que permite analizar los ponderadores.

Siguiendo la misma estructura y contenido que el expuesto en la **Tabla 3** y **Tabla 13**, en la **Tabla 27** se muestra el comportamiento de las encuestas de Re-Entrevista realizadas a quienes respondieron el cuestionario de Vivos en la encuesta Presencial. Dentro de este grupo, se dieron 4 decesos y 28 entrevistados que pasaron a tener la condición de Impedidos. A cada uno de ellos se le aplicó un cuestionario especialmente diseñado para la circunstancia, que fue respondido por una persona conocedora del entrevistado fallecido o impedido.

Nótese que se obtuvieron 2.082 observaciones (cuestionarios administrados a Vivos) que forman panel con la encuesta Presencial y que pueden ser empleados en el análisis antes-después de corto plazo de los efectos de la pandemia en Chile. Para esta sub-muestra, se estimaron factores de expansión longitudinales, según se detalla en la sección 5.2.

Las secciones finales de la **Tabla 27** muestran la poblaciones representadas en el momento de la selección, después del ajuste por no respuesta, y tras la post-estratificación. Se volverá sobre estos totales más adelante.

Tabla 27. Encuesta de Re-Entrevista a Vivos: comportamiento de la muestra

Tipo de Muestra	Afiliados 2002	Nuevos Afiliados 2004	No Afiliados 2004	Nuevos 2015	Total
Muestra Total	4.505	245	805	2.245	7.800
Muestra con Cobertura Telefónica	4.473	245	800	2.245	7.763
Muestra Nominal	3.197	195	604	1.573	5.569
Muestra Efectiva	3.174	195	596	1.572	5.537
Contactado	1.480	93	280	670	2.523
Responde Contactado	1.239	74	246	580	2.139
Responde Completa Responde, Contactado	1.224	74	243	573	2.114
Vivos	3.174	195	596	1.572	5.537
Contactado	1.457	93	272	669	2.491
Responde Contactado	1.216	74	238	579	2.107
Responde Completa Responde, Contactado	1.201	74	235	572	2.082
Fallecidos	4	0	0	0	4
Contactado	4	0	0	0	4
Responde Contactado	4	0	0	0	4
Responde Completa Responde, Contactado	4	0	0	0	4
Impedidos	19	0	8	1	28
Contactado	19	0	8	1	28
Responde Contactado	19	0	8	1	28
Responde Completa Responde, Contactado	19	0	8	1	28
Total Encuestas a VIVOS	1.224	74	243	573	2.114
Población representada originalmente	8.074.667	527.914	3.994.879	3.250.561	15.848.021
Factor de expansión base promedio	6596,950	7133,973	16439,831	5672,881	7496,699
Población representada con ajuste por NR	6.010.759	466.318	3.149.549	3.814.800	13.441.426
Factor de expansión ajustado por NR promedio	5004,795	6301,596	13402,336	6669,231	6456,016
Población representada post-estratificada	6.165.145	460.681	3.148.114	3.866.617	13.640.557
Factor de expansión post-estratificado promedio	5133,343	6225,414	13396,230	6759,820	6551,660

Nota: la encuesta de Re-Entrevista a Vivos se refiere a la sub-muestra de esta encuesta que fue obtenida de la muestra de entrevistados que respondieron el cuestionario de Vivos en la encuesta Presencial. Dado el paso del tiempo y las circunstancias, algunos de estos entrevistados fallecieron y otros pasaron a una situación de impedimento. Esa es la razón por la que se administraron cuestionarios de Fallecidos e Impedidos, además de los de Vivos, en esta sub-muestra.

Fuente: Elaboración propia.

Por su parte, el tratamiento de la muestra de impedidos permitió obtener un total de 111 cuestionarios completos que incluyen a tres fallecidos. Esto se puede observar en la **Tabla 28**. Los 108 cuestionarios a Impedidos obtenidos en esta sub-muestra, sumados a los 28 obtenidos entre los administrados a la sub-muestra de Vivos, permitió obtener 136 cuestionarios para estudiar las particularidades de este grupo de la población.

Tabla 28. Encuesta de Re-Entrevista a Impedidos: comportamiento de la muestra

Tipo de Muestra	Afiliados 2002	Nuevos Afiliados 2004	No Afiliados 2004	Nuevos 2015	Total
Muestra Total	245	2	64	26	337
Muestra con Cobertura Telefónica	245	2	64	26	337
Muestra Nominal	245	2	64	26	337
Muestra Efectiva	245	2	64	26	337
Contactado	95	1	28	12	136
Responde Contactado	83	0	20	9	112
Responde Completa Responde, Contactado	82	0	20	9	111
Fallecidos	2	0	1	0	3
Contactado	2	0	1	0	3
Responde Contactado	2	0	1	0	3
Responde Completa Responde, Contactado	2	0	1	0	3
Impedidos	80	0	19	9	108
Contactado	80	0	19	9	108
Responde Contactado	80	0	19	9	108
Responde Completa Responde, Contactado	80	0	19	9	108
Total Encuestas a IMPEDIDOS	82	0	20	9	111

Nota: la encuesta de Re-Entrevista a Impedidos se refiere a la sub-muestra de esta encuesta que fue obtenida de la muestra de entrevistados que respondieron el cuestionario de Impedidos en la encuesta Presencial. Dado el paso del tiempo y las circunstancias, algunos de estos entrevistados fallecieron. Esa es la razón por la que se administraron cuestionarios de Fallecidos, además de los de Impedidos, en esta sub-muestra.

Fuente: elaboración propia.

4.3.2. Ajustes por cobertura telefónica

Para realizar los ajustes a los factores de expansión finales, se siguió el procedimiento descrito a continuación:

PASO 1: Se tomaron los factores de post-estratificación de la encuesta Presencial como factor base para esta encuesta. Dado que el marco muestral de la encuesta de Re-Entrevista es el conjunto de no fallecidos de la encuesta Presencial, estos factores son los adecuados como punto de partida para realizar la construcción de los factores finales.

PASO 2: El primer ajuste realizado corresponde a un ajuste por cobertura telefónica. Por lo tanto, dado que ese es el canal de implementación de la encuesta, se estimó la probabilidad de tener un número de contacto a través del cual acceder al entrevistado. Para realizar este ajuste, se estimaron varias especificaciones de modelos de *propensity score* empleando regresiones logísticas. De las 7.800 observaciones disponibles, 37 no poseían teléfono.

Los efectos marginales del modelo elegido se muestran en la **Tabla 29**.

Tabla 29. Estimación de la probabilidad de tener cobertura telefónica en la encuesta de Re-Entrevista, sub-muestra de Vivos

Logistic regression			Number of obs	=	7,800	
			Wald chi2(16)	=	118.12	
			Prob > chi2	=	0.0000	
Log pseudolikelihood = -340648.61			Pseudo R2	=	0.1780	

d_TieneFono	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	

sexo_1	-.7958118	.4246107	-1.87	0.061	-1.628033	.0364098
region_2	-.2664719	.8186372	-0.33	0.745	-1.870971	1.338028
region_4	.259609	1.069394	0.24	0.808	-1.836365	2.355583
region_5	-.6923727	.663598	-1.04	0.297	-1.993001	.6082554
region_6	-.6088129	.9533012	-0.64	0.523	-2.477249	1.259623
region_7	-.0458924	1.074347	-0.04	0.966	-2.151575	2.05979
region_8	-.0429782	.7375484	-0.06	0.954	-1.488547	1.40259
region_10	-1.237641	.5189209	-2.39	0.017	-2.254707	-.2205748
region_12	-1.413371	1.065522	-1.33	0.185	-3.501755	.6750139
t_educ_2	.1873316	.4833682	0.39	0.698	-.7600527	1.134716
grupo_etario_2020_4	-1.496586	1.270822	-1.18	0.239	-3.987351	.9941793
grupo_etario_2020_5	-2.038621	1.092929	-1.87	0.062	-4.180722	.1034794
grupo_etario_2020_6	-2.676714	1.099801	-2.43	0.015	-4.832284	-.5211432
grupo_etario_2020_7	-3.957627	1.042606	-3.80	0.000	-6.001098	-1.914156
pensionado						
Si	.2561915	.5919859	0.43	0.665	-.9040795	1.416462
cotizando						
Si	.7489814	.5790804	1.29	0.196	-.3859953	1.883958
_cons	7.88094	1.106731	7.12	0.000	5.711788	10.05009

Fuente: elaboración propia.

La distribución de la probabilidad predicha de poseer teléfono se muestra en la **Tabla 30**.

Tabla 30. Distribución de la probabilidad estimada de tener teléfono en la encuesta de Re-Entrevista, sub-muestra Vivos

Pr(d_TieneFono)				
	Percentiles	Smallest		
1%	.9365125	.8687396		
5%	.9781102	.8687396		
10%	.9849247	.8687396	Obs	7,800
25%	.9946953	.8687396	Sum of Wgt.	7,800
50%	.9982345		Mean	.9943508
		Largest	Std. Dev.	.0119564
75%	.9993844	.9998857		
90%	.9996868	.9998857	Variance	.000143
95%	.9998214	.9998857	Skewness	-5.540856
99%	.9998519	.9998857	Kurtosis	45.28061

Fuente: elaboración propia.

Para corregir los factores de expansión base por cobertura telefónica, se empleó el método directo de ponderación por propensity. Este método fue el que mostró las mejores características estadísticas. Posiblemente se deba a la baja proporción de casos en que había teléfono disponible, por lo que la corrección realizada por este ajuste es muy pequeña.

4.3.3. Ajustes por atrición y no respuesta

Los siguientes pasos son similares a los PASOS 2 a 9 de la encuesta Presencial, aunque sólo se aplican a una muestra. Se evita la redundancia de su presentación y comentarios, directamente al último paso, referido a la post-estratificación.

Nótese que las estimaciones para el ajuste por no respuesta corresponden, ahora, a la probabilidad de ser contactado, a la probabilidad de responder, condicional en ser contactado, y a la probabilidad de responder completa la encuesta, condicional en responder y haber sido contactado. La metodología empleada es idéntica a la descrita para la encuesta Presencial.

Por lo tanto, los factores de expansión ajustados por atrición y no respuesta en la encuesta de Re-Entrevista, se obtienen calculando:

$$\begin{aligned}
 FE_{Re-Entrevista}^{Ajustado} &= FE_{Re-entrevista}^{Base} \cdot \Pr(\text{Contacto})^{-1} \cdot \Pr(\text{Responder}|\text{Contacto})^{-1} \\
 &\quad \cdot \Pr(\text{Respuesta completa}|\text{Responder, Contacto})^{-1}
 \end{aligned}$$

4.3.4. Post-estratificación de los factores de expansión de corte transversal

PASO 10: los factores de expansión ajustados por no respuesta aún no logran representar a la población contingente del momento de la encuesta. Por lo tanto, se realiza un ajuste adicional de calibración por celdas, es decir, post-estratificación por sexo, grupo etario y macrozonas. Los grupos etarios son similares a los de la encuesta Presencial, y las regiones se agruparon en macrozonas para lograr mayor densidad y mantener la cobertura nacional. Las macrozonas se definieron como:

- Norte: regiones 15, 1, 2, 3 y 4.
- Centro: 5, 6, 7, 16 y 8.
- Sur: 9, 14, 10, 11 y 12.
- R. Met.: 13.

Los totales poblacionales se tomaron de las proyecciones censales del INE, disponibles públicamente. La fecha que se tomó como referencia fue el mes de julio de 2020, periodo anterior al inicio del levantamiento, y los totales poblacionales para los que se post-estratificó se muestran en la **Tabla 31**. La interpolación se hizo con la misma metodología de la encuesta Presencial, siguiendo a Sadinle (2008).

Tabla 31. Proyección de población de 23 años y más a Julio de 2020, según macrozona, sexo y grupo etario

Macrozona	Hombres				Mujeres				TOTAL
	23-40	41-55	56+	TOTAL	23-40	41-55	56+	TOTAL	Pobl.
Norte	366.519	236.149	222.820	825.488	357.178	235.166	255.821	848.165	1.673.653
Centro	811.464	611.509	700.163	2.123.136	815.794	636.327	827.676	2.279.797	4.402.933
Sur	336.988	265.166	280.325	882.479	344.512	263.748	318.438	926.698	1.809.177
R. Met.	1.304.421	775.718	727.380	2.807.519	1.225.462	788.746	933.065	2.947.273	5.754.792
Total	2.819.392	1.888.542	1.930.688	6.638.622	2.742.946	1.923.987	2.335.000	7.001.933	13.640.555

Nota: Interpolación según Sadinle (2008). Macrozonas: Norte: regiones 15, 1, 2, 3 y 4. Centro: 5, 6, 7, 16 y 8. Sur: 9, 14, 10, 11 y 12. R. Met.: 13.

Fuente: elaboración propia en base a proyecciones INE a Junio de 2020.

En la **Tabla 32**, se puede observar la cantidad de observaciones disponibles en cada celda de post-estratificación. Con la agrupación por macrozonas, se evitó tener celdas con menos de 10 observaciones, para evitar exacerbaciones de los factores de expansión.

Tabla 32. Muestra de entrevistados Vivos de la encuesta de Re-Entrevista, según macrozona, sexo y grupo etario

Macrozona geográfica	Hombres				Mujeres				TOTAL
	23-40	41-55	56+	TOTAL	23-40	41-55	56+	TOTAL	Pobl.
Norte	29	20	35	84	48	38	58	144	228
Centro	72	54	127	253	134	101	177	412	665
Sur	29	21	45	95	54	49	85	188	283
R. Met.	94	80	139	313	196	148	249	593	906
Total	224	175	346	745	432	336	569	1.337	2.082

Fuente: elaboración propia.

Por su parte, la muestra, después de los diversos ajustes, estaba representando la siguiente distribución de población:

Tabla 33. Muestra expandida de entrevistados Vivos de la encuesta de Re-Entrevista, según macrozona, sexo y grupo etario

Macrozona geográfica	Hombres				Mujeres				TOTAL
	23-40	41-55	56+	TOTAL	23-40	41-55	56+	TOTAL	Pobl.
Norte	336.094	174.557	270.854	781.504	345.268	268.454	272.963	886.686	1.668.190
Centro	807.578	536.932	706.093	2.050.603	887.897	663.635	768.156	2.319.688	4.370.290
Sur	361.100	236.056	306.604	903.760	338.375	235.808	340.084	914.267	1.818.027
R. Met.	1.208.679	800.625	724.973	2.734.277	1.217.076	632.869	1.000.696	2.850.642	5.584.919
Total	2.713.451	1.748.170	2.008.524	6.470.144	2.788.616	1.800.767	2.381.900	6.971.282	13.441.426

Fuente: elaboración propia.

En la **Tabla 34**, se puede ver el factor de expansión de post-estratificación para cada celda. Este factor multiplica al factor de expansión de cada observación para conseguir los totales poblacionales post-estratificados.

Tabla 34. Coeficientes de ajuste de post-estratificación para entrevistados Vivos de la encuesta de Re-Entrevista, según macrozona, sexo y edad

Macrozona geográfica	Hombres			Mujeres		
	23-40	41-55	56+	23-40	41-55	56+
Norte	1,090526	1,352850	0,822658	1,034495	0,876000	0,937199
Centro	1,004812	1,138894	0,991602	0,918794	0,958851	1,077484
Sur	0,933225	1,123320	0,914291	1,018138	1,118485	0,936350
R. Met.	1,079212	0,968891	1,003320	1,006890	1,246302	0,932416

Fuente: elaboración propia.

4.3.5. Factores de expansión finales de corte transversal

Después de aplicar los factores de post-estratificación de la tabla anterior a cada entrevistado, se obtienen los factores finales corregidos de corte transversal para la encuesta de Re-Entrevista. En la **Tabla 35**, se muestra la descripción estadística de estos factores para la muestra total, junto a la población por cluster, como medidas informativas del comportamiento de los factores de expansión finales de la encuesta.

Tabla 35. Comportamiento de los factores de expansión de corte transversal de la encuesta de Re-Entrevista, y de las poblaciones representadas por cada uno					
	Obs	Media	Desv. Std.	Min	Max
Población por cluster					
Base	2.082	74.268	57.334	1.538	283.247
Ajustada por Cobertura Telefónica	2.082	178.273	100.812	2.386	411.106
Muestra Nominal	2.082	178.273	100.812	2.386	411.106
Muestra Efectiva	2.082	178.273	100.812	2.386	411.106
Ajustada por No Respuesta	2.082	179.972	112.738	2.310	500.149
Post-estratificada	2.082	183.390	116.814	2.215	519.379
Factores de Expansión					
Base	2.082	621,6	364,9	200,9	2.571,6
Base de encuesta de Re-Entrevista	2.082	1.661,6	2.227,1	334,5	67.337,0
Ajustados por Cobertura Telefónica	2.082	1.669,8	2.231,7	335,9	67.383,8
Ajustados por Muestra Nominal	2.082	2.292,9	2.781,9	369,9	82.908,3
Ajustados por Muestra Efectiva	2.082	2.304,3	2.790,0	369,9	82.908,3
Ajustados por No Respuesta	2.082	6.456,0	9.539,7	715,5	293.283,9
Post-estratificados	2.082	6.551,7	9.980,8	701,2	316.515,5
Fuente: elaboración propia.					

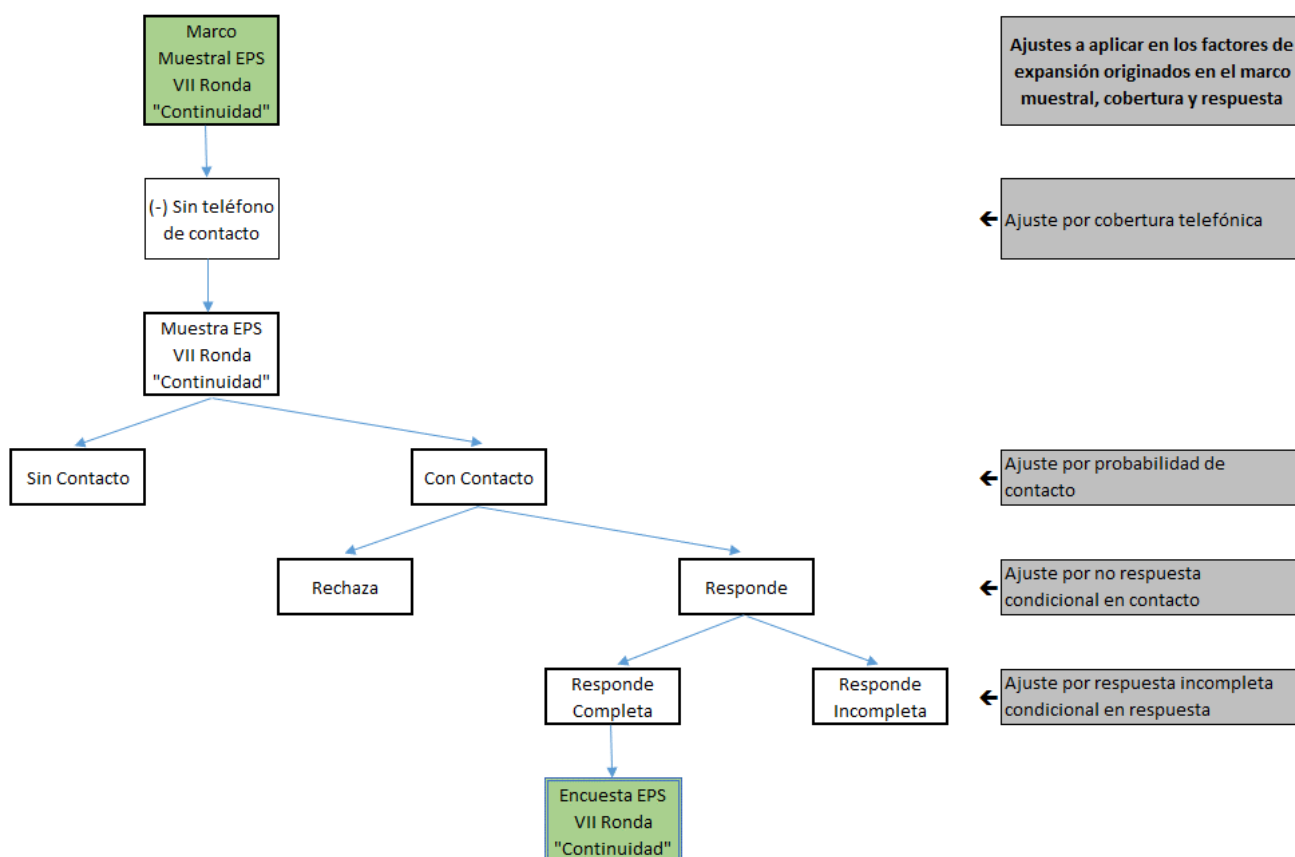
En la parte final de la **Tabla 27**, se puede observar el comportamiento que tuvieron los factores de expansión, teniendo en cuenta las muestras originales de las que provenían.

4.4. Factores de expansión de corte transversal de la encuesta de Continuidad

4.4.1. Configuración de la muestra de la encuesta de Continuidad

En la Figura 3, se muestra esquemáticamente cómo se espera que se desempeñe la muestra de la encuesta de continuidad. Al lado derecho, en los cuadros sombreados de color gris, nuevamente se presentan los ajustes sugeridos a los factores de expansión según el detalle a continuación.

Figura 19. Configuración de la muestra de la encuesta de Continuidad de la EPS VII Ronda



Fuente: elaboración propia.

La muestra de esta encuesta se desprende de todos los entrevistados que no respondieron la encuesta Presencial, como se explicó en la sección 2.6. Por lo tanto, la muestra requiere un conjunto de ajustes que tenga en cuenta el origen de la muestra, la cobertura telefónica, la no respuesta, y un ajuste final para que la encuesta represente la población contingente.

La **Tabla 36** muestra el comportamiento de la muestra de esta encuesta. Se tuvieron a disposición 16.227 observaciones, las cuales provenían de tres grupos mutuamente excluyentes que se originaron en la aplicación de la encuesta Presencial: 10.695 no fueron visitados en la encuesta presencial, 4.812 fueron visitados pero no llegaron a completar el conjunto de visitas y protocolos para cerrar con una respuesta o un rechazo a la encuesta Presencial, y finalmente, 720 rechazaron la realización de la encuesta presencial, y fueron incorporados a la muestra de la encuesta de Continuidad. De este modo, la muestra de la encuesta de Continuidad involucra a todos los entrevistados del marco muestral de la EPS VII Ronda que no respondieron la encuesta Presencial.

De este conjunto de 16.227 entrevistados, 15.506 tenían algún número de contacto telefónico, como fue explicado en la sección 2.6. Un total de 6.696 entrevistados lograron ser contactados, de los cuales 5.325 aceptaron responder la encuesta. 29 de ellos abandonaron la entrevista antes de finalizar completamente el cuestionario, por lo que, finalmente, se obtuvieron 5.296 encuestas que incluyeron 5.031 cuestionarios de vivos, 103 cuestionarios de fallecidos, y 162 cuestionarios de impedidos.

Tabla 36. Encuesta de Continuidad: Comportamiento de la muestra

Tipo de Muestra	Afiliados 2002	Nuevos Afiliados 2004	No Afiliados 2004	Nuevos 2015	Total
Muestra Nominal	9.478	665	1.584	4.500	16.227
No visitados en Presencial	6.168	442	1.127	2.958	10.695
No revisitados en Presencial	2.795	205	409	1.403	4.812
Rechazos en Presencial	515	18	48	139	720
Muestra con Cobertura Telefónica	8.961	647	1.416	4.482	15.506
Muestra Efectiva	8.730	636	1.379	4.464	15.209
Contactado	3.848	282	598	1.968	6.696
Responde Contactado	2.952	222	488	1.663	5.325
Responde Completa Responde, Contactado	2.932	221	486	1.657	5.296
Vivos	8.544	631	1.329	4.440	14.944
Contactado	3.662	277	548	1.944	6.431
Responde Contactado	2.766	217	438	1.639	5.060
Responde Completa Responde, Contactado	2.746	216	436	1.633	5.031
Fallecidos	85	2	14	2	103
Impedidos	101	3	36	22	162
Total Encuestas	2.932	221	486	1.657	5.296
Población representada originalmente	8.074.667	527.914	3.994.879	3.250.561	15.848.021
Factor de expansión base promedio	2753,979	2388,751	8219,916	1961,715	2992,451
Población representada con ajuste por NR	8.001.575	531.875	3.993.678	3.214.477	15.741.605
Factor de expansión ajustado por NR promedio	2913,902	2462,384	9159,812	1968,449	3128,922
Población representada post-estratificada	6.261.053	592.724	2.978.394	3.828.663	13.660.834
Factor de expansión post-estratificado promedio	2280,063	2744,092	6831,179	2344,558	2715,332

Fuente: elaboración propia.

A continuación, se explican los pasos más relevantes en el ajuste de los factores de expansión de la encuesta de Continuidad. Como se verá, algunos de ellos coinciden con los ya explicados para la encuesta Presencial, por lo que no serán discutidos en profundidad nuevamente.

4.4.2. Ajustes por selección de la muestra

Dado que la muestra de la encuesta de Continuidad es el complemento de la muestra de la encuesta Presencial, usa los mismos fundamentos, es decir, los factores de expansión base son los mismos que los

de la encuesta Presencial. A partir de estos factores de expansión, se realizan un conjunto de ajustes que se detallan a continuación, comenzando con el ajuste por selección de la muestra.

Tal como se hizo con la muestra de la encuesta Presencial, cada sub-muestra se trata por separado, manteniendo el esquema general de ajustes. Así, se realizan ajustes separados para las sub-muestras de Afiliados 2002, Nuevos Afiliados 2004, No Afiliados 2004, y Nuevos 2015.

Cada una de las sub-muestras, a su vez, proviene de tres posibles fuentes diferentes y mutuamente excluyentes, dependiendo de la situación de los entrevistados en la encuesta Presencial:³⁶

- Entrevistados que no fueron visitados durante la encuesta Presencial.
- Entrevistados que fueron visitados, pero no se pudo dar un cierre a su situación de participación en la encuesta Presencial, es decir, no se obtuvo finalmente una respuesta o un rechazo a la realización del cuestionario.
- Entrevistados que rechazaron realizar la encuesta Presencial.

El ajuste por selección de la muestra se realiza para incorporar el control de los posibles sesgos a los que estos tres grupos de entrevistados estuvo sujeto. Entonces, a los factores de expansión base correspondientes al marco muestral de esta encuesta, se les deben practicar los ajustes correspondientes a los tres grupos.

Para realizar este ajuste, se debe tener en cuenta lo siguiente:

- Para los entrevistados que nunca fueron visitados en la encuesta Presencial: la probabilidad de no haber sido visitado es el complemento de la probabilidad de haber sido visitado que se estimó para la encuesta presencial. Por lo tanto, se puede estimar la probabilidad de ser visitado en la encuesta Presencial y emplear esa probabilidad para estimar el ajuste por no haber sido visitado en la encuesta Presencial:

$$FE_{Continuidad}^{Base, \widehat{No\ Visitado}} = FE_{Presencial}^{Base} \cdot [1 - \Pr(Visitado)]^{-1}$$

- Para los entrevistados no revisitados en la encuesta Presencial: se trata del complemento de aquellos que, habiendo sido visitados en la encuesta Presencial, no llegaron a responder o rechazar la encuesta Presencial. Por lo tanto, condicional en haber sido visitado en la encuesta Presencial, la probabilidad de no ser revisitado es el complemento de la probabilidad de haber respondido o rechazado la encuesta Presencial:

$$\begin{aligned} FE_{Continuidad}^{Base, \widehat{No\ Revisitado}} &= FE_{Presencial}^{Base} \cdot [1 - \Pr(Rechazar \cup Responder | Visitado)]^{-1} \\ &\quad \cdot \Pr(Visitado)^{-1} \end{aligned}$$

³⁶ Esto implica que, al igual que las sub-muestras que conforman la EPS VII Ronda, no hay superposición de marcos muestrales entre los tres grupos de entrevistados que conforman cada sub-muestra de la encuesta de Continuidad.

- Para los entrevistados que rechazaron la encuesta Presencial: estos entrevistados fueron visitados nuevamente y se consiguió realizar conversiones exitosas de rechazos (25% de los casos).³⁷ Por lo tanto, la probabilidad de haber rechazado la encuesta Presencial, condicional en haber respondido o rechazado, y en haber sido visitado, es el complemento de la probabilidad de responder la encuesta Presencial, con los mismos condicionantes, es decir:

$$\begin{aligned} FE_{Continuidad}^{Base,Rechazo} &= FE_{Presencial}^{Base} \\ &\cdot [1 - \Pr(\text{Responder} | \text{Responder} \cup \text{Rechazar}, \text{Visitado})]^{-1} \\ &\cdot \Pr(\text{Rechazar} \cup \text{Responder} | \text{Visitado})^{-1} \cdot \Pr(\text{Visitado})^{-1} \end{aligned}$$

Para ajustar por la probabilidad de ser parte de alguno de estos tres grupos dentro la muestra de la encuesta de Continuidad, se estimaron modelos de propensión (i) a ser visitado en la encuesta Presencial, (ii) a responder o rechazar la encuesta Presencial, condicional en haber sido visitado, y (iii) a responder la encuesta Presencial, condicional en responder o rechazar esa encuesta y haber sido visitado. Estos modelos se estimaron con regresiones logísticas sobre la muestra nominal de la encuesta Presencial.³⁸

Luego, se realizaron los ajustes, como se explicó anteriormente, para obtener $FE_{Continuidad}^{Base,No Visitado}$, $FE_{Continuidad}^{Base,No Revisitado}$ y $FE_{Continuidad}^{Base,Rechazo}$.

4.4.3. Ajustes por cobertura telefónica

Dado que la encuesta de continuidad se aplicó a través de llamados telefónicos, se debe ajustar por la cobertura telefónica, como se discutió en la encuesta de Re-Entrevista. Tal como allí, dado que el marco muestral no condicionó en la posesión de un teléfono de contacto, la forma de preceder es realizar un ajuste a los factores de expansión por la probabilidad de tener registrado un número de contacto telefónico.

La literatura sugiere que, para resolver este problema, se estime un modelo de propensión a tener teléfono para corregir por las posibles diferencias sistemáticas entre quienes tienen registrado un número telefónico

³⁷ De 720 entrevistados que rechazaron la encuesta Presencial, 705 tenían teléfono de contacto. De éstos, se intentó contactar a 694, de los cuáles 299 fueron contactados exitosamente. 177 de ellos aceptaron realizar la encuesta, pero 3 abandonaron la entrevista antes de finalizar el cuestionario. En definitiva, 174 entrevistados sobre 705 que podían ser contactados, respondieron la encuesta de Continuidad, siendo que habían rechazado participar en la encuesta Presencial.

³⁸ Nótese que, al contrario de las estimaciones de ajustes por no respuesta en la encuesta Presencial, que se realizan sobre la muestra efectiva (es decir, la muestra nominal excluidos los fallecidos e impedidos), aquí se mantienen esas observaciones dentro de la muestra. Esto se debe a que las probabilidades de selección en la muestra de la encuesta de Continuidad no están condicionadas por el fallecimiento o el impedimento que fuera información obtenida en la otra encuesta. Es decir, la información acerca del fallecimiento o impedimento de un entrevistado no altera el marco muestral de la encuesta de Continuidad. Y, por lo tanto, esa Información no debe ser incorporada en los ajustes.

públicamente (en alguna fuente de información) y quienes no cuentan con teléfono, comparten línea con otras personas y no lo reportan como propio, o niegan la información de contacto.³⁹

Para realizar el ajuste, se estimó un modelo de propensión a tener un número de teléfono de contacto basado en regresión logística para obtener la probabilidad de tener un teléfono de contacto registrado, $Pr(\text{Tener teléfono})$. La **Tabla 37** muestra los resultados de la estimación de la especificación del modelo elegido.

Tabla 37. Estimación de la probabilidad de tener cobertura telefónica en la encuesta de Continuidad

Logistic regression			Number of obs		=	1,584
			Wald chi2(23)		=	127.22
			Prob > chi2		=	0.0000
Log pseudolikelihood = -813010.24			Pseudo R2		=	0.2645

d_TieneFono	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	

grupo_etario_2020_2	-1.389325	.987633	-1.41	0.160	-3.32505	.5464001
grupo_etario_2020_3	-1.475518	.9398901	-1.57	0.116	-3.317669	.3666324
grupo_etario_2020_4	-1.595308	.9306439	-1.71	0.086	-3.419337	.2287205
grupo_etario_2020_5	-2.074709	.9156486	-2.27	0.023	-3.869348	-.2800711
grupo_etario_2020_6	-3.228812	.9148073	-3.53	0.000	-5.021801	-1.435822
region_1	-.7363825	.8041796	-0.92	0.360	-2.312546	.8397806
region_2	-.6496028	.6101094	-1.06	0.287	-1.845395	.5461897
region_4	.4124505	.5525732	0.75	0.455	-.670573	1.495474
region_5	.1974071	.4855424	0.41	0.684	-.7542385	1.149053
region_6	.1333806	.677078	0.20	0.844	-1.193668	1.460429
region_7	-1.086619	.3423689	-3.17	0.002	-1.75765	-.4155882
region_8	.1051754	.3553512	0.30	0.767	-.5913002	.801651
region_9	.4350826	.4570288	0.95	0.341	-.4606774	1.330843
region_10	-.9019867	.468218	-1.93	0.054	-1.819677	.0157037
region_11	-1.707542	.7385482	-2.31	0.021	-3.15507	-.2600146

sexo						
Hombre	-.4466563	.251267	-1.78	0.075	-.9391306	.0458181

t_educ_2	-.185355	.264673	-0.70	0.484	-.7041045	.3333946
t_educ_3	1.432071	.7410877	1.93	0.053	-.0204344	2.884576
t_educ_4	1.751333	.7376245	2.37	0.018	.3056152	3.19705
ultima_EPS_2	-2.73942	.4735624	-5.78	0.000	-3.667586	-1.811255
ultima_EPS_3	-2.522068	.3281714	-7.69	0.000	-3.165272	-1.878864
ultima_EPS_4	-2.397543	.3196714	-7.50	0.000	-3.024088	-1.770999

pensionado						
Si	-.8193531	.53414	-1.53	0.125	-1.866248	.2275421

_cons	6.281462	.9872668	6.36	0.000	4.346454	8.216469

Fuente: elaboración propia.

Por su parte, la **Tabla 38** muestra la distribución de la probabilidad estimada de poseer algún número de contacto.

³⁹ Para mayor detalle sobre la conveniencia y procedimiento de ajuste, véase Cobben y Bethlehem (2005), y Cobben (2009, capítulo 7).

Tabla 38. Distribución de la probabilidad estimada de cobertura telefónica en la encuesta de Continuidad

Pr(d_TieneFono)				
	Percentiles	Smallest		
1%	.3644649	.1195905		
5%	.6296188	.1547892		
10%	.7460142	.2336082	Obs	1,584
25%	.8969777	.2566367	Sum of Wgt.	1,584
50%	.9703367		Mean	.914435
		Largest	Std. Dev.	.1299331
75%	.9890217	.9991577		
90%	.9928671	.9991577	Variance	.0168826
95%	.9978025	.9991577	Skewness	-2.57666
99%	.9987371	.9994928	Kurtosis	10.36323

Fuente: elaboración propia.

Por lo tanto, el factor de expansión base de la encuesta de continuidad (para cada una de las observaciones), ajustado por cobertura telefónica, está dado por:

$$FE_{Continuidad}^{\widehat{Base,j}} = \frac{FE_{Continuidad}^{\widehat{Base,j}}}{Pr(Tener\ teléfono)} \quad \text{con } j = \{No\ Visitado, No\ Revisitado, Rechazo\}$$

Estos factores de expansión ajustados por cobertura telefónica, se utilizarán a continuación para realizar los ajustes por atrición y no respuesta.

4.4.4. Ajustes por atrición y no respuesta

Siguiendo el esquema de ajustes de la encuesta Presencial, se aplicó el ajuste del paso 3: se ajustó proporcional al tamaño para obtener la muestra efectiva, es decir, se excluyeron los fallecidos e impedidos.

A continuación, se realizaron ajustes por no respuesta que, se asume, tienen algún vínculo con las características informativas de la encuesta. Por lo tanto, esta posible autoselección en la no respuesta, se corrige mediante ajustes similares a los realizados en la encuesta de Re-Entrevista: se controla por tres posibles fuentes de sesgo.

En primer término, quienes tienen número de teléfono, no necesariamente contestan o son localizables o el número es correcto o asociado a la persona buscada, por lo que el contacto no está asegurado aun cuando esté disponible un número de teléfono. Esto se ajusta mediante la estimación de un modelo de propensión a tener contacto con el entrevistado, condicional en tener teléfono disponible, basado en regresión logística para obtener la probabilidad de contacto condicional en tener teléfono, $Pr(Contacto|Tener\ teléfono)$.

En segundo término, todos aquellos que pueden ser contactados no necesariamente responden la encuesta. Algunos de ellos, de hecho, rechazan la posibilidad de realizarla. Por lo tanto, se ajustan los factores de expansión por la probabilidad de responder la encuesta, $Pr(Responder|Contacto, Tener teléfono)$, que se obtiene de la estimación de un modelo de propensión a responder condicional en haber logrado el contacto y tener teléfono, basado en regresión logística.

Finalmente, el hecho de que una persona acepte y comience a responder una encuesta no asegura que la finalice, en especial, cuando se trata de encuestas telefónicas.⁴⁰ Por lo tanto, existe una probabilidad distinta de cero de que las encuestas sean abandonadas por el entrevistado antes de su finalización, lo que obliga a prescindir de las mismas dependiendo de la riqueza del contenido informativo logrado. Por lo tanto, se ajustan los factores de expansión para dar cuenta de la probabilidad de respuesta completa/incompleta de la encuesta:

$$Pr(Respuesta completa|Responder, Contacto, Tener teléfono)$$

Esta probabilidad se estima mediante un modelo de propensión a responder de forma completa la encuesta, condicional en aceptar responder la encuesta, haber logrado contacto y tener teléfono. La estimación se realiza con una regresión logística.

Por lo tanto, el factor de expansión ajustado por atrición y no respuesta en la encuesta de Continuidad se obtiene como:

$$FE_{Continuidad}^{AjNR,j} = FE_{Continuidad}^{Base,j} \cdot Pr(Tener teléfono)^{-1} \cdot Pr(Contacto|Tener teléfono)^{-1} \cdot Pr(Respuesta completa|Responder, Contacto, Tener teléfono)^{-1}$$

con $j = \{No Visitado, No Revisitado, Rechazo\}$

4.4.5. Post-estratificación

El último ajuste a los factores de expansión corresponde a la calibración por celdas o post-estratificación por sexo, grupo etario y regiones. Estas últimas fueron agrupadas, en algunos casos, como en la encuesta Presencial (véase la sección 4.2.3 para más detalles). Los grupos etarios son similares a los de la encuesta Presencial y de Re-Entrevista.

Los totales poblacionales se tomaron de las proyecciones censales del INE, disponibles públicamente. La fecha que se tomó como referencia fue el mes de agosto de 2020, periodo anterior al inicio del levantamiento, y los totales poblacionales para los que se post-estratificó se muestran en la **Tabla 39**. La interpolación se hizo con la misma metodología de la encuesta Presencial, siguiendo a Sadinle (2008).

⁴⁰ Véase Häder et al. (2012) y Peytchev et al. (2011). Respecto a los ajustes que se pueden aplicar a la información recolectada para evitar sesgos, véase Biemer (2001).

Tabla 39. Proyección de población de 23 años y más a Agosto de 2020, según macrozona, sexo y grupo etario

Región	Hombres				Mujeres				TOTAL
	23-40	41-55	56+	TOTAL	23-40	41-55	56+	TOTAL	Pobl.
I / XV	98.874	58.526	53.131	210.531	93.627	57.358	60.912	211.897	422.428
II	112.513	67.896	52.566	232.975	109.128	65.455	59.438	234.021	466.996
III	44.768	30.514	30.718	106.000	42.743	28.789	33.942	105.474	211.474
IV	110.889	79.533	87.048	277.470	112.220	83.875	102.269	298.364	575.834
V	265.522	180.058	221.098	666.678	258.158	189.045	274.412	721.615	1.388.293
VI	130.042	103.687	113.520	347.249	126.382	102.243	124.855	353.480	700.729
VII	142.069	112.838	129.104	384.011	146.154	116.854	146.492	409.500	793.511
VIII / XVI	274.110	215.082	238.531	727.723	285.413	228.364	284.224	798.001	1.525.724
IX	126.556	98.905	111.769	337.230	133.043	101.940	129.453	364.436	701.666
X / XIV	169.319	137.143	138.648	445.110	171.437	133.942	157.633	463.012	908.122
XI / XII	41.166	29.157	30.752	101.075	40.134	27.947	32.203	100.284	201.359
RM	1.306.328	776.892	729.454	2.812.674	1.227.221	789.463	935.340	2.952.024	5.764.698
Total	2.822.156	1.890.231	1.936.339	6.648.726	2.745.660	1.925.275	2.341.173	7.012.108	13.660.834

Nota: Interpolación según Sadinle (2008). Macrozonas: Norte: regiones 15, 1, 2, 3 y 4. Centro: 5, 6, 7, 16 y 8. Sur: 9, 14, 10, 11 y 12. R. Met.: 13.

Fuente: elaboración propia en base a proyecciones INE a Junio de 2020.

En la **Tabla 40**, se puede observar la cantidad de observaciones disponibles en cada celda de post-estratificación. Con la agrupación por macrozonas, se evitó tener celdas con menos de 10 observaciones, para evitar exacerbaciones de los factores de expansión.

Tabla 40. Muestra de entrevistados Vivos de la encuesta de Continuidad, según macrozona, sexo y grupo etario

Región donde vive/vivia el entrevistado	Hombres				Mujeres				TOTAL
	23-40	41-55	56+	TOTAL	23-40	41-55	56+	TOTAL	Pobl.
I / XV	30	15	22	67	38	14	18	70	137
II	17	15	10	42	26	17	13	56	98
III	22	13	12	47	20	19	10	49	96
IV	41	39	36	116	57	34	42	133	249
V	62	76	105	243	81	92	133	306	549
VI	102	32	41	175	124	36	74	234	409
VII	86	63	64	213	118	57	83	258	471
VIII / XVI	116	75	106	297	128	76	100	304	601
IX	82	62	75	219	106	49	100	255	474
X / XIV	47	38	46	131	59	40	35	134	265
XI / XII	19	10	16	45	16	12	11	39	84
RM	275	206	247	728	314	235	321	870	1598
Total	899	644	780	2323	1087	681	940	2708	5031

Fuente: elaboración propia.

Por su parte, la muestra, después de los diversos ajustes, estaba representando la siguiente distribución de población:

Tabla 41. Muestra expandida de entrevistados Vivos de la encuesta de Re-Entrevista, según macrozona, sexo y grupo etario

Región donde vive/vivía el entrevistado	Hombres				Mujeres				TOTAL Pobl.
	23-40	41-55	56+	TOTAL	23-40	41-55	56+	TOTAL	
I / XV	71438	41692	110303	223433	106804	32569	75226	214599	438032
II	83517	58028	52717	194262	62384	184516	82500	329400	523662
III	57020	32433	42951	132405	44081	35911	34203	114195	246600
IV	102126	104399	102585	309110	160254	144695	141775	446724	755834
V	145040	222843	260924	628807	118148	268038	466925	853110	1481917
VI	219929	84538	83023	387490	247159	74200	254554	575913	963403
VII	149671	202162	205155	556987	141827	146712	210475	499014	1056001
VIII / XVI	204430	225815	425436	856681	214298	253104	420858	888260	1743941
IX	78462	97799	134671	310931	117945	76050	349106	543102	854033
X / XIV	202817	252528	402988	858333	271888	217022	186062	674973	1533305
XI / XII	29898	9113	65083	104094	18167	24512	114356	157035	261129
RM	846451	831104	1001595	2679149	892159	850937	1461502	3204598	5883748
Total	2190799	2162453	2887431	7240683	2395115	2308265	3797542	8500922	15741605

Fuente: elaboración propia.

En la **Tabla 42**, se puede ver el factor de expansión de post-estratificación para cada celda. Este factor multiplica al factor de expansión de cada observación para conseguir los totales poblacionales post-estratificados.

Tabla 42. Coeficientes de ajuste de post-estratificación para entrevistados Vivos de la encuesta de Continuidad, según macrozona, sexo y edad

Región donde vive/vivía el entrevistado	Hombres			Mujeres		
	23-40	41-55	56+	23-40	41-55	56+
I / XV	1,384055	1,403770	0,481681	0,876623	1,761144	0,809720
II	1,347180	1,170059	0,997137	1,749291	0,354739	0,720463
III	0,785124	0,940834	0,715179	0,969651	0,801668	0,992379
IV	1,085806	0,761817	0,848548	0,700262	0,579668	0,721346
V	1,830685	0,808004	0,847364	2,185045	0,705293	0,587700
VI	0,591290	1,226508	1,367332	0,511339	1,377940	0,490486
VII	0,949212	0,558157	0,629301	1,030507	0,796487	0,696007
VIII / XVI	1,340847	0,952470	0,560674	1,331854	0,902253	0,675344
IX	1,612962	1,011313	0,829943	1,128007	1,340430	0,370813
X / XIV	0,834836	0,543081	0,344050	0,630542	0,617181	0,847208
XI / XII	1,376872	3,199628	0,472501	2,209146	1,140146	0,281602
RM	1,543301	0,934771	0,728293	1,375563	0,927757	0,639985

Fuente: elaboración propia.

4.4.6. Factores de expansión finales de corte transversal

Después de aplicar los factores de post-estratificación de la tabla anterior a cada entrevistado, se obtienen los factores finales corregidos de corte transversal para la encuesta de Continuidad. En la **Tabla 35**, se muestra la descripción estadística de estos factores para cada sub-muestra y la muestra total, junto a la población por cluster, como medidas informativas del comportamiento de los factores de expansión finales de la encuesta.

Tabla 43. Comportamiento de los factores de expansión de corte transversal de la encuesta de Re-Entrevista, y de las poblaciones representadas por cada uno

	Obs	Media	Desv. Std.	Min	Max
Afiliados 2002					
Población por cluster					
Base	2.746	99.416	57.890	6.824	283.247
Nominal inicial (en encuesta Presencial)	2.746	99.416	57.890	6.824	283.247
Ajustada por selección de muestra	2.746	99.416	57.890	6.824	283.247
Ajustada por Cobertura Telefónica	2.746	99.427	58.140	6.932	292.394
Efectiva	2.746	99.416	57.890	6.824	283.247
Ajustada por No Respuesta	2.746	104.688	63.415	3.160	323.988
Post-estratificada	2.746	178.250	101.997	12.117	533.624
Factores de Expansión					
Base	2.746	537,6	209,0	363,5	2.644,3
Nominal inicial (en encuesta Presencial)	2.746	552,5	213,7	369,5	2.644,3
Ajustados por selección de muestra	2.746	815,9	1.239,9	133,2	21.862,9
Ajustados por Cobertura Telefónica	2.746	837,8	1.293,1	134,5	22.065,8
Ajustados por Muestra Efectiva	2.746	883,4	1.342,6	131,9	23.031,7
Ajustados por No Respuesta	2.746	2.913,9	4.812,3	221,5	107.913,3
Post-estratificados	2.746	2.280,1	3.674,9	101,2	100.874,3
Nuevos Afiliados 2004					
Población por cluster					
Base	216	6.561	5.283	1.128	45.633
Nominal inicial (en encuesta Presencial)	216	6.561	5.283	1.128	45.633
Ajustada por selección de muestra	216	6.597	5.312	1.134	45.885
Ajustada por Cobertura Telefónica	216	6.703	5.424	1.134	45.917
Efectiva	216	6.597	5.312	1.134	45.885
Ajustada por No Respuesta	216	8.737	7.624	471	37.053
Post-estratificada	216	176.476	96.421	22.915	533.624
Factores de Expansión					
Base	216	567,9	630,4	200,9	6.519,0

Nominal inicial (en encuesta Presencial)	216	564,1	620,3	200,9	6.519,0
Ajustados por selección de muestra	216	821,9	1.037,7	5,9	10.740,0
Ajustados por Cobertura Telefónica	216	827,8	1.038,3	6,0	10.742,4
Ajustados por Muestra Efectiva	216	881,5	1.371,5	6,0	15.533,0
Ajustados por No Respuesta	216	2.462,4	3.546,2	21,7	24.723,6
Post-estratificados	216	2.744,1	4.281,4	33,6	38.156,0
No Afiliados 2004					
Población por cluster					
Base	436	49.980	22.454	2.640	99.044
Nominal inicial (en encuesta Presencial)	436	49.980	22.454	2.640	99.044
Ajustada por selección de muestra	436	49.980	22.454	2.640	99.044
Ajustada por Cobertura Telefónica	436	49.800	22.329	2.703	97.017
Efectiva	436	49.980	22.454	2.640	99.044
Ajustada por No Respuesta	436	56.955	31.938	2.495	176.211
Post-estratificada	436	183.119	109.295	22.915	533.624
Factores de Expansión					
Base	436	1.484,5	239,8	1.159,1	2.640,0
Nominal inicial (en encuesta Presencial)	436	1.561,7	224,0	1.118,9	2.711,4
Ajustados por selección de muestra	436	2.533,4	3.164,5	260,1	30.268,5
Ajustados por Cobertura Telefónica	436	2.630,6	3.281,7	260,9	32.117,8
Ajustados por Muestra Efectiva	436	2.764,0	3.310,9	267,0	30.400,2
Ajustados por No Respuesta	436	9.159,8	13.771,5	451,2	165.506,4
Post-estratificados	436	6.831,2	8.679,8	196,3	59.937,1
Nuevos 2015					
Población por cluster					
Base	1.633	38.203	22.394	6.461	117.213
Nominal inicial (en encuesta Presencial)	1.633	38.203	22.394	6.461	117.213
Ajustada por selección de muestra	1.633	38.203	22.394	6.461	117.213
Ajustada por Cobertura Telefónica	1.633	38.208	22.392	6.486	117.342
Efectiva	1.633	38.203	22.394	6.461	117.213
Ajustada por No Respuesta	1.633	44.770	25.572	1.195	132.094
Post-estratificada	1.633	173.182	95.530	12.117	533.624
Factores de Expansión					
Base	1.633	470,9	169,6	262,8	2.916,7
Nominal inicial (en encuesta Presencial)	1.633	470,9	169,6	262,8	2.916,7
Ajustados por selección de muestra	1.633	606,1	762,0	79,0	11.259,7
Ajustados por Cobertura Telefónica	1.633	607,8	763,8	79,5	11.271,2
Ajustados por Muestra Efectiva	1.633	611,3	768,7	78,9	11.232,6
Ajustados por No Respuesta	1.633	1.968,4	2.834,8	136,3	57.942,6
Post-estratificados	1.633	2.344,6	3.673,2	127,3	89.422,8

Muestra Completa					
Población por cluster					
Ajustada por No Respuesta	5.031	76.983	59.061	471	323.988
Post-Estratificado	5.031	176.951	100.388	12.117	533.624
Factores de Expansión					
Ajustados por No Respuesta	5.031	3.128,9	5.984,8	21,7	165.506,4
Post-Estratificados	5.031	2.715,3	4.545,7	33,6	100.874,3
Fuente: elaboración propia.					

En la parte final de la **Tabla 36**, se puede observar el comportamiento que tuvieron los factores de expansión, teniendo en cuenta las muestras originales de las que provenían.

5. Factores de expansión longitudinales

5.1. Factores de expansión longitudinales de la encuesta Presencial

Dado que la EPS ha acumulado 7 rondas, las combinaciones de muestras longitudinales que se pueden generar son abundantes. Por lo tanto, es necesario decidir qué muestras longitudinales pueden ser interesantes de estudiar. Para ellas, se deben crear los factores de expansión longitudinales correspondientes.

En el presente contexto, sabiendo que las incorporaciones de muestra se dieron en las olas de 2002, 2004 y 2015, se propuso generar los siguientes factores longitudinales:

1. Factor longitudinal para la muestra presente en 2002-2004-2006-2009-2015-2019
2. Factor longitudinal para la muestra presente en 2004-2006-2009-2015-2019
3. Factor longitudinal para la muestra presente en 2015-2019

Para generar los factores longitudinales, cada una de las sub-muestras originales es expandida para representar la población total que le corresponda (Participantes en 2002, Participantes en 2004, Participantes en 2015). Los factores, además, son corregidos por la propensión marginal a participar en cada ola, es decir, $Pr(participar\ en\ ola\ W)$, donde W es la ola particular o conjunto de olas, según corresponda.

5.1.1. Panel 2002-2004-2006-2009-2015-2019

Se estimó un modelo de propensión a participar en las siete olas de la EPS, alcanzando a un total 13.886 entrevistados, de los cuales 2.495 participaron del panel. La **Tabla 44** muestra las estimaciones de cinco especificaciones del modelo de probabilidad de participación.

Tabla 44. Estimación de la probabilidad de participar en el panel 2002-2004-2006-2009-2015-2019, encuesta Presencial

	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
Participó en el panel					
Edad 45-54	0.420*** (0.0935)	0.439*** (0.0940)	0.475*** (0.0912)	0.471*** (0.0906)	0.435*** (0.0936)
Edad 55-64	0.750*** (0.0964)	0.729*** (0.0967)	0.769*** (0.0906)	0.761*** (0.0901)	0.719*** (0.0963)
Edad 65-74	0.976*** (0.100)	0.943*** (0.102)	0.976*** (0.0967)	0.958*** (0.0959)	0.925*** (0.101)
Edad 75 o más	0.768*** (0.119)	0.685*** (0.122)	0.676*** (0.117)	0.659*** (0.116)	0.666*** (0.122)
I. Tarapacá	-0.199 (0.164)	-0.208 (0.166)	-0.221 (0.166)		
II. Antofagasta	0.103 (0.155)	0.0894 (0.155)	0.0826 (0.154)		
III. Atacama	0.0991 (0.168)	0.0222 (0.168)	0.0159 (0.168)		
IV. Coquimbo	-0.107 (0.116)	-0.159 (0.117)	-0.153 (0.117)		
V. Valparaíso	-0.291*** (0.0876)	-0.322*** (0.0881)	-0.329*** (0.0879)		
VI. O'Higgins	-0.0135 (0.104)	-0.117 (0.105)	-0.119 (0.105)		
VII. Maule	-0.306** (0.107)	-0.375*** (0.108)	-0.355** (0.108)		
VIII. Biobío	0.456*** (0.0711)	0.404*** (0.0714)	0.409*** (0.0712)		
IX. Araucanía	-0.533*** (0.139)	-0.578*** (0.139)	-0.574*** (0.139)		
X. Los Lagos	0.638*** (0.0830)	0.559*** (0.0837)	0.571*** (0.0833)		
XI. Aysén	-0.742* (0.330)	-0.814* (0.332)	-0.816* (0.334)		
XII. Magallanes	0.297 (0.206)	0.302 (0.213)	0.332 (0.212)		
Hombre	-0.420*** (0.0483)	-0.468*** (0.0490)	-0.420*** (0.0482)	-0.413*** (0.0479)	-0.463*** (0.0487)
Conviviente	-0.0515 (0.0759)	-0.123 (0.0763)			-0.146 (0.0755)
Anulado(a)	-0.467 (0.364)	-0.365 (0.367)			-0.413 (0.362)
Separado(a) de uni~l	-0.453*** (0.118)	-0.459*** (0.117)			-0.475*** (0.117)
Separado(a) de uni~o	-0.0194 (0.114)	-0.0580 (0.115)			-0.0974 (0.113)
Viudo(a)	-0.442** (0.144)	-0.492*** (0.143)			-0.493*** (0.142)
Soltero(a)	-0.157* (0.0664)	-0.126 (0.0665)			-0.131* (0.0662)
Sí	-0.0437 (0.103)	0.0271 (0.103)	0.0159 (0.103)	0.00475 (0.103)	0.0161 (0.103)
Sí	-0.166*** (0.0502)	-0.0381 (0.0517)	-0.0377 (0.0516)	-0.0332 (0.0511)	-0.0330 (0.0512)
Educación Básica		0.632** (0.212)	0.671** (0.213)	0.691** (0.215)	0.647** (0.214)
Educación Media		0.452* (0.213)	0.504* (0.215)	0.514* (0.217)	0.456* (0.215)
Educación Terciaria		0.00291 (0.227)	0.0513 (0.229)	0.0599 (0.230)	0.00509 (0.229)
Educación Universi~a		-0.349 (0.228)	-0.295 (0.229)	-0.284 (0.231)	-0.345 (0.230)
Norte				-0.0737 (0.0827)	-0.0691 (0.0829)
Centro				-0.0112	-0.0155

Sur				(0.0546)	(0.0548)
				0.122	0.111
Constant	-1.824***	-2.187***	-2.370***	(0.0709)	(0.0711)
	(0.0999)	(0.236)	(0.230)	(0.232)	(0.237)
Observations	13886	13886	13886	13886	13886
Pseudo R-squared	0.035	0.047	0.045	0.033	0.036
Standard errors in parentheses					
* p<0.05, ** p<0.01, *** p<0.001					
Fuente: elaboración propia.					

En la **Tabla 45**, se puede observar el comportamiento de la muestra para este panel, en que se rescatan 2.495 observaciones. También se puede ver el comportamiento promedio de los factores de expansión para la muestra seleccionada, después de ajuste por probabilidad de participación en el panel, y post-estratificados.

Tabla 45. Panel 2002-2004-2006-2009-2015-2019, encuesta Presencial: Comportamiento de la muestra y los factores de expansión

Tipo de Muestra	Participantes 2002
Original 2002	17.246
Nominal 2019	14.632
Efectiva 2019	13.886
Responden 2002 2004 2006 2009 2015 2019	2.495
Población representada originalmente	8.074.667
Factor de expansión base promedio	3236,339
Población representada con ajuste por NR	8.073.673
Factor de expansión ajustado por NR promedio	3235,941
Población representada post-estratificada	4.331.310
Factor de expansión post-estratificado promedio	1735,996

Fuente: elaboración propia.

La post-estratificación empleó los totales poblacionales de la **Tabla 46**. Nótese que los tramos de edad comienzan en 35 años, ya que este panel corresponde a los que tenían como mínimo 18 años en 2002. Por lo tanto, esto muestra la consistencia en el envejecimiento del panel. Geográficamente, se estratificó por macrozonas debido a la baja densidad de la muestra de panel en algunas regiones. De esta manera, se asegura que al menos haya 10 entrevistados en cada celda.

Tabla 46. Proyección de población de 35 años y más a Noviembre de 2019, según macrozona, sexo y grupo etario

Macrozona	Hombres				Mujeres				TOTAL
	35-50	51-65	66+	TOTAL	35-50	51-65	66+	TOTAL	Pobl.
Norte	117.128	86.803	46.793	250.724	115.408	89.659	59.379	264.446	515.170
Centro	286.573	253.143	157.651	697.367	295.787	270.072	203.440	769.298	1.466.665
Sur	124.163	104.945	62.142	291.250	124.751	105.627	78.618	308.996	600.246
R. Met.	390.292	283.373	153.828	827.494	382.373	315.724	223.636	921.734	1.749.227
Total	918.156	728.265	420.414	2.066.835	918.319	781.083	565.073	2.264.474	4.331.309

Nota: Interpolación según Sadinle (2008). Macrozonas: Norte: regiones 15, 1, 2, 3 y 4. Centro: 5, 6, 7, 16 y 8. Sur: 9, 14, 10, 11 y 12. R. Met.: 13.

Fuente: elaboración propia en base a proyecciones INE a Junio de 2019.

Los factores de ajuste por post-estratificación resultantes son los siguientes:

Tabla 47. Coeficientes de ajuste de post-estratificación: Muestra Panel 2002 | 2004 | 2006 | 2009 | 2015 | 2019, encuesta Presencial, según macrozona, sexo y grupo etario

Macrozona geográfica	Hombres			Mujeres		
	35-50	51-65	66+	35-50	51-65	66+
Norte	1,151584	0,371979	0,365287	1,142302	0,447843	0,475164
Centro	0,678604	0,40489	0,322346	0,750907	0,594522	0,747705
Sur	0,973466	0,502768	0,403123	0,756995	0,583786	0,536469
R. Met.	0,679213	0,474964	0,285359	0,638525	0,426682	0,451201

Fuente: elaboración propia.

5.1.2. Panel 2004-2006-2009-2015-2019

El panel 2004-2006-2009-2015-2019 recibió el mismo tratamiento. Las tres sub-muestras que conforman el panel (Afiliados 2002, Nuevos Afiliados 2004, No Afiliados 2004) fueron tratadas en conjunto debido al bajo volumen de casos por clusters. La presenta el comportamiento de la muestra y los factores de expansión longitudinales.

Tabla 48. Panel 2004-2006-2009-2015-2019, encuesta Presencial: Comportamiento de la muestra y los factores de expansión

Tipo de Muestra	Afiliados 2002	Nuevos Afiliados 2004	No Afiliados 2004	Total
Original 2004	13.190	944	2.860	16.994
Nominal 2019	11.743	918	2.554	15.215
Efectiva 2019	11.139	907	2.371	14.417
Responden 2004 2006 2009 2015 2019	2.495	148	563	3.206
Población representada originalmente	8.074.667	527.914	3.994.879	12.597.460
Factor de expansión base promedio	3236,339	3566,986	7095,700	3929,339
Población representada con ajuste por NR	8.178.253	562.602	3.870.322	12.611.177
Factor de expansión ajustado por NR promedio	3277,857	3801,366	6874,462	3933,617
Población representada post-estratificada	6.126.705	508.500	2.706.549	9.341.754
Factor de expansión post-estratificado promedio	2455,593	3435,813	4807,369	2913,835

Fuente: elaboración propia.

La post-estratificación empleó los totales poblacionales de la **Tabla 49**. Nótese que los tramos de edad comienzan en 34 años, ya que este panel corresponde a los que tenían como mínimo 18 años en 2004. Por lo tanto, esto muestra la consistencia en el envejecimiento del panel. Geográficamente, se estratificó por macrozonas debido a la baja densidad de la muestra de panel en algunas regiones. De esta manera, se asegura que al menos haya 10 entrevistados en cada celda.

Tabla 49. Proyección de población de 34 años y más a Noviembre de 2019, según macrozona, sexo y grupo etario

Macrozona	Hombres				Mujeres				TOTAL
	34-50	51-65	66+	TOTAL	34-50	51-65	66+	TOTAL	Pobl.
Norte	252.621	187.217	100.923	540.761	248.911	193.377	128.069	570.357	1.111.118
Centro	618.080	545.978	340.021	1.504.079	637.952	582.490	438.778	1.659.220	3.163.299
Sur	267.794	226.345	134.028	628.167	269.063	227.817	169.562	666.442	1.294.609
R. Met.	841.781	611.179	331.776	1.784.736	824.701	680.953	482.338	1.987.992	3.772.728
Total	1.980.276	1.570.719	906.748	4.457.743	1.980.627	1.684.637	1.218.747	4.884.011	9.341.754

Nota: Interpolación según Sadinle (2008). Macrozonas: Norte: regiones 15, 1, 2, 3 y 4. Centro: 5, 6, 7, 16 y 8. Sur: 9, 14, 10, 11 y 12. R. Met.: 13.

Fuente: elaboración propia en base a proyecciones INE a Junio de 2019.

Los factores de ajuste por post-estratificación resultantes se muestran en la **Tabla 50**.

Tabla 50. Coeficientes de ajuste de post-estratificación: Muestra Panel 2004|2006|2009|2015|2019, encuesta Presencial, según macrozona, sexo y grupo etario

<i>Macrozona geográfica</i>	<i>Hombres</i>			<i>Mujeres</i>		
	34-50	51-65	66+	34-50	51-65	66+
Norte	1,117763	0,682523	0,564897	1,326012	0,6893	0,523768
Centro	0,981619	0,745515	0,531715	0,839296	0,69882	0,585071
Sur	1,297535	0,868211	0,485317	0,941984	0,667104	0,484335
R. Met.	1,055341	0,94743	0,469069	0,890711	0,611354	0,500294

Fuente: elaboración propia.

5.1.3. Panel 2015-2019

Se aplicaron los mismos procedimientos que en los paneles anteriores. La muestra y los factores de expansión longitudinales presentaron el siguiente comportamiento:

Tabla 51. Panel 2015-2019, encuesta Presencial: Comportamiento de la muestra y los factores de expansión

Tipo de Muestra	Afiliados 2002	Nuevos Afiliados 2004	No Afiliados 2004	Nuevos 2015	Total
Original	8.634	475	1.711	6.805	17.625
Nominal 2019	8.063	464	1.587	6.778	16.892
Efectiva 2019	7.666	457	1.484	6.742	16.349
Responden 2015 2019	3.223	175	641	2.245	6.284
Población representada originalmente	8.074.667	527.914	3.994.879	3.250.561	15.848.021
Factor de expansión base promedio	2505,326	3016,651	6232,261	1447,911	2521,964
Población representada con ajuste por NR	8.072.751	505.797	3.986.971	3.289.106	15.854.625
Factor de expansión ajustado por NR promedio	2504,732	2890,266	6219,924	1465,081	2523,015
Población representada post-estratificada	6.221.192	513.338	2.739.482	3.943.835	13.417.847
Factor de expansión post-estratificado promedio	1930,249	2933,361	4273,763	1756,719	2135,240

Fuente: elaboración propia.

La post-estratificación empleó los totales poblacionales de la **Tabla 52**. Nótese que los tramos de edad comienzan en 23 años, ya que este panel corresponde a los que tenían como mínimo 18 años en 2015. Por lo tanto, esto muestra la consistencia en el envejecimiento del panel. Geográficamente, se estratificó por macrozonas para mantener la estrategia similar a los demás paneles que tienen menor cantidad de observaciones.

Tabla 52. Proyección de población de 23 años y más a Noviembre de 2019, según macrozona, sexo y grupo etario

Macrozona	Hombres					Mujeres					TOTAL
	23-35	36-50	51-65	66+	TOTAL	23-35	36-50	51-65	66+	TOTAL	Pobl.
Norte	268.484	252.621	187.217	100.923	809.245	260.570	248.911	193.377	128.069	830.927	1.640.172
Centro	595.617	618.080	545.978	340.021	2.099.696	594.956	637.952	582.490	438.778	2.254.176	4.353.872
Sur	245.822	267.794	226.345	134.028	873.989	250.990	269.063	227.817	169.562	917.432	1.791.421
R. Met.	959.756	841.781	611.179	331.776	2.744.492	899.899	824.701	680.953	482.338	2.887.891	5.632.383
Total	2.069.679	1.980.276	1.570.719	906.748	6.527.422	2.006.415	1.980.627	1.684.637	1.218.747	6.890.426	13.417.848

Nota: Interpolación según Sadinle (2008). Macrozonas: Norte: regiones 15, 1, 2, 3 y 4. Centro: 5, 6, 7, 16 y 8. Sur: 9, 14, 10, 11 y 12. R. Met.: 13.

Fuente: elaboración propia en base a proyecciones INE a Junio de 2019.

Los factores de ajuste por post-estratificación resultantes se muestran en la **Tabla 53**.

Tabla 53. Coeficientes de ajuste de post-estratificación: Muestra Panel 2015 | 2019, encuesta Presencial, según macrozona, sexo y grupo etario

Macrozona geográfica	Hombres				Mujeres			
	23-35	36-50	51-65	66+	23-35	36-50	51-65	66+
Norte	1,16955	1,63621	0,661758	0,537706	1,273234	1,391745	0,662309	0,446608
Centro	1,052354	1,291043	0,825116	0,534175	1,176775	0,946169	0,646195	0,512649
Sur	0,835222	1,632406	0,90707	0,477676	0,968248	1,029975	0,607477	0,454615
R. Met.	1,520097	1,39819	0,969094	0,466158	1,281783	0,944909	0,556065	0,428549

Fuente: elaboración propia.

5.2. Factores de expansión longitudinales para la encuesta Presencial y de Re-entrevista

Se calcularon factores de expansión longitudinales para estas dos encuestas, siendo la encuesta de Re-Entrevista la que generó la restricción en la muestra. Sin embargo, la muestra de la encuesta de Re-Entrevista (que forma panel con una parte de la muestra de la encuesta Presencial), representa a la población total mediante el ajuste de factores en base a la probabilidad de participar en ambas encuestas. Esto permite realizar análisis antes-después de la pandemia sobre ese subconjunto de observaciones. Por lo tanto, para estas dos encuestas se creó un factor longitudinal siguiendo los pasos ya descritos anteriormente.

Nótese aquí también que la combinación de estas dos encuestas implica combinar encuestas con distinta modalidad que requiere ser analizada en profundidad. Sin embargo, en este caso, hay evidencia más fuerte

que señala que se pueden lograr buenos resultados al alternar modalidades de entrevista, en especial, cuando las encuestas tienen distinto carácter, aunque vayan dirigidas a la misma población.⁴¹

Los factores de expansión ajustados presentaron el siguiente comportamiento:

Tabla 54. Panel encuesta Presencial y encuesta de Re-Entrevista: Comportamiento de la muestra y los factores de expansión

Tipo de Muestra	Total
Original	7.800
Nominal	5.569
Efectiva	5.537
Responden Encuestas Presencial y Re-Entrevista	2.082
Población representada originalmente	13.417.816
Factor de expansión base promedio	6444,676
Población representada con ajuste por NR	13.441.426
Factor de expansión ajustado por NR promedio	6456,016
Población representada post-estratificada	13.640.555
Factor de expansión post-estratificado promedio	6551,659

Fuente: elaboración propia.

La post-estratificación empleó los totales poblacionales de la **Tabla 55**.

Tabla 55. Proyección de población de 23 años y más a Julio de 2020, según macrozona, sexo y grupo etario

Macrozona	Hombres				Mujeres				TOTAL Pobl.
	23-40	41-55	56+	TOTAL	23-40	41-55	56+	TOTAL	
Norte	366.519	236.149	222.820	825.488	357.178	235.166	255.821	848.165	1.673.653
Centro	811.464	611.509	700.163	2.123.136	815.794	636.327	827.676	2.279.797	4.402.933
Sur	336.988	265.166	280.325	882.479	344.512	263.748	318.438	926.698	1.809.177
R. Met.	1.304.421	775.718	727.380	2.807.519	1.225.462	788.746	933.065	2.947.273	5.754.792
Total	2.819.392	1.888.542	1.930.688	6.638.622	2.742.946	1.923.987	2.335.000	7.001.933	13.640.555

Nota: Interpolación según Sadinle (2008). Macrozonas: Norte: regiones 15, 1, 2, 3 y 4. Centro: 5, 6, 7, 16 y 8. Sur: 9, 14, 10, 11 y 12. R. Met.: 13.

Fuente: elaboración propia en base a proyecciones INE a Junio de 2020.

Los factores de ajuste por post-estratificación resultantes se muestran en la **Tabla 56**.

⁴¹ Véase Dabalen et al. (2016).

Tabla 56. Coeficientes de ajuste de post-estratificación: Panel encuesta Presencial y encuesta de Re-Entrevista, según macrozona, sexo y grupo etario

<i>Macrozona geográfica</i>	<i>Hombres</i>			<i>Mujeres</i>		
	23-40	41-55	56+	23-40	41-55	56+
Norte	1,323050	1,313982	0,924768	0,994922	0,836955	0,888254
Centro	1,136188	1,087978	0,958367	0,928409	0,931318	1,029343
Sur	1,005448	1,085613	0,896630	1,070718	1,085370	0,906064
R. Met.	1,020595	0,934903	1,019729	1,018903	1,218711	0,944917

Fuente: elaboración propia.

5.3. Factores de expansión longitudinales de la encuesta de Continuidad

En el presente contexto, sabiendo que las incorporaciones de muestra se dieron en las olas de 2002, 2004 y 2015, y que la encuesta de Continuidad presenta las mismas características que la encuesta Presencial, se propuso generar los siguientes factores longitudinales:

1. Factor longitudinal para la muestra presente en 2002-2004-2006-2009-2015-2019
2. Factor longitudinal para la muestra presente en 2004-2006-2009-2015-2019
3. Factor longitudinal para la muestra presente en 2015-2019

Para generar los factores longitudinales, cada una de las sub-muestras originales es expandida para representar la población total que le corresponda (participante en 2002, participante en 2004, participante en 2015). Los factores, además, son corregidos por la propensión marginal a participar en cada ola, es decir, $\Pr(\text{participar en ola } W)$, donde W es la ola particular o conjunto de olas, según corresponda.

5.3.1. Panel 2002-2004-2006-2009-2015-2019

En la **Tabla 57**, se puede observar el comportamiento de la muestra para este panel, en que se rescatan 1.323 observaciones. También se puede ver el comportamiento promedio de los factores de expansión para la muestra seleccionada, después de ajuste por probabilidad de participación en el panel, y post-estratificados.

Tabla 57. Panel 2002-2004-2006-2009-2015-2019, encuesta de Continuidad: Comportamiento de la muestra y los factores de expansión

Tipo de Muestra	Afiliados 2002
Original 2002	17.246
Nominal 2019	14.632
Efectiva 2019	8.544
Responden 2002 2004 2006 2009 2015 2019	1.323
Población representada originalmente	8.074.667
Factor de expansión base promedio	6103,301
Población representada con ajuste por NR	8.104.822
Factor de expansión ajustado por NR promedio	6126,094
Población representada post-estratificada	4.566.095
Factor de expansión post-estratificado promedio	3451,319

Fuente: elaboración propia.

La post-estratificación empleó los totales poblacionales de la **Tabla 58**. Nótese que los tramos de edad comienzan en 35 años, ya que este panel corresponde a los que tenían como mínimo 18 años en 2002. Por lo tanto, esto muestra la consistencia en el envejecimiento del panel. Geográficamente, se estratificó por macrozonas debido a la baja densidad de la muestra de panel en algunas regiones. De esta manera, se asegura que al menos haya 10 entrevistados en cada celda.

Tabla 58. Proyección de población de 35 años y más a Agosto de 2020, según macrozona, sexo y grupo etario

Macrozona	Hombres				Mujeres				TOTAL
	35-50	51-65	66+	TOTAL	35-50	51-65	66+	TOTAL	Pobl.
Norte	123.472	91.262	50.297	265.032	121.793	94.283	63.609	279.685	544.717
Centro	297.853	265.658	169.102	732.613	307.560	283.144	217.408	808.112	1.540.724
Sur	128.784	110.330	66.498	305.612	129.664	110.910	83.687	324.261	629.873
R. Met.	414.448	298.220	164.835	877.502	403.913	331.037	238.329	973.279	1.850.781
Total	964.557	765.470	450.731	2.180.758	962.930	819.374	603.033	2.385.337	4.566.095

Nota: Interpolación según Sadinle (2008). Macrozonas: Norte: regiones 15, 1, 2, 3 y 4. Centro: 5, 6, 7, 16 y 8. Sur: 9, 14, 10, 11 y 12. R. Met.: 13.

Fuente: elaboración propia en base a proyecciones INE a Junio de 2020.

Los factores de ajuste por post-estratificación resultantes son los siguientes:

Tabla 59. Coeficientes de ajuste de post-estratificación: Muestra Panel**2002 | 2004 | 2006 | 2009 | 2015 | 2019, encuesta de Continuidad, según macrozona, sexo y grupo etario**

<i>Macrozona geográfica</i>	<i>Hombres</i>			<i>Mujeres</i>		
	35-50	51-65	66+	35-50	51-65	66+
Norte	0,789351	0,529423	0,358552	0,492149	0,472618	1,339045
Centro	0,780343	0,425497	0,391679	0,968775	0,619704	0,703584
Sur	1,031674	0,579865	0,303641	0,814414	0,49837	0,718347
R. Met.	0,876165	0,457595	0,286297	0,646455	0,396447	0,557545

Fuente: elaboración propia.

5.3.2. Panel 2004-2006-2009-2015-2019

El panel 2004-2006-2009-2015-2019 recibió el mismo tratamiento. Las tres sub-muestras que conforman el panel (Afiliados 2002, Nuevos Afiliados 2004, No Afiliados 2004) fueron tratadas en conjunto debido al bajo volumen de casos por clusters. La presenta el comportamiento de la muestra y los factores de expansión longitudinales.

Tabla 60. Panel 2004-2006-2009-2015-2019, encuesta de Continuidad: Comportamiento de la muestra y los factores de expansión

Tipo de Muestra	Afiliados 2002	Nuevos Afiliados 2004	No Afiliados 2004	Total
Original 2004	13.190	944	2.860	16.994
Nominal 2019	11.743	918	2.554	15.215
Efectiva 2019	6.663	631	1.366	8.660
Responden 2004 2006 2009 2015 2019	1.323	104	293	1.720
Población representada originalmente	8.074.667	527.914	3.994.879	12.597.460
Factor de expansión base promedio	6103,301	5076,096	13634,399	7324,105
Población representada con ajuste por NR	8.227.798	513.577	3.849.039	12.590.414
Factor de expansión ajustado por NR promedio	6219,046	4938,243	13136,652	7320,008
Población representada post-estratificada	6.258.864	492.437	2.762.449	9.513.750
Factor de expansión post-estratificado promed	4730,812	4734,967	9428,154	5531,250

Fuente: elaboración propia.

La post-estratificación empleó los totales poblacionales de la **Tabla 61**. Nótese que los tramos de edad comienzan en 34 años, ya que este panel corresponde a los que tenían como mínimo 18 años en 2004. Por lo tanto, esto muestra la consistencia en el envejecimiento del panel. Geográficamente, se estratificó por

macrozonas debido a la baja densidad de la muestra de panel en algunas regiones. De esta manera, se asegura que al menos haya 10 entrevistados en cada celda.

Tabla 61. Proyección de población de 34 años y más a Agosto de 2020, según macrozona, sexo y grupo etario

Macrozona	Hombres				Mujeres				TOTAL
	35-50	51-65	66+	TOTAL	35-50	51-65	66+	TOTAL	Pobl.
Norte	257.262	190.151	104.797	552.210	253.764	196.444	132.534	582.742	1.134.952
Centro	620.596	553.515	352.334	1.526.445	640.820	589.949	452.983	1.683.752	3.210.197
Sur	268.329	229.880	138.552	636.761	270.164	231.088	174.367	675.619	1.312.380
R. Met.	863.528	621.360	343.444	1.828.332	841.579	689.736	496.573	2.027.888	3.856.220
Total	2.009.715	1.594.906	939.127	4.543.748	2.006.327	1.707.217	1.256.457	4.970.001	9.513.749

Nota: Interpolación según Sadinle (2008). Macrozonas: Norte: regiones 15, 1, 2, 3 y 4. Centro: 5, 6, 7, 16 y 8. Sur: 9, 14, 10, 11 y 12. R. Met.: 13.

Fuente: elaboración propia en base a proyecciones INE a Junio de 2020.

Los factores de ajuste por post-estratificación resultantes se muestran en la **Tabla 62**.

Tabla 62. Coeficientes de ajuste de post-estratificación: Muestra Panel 2004|2006|2009|2015|2019, encuesta de Continuidad, según macrozona, sexo y grupo etario

Macrozona geográfica	Hombres			Mujeres		
	35-50	51-65	66+	35-50	51-65	66+
Norte	1,446430	0,892900	0,672729	0,798886	0,641441	0,607188
Centro	1,060109	0,694663	0,677911	0,982976	0,627667	0,672056
Sur	1,112789	0,910776	0,429856	0,949183	0,652548	0,547364
R. Met.	1,056812	0,987894	0,464064	0,906676	0,537191	0,575849

Fuente: elaboración propia.

5.3.3. Panel 2015-2019

Se aplicaron los mismos procedimientos que en los paneles anteriores. La muestra y los factores de expansión longitudinales presentaron el siguiente comportamiento:

Tabla 63. Panel 2015-2019, encuesta de Continuidad: Comportamiento de la muestra y los factores de expansión

Tipo de Muestra	Afiliados 2002	Nuevos Afiliados 2004	No Afiliados 2004	Nuevos 2015	Total
Original	8.634	475	1.711	6.805	17.625
Nominal 2019	8.063	464	1.587	6.778	16.892
Efectiva 2019	4.268	277	785	4.458	9.788
Responden 2015 2019	1.736	128	326	1.633	3.823
Población representada originalmente	8.074.667	527.914	3.994.879	3.250.561	15.848.021
Factor de expansión base promedio	4651,306	4124,328	12254,230	1990,546	4145,441
Población representada con ajuste por NR	8.180.624	424.521	3.796.340	3.228.302	15.629.787
Factor de expansión ajustado por NR promedio	4712,341	3316,570	11645,215	1976,915	4088,357
Población representada post-estratificada	6.556.471	468.739	2.693.399	3.942.216	13.660.825
Factor de expansión post-estratificado promedio	3776,769	3662,022	8261,960	2414,094	3573,326

Fuente: elaboración propia.

La post-estratificación empleó los totales poblacionales de la **Tabla 64**. Nótese que los tramos de edad comienzan en 23 años, ya que este panel corresponde a los que tenían como mínimo 18 años en 2015. Por lo tanto, esto muestra la consistencia en el envejecimiento del panel. Geográficamente, se estratificó por macrozonas para mantener la estrategia similar a los demás paneles que tienen menor cantidad de observaciones.

Tabla 64. Proyección de población de 23 años y más a Agosto de 2020, según macrozona, sexo y grupo etario

Macrozona	Hombres					Mujeres					TOTAL
	23-35	36-50	51-65	66+	TOTAL	23-35	36-50	51-65	66+	TOTAL	Pobl.
Norte	274.765	257.262	190.151	104.797	826.975	267.014	253.764	196.444	132.534	849.756	1.676.731
Centro	599.217	620.596	553.515	352.334	2.125.662	598.845	640.820	589.949	452.983	2.282.597	4.408.259
Sur	246.658	268.329	229.880	138.552	883.419	252.114	270.164	231.088	174.367	927.733	1.811.152
R. Met.	984.332	863.528	621.360	343.444	2.812.664	924.130	841.579	689.736	496.573	2.952.018	5.764.682
Total	2.104.972	2.009.715	1.594.906	939.127	6.648.720	2.042.103	2.006.327	1.707.217	1.256.457	7.012.104	13.660.824

Nota: Interpolación según Sadinle (2008). Macrozonas: Norte: regiones 15, 1, 2, 3 y 4. Centro: 5, 6, 7, 16 y 8. Sur: 9, 14, 10, 11 y 12. R. Met.: 13.

Fuente: elaboración propia en base a proyecciones INE a Junio de 2020.

Los factores de ajuste por post-estratificación resultantes se muestran en la **Tabla 65**.

Tabla 65. Coeficientes de ajuste de post-estratificación: Muestra Panel 2015 | 2019, encuesta de Continuidad, según macrozona, sexo y grupo etario

Macrozona geográfica	Hombres				Mujeres			
	23-35	36-50	51-65	66+	23-35	36-50	51-65	66+
Norte	1,300209	1,399927	0,966206	0,380419	1,27483	1,031909	0,666891	0,530842
Centro	1,176229	1,554303	0,793484	0,639509	1,084673	1,045006	0,595797	0,583447
Sur	1,084892	1,529668	0,90392	0,510623	0,945054	1,08213	0,636905	0,454288
R. Met.	1,346832	1,263397	0,972963	0,437745	1,364982	1,178288	0,576348	0,469963

Fuente: elaboración propia.

5.4. Ajustes adicionales a los factores de expansión

Los factores de expansión son un componente importante en el uso de una encuesta. La mayoría de los investigadores producen sus propios factores de expansión para las encuestas que analizan, dado que generalmente se trabaja con porciones de la muestra que tiene sus propios factores de expansión.

Uno de los problemas que tratan de solucionar los investigadores es la aparición de factores con valores extremos: es posible que los factores de expansión sean muy grandes o muy pequeños. Algunos investigadores prefieren "recortar" los factores de expansión cuando tienen tamaños extremos notorios. Este proceso se conoce en la literatura como "*trimming*". Esto se realiza por varias razones: estabilidad (menor varianza) de los estimadores, menor impacto de algunas observaciones (eliminación de *outliers* en la cola superior de la distribución, "los más grandes"), y reconsideración de la contribución informativa de otras observaciones (eliminación de *outliers* en la cola inferior de la distribución, "los más pequeños").

Uno de los problemas del *trimming* es que, al recortar los factores de expansión, se puede reducir la representatividad de la muestra, ya que la contribución informativa de algunas observaciones se ve alterada en favor de un comportamiento más homogéneo. Esto implica forzar un comportamiento más homogéneo también en la inferencia, alterando la representatividad. En todo caso, existe un *trade-off* entre mayores niveles de estabilidad en los factores de expansión y una mayor precisión en la representatividad de la muestra respecto a la población.

Muchas técnicas se han propuesto para reducir los tamaños extremos de los factores en la distribución y, aun así, mantener la representatividad de la muestra hasta cierto grado:

- colapsar categorías o celdas poblacionales: este tratamiento incluye la creación de pseudo-estratos y pseudo-clusters.
- fijar restricciones en los tamaños de los factores que se obtienen en procesos de ajustes poblacionales iterativos, como en el caso de la calibración iterativa en los márgenes ("*iterative proportional fitting*").

- aplicar métodos bayesianos y de simulaciones de cadenas de Markov (Montecarlo Markov Chain, o MCMC), aunque menos usados debido a su sofisticación en el cálculo en el uso posterior de los factores resultantes.

En todo caso, se sugiere brindar a los investigadores información suficiente, más allá de los datos propios de la encuesta, para que los investigadores puedan calcular y ajustar sus propios sets de factores de expansión ad-hoc al problema que están tratando.

6. Estimación de varianza y errores muestrales

En esta sección, se discute el impacto del uso de muestreo complejo para encuestas y cómo esto afecta a la EPS VII Ronda. Lo primero que se debe reconocer es que no existe encuesta (ni siquiera un censo) que no tenga problemas en la recolección de información. Por lo tanto, las encuestas requieren realizar ajustes a su representatividad y calidad de los datos y, además, deben ser capaces de representar los casos que no fueron efectivamente reportados en la encuesta.

El error de muestreo representa la imprecisión con que se mide un parámetro poblacional, θ , a partir de una muestra en vez de la población completa. Es decir, es la incertidumbre asociada al estimador $\hat{\theta}$ del parámetro poblacional. El parámetro poblacional puede ser el total de una característica en la población, medias, conteos, ratios y proporciones, cuantiles de la distribución de la característica, etc. Lo importante es que, cualquiera sea el parámetro que se esté estimando, el error de muestreo se puede definir como:

$$\text{Error de muestreo} = \sqrt{V(\hat{\theta})}$$

donde $V(\cdot)$ es la varianza del estimador.

El error de muestreo depende de muchos factores. Algunos están asociados a la manera en que diseña el muestreo, como, por ejemplo, el método de muestreo, el tamaño de la muestra, la naturaleza y distribución de la característica en la población, entre otros aspectos. Otros factores tienen que ver con la manera en que se recolectan los datos. Platek (1977) plantea que, al menos, hay tres grandes grupos de fuentes adicionales de incertidumbre:

- Los encuestadores: éstos se ven afectados por las competencias, motivación, entrenamiento, carga de trabajo y método de recolección de datos.
- Los entrevistados: sufren la múltiple influencia de factores que determinan su respuesta, como la disponibilidad, el esfuerzo en responder, la motivación, y la posibilidad de ser identificado dentro de la encuesta (secreto estadístico).
- El contenido de la encuesta: ésta puede implicar revelar información demográfica, económica o socio-económica de distinta sensibilidad para el entrevistado, lo que puede generar no respuesta a la encuesta o al ítem.

Todas estas situaciones (y otras más estudiadas en la literatura⁴²) introducen error en los resultados y la inferencia que se pueda realizar a partir de la información de encuestas.

La complejidad que presenta la EPS en su conjunto lleva estos inconvenientes al límite:

- La encuesta Presencial implica muestreo por *clusters* en dos etapas, sin reemplazos, con probabilidades desiguales de muestreo, ya que la probabilidad de ser muestreado depende no sólo del *cluster*, sino que también de la inversa de la proporción de población aportada por cada comuna a cada *cluster*, y cada *cluster* a la muestra total. La excepción está dada por las comunas auto-representadas que tienen probabilidad 1 de ser seleccionadas, pero igualmente la probabilidad de selección de cada observación depende del tamaño poblacional, que ha sido identificado como factor clave para determinar la contribución informativa a la muestra.
- Por su parte, la encuesta de Re-Entrevista procura información a partir de quienes respondieron la encuesta Presencial. Los tópicos allí relevados también difieren en algunos aspectos de la encuesta Presencial. Esto lleva a interpretar que se trata de una encuesta en dos fases, en la que la primera fase estuvo constituida por la encuesta Presencial, y la segunda fase es la encuesta de Re-Entrevista en sí misma. Nuevamente, se trata de un muestreo por *clusters* en dos etapas para cada una de las fases, sin reemplazo, y con desigual probabilidad de participación.
- Finalmente, la encuesta de Continuidad se realizó en base a la muestra residual de la encuesta Presencial, por lo que se debió ajustar la probabilidad de haber quedado en alguno de los tres grupos o sub-muestras de esa encuesta (no visitados, no re-visitados, y rechazo en encuesta Presencial). Luego de realizar estos ajustes, esta encuesta se puede interpretar como la encuesta Presencial, en que se realiza un muestreo por *clusters* en dos etapas, sin reemplazos, con probabilidades desiguales de selección.

Desde un punto de vista técnico, si bien la obtención de los estimadores como medias, proporciones y totales, puede ser algo más trivial, la situación es más dificultosa para determinar formas cerradas de la varianza (y, por ende, de los errores estándar) y la obtención de parámetros fiables para inferencia.

Sin embargo, a partir de los desarrollos de la literatura en la materia, en lo que sigue, se plantean maneras de obtener la varianza para cada una de las encuestas y estimadores habituales, con una forma cerrada basada en expansiones de Taylor (linealizaciones).^{43,44} También se discute la necesidad y metodología de los ajustes por población finita para la encuesta de Re-Entrevista.

⁴² Véase Lohr (2010), capítulo 1.

⁴³ Las derivaciones que siguen se basan, en gran parte, en Lohr (2010). También Horvitz y Thompson (1952), Sen (1953), Yates y Grundy (1953), Särndal et al. (1992).

⁴⁴ En la literatura se han explorado largamente metodologías heurísticas como *bootstrap*, *jackknife*, *balanced repeated replications*, etc., que también pueden ser usadas con la EPS VII Ronda.

6.1. Errores muestrales en la encuesta Presencial y de Continuidad

Como se mencionó anteriormente, para el caso de la encuesta Presencial y la encuesta de Continuidad (después de ajustar la muestra por la probabilidad de haber quedado en alguno de los grupos residuales de la encuesta Presencial), la selección de los entrevistados corresponde a un muestreo por *clusters* en dos etapas, sin reemplazos, con probabilidades desiguales de muestreo, ya que la probabilidad de ser muestreado depende no sólo del *cluster*, sino que también de la inversa de la proporción de población aportada por cada comuna a cada *cluster*, y cada *cluster* a la muestra total. La excepción está dada por las comunas auto-representadas que tienen probabilidad 1 de ser seleccionadas, pero igualmente la probabilidad de selección de cada observación dentro de la comuna depende del tamaño poblacional, que es la clave para entender la contribución informativa de cada observación a la muestra.⁴⁵

Para las comunas auto-representadas, la probabilidad de ser incluidas en la muestra es 1. Para las comunas no-auto-representadas, se forman estratos de aproximadamente 100.000 habitantes, según consta en las explicaciones del Diseño 2015. Por lo tanto, la probabilidad de que una comuna del estrato h sea incluida en la muestra es n_h/N_h , donde n_h es la población de tal comuna, y N_h es la población del estrato h . Así, cada comuna elegida en la muestra representa a N_h/n_h comunas de la población. Esto nos genera probabilidades distintas de selección para cada comuna, que se traducirán en probabilidades distintas de selección para cada unidad observacional.

Ahora, podemos definir la probabilidad de que un individuo forme parte de la muestra: dado que la muestra es sin reemplazos, la probabilidad de que un individuo i sea elegido en primera instancia es:

$$\psi_i = \text{Pr}(\text{seleccionar la unidad } i \text{ en primera instancia})$$

La probabilidad de que un individuo k sea seleccionado en segunda instancia, por lo tanto, dependerá de la probabilidad de haber sido seleccionado en primera instancia:

$$\begin{aligned} & \text{Pr}(\text{unidad } i \text{ elegida primera, unidad } k \text{ elegida segunda}) = \\ &= \text{Pr}(\text{unidad } i \text{ elegida primera}) \text{Pr}(\text{unidad } k \text{ elegida segunda} | \text{unidad } i \text{ elegida primera}) = \\ &= \psi_i \frac{\psi_k}{1 - \psi_i} \end{aligned}$$

Nótese que, dado que no hay reemplazo, el orden de elección importa, por lo que la probabilidad de que el individuo k sea elegido primero no es igual a la probabilidad de que el individuo i sea elegido primero. De hecho, para el individuo k , la probabilidad está dada por:

⁴⁵ Nótese que en la EPS se habla de estratos y de *clusters*, siendo los primeros las agrupaciones de comunas no-auto-representadas a partir de las cuales se elige un *cluster*. Sin embargo, dado que en cada estrato sólo se selecciona una comuna, hay una coincidencia entre lo que se denomina *cluster* y los estratos. Por lo tanto, la agrupación principal es el *cluster*, en que se seleccionan las observaciones finales.

$$\Pr(\text{unidad } k \text{ elegida primera, unidad } i \text{ elegida segunda}) = \psi_k \frac{\psi_i}{1 - \psi_k}$$

Si sólo se debieran elegir dos unidades de la muestra, entonces la suma de estas probabilidades darían como resultado la probabilidad de que un determinado individuo i fuese parte de la muestra:

$$\Pr(\text{unidades } i \text{ y } k \text{ estén en la muestra}) = \pi_{ik} = \psi_i \frac{\psi_k}{1 - \psi_i} + \psi_k \frac{\psi_i}{1 - \psi_k}$$

En forma más general, la probabilidad de que la unidad primaria de muestreo i sea parte de la muestra es:

$$\pi_i = \sum_{\mathcal{S}: i \in \mathcal{S}} \Pr(\mathcal{S})$$

donde \mathcal{S} es la muestra de n individuos tomada del universo finito $\mathcal{U} = \{1, 2, \dots, N\}$ de N individuos.

Sin embargo, dado que la muestra de las encuestas Presencial y de Continuidad son en dos etapas, se extiende este resultado de la siguiente manera: así, en la primera etapa se seleccionan *clusters*, por lo que el universo está constituido por N *clusters* tal que $i \in \mathcal{U} = \{1, 2, \dots, N\}$. A su vez, dentro de cada *cluster* se seleccionan observaciones j que corresponden a los entrevistados.

Asumiendo que y_{ij} es una característica asociada a la unidad j dentro del cluster i de la población, se desea estimar (inicialmente) el total de esta característica poblacional a través de:

$$t = \sum_{i=1}^N t_i = \sum_{i=1}^N \sum_{j=1}^{M_i} y_{ij} = \text{total poblacional}$$

donde N es el número de clusters en la población y M_i es el número de observaciones seleccionadas dentro del cluster i .

A partir de Horvitz y Thompson (1952), se puede obtener un estimador de la característica en la población total para una muestra como la de las encuestas Presencial y de Continuidad. Dado que la muestra es en dos etapas, el estimador de la población en la primera etapa está dado por:

$$\hat{t}_{HT} = \sum_{i \in \mathcal{S}} \frac{t_i}{\pi_i} = \sum_{i=1}^N Z_i \frac{t_i}{\pi_i}$$

donde $Z_i = 1$ indica que el individuo i pertenece a la muestra, y 0 en otro caso. Se puede mostrar que este estimador es insesgado siguiendo el Teorema 6.2 (Lohr, 2010, pp.255) basado en Horvitz y Thompson (1952), es decir:

$$E[\hat{t}_{HT}] = \sum_{i=1}^N \pi_i \frac{t_i}{\pi_i} = t$$

La varianza de este estimador está dada por:

$$V(\hat{t}_{HT}) = \sum_{i=1}^N \frac{1-\pi_i}{\pi_i} t_i^2 + \sum_{i=1}^N \sum_{\substack{k=1 \\ k \neq i}}^N \frac{\pi_{ik} - \pi_i \pi_k}{\pi_i \pi_k} t_i t_k = \frac{1}{2} \sum_{i=1}^N \sum_{\substack{k=1 \\ k \neq i}}^N (\pi_i \pi_k - \pi_{ik}) \left(\frac{t_i}{\pi_i} - \frac{t_k}{\pi_k} \right)^2$$

Donde π_{ik} es la probabilidad conjunta de que i y j sean parte de la muestra, y se puede estimar mediante:

$$\sum_{\substack{k=1 \\ k \neq i}}^N \pi_{ik} = (n-1)\pi_i$$

apoyado en el Teorema 6.1 en Lohr (2010, pp. 255).

Un estimador de esta varianza, sugerida por Horvitz y Thompson (1952), es:

$$\hat{V}_{HT}(\hat{t}_{HT}) = \sum_{i \in \mathcal{S}} (1 - \pi_i) \frac{t_i^2}{\pi_i^2} + \sum_{i \in \mathcal{S}} \sum_{\substack{k \in \mathcal{S} \\ k \neq i}} \frac{\pi_{ik} - \pi_i \pi_k}{\pi_i \pi_k} \frac{t_i}{\pi_i} \frac{t_k}{\pi_k}$$

Sustituyendo el estimador insesgado de la primera etapa para el *cluster* i , \hat{t}_i , iterativamente en el estimador de totales de Horvitz y Thompson (1952), se obtiene el estimador del total para la muestra en dos etapas:

$$\hat{t}_{HT}^{2S} = \sum_{i \in \mathcal{S}} \frac{\hat{t}_i}{\pi_i} = \sum_{i=1}^N Z_i \frac{\hat{t}_i}{\pi_i}$$

Horvitz y Thompson (1952) demostraron que este es un estimador insesgado de t así como también lo es $E[\hat{t}_i] = t_i$ para cada *cluster*.

La varianza de este estimador está dada por:

$$\begin{aligned} V(\hat{t}_{HT}^{2S}) &= \sum_{i=1}^N \frac{1-\pi_i}{\pi_i} t_i^2 + \sum_{i=1}^N \sum_{\substack{k=1 \\ k \neq i}}^N \frac{\pi_{ik} - \pi_i \pi_k}{\pi_i \pi_k} t_i t_k + \sum_{i=1}^N \frac{V(\hat{t}_{HT})}{\pi_i} \\ &= \frac{1}{2} \sum_{i=1}^N \sum_{\substack{k=1 \\ k \neq i}}^N (\pi_i \pi_k - \pi_{ik}) \left(\frac{t_i}{\pi_i} - \frac{t_k}{\pi_k} \right)^2 + \sum_{i=1}^N \frac{V(\hat{t}_{HT})}{\pi_i} \end{aligned}$$

que es similar a la de una etapa, pero incorpora incertidumbre por estimar los t_i en vez de medirlos directamente (variabilidad aportada por la *clusterización*).

El estimador de la varianza anterior es:

$$\hat{V}_{HT}(\hat{t}_{HT}^{2S}) = \sum_{i \in \mathcal{S}} (1 - \pi_i) \frac{\hat{t}_i^2}{\pi_i^2} + \sum_{i \in \mathcal{S}} \sum_{\substack{k \in \mathcal{S} \\ k \neq i}} \frac{\pi_{ik} - \pi_i \pi_k}{\pi_i \pi_k} \frac{\hat{t}_i}{\pi_i} \frac{\hat{t}_k}{\pi_k} + \sum_{i=1}^N \frac{\hat{V}_{HT}(\hat{t}_{HT})}{\pi_i}$$

Aun cuando la derivación teórica es directa, la aplicación de estos estimadores de la varianza requieren el conocimiento de la distribución conjunta de i y j para determinar

$$\pi_{ik} = \Pr(\text{clusters } i \text{ y } k \text{ están incluidos en la muestra})$$

y esto no suele ser trivial de computar. Por ello, generalmente se sugiere emplear el estimador de la varianza con reemplazo en su lugar:⁴⁶

$$\hat{V}_{HT,WR}(\hat{t}_{HT}^{2S}) = \frac{1}{n} \frac{1}{n-1} \sum_{i \in \mathcal{S}} \left(\frac{n\hat{t}_i}{\pi_i} - \hat{t}_{HT} \right)^2 = \frac{n}{n-1} \sum_{i \in \mathcal{S}} \left(\frac{\hat{t}_i}{\pi_i} - \frac{\hat{t}_{HT}}{n} \right)^2$$

Dado que, habitualmente, los analistas utilizan directamente los factores de expansión para realizar estos cálculos, podemos reescribir estos estimadores en términos esos factores. Así, el factor de expansión de la primera etapa para el cluster i está dado por:

$$w_i = \frac{1}{\pi_i}$$

Y el estimador del total poblacional se puede reescribir como:

$$\hat{t}_{HT} = \sum_{i \in \mathcal{S}} w_i \hat{t}_i$$

Para la segunda etapa, sin reemplazo, el factor de expansión está dado por:

$$w_{ij} = \frac{1}{\pi_{j|i}\pi_i}$$

donde

$$\pi_{j|i} = \Pr(\text{unidad } j \text{ del cluster } i \text{ está en la muestra} \mid \text{cluster } i \text{ está en la muestra}).$$

Por lo tanto, el estimador del total poblacional viene dado por:

$$\hat{t}_{HT}^{2S} = \sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{S}_i} w_{ij} y_{ij}$$

Donde $\mathcal{S}_i \in \mathcal{S}$ es la muestra correspondiente al *cluster* i . Reemplazando este estimador en el estimador de la varianza de dos etapas anterior, $\hat{V}_{HT,WR}(\hat{t}_{HT}^{2S})$, se tiene la correspondiente expresión para el mismo estimador, pero en términos de los factores de expansión.⁴⁷

⁴⁶ Esta varianza sólo depende de la Información de la primera etapa de selección, es decir, de la clusterización, por lo que el submuestreo al interior de cada *cluster* pasa a ser irrelevante.

⁴⁷ Los paquetes estadísticos estándar que trabajan con datos de encuestas, generalmente, tienen implementadas rutinas para resolver esta varianza a través del método de linealización con estimadores robustos (Sandwich).

Un ejemplo de comando en Stata 14 que instruye al sistema sobre las complejidades de la encuesta Presencial o la de Continuidad, es el siguiente:

```
svyset cluster_ID [pweight=factor_XS], strata(strata_ID), singleunit(certainty)
```

donde *cluster_ID* es el indicador de cluster, *strata_ID* es el indicador de cluster, y *factor_XS* es el factor de corte transversal post-estratificado.

El error estándar del estimador \hat{t}_{HT}^{2S} es la raíz cuadrada de la varianza estimada:

$$SE(\hat{t}_{HT}^{2S}) = \sqrt{\hat{V}_{HT,WR}(\hat{t}_{HT}^{2S})}$$

6.2. Errores muestrales en la encuesta de Re-Entrevista

Teniendo en cuenta que la encuesta de Re-Entrevista corresponde a la segunda fase de la encuesta Presencial (o, al menos, se puede interpretar estadísticamente así), se pueden obtener los errores muestrales a partir de las contribuciones de Särndal y Swensson (1987) y Legg y Fuller (2009). El análisis se realiza fase a fase.

En la primera fase, las unidades muestrales son seleccionadas siguiendo la variable aleatoria Z_{ij} , como antes:

$$Z_{ij} = \begin{cases} 1 & \text{si la unidad } i \text{ se incluye en la fase 1 del muestreo} \\ 0 & \text{en otro caso} \end{cases}$$

Los factores de esta primera fase, $w_{ij}^{(1)}$ para los entrevistados j del cluster i , vienen dados por:

$$w_{ij}^{(1)} = \frac{1}{P(Z_i = 1)} = w_{ij} = \frac{1}{\pi_{j|i}\pi_i}$$

Desde el punto de vista del observador, existe un conjunto de características auxiliares,

$$\mathbf{x}_{ij} = (x_{ij1}, x_{ij2}, \dots, x_{ijk})'$$

para cada observación de la muestra de fase 1. Usando el estimador Horvitz-Thompson, podemos estimar el total poblacional de una variable de interés k :

$$\hat{t}_{HT,x_k}^{2S(1)} = \sum_{i \in \mathcal{S}^{(1)}} \sum_{\substack{j \in \mathcal{S}_i^{(1)} \\ i \neq j}} w_{ij}^{(1)} x_{ijk} = \sum_{i=1}^N \sum_{j \neq i}^N Z_{ij} w_{ij}^{(1)} x_{ijk}$$

En la segunda fase, la variable D_{ij} indica si la observación es parte de la muestra, es decir,

$$D_{ij} = \begin{cases} 1 & \text{si la unidad } i \text{ se incluye en la fase 2 del muestreo} \\ 0 & \text{en otro caso} \end{cases}$$

La probabilidad de que una unidad sea parte de la fase 2, depende de que haya sido parte de la fase 1. Por lo tanto, los factores de la fase 2 vienen dados por:

$$w_{ij}^{(2)} = w_{ij}^{(2)}(\mathbf{Z}) = \begin{cases} \frac{1}{\Pr(D_{ij} = 1|\mathbf{Z})} & \text{si } Z_{ij} = 1 \\ 0 & \text{si } Z_{ij} = 0 \end{cases}$$

Un estimador similar al de una fase mostrado anteriormente para las otras dos encuestas, se puede aplicar a la encuesta de Re-Entrevista en la segunda fase:

$$\hat{t}_y^{(2)} = \sum_{i,j \in \mathcal{S}^{(1)}} Z_{ij} w_{ij}^{(1)} \sum_{i,j \in \mathcal{S}_i^{(2)}} D_{ij} w_{ij}^{(2)} y_{ij}$$

Asumiendo que la muestra de la primera fase es conocida, se puede hacer inferencia acerca del comportamiento de la segunda fase de manera más simple. Así, se define el total poblacional de primera fase:

$$\hat{t}_y^{(1)} = \sum_{i,j \in \mathcal{S}^{(1)}} w_{ij}^{(1)} y_{ij} = \sum_{i=1}^N \sum_{j=1}^{M_i} Z_{ij} w_{ij}^{(1)} y_{ij}$$

Aun cuando no se conozca $\hat{t}_y^{(1)}$ porque sólo se observan los valores de la fase 2 para y_{ij} , se lo puede asumir como un estimador de la población total en la fase 2, si es que se conocen los valores para y_{ij} de la muestra de la fase 1. Por lo tanto, tratando la muestra de la fase 1 como conocida, se puede mostrar que:

$$E[\hat{t}_y^{(1)} | \mathbf{Z}] = \sum_{i=1}^N \sum_{j=1}^{M_i} Z_{ij} w_{ij}^{(1)} y_{ij} E[D_{ij} | \mathbf{Z}] = \sum_{i,j \in \mathcal{S}^{(1)}} Z_{ij} w_{ij}^{(1)} y_{ij} = \hat{t}_y^{(1)}$$

Condicionando nuevamente en la fase 1, se obtiene el estimador insesgado de la población a partir del conocimiento de la fase 2:

$$E[\hat{t}_y^{(2)}] = E\{E[\hat{t}_y^{(1)} | \mathbf{Z}]\} = E\left[\sum_{i,j \in \mathcal{S}^{(1)}} Z_{ij} w_{ij}^{(1)} y_{ij}\right] = t_y$$

La varianza se obtiene de la misma manera:

$$V(\hat{t}_y^{(2)}) = V(E[\hat{t}_y^{(2)} | \mathbf{Z}]) + E(V[\hat{t}_y^{(2)} | \mathbf{Z}]) = V(\hat{t}_y^{(1)}) + E(V[\hat{t}_y^{(2)} | \mathbf{Z}])$$

Nótese que el sesgo (segunda expresión del lado derecho) depende del comportamiento de la primera fase, es decir, de la variable de selección Z .

Un problema adicional que se enfrenta es que la población muestreada en la segunda fase es relativamente grande respecto a la población disponible en la primera fase. En el caso de la encuesta de Re-Entrevista, se trata de 5.529 observaciones seleccionadas sobre un total de 7.763 disponibles con cobertura telefónica.⁴⁸

⁴⁸ Cabe recordar que la selección de la encuesta de Re-Entrevista se hizo condicional en tener teléfono de contacto disponible. Véase la sección 2.5.5.

Esto obliga a recurrir a una corrección por población finita que intuitivamente ajusta las ponderaciones teniendo en cuenta que la muestra incluye un porcentaje alto de la población muestreada. La corrección⁴⁹ viene dada, para el cluster i , por:

$$f_i = \frac{M_i}{N_i}$$

donde M_i es la cantidad de unidades seleccionadas dentro del cluster i , y N_i es la población total del cluster i .⁵⁰

Rao (1973) propone un estimador general de la varianza para muestreo en dos fases, que incluye estratificación y corrección por población finita. Teniendo en cuenta las características de la encuesta de Re-Entrevista, se obtuvo una forma cerrada para el estimador de la varianza $V(\hat{t}_y^{(2)})$ a partir de la contribución de Rao (1973):

$$\hat{V}(\hat{t}_y^{(2)}) = \sum_{i=1}^N (\hat{t}_i - \bar{\hat{t}}_y)^2 + \sum_{i=1}^N (1 - f_i) \left(\frac{M_i}{M_i - 1} \right) \sum_{j=1}^{M_i} (\hat{t}_{ij} - \bar{\hat{t}}_i)^2$$

donde

$$f_i = \frac{M_i}{N_i}$$

$$\hat{t}_i = \sum_{j=1}^{M_i} Z_{ij} w_{ij}^{(2)} y_{ij}$$

$$\bar{\hat{t}}_y = \frac{1}{N} \sum_{i=1}^N \hat{t}_i$$

$$\hat{t}_{ij} = \sum_{k=1}^{M_{ij}} Z_{ijk} w_{ijk}^{(2)} y_{ijk}$$

⁴⁹ Cochran, W.G. (1977), Sampling Techniques, third edition, Wiley, New York.

⁵⁰ Si bien en la encuesta Presencial no fue necesario hacer este ajuste porque el marco muestral era la población total, en la encuesta de Re-Entrevista, el marco muestral de Vivos es el conjunto de personas que respondió el cuestionario de Vivos en la encuesta Presencial. Es fácil ver que, en el caso de la encuesta de Re-Entrevista, el factor de corrección medio ronda el 0.4 (o 40%) para la mayoría de los clusters, mientras que en la encuesta Presencial es cercano a 0 (para la sub-muestra 2002, es 0,003802012; para la sub-muestra nuevos afiliados 2004, es 0,003746822; para la sub-muestra no afiliados 2004, es 0,00085109; para la sub-muestra 2015, es 0,004614588). Esto quiere decir que, en promedio, se seleccionaron muestras intra-cluster de aproximadamente el 60% de la población disponible para la encuesta de Re-Entrevista. El factor de corrección de la encuesta de Re-Entrevista presenta la siguiente distribución: $\bar{x} = 0,3936799$, $\sigma = 0,0827845$, $\min(x) = 0,1666667$, $\max(x) = 0,7857143$.

$$\bar{\hat{t}}_i = \frac{1}{M_i} \sum_{j=1}^{M_i} \hat{t}_{ij}$$

En la expresión de la varianza, puede verse que la primera parte del lado derecho corresponde a la incertidumbre proveniente de la primera fase, mientras que la segunda parte corresponde a la incertidumbre de la segunda fase.⁵¹

El error estándar del estimador $\hat{t}_y^{(2)}$ es la raíz cuadrada de la varianza estimada:

$$s(\hat{t}_y^{(2)}) = \sqrt{\hat{v}(\hat{t}_y^{(2)})}$$

6.3. Coeficiente de variación

El coeficiente de variación de un estadístico $\hat{\theta}$ está dado por la relación entre el error estándar y la media esperada del estimador:

$$CV(\hat{\theta}) = \frac{\sqrt{V(\hat{\theta})}}{E(\hat{\theta})}$$

Para los estimadores de totales poblacionales mostrados anteriormente, los coeficientes de variación están dados por:

- Coeficiente de variación del estimador de totales en la encuesta Presencial y de Continuidad:

$$CV(\hat{\theta}) = \frac{\sqrt{\hat{V}_{HT,WR}(\hat{t}_{HT}^{2S})}}{E[\hat{t}_{HT}^{2S}]}$$

- Coeficiente de variación del estimador de totales en la encuesta de Re-Entrevista:

$$CV(\hat{\theta}) = \frac{\sqrt{\hat{V}(\hat{t}_y^{(2)})}}{E[\hat{t}_y^{(1)}|\mathbf{Z}]}$$

⁵¹ Un ejemplo de comando en Stata 14 que instruye al sistema sobre las complejidades de la encuesta de Re-Entrevista, es el siguiente:

```
svyset cluster_ID [pweight=factor_XS], strata(strata_ID) fpc(fpc1) ||
cluster_ID, fpc(fpc2) , singleunit(certainty)
```

donde `cluster_ID` es el indicador de cluster, `strata_ID` es el indicador de cluster, `factor_XS` es el factor de corte transversal post-estratificado, `fpc1` y `fpc2` son los tamaños relativos de la cantidad de unidades muestreadas respecto a la cantidad de unidades representadas, es decir, los factores de corrección por muestra finita.

6.4. Intervalos de confianza

Para los resultados obtenidos en esta sección para los tres componentes de la EPS VII Ronda, bajo condiciones bastante generales, se espera que los estimadores sigan asintóticamente una distribución normal estándar⁵², es decir,

$$\frac{(\hat{\theta} - \theta)}{\sqrt{\hat{V}(\hat{\theta})}} \underset{\sim}{\text{asym.}} \mathcal{N}(0,1)$$

Por lo tanto, se puede construir un intervalo de confianza de aproximadamente 95% para el parámetro θ empleando:

$$\hat{\theta} \pm 1,96\sqrt{\hat{V}(\hat{\theta})}$$

Donde la varianza $\hat{V}(\hat{\theta})$ es alguna de las encontradas anteriormente.

6.5. Efecto diseño

El efecto diseño es una medida que muestra el efecto del diseño de una encuesta sobre la varianza de un estimador (Gabler et al., 2006; Skinner, 1986). Esta es una medida importante en el diseño de las encuestas, en especial, cuando se plantea el uso de agrupaciones de muestra (estratos y clusters). Kish (1965) lo definió como:

$$deff(plan, \hat{\theta}) = \frac{\hat{V}_{plan}(\hat{\theta})}{\hat{V}_{SRS}(\hat{\theta})}$$

donde *plan* indica el plan de muestreo particular de la encuesta analizada, y *SRS* indica un plan alternativo hipotético basado en muestreo aleatorio simple. El estadístico $\hat{\theta}$ puede ser cualquiera de las medidas habituales, como media, total, proporción, ratio, cuantil, etc. Por lo tanto, esta es una medida que se debe estimar para cada variable e indicador (y subpoblación, si corresponde).

La interpretación es que si $deff > 1$, entonces el plan de muestreo genera un incremento en la incertidumbre versus una situación en que se habría aplicado un muestreo aleatorio simple. En el caso de los componentes de la EPS VII Ronda, dado que incluyen clusterización y probabilidades desiguales de selección, esto impacta incrementando la varianza por encima del muestreo aleatorio simple.⁵³

⁵² Véase Binder (1983).

⁵³ Nótese que este resultado no es general. En ocasiones, estrategias de clusterización y utilización de probabilidades desiguales pueden llevar a mejorar la varianza (Lohr, 2010, capítulo 6). Sin embargo, no es un resultado habitual.

El efecto diseño se puede relacionar con la inflación que sufre la varianza y los intervalos de confianza derivados de ellos cuando se aplican planes de muestreo complejo. Así, el intervalo de confianza ajustado por efecto diseño puede estimarse como:

$$\hat{\theta} \pm 1,96\sqrt{deff}\sqrt{\hat{V}(\hat{\theta})}$$

Por lo tanto, si $deff > 1$, el intervalo de confianza con el plan de muestreo complejo será más amplio (mayor imprecisión en la inferencia) que en un muestreo aleatorio simple.

Kish (1995) aproxima esta definición de \sqrt{deff} con

$$deft(\hat{\theta}) = \frac{SE_{plan}(\hat{\theta})}{\frac{s}{\sqrt{n}}}$$

Donde s es el error estándar del estimador proveniente de un muestreo aleatorio simple. Por lo tanto, $deft$ puede ser usado en lugar de \sqrt{deff} con similares resultados.

Algunos analistas suelen reportar el efecto de error de especificación del muestreo, ya que es un cálculo que suele estar disponible en los paquetes de software. Esta medida y su raíz cuadrada están dadas por:

$$meff(plan, \hat{\theta}) = \frac{\hat{V}_{plan}(\hat{\theta})}{\hat{V}(\hat{\theta})}$$

$$mefl(plan, \hat{\theta}) = \sqrt{meff}$$

Donde $\hat{V}_{plan}(\hat{\theta})$ corresponde a la varianza estimada según las particularidades del plan de muestreo, y $\hat{V}(\hat{\theta})$ corresponde a una estimación *naive* de la varianza, es decir, una estimación que no tiene en cuenta factores de expansión, estratificación y/o clusterización.

Al igual que $deff$ y $deft$, estas medidas son estimadas para un estadístico particular, por lo que cambian según el análisis que se realice. En la siguiente sección, se muestran algunos ejemplos para diferentes indicadores.

Referencias

- AAPOR (2010), "New Considerations for Survey Researchers When Planning and Conducting RDD Telephone Surveys in the U.S. With Respondents Reached via Cell Phone Numbers," AAPOR Cell Phone Task Force Report.
- AAPOR (2014), "Mobile Technologies for Conducting, Augmenting and Potentially Replacing Surveys: Report of the AAPOR Task Force on Emerging Technologies in Public Opinion Research," AAPOR Task Force on Emerging Technologies Report.
- Albert, P. S. (2000). "A transitional model for longitudinal binary data subject to non-ignorable missing data." *Biometrics*, Vol. 56:602–608.
- Albert, P. S., Follmann, D. A., Wang, S. A. y Suh, E. B. (2002). "A latent autoregressive model for longitudinal binary data subject to informative missingness." *Biometrics*, Vol. 58:631–642.
- Biemer, P. P. (2001), "Nonresponse bias and measurement bias in a comparison of face to face and telephone interviewing," *Journal of Official Statistics*, vol. 17(2):295-320.
- Binder, D. A. (1983), "On the variances of asymptotically normal estimators from complex surveys," *International Statistical Review*, Vol. 51:279–292.
- Buuren, S.V. (2018). *Flexible Imputation of Missing Data*, segunda edición, CRC Press.
- Cameron, A. C., y P. K. Trivedi (2005), *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.
- Cameron, A. C., y P. K. Trivedi (2010), *Microeconometrics Using Stata*. Revised edition, College Station, TX: Stata Press.
- CEPAL (2020a), "Recomendaciones para eliminar el sesgo de selección en las encuestas de hogares en la coyuntura de la enfermedad por coronavirus (COVID-19)," *Informes COVID-19*, Mayo, División de Estadísticas.
- CEPAL (2020b), "Recomendaciones para la publicación de estadísticas oficiales a partir de encuestas de hogares frente a la coyuntura de la enfermedad por coronavirus (COVID-19)," *Informes COVID-19*, Abril, División de Estadísticas.
- Chen, C. (2004), "An Introduction to Quantile Regression and the QUANTREG Procedure," Paper 213-30, SUGI 30, SAS Institute.
- Cobben, F. (2009). *Nonresponse in sample surveys : methods for analysis and adjustment*. Den Haag: Statistics Netherlands.

- Cobben, F. y J. Bethlehem (2005), "Adjusting undercoverage and nonresponse bias in telephone surveys," Statistics Netherlands, Discussion paper 05006.
- Courbois, J.-Y. P., y N. S. Urquhart (2004), "Comparison of survey estimates of the finite population variance," *Journal of Agricultural, Biological, and Environmental Statistics*, vol. 9(2):236–251.
- Dabalen, A., A. Etang, J. Hoozevee, E. Mushi, Y. Schipper, y J. von Engelhardt (2016), "Mobile Phone Panel Surveys in Developing Countries: A Practical Guide for Microdata Collection." International Bank for Reconstruction and Development / World Bank Group, Washington, DC.
- Daniels, M. J. y J. W. Hogan (2008). *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. Monographs on Statistics and Applied Probability 109, Chapman & Hall/CRC: Boca Raton, FL.
- Demnati, A. y J.N.K. Rao (2004), "Linearization Variance Estimators for Survey Data," *Survey Methodology*, vol. 30(1):17-26.
- Díaz de Rada, V. (2019), "Calidad de respuesta y modo de administración del cuestionario," *Revista Internacional de Sociología*, Vol. 77, No. 1.
- Eltinge, J. L. y W. M. Sribney (1996), "Estimation of means, totals, ratios, and proportions for survey data," *Stata Technical Bulletin*, May, STB-31.
- Gabler, S., S. Häder y P. Lynn (2006), "Design effects for multiple design samples," *Survey Methodology*, Vol. 32(1):115-120.
- Gosset, J. M., P. Simpson, J.G. Parker y W.L. Simon (2002), "How Complex Can Complex Survey Analysis Be with SAS?" Paper 266-27, SUGI 27, SAS Institute.
- Häder, S., M. Häder y M. Kühne (Eds., 2012), *Telephone Surveys in Europe: Research and Practice*. Springer:Berlin.
- Horvitz, D. G., y D. J. Thompson (1952), "A generalization of sampling without replacement from a finite universe," *Journal of the American Statistical Association*, vol. 47:663-685.
- Huang, Q., H. Zhang, J. Chen y M. He (2017), "Quantile Regression Models and Their Applications: A Review," *Journal of Biomtrics & Biostatitics*, vol 8(3).
- INE Chile (2020), "Nota Técnica: Encuesta Nacional de Empleo (ENE) - Contingencia COVID-19.", Encuesta Nacional de Empleo, INE Chile.

- Kennedy, C. y H. Hartig (2019), "Response rates in telephone surveys have resumed their decline," Pew Research Center, in <https://www.pewresearch.org/fact-tank/2019/02/27/response-rates-in-telephone-surveys-have-resumed-their-decline/>
- Kish, L. (1965). Survey sampling. NewYork:Wiley.
- Kish, L. (1995). Methods for design effects. Journal of Official Statistics, 11, 55–77.
- Lavrakas, P. J. (2008), "Methods for sampling and interviewing in telephone surveys," capítulo 16 en L. Bickman y D. J. Rog, editores, "The SAGE Handbook of Applied Social Research Methods," 2da. edición, SAGE Publications:US.
- Legg, J. C., y W. A. Fuller (2009), "Two-phase sampling," en D. Pfeffermann y C. R. Rao (editores), Handbook of Statistics: Vol. 29A. Sample surveys: Design, methods and applications, pp. 55–70, Amsterdam: North Holland.
- Lewis, T. H. (2017), Complex Survey Data Analysis with SAS, Chapman and Hall/CRC, .
- Lin, H., C. E. McCulloch y R. A. Rosenheck (2004). "Latent pattern mixture models for informative intermittent missing data in longitudinal studies." Biometrics, Vol.60:295–305.
- Little, R. y D. Rubin (2014). Statistical analysis with missing data. John Wiley & Sons:Boston, MA.
- Lohr, S. L. (2008), "Coverage and sampling," capítulo 6 en De Leeuw, E. D., J. J. Hox y D. A. Dillman & European Association of Methodology (editores). International handbook of survey methodology. New York: Lawrence Erlbaum Associates.
- Lohr, S. L. (2010), "Sampling: design and analysis", 2da. edición, Brooks/Cole Cengage Learning.
- Peytchev, A., L. R. Carley-Baxter y M. C. Black (2011), "Multiple Sources of Nonobservation Error in Telephone Surveys: Coverage and Nonresponse," Sociological Methods & Research, vol. 40(1):138–168.
- Platek, R. (1977), "Some factors affecting non-response," Survey Methodology, vol. 3:191-214.
- Rodriguez, R. N. y Y. Yao (2017), "Five Things You Should Know about Quantile Regression," Paper SAS525-2017, SAS Institute.
- Rotnitzky, A., D. O. Scharfstein, T.-L. Su y J. M. Robins (2001). "Methods for conducting sensitivity analysis of trials with potentially nonignorable competing causes of censoring." Biometrics, Vol. 57:103–113.
- Rubin, D. (1976). Inference and missing data. Biometrika, Vol. 581–592.
- Sadinle, M. (2008), "Metodología para interpolar tamaños poblacionales," Documentos de CERAC, No. 7.

- Särndal, C.-E., B. Swensson y J. Wretman (1992), *Model assisted survey sampling*. New York: Springer-Verlag.
- Särndal, C.-E., y B. Swensson (1987), "A general view of estimation for two phases of selection with applications to two-phase sampling and nonresponse," *International Statistical Review*, vol. 55:279–294.
- Sen, A. R. (1953), "On the estimate of the variance in sampling with varying probabilities," *Journal of the Indian Society of Agricultural Statistics*, vol. 5:119–127.
- Shah, B. V. (2004), "Comment to 'Demnati and Rao: Linearization Variance Estimators for Survey Data'," *Survey Methodology*, vol. 30(1):29.
- Skinner, C. J. (1986), "Design Effects of Two-Stage Sampling," *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 48(1):89-99.
- Skinner, C. y J. Wakefield (2017), "Introduction to the Design and Analysis of Complex Survey Data," *Statistical Science*, vol. 32(2):165-175.
- Steeh, C. (2008), "Telephone Surveys," capítulo 12 en De Leeuw, E. D., J. J. Hox y D. A. Dillman & European Association of Methodology (editores). *International handbook of survey methodology*. New York: Lawrence Erlbaum Associates.
- Troxel, A. B., S. R. Lipsitz y D. P. Harrington (1998). "Marginal models for the analysis of longitudinal measurements with nonignorable non-monotone missing data." *Biometrika*, Vol. 85:661–672.
- United Nations (2005), "Household Sample Surveys in Developing and Transition Countries," Department of Economic and Social Affairs Statistics Division, *Studies in Methods*, Series F No. 96, United Nations Publication, New York:NY.
- Valliant, R. y J. A. Dever (2018), *Survey Weights: A Step-by-Step Guide to Calculation*. Stata Press Publication, College Station:TX.
- Valliant, R., Dever, J. A., & Kreuter, F. (2018). *Practical tools for designing and weighting survey samples*. Cham, Switzerland: Springer.
- Wood, J., G. Matthews, J. Pellowski y O. Harel (2019). "Comparing Different Planned Missingness Designs in Longitudinal Studies." *Sankhya B*, Vol. 81:226–250.
- Yates, F., y P. M. Grundy (1953), "Selection without replacement from within strata with probability proportional to size," *Journal of the Royal Statistical Society, Series B*, vol. 109:12-30.

