



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

FACULTAD DE CIENCIAS SOCIALES

INSTITUTO DE SOCIOLOGÍA

MAGÍSTER EN DISEÑO Y ANÁLISIS DE ENCUESTAS SOCIALES

PENSIONES DE VEJEZ:

CUANTIFICACIÓN DE LA DISCREPANCIA ENTRE
LA ENCUESTA DE PROTECCIÓN SOCIAL E
HISTORIAL PREVISIONAL DE AFILIADOS,
PENSIONADOS Y FALLECIDOS DEL AÑO 2015

POR

VALERIA ANDREA PEÑAILILLO SEGUEL

Tesis presentada al Instituto de Sociología de la Pontificia Universidad Católica de Chile para optar al grado académico de magíster.

Comisión Carolina Casas Cordero – Matías Bargsted –

Luis Maldonado

Diciembre 2020 Santiago, Chile

©2020 VALERIA ANDREA PEÑAILILLO SEGUEL

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento incluyendo la cita bibliográfica del documento.

Índice

Resumen.....	6
1. Introducción	7
2. Marco Teórico.....	10
2.1 Errores de medición desde la metodología de encuestas.....	10
2.2 Estudios sobre cuantificación de discrepancia entre encuestas y registros administrativos	13
3. Objetivos de investigación	17
4. Datos y Métodos.....	18
4.1 Fuentes de datos	18
4.1.1 Encuesta de Protección Social 2015.....	18
4.1.2 Historial Previsional de pensionados y fallecidos 2015	21
5. Variables	22
5.1 Auto reporte de Ingresos por pensiones de vejez -EPS.....	22
5.2 Registro objetivo de pensiones de vejez – HPA.....	23
5.3 Sub/sobre reporte de ingresos de pensiones de vejez.....	24
5.4 Límites de tolerancia para discrepancia	25
5.5 Otras Variables.....	26
6. Base de datos Analítica	29
6.1 Identificación de pensionados por vejez -edad y anticipada- en modalidad de retiro vitalicia en EPS 2015	30
6.2 Identificación de pensionados por vejez -edad y anticipada- en modalidad de retiro vitalicia en Historial Previsional Administrativo de Afiliados Activos, Pensionados y Fallecidos del Sistema de AFP – 2015	31
6.3 Match de datos ESP – HPA 2015.....	33
6.4 Depuración base de datos.....	34
6.5 Estrategia analítica	38
7. Análisis y resultados.....	39
7.1 Cuantificación discrepancia EPS – HPA 2015 respecto a ingresos por pensiones de vejez	39
7.2 Factores que influyen en el sub y sobre reporte de pensiones de vejez.....	41
8. Conclusiones	43
9.1 Discusión.....	45
Bibliografía	48
Anexos.....	53

Porcentaje de casos perdidos (missing), según conocimiento del sistema de pensiones y características demográficas de los entrevistados (n=499)	53
Figura N°1 Encuesta de Protección Social 2015.....	54
Figura N°2 Historial Previsional Administrativo (HPA) de Afiliados Activos, Pensionados y Fallecidos del Sistema de AFP – 2015	55

Tabla 1: Módulos de la Encuesta Protección Social 2015 – Subsecretaria de Previsión Social	20
Tabla 2: Historial Previsional de Afiliados, Pensionados y Fallecidos– Superintendencia de Pensiones	21
Tabla 3: Descripción variables modelos regresión logística	26
Tabla 4: Variables dicotómicas utilizadas en estudio	27
Tabla 5: Estadísticas univariadas de variables utilizadas (no ponderadas)	35
Tabla 6: Modelos de regresión Logística de los indicadores de sub/sub reporte de pensiones. Análisis sin ponderar	42
Figura 1: Proceso construcción base de datos	37
Figura 2: Scatter plot entre EPS y HPA 2015	39
Figura 3: : Scatter plot Discrepancia - HPA 2015 (valor objetivo)	40
Figura 4: Box Plot EPS - HPA 2015	41

Resumen

Ante las críticas respecto a la presencia de errores en las encuestas sociales y el debate respecto a su idoneidad para la representación de la población, la presente investigación realizó un análisis entre las diferencias existentes entre encuestas y registros administrativos. En concreto, este estudio cuantificó la discrepancia existente entre la Encuesta de Protección Social (EPS) y el Historial Previsional de Afiliados, Pensionados y Fallecidos (HPA) del año 2015 en lo que respecta a los ingresos por pensiones de vejez por edad y anticipada en modalidad de retiro vitalicia. Para cada persona de la muestra se realizó un matching entre la EPS y el HPA (Deterministic Record Linkage). Los principales resultados evidencian una brecha absoluta de \$43.124 CLP (media) entre lo declarado en la EPS y lo registrado en HPA respecto a pensiones de vejez por edad y anticipada en modalidad de retiro vitalicia. Los modelos de regresión logística aplicados permiten evidenciar que existen dos factores asociados al sub-reporte de pensiones de vejez en EPS 2015, el primero de ellos es el conocimiento de la Pensión Básica Solidaria (PBS) y el segundo es encontrarse emparejado. En otras palabras, las personas que no conocen la PBS tienden a sub reportar y aquellos que están casados también. Un importante hallazgo fue que ambas fuentes de información no están exentas de errores y omisiones, frente a este escenario se sugiere medir y mejorar continuamente los sistemas de recolección de información ya sean estos encuestas o registros administrativos para de esta manera contar con fuentes de información confiables y completas.

Palabras claves: Encuestas, Registros Administrativos, Discrepancia, Pensiones de Vejez, Modalidad Vitalicia.

1. Introducción

Las encuestas son un método de recolección de sistemática de datos de una muestra de una población (Groves y cols., 2004), que permiten orientar la toma de decisiones y el diseño de políticas públicas en diversas materias, dado que se fundamentan en métodos científicos de muestreo, construcción del instrumento y análisis de la información recopilada (Traugot, 2012). Sin embargo, con el transcurso de los años las encuestas han sido criticadas por las inminentes fuentes de error que presentan. Este escenario ha incrementado la desconfianza e incertidumbre respecto a la veracidad, utilidad y capacidad de representatividad de los datos que se recopilan a través de este tipo de instrumentos (Smith, 2011). Ya desde el año 1994 se impulsaba el concepto Error Total de Encuestas (TSE¹) (Smith, 2011), que permite interpretar las múltiples formas a través de las cuales los valores observados de una encuesta resultan erróneos al desviarse de los valores verdaderos; hay algunos errores que se atribuyen al proceso de muestreo y otros al proceso de aplicación del instrumento. Por lo que hoy por hoy, el TSE es constantemente utilizado como estrategia metodológica para reducir los errores que emergen en los procesos de diseño, aplicación y análisis de las encuestas.

Los registros administrativos se definen como *“colecciones de datos del gobierno, utilizadas con el propósito de administrar impuestos, beneficios o servicios”* (United Nations Economic Commission for Europe, p.1, 2010). El principal incentivo asociado al uso de registros administrativos es que son considerados como Gold Standard (Villatoro et al. 2018), lo que permite decir que la información reportada por la mencionada fuente es representativa, consistente y confiable. Si bien, los datos contenidos en registros administrativos cuentan con información respaldada metodológicamente, presentan una gran desventaja en comparación a las encuestas: la mayoría no permite probar hipótesis ni tampoco responder preguntas de investigación porque no cuentan con la riqueza y profundidad de información necesarias para las mencionadas labores (Ibid. 2018). Dado este escenario, es que la estrategia de complementar y sustituir la información proveniente de las encuestas con la de los registros administrativos, se ha tornado una práctica recurrente entre investigadores de diferentes áreas para obtener resultados fundados en la unificación de diferentes fuentes de

¹ Total Survey Error

información. Se considera que los registros administrativos asumen un rol significativo cuando complementan la información proveniente de las encuestas (Groen, 2011). Es por esto, que la complementariedad de ambas fuentes de información ofrece gran utilidad para la investigación aplicada vinculada a las políticas públicas; dado que los registros administrativos y las encuestas proporcionan un conjunto integrado de datos confiables para desarrollar diferentes tipos de análisis, desde la evaluación de políticas públicas y programas sociales, como también el funcionamiento de determinados servicios ofrecidos por determinadas instituciones (Künn,2015).

Es importante señalar que existen registros administrativos a nivel individual, como el caso del Historial Previsional de Afiliados, Pensionados y Fallecidos del año 2015 (HPA), y a nivel agregado como las cuentas nacionales de los países en las que no es posible identificar casos a nivel individual o desagregadamente. La complementariedad entre los registros administrativos y las encuestas está sujeta a dificultades que se vinculan directamente con las discrepancias de información que se obtiene de una y la otra fuente. Es decir, a partir del proceso de enlace entre las fuentes de datos mencionadas emergen diferencias que deben ser analizadas desde una perspectiva metodológica. La literatura considera diversas hipótesis para estudiar las causas de las discrepancias entre las encuestas y los registros administrativos. Las principales, guardan relación con los errores de cobertura identificados en cuentas nacionales, registros individuales ofrecidos por instituciones públicas y privadas (Groen, 2012; Kapteyn & Ypma, 2006). Por otro lado, están aquellos supuestos que encausan la discrepancia por fracasos en la fase de muestreo de las encuestas, una inadecuada aplicación de los instrumentos por parte de los encuestadores como también por falseamiento u olvido de datos declarados por parte de los encuestados (CEPAL, 2003).

En la literatura especializada, los esfuerzos se concentran en determinar en términos porcentuales los niveles de discrepancia que emergen de la unificación de los registros administrativos con las encuestas en lo que respecta a determinadas variables de interés. La presente investigación afrontó el desafío de cuantificar la discrepancia individual entre lo declarado en la Encuesta de Protección Social y lo registrado en el Historial Previsional de Afiliados, Pensionados y Fallecidos del año 2015 en relación a los ingresos por pensiones de vejez en modalidad vitalicia de las unidades encuestadas. Se decidió utilizar la Encuesta de

Protección Social 2015 porque está vinculada a los registros administrativos elaborados por la Superintendencia de Pensiones. En otras palabras, la encuesta permite acceder a los datos declarados individualmente y al mismo tiempo a la información de datos administrativos retrospectiva e innominada sobre los individuos encuestados.

El presente análisis se orientó en cuantificar la discrepancia individual entre las fuentes de información referidas en relación con los ingresos de vejez en modalidad de retiro vitalicia. Posteriormente, se identificaron las unidades encuestadas que sub y sobre reportan pensiones de vejez en modalidad de retiro vitalicia en la Encuesta de Protección Social al ser comparadas con el Historial Previsional de Afiliados, Pensionados y Fallecidos del año 2015, para luego caracterizarlas en función de variables sociodemográficas y conocimiento de determinados elementos del Sistema de Previsión Social de Chile. El principal resultado analítico, es que la Encuesta de Protección Social (EPS) del año 2015 respecto a los ingresos por pensiones de vejez en modalidad vitalicia está capturando apropiadamente el ingreso declarado por las unidades encuestadas, al ser comparada con el Historial Previsional de Afiliados, Pensionados y Fallecidos (HPA) que contiene información sobre los montos realmente recibidos por las unidades encuestadas.

En Chile no se han desarrollado investigaciones que cuantifiquen la discrepancia de información entre registros administrativos y encuestas respecto a ingresos por pensiones de vejez. Por tanto, la presente tesis busca contribuir en revelar las precauciones metodológicas a considerar al momento de realizar la complementariedad entre ambos tipos de fuentes de datos. Como también, a incentivar el desarrollo de investigaciones del mismo tipo y sobre todo aquellas que utilicen como fuentes de información la EPS y HPA.

El documento se estructura así: primero se expone una revisión de literatura sobre investigaciones que han desarrollado la comparación de registros administrativos con encuestas para identificar la discrepancia resultante. En segundo lugar, se presenta una revisión sobre los errores asociados a la discrepancia que emerge del matching entre las encuestas y los registros administrativos. En tercer lugar se expone el diseño metodológico del análisis, descripción de las fuentes de datos, plan de análisis; y finalmente los resultados y conclusiones.

2. Marco Teórico

En primera instancia, se presentan una serie errores de medición desde la metodología de encuestas (sección 3.1). Luego, se exponen distintas investigaciones que cuantifican la discrepancia entre encuestas y registros administrativos a nivel individual, y los posibles factores asociados a la discrepancia resultante (sección 3.2).

2.1 Errores de medición desde la metodología de encuestas

El Error Total de Encuestas identifica la acumulación de errores que emergen a partir de la implementación y desarrollo de las diferentes etapas de la encuesta: el diseño, recopilación, procesamiento y análisis de los datos obtenidos (*Public Opinion Quarterly*, 2010). El error total de encuestas, de ahora en adelante TSE² por sus siglas en inglés, es un concepto introducido por primera vez por Deming en 1994; que permite interpretar las múltiples formas a través de las cuales los valores observados de una encuesta resultan erróneos al desviarse de los valores verdaderos. Si bien, la principal función del TSE es identificar cuán alejados están los valores observados del valor verdadero, es frecuente que también se utilice para comparar valores verdaderos de diferentes conjuntos de datos. Una metodología para comparar datos observados con los llamados verdaderos es la utilización de los Gold Standard; que son un marco de comparación que tiene como utilidad ser referentes para evaluar el movimiento de los valores observados de las encuestas. En otras palabras, cuantificar cuán alejados se encuentran los valores observados del Gold Standard. (Smith, 2011).

A partir de lo expuesto por Biemer (2010), el TSE es un marco referencial que permite optimizar las encuestas en función de dos caminos. El primero guarda relación con el incremento de la calidad de los datos a través de la reducción de los errores, y el segundo busca vincular la calidad de los datos ajustándolos a un presupuesto determinado. En otras palabras, el TSE es una estrategia de diseño que perfecciona el proceso de construcción, aplicación y análisis de los instrumentos. El TSE forma parte de los componentes principales de la calidad total de la encuesta, de ahora en adelante TSQ³ por sus siglas en inglés. La TSQ

² TSE: Total Survey Error (Smith, 2011)

³ TSQ: Total Survey Quality (Whiteley, 2014)

permite evaluar y determinar que una encuesta es apta para medir los fenómenos para la cual fue construida, enfocándose en la relevancia, credibilidad y accesibilidad de los datos levantados. En este escenario la TSE tiene como objetivo analizar la precisión de los datos y reducir los errores (Whiteley, 2014). El TSE identifica dos fuentes de variación entre los valores observados respecto a los verdaderos, la primera es la variación que guarda relación con los errores aleatorios, y la otra es el sesgo o error sistemático. La primera no provoca un impacto en los valores medios dada su relación con la varianza; en cambio la segunda altera las estimaciones medias por su carácter direccional. Cabe señalar que, estas dos formas de interpretar el TSE presentan diferencias estadísticas, por tanto, deben ser interpretadas de manera independiente (Biemer, 2010).

Los errores aleatorios o de muestreo, son fuentes de error atribuibles al esquema de muestreo o proceso de construcción de la muestra; se clasifican en tres tipos: construcción de la muestra, tamaño de la muestra y elección del estimador. Esta clasificación de errores genera una variación en las observaciones de la muestra, y no introducen sesgos (Smith, 2010). La *construcción de la muestra* se asocia a la omisión, duplicación e incorporación de observaciones erróneas; *el tamaño de la muestra* está vinculada directamente con las observaciones seleccionadas para aplicar el instrumento, por ejemplo, puede surgir un error por una mala implementación del método de selección al azar de quien responde en el hogar; *la elección del estimador* es la ponderación inadecuada de la muestra final, o también el uso inadecuado de una ponderación correcta.

A partir de lo expuesto por Biemer (2010), los errores sistemáticos o sesgos se clasifican en: error de especificación, de medición, de no respuesta, de procesamiento de datos y de marco de muestreo. La *especificación* surge cuando en la encuesta se emplea un concepto que no está vinculado con los objetivos de recolección de datos, esto quiere decir que, se recolectan datos erróneos porque se utiliza una concepción distinta a lo que se planificó; un ejemplo de la especificación es cuando se pregunta sobre conceptos de ingresos laborales de la principal ocupación del entrevistado, y al momento de detallar el ingreso el encuestado considera los aportados por actividades u ocupaciones secundarias.

Los errores de medición de las encuestas se refieren a la discrepancia que emerge entre el valor verdadero y el valor observado de una variable determinada (Biemer,2010). El *error de medición* ha sido estudiado profundamente por investigadores y se ha concluido que es el principal factor que influye negativamente en los análisis de datos basados en encuestas (Groves et al. 2009). Este tipo de error da cuenta de la diferencia entre el valor observado y el valor verdadero, y puede surgir por: aplicar inadecuadamente el instrumento, porque los entrevistados proporcionan información incorrecta, falseamiento de datos por contextos o preguntas delicadas, por efectos del encuestador, y también por el incumplimiento de los protocolos a seguir para la aplicación de las encuestas.

El *error de no respuesta* se puede originar cuando una unidad no responde la totalidad del cuestionario, como también, cuando una unidad no responde una pregunta del instrumento. Ambas situaciones se traducen en la reducción de la muestra que trae como consecuencia una reducción en la representatividad de los datos.

Por otro lado, el *error de procesamiento de datos* guarda relación con el ingreso, codificación de los datos, y la asignación de ponderadores al instrumento (Biemer, 2010); en reiteradas ocasiones este tipo de errores genera efectos negativos sobre las estimaciones de las encuestas.

Para evidenciar si una encuesta está capturando adecuadamente determinado tipo de información, es indispensable utilizar una metodología que permita comparar datos observados *-encuesta-* con los *registros administrativos*, que son un caso particular donde se asume que son fuentes de información libres de error o, que tendrían menos errores que las encuestas. En otras palabras, aquellos datos considerados libres de error o con menores niveles de error que los datos de las encuestas, se les denomina Gold Standard. En este sentido, el uso de los registros administrativos como Gold Standard permite obtener un marco de comparación que tiene como utilidad ser un referente para evaluar la validez de los valores observados de las encuestas. En otras palabras, cuantificar cuan alejados se encuentran los *valores observados* de los *valores verdaderos*. (Smith, 2011).

2.2 Estudios sobre cuantificación de discrepancia entre encuestas y registros administrativos

Para realizar una cuantificación de la discrepancia entre encuestas y registros administrativos existen dos estrategias. La primera de ellas es utilizar registros administrativos y encuestas que brinden información a nivel agregado, en aquellos casos en que no se puede obtener información individual. La segunda, es utilizar fuentes de información que cuenten con alguna variable que identifique a nivel individual la información que suministran.

Kapteyn e Ypma (2006), desarrollan un estudio empírico a nivel individual utilizando los datos de LINDA⁴ del año 2002 que contiene datos administrativos de los ingresos de las personas y los datos de una encuesta anual⁵ aplicada en Suecia sobre ingresos, gastos de hogar, educación, salud, sistema de pensiones. Para realizar el matching entre las fuentes de información mencionadas, los investigadores utilizaron el número de seguro social disponible en la encuesta y en los datos administrativos. Para cuantificar la discrepancia utilizaron la pregunta “¿Tuvo algún ingreso por concepto de trabajo en el año 2002?”, que fue respondida afirmativamente por 435 encuestados. Al comparar esta respuesta con los registros de LINDA, se identificó que 17 (4,6%) encuestados no cuentan con ingresos a pesar de responder en la encuesta que sí los obtuvieron, 93 (18,2%) encuestados declararon no tener ingresos y en LINDA se observó que sí recibieron ingresos. Los autores atribuyen que las diferencias entre ambas fuentes de datos pueden emerger por los errores de aplicación de la encuesta, también porque el encuestado aproxima de manera errónea o por la facilidad de olvidar el monto exacto. Para los datos sobre ingresos por conceptos de pensiones se detectó un patrón similar de información con discrepancias. Un 12,6% de los encuestados declara que no reciben pensiones, sin embargo, en los registros administrativos se observa que sí reciben ingresos por conceptos de pensión. Los autores concluyen que los desajustes se deben a que ambas fuentes de información miden conceptos diferentes sobre ingresos, por lo que afirman que los datos administrativos no deben ser considerados una verdad absoluta, sino que deben ser evaluados antes de ser utilizados como fuentes de datos estadísticos.

⁴ LINDA: Longitudinal Individual data Base 1968 /2018 – Sweden

⁵ Encuesta similar a CASEN en Chile

Asimismo, Groen (2012) para realizar una comparación entre registros administrativos y encuestas a nivel individual, utiliza por un lado los microdatos del censo trimestral QCEW⁶ en donde las empresas presentan sus contribuciones y salarios otorgados a trabajadores trimestralmente. Por otro lado, consideró los datos del CES⁷ sobre empleo y salarios de la industria no agrícola, recopilados a través de encuestas telefónicas (CATI). Los datos que comparó fluctúan entre enero de 2006 a marzo de 2007, y otorgan información sobre los empleos que ofrece la industria privada no agrícola. A partir de ambas fuentes de datos, se logró una coincidencia de 155 registros entre ambas fuentes de información. Las principales causas que el autor atribuyó a las discrepancias identificadas en el crecimiento del empleo se deben al error de cobertura (41%) y la no respuesta (32%). A pesar de que ambas fuentes de datos cuentan con la misma conceptualización de empleo, los errores y discrepancias emergen cuando los empleadores declaran los datos de salarios de sus trabajadores, ya que muchas veces no diferencian a los trabajadores con licencia médica o despedidos. La alta tasa de no respuesta en CES genera un incremento de la varianza y sesgos, al igual que la imputación de datos en QCEW por el hecho de que los empleadores no declaran información en función del trimestre correspondiente. Particularmente en el estudio realizado por Groen (2012) se observa que la reducción de casos al comparar las fuentes de información se origina por la no respuesta en las mismas fuentes.

La investigación desarrollada por Guimarães (2013) analiza las discrepancias entre los registros administrativos de los beneficiarios del programa PBF⁸ y la encuesta PNAD⁹ a nivel individual entre los años 2006 y 2009. Primero realiza un cálculo de representatividad en base a PBF, y de captación o beneficiarios reales en base a PNAD, lo que permitió evidenciar que el número de familias que declaran ser beneficiarias del PBF en la PNAD es bastante inferior a lo que se observa en los registros administrativos de beneficiarios del programa PBF. El autor explica que este error en las declaraciones se debe en primera instancia por la

⁶ QCEW: Quarterly Census of Employment and Wages – E.E.U.U

⁷ CES: Current Employment Statics – E.E.U.U

⁸ PBF: Programa de transferencias monetarias para población que vive en situación de pobreza o extrema pobreza en Brasil.

⁹ PNAD: Encuesta nacional por Muestra de Domicilios. Similar a CASEN en Chile.

confusión de los encuestados ante la unificación de distintos programas sociales en un gran programa y/o beneficios.

Dinamarca cuenta con uno de los registros administrativos más confiables y representativos a nivel internacional, situación que favorece los procesos de comparación con los datos obtenidos a partir de encuestas. Conjuntamente, el gobierno del mencionado país otorga un registro único de identificación¹⁰ a cada persona, escenario que facilita la combinación o pareo de datos a nivel individual con encuestas de diversa índole. Es por las mencionadas razones que los investigadores Kreiner, Dreyey y Petersen (2012), decidieron utilizar la información obtenida a partir de registros administrativos sobre impuestos a la renta para validar los datos suministrados por la encuesta DES¹¹ entre 1994 y 1996. Los resultados arrojaron la existencia de diferencias entre las respuestas de la encuesta sobre gastos realizados y los registros administrativos, situación que los autores atribuyen a que las personas no recuerdan detalladamente los gastos que realizan diariamente y lo que declaran es un promedio. En otras palabras, las compras que declaran no necesariamente se realizaron durante el periodo de tiempo consultado en la encuesta; por otro lado, las diferencias resultantes las asocian a las imputaciones de datos que se realizan a los gastos de las familias.

Domingo Sanz (2011) desarrolló una investigación en la cual compara los datos de la EPA con el fichero de Afiliación de Seguridad Social de España, el registro administrativo permite obtener información respecto a la cifra mensual de afiliación y cotizaciones individuales. La investigadora identificó elevadas diferencias en aquellas personas que declaran ser activas en la EPA¹² versus la información del fichero de afiliación. Esta diferencia la atribuye a las diferentes conceptualizaciones sobre lo que significa “estar ocupado” en ambas fuentes de datos. Respecto a los ingresos y egresos del sistema de cotización, la ficha y la EPA presentan discrepancias que son explicadas por las diferentes temporalidades en las cuales las fuentes de información ingresan o recopilan los datos.

¹⁰ CPR: Personal identification number, equivalente al número de cedula de identidad en Chile.

¹¹ DES: Danish Family Expenditure Survey

¹² EPA: Encuesta de Población Activa en España

Hace 25 años el Ministerio de Desarrollo Social de Chile ha encargado a la Comisión Económica para América Latina (CEPAL) el proceso de corrección de los errores de medición en los reportes de ingresos declarados en la encuesta CASEN. Para esta labor la CEPAL ejecuta tres procesos: corrección por no respuesta, evaluación de error de medición y ajuste respecto a las Cuentas Nacionales (Ministerio de Desarrollo Social, 2011). El ajuste de la tabla 6 a las Cuentas Nacionales permite identificar los sesgos de sub y sobre estimación de los ingresos declarados en la mencionada encuesta. Para la CASEN 2011 se evidenció el sub-reporte de ingresos de fuentes laborales, por pensiones y por ingreso de rentas de propiedad. El trabajo realizado por la CEPAL permite identificar errores de medición a nivel agregado entre la CASEN y los registros administrativos provenientes de las Cuentas Nacionales. A partir de CASEN 2013 la CEPAL publica los resultados del análisis del error de medición, y no realiza la corrección de los ingresos recopilados a través de la encuesta en función de las cuentas nacionales.

En resumen, la literatura indica que existen factores asociados a la discrepancia resultante entre la comparación de encuestas y registros administrativos a nivel individual de los datos como también a nivel agregado de las estimaciones. A partir de la literatura, es posible inferir que la discrepancia que emerge al comparar encuestas con registros administrativos se produce por diferentes mecanismos¹³. La primera que se mencionará es el error de recuerdo (olvido – memoria), también es importante considerar la mentira deliberada (ocultamiento de información consciente para por ejemplo evitar la pérdida de beneficios sociales), aproximación errónea (redondeo de cifras – facilidad para olvidar el monto exacto de ingresos/gastos), confusión con otros conceptos.

¹³ Permite explicar las discrepancias entre valor verdadero-encuestas y registros administrativos

3. Objetivos de investigación

En la presente investigación se utilizarán como fuentes de información la Encuesta de Protección Social e Historial Previsional de Afiliados, Pensionados y Fallecidos del año 2015. La decisión de utilizar las fuentes de datos mencionadas se basó en que contienen información desagregada, lo que permitirá analizar desde un contexto individual la discrepancia entre la encuesta y registro administrativo. La encuesta está vinculada al Historial Previsional de Afiliados, Pensionados y Fallecidos de las personas encuestadas. Por tanto, parear ambas fuentes de información permitirá contar con información retrospectiva de las unidades encuestadas y además información del comportamiento laboral registrado por los archivos oficiales de la Superintendencia de Pensiones.

Se decidió seleccionar solo a los pensionados de vejez por edad y vejez anticipada con retiro en modalidad vitalicia porque es un grupo homogéneo, y además la mencionada modalidad seleccionada considera un monto fijo mensual, por tanto, esto podría disminuir los errores asociados a la declaración de montos no exactos o alejados del valor objetivo provisto por el HPA 2105.

A continuación, se exponen los objetivos que guiaron el estudio:

- a. Elaborar bandas de tolerancia para descartar de los análisis ejecutados sobre sub/sobre reportes de estimación, errores que puedan ser considerados “al azar” de aquellos que puedan ser considerados “no aleatorios” o sistemáticos
- b. Identificar la existencia de diferencias significativas en aquellas unidades en las que se observan *missing data* en las variables críticas (monto de pensión de vejez por edad y anticipada modalidad de retiro vitalicia) respecto al grupo de personas que cuenta con información, para identificar si determinadas variables sociodemográficas y de conocimiento sobre el Sistema Previsional se relacionan con el no reporte de ingresos por conceptos de pensión de vejez en modalidad vitalicia en la EPS 2015
- c. Cuantificar la discrepancia a nivel individual entre Encuesta de Protección Social e Historial Previsional de Afiliados, Pensionados y Fallecidos del año 2015 en relación

a los ingresos por pensión de vejez por edad y anticipada de las personas con modalidad de retiro vitalicia¹⁴.

- d. Determinar porcentualmente la disponibilidad de información en variables críticas (monto pensión de vejez por edad y anticipada modalidad de retiro vitalicia) para la Encuesta de Protección Social e Historial Previsional de Afiliados, Pensionados y Fallecidos del año 2015.
- e. Caracterizar el perfil sociodemográfico y sobre educación financiera de las personas que sub o sobre reportar montos de pensiones de vejez por edad y anticipada.

4. Datos y Métodos

4.1 Fuentes de datos

A continuación, se describirán las fuentes de datos a utilizar para determinar la brecha existente entre la Encuesta de Protección Social y el Historial Previsional de Afiliados, Pensionados y Fallecidos del año 2015 en lo que respecta a ingresos por conceptos de pensiones de vejez en modalidad vitalicia, y además para caracterizar los perfiles de las personas que sub y sobre reportan ingresos en la Encuesta de Protección Social.

4.1.1 Encuesta de Protección Social 2015

La Encuesta de Protección Social, de ahora en adelante EPS, es la primera encuesta longitudinal implementada en Chile a partir del año 2002, y representa una fuente de información relevante sobre el mercado laboral y seguridad social (Subsecretaría de Previsión Social Dirección de Estudios Previsionales Observatorio Previsional, 2016). La encuesta es realizada por encargo de la Subsecretaria de Previsión Social del Ministerio del Trabajo y Previsión Social, y sus principales objetivos tienen que ver con la proporción de información al Estado de Chile y al público en general sobre temáticas de previsión social. La EPS es una herramienta de diagnóstico que permite orientar la toma de decisiones y diseño

¹⁴ Modalidad vitalicia: Pensión con pago de renta mensual fija en UF para toda la vida del afiliado.

de políticas públicas en materia de previsión social, dada la perspectiva longitudinal por el seguimiento del mismo grupo de personas a lo largo del tiempo. Es importante señalar que la EPS es la única encuesta en Chile que se encuentra vinculada a registros administrativos, es decir, el auto reporte sobre la historia laboral es complementado con registros administrativos sobre la situación previsional del encuestado (Bravo, 2010).

La población objetivo de la primera ronda de la EPS consideró a la población chilena afiliada al sistema de pensiones, incluyendo a las personas afiliadas al antiguo sistema de pensiones (sistema de reparto o INP) y a los del sistema de pensiones vigente desde 1981 (sistema de capitalización individual o AFP). Se excluyó a los miembros de las Fuerzas Armadas y del orden y a un pequeño porcentaje de la población chilena que residen en áreas de difícil acceso (Diseño Muestral y Factores de Expansión VI Ronda EPS, 2016).

Se han realizado 6 rondas: 2002, 2004, 2006, 2009, 2012 y 2015, en las cuales se considera el historial laboral y previsional de las personas mayores de 18 años e información sobre el hogar. El marco muestral de la ronda del año 2002 de la EPS se construyó en base a los 8.074.673 afiliados al sistema de pensiones¹⁵, cuya información se obtuvo a partir de los datos otorgados por el Instituto de Previsión Social y la Superintendencia de Administradora de Fondos de Pensiones. Para obtener una muestra representativa y evitando la excesiva dispersión geográfica, se consideró un diseño muestral en dos etapas estratificado por clúster, el cual permite que cada afiliado posea la misma probabilidad de ser seleccionado (Kish, 1965). En una primera etapa, se selecciona comunas como Unidades Primarias de Muestreo. En la segunda etapa se estratificó a los afiliados de las comunas seleccionadas en la primera etapa, en base a las siguientes categorías: condición de afiliado, sexo y edad para luego seleccionar a afiliados y no afiliado de las comunas seleccionadas en la primera etapa. A partir del año 2004 se incluye a la población mayor de 18 años no afiliada al sistema de pensiones. La muestra de afiliados se selecciona en las mismas comunas (unidades primarias de muestreo) de la muestra de afiliados de EPS.

¹⁵ Mayores de 18 años. La encuesta cuenta con los RUT de los afiliados, y aquellos no afiliados se les pregunta en la encuesta. Esto se realiza para vincular la información proveniente de la EPS con los registros administrativos del HPA.

Para efectos del presente análisis se utilizarán los datos de la EPS 2015 cuenta 16.906 encuestados, se decidió utilizar el 2015 porque es la fuente de información más reciente. Para el año 2015 el cuestionario se diseñó en formato papel y posteriormente se programó para ser aplicado en dispositivo electrónico. El cuestionario final constó de los siguientes módulos

Tabla 1: Módulos de la Encuesta Protección Social 2015 – Subsecretaría de Previsión Social

Módulo EPS 2015	Información contenida en el módulo
Módulo 1	Información general de los integrantes del hogar
Módulo A	Educación
Módulo B	Historia Laboral
Módulo C	Ingresos del entrevistado e ingreso total del hogar
Módulo D	Activos y patrimonio del entrevistado y del hogar
Módulo E	Protección social
Módulo F	Salud
Módulo G	Capacitación
Módulo I	Historia Individual
Módulo J	Calidad de vida
Módulo Q	Coyuntura previsional y opinión
Módulo H	Historia familiar Situación laboral, de salud y previsional del resto de los integrantes del hogar

Fuente: Elaboración propia a partir de Manual de Usuarios EPS 2015

La EPS 2015 se realizó en Chile continental durante los meses de abril a julio del año 2016; en esta fase se lograron 18.007¹⁶ entrevistas. Para acceder a las bases de datos de la EPS, se debe solicitar a través de la plataforma de transparencia de la subsecretaría de Previsión Social del Ministerio del Trabajo y Previsión Social¹⁷.

Es importante mencionar las tasas de respuesta de la EPS 2015, la tasa de cooperación¹⁸ (COOP) fue de un 89,5% en la muestra panel y un 88,2% en la muestra refresco, la tasa de rechazo¹⁹ (REF) es de un 5,2% en la muestra panel y un 3,8% en la muestra refresco.

¹⁶ N:18.007. Casos fallecidos n:997, los casos impedidos para responder el cuestionario n:104, y los casos de entrevistados vivos n:16.906.

¹⁷ Enlace para elevar solicitud de datos de rondas de EPS:

<https://www.previsionsocial.gob.cl/sps/biblioteca/encuesta-de-proteccion-social/bases-de-datos-eps/>

¹⁸ Proporción de personas entrevistadas sobre el total de personas contactadas, se excluye a fallecidos e impedidos

¹⁹ Proporción de la muestra que rechaza el estudio, se excluye a fallecidos e impedidos

4.1.2 Historial Previsional de pensionados y fallecidos 2015

El Historial Previsional de Afiliados, Pensionados y Fallecidos, ahora en adelante HPA, es una base de datos facilitada por la Superintendencia de Pensiones de Chile, que contiene información sobre el comportamiento laboral y previsional de la población afiliada al sistema de pensiones (Morales, 2015).

En el HPA contiene información sobre el historial de ingresos disponibles y sus fechas de devengamiento o cobro de montos por parte de los pensionados, lo cual permite contabilizar la cantidad de meses en los cuales efectivamente las personas han recibido ingresos por conceptos de pensiones de vejez y además el monto que reciben. Las bases de datos del HPA se solicitan a través de la plataforma de transparencia de la subsecretaría de Previsión Social del Ministerio del Trabajo y Previsión Social²⁰.

Para el año 2015 el HPA está desplegado en diferentes bases de datos que contienen la siguiente información.

Tabla 2: Historial Previsional de Afiliados, Pensionados y Fallecidos– Superintendencia de Pensiones

N°	Nombre BBDD	Información contenida en la BBDD	Número de casos	Formato base de datos
1	Características afiliadas	Información Individual	28.135	DTA
2	Información mensual	Registro histórico en Cuenta de Ahorro de Indemnización	60.045	DTA
3	Información mensual	Registro histórico en Cuenta de Ahorro Voluntario	209.564	DTA
4	Información mensual	Registro histórico en Cuenta de Capitalización Individual de Cotizaciones Obligatorias	3.852.473	DTA
5	Información mensual	Registro histórico en Cuenta de Capitalización Individual de Cotizaciones Voluntarias	58,522	DTA
6	Información mensual	Registro histórico de Capitalización Individual de Depósitos Convenidos	13.365	DTA
7	Información mensual	Registro histórico de Pagos de APS	80.910	DTA
8	Información mensual	Registro histórico de Saldos de Cuentas	3.746.565	DTA
9	Información mensual	Registro histórico en cuenta de Capitalización Individual de	435	DTA

²⁰ Enlace para elevar solicitud de datos de rondas de HPA:

<https://www.previsionsocial.gob.cl/sps/biblioteca/encuesta-de-proteccion-social/bases-de-datos-eps/>

		Afiliado Voluntario		
--	--	---------------------	--	--

Fuente: Elaboración propia a partir de Descripción de Historias previsionales 2016

Se optó por utilizar sólo la base de datos “Características Afiliados” puesto que otorga información respecto a la modalidad de retiro vitalicia, modalidad seleccionada para el presente análisis, y además corresponde a los pagos de los afiliados para el año 2016. La base del HPA cuenta con 28.135 personas registradas y contempla a aquellas que fueron encuestadas en alguna de las rondas de la EPS a partir del año 2009, es importante señalar que esta submuestra se construye en base a la muestra histórica de la EPS 2002. La muestra nominal EPS 2015 está compuesta de la siguiente manera:

- Afiliados 2002: 15.060
- No afiliados 2004: 2.673
- Nuevos afiliados 2004: 916
- Refresco 2015: 14.995

Un total de 33.644 casos para la muestra nominal EPS 2015, es importante señalar que para el HPA 2015 solo está disponible para 28.135 de los casos provenientes de la muestra nominal EPS 2015.

5. Variables

5.1 Auto reporte de Ingresos por pensiones de vejez -EPS

Los auto reportes sobre ingresos por conceptos de vejez fueron recolectados en la EPS a través de la pregunta “¿Cuál es el monto de su pensión?”²¹. Este monto fue registrado por los encuestadores en pesos chilenos corrientes del año 2015 y 2016 (según fecha de entrevista), y es una de las variables de interés en la presente investigación, puesto que con la información contenida en esta variable se calculó la brecha que emergió de la comparación de la EPS y el HPA 2015 en lo que refiere a pensiones de vejez en modalidad de retiro vitalicia.

²¹ La EPS 2015 en la pregunta ¿Cuál es el monto de su pensión? No se alude un periodo de tiempo de referencia para que los encuestados respondan.

5.2 Registro objetivo de pensiones de vejez – HPA

En el HPA la variable de interés está nombrada como “*montopen*”, que refiere a los montos de pensión de vejez recibidas en *modalidad vitalicia*²² por las personas registradas. Se decidió utilizar esta modalidad (vitalicia) ya que otorga un monto fijo mensual, por tanto, se tiende a suponer que las personas podrán recordar fácilmente los ingresos recibidos. Este escenario podría permitir suponer que las declaraciones realizadas en la EPS 2015 sobre ingresos de pensiones de vejez podrían contar con una menor presencia de errores asociados al no recuerdo del monto exacto por parte de los entrevistados, esto por la inexistencia de fluctuaciones en los montos recibidos a diferencia de lo que acontece con la modalidad de retiro programada.

Fue necesario convertir la variable “*montopen*” del HPA 2015 a pesos chilenos del año 2016, dado que esta última se encontraba en UF. Se tomó esta decisión ya que los ingresos por pensiones de vejez en la EPS se encuentran registrados en pesos chilenos; al homologar las unidades de medidas se pudo realizar una comparación adecuada de ambas fuentes de información.

A continuación, la expresión que permite visualizar esta relación (2)

$$y_i = montopen_i \times UF \quad (1)$$

El término y_i corresponde al monto de pensión en pesos chilenos para cada persona de la base de datos del HPA del año 2015; el segundo término $montopen_i$ es el monto de ingreso de pensión de cada persona registrado en el HPA en UF. Finalmente, el último término corresponde al promedio de la Unidad de Fomento entre abril y julio del 2016 por la cual se multiplicará el $montopen_i$, para obtener el monto de la pensión en pesos al igual que en la EPS.

²² Existen dos modalidades de retiro de pensiones de vejez en Chile. La primera es la *renta vitalicia*, en la cual se otorga una pensión con monto fijo en UF y una Compañía de Seguros es la encargada de la transferencia monetaria. A diferencia de la *renta programada*, en la cual el monto de pensión se calcula año a año en función del saldo de la cuenta individual y otros factores, y la transferencia monetaria la realiza la AFP que está a cargo de la cuenta de capitalización individual del afiliado.

5.3 Sub/sobre reporte de ingresos de pensiones de vejez

Para cuantificar la discrepancia entre la EPS y el HPA del año 201 en relación a los ingresos por pensión de vejez en modalidad de retiro vitalicia de las personas, se construyeron dos indicadores que serán definidos a continuación:

- 1) Brecha Absoluta: Es la discrepancia observada en pesos chilenos del año 2016. Corresponde a la resta simple entre los ingresos declarados en la EPS-*valor observado*- y la información registrada en el HPA-*valor objetivo*-. La ecuación (2) permitió calcular la brecha a nivel individual y así cuantificar la discrepancia entre ambas fuentes de datos:

$$e_i = | (y_i - t_i) | \quad (2)$$

La ecuación (2) corresponde al valor absoluto de la diferencia entre el monto de pensión vitalicia declarado por la persona i en la EPS 2015 y_i y el monto de pensión vitalicia registrado a nombre de la persona i en el HPA 2015. El término e^i se refiere a la brecha absoluta que existe para cada unidad de la base de datos en relación a los montos declarados y conceptos de pensiones de vejez en modalidad de retiro vitalicia. El término y^i da cuenta del *valor observado* o declarado por la unidad encuestada en la EPS 2015 respecto al ingreso por pensión de vejez. El término t^i da cuenta del *valor verdadero* que corresponde a los registros administrativos sobre pensiones de vejez recibidas que proporciona el HPA.

- 2) Brecha Relativa: Corresponde al porcentaje de la brecha absoluta en base a lo registrado en el HPA 2015. La ecuación (3) permitió conocer el porcentaje de la brecha absoluta calculada para cada individuo de la muestra.

$$BR = \frac{e^i}{t^i} \times 100 \quad (3)$$

La ecuación (3) es el porcentaje individual de la brecha respecto a lo registrado como *valor verdadero o gold standard* en el HPA 2015 respecto a los montos por conceptos

de pensiones de vejez en modalidad de retiro vitalicia. El termino e^i corresponde a la brecha absoluta para cada unidad encuestada en EPS y registrada en HPA 2015; el termino t^i da cuenta de los registros administrativos individuales del HPA del 2015.

5.4 Límites de tolerancia para discrepancia

En todo ejercicio de medición existe espacio para que emerjan errores aleatorios, en el caso del reporte de montos de ingresos este podría provenir de errores del recuerdo/memoria o a errores de redondeo de cifras. Por esta razón, estudios que analizan discrepancias entre dos valores asociados al mismo fenómeno consideran “límites de tolerancia” para los errores aleatorios, que permitan descartar de los análisis los errores de sub/sobre reportes de estimación que puedan ser considerados “al azar” de aquellos errores que van mas allá de lo razonable. Existen diversas investigaciones que establecen límites o bandas de tolerancia para analizar las discrepancias entre los valores observados y valores objetivos, en el caso del presente estudio se consideró valor observado lo declarado individualmente en la EPS 2015 y valor objetivo lo registrado en el HPA 2015. Bowring (2012) realizó un análisis comparado sobre el peso declarado y la medición objetiva realizada por profesionales del área de la salud. Estableció como concordantes aquellas mediciones en las que se obtuvo una diferencia de menos de 2 kilos para peso. La estrategia de determinar límites de tolerancia permite depurar y reducir las discrepancias que emerjan por diferentes tipos de errores de medición y técnicos.

En el presente estudio también se establecieron límites de tolerancia para la discrepancia resultante entre la comparación de las fuentes de información mencionadas respecto a los ingresos de pensiones de vejez en modalidad de retiro vitalicia para el año 2015. El límite se elaboró utilizando el SE ²³, construyendo de esta manera un margen de error que permitió calcular las bandas de tolerancia. A continuación, la ecuación (4) expone la construcción de los límites de tolerancia:

²³Error estándar (SE).

$$\text{Tolerancia de la Brecha Absoluta} = SE(y_i) \times 1.96 \quad (4)$$

Con el resultado de la ecuación (4) se determinaron bandas que permitieran diferenciar aquellas personas que sub reportaron, sobre reportaron y aquellas que coincidieron²⁴ en los montos declarados en la ESP respecto al HPA del año 2015. Aquellos montos de la Brecha $\leq \$ 6.933$ serán considerados como Sub reportes, aquellos $\geq \$- 6.933$ serán considerados como Sobre reportes, y los que se encuentren entre $\geq \$- 6.933$ y $\$ 6.933$ serán catalogados como coincidencias o concordancias.

5.5 Otras Variables

Para la ejecución de los modelos de regresión logística se construyeron variables que permitieron comparar a las personas que Sub reportan respecto a las coincidentes (1), a las que Sobre reportan respecto a las coincidentes (2) y las que Sub y Sobre (3) reportan respecto a las que coinciden en los montos de ingresos de vejez según los límites de tolerancias establecidos anteriormente. A continuación, se expone una tabla resumen de las variables mencionadas:

Tabla 3: Descripción variables modelos regresión logística

Variable	Categorías	N° de pregunta EPS y variable HPA	Descripción
(1) Sub reporte respecto a las coincidentes	Sub reporte:1 Coincidentes:0	EPS: c19m_1 HPA: monto_pen	Personas que declaran montos de ingresos por pensiones de vejez en EPS 2015 por debajo de \$ 6.933 pesos chilenos respecto a las personas que declaran montos de ingresos por pensiones de vejez en EPS 2015 entre $\geq \\$- 6.933$ y \$ 6.933 pesos chilenos.

²⁴ Coincidencia según el límite de tolerancia estipulado ($\geq \$-6.933$ y 6.933)

(2) Sobre reporte respecto a las coincidentes	Sobre reporte:1 Coincidentes:0	EPS: c19m_1 HPA: monto_pen	Personas que declaran montos de ingresos por pensiones de vejez en EPS 2015 por sobre de \$ 6.933 pesos chilenos respecto a las personas que declaran montos de ingresos por pensiones de vejez en EPS 2015 entre \geq \$- 6.933 y \$ 6.933 pesos chilenos
(3) Sub/sobre reporte respecto a las coincidentes	Sub/sobre reporte:1 Coincidentes: 0	EPS: c19m_1 HPA: monto_pen	Personas que declaran montos de ingresos por pensiones de vejez en EPS 2015 por sobre de \$ 6.933 y por debajo de \$ 6.933 pesos chilenos respecto a las personas que declaran montos de ingresos por pensiones de vejez en EPS 2015 entre \geq \$- 6.933 y \$ 6.933 pesos chilenos

Fuente: Elaboración propia

Conjuntamente se construyeron variables dicotómicas, la tabla 4 expone un resumen de las variables mencionadas:

Tabla 4: Variables dicotómicas utilizadas en estudio

Variable	Categorías	Nº de pregunta EPS 2015	Descripción
Sexo	Hombre: 1 Mujer: 0	a8	Variable para identificar brechas existentes entre los géneros estipulados por norma social. Se utilizó para determinar si el sexo influye en la

			conducta de sub/sobre reportar ingresos y además en la conducta de no reportar ingresos en la EPS 2015. Además, existen estudios que vinculan el sexo masculino con mayor conocimiento sobre Sistema Previsional y Financiero ²⁵ .
Edad	Tramo 58 -74 años: 1 Tramo 75 a 93 años: 0	a9	La edad permite identificar y clasificar a las personas respecto a dos segmentos.
Estado Civil	Tiene pareja:1 No tiene pareja:0	i1	Esta variable se utilizó para determinar si el Estado Civil de las personas determina las conductas de sub/sobre reportar, como también la conducta de no declarar ingresos en la EPS 2015. Las personas casadas cuentan con mayor conocimiento sobre el Sistema de Pensiones y financiero.
Nivel Educativo	Educación superior:1 Educación media:0	a12n	El nivel educativo podría estar vinculado a las conductas de no reportar ingresos en la EPS 2015 o no contar con información en HPA 2015.

²⁵ Lusardi, 2012

Conocimiento sobre el porcentaje de cobro de AFP para administrar fondos	Conocimiento sobre porcentaje de cobro AFP: 1 Sin conocimiento sobre porcentaje de cobro AFP: 0	e12	Escaso conocimiento sobre el Sistema de Pensiones y financiero está asociado al nivel socioeconómico, por tanto, se podría suponer que personas con mayor educación cuentan con un nivel más alto de conocimiento sobre el sistema; y personas con escasa educación son clasificadas en un nivel bajo de alfabetización financiera ²⁶ . Estas variables podrían estar vinculadas al error de medición, particularmente, el error de aproximación del monto exacto /falta de memoria.
Conocimiento sobre la Pensión Básica Solidaria (PBS)	Conocimiento sobre la PBS: 1 Sin conocimiento sobre la PBS: 0	e58	
Conocimiento sobre el Aporte Previsional Social de Vejez (APSV)	Conocimiento sobre APSV: 1 Sin conocimiento sobre APSV: 0	e63	

Fuente: Elaboración propia

6. Base de datos Analítica

En la presente sección se describirá detalladamente la información proveniente de la Encuesta de Protección Social y la del Historial Previsional Administrativo (HPA) de Afiliados Activos, Pensionados y Fallecidos del Sistema de AFP ambas fuentes de información del año 2015. Posteriormente, se describe el matching realizado entre la EPS y HPA 2015, y la construcción de la base de datos que se utilizó para la realización de la presente investigación.

²⁶ Ibid.

6.1 Identificación de pensionados por vejez -edad y anticipada- en modalidad de retiro vitalicia en EPS 2015

Para la construcción del diagrama que ilustra la figura 1 de la EPS²⁷ se estudiaron los documentos metodológicos y los 14 módulos de la encuesta. La EPS es aplicada a un mismo grupo de sujetos en diferentes periodos de tiempo; por tanto, es probable que existan personas que fallecieron, se encuentren imposibilitadas por alguna enfermedad para responder la encuesta y aquellos que continúan vivos para responder la encuesta.

En la EPS 2015 se observan tres tipos de casos: los fallecidos, los impedidos para contestar y los que continúan vivos. Los primeros y segundos no participaron de la ola 2015, pero son registrados en los manuales metodológicos y bases de datos en el caso de los fallecidos. Para el caso de los vivos, se les aplicó el instrumento y la información de este grupo se encuentra distribuida en los 14 módulos de la EPS 2015.

Para contabilizar e identificar los casos fallecidos en la muestra, se recurrió a la base de datos que cuenta con información exclusiva para estas personas; para aquellos que se encontraban impedidos al momento de aplicar la encuesta se recurrió al manual metodológico de la EPS que presenta la cuantificación de este grupo en particular, es importante clarificar que no existe una base de datos pública disponible para analizar las características del mencionado grupo. Los casos logrados en la EPS 2015 son $n=18.007$, los casos fallecidos $n=997$, los casos impedidos para responder el cuestionario $n=104$, y los casos de entrevistados vivos $n=16.906$.

Los casos entrevistados *vivos* ($n=16.906$) contempla a dos grupos de personas: Afiliados al Sistema Previsional y los No afiliados al Sistema Previsional. Para determinar la cantidad de personas afiliadas, se manipuló el módulo E (Protección Social) de la EPS 2015 que contempla información vinculada a la afiliación al sistema previsional del encuestado, conocimientos de diferentes beneficios, hábitos de consumos, gastos, entre otros. Se utilizó particularmente la pregunta e2 (¿A qué sistema previsional está afiliado?), obteniendo de esta manera que los *afiliados al sistema previsional* ($n=12.241$) y aquellos que *no se encuentran afiliados* ($n=4.665$).

²⁷Ver anexos figura N° 1

La categoría de *afiliados al sistema previsional* (n=12.241), agrupa a dos subcategorías los *afiliados al sistema previsional Activos* (n=9.410) y los *afiliados al sistema previsional Pensionados* (n=2.831). (¿Está usted jubilado o pensionado por alguno de los siguientes motivos?), lo cual permitió identificar a los pensionados de *vejez por edad* (n=2.078), *vejez anticipada* (n=268), *de invalidez* (n=145) y *de sobrevivencia* (n=340).

Para que las personas pensionadas del sistema puedan retirar sus ahorros, existen dos modalidades: Modalidad de retiro Vitalicia (1) y Modalidad de retiro Programada (2), y en la EPS 2015 esta información se obtiene de la pregunta e87 (¿Qué modalidad de pensión tiene usted?). En la EPS aquellos *pensionados de vejez por edad* que declaran una *modalidad de retiro Vitalicia* son n=516, aquellos *pensionados de vejez por edad* que declaran *otra modalidad de retiro* son n= 410, y aquellos que aún retiran *bajo el sistema antiguo* son n=563. Conjuntamente, aquellos *pensionados de vejez anticipada* que declaran una *modalidad de retiro Vitalicia* son n=163, aquellos *pensionados de vejez anticipada* que declaran *otra modalidad de retiro* son n=53, y aquellos que aún retiran *bajo el sistema antiguo* son n=23.

Para efectos de la presente investigación se utilizará la información proveniente de los “Afiliados al Sistema Previsional, Pensionados por vejez de edad y vejez anticipada que declaren modalidad de retiro Vitalicia”, obteniendo un **n=679** (pensionados de vejez por edad - modalidad de retiro Vitalicia **n= 516** - pensionados de vejez anticipada - modalidad de retiro Vitalicia **n=163**).

6.2 Identificación de pensionados por vejez -edad y anticipada- en modalidad de retiro vitalicia en Historial Previsional Administrativo de Afiliados Activos, Pensionados y Fallecidos del Sistema de AFP – 2015

El Historial Previsional Administrativo (HPA) de Afiliados Activos, Pensionados y Fallecidos del Sistema de AFP, es un registro administrativo elaborado por la Superintendencia de Pensiones que contiene información sobre las personas afiliadas al Sistema Previsional. Los registros administrativos correspondientes al año 2015 son la muestra original que corresponde a la muestra teórica de la Encuesta de Previsión Social 2015; por lo que es posible que no todas las personas estén afiliadas al sistema de AFP

Los registros administrativos del HPA²⁸ 2015 están contenidos en 9 archivos que cuentan con la siguiente información: AFP del afiliado, los saldos en diferentes tipos de fondos A-B-C-D-E, las cotizaciones, existencia de cuenta de capitalización individual de depósitos convenidos, Valor actualizado total del (los) bono(s) de reconocimiento del afiliado, Indicador de existencia de cuenta de ahorro voluntario, Indicador de existencia de cuenta de capitalización individual de cotizaciones voluntarias, Indicador de existencia de cuenta de ahorro e indemnización, Indicador de existencia de cuenta de capitalización individual de depósitos convenidos, Indicador de existencia de cuenta de capitalización individual de afiliado voluntario, Indicador de existencia de cuenta de ahorro voluntario colectivo, Indicador para cantidad de pensiones, Fecha de solicitud de pensión, fecha de cambio de modalidad de pensión de RP a RV, Modalidad de pensión, Monto de la pensión de la RV contratada en UF, Monto de la prima traspasada en UF, Indicador Saldo Fondo A igual a cero, Indicador Saldo Fondo B igual a cero, Indicador Saldo Fondo C igual a cero, Indicador Saldo Fondo D igual a cero, Indicador Saldo Fondo E igual a cero, Indicador Bono Reconocimiento igual a cero, Indicador Monto Pensión igual a cero, Indicador Monto Prima igual a cero, sexo, id de persona.

Para la construcción del diagrama que identifica a los pensionados por Vejez-por edad y anticipada-en el HPA 2015 que se ilustra en la figura 2 se utilizó la base de datos “características_afiliados.csv”, la cual contiene información individual de cada afiliado a la fecha: Fecha nacimiento, fecha fallecimiento, fecha afiliación, AFP afiliado, Saldo en los diferentes tipos de fondos, Tipo de pensión, monto de pensión recibida, tipo de modalidad de retiro, entre otros. El número de filas (casos) de la base en cuestión es de 28.135, y contempla a aquellas personas que fueron encuestadas en alguna de las rondas de la EPS a partir del año 2009, es importante señalar que esta submuestra se construye en base a la muestra histórica de la EPS 2002.

Al igual que en la EPS 2015, es posible identificar a los *afiliados al sistema previsional Activos* ($n=23.234$) y *afiliados al sistema previsional Pensionados* ($n=4.901$). Los primeros, son aquellas personas que aún se encuentran laboralmente activas en el mercado formal del trabajo, y los segundos son las personas que reciben pensiones. Para identificar a los

²⁸ Ver anexos figura N° 2

pensionados por diferentes motivos se utilizó la variable **tipopen**, la cual contiene información sobre el tipo de pensión de las unidades registradas: *pensionados por vejez por edad* (n=2.886), *pensionados por vejez anticipada* (n=831), *pensionados por invalidez* (n=615) y *pensionados por sobrevivencia* (n=615).

De la misma manera que en la EPS 2015, se identificaron las unidades registradas en el HPA que cuenten con modalidad de retiro vitalicia y modalidad de retiro programada. Para identificar a las unidades registradas respecto a la modalidad de retiro de pensiones en el HPA se utilizó variable **modapen**. Aquellos pensionados *de vejez por edad* que declaran una *modalidad de retiro Vitalicia* son n=2.327, aquellos *pensionados de vejez por edad* que declaran *otra modalidad de retiro* son n=559. Simultáneamente, aquellos *pensionados de vejez anticipada* que declaran una *modalidad de retiro Vitalicia* son n=486, aquellos *pensionados de vejez anticipada* que declaran *otra modalidad de retiro* son n=345. Es importante señalar que no existe información sobre el sistema antiguo en la base de datos utilizada.

Para efectos de la presente investigación se utilizará la información proveniente de los “Afiliados al Sistema Previsional, Pensionados por vejez de edad y vejez anticipada que declaren modalidad de retiro Vitalicia”, obteniendo un **n=2.813** (pensionados de vejez por edad - modalidad de retiro Vitalicia **n=2.327**- pensionados de vejez anticipada - modalidad de retiro Vitalicia **n=486**).

6.3 Match de datos ESP – HPA 2015

Para cuantificar la discrepancia de pensiones de vejez entre la EPS y el HPA 2015, se llevó a cabo un match entre ambas fuentes de información. La variable común en las bases de datos de la EPS y el HPA fue “folio”²⁹, que es un número identificador para cada una de las personas registradas en el HPA y entrevistadas en la EPS 2015. La mencionada variable se

²⁹ El folio identificador para cada unidad encuestada en la EPS está basado en los RUT de las personas, que están disponibles en el marco de muestreo a partir del cual se seleccionaron los afiliados (BBDD de la Subsecretaría de Previsión Social). Esta variable también se encuentra disponible para los registros administrativos: HPA.

utilizó para realizar la complementariedad de las fuentes de información citadas. A partir de este primer procedimiento se generó una base de datos con 14.298 unidades ligadas.

Ambas fuentes de información contienen una variable en común e innominada, que permite identificar en ambas bases de datos a una misma unidad encuestada en EPS y registrada en HPA 2015. Dado este escenario-*contar con una variable en común*- se realizó un proceso de matching entre ambas fuentes de datos a través de la variable en común folio³⁰ y empleando para este procedimiento el software STATA. Este tipo de procedimiento de vinculación de datos es apuntado en la literatura como *Deterministic record linkage* (Ariel, et al, 2014), el cual se caracteriza por enlazar mediante una comparación exacta información proveniente de diferentes fuentes de datos utilizando una variable en común. Cada valor de las variables utilizadas para enlazar ambas bases de datos, en este caso el folio identificador, serán comparados por el software con sus pares. Esto quiere decir que, cuando los registros o información coincide entre ambas fuentes de datos se fundará un enlace completo.

Se considerará el deterministic record linkage a partir de la siguiente ecuación:

$$y_{kij} = \begin{cases} 1 \\ 0 \end{cases}$$

Los términos expuestos corresponden a los registros enlazados (1) y los no enlazados (0), la información de la EPS (*i*) y la proveniente del HPA (*j*) son enlazadas a través de la variable en común (*k*), que en este caso corresponde al folio identificador que se encuentra presente en ambas fuentes de información.

6.4 Depuración base de datos

Desde el HPA se seleccionaron aquellos casos que eran pensionados de vejez por edad y anticipada desde el HPA 2015, es importante señalar que la información proveniente del HPA contiene datos sobre afiliados y pensionados que optan por una modalidad de retiro vitalicia. Este procedimiento permitió identificar la existencia de 2.327 pensionados de vejez por edad y 486 pensionados anticipadamente con modalidad de retiro vitalicia desde el HPA 2015.

³⁰ La variable folio permite identificar a un mismo individuo en EPS y HPA 2015, y con esto obtener información de dos fuentes de información diferentes.

Posteriormente, se realizó el mismo procedimiento, pero con la información proveniente de la EPS 2015 que da cuenta de los pensionados de vejez por edad y vejez anticipada, y además se consideró solo aquellos casos que declaran optar por una modalidad de retiro vitalicia. Este método permitió identificar que desde la EPS 2015 existen 516 pensionados de vejez por edad y 163 pensionados anticipadamente con modalidad de retiro vitalicia.

Se logró una base de datos con 499 casos ligados³¹ a partir de ambas fuentes de información. Al eliminar los casos sin información, los n/s y n/r desde la EPS y HPA en la base con 499 casos, se obtiene finalmente una reducción del **n a 371** casos cuyas características principales son las siguientes: **“Afiliados al Sistema Previsional, Pensionados por vejez de edad y vejez anticipada que declaren/estén registrados en modalidad de retiro Vitalicia, según información de EPS y HPA 2015”**. Es importante clarificar que se identificó que existe un 20% de personas registradas en el HPA 2015 que no cuentan con información para las variables de interés, y un 8% para le EPS 2015.

A continuación, la tabla 5 expone un resumen con las variables utilizadas en el estudio:

Tabla 5: Estadísticas univariadas de variables utilizadas (no ponderadas)

Tipo de variables	Variable	Muestra	Categorías	n	min.	max.	mean	sd
Variables dependientes	Pensión EPS 2015	Todos los elegibles	-	620	29000	1.700.000	216448.7	159298.9
	Pensión HPA 2015	Todos los elegibles	-	1543	25982	1.673.761	257313.1	186089.1
	Pensión EPS 2015	Casos matching con información en monto pensión de vejez	-	461	29000	1.130.000	214558.9	147361.8
	Pensión HPA 2015	Casos matching con información en monto pensión de vejez	-	397	0	1.125.021	248115.7	173926.8
	Pensión EPS 2015	base final (371)	-	371	38000	1.130.000	236467.4	145934.8
	Pensión HPA 2015	base final (371)	-	371	0	1125021	247323.4	173781.5

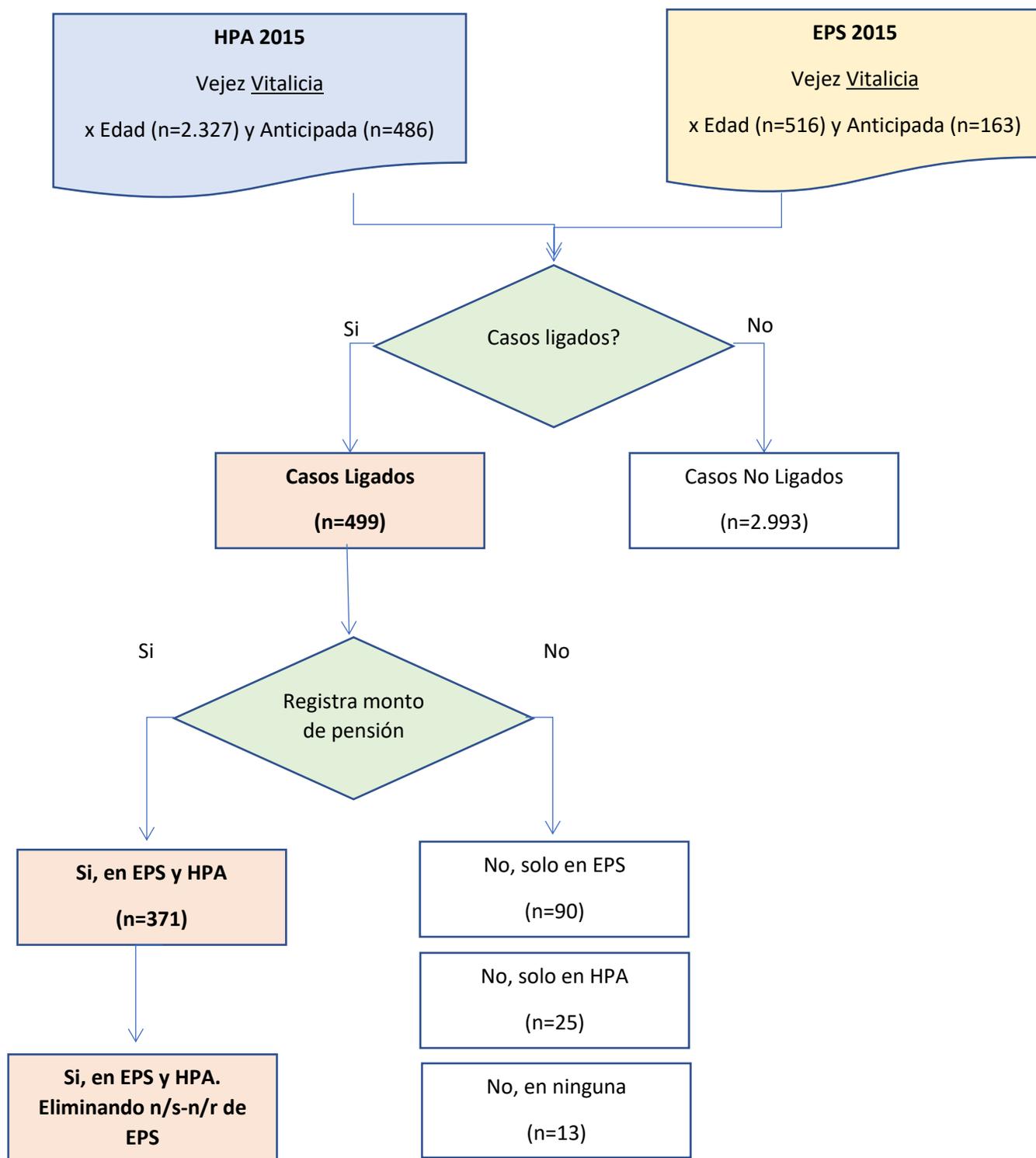
³¹ Pensionados de vejez por edad y anticipada, bajo la modalidad de retiro vitalicia año 2015.

	Sub-reporte	base final (371)	-	159	89000	1.130.000	235375.8	133403.4
	Sobre -reporte	base final (371)	-	142	38000	950.000	258.103	165668.5
	Coincidentes	base final (371)		70	87000	700.000	195057.6	121072.8
	Brecha Relativa	base final (371)	-	371	-952.910	278.115	-318.479	80665.01
	Brecha Absoluta	base final (371)	-	371	47.5	952910	43124.08	68133.93
Variables independientes	Conocimiento PBS	Casos matching	Conoce (1)	202	-	-	54,40%	-
			No Conoce (0)	169	-	-	45,50%	-
	Conocimiento APSV	Casos matching	Conoce (1)	78	-	-	21%	-
			No Conoce (0)	293	-	-	78,90%	-
	Conocimiento cobro AFP	Casos matching	Conoce (1)	13	-	-	4,50%	-
			No Conoce (0)	273	-	-	95,40%	-
Controles demográficos	Edad	Casos matching	Tramo 58 - 74 (1)	273	-	-	73,50%	-
			Tramo 75 - 93 (0)	98	-	-	26,40%	-
	Sexo	Casos matching	Hombre (1)	262	-	-	70,60%	-
			Mujer (0)	109	-	-	29,30%	-
	Educación	Casos matching	Media (0)	287	-	-	80%	-
			Superior (1)	71	-	-	19,80%	-
	Estado Civil	Casos matching	Con pareja (1)	264	-	-	73,70%	-
			Sin pareja (0)	107	-	-	29,80%	-

Fuente: Elaboración propia a partir de EPS y HPA 2015.

A continuación, la figura 1 expone gráficamente el proceso desarrollado para la depuración de la base de datos descrito anteriormente.

Figura 1: Proceso construcción base de datos



Fuente: Elaboración propia a partir de EPS y HPA 2015.

6.5 Estrategia analítica

En primera instancia, se realizaron test de medias para identificar si existen diferencias significativas en aquellas unidades en las que se observan *missing data* en las variables críticas. Se procuró identificar si determinadas variables sociodemográficas y de conocimiento sobre el Sistema Previsional afectan el no reporte de ingresos por conceptos de pensión de vejez en modalidad vitalicia en la EPS 2015, además de observar si existen diferencias en la distribución entre ambas poblaciones. Las variables que fueron utilizadas para el test de medias son las siguientes: Sexo, Estado civil, Educación media completa, Conocimiento sobre el porcentaje de cobro de AFP para administrar fondos, Conocimiento sobre la Pensión Básica Solidaria y Conocimiento sobre el Aporte Previsional Solidario de Vejez.

Conjuntamente, se realizó un análisis respecto a la disponibilidad de datos que ofrecía cada una de las fuentes de información utilizadas en el presente análisis respecto a las variables de interés³². Esto permitió comprender de manera más detallada la pérdida de casos al momento de construir y depurar la base de datos.

Es importante considerar el promedio de la discrepancia individual entre la EPS y el HPA 2015 respecto a pensiones de vejez en modalidad de retiro vitalicia, esto permitirá evaluar si la encuesta está recopilando información adecuada. Además, con los límites de tolerancia estipulados y la clasificación de las unidades encuestadas/registradas en Sub, Sobre reporte y Coincidencia, se logró determinar el porcentaje de los individuos que recae en las diferentes categorías.

Respecto al análisis de la variable dependiente Brecha, se estimó un modelo de regresión logística que permitió identificar si las personas que sub, sobre reportan y coinciden en los montos de ingresos de vejez en modalidad de retiro vitalicia está asociada a ciertas características sociodemográficas de las personas y/o su conocimiento del sistema de pensiones.³³.

³² Montos de pensión de vejez desde HPA- Ingresos por montos de pensión de vejez declarados en EPS 2015

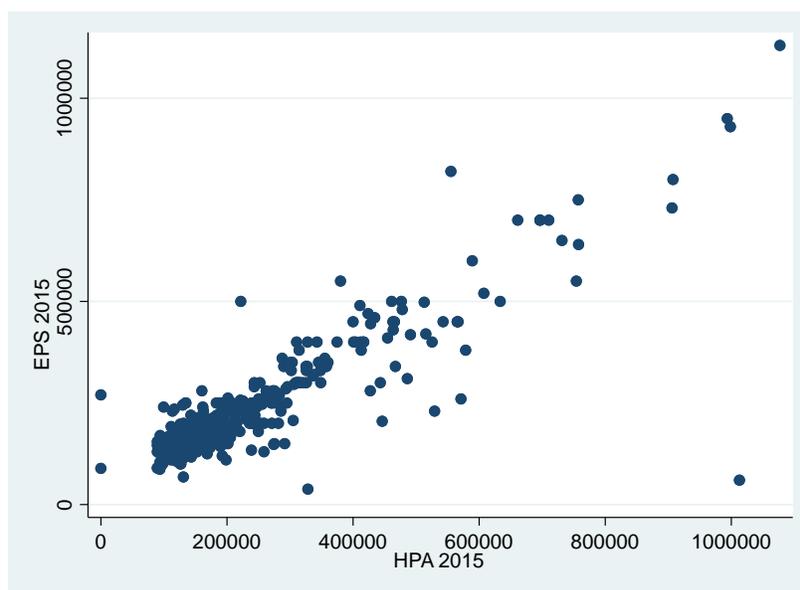
³³ Sexo, Estado civil, Conocimiento sobre el porcentaje de cobro de AFP para administrar fondos, Conocimiento sobre la Pensión Básica Solidaria y Conocimiento sobre el Aporte Previsional Solidario de Vejez.

7. Análisis y resultados

7.1 Cuantificación discrepancia EPS – HPA 2015 respecto a ingresos por pensiones de vejez

La figura 2 presenta la relación que existe entre la EPS y HPA 2015 respecto a los ingresos de pensiones de vejez en modalidad de retiro vitalicia. Se puede observar una correlación positiva ($\text{corr}=0.87$) entre lo declarado en la encuesta EPS 2015 (eje y) y los montos registrados en el HPA 2015 (eje x)

Figura 2: Scatter plot entre EPS y HPA 2015



Fuente: Elaboración propia a partir de EPS y HPA 2015.

La Brecha individual resultante entre los ingresos declarados en la EPS y los recibidos del HPA 2015 cuenta con un promedio de 43.124 pesos chilenos, es decir, las unidades encuestadas sub y sobre reportan en torno a 43.124 pesos chilenos.

A partir de la construcción de las bandas de tolerancia³⁴ que permiten interpretar la variable Brecha, fue posible identificar que un 43% (159) de las personas Sub reporta ingresos por

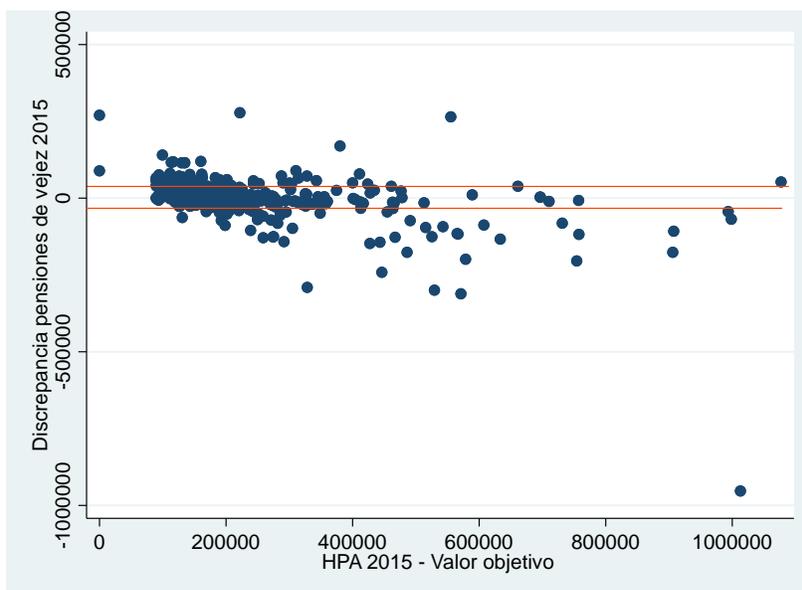
³⁴ Subreporte <6.933

Sobrereporte >6.933

Coincidencia >-6.933 & <6.933

pensiones de vejez, un 38% (142) Sobre reporta y un 19% (70) coincide en los montos declarados respecto a los realmente recibidos. A continuación, se expone en un gráfico de dispersión en la figura 3, en el cual las líneas rojas dan cuenta del 19% de las personas que coinciden en los ingresos declarados en EPS (>-6.933 & <6.933)

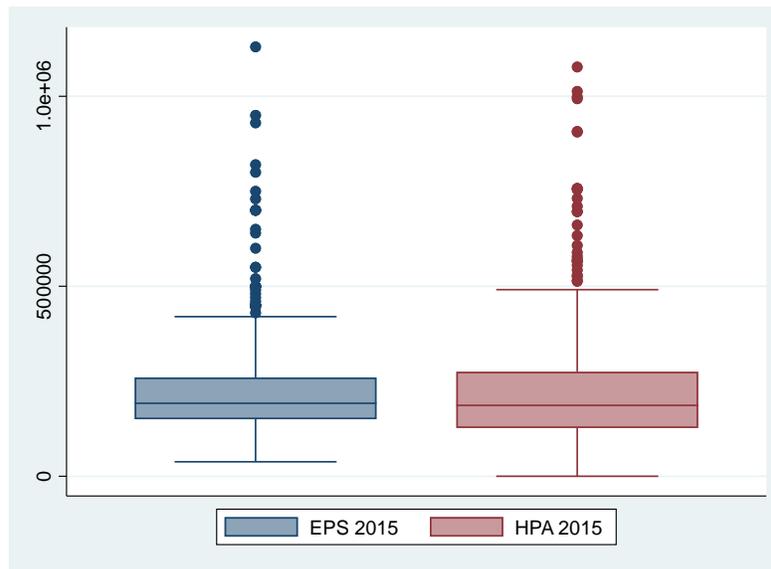
Figura 3: : Scatter plot Discrepancia - HPA 2015 (valor objetivo)



Fuente: Elaboración propia a partir de EPS y HPA 2015.

La figura 4 permite evidenciar que la mediana de las pensiones declaradas (EPS 2015) al ser comparados con los valores objetivos (HPA 2015), cuentan con valores similares. Esto podría indicar que la EPS 2015 está capturando de manera adecuada ingresos por pensiones de vejez en modalidad vitalicia.

Figura 4: Box Plot EPS - HPA 2015



Fuente: Elaboración propia a partir de EPS y HPA 2015.

Es importante señalar que los datos de la EPS 2015 cuentan con menos dispersión que los del HPA 2015, esto podría ser considerado un resultado intuitivo dado que los resultados de las encuestas tienden a ser redondeados, lo cual trae como consecuencia la reducción de la variabilidad de los montos.

7.2 Factores que influyen en el sub y sobre reporte de pensiones de vejez

La tabla 5 presenta los modelos de regresión logística ejecutados para las (1) personas que Sub reportan y Sobre reportan respecto a las personas que son catalogadas como coincidentes en los montos que declaran; (2) personas que Sobre reportan y aquellas que coinciden y (3) quienes Sub reportan respecto de las personas que coinciden en los montos de pensiones de vejez en modalidad vitalicia que declaran en la EPS 2015.

Tabla 6: Modelos de regresión Logística de los indicadores de sub/sub reporte de pensiones. Análisis sin ponderar

Variables		Log Sub/sobre reporte (1)	Log Sobre reporte (2)	Log Sub reporte (3)
Constant	Or	1.679	1.357	0.298
	Std. Err.	(1.119)	(1.020)	(0.269)
Sabe % de cobro de AFP por administrar fondos	Or	2.719	2.169	3.693
	Std. Err.	(2.904)	(2.442)	(4.112)
Sabe sobre la Pensión Básica Solidaria (PBS)	Or	0.397***	0.386**	0.399**
	Std. Err.	(0.136)	(0.144)	(0.158)
Sabe sobre el APS de Vejez	Or	1.524	1.847	1.158
	Std. Err.	(0.634)	(0.825)	(0.560)
En pareja	Or	2.115**	1.561	3.834***
	Std. Err.	(0.819)	(0.629)	-1.856
Edad	Or	0.474	0.627	1.223
	Std. Err.	(0.319)	(0.255)	(.0547)
Hombre	Or	1.785	1.357	2.248
	Std. Err.	(0.743)	(1.020)	(1.130)
Educación superior	Or	2.115	2.014	2.117
	Std. Err.	(1.063)	(1.083)	(1.142)
Observations		276	171	161
R2		0.057	0.060	0.092

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

Fuente: Elaboración propia a partir de EPS y HPA 2015

La tabla 6 permite evidenciar que controlando por sexo, edad y educación la única variable que resulto estadísticamente significativa en los tres modelos es “Sabe sobre la Pensión Básica Solidaria (PSB)”; conjuntamente tanto el modelo 1 como en el 3 la variable “En pareja” cuenta con significancia estadística. De los tres modelos presentados el que posee mejores resultados es el número 3. Mientras el modelo 2 solo posee una variable con significancia estadística, tanto el modelo 1 como el modelo 3 poseen dos variables estadísticamente significativas. Sin embargo, el modelo 1 tiene como variable dependiente el promedio de sub y sobre reportar y quienes que coinciden, es decir, mide la probabilidad de errar sin diferenciar entre si es por estimar menos o estimar más de lo que reciben. Por lo tanto, no se podría realizar un análisis específico del sub o sobre reporte. El modelo 3, por su parte, posee como variable dependiente solo a quienes sub reportan y quienes aciertan. Por ello, la información que entrega este modelo especifica directamente sobre quienes se

equivocan exclusivamente por sub reportar sus pensiones de vejez.

En la tabla 6 y en el modelo 3 se observa que las variables “Sabe sobre Pensión Básica Solidaria (PBS)” y “En pareja” cuentan con significancia estadística al 99%. En el caso de la variable sobre conocimiento de la PBS, las odds de sub- reportar de aquellos que no tienen conocimiento sobre la PBS es de 2,5 veces más altas respecto a quienes sí conocen la Pensión Básica Solidaria. Conjuntamente, la odds de sub- reportar para aquellos que sí tienen pareja es de 3,8 veces más altas que para los que no tienen pareja respecto a las personas que coinciden en los montos de ingresos; por tanto, se podría decir que tener pareja incide considerablemente en sub reportar ingresos de pensiones de vejez.

8. Conclusiones

El objetivo general del presente análisis fue cuantificar la discrepancia a nivel individual entre la Encuesta de Protección Social e Historial Previsional de Afiliados, Pensionados y Fallecidos del año 2015 en relación con los ingresos por pensiones de vejez por edad y anticipada en modalidad de retiro vitalicia. Frente a este desafío, se realizó una estrategia analítica basada en la comparación de dos fuentes de información y además de una revisión de literatura que sugiere interesantes conclusiones.

Se identificó que la mediana de la brecha entre el *valor observado* proveniente de la EPS 2015 al ser comparado con los *valores objetivos* del HPA 2015 es de 43.124 pesos chilenos. Es importante contrastar la brecha con la mediana de las pensiones de vejez registradas en el HPA 2015 que son 236.785 pesos chilenos. La brecha resultante (43.124 pesos chilenos) representa un 18% del valor de las pensiones registradas en el HPA 2015. Por tanto, es posible decir que la discrepancia o brecha resultante a partir de la comparación de ambas fuentes de información podría ser considerada un monto importante. La mencionada discrepancia, podría explicarse por la presencia de elementos asociados a la capacidad de memoria de las personas encuestadas, de las aproximaciones erróneas que realizan (Kapteyn e Ypma, 2006) y por aquellos factores no observados y no controlados en la presente investigación. Este escenario se relaciona directamente con el error de medición expuesto en la revisión de literatura, la diferencia evidenciada entre la EPS y el HPA 2015 podría atribuirse, entre otros factores, a la escasa capacidad de recordar montos exactos de ingresos de pensiones de vejez,

como también a la confusión del ingreso de pensión de vejez por edad o anticipada con otro tipo de pensiones u otro tipo de ingresos.

Es importante señalar, que sí existe la posibilidad de que las concepciones diferentes de ingresos o la confusión entre tipos de pensiones puede generar un incremento en la brecha que existe entre ambas fuentes de información. La investigación desarrollada por Guimarães (2013) evidencia que los errores en las declaraciones de los encuestados sí son factores que pueden generar diferencias entre encuestas y registros administrativos. Al mismo tiempo, la investigación realizada por Kreiner, Dreyey y Petersen (2012) también da cuenta que la existencia de factores asociados a la memoria de las personas podría explicar parte de la discrepancia entre las fuentes de información analizadas y comparadas, ya que una alta proporción de los encuestados no recuerda detalladamente los ingresos y lo que declaran es un promedio de estos. Este panorama también fue comprobado por Domingo Sanz (2011), quien en su investigación observó que las diferentes conceptualizaciones sobre lo que significa “estar ocupado” en ambas fuentes de datos incrementaba la brecha entre las fuentes de información que utilizó; además las discrepancias también son explicadas por las diferentes temporalidades en las cuales las fuentes de información ingresan o recopilan los datos. Es posible decir entonces que la discrepancia mínima resultante en la presente investigación también podría ser explicada por el no recuerdo del monto exacto de los ingresos por pensiones de vejez declarados por las personas en la EPS 2015 y también por otros factores.

Los modelos de regresión logística ejecutados permiten evidenciar que controlando por sexo, edad y educación la única variable que resulto estadísticamente significativa en los tres modelos es “Sabe sobre la Pensión Básica Solidaria (PSB)”, conjuntamente tanto el modelo 1 como en el 3 la variable “En pareja” cuenta con significancia estadística. En el caso de la variable sobre conocimiento de la PBS, las odds de sub-reportar para aquellos que no tienen conocimiento sobre la PBS es de 2,5 veces más altas respecto a quienes sí conocen la Pensión Básica Solidaria; las odds de sub-reportar para aquellos que sí tienen pareja es de 3,8 veces más altas que para los que no tienen pareja respecto a las personas que coinciden en los montos de ingresos.

Al construir y depurar la base de datos que permitió realizar el presente análisis, se redujo la muestra desde 499 a 371 casos con información a nivel individual, este escenario podría estar vinculado al error de no respuesta en ambas fuentes de información (Groen, 2012). Se detectó que 128 casos (26%) no contenían información sobre ingresos de pensiones declarados en la EPS o en los montos recibidos registrados en HPA 2015. Un 82³⁵% de estos casos (missing) provenían de los registros administrativos HPA, cabe recordar que la base de datos utilizada para los efectos de la presente investigación es del año 2015 que presenta las falencias expuestas y se recomienda continuar trabajando en mejoras de disponibilidad, completitud y exactitud de los registros administrativos para la producción de futuras bases de datos estadísticas. Las mejoras que podrían implementarse dada la revisión constante de este tipo de fuentes de información, podrían incentivar el uso de registros administrativos para la realización de diferentes análisis estadísticos.

9.1 Discusión

Es interesante destacar la principal contribución del presente estudio la cual se vincula directamente con la identificación de errores, omisiones por parte de las dos fuentes de información analizadas. No existe evidencia concreta para saber cuál de las dos fuentes está en lo correcto, puesto que como ya se mencionó, ambas cuentan con vacíos y errores de uno u otro tipo. En este escenario, se sugiere continuar la revisión de los sistemas de información que existen en Chile, ya sean estas encuestas o registros administrativos. El panorama idóneo es evaluar las distintas encuestas implementadas en nuestro país a través de la estrategia de comparación de información respecto a registros administrativos que guarden relación con ellas. Esto permitiría complementar la información, evaluar la eficacia e identificar la capacidad de recopilar información actualizada por parte de las encuestas y registros administrativos. Conjuntamente, realizar evaluaciones respecto a la producción de registros administrativos a través del análisis y monitoreo constante de este tipo de información.

En Chile la mayoría de las instituciones públicas generan registros administrativos que se

³⁵ El 82% de unidades registradas en el HPA no cuenta con información en la variable de interés, panorama que implicó una reducción considerable de la muestra a estudiar

conforman, estructuran y almacenan en base a soportes computacionales. Si bien, hoy por hoy existe un gran avance en el desarrollo tecnológico en el manejo de información, en Chile no existe un protocolo coordinado entre las diferentes instituciones que permita dar cuenta que la información recopilada en los registros administrativos sea fidedigna y de calidad para ser transformada en datos estadísticos. Las instituciones en Chile no cuentan con convenios, ni procesos coordinados de validación y control completos que permitan obtener productos estadísticos coherentes, comparables y claros. Dada las condiciones actuales del país y los resultados del análisis, es imprescindible contar con registros administrativos estandarizados. Esto quiere decir que, todas aquellas instituciones que produzcan registros administrativos de cualquier índole debiesen contar con un protocolo que permita ingresar de manera estandarizada la información, como también se debiese velar por el constante monitoreo de la producción de este tipo de información. Esto permitiría contar con el desarrollo de registros administrativos claros, confiables y coherentes para que posteriormente sean utilizados con fines de evaluación de programas, políticas públicas, investigación, etc. El escenario ideal bajo los parámetros internacionales es contar con un repositorio centralizado e integrado de registros administrativos provenientes de diferentes fuentes de información o de diferentes instituciones públicas, los cuales fuesen monitoreados constantemente para velar por su confiabilidad, representatividad y coherencia.

Conjuntamente se evidenció la existencia de dos efectos que explican el sub-reporte de pensiones de vejez. El primero guarda relación con aquellos encuestados que no conocen la Pensión Básica Solidaria (PBS). Aquellos que no tienen conocimiento sobre la PBS es más probable que se equivoquen al tratar de aproximar el monto de pensión de vejez; por tanto, es posible evidenciar que el nulo o escaso conocimiento sobre el sistema podría afectar la declaración de un monto exacto o que se encuentre dentro de los límites de tolerancia estipulados. Esto quiere decir que a medida que las personas cuenten con un mayor conocimiento sobre el Sistema Previsional se reducen las posibilidades de errar en los montos de pensiones de vejez declarados en la EPS. En otras palabras, una persona que se encuentra más informada respecto a la Pensión Básica Solidaria tiene mayores probabilidades de coincidir en el monto declarado en la ESP respecto a lo realmente recibido y registrado en el HPA 2015. El segundo factor se vincula con que las personas que se encuentran emparejadas tienden a sub reportar respecto a aquellos que no están emparejados,

este fenómeno no es analizado en las investigaciones mencionadas en la presente tesis, a pesar de esto es posible aventurarse con una hipótesis: algunas personas emparejadas al no manejar información de manera individual (cuentas conjuntas) tienden a equivocarse y a olvidar los montos exactos de la pensión de vejez que recibe de manera personal. La investigación realizada por Kreiner, Dreyey y Petersen (2012) evidencia que existen factores asociados a la memoria de las personas que podrían explicar parte de la discrepancia entre encuestas y registros administrativos.

Si bien, el presente análisis ha otorgado resultados relevantes es imprescindible indicar la limitante del estudio. El análisis sólo se centra en pensiones de vejez, por tanto, sería interesante evaluar el comportamiento de la encuesta respecto a otro tipo de pensiones. Se decidió analizar a la población que decide retirar sus pensiones en modalidad vitalicia, al igual que el punto anterior es impórtate considerar otras modalidades de retiro para desarrollar una evaluación más extensa de la EPS. Se sugiere el desarrollo de nuevos análisis que consideren un espectro amplio de las diversas pensiones que se otorgan en nuestro país, para con esto demostrar la eficacia de la EPS al recopilar este tipo de información sensible.

Continuando en la misma línea, es importante que futuras investigaciones consideren la revisión de los casos sin información provenientes del HPA. Este escenario pone de manifiesto que los datos administrativos cuentan con omisiones que al no ser subsanadas en el tiempo, podrían distorsionar la evidencia empírica que se utiliza para la toma de decisiones en política pública. Si bien las encuestas no están exentas de errores, es imprescindible que se continúen realizando esfuerzos para identificar los factores que podrían inducir a errores en este tipo de instrumentos.

Bibliografía

- Ariel, A., Bakker, B., Groot, M., van Grootheest, G., der Laan, J., Smit, J., & VerkerkB. (2014). *Record Linkage in Health Data: a simulation study* (pp. 3 -63). Netherlands: CBS.
- Banco Central (2017). *Segunda Conferencia de Estadística: Midiendo la economía en un mundo globalizado*
- Bennett, D. (2001). How can I deal with missing data in my study?. *Australian And New Zealand Journal Of Public Health*, (25), 464 -469. Retrieved from <https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1467-842X.2001.tb00294.x>
- Biemer, P. (2011). TOTAL SURVEY ERROR DESIGN, IMPLEMENTATION, AND EVALUATION. *Public Opinion Quarterly*, (74), 817 - 848. Retrieved from <https://pdfs.semanticscholar.org/e032/7460b101f1aceb27de92d2ca4a133da84ea5.pdf>
- Bravo, D. (2010). *Encuestas Longitudinales en Chile* (pp. 1 - 24). Santiago.
- Bowring, A. L., Peeters, A., Freak-Poli, R., Lim, M. S., Gouillou, M., y Hellard, M. (2012). Measuring the accuracy of self-reported height and weight in a community-based sample of young people. *BMC medical research methodology*.
- Calderwood L. & Lessof C. (2009). Chapter 4 “Enhancing Longitudinal Surveys by Linking to Administrative data”. *Methodology of Longitudinal Surveys*.
- Centro UC Encuestas y Estudios Longitudinales. (2016). Informe Final IV Ronda EPS. *Ministerio Del Trabajo y Previsión Social*.
- CEPAL (2003) Registros administrativos, calidad de los datos y credibilidad pública: presentación y debate de los temas sustantivos de la segunda reunión de la Conferencia Estadística de las Américas de la CEPAL. División de Estadística y Proyecciones Económicas.

- CEPAL (2017). *Informe: Potenciando el uso de registros administrativos con fines estadísticos para el seguimiento de la agenda 2030*.
- Diseño Muestral y Factores de Expansión VI Ronda EPS – Centro UC de Encuestas Longitudinales, Ministerio del Trabajo y Previsión Social. Santiago de Chile, 2016
- Domingo Sanz, M. (2011). El Empleo y el Paro: Comparación de resultados de la EPA con los de Fuentes Administrativas. *Economía Española Y Protección Social*, (3), 1 - 26.
- Echegoyen, G. (2003). Registros administrativos, calidad de los datos y credibilidad pública: presentación y debate de los temas sustantivos de la segunda reunión de la Conferencia Estadística de las Américas de la CEPAL. *Serie De Estudios Estadísticos Y Prospectivos*,(21), 2 - 46. Retrieved from https://repositorio.cepal.org/bitstream/handle/11362/4725/1/S0311768_es.pdf
- Encuestas y Estudios Longitudinales UC. (2016). *Imputaciones de Corte Transversal de la VI Ronda de la Encuesta de Protección Social* (pp. 8 - 29). Santiago.
- Gazola Hellmann, A. (2015). *¿Cómo funciona Bolsa Familia?* (pp. 4 - 30). Banco Interamericano de Desarrollo – División de Protección y Salud. Retrieved from https://publications.iadb.org/bitstream/handle/11319/7210/como_funciona_bolsa_familia.pdf
- Groen, J. (2012). Sources of Error in Survey and Administrative Data: The Importance of Reporting Procedures. *Journal Of Official Statistics*, (28), 173 - 198. Retrieved from <http://www.sverigeisiffror.scb.se/contentassets/f6bcee6f397c4fd68db6452fc9643e68/sources-of-error-in-survey-and-administrative-data-the-importance-of-reporting-procedures.pdf>

- Groves, R., Fowler Jr, F., Couper, M., Lepkowski, J., Singer, E., y Tourangeau, R. (2004). *Survey methodology*. J. Wiley.
- Groves, Robert, Floyd J Fowler, Mick Couper, James Lepkowski, Eleanor Singer, and Roger Tourangeau. 2009. *Survey Methodology*. 2nd ed. Hoboken, NJ: John Wiley & Sons.
- Guimarães Ferreira de Souza*, P. (2013). Uma metodologia para explicar diferenças entre dados administrativos e pesquisas amostrais, com aplicação para o Bolsa Família e o Benefício de Prestação Continuada na PNAD. *R. Bras. Est. Pop., Rio De Janeiro*, (30), 299 - 315. Retrieved from <http://www.scielo.br/pdf/rbepop/v30n1/v30n1a14.pdf>
- Kapteyn, A., & Yeb Ypma, J. (2006). *Measurement Error and Misclassification: A Comparison of Survey and Register Data* (pp. 1 - 38). Santa Monica. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.574.8573&rep=rep1&type=pdf>
- Kish, L. (1965). *Survey sampling*. New York: Chichester.
- Künn, S. (2015). The challenges of linking survey and administrative data. *I Z A World Of Labor: Evidence -Based Policy Making*, 1 - 10.
- Madow, W. Nisselson, H. Olkin, I & Rubin, D. (1983). *Incomplete Data in Sample Surveys*, vols 1 -3. New York
- Lusardi, A. (2012), Numeracy, Financial Literacy and Financial Decision-Making, National Bureau of Economic Research Working Paper N° 17821
- Marsden, P. and Wright, J. (2010). *Handbook of Survey Research*. 2nd ed. USA
- Ministerio de Desarrollo Social (2011). *Diseño y Evaluación de Preguntas Módulo de*

Ingresos Encuesta de Caracterización Socioeconómica Nacional – CASEN.

- Morales, Lema (2015). *Incentivos Financieros y de Salud al Retiro* (Tesis Magister). Facultad de Economía y Negocios, Universidad de Chile. Santiago, Chile
- Paul P. Biemer; Total Survey Error: Design, Implementation, and Evaluation, *Public Opinion Quarterly*, Volume 74, Issue 5, 1 January 2010, Pages 817–848, <https://doi.org/10.1093/poq/nfq058>
- Sakshaug, J & Bender, B (2014). Using Administrative Data: Strengths and Weaknesses. Institute for Employment Research (IAB) Nuremberg, Germany SRMS Webinar.
- Sakshaug, J, Couper, M, Beth, M Ofstedal and David R. Weir (2012). Linkin Survey and Administrative records: Mechanisms of consent. Published in final edited form as: *Sociol Methods Res.* 2012 November; 41(4): 535–569.
- Smith, T. (2011). Refining the Total Survey Error Perspective. *International Journal Of Public Opinion Research*, (23), 464–484.
- Subsecretaría de Previsión Social Dirección de Estudios Previsionales Observatorio Previsional, M. (2016). *VI RONDA EPS 2015 Una mirada a la protección social en Chile*. Presentation, Santiago, Chile.
- Rodríguez-Oreggia, Eduardo, & López-Videla, Bruno. (2015). Imputación de ingresos laborales. Una aplicación con encuestas de empleo en México. *El trimestre económico*, 82(325), 117-146.
- Traugott, Michael W. (2012). “Data Quality from Low Cost Data Collection Methodologies.” Ponencia presentada en la conferencia anual de la AAPOR, Orlando, FL.

- Thustrup Kreiner, C., Dreyer Lassen, D., & Leth-Petersen, S. (2012). *Measuring the Accuracy of Survey Responses using Administrative Register Data: Evidence from Denmark*. Lecture, Washington, DC.
- UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE. (2010). *Micro-integration: State of the Art Note by the Statistics Netherlands* (pp. 1 - 28). The Hague, The Netherlands.
- Villatoro, P., & Cecchini, S. - CEPAL (2018). ¿Cuál es el alcance de las transferencias no contributivas en América Latina? Discrepancias entre encuestas y registros. *Serie Estudios Estadísticos*, (96), 3 - 43. Retrieved from https://repositorio.cepal.org/bitstream/handle/11362/43642/1/S1800299_es.pdf
- Whiteley, S. (2014). *Total Survey Error & Institutional Research: A case study of the University Experience Survey* (pp. 1 33). Melbourne, Australia: Australasian Association for Institutional Research (AAIR).
- Zhang, L. (2012). Topics of statistical theory for register-based statistics and data integration. *Statistica Neerlandica*

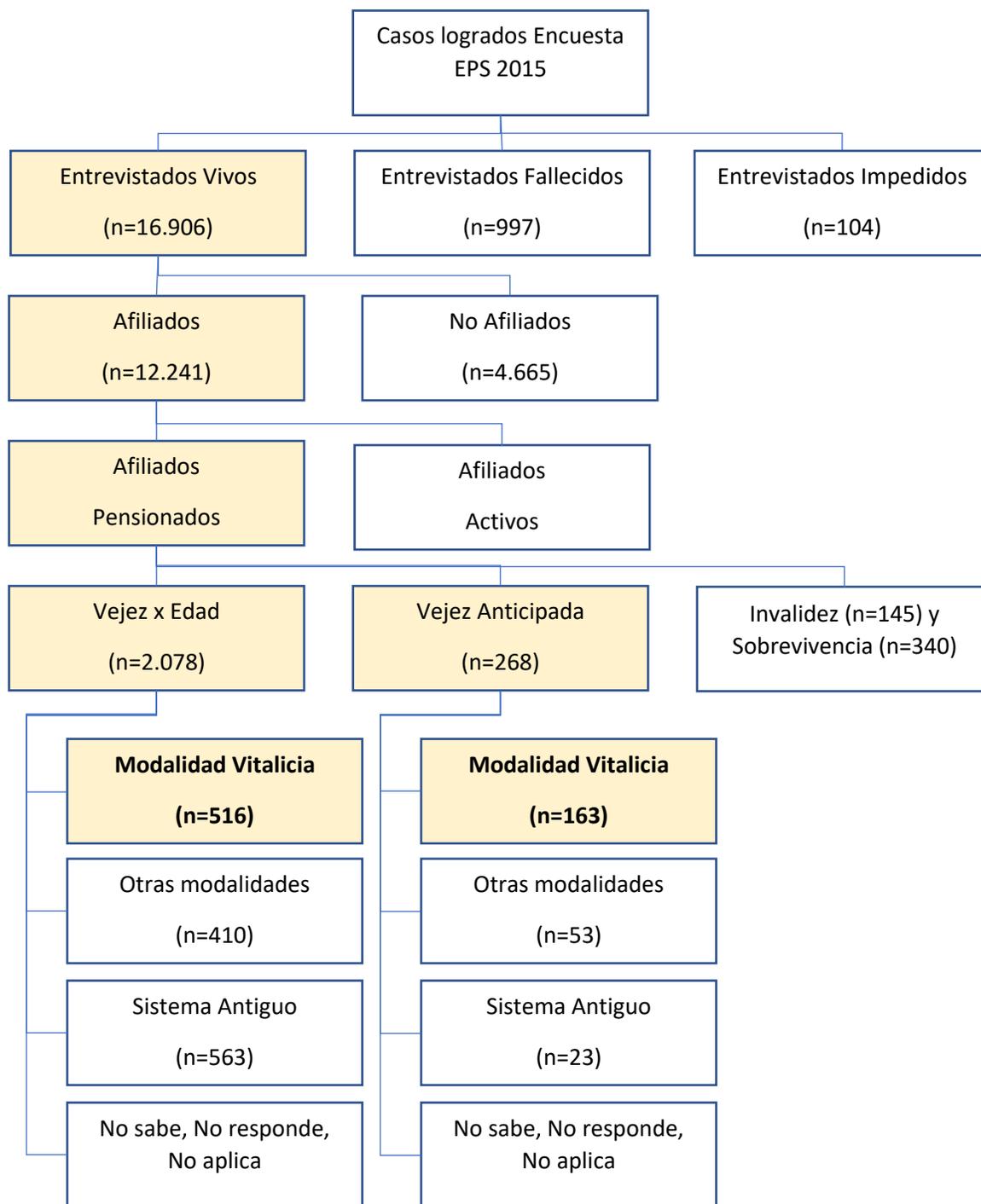
Anexos

Porcentaje de casos perdidos (missing), según conocimiento del sistema de pensiones y características demográficas de los entrevistados (n=499)

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]		t	Pr(T > t)
Sexo								
Hombre	331	.8308157	.0206384	.3754822	.7902163	.8714151	40.738	0.0001
Mujer	168	.672619	.0363122	.4706604	.6009288	.7443092		
Combined	499	.7775551	.0186364	.4163057	.7409394	.8141708		
Edad								
58-74	357	.8011204	.0211553	.3997174	.7595154	.8427255	-1.5028	0.1335
75-93	138	.7391304	.0375156	.4407086	.6649459	.813315		
Combined	495	.7838384	.0185199	.4120422	.7474509	.8202259		
Estado Civil								
Con pareja	334	.6787879	.036462	.4683635	.6067923	.7507835	-3.7741	0.0002
Sin pareja	165	.8263473	.0207587	.3793789	.7855126	.867182		
Combined	499	.7775551	.0186364	.4163057	.7409394	.8141708		
Educación								
Media	384	.7786458	.0212136	.4157	.7369361	.8203555	0.2331	0.8158
Superior	99	.7676768	.0426602	.4244632	.6830191	.8523345		
Combined	483	.7763975	.0189783	.4170906	.7391071	.8136879		
Conocimiento PBS								
Conoce	262	.8091603	.0243238	.3937151	.7612644	.8570562	-17.870	0.0745
No Conoce	237	.742616	.0284588	.438118	.6865503	.7986818		
Combined	499	.7775551	.0186364	.4163057	.7409394	.8141708		
Conocimiento APS de Vejez								
Conoce	91	.754902	.0213215	.4306734	.7129879	.796816	-25.885	0.0099
No Conoce	408	.8791209	.034362	.3277928	.8108547	.947387		
Combined	499	.7775551	.0186364	.4163057	.7409394	.8141708		
Maneja % de cobro de AFP por administrar fondos								
Conoce	20	.7	.1051315	.4701623	.4799572	.9200428	11.336	0.2577
No Conoce	353	.8045326	.0211367	.397123	.7629624	.8461027		
Combined	373	.7989276	.0207806	.4013404	.7580654	.8397898		

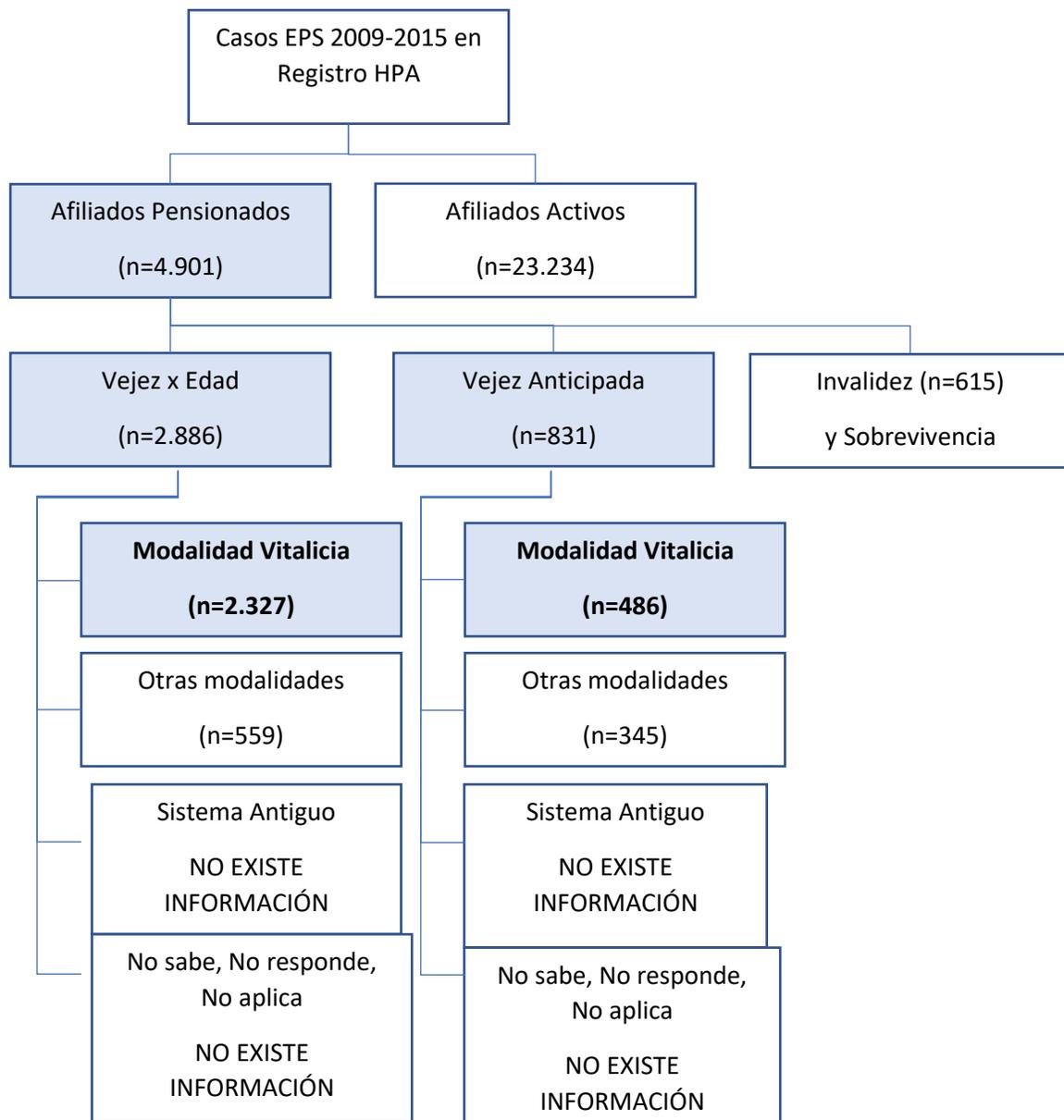
Fuente: Elaboración propia

Figura N°1 Encuesta de Protección Social 2015



Fuente: Elaboración propia

Figura N°2 Historial Previsional Administrativo (HPA) de Afiliados Activos, Pensionados y Fallecidos del Sistema de AFP – 2015



Fuente: Elaboración propia

