

Tipo de documento: Tesis de maestría

Master in Management + Analytics

Un modelo de segmentación de Financiadores de la Salud en Argentina: Propuesta de descuentos diferenciales para maximizar el acceso a los medicamentos de alto precio

Autoría: *Castro Aranda, Facundo*

Año académico: *2023*

¿Cómo citar este trabajo?

Castro Aranda, F. (2023) "Un modelo de segmentación de Financiadores de la Salud en Argentina: Propuesta de descuentos diferenciales para maximizar el acceso a los medicamentos de alto precio". [*Tesis de maestría. Universidad Torcuato Di Tella*]. Repositorio Digital Universidad Torcuato Di Tella

<https://repositorio.utdt.edu/handle/20.500.13098/12085>

El presente documento se encuentra alojado en el Repositorio Digital de la Universidad Torcuato Di Tella bajo una licencia Creative Commons Atribución-No Comercial-Compartir Igual 2.5 Argentina (CC BY-NC-SA 2.5 AR)
Dirección: <https://repositorio.utdt.edu>



**UNIVERSIDAD
TORCUATO DI TELLA**

MASTER IN MANAGEMENT + ANALYTICS

**UN MODELO DE SEGMENTACIÓN DE
FINANCIADORES DE LA SALUD EN ARGENTINA:
PROPUESTA DE DESCUENTOS DIFERENCIALES
PARA MAXIMIZAR EL ACCESO A LOS
MEDICAMENTOS DE ALTO PRECIO**

TESIS

Facundo Castro Aranda

Mayo 2023

Tutoras: Magdalena Cornejo y Victoria Sáenz

Resumen

El mercado de medicamentos de alto precio¹ está integrado por laboratorios a nivel mundial que llevan a cabo la investigación y desarrollo de tratamientos encaminados a mejorar la salud de los pacientes y ampliar la cobertura por parte de sus clientes, los financiadores. Su objetivo se da en un contexto de constante transformación y con competencia cada vez más intensa. En Argentina, además, hay gran fragmentación en la cobertura de la salud, con más de 441 organismos públicos y privados² que se ocupan de financiar la salud de la población con beneficiarios de características demográficas y epidemiológicas muy dispares entre sí. En este trabajo se busca reflejar una posible transformación en la toma de decisiones sobre el precio de comercialización de los mencionados productos farmacéuticos a los distintos financiadores de salud en Argentina, a partir de una propuesta de segmentación. Esta propuesta culmina en un modelo de diferenciación de precios, un mecanismo que busca resolver la presencia de inequidad en el acceso a los tratamientos innovadores. Para trabajar en el problema de negocio, se utiliza la metodología CRISP DM³. En ese marco, se describe en profundidad el negocio, las herramientas y datos disponibles para encarar el análisis, las distintas metodologías y métricas de performance. Por último, se propone un posible plan de implementación.

El trabajo indaga en las variables que hoy definen la cobertura o rechazo de los productos por los financiadores y el criterio en los descuentos para cada financiador. Luego, organiza a los distintos financiadores en clústeres, utilizando métodos de *machine learning*. La mejor performance es de 0.40. en promedio, en anchura de silueta, y se obtiene al seleccionar una estrategia mixta de segmentación para dos subtipos de patologías: *K-means clustering* para el portfolio de enfermedades de baja incidencia y *Agglomerative Hierarchical Clustering* para patologías oncológicas. Con esos segmentos, luego se presenta una propuesta de implementación de precios diferenciales, a través de un ejercicio de optimización lineal entera.

Si bien la estrategia sugerida por este proyecto no provee un modelo diferencial, basado en valor, para la determinación de los precios de los medicamentos para cada paciente, se pretende plantear un primer paso en el camino hacia esa situación. La generación de un mecanismo innovador, basado en datos e información del mundo real para definir la asignación de descuentos, podría implicar un acceso más temprano y universal a los productos innovadores y, a la vez, garantizar un nivel sustentable de retornos para las compañías farmacéuticas.

¹ Los medicamentos de alto precio, también conocidos como medicamentos de alto costo en la literatura (OPS 2009; Tobar, 2014; Marin y otros, 2022), son aquellos cuyo precio implica un alto esfuerzo económico asociado a su adquisición, para quien financia el tratamiento. Según la Organización Panamericana de la Salud (OPS, 2009), las características que definen a este tipo de medicamentos son: su costo de financiamiento, exclusividad, novedad y tipo de enfermedad para el que son utilizados.

En general pertenecen a esta categoría, aquellos medicamentos destinados al tratamiento de enfermedades catastróficas, un conjunto de patologías definidas de ese modo por su gravedad médica y elevado costo en medicamentos y uso de recursos sanitarios. Algunas patologías con medicamentos de alto precio son: para el cáncer (Pembrolizumab, Trastuzumab); atrofia muscular espinal (Nusinersen); la artritis reumatoide (Adalimumab); entre otras.

En particular en este trabajo se pretende incluir en el análisis, el detalle de prevalencia e incidencia de algunas patologías de relevancia para un laboratorio farmacéutico de alto precio. Se incorporan en el análisis a las siguientes patologías: HER2 (cáncer de mama con ese subtipo), Cáncer de Hígado, Hemofilia, Atrofia muscular espinal - AME-, cáncer de pulmón, esclerosis múltiple, linfoma de *Non Hodgkin*.

² Según el informe de Cobertura de Salud en Argentina (Ministerio de Salud Argentina, 2023)

³ CRISP DM: por las siglas en inglés de *Cross Industry Standard Process for Data Mining*, es una herramienta estándar para el desarrollo de proyectos relacionados a la ciencia de datos que se utiliza desde 1997 y fue concebido por 5 empresas: SPSS, Teradata, Daimier AG, NCR y Ohra.

Abstract

The high-priced medication⁴ market is comprised of global laboratories that conduct research and development of new treatments aimed at improving patient health and expanding coverage by healthcare payers. They pursue their objective in a context of constant transformation and increasingly intense competition. In Argentina, there are over 441 public and private entities⁵ responsible for financing the population's healthcare, with beneficiaries presenting diverse demographic and epidemiologic characteristics. This work aims to reflect a potential transformation in decision-making regarding the pricing of pharmaceutical products for different healthcare payers in Argentina, based on a segmentation proposal. This project concludes with a price differentiation model, as a mechanism to address inequities in access to innovative treatments. We use the CRISP DM Methodology⁶ to structure the business problem. To define it, we developed a deep understanding of the business, available analysis tools, various methodologies, and performance metrics. Finally, we propose a potential implementation plan.

We first analyze the variables that today define the coverage or rejection of the products by the financiers and the discount criteria for each payer. Then, we organize the various payers into clusters, using machine-learning methods. We obtain a performance of 0.40, using silhouette scoring. We obtain this result by selecting a mixed segmentation strategy for two subtypes of pathologies: K-means clustering for the portfolio of low-incidence diseases and Agglomerative Hierarchical Clustering for oncological pathologies. With these segments, we present a proposal for the implementation of differential prices, through an integer linear optimization exercise.

While the strategy suggested by this work is likely insufficient to guarantee a differential value-based model for determining medication prices for each healthcare payer, it aims to propose an initial step towards that situation. The generation of an innovative mechanism based on real-world data and information to define discount allocation could lead to earlier and universal access to innovative products while ensuring a sustainable level of returns for pharmaceutical companies.

⁴ High-priced medications, also known as high-cost drugs in the literature (PAHO 2009; Tobar, 2014; Marin et al., 2022), are those whose price implies a high economic effort associated with their acquisition for those who pay the treatment. According to the World Health Organization (WHO, 2009), the characteristics that define this type of medication are: cost, exclusivity, novelty, and type of disease for which they are used.

In general, drugs intended for the treatment of catastrophic diseases belong to this category, a group of pathologies defined in this way by their medical severity and their high cost of drugs and use of health resources.

Some pathologies with high-priced drugs are for cancer (Pembrolizumab, Trastuzumab); spinal muscular atrophy (Nusinersen); rheumatoid arthritis (Adalimumab).

We will include in the analysis the prevalence and incidence of some specific pathologies with high priced medicines: HER2 (breast cancer), Liver cancer, Hemophilia, Spinal Muscular Atrophy, lung cancer, multiple sclerosis and Non Hodgkin Lymphoma

⁵ According to: informe de Cobertura de Salud en Argentina (Ministerio de Salud Argentina, 2023)

⁶ CRISP DM: *Cross Industry Standard Process for Data Mining*, is a framework to structure data mining projects on the industry. It is available since 1997 was conceived by: SPSS, Teradata, Daimier AG, NCR y Ohra.

A Lucre, por su ayuda siempre, su compañía y amor infinito.

A mi familia, una fuente de inspiración tremenda.

TABLA DE CONTENIDOS

1. Introducción	9
I. Uso de estrategias analíticas en la gestión de la salud	10
II. Métodos de <i>clusterización</i> de <i>machine learning</i> para identificar segmentos de financiadores.....	12
III. Objetivo	12
2. Métodos y procedimientos	13
a. Comprensión del negocio	15
I. El sector de la salud como mercado imperfecto	15
II. Evolución en el mercado representativo: Las transformaciones en la salud originadas en la transición epidemiológica y demográfica de la población de Argentina	17
III. El sistema de Salud Argentino	19
IV. El proceso de incorporación de nuevas tecnologías sanitarias para la población	31
V. Definición de precios para los medicamentos innovadores y la compañía en ese contexto	35
b. Exploración y preparación de los datos	38
I. Descripción de variables iniciales	38
II. Exploración de los datos de la venta	41
III. Técnicas empleadas en el análisis de datos y selección de variables	44
IV. Análisis de variables disponibles y transformación de los datos	48
V. Modelos seleccionados para ser testeados en el análisis de <i>clustering</i>	65
c. Modelado: Selección de modelos a testear	71
I. Modelos de clustering	71
II. Métricas de Rendimiento	73
III. Modelos de determinación de precios diferenciales	74
3. Resultados.....	76
d. Evaluación de resultados	76
IV. Modelos de clustering.....	76
V. Análisis de resultados y selección del modelo definitivo	85
e. Plan de Implementación	85
VI. Modelos de determinación de precios diferenciales	85
VII. Recomendaciones de implementación del modelo	91
4. Conclusiones y próximos pasos.....	94
5. Referencias.....	98
Anexo 1: Construcción de datos de Financiadores	101
Anexo 2: Construcción de datos para modelos. Descripción de procesamiento de los datos	109
Anexo 3: Análisis de variables categóricas mediante técnica FAMD. Profundización en el análisis.....	113

Índice de Tablas

Tabla 1. Muestra de la base con información para los distintos financiadores de salud de Argentina	40
Tabla 2: Resultados de experimentación del modelo 1 con variables seleccionadas previamente. Selección del modelo de <i>k means</i> para generar k clústeres de financiadores	77
Tabla 3: Representación de valores promedios por clúster determinado	79
Tabla 4: Análisis de segmentación 1. Análisis de la Incidencia enfermedades de baja incidencia.....	80
Tabla 5: Representación de grupos de modelo para patologías de baja incidencia en k=3 clústeres	82
Tabla 6: Análisis de segmentación 2: Patologías más asociadas a grupos de mayor edad (oncología, principalmente).....	83
Tabla 7: Base inicial de población estimada para cada financiador por rango de edad y género....	109
Tabla 8: Variables construidas a partir de información base de la tabla 7 que fueron incluidas en el análisis del modelo	110
Tabla 9: Variables construidas a partir de información base de la tabla 7 combinadas con fuentes de incidencia y prevalencia extraídas de diversas fuentes de información (Globocan, Global Burden of Disease y a partir de opinión de expertos)	111
Tabla 10: Base Financiadores y estructura de costos y de gestión (por disponibilidad o no de centros propios de salud que brindan cobertura exclusivamente a los afiliados y que, en su mayoría cubren a otros financiadores de salud).....	112
Tabla 11: Variables descriptivas de los financiadores	113

Índice de Figuras

Figura 1: Representación gráfica del proceso de implementación de modelos de datos en técnica CRISP - DM	13
Figura 2: Representación del sistema de salud en sus tres subsectores en Argentina	19
Figura 3: Distribución de afiliados por género y edad para Obras Sociales Nacionales. Ejemplos	21
Figura 4: Distribución de afiliados por género y edad para Obras Sociales provinciales. Ejemplos...	23
Figura 5: Distribución de afiliados por género y edad para el PAMI (INSSJP)	25
Figura 6: Distribución de afiliados por género y edad para Obras sociales especiales	26
Figura 7: Distribución de afiliados por género y edad para Empresas de Medicina Prepaga	27
Figura 8: Distribución de beneficiarios por género y edad para la población con cobertura exclusiva del estado (según registros de empadronamiento en programa SUMAR)	29
Figura 9: Proceso de distribución de medicamentos. Actores relevantes.....	33
Figura 10: Cadena de distribución de medicamentos de alto precio	34
Figura 11: Esquema simplificado del proceso de cobertura de un medicamento de alto precio para un paciente	35
Figura 12: Representación gráfica de los objetivos que se buscan balancear en la determinación del precio de un medicamento de alto precio.....	36
Figura 13. Análisis de Precio pagado y de ingreso promedio por afiliado para el financiador.....	41

Figura 14. Análisis de Precio pagado y cantidad comprada por financiador directo	42
Figura 15: Matriz de correlaciones de información de precios y ventas de la compañía en análisis en la actualidad en para el tratamiento del cáncer de mama	44
Figura 16: Representación de la cantidad de afiliados por financiador y su transformación logarítmica 48	
Figura 17: Representación gráfica de transformación logarítmica en funciones que representan población total por rango etario.....	50
Figura 18: Análisis de la edad promedio en financiadores argentinos	51
Figura 19: Representación gráfica de la variable 'ingreso por afiliado' para los distintos financiadores y su comportamiento post aplicación de log normalización de la variable.....	52
Figura 20: Representación gráfica de la variable Aumento y su comportamiento post aplicación de <i>log normalización</i> de la variable	53
Figura 21: Representación gráfica de variable edad promedio e ingreso promedio por afiliado	53
Figura 22: Representación de la incidencia o prevalencia específica por afiliado para los distintos financiadores de Argentina y su versión en función logarítmica.....	54
Figura 23: Ingreso promedio por afiliado contra incidencia de cáncer de mama HER2.....	57
Figura 24: Ingresos económicos e incidencia de patologías estratégicas.....	58
Figura 25: Prevalencia específica de AME y Hemofilia para los distintos financiadores	59
Figura 26: Incidencia potencial de pacientes del portfolio de la compañía analizada en los distintos financiadores.....	60
Figura 27: Incidencia del portfolio de la compañía con ajuste de <i>log normalización</i> de la variable para corregir el sesgo de los datos y estandarizar la variable	60
Figura 28: Representación gráfica del ingreso promedio por afiliado y la incidencia potencial del portafolio de productos estratégicos de la compañía en análisis	61
Figura 29: Disponibilidad de Centros propios y Cantidad de centros propios por beneficiario	62
Figura 30: Representación de cantidad de centros propios corregido por la cantidad de millones de afiliados cubiertos por financiador y representación logarítmica para aquellos con disponibilidad de centros propios	63
Figura 31: Análisis gráfico de mapa de correlaciones de variables	64
Figura 32: Análisis de componentes principales de variables con alta correlación entre sí.....	66
Figura 33: Análisis de la contribución de las distintas variables con alta correlación en la dimensión 1, 2 y 3	67
Figura 34: Representación de matriz de correlaciones de primer modelo a ser testado para la segmentación de financiadores	68
Figura 35. <i>Dataframe</i> Portfolio. Representación de análisis de componentes principales.....	69
Figura 36. <i>Dataframe</i> baja incidencia: Representación del análisis de componentes principales.....	70
Figura 37. <i>Dataframe</i> Edad: Representación del análisis de componentes principales.....	70
Figura 38: Diferenciación de precios a nivel país con efectos positivos en el ingreso y la cobertura de la población.....	75

Figura 39: Representación gráfica de métricas de rendimiento analizadas en el primer modelo de clústeres.....	78
Figura 40: Representación gráfica de métricas de rendimiento analizadas en el modelo de análisis específico para enfermedades de baja incidencia.....	81
Figura 41: Representación gráfica de métricas de performance en análisis modelo portfolio oncología y medicamentos para mayor prevalencia por edad	84
Figura 42: Dendograma de la distribución de los datos a partir de la definición de la estrategia de segmentación para patologías más asociadas a la edad	85
Figura 43: Gráfico extraído de Yadav (2010) donde se muestra la posibilidad de tener canales individualizables en la implementación de una estrategia de precios diferenciales para distintos niveles socioeconómicos.....	92
Figura 44: Arquetipos de información para la implementación de una estrategia diferencial de precios para productos farmacéuticos.....	93
Figura 45: Cuadro de asignaciones en concepto de matriz de riesgo SANO para obras sociales nacionales	104
Figura 46: Representación de modelo de base de cálculo de datos de financiamiento de Obras Sociales Nacionales	104
Figura 47: Representación de las dimensiones 1 y 2 de los componentes principales en forma vectorial y de las variables categóricas.....	114
Figura 48: Contribución en la variabilidad de los datos de los distintos factores identificados por el modelo FAMD	115
Figura 49: Contribución en la variabilidad de los datos de los distintos factores identificados por el modelo FAMD	116
Figura 50: Representación de los aportes en la variabilidad determinada por el cálculo del análisis FAMD para los datos.....	116
Figura 51: Modelo de tipo FAMD con variables logarítmicas - representación vectorial de los datos, contribución de las dimensiones determinadas por el modelo y la descripción de las variables que contribuyen en mayor y menor medida en cada una de esas dimensiones determinadas	117

1. Introducción

En este trabajo, se pretende analizar las decisiones de financiamiento dentro del consumo de medicamentos denominados de “alto precio” en salud⁷. En particular, se busca redefinir la estrategia de determinación de los precios a los que una compañía farmacéutica podría comercializar sus medicamentos a los distintos financiadores de salud, en el contexto del sistema de salud de Argentina.

Para encaminarse a modelar dicha estrategia, se pretende entender qué define la capacidad y voluntad de pago de aquellas entidades públicas o privadas que definen la compra de medicamentos, (llamados en este trabajo y en la literatura, financiadores de la salud). A partir de eso, se busca proponer un modelo que busque maximizar el acceso a los tratamientos de alto precio.

En otras palabras, se busca identificar un mecanismo racional que permita definir la asignación de los recursos de la compañía, entendidos como los descuentos específicos que asigna a los distintos segmentos de financiadores, garantizando el máximo acceso a los tratamientos posible y también la equidad en el acceso para todos los pacientes que integran el sistema de salud. A través de esos mencionados mecanismos de descuentos o reembolsos, se pretende generar un esquema de precios diferenciales.

En particular, se pretende trabajar con algunas patologías seleccionadas que tienen medicamentos de alto precio en el mercado farmacéutico. Se han elegido dichas enfermedades ya que son de relevancia para una compañía farmacéutica para la que se realiza el análisis (con la que se ha firmado un acuerdo de confidencialidad). Las patologías incorporadas son: cáncer de mama con marcador tumoral HER2, cáncer de Hígado, Hemofilia A, Atrofia muscular espinal, Cáncer de pulmón, Esclerosis Múltiple y Linfoma de *Non Hodgkin*.

Se argumenta que la generación de un mecanismo innovador de asignación de descuentos podría implicar un acceso temprano y universal a los productos farmacéuticos innovadores y, a la vez, garantizar un nivel sustentable de retornos para la compañía farmacéutica. En conclusión, el objetivo de este trabajo podría implicar un beneficio muy significativo tanto para la sociedad como para la compañía en análisis en cuanto al tiempo al acceso y a la incorporación sustentable de tecnologías innovadoras en el sistema de salud.

⁷ Los medicamentos de alto precio, también conocidos como medicamentos de alto costo en la literatura (OPS 2009; Tobar, 2014; Marin y otros, 2022), son aquellos cuyo precio implica un alto esfuerzo económico asociado a su adquisición para quien financia el tratamiento. Según la Organización Panamericana de la Salud (OPS, 2009), las características que definen a este tipo de medicamentos son: su costo de financiamiento, exclusividad, novedad y tipo de enfermedad para el que son utilizados.

En general pertenecen a esta categoría, aquellos medicamentos destinados al tratamiento de enfermedades catastróficas, un conjunto de patologías definidas de ese modo por su gravedad médica y su elevado costo en medicamentos y uso de recursos sanitarios.

Algunas patologías con medicamentos de alto precio son para el cáncer (Pembrolizumab, Trastuzumab); atrofia muscular espinal (Nusinersen); la artritis reumatoide (Adalimumab); o la esclerosis múltiple (Ofatumumab, Ocrelizumab).

En particular en este trabajo se pretende incluir en el análisis, el detalle de prevalencia e incidencia de algunas patologías de relevancia que disponen de tratamientos de alto precio con el fin de tratar a su población. Se incorporan en el análisis a las siguientes patologías: HER2 (cáncer de mama con ese subtipo), Cáncer de Hígado (en particular para una primera línea en cierto escenario), Hemofilia, Atrofia muscular espinal - AME-, cáncer de pulmón, esclerosis múltiple, linfoma de Non Hodgkin

Como punto de partida para enmarcar el problema de investigación, se aborda al negocio de los medicamentos de alto precio desde lo más general hacia lo particular. Es decir, se define al negocio de los productores de tecnologías sanitarias dentro de la industria de la salud en el contexto de la prestación de servicios sanitarios. Luego, se presentan las transformaciones más relevantes dentro de la salud poblacional, el abordaje de la salud y la producción de tecnologías sanitarias. A continuación, en un segundo nivel de análisis, se presenta el sistema de salud argentino, con sus respectivos actores y se profundiza, en particular, en el proceso de incorporación de tecnologías sanitarias innovadoras para los pacientes pasibles de esos tratamientos. En un tercer nivel de análisis, se procede a presentar la estrategia corporativa que se busca resolver y luego se introducen las herramientas que luego son usadas en el modelo que se propone implementar.

El trabajo luego se focaliza en explorar la posible implementación de una estrategia de segmentación de financiadores para lo que se analizan distintas experiencias de utilización de herramientas de *clustering* en la salud, se modelan y comprenden los datos disponibles, se testean los modelos y se analiza el rendimiento de las técnicas analizadas. Por último, se propone la implementación de un posible modelo de optimización lineal entera simple, en el contexto del negocio analizado y de las tareas analíticas de la gestión de ingresos y segmentación de clientes, donde se refleje el objetivo de maximización del acceso a los tratamientos analizados. Todos estos pasos se ejecutan siguiendo una metodología de trabajo de la minería de datos – *data mining*– para garantizar un correcto y completo abordaje del negocio bajo análisis.

I. Uso de estrategias analíticas en la gestión de la salud

En el sector de la salud se han utilizado distintos mecanismos que permitan hacer una asignación más eficiente de recursos⁸. En el ámbito de la investigación, se ha buscado principalmente identificar cuáles son las características que se asocian en mayor medida a la probabilidad de suceso de patologías – factores de riesgo- cuyo costo implicado sea alto. Hay diversos trabajos donde se utilizan las características poblacionales para definir una asignación eficiente de recursos sanitarios.

Los financiadores privados de salud son los que históricamente han utilizado, en el valor de la prima que cobran, mediciones del riesgo que un cierto individuo, por sus características, tiene asociado. En tal sentido, como se describe en un trabajo de análisis de medición del riesgo en la cobertura de salud, el ajuste de riesgo de severidad de una población ha sido estudiado por muchos años como un mecanismo para proveer resultados que involucran información de cuidado de la salud (Hileman et al., 2016). En varios trabajos, se ha buscado identificar las características que definen el riesgo de poblaciones específicas de tener requerimientos específicos. Los métodos han sido utilizados para definir la severidad, la población pertinente, el rol del diagnóstico, el rol de las cirugías mayores, los requerimientos de información y la forma en que se hacen las mediciones, el tiempo de medición y los criterios de clasificación.

En la propuesta realizada para elaborar mediciones de riesgo por los mencionados autores (Hileman et al., 2016), se define que la demografía es un insumo clave para la mayoría de los modelos de *scoring* de riesgo. En este sentido, la edad y el sexo son claves en modelos prospectivos donde se estén evaluando las condiciones del diagnóstico para el nivel de riesgo. Estas son herramientas útiles

⁸ La asignación de precios, a través de descuentos, no necesariamente podría ser vista como un mecanismo de asignación eficiente de recursos. Sin embargo, en este trabajo se utilizó dicha conceptualización para la elaboración de la propuesta que se realiza ya que se alinea en cuanto a la forma en que se da uso a la información relevante de los pacientes potenciales.

para diferenciar el riesgo contribuido por ciertas condiciones. En ese sentido, los autores señalan, *“suponga que las variables independientes en un modelo lineal incluyen características de una enfermedad, como edad o sexo. Una condición específica podría tener un impacto para un rango etario por sobre otro”*. (Hileman et al., 2016)

Asimismo, se han desarrollado también modelos que permiten identificar poblaciones con determinadas características, tal como comorbilidades o grupos etarios, para prevenir enfermedades crónicas a partir del uso de un perfil digital de salud y un seguimiento personalizado según sus características. Un ejemplo de este enfoque está presentado en el trabajo de Nnoaham y Cann (2020), donde a partir de datos de cuidado crean una base con información sociodemográfica, morbilidad, utilización de cuidados, costos y factores de riesgo. Luego, utilizan un método de clúster para agrupar la población en segmentos con distintos patrones de utilización y consumo de salud basado en la mejor estrategia para cada grupo determinado (Nnoaham & Cann, 2020).

En el trabajo de Hadian & Fattahi (2019), se busca aplicar técnicas de clusterización para identificar a grupos de hogares con necesidades de financiamiento similares en Irán. Se utilizan los datos de la encuesta de Ingresos y Gastos de Hogares y se aplican técnicas de *k means* para identificar grupos con diferentes necesidades de financiamiento en Salud (Hadian & Fattahi, 2019). En ese caso, el trabajo busca establecer distintas estrategias de financiamiento, identificando tres grupos principales según la mayor o menor necesidad de recursos en salud. Se trabaja en el precio para cada paciente individual.

Sin embargo, la utilización de datos demográficos y de ingresos económicos para la determinación de precios eficientes para medicamentos de alto precio donde el financiamiento no es realizado por las personas en sí, sino por diversos financiadores de salud, no ha sido abordado en profundidad en la literatura relevada. Asimismo, no se ha encontrado información sobre la utilización de estos datos para el establecimiento de precios diferenciales para medicamentos de alto precio. En los distintos trabajos que se consultaron que proponían el establecimiento de precios diferenciales entre países, se utilizan medidas de pobreza y de inversión en salud, pero no se incorpora el riesgo asociado a tener incidencia de algún producto en particular por encima que el resto en función de sus características demográficas.

Asimismo, si bien se han identificado diversas propuestas de modelos teóricos que establecen esquemas de fijación de precios de medicamentos diferenciales entre países, no existe gran profundidad en la posible implementación de estas herramientas al interior de países con sistemas muy fragmentados de salud y con poblaciones que son muy diversas en cuanto a sus características demográficas y de tamaño, como sucede en el caso de Argentina. En la revisión de literatura, se han identificado dos herramientas estratégicas que son utilizadas con el fin de garantizar mayor capacidad de acceso por parte de los financiadores. Por un lado, a través de la fijación del precio al que se considere que un financiador puede incorporar de manera sustentable una cierta tecnología sanitaria. Por el otro, a través de la implementación de mecanismos innovadores de pago tales como acuerdos de riesgo compartido, suscripciones, modelos capitados, entre otros, que requieren cierta estimación de los riesgos involucrados para garantizar la sustentabilidad del sistema de salud.

Las dos estrategias podrían enriquecerse de realizar un análisis en profundidad de los datos poblacionales orientados a identificar el riesgo demográfico y de tamaño asociado a los financiadores. Con ello, se vuelve posible el diseño de acuerdos que garanticen la reducción de los riesgos de incurrir en gastos que no resulten sustentables para cada financiador.

II. Métodos de *clusterización* de *machine learning* para identificar segmentos de financiadores

Con el fin de identificar grupos de financiadores con características similares entre sí, en este trabajo se pretenden utilizar métodos de *clusterización*. Éstos son un conjunto de técnicas que se utilizan para encontrar agrupaciones de observaciones en un conjunto de datos. El objetivo de clasificar en clústeres las observaciones de un conjunto de datos es dividirlos en distintos grupos buscando que las observaciones dentro de cada grupo sean similares entre sí y que las observaciones en diferentes grupos sean diferentes entre sí.

Existen muchos algoritmos que difieren significativamente en la forma en que comprenden la constitución de cada grupo y en la eficiencia para encontrarlos, permiten ejecutar tareas para obtener clústeres. En ese contexto, el análisis de clústeres no es una tarea automática, sino que funciona como un proceso iterativo que implica prueba y fracaso. En consecuencia, es frecuente que sea necesario modificar la forma en que se hace el procesamiento de datos y los parámetros del modelo hasta que el resultado alcance las propiedades deseadas.

Muchos problemas de negocios se resuelven a través de estas técnicas de *machine learning* ya que permite identificar las particularidades que caracterizan a cada grupo. Sin embargo, no se ha encontrado bibliografía que reporte la ejecución de este mecanismo de aprendizaje en particular para precios de medicamentos de alto precio para financiadores de salud diversos dentro de un mismo país. En el caso particular de este trabajo se busca incorporar variables económicas, demográficas, de riesgo de tener ciertas patologías con vínculo a los productos ofrecidos por la compañía farmacéutica, sobre la disponibilidad de centros propios de salud, etcétera.

III. Objetivo

En función a la problemática planteada y el breve relevamiento que se ha realizado sobre el tema en la literatura especializada, el objetivo de este trabajo es proponer una redefinición de la estructura actual con que se definen los precios a los que cada financiador accede a los tratamientos. En línea con la perspectiva corporativa y los desafíos de la sustentabilidad en el sistema de salud y la innovación, se busca diseñar un modelo encaminado a la determinación de precios diferenciales basados en valor. Con esto, en función a los argumentos mencionados, podría permitirse un acceso más temprano a las nuevas tecnologías sanitarias.

Para la consecución del mencionado objetivo, se construyen estimaciones de población cubierta, financiamiento y estructura de los financiadores. Además, se realiza un análisis de las ventas actuales de la compañía y se identifican las características que reúnen los financiadores que efectivamente autorizan el uso de los productos. Luego, se identifican características que diferencian a los financiadores que integran el sistema de salud y, por último, se propone un modelo de precios diferenciales que promueva criterios de equidad y facilite un acceso más masivo a los tratamientos innovadores.

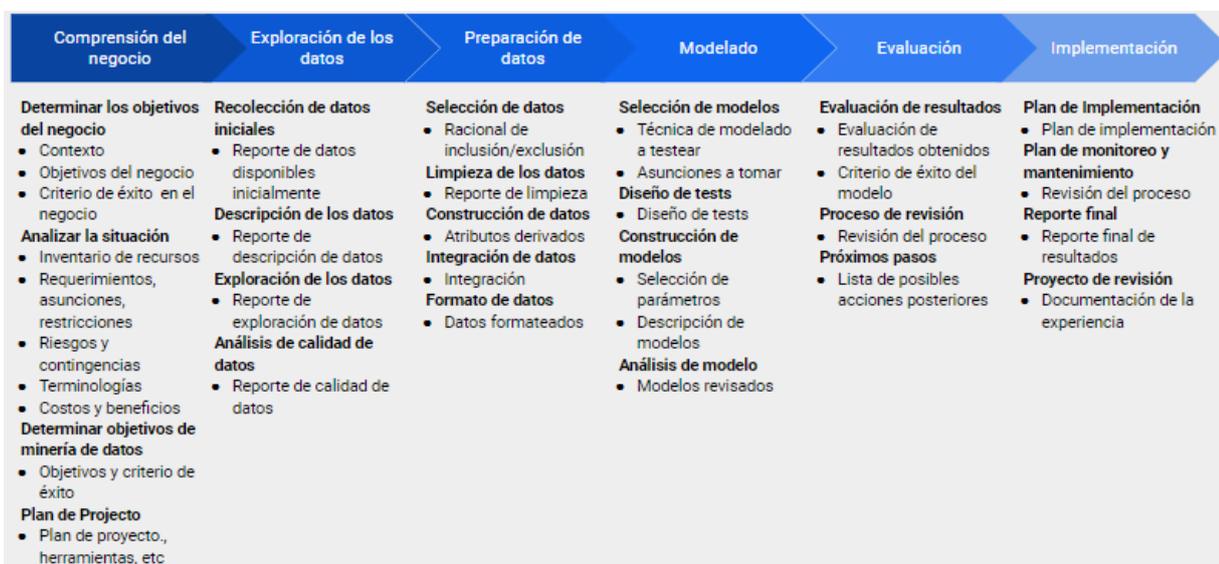
En resumen, el objetivo se alcanza siguiendo un proceso de tres etapas que permiten entender la situación actual, organizar a los financiadores que compran o potencialmente podrían comprar productos de la compañía y, por último, se propone un modelo de precios diferenciales a aplicar para dichos financiadores. Como resultados, se obtiene el modelo de datos disponible para estimar la población potencial, se organiza a los financiadores en clústeres o segmentos, y se diagrama una propuesta de modelo de optimización lineal para determinar el diferencial de precio. La forma de

implementación de esa estrategia, se plantea, es a través de descuentos comerciales específicos para cada financiador. Sin embargo, la implementación de ese descuento no forma parte de este trabajo, pero se incluirán recomendaciones sobre su posible implementación en las etapas finales.

2. Métodos y procedimientos

En este trabajo se pretende resolver un problema de negocio. A tal fin, se propone utilizar como marco general metodológico, al modelo conocido como CRISP-DM (por sus siglas en inglés de *Cross Industry Standard Process for Data Mining*). Esta es una herramienta estándar para el desarrollo de proyectos relacionados a la ciencia de datos que se utiliza desde 1997 y que ayuda a ordenar las distintas etapas requeridas para implementar alguna transformación en una industria, a partir de la utilización de datos. Se incorpora una versión esquemática de las etapas que propone este modelo, en la figura 1.

Figura 1: Representación gráfica del proceso de implementación de modelos de datos en técnica CRISP - DM



Fuente: Traducción del autor del trabajo de SPSS (2000,15) de tareas genéricas y etapas del proceso CRISP DM⁹

En esa línea, se enumeran las 6 siguientes etapas que se están cubiertas a lo largo de este trabajo y que dan fundamento a la estrategia que se propone:

1. **Comprensión del negocio:** En esta etapa se busca comprender los objetivos del negocio y se definen los requerimientos del proyecto. Se identifican los problemas a resolver y se definen criterios de éxito.

Esta etapa se aborda completamente en la siguiente sección de este trabajo. Se busca definir completamente el contexto del negocio y de la industria para la que se busca implementar el proyecto, las particularidades del sistema de salud argentino, la evolución reciente del mercado representativo, el proceso de toma de decisión por parte de los financiadores en la definición de cobertura de los medicamentos.

⁹ Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R. (2000). CRISP-DM 1.0: Step-by-Step Data Mining Guide. Copenhagen: SPSS.

2. **Exploración de los datos:** en esta etapa se recopilan y exploran todos los datos disponibles para el proyecto. Se hace un análisis exploratorio de los datos con el fin de comprender su calidad, estructura y contenido. Asimismo, se identifican posibles problemas y se seleccionan los datos relevantes para el proyecto.
3. **Preparación de los datos:** Se realiza, en esta etapa, la limpieza, transformación y enriquecimiento de los datos. Se definen acciones para datos faltantes, se eliminan los datos irrelevantes o que hacen ruido y se realiza la transformación e integración de datos en el formato ideal para el análisis.

La ejecución de las secciones 2 y 3, se ha integrado en un capítulo de este trabajo y se complementa la ejecución de sus tareas en el Anexo 1, 2 y 3.

Para ejecutar las tareas propuestas en esta etapa, en primer lugar, se busca entender cuáles son las variables que hoy definen la cobertura o rechazo de los productos por los financiadores de Argentina, así como también el criterio del otorgamiento de descuentos para los distintos clientes. Esto es posible realizarlo a través de un análisis retrospectivo de las ventas de la compañía a aquellos financiadores que compran directo a la compañía con algunas visualizaciones, que permiten identificar las variables involucradas en la definición de cobertura de los tratamientos.

Se organiza este capítulo a partir de la descripción inicial de las variables y la exploración de los datos de la venta actual de la compañía farmacéutica. Luego, se justifican y ejecutan las tareas de ingeniería de datos y selección de variables y se concluye con los modelos a ser testeados.

4. **Modelado:** en esta etapa, CRISP DM propone seleccionar y aplicar técnicas para adaptar y utilizar los datos con el fin de construir los modelos que se construyen. En este sentido, se ajustan y evalúan los distintos modelos y se selecciona el mejor para ser implementado.

Para esta sección del trabajo, se definen los modelos y las métricas que se utilizan para ejecutar la segmentación propuesta y evaluar su *performance*. Luego, en el capítulo de evaluación de resultados, se presentan los resultados de los distintos modelos. Para garantizar el correcto modelado, se propone la organización de los financiadores en clústeres, utilizando distintos métodos de *machine learning* que permiten asignarlos a grupos en función de sus características de disparidad territorial, género, grupos etarios, y por capacidad e ingresos económicos.

5. **Evaluación:** se realiza la evaluación detallada del modelo construido, se validan los resultados y se analiza la precisión y robustez. En esta etapa se comparan los resultados con los objetivos previstos para el proyecto y se toman decisiones sobre su viabilidad. Con el fin de ejecutar las tareas previstas para esta etapa, se abordan los resultados obtenidos y se selecciona el modelo en la sección planteada. Luego, en las conclusiones se enumeran distintas cuestiones relevantes en torno a la robustez y precisión de los datos y modelos utilizados. Esta etapa está detallada en el capítulo de resultados de este trabajo.
6. **Implementación:** Se implementa el modelo en el entorno operativo del negocio y se realiza el seguimiento del rendimiento, se documenta y comunica el resultado del proyecto y se planifican las acciones a seguir para su mantenimiento y mejora.

Esta última etapa es abordada a través de una posible implementación de la segmentación propuesta y los mecanismos mediante los cuales se busca llegar a tal fin. Se propone un

modelo de asignación de descuentos comerciales (con el fin de que con dichos descuentos se establezca un mecanismo de determinación de precios diferenciales) para éstos financiadores, que permita un mayor acceso a los tratamientos innovadores para los distintos segmentos de financiadores. Para tal fin, se propone un modelo de optimización lineal donde se analiza la sustentabilidad en el precio neto promedio al que comercializa la compañía farmacéutica y, a su vez, se satisfacen las variables relevantes. Sin embargo, en este trabajo no se aborda la implementación del modelo sino más bien se hace una breve recomendación sobre su posible implementación. Esta etapa también es abordada en el capítulo de Resultados de este trabajo.

Con el objetivo de garantizar un correcto abordaje del problema de negocio que se quiere resolver, se estructuran los próximos capítulos del trabajo, utilizando las etapas en el marco presentado por el mencionado modelo CRISP DM. Si bien se ha optado por no hacer una utilización exhaustiva de los distintos puntos mencionados por dicha metodología, se pretende organizar el abordaje del problema de negocio mediante esa estructura con el objetivo de Identificar posibles próximos pasos y tareas pendientes en la implementación del modelo.

a. Comprensión del negocio

I. El sector de la salud como mercado imperfecto

El sistema de salud está integrado por todas las organizaciones, instituciones y recursos cuyo principal objetivo es llevar a cabo actividades para mejorar la salud. Dentro de este sistema, se identifican cuatro funciones principales que son: la provisión de servicios (internación, atención de pacientes, prestación de servicios de salud en centros, etc.), la generación de recursos (nuevas tecnologías sanitarias que permiten mejorar la calidad de vida de los pacientes), la financiación y la gestión (cobertura de los servicios, seguros de salud, etc.).

En la ejecución de esas funciones, a diferencia de los mercados competitivos, el mercado de la salud tiene varias fallas de mercado en cuanto a los mecanismos de asignación de recursos y la provisión de sus servicios. Si bien existe una demanda (los pacientes) y una oferta (profesionales de la salud, farmacéuticos, equipos de investigación, financiadores, entre otros), los vínculos y flujos de información hacen que su comportamiento sea particular.

En el trabajo de Mwachofi et al. (2011), se definen los componentes que identifican a un mercado competitivo para explicar la imperfección en el mercado de la salud. En primer lugar, detallan que, en condiciones ideales, un mercado perfecto implica que la producción de los bienes y servicios ofrecidos sea eficiente (es decir que el costo por unidad sea el menor posible) y que el consumo, por otro lado, sea eficiente (los consumidores obtienen el mejor valor por su dinero para maximizar su utilidad dada su restricción presupuestaria). En contraposición, cuando se analiza el mercado de la salud, se observan distintas fallas de mercado. Entre otras se describe la asimetría en la información donde, el médico conoce sobre la enfermedad de sus pacientes y es el mismo que está en posición de definir la demanda de su servicio. Por otro lado, se describen en el mismo trabajo (Mwachofi et al., 2011), dos fallas de mercado asociadas a la relación entre los pacientes contratando servicios de seguro de salud: selección adversa y riesgo moral. La primera implica que existen incentivos para que los pacientes que tengan peor salud sean más proclives a contratar una cobertura de salud que aquellos que están sanos. Primas de salud más altas implican que aquellos con menor riesgo de enfermedad prefieran no pagar esos servicios. Por otro lado, aquellos individuos que tienen cobertura de salud tienen incentivos a sobre utilizar los recursos por tener cobertura. Adicionalmente

a estas fallas de mercado, se presenta como desafío adicional según el trabajo de Roberts (1998), el hecho de que el valor del tratamiento para el paciente sea contingente a diversos aspectos del cuidado como los tests y el tratamiento, la adherencia, entre otros (Roberts, 1998). Esto se traduce, evidentemente, a que no exista un 'precio de mercado' como tal ya que, en general, no existe un mecanismo de retroalimentación que muestre el valor de los recursos utilizados.

En particular, en el caso de los medicamentos de alto precio, se presenta la falla de mercado de principal – agente, como consecuencia de que si bien las perspectivas de los consumidores, productores y de la sociedad convergen en un mercado competitivo, el valor incurrido por los pacientes (los consumidores), en comparación con los costos incurridos por aquellos que financian la salud difieren en gran medida. En este sentido, en otro trabajo se explica que, si bien el dolor y el sufrimiento son una preocupación del paciente, podría pasar que no sean relevantes desde un análisis puramente de costos para aquel que financia la provisión de salud o de quien provee los servicios (Haddix et al., 1996, p. 12-26).

En definitiva, si bien el impacto en las decisiones de la salud es directo para el paciente, la asignación de un tratamiento específico depende de múltiples integrantes del sistema de salud: el médico que prescribe, las instituciones que proveen los servicios de salud, el financiador de los tratamientos, los productores de tecnologías sanitarias; también se suma la cadena de comercialización y distribución de medicamentos e insumos, entre otros.

Dentro del sistema de salud, el sector farmacéutico constituye una parte importante por su colaboración en la búsqueda de mejores resultados sanitarios, como así también, por su cuota de participación en la estructura de costos de la sociedad. Es una industria, como ha sido mencionado previamente, caracterizada por presentar diversas fallas de mercado. La asimetría en la información entre el principal (el paciente) y el agente (el médico prescriptor) es muy evidente en el proceso de prescripción y cobertura. Adicionalmente, suele estar muy regulada a nivel estatal¹⁰. Las regulaciones gubernamentales se preocupan por la seguridad y eficacia del producto, los estudios clínicos que los avalan, la sustitución genérica, la cobertura nacional o no, entre otras. La rentabilidad de las compañías farmacéuticas depende de la innovación. Los laboratorios que inventan y desarrollan nuevas alternativas de tratamiento utilizan el mecanismo de patentes como forma de protección del mercado que les permita obtener un retorno de la inversión en investigación y desarrollo.

Como resultado de estas características, la industria farmacéutica se asocia a una estructura de precios, en general, con menor elasticidad precio en comparación con otros bienes que no son tan esenciales. Se ha descrito que no existe 'soberanía del consumidor' de los pacientes para decidir si compran o rechazan un producto médico (Kumra, 2006). Adicionalmente a estas circunstancias, como ya ha sido indicado, para la mayoría de los países, la decisión sobre la cobertura en el uso del producto es transferida gradualmente del médico al financiador. Como consecuencia, el proceso sucede de la siguiente manera: el médico prescribe, el paciente consume y el financiador (el estado, las obras sociales, o empresas de medicina prepagas) paga el tratamiento. Esta relación implica que los incentivos de cada miembro involucrado sean desiguales y que, en consecuencia, se generen limitaciones en la voluntad de pago que conlleva al sacrificio de obtener un balance entre los incentivos de la industria y el beneficio de la sociedad. En resumen, como ha sido indicado, en

¹⁰ La existencia de las detalladas fallas de mercado, asociadas a la comercialización de medicamentos como bienes meritorios, y el hecho de que la atención de la salud sea considerada por múltiples tratados y convenciones internacionales como un derecho básico de los ciudadanos (por solo mencionar un ejemplo, el Pacto Internacional para los Derechos Sociales y Culturales reconoce en su artículo 12 el "derecho de toda persona al disfrute del más alto nivel posible de salud física y mental") hacen que el estado intervenga dentro de sus funciones económico sociales como regulador, proveedor y financiador de los medicamentos.

términos económicos se constituyen fallas de mercado por la existencia de información asimétrica en el consumo del tratamiento y el vínculo entre el principal y el agente.

Con el objetivo de definir en particular al mercado relevante, dentro de la industria farmacéutica, se pueden caracterizar a los medicamentos según el precio para su cobertura, por ser de “*alto precio*” - para tratamientos especiales o crónicos- y “*ambulatorios*”, cuyo precio es significativamente menor y en general una parte del pago está a cargo del consumidor final. El mercado de medicamentos de alto precio está integrado por pocos laboratorios a nivel mundial que llevan a cabo la investigación y desarrollo de nuevos tratamientos. Su desafío es mejorar los tratamientos disponibles para los pacientes y ampliar la cobertura por parte de los financiadores que integran el sistema de salud. En esta línea, como recomienda el grupo de trabajo en Acceso a Medicinas Esenciales de la ONU, se deben identificar nuevos mecanismos de interacción entre los distintos actores involucrados que permitan que las nuevas medicaciones estén disponibles (Paluzzi et al., 2005).

II. Evolución en el mercado representativo: Las transformaciones en la salud originadas en la transición epidemiológica y demográfica de la población de Argentina

Más allá de las particularidades naturales que definen a la industria farmacéutica, múltiples tendencias están transformando su comportamiento. El sistema de salud se encuentra en constante transformación como consecuencia de que la sociedad evoluciona en tres sentidos fundamentales que están transformando la dinámica en el uso de los servicios de salud y los costos implicados.

Por un lado, la expectativa de vida está aumentando y hay mayor población en el segmento de más de 65 años. Esta variable está afectando la dinámica en el volumen y estructura del gasto en la salud. En el caso de Argentina, si se analiza la proporción de población de más de 60 años en el censo 2010 era de 14,1% y de 16,2% en la proyección poblacional del 2023. En consecuencia, la proporción de adultos mayores se esperaba que creciera en dos puntos porcentuales entre esos años. Para este grupo poblacional en particular, se identifica que las causas más frecuentes de mortalidad son las enfermedades del sistema circulatorio -32,6%-, las enfermedades del sistema respiratorio -19,3%- y los tumores malignos -17,7%- (Ministerio de Salud Argentina, 2022).

Adicionalmente, existe aumento significativo de la población que implica que el consumo de insumos y servicios de salud se vea incrementado ya que hay mayor población incidente para las patologías. Si se comparan los datos poblacionales del Censo 2010¹¹, en comparación con los del 2022, para Argentina, se muestra un crecimiento poblacional de 14.7 puntos porcentuales (40.117.096 a 46.044.703).

Por último, la prevalencia de las enfermedades crónicas está en crecimiento como consecuencia de la existencia de nuevas tecnologías sanitarias que permiten prolongar la vida de los pacientes y mejorar su calidad de vida. Esto también implica un incremento significativo en los costos de los financiadores de salud. La pandemia, por otro lado, generó un aumento en diagnósticos tardíos, aumentos en la prevalencia de patologías crónicas asociadas al sedentarismo, consumo de medicamentos estacionales, vacunas, entre otros (IQVIA, 2022). En particular para los medicamentos de alto precio, la pandemia implicó cambios en el estadio de diagnóstico de la población y, por lo tanto, en la posibilidad de obtener un mayor beneficio de los tratamientos.

Además, se suman a estas circunstancias, la transformación de las formas y dinámicas de vida, el medio ambiente y los riesgos de factores genéticos. Además de las cuestiones que han transformado a la demanda de medicamentos, cuando se analiza en particular la oferta en el mercado

¹¹ Datos provisorios del Censo 2022, consultados en julio 2023 en la página web del INDEC (Indec.gob.ar)

farmacéutico, tal como se explica en el informe anual de Tendencias Globales en Salud (IQVIA, 2022), ha habido un incremento significativo en el desarrollo de nuevas sustancias activas. Según el informe, han sido comercializadas 84 nuevas soluciones donde al menos un componente, entidad biológica o combinación, es innovadora. Ese valor representa el doble de los lanzamientos del año 2016, tan solo cinco años atrás. En la misma publicación (IQVIA, 2022), se indica que, con el total de lanzamientos del 2021, se alcanzó un número de lanzamientos récord de 883 en los últimos 20 años. En esta misma línea, se describe que, actualmente, se encuentran en desarrollo un total de nuevos productos que excede 6000 estudios clínicos en humanos, más de 67 puntos porcentuales adicionales al nivel de productos en desarrollo del año 2016.

Si se observa en profundidad el tipo de desarrollos que se están llevando a cabo en la industria farmacéutica, se presenta en el informe que, en los tratamientos oncológicos, la medicina de precisión domina significativamente el desarrollo de nuevos productos. Las terapias para poblaciones target de baja incidencia, con características muy específicas, justifican más del 40% del desarrollo del *pipeline* en oncología con tratamientos de próxima generación donde se incluyen principalmente las terapias génicas. Existen más de 800 terapias biogénicas de próxima generación en investigación y desarrollo con incrementos significativos de la actividad CAR-T y de terapias celulares NK, así como también las terapias de edición genética y de RNA (IQVIA, 2022). Estos tratamientos están asociados a tecnologías de muy costosa investigación para tratar a poblaciones de pacientes cada vez más pequeñas y específicas (que incluso, en múltiples casos donde no había tratamiento previo, resultan hasta difíciles de identificar).

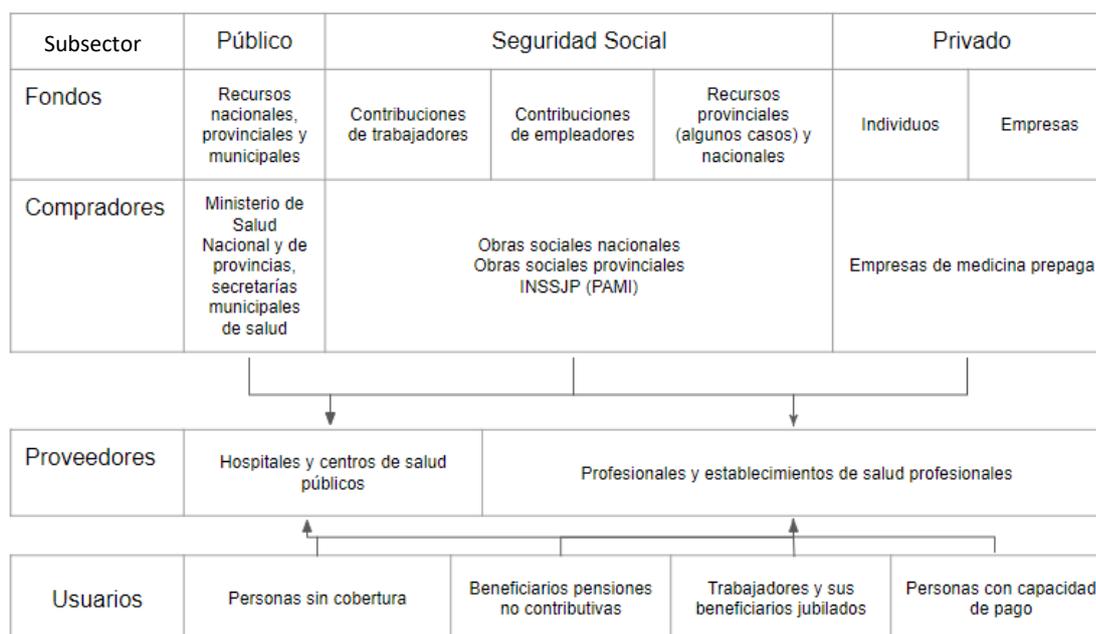
La utilización de nuevas herramientas para desarrollar e investigar la efectividad de los tratamientos que se aprueban para su comercialización también está en proceso de constante transformación. Como consecuencia de la baja incidencia de las enfermedades para las que surgen los medicamentos nuevos, la dificultad de identificar a los pacientes con estas enfermedades de baja incidencia en los ensayos clínicos y también la falta de comparadores establecidos en algunos casos de productos nuevos, se están implementando nuevas estrategias para acelerar el acceso a estas tecnologías para poblaciones huérfanas de tratamientos, mediante testeos innovadores de la eficacia de los productos. Tal como ha sido indicado por Panageas (2015), si bien existen mecanismos para promover la recopilación de datos de estudios clínicos aleatorizados que involucran a poblaciones pequeñas, su finalización puede llevar una década y aun así no lograr demostrar un beneficio clínico significativo (Panageas, 2015). En ese sentido, las agencias regulatorias están implementando aprobaciones condicionales para los productos que implican la necesidad de analizar la efectividad de los tratamientos luego de obtener la aprobación regulatoria. Adicionalmente, el rechazo por parte de los financiadores a las prescripciones de los médicos ha ido en aumento como consecuencia, en parte, de un desbalance entre el precio de las tecnologías y la incertidumbre asociada a su efectividad para mejorar la vida de los pacientes.

En un contexto muy desafiante y con grandes transformaciones, otra tendencia de gran relevancia en la evolución de las dinámicas existentes dentro del sistema de salud, es la producción y el uso de datos. El sistema de la salud no está aislado de los avances en ese sentido. Justamente en esa línea, el uso de herramientas digitales que permiten sistematizar el análisis de patrones de salud y seguimiento de la población vienen en crecimiento. En este sentido, se ha propuesto que el desarrollo del *big data* y las técnicas predictivas analíticas pueden contribuir a mejorar la precisión de la asignación en salud pública a través de la provisión de un sistema de supervisión y evaluación de los pacientes (Pastorino et al., 2019). El análisis de una gran cantidad de datos constituye una fuente inestimable de recursos a ser utilizados como base de investigación poblacional, evaluación de las intervenciones en la población y la toma de decisiones de política informadas.

En el mismo trabajo de Pastorino et al. (2019), se explica que la implementación de la medicina de precisión sigue siendo contingente a la adquisición de información y al análisis a tiempo para determinar la base más apropiada sobre la cual optimizar la prevención individual, el diagnóstico y el tratamiento oportuno. Luego, se expresa que *“alcanzando una gobernanza efectiva y proporcionada de la información relevante para la salud será esencial para los futuros sistemas de salud y se requiere que los distintos participantes del sistema colaboren y adaptan el diseño y efectividad de sus sistemas para alcanzar el máximo potencial innovador de información y de tecnologías de la innovación en salud”* (Pastorino et al., 2019, 26).

III. El sistema de Salud Argentino

Figura 2: Representación del sistema de salud en sus tres subsectores en Argentina



Fuente: Elaborado por el autor a partir de Torres et al. (2020)¹²

En la Argentina, la población tiene acceso a la salud a través de tres subsectores principales representados en la figura 2: la seguridad social, el subsector privado y el subsector público. Éstos son los que se ocupan de financiar los servicios de salud que se proveen a la población y los medicamentos que demandan sus beneficiarios. Se diferencian entre sí por su forma de financiamiento, mecanismos de compra, proveedores de salud que los integran y, en cuanto a los usuarios para los que brindan cobertura.

Dentro de la seguridad social, se da cobertura a través de múltiples financiadores de salud que son: las obras sociales nacionales, las obras sociales provinciales, las obras sociales especiales y el PAMI (INSSJP). Si bien están todos enmarcados dentro de la Seguridad Social, las distintas familias de financiadores tienen diversidad en cuanto a su criterio de gestión, reglamentación y cobertura de sus poblaciones beneficiarias. En los siguientes párrafos se busca presentar brevemente a las distintas familias de financiadores, con sus particularidades legislativas y de gobernanza, que sirve de punto de partida fundamental para describir brevemente a la población analizada en este trabajo.

¹² Torres, R., Jorgensen, N., Robba, M. (2020). Mitos y realidades de las obras sociales. Universidad ISalud. Buenos Aires. Argentina.

En primer lugar, integran al Subsector de la Seguridad Social, a las Obras Sociales Nacionales. En el informe de Cobertura de Salud en Argentina (Ministerio de Salud Argentina, 2023), se indica que las Obras Sociales Nacionales (OSN) brindan cobertura de salud a los trabajadores asalariados y sus familias según ramas de actividad, que representan a los 32,13 puntos porcentuales de la población (14.795.376 personas sobre el total de población del Censo 2022). Además, del mismo informe (Ministerio de Salud Argentina, 2023), se indica que este sector está integrado por 291 obras sociales (sindicales y de dirección) que presentan grandes diferencias entre sí como consecuencia de su composición demográfica y su tamaño. Por un lado, en cuanto a su tamaño, se diferencian ya que tienen poblaciones que van desde menos de 1000 hasta más de un millón y medio de afiliados. Por otro lado, en relación a las condiciones epidemiológicas, distribución geográfica y capacidad económica, se observa que sus ingresos estimados por afiliado, a la fecha de los datos utilizados¹³, van desde \$2000 hasta \$40000 por mes. Estas obras sociales están reguladas por la Superintendencia de Servicios de Salud y tienen la obligatoriedad en la cobertura del Plan Médico Obligatorio (PMO).

El financiamiento de las OSN está dado por los aportes y contribuciones de los trabajadores y sus empleadores, principalmente. Obtienen ingresos, además, a través de diversos programas coordinados por la Superintendencia de Servicios de Salud. En particular, para la incorporación de tecnologías sanitarias de alto precio, las Obras sociales nacionales obtienen financiamiento a través del programa nacional SURGE¹⁴. Según el decreto que reglamenta su implementación (Decreto 731/2023 de la Superintendencia de Servicios de Salud), se actualiza la forma de gestionar, reglamentar y administrar los recursos provenientes del Fondo Solidario de Redistribución, cuyo objetivo es fortalecer la atención de la salud de los beneficiarios del Sistema Nacional del Seguro de Salud. Por otro lado, las OSN obtienen financiamiento a través de otros programas implementados para el reparto de fondos (SUMA, SUMARTE, SUMA 65, SANO¹⁵). A partir de un análisis para la

¹³ Se utilizan los datos disponibles en el Boletín Anual de Seguridad Social de la AFIP con información para el año 2021. Para más datos al respecto de esta información remitirse al anexo 1.

¹⁴ El Sistema único de Reintegro por Gestión de enfermedades (SURGE), es un programa que modifica al Sistema único de Reintegros (SUR) y que se ocupa de gestionar los recursos de los fondos destinados a apoyar a los Agentes del Seguro de Salud para la cobertura de subsidios a través de reintegros por prestaciones de alto impacto económico y que demanden una cobertura prolongada en el tiempo. El fin del programa es asegurar el otorgamiento de prestaciones de salud igualitarias, garantizando a los beneficiarios la obtención del mismo tipo y nivel de prestaciones. La modificación del programa, implementada en marzo de 2023, propone efectuar cambios sustanciales y operativos que aporten mayores ventajas, tanto en la simplificación de la carga de información que efectúan los Agentes del Seguro de Salud, como en el proceso de validación y análisis de los datos por parte de las áreas con competencia específica para la consecución de los objetivos y finalidad del sistema de reintegros a implementar. En la forma de gestionar los recursos disponibles por el Fondo Solidario de Redistribución, se incorpora una visión estratégica de economía de la salud y evaluación de tecnologías sanitarias y se favorece un enfoque de gestión de enfermedades orientado a contar con valiosa información de índole sanitaria, científica, epidemiológica y estadística. Con ello, se busca mejorar la implementación de las políticas sanitarias que requieran de los recursos que gestiona la Superintendencia de Servicios de Salud.

¹⁵ El Subsidio de Mitigación de Asimetrías (SUMA) es un programa aprobado para resolver asimetrías en la población cubierta por los financiadores de Salud.

El Subsidio de Mitigación de Asimetrías para el Régimen de Trabajo Especial (SUMARTE) busca compensar los aportes realizados por grupos de trabajadores que tienen regímenes de trabajo especiales como por las empleadas domésticas o monotributistas.

El Subsidio de Mitigación de Asimetrías para mayores de 65 años (SUMA 65) busca compensar a aquellas obras sociales nacionales que dan cobertura a grupos poblacionales de 65 años o más.

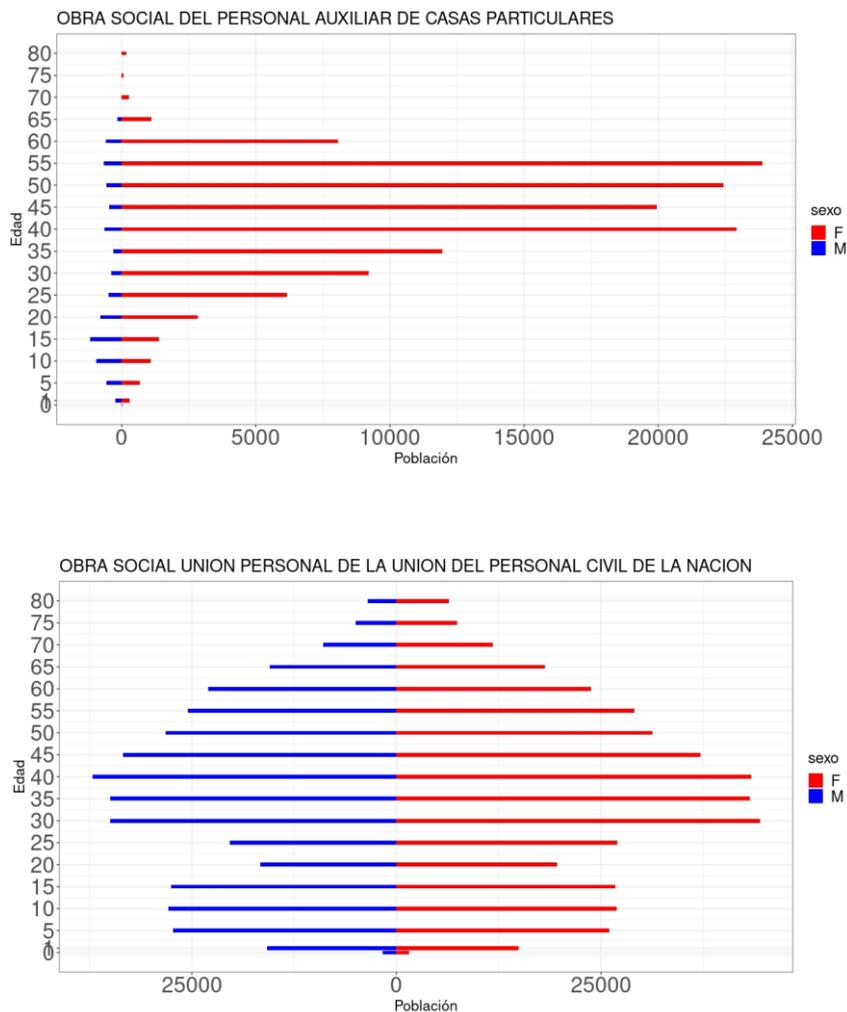
El Subsidio Automático Nominativo de Obras Sociales (SANO) está destinado a transferir recursos a las Obras Sociales en función de la edad y el sexo de sus afiliados y el ingreso promedio por grupo familiar.

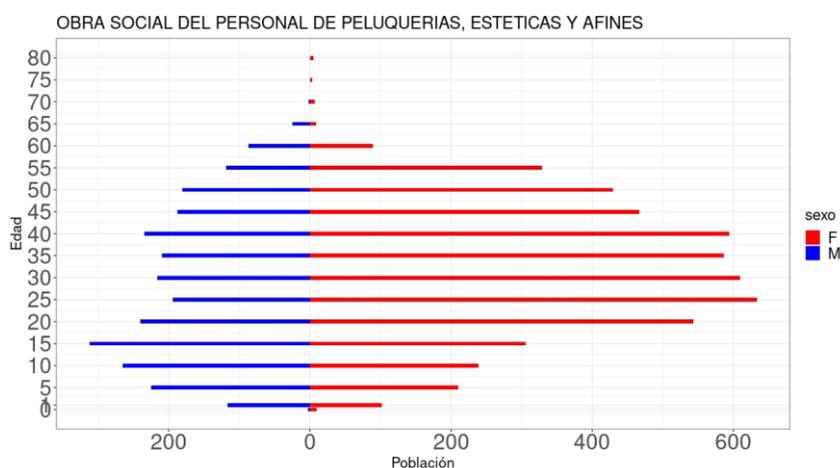
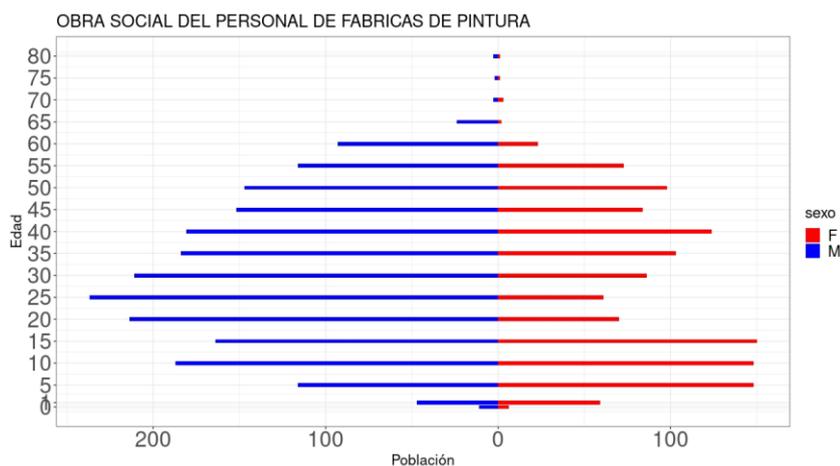
Se incluye un análisis sobre el financiamiento en el Anexo 1.

estimación de ingresos promedio, realizado en este trabajo para el año 2021, se estimaron los ingresos porcentuales correspondientes a cada vía de financiamiento para las obras sociales Nacionales. Se incorpora un análisis detallado sobre el financiamiento a través de sus diversos mecanismos en el Anexo 1.

Tal como se muestra en la figura 3, los grupos poblacionales con cobertura a través de estos financiadores, como consecuencia de representar a asalariados por rama de actividad, muestran las particularidades de cada sector de trabajadores. Por ejemplo, la obra social que brinda cobertura al personal de peluquerías está representada, en su gran mayoría, por mujeres jóvenes. Esto mismo ocurre, en mayor dimensión incluso, para la obra social del personal Auxiliar de Casas Particulares. Por el contrario, financiadores de sectores industriales, como el personal de fábricas de pintura, representa en mayor medida a beneficiarios de sexo masculino. Por último, la obra social del personal civil de la nación (UPCN) está representada por una pirámide poblacional más balanceada entre hombres y mujeres y en los distintos rangos etarios representados. Asimismo, se evidencia poca población de edad mayor a 65 años, como consecuencia de que dicha población pasa a estar cubierta, en su mayoría, por el INSSJP.

Figura 3: Distribución de afiliados por género y edad para Obras Sociales Nacionales. Ejemplos





Fuente: Superintendencia de Servicios de Salud (abril 2022) ¹⁶

Por otro lado, cada provincia cuenta con una Obra Social Provincial (OSP) que asegura en forma obligatoria a los empleados públicos activos y jubilados que dependen de los gobiernos provinciales (docentes, policías y empleados administrativos del estado). Cubren en conjunto, según surge de los datos publicados en el informe de Cobertura de Salud (Ministerio de Salud Argentina, 2023), a 15,58 puntos porcentuales (7.176.619 de personas) dentro del total de población de Argentina. Como todas las provincias tienen su propia obra social, además de la Ciudad Autónoma de Buenos Aires, hay 24 en total. Entre ellas, IOMA, que cubre a los trabajadores de la provincia de Buenos Aires, es la más grande con más de 2 millones de afiliados. En la figura 43, se muestra la distribución poblacional de las dos obras sociales provinciales con más afiliados del país donde se puede destacar la gran participación del género femenino y, por otro lado, a diferencia de las OSN, la presencia de adultos mayores dentro de la población cubierta.

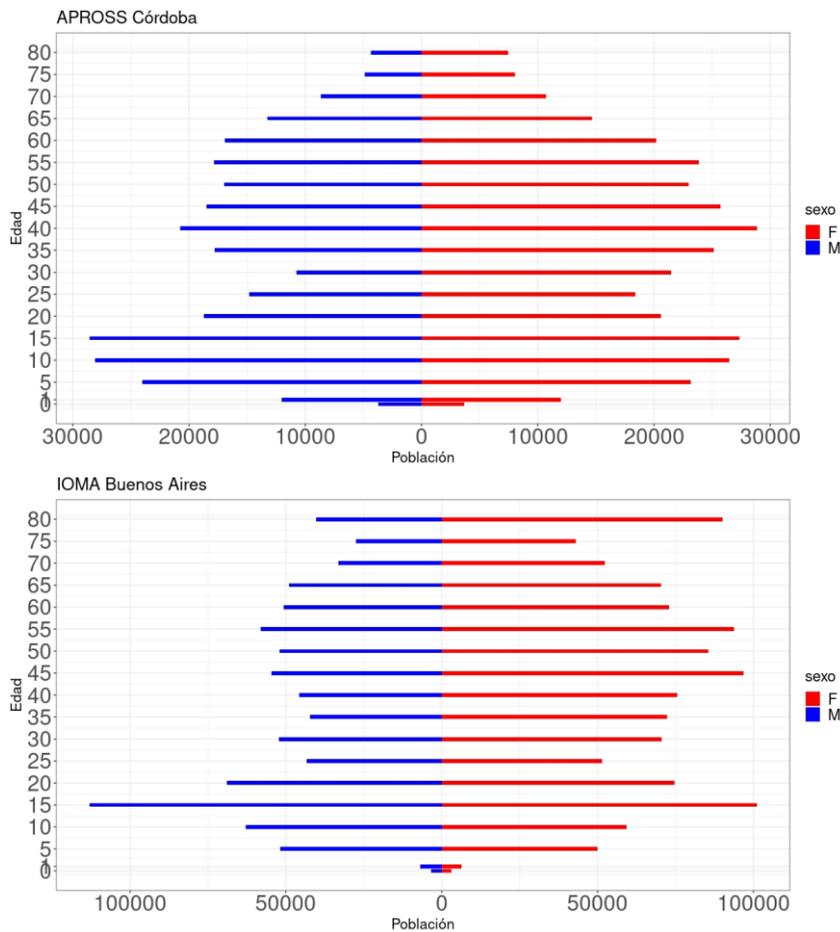
Las OSP participan del Consejo de Obras y Servicios Sociales Provinciales de la República Argentina (COSSPRA) que fue creado en 1966 y cuyo objetivo es la interrelación entre las obras sociales y la coordinación de convenios de atención recíprocos. A pesar de que están agrupadas a través del mencionado organismo, este grupo de financiadores está regulado por legislaciones provinciales y, en consecuencia, tiene mecanismos diversos de incorporación de las distintas prestaciones que ofrecen a sus beneficiarios. Asimismo, deciden de manera autónoma cuáles y en qué momento brindar cobertura a nuevas tecnologías sanitarias. En este sentido, la principal diferencia con

¹⁶ Superintendencia de Servicios de Salud (2022). Padrón de afiliados a obras sociales y empresas de medicina prepaga. Accedido por un pedido de acceso a la información con fecha abril 2022.

respecto a las OSN se da como consecuencia de que no están reguladas por la Ley 23.661 de Agentes del Sistema Nacional del Seguro de Salud y, por otro lado, no están obligados a cumplir con el Programa Médico Obligatorio (PMO). Por otro lado, carecen de Fondos Solidarios que gestionen de manera colectiva.

Al igual que las obras sociales nacionales, las OSP se financian principalmente de los aportes y contribuciones de los trabajadores y sus empleadores, los gobiernos provinciales. Por lo que indica Torres et al (2020), para el financiamiento de las OSP, los trabajadores aportan recursos en proporción a sus salarios y el Estado Provincial, en su rol de agente de retención, obtiene ingresos y, a la vez aporta contribuciones en su rol de empleador. El mencionado autor (Torres et al, 2020), indica que existe una relación especial entre el Estado Provincial y las OSP donde el primero puede, en algunas situaciones, obtener financiamiento de las OSP a través de generación de deuda. Asimismo, en algunos casos y según lo indica el mismo autor (Torres et al, 2020), en algunos casos perciben ingresos adicionales de gobiernos provinciales o municipales para fortalecer sus prestaciones.

Figura 4: Distribución de afiliados por género y edad para Obras Sociales provinciales. Ejemplos



Fuente: Elaboración propia en base a bases diversas y técnicas de modelado¹⁷

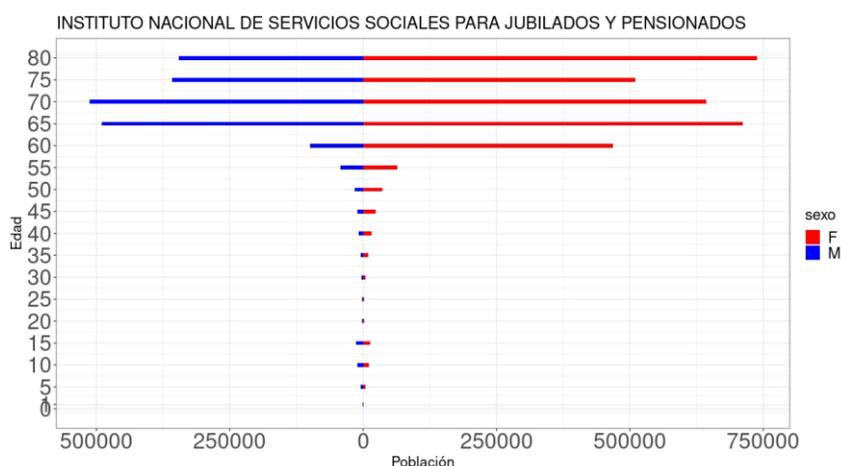
¹⁷ Elaborado en base a datos obtenidos de PUCO (2021), Estimaciones de distribución por edad y género construidas por el autor. Remitirse al anexo 1 para una descripción en detalle del mecanismo mediante el cual se ha estimado esta información

Otra institución de gran relevancia dentro de la seguridad social, es el Instituto Nacional de Servicios Sociales para Jubilados y Pensionados, también conocido como Programa de Asistencia Médica Integral (INSSJP - PAMI). Fue creado en el año 1971 mediante el decreto Ley 19032 y sus modificatorias. Brinda cobertura a los jubilados del sistema nacional de previsión y sus familias. Según indica el informe de Coberturas de Salud en Argentina (2023), cubre a 5.189.599 millones de beneficiarios, 11,27 puntos porcentuales dentro del total de la población de Argentina. Su financiamiento se obtiene por aportes de los trabajadores en actividad y retirados a través de la percepción de ingresos generados por leyes que lo rigen. En particular, tal como se presenta en el informe de Coberturas de Salud (2023), el financiamiento del INSSJP está dado por: aportes de los trabajadores en actividad (público, privado y autónomo), sus respectivos empleadores y las retenciones que se hacen a los jubilados y pensionados de sus haberes; por otro lado, lo compone lo producido de los aranceles que cobre por los servicios que preste, las donaciones, legados y subsidios que reciba, todos los intereses y rentas de los bienes que integran su patrimonio y el producido de la venta de esos bienes y los aportes del Tesoro que determina la ley de Presupuesto Nacional cada año. Al igual que las OSP, el INSSJP no está incluido dentro de los Agentes del Sistema Nacional del Seguro de Salud regulados por la Ley 23.661 y no tiene la obligatoriedad en el cumplimiento del Programa Médico Obligatorio (PMO).

Según las leyes que lo regulan, se trata de un organismo no estatal cuya personalidad jurídica está separada de la administración pública nacional y, está dotado de “individualidad financiera y administrativa”, tal como se indica en la ley 19032, que lo regula. Si bien el INSSJP es un ente público no estatal, su directorio es designado por el Presidente de la Nación. Es decir, según se indica en las regulaciones que lo definen, el Estado tiene un rol de contralor de la gestión administrativa de las prestaciones de la seguridad social que otorga el Instituto. En consecuencia, el Instituto funciona como un organismo descentralizado del Ministerio de Salud de la Nación (MSN) pero que toma decisiones de cobertura que son autónomas al mencionado ministerio.

Como ha sido indicado, su cobertura está ligada principalmente a las personas jubiladas y pensionadas y también a los ex combatientes de Malvinas y a otros grupos elegibles. Dentro de su población, 85 puntos porcentuales tienen más de 60 años y, por otro lado, da cobertura a 75 puntos porcentuales dentro de la población total nacional mayor a 65 años (Coberturas de Salud en Argentina, 2022). Tal como se puede ver en la figura 5, su población principalmente es mayor de 65 años. Asimismo, se evidencia en su pirámide poblacional, una mayor proporción de mujeres (lo cual evidencia la mayor expectativa de vida de este género por sobre los hombres).

Figura 5: Distribución de afiliados por género y edad para el PAMI (INSSJP)

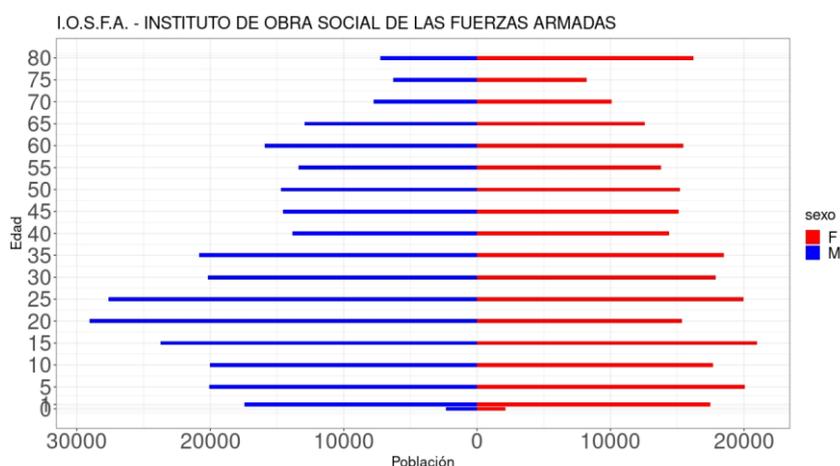


Fuente: Portal de Datos abiertos PAMI¹⁸

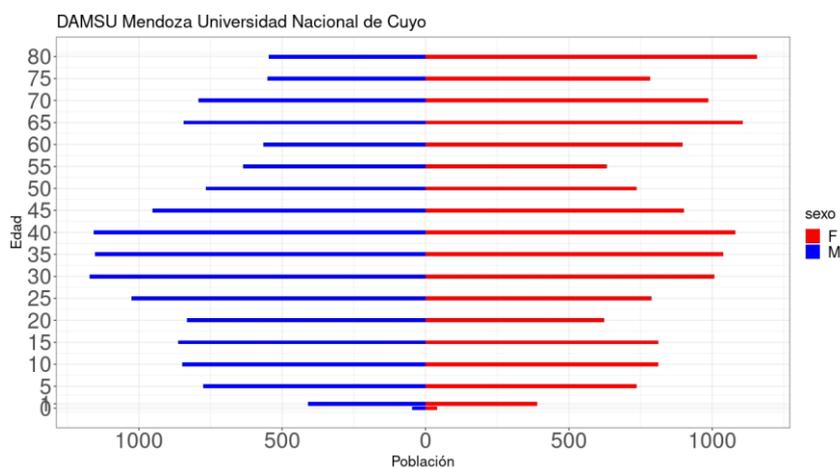
La seguridad social en los financiadores de salud de Argentina, se termina de integrar con un conjunto de financiadores que tienen regulaciones especiales y que brindan cobertura a algunos grupos específicos de trabajadores. Son obras sociales que fueron creadas por leyes especiales y cuyo programa médico, financiamiento, etc. difiere del resto de las OSN (por ejemplo, no están obligadas a cubrir el PMO). En este sentido, poseen diversas fuentes de ingresos y su cobertura no está sujeta a la gobernanza de la Superintendencia de Servicios de Salud. Es por ello que tienen autonomía en la toma de ciertas decisiones de cobertura. Se destacan: Fuerzas de Seguridad y militares (IOSFA; 600.000 beneficiarios); Universidades Nacionales (agrupadas en COSUN, 300.000 beneficiarios); Poder Legislativo (DAS, 28.000 beneficiarios); Poder Judicial de la Nación (OSPJN, 47.000). En la figura 6, se presenta el comportamiento de cobertura en particular para IOSFA, donde se destaca la en mayor medida población masculina y también se observa la presencia, significativa, de personas mayores. Por otro lado, en la misma figura se muestra la pirámide poblacional construida para una obra social universitaria donde se observa cobertura de personas mayores y una pirámide poblacional similar a varias de las obras sociales nacionales.

¹⁸ Se utiliza la información disponible en el portal de datos abiertos del organismo que corresponde a marzo de 2022 (fecha de consulta: 15 de agosto de 2022)

Figura 6: Distribución de afiliados por género y edad para Obras sociales especiales



Fuente: Informe de afiliados IOSFA (2022)¹⁹



Fuente: Elaboración propia²⁰

En su conjunto, tal como se puede concluir a partir del informe de Cobertura de Salud (Ministerio de Salud, 2023), el subsector de la Seguridad Social brinda cobertura a 19.516.522 personas, lo cual representa a 42.38 puntos porcentuales de la población²¹. La prestación de los servicios de salud para esta población está integrada por profesionales y establecimientos de salud profesionales (que podrían o no, según el caso, depender o ser exclusivos para los beneficiarios de algún financiador de los mencionados en particular).

Por otro lado, el subsector privado de la salud está integrado por las empresas de medicina prepaga (EMP). Según el informe de Cobertura de Salud (Ministerio de Salud, 2023), tienen acceso a la salud

¹⁹ Informe de afiliados de IOSFA elaborado con datos al 2 de junio de 2022.

²⁰ Para mayor información sobre la estimación de estos grupos poblacionales remitirse al anexo 1

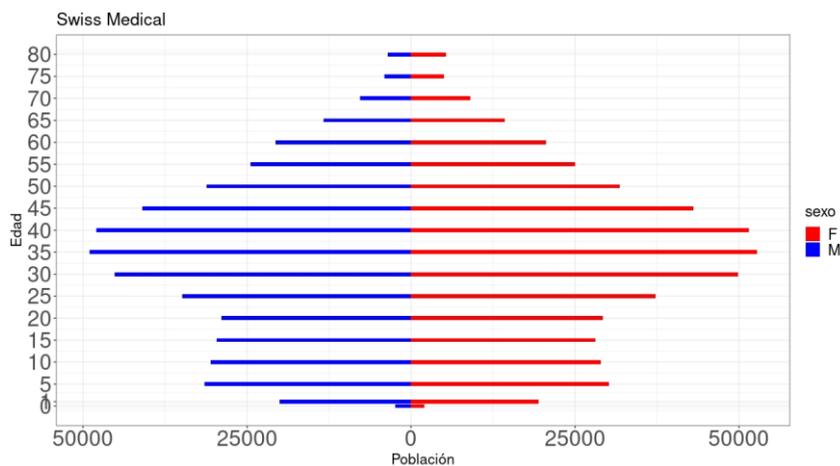
²¹ Es muy complejo estimar con exactitud este porcentaje como consecuencia de la existencia de cobertura múltiple, o también conocida como doble cobertura. Esta problemática implica que las personas puedan ser contadas en dos o más financiadores en simultáneo y se da cuando, por ejemplo, en una familia con más de un trabajador se obtiene cobertura para la pareja. Asimismo, se da cuando una persona trabaja para más de un empleador y no le está permitido unificar sus aportes como consecuencia de las leyes de cobertura obligatoria que tienen ciertos financiadores. En estos casos podría haber dificultades para definir quién se hace efectivamente responsable de proveer los tratamientos para ese conjunto de usuarios contados de manera repetida.

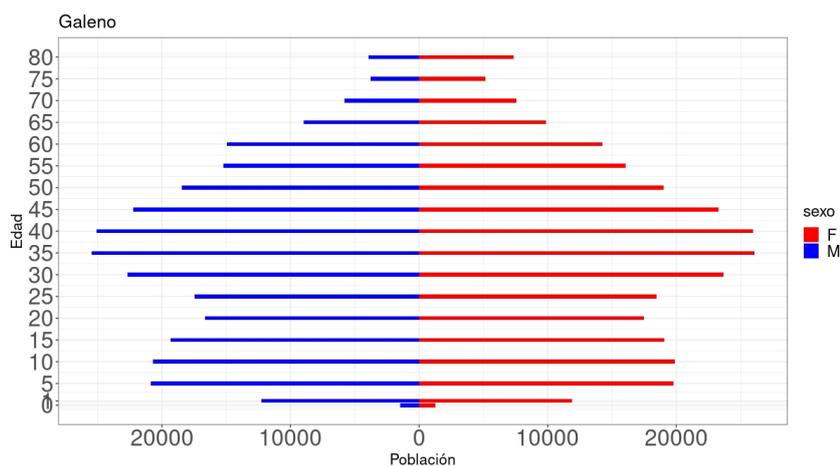
a través de este régimen, un total de 6.796.690 personas, alrededor de 14 puntos porcentuales dentro de la población de Argentina. La población de beneficiarios está integrada por personas afiliadas de forma individual o familiar que pagan mensualmente una cuota a estos financiadores. Según consta en el informe de Coberturas (Ministerio de Salud, 2023), a agosto de 2022, el Registro Nacional de Entidades de Medicina Prepaga (RNEMP) de la SSS contabiliza 674 Entidades de Medicina Prepaga. Dentro del total de entidades, tal como ha sido analizado por el autor a partir de los datos obtenidos de la Superintendencia de Servicios de Salud, las primeras 11 concentran a cerca de 90 puntos porcentuales dentro de la población total en este subsector. El financiamiento está dado por las cuotas que aportan quienes se afilian voluntariamente a estas compañías, por algunas empresas que brindan cobertura a sus trabajadores y por la derivación que realizan algunos trabajadores asalariados desde su obra social nacional. Bajo el concepto de la desregulación, la mayor cantidad de beneficiarios de las prepagas son trabajadores que ejercen la opción de cambio desde su Obra Social de origen a otras que brindan cobertura a través de convenios con EMP, o de convenios de empresas directamente con las EMP. Con el objetivo de analizar la distribución demográfica y poblacional dentro de este grupo, se ha utilizado información obtenida a través de un pedido de información pública a la Superintendencia de Servicios de Salud.

Es de gran relevancia indicar que estas entidades se encuentran regidas por la Superintendencia de Servicios de Salud (SSS) y están legalmente obligadas a garantizar a sus afiliados el piso mínimo de prestaciones establecido en el Programa Médico Obligatorio (PMO).

En la figura 7 se presenta la distribución poblacional de dos financiadores de gran relevancia dentro de este subsector. A partir del análisis gráfico, se observa gran concentración de poblaciones económicamente activas (entre 20 y 45 años), y la presencia de muchos nacimientos y niños cubiertos por empresas de medicina prepaga. Asimismo, se evidencia cómo no es muy significativa la presencia de adultos mayores dentro de estos financiadores.

Figura 7: Distribución de afiliados por género y edad para Empresas de Medicina Prepaga





Fuente: Superintendencia de Servicios de Salud (diciembre de 2021) ²²

En Argentina, el acceso a la protección de la salud es un derecho constitucional. En tal sentido, el último gran subsector significativo dentro del financiamiento de la salud, es el estado, que tiene la responsabilidad última de proteger y garantizar el derecho a la salud. A partir de esto, ejecuta presupuesto a través de sus distintos niveles de gobierno. El Ministerio de Salud de la Nación es la máxima autoridad nacional en materia de salud y ejecuta y elabora diversos programas, normas y acciones que permiten la coordinación entre los distintos subsectores (Bisang & Cetrángolo, 1997). Sin embargo, tal como menciona el trabajo de Bisang y Cetrángolo (1997), dada la estructura federal del país, los gobiernos provinciales cuentan con total autonomía en materia de políticas de salud, lo que hace que los lineamientos del nivel nacional tengan solamente un valor indicativo, estando la adhesión de los mismos condicionada a las coincidencias en el plano político más general (Bisang & Cetrángolo, 1997). En definitiva, cuando se observa el financiamiento de la prestación de medicamentos para la población con cobertura exclusiva, se observa que la población tiene acceso a la salud a través de los hospitales públicos, de algunos municipios que proveen servicios en sus hospitales o centros de salud, a través de los ministerios de salud provinciales y algunos efectores municipales. Las personas que no cuentan con ninguna cobertura formal, en consecuencia, disponen de Cobertura Pública Exclusiva (CPE). Según surge del informe de Cobertura en Salud (Ministerio de Salud, 2023), en total 39 puntos porcentuales del total de población de Argentina, tiene cobertura pública exclusiva. Es decir, presenta el mismo informe, en total brinda cobertura a 19.921.618 personas (Ministerio de Salud, 2023). El financiamiento de este sector de salud está dado por los recursos disponibles a partir de la recaudación de los distintos entes provinciales, municipales y nacionales. La prestación de los servicios de salud se realiza a través de hospitales y centros de salud públicos. Por otro lado, el subsector público provee los medicamentos de alto precio para sus beneficiarios a través de un Banco Nacional de Drogas y los bancos de drogas provinciales.

Dentro de este sector, es relevante mencionar la situación particular de los titulares de pensiones no contributivas y sus familiares que son, en su mayor parte, una población con grandes carencias, necesidades especiales y con dificultades de acceso a atención médica. El financiamiento, la implementación y la coordinación de la atención médica que se brinda a esta población está a cargo del Programa Nacional INCLUIR SALUD, dependiente de la Agencia Nacional de Discapacidad. Las provincias, de todos modos, reciben transferencias para efectivizar la implementación de ese

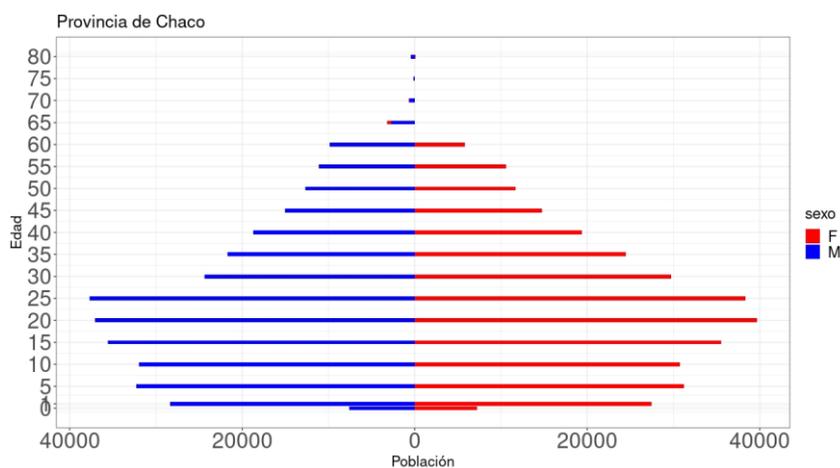
²² Superintendencia de Servicios de Salud (2022). Padrón de afiliados de empresas de medicina prepaga. Accedido por un pedido de acceso a la información con fecha diciembre de 2021.

programa en sus territorios. Asimismo, el INSSJP brinda cobertura a un subgrupo de esta población (ex combatientes de Malvinas, principalmente).

Otro programa de gran relevancia en el análisis de la población con cobertura pública exclusiva, es el Programa Sumar. Es una política pública que tiene como beneficiarios a la población cobertura pública exclusiva y que promueve un acceso equitativo y de calidad a los servicios de salud para toda la población que no posee cobertura formal en salud. Se trata de la ampliación del Plan Nacer, iniciado en el 2004 en las provincias del norte argentino y extendido en el 2007 al resto del país para brindar cobertura de salud a la población materno-infantil. A lo largo de los años, fue ampliando su extensión progresivamente, logrando incorporar a toda la población en el año 2020. A través de un esquema de diseño de financiamiento por resultados, financia cerca de 800 prestaciones y servicios de salud, de APS, y de alta complejidad (Cardiopatía congénitas, IAM, Módulos COVID de internación, entre otros). Ejecuta transferencias que están a cargo de la Dirección Nacional de los Sistemas Provinciales a las distintas provincias que luego ejecutan las tareas de provisión de los servicios de salud.

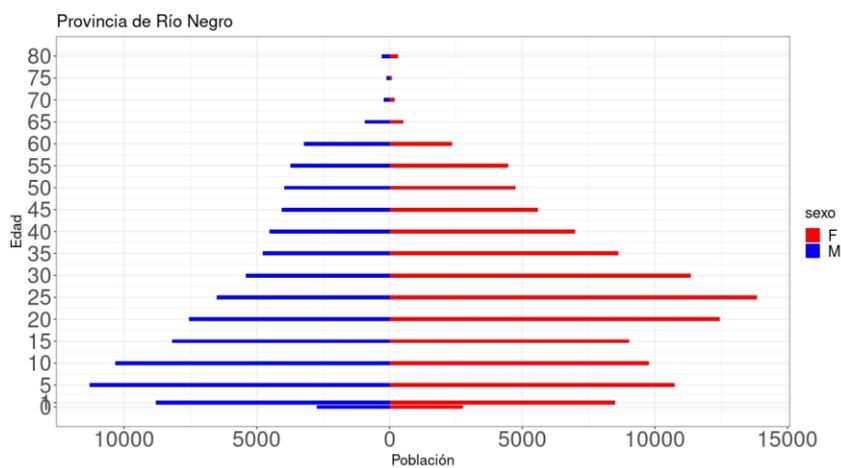
En el caso particular de este trabajo, resultan de gran relevancia los registros del programa, a los que se accedió a través de un pedido de acceso a la información pública (en el Anexo 1, se indican más detalles), con el fin de entender la composición de la población dentro de este subsector. Se observa que, si bien sus características demográficas y de total de población cubierta varía entre provincias, en general hay alta cobertura de población joven, y de bebés recién nacidos, niñas y niños. Se presenta en la figura 8 la distribución poblacional para algunos financiadores de ejemplo dentro de la prestación de cobertura pública.

Figura 8: Distribución de beneficiarios por género y edad para la población con cobertura exclusiva del estado (según registros de empadronamiento en programa SUMAR)

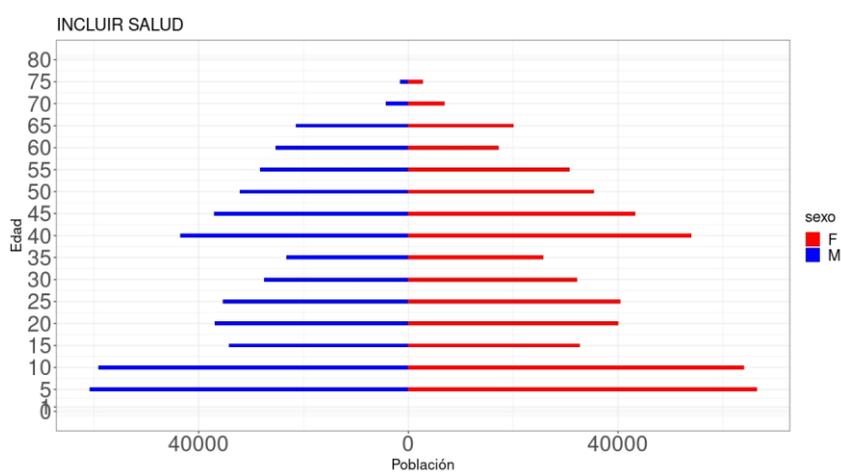


Fuente: SUMAR (2022)²³

²³Ministerio de Salud Argentina (2022). Beneficiarios programa SUMAR. Datos provistos por consulta de acceso a la información con datos de beneficiarios a diciembre de 2021. Más detalles al respecto en anexo 1.



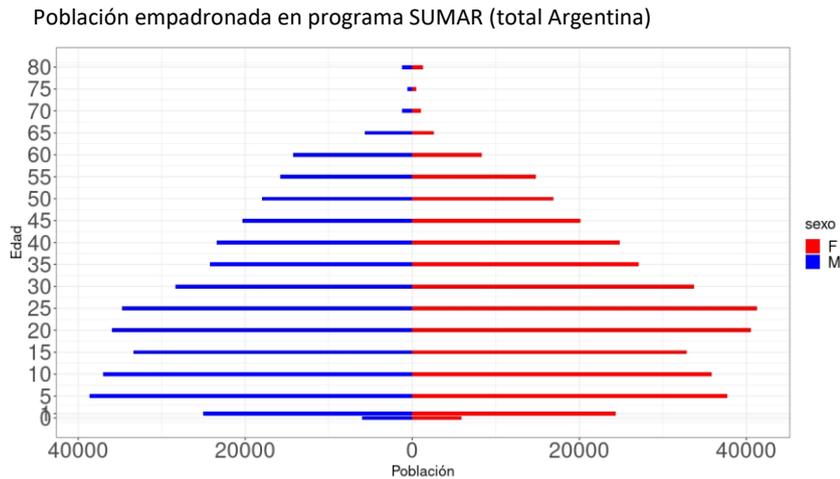
Fuente: SUMAR (2022)²⁴



Fuente: Registro Agencia Nacional de Discapacidad (marzo 2022)²⁵

²⁴Ministerio de Salud Argentina (2022). Beneficiarios programa SUMAR. Datos provistos por consulta de acceso a la información con datos de beneficiarios a diciembre de 2021. Más detalles al respecto en anexo 1.

²⁵Se accede a padrón informado por la Agencia Nacional de Discapacidad mediante un pedido de acceso a la información con datos actualizados a agosto de 2022 (EX-2022-84660323- -APN-DNAIP#AAIP).



Fuente: SUMAR (2022)²⁶

En consecuencia, contemplando todos los subsectores presentados, existen más de 441 organismos públicos y privados que se ocupan de financiar la salud de la población. Los usuarios de los servicios de salud, como se muestra en las figuras anteriores, presentan características muy dispares entre sí.

Como resulta evidente, el sistema de salud argentino es un sistema muy fragmentado con gran disparidad territorial, de género, por grupos etarios, y por capacidad e ingresos económicos. En consecuencia, el reparto del riesgo de tener que enfrentar gastos más altos para dar cobertura a sus poblaciones también resulta muy dispar. Si se analizan en profundidad las estadísticas publicadas por el Observatorio Global del Cáncer (Globocan, 2020), dependiente de la organización Mundial de la Salud, la prevalencia esperada de tener cáncer de mama para las mujeres es mucho más alta que para los varones. En consecuencia, se podría esperar que si un financiador representa a una población más alta de mujeres, su incidencia esperada de cáncer de mama será más alta. Por otro lado, la prevalencia del cáncer de pulmón, según las mismas estadísticas (Globocan, 2020), es mucho mayor en la población mayor a 65 años. En consecuencia, los financiadores con edades promedio más altas tienen más riesgo de tener más incidencia de esta patología entre sus beneficiarios.

IV. El proceso de incorporación de nuevas tecnologías sanitarias para la población

A grandes rasgos, existen tres entornos de validación relevantes para la incorporación de una tecnología sanitaria innovadora: científico, regulatorio y de cobertura. El primero es ejecutado por los laboratorios productores de tecnologías. Son los primeros que se ocupan de determinar con rigurosidad científica que se cumplan los objetivos buscados en torno a una tecnología y de identificar que existe suficiente evidencia para continuar con una investigación. En segundo lugar, en el entorno regulatorio, se identifica que existe suficiente evidencia como para que la tecnología sea registrada. Por último, se define la cobertura donde se determina que la evidencia y el impacto económico implican la aprobación para los beneficiarios de un financiador específico.

Dentro de ese contexto, se detalla brevemente la situación actual en Argentina. Una vez que el laboratorio productor de tecnologías culmina un estudio clínico donde se analiza la eficacia y seguridad del medicamento, éste es presentado para su aprobación regulatoria a la Administración Nacional de Medicamentos, Alimentos y Tecnología Médica (ANMAT). Esta institución se ocupa de revisar la evidencia disponible de las distintas tecnologías que presentan los laboratorios farmacéuticos y define la aprobación regulatoria. A partir de ese momento, las tecnologías pasan a

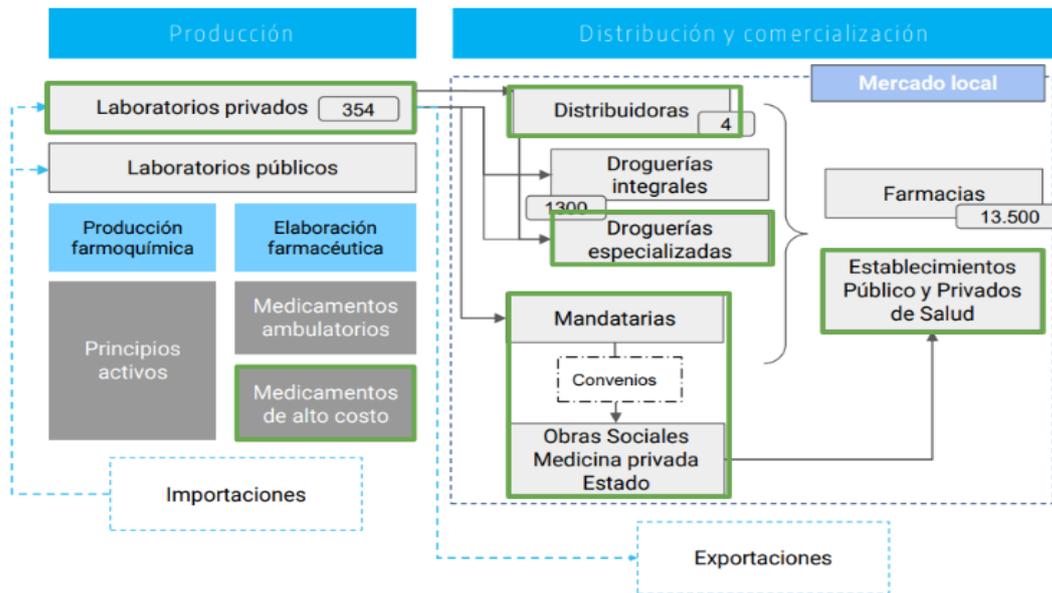
²⁶Ministerio de Salud Argentina (2022). Beneficiarios programa SUMAR. Datos provistos por consulta de acceso a la información con datos de beneficiarios a diciembre de 2021. Más detalles al respecto en anexo 1.

estar disponibles y los distintos financiadores definen si aprueban su cobertura para sus beneficiarios (luego de que el médico genere la prescripción para un paciente candidato). Como consecuencia del reparto desigual de riesgo epidemiológico y de los recursos económicos disponibles, además de la disparidad en el acceso a los servicios de salud, existe gran inequidad en cuanto a las posibilidades de que cada ciudadano pueda acceder a los tratamientos innovadores. Dentro de este contexto, los laboratorios tienen la posibilidad de negociar con cada uno de los financiadores los descuentos (precios netos de compra) a los que accederán a los productos. La comercialización de los productos en general se realiza a través de un canal directo (venta directa del laboratorio) o a través de una droguería (intermediarios entre los laboratorios farmacéuticos y los consumidores finales).

La determinación de la cobertura está también afectada por el análisis de otro organismo relevante en Argentina. Desde el año 2018, se creó la Comisión Nacional de Evaluación de Tecnologías de Salud (CONETEC), con el fin de ser una instancia de consulta permanente respecto de las decisiones sobre tecnologías sanitarias para la República Argentina a través de publicaciones científicas. En este sentido, y considerando la relevancia que vienen tomando en el proceso de priorización e incorporación de tecnologías sanitarias, la metodología que utiliza está basada en la Evaluación de Tecnologías Sanitarias (ETS). Dichas evaluaciones contribuyen al acceso universal a la salud a través de la mejora de la calidad de la atención sanitaria –ampliación de cobertura, eficacia, seguridad, costo efectividad y racionalidad en su uso-, a la difusión de innovaciones terapéuticas y al aumento en la eficiencia de los gastos. En ese contexto, en el último mes (julio de 2023), se aprobó la creación de la Comisión Nacional De Evaluación De Tecnologías Sanitarias y Excelencia Clínica (CONETEC) como organismo desconcentrado dependiente de la Secretaría de Acceso a la Salud, dependiente del Ministerio de Salud, con el objetivo de realizar evaluaciones de tecnologías sanitarias según criterios de calidad de la evidencia, beneficio clínico, impacto económico en la equidad y la salud pública, entre otros criterios que pudieran incorporarse y publicar los informes correspondientes. Se integra con representantes del Ministerio De Salud, del Consejo Federal De Salud (Co.Fe.Sa.), de la Superintendencia De Servicios de Salud, de la Administración Nacional De Medicamentos, Alimentos y Tecnología Médica (ANMAT), del Instituto Nacional De Servicios Sociales Para Jubilados y Pensionados (INSSJP), Obras Sociales Nacionales, Obras Sociales de las distintas jurisdicciones, Entidades De Medicina Prepaga, Defensoría Del Pueblo De La Nación y Asociaciones De Pacientes. El organismo tiene distintas responsabilidades alineadas a la definición de la aprobación o rechazo en la cobertura de las tecnologías sanitarias.

Como se muestra en la figura 9, extraída de un informe sobre la cadena de valor de la industria farmacéutica elaborado por el Ministerio de Economía de Argentina, la estructura que interviene en el acceso de las tecnologías sanitarias tiene cierta complejidad por el lado de la oferta y los canales de distribución. Se han señalado con verde los actores relevantes para la distribución y comercialización de medicamentos de alto precio.

Figura 9: Proceso de distribución de medicamentos. Actores relevantes

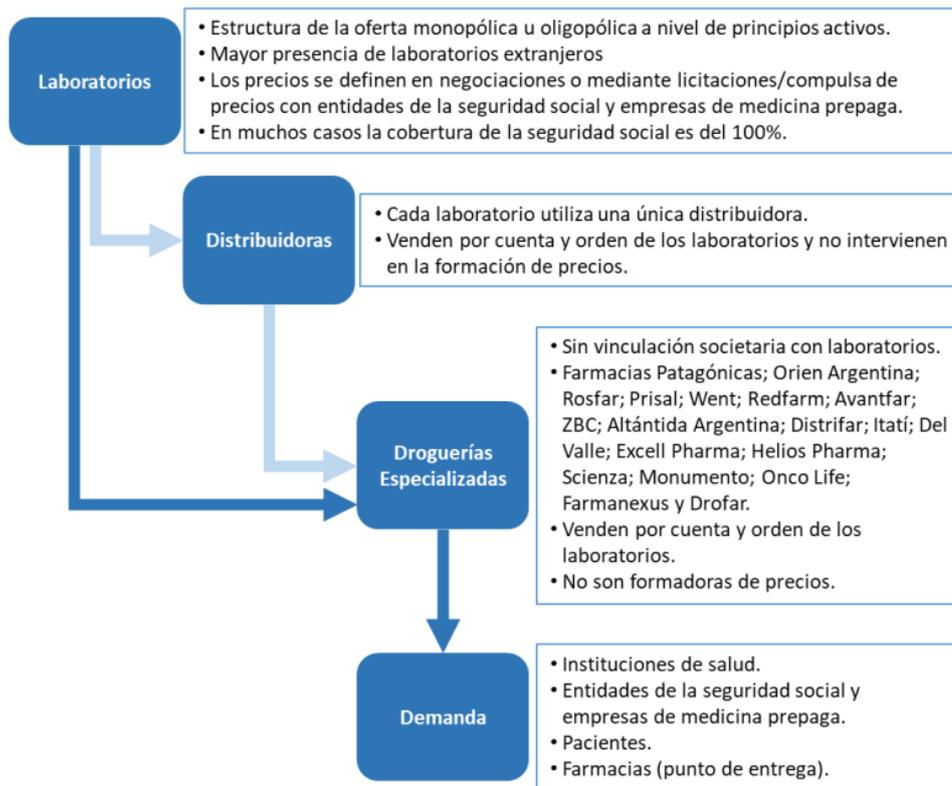


Fuente: Ministerio de Economía de Argentina (2019)²⁷

Asimismo, en un informe consultado de la Comisión Nacional de Defensa de la Competencia, se presenta un esquema sobre la participación de los distintos integrantes que forman parte de la “oferta” dentro del sistema de salud. Se presenta esta información en la Figura 10 (Comisión Nacional de Defensa de la Competencia Argentina, 2019).

²⁷ Ministerio de Hacienda, “Industria Farmacéutica”, Serie Informes de cadenas de valor, Año 3, N° 21, agosto 2018.

Figura 10: Cadena de distribución de medicamentos de alto precio

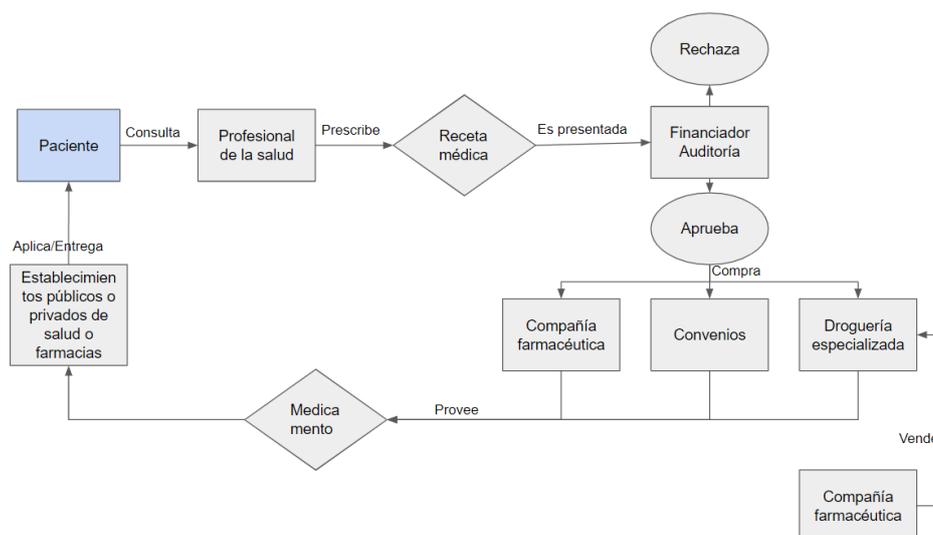


Fuente: Comisión Nacional de Defensa de la Competencia Argentina (2019) ²⁸

En este trabajo, como ha sido mencionado, se pretende analizar en particular la cobertura de medicamentos de alto precio. En este sentido, en la figura 11, se presenta un resumen del camino que comienza con la identificación de una necesidad médica para un paciente, efectuada por un profesional de la salud, y finaliza con la entrega del medicamento de alto precio a ese mismo paciente.

²⁸ IM 4 - Informe técnico sobre las condiciones de competencia en el mercado de medicamentos. Comisión Nacional de Defensa de la Competencia. Argentina. Año 2019

Figura 11: Esquema simplificado del proceso de cobertura de un medicamento de alto precio para un paciente



Fuente: Elaboración propia

Según el subsector, el tipo de tratamiento, las indicaciones y el escenario particular del que se trate, el acceso efectivo del paciente al medicamento implica mayores o menores complejidades y objeciones por parte de los financiadores para cubrir los tratamientos. En función de la burocracia interna, las características socioculturales de la población cubierta, la capacidad económica, la incidencia potencial que el financiador pudiera tener de cierta patología, la evidencia de la tecnología, los estudios diagnósticos que se hayan utilizado para justificar la prescripción, entre otros, se influye en la aprobación y también en los plazos que conlleva dicha aprobación. La compañía farmacéutica entabla un vínculo, por un lado, con el profesional de la salud, a quien capacita sobre su tecnología y otorga información que luego es utilizada al momento de la decisión prescriptiva para un paciente pasible. Por otro lado, negocia el precio al que comercializa su producto con los clientes directos – financiadores que compran directamente al laboratorio- y con las droguerías especializadas que luego venden y proveen los medicamentos a los financiadores.

V. Definición de precios para los medicamentos innovadores y la compañía en ese contexto

La definición del precio a la que se venden los productos farmacéuticos implica mayor complejidad que en otros mercados donde la información es libre para el consumidor y éste puede decidir si consume o no el bien o servicio como consecuencia de las fallas de mercado que se han presentado previamente. Asimismo, el precio es una variable clave en la decisión de rechazo o cobertura del producto por parte del financiador.

Como ha sido descrito, es complejo definir eficientemente el precio de un producto farmacéutico. Por un lado, las empresas farmacéuticas enfrentan costos hundidos sumamente elevados en el proceso de investigación y desarrollo de sus tecnologías. Luego, la compañía farmacéutica, cuando demuestra que la tecnología funciona, recupera la inversión y obtiene ganancias, a través de precios más altos a su costo marginal gracias a que disponen de una patente por tiempo limitado. En este contexto, cuando los precios de los productos son muy bajos, no se generan los incentivos suficientes para investigar en innovación (a este concepto, Chalkidou (2020), lo define como *ineficiencia dinámica*). Por otro lado, cuando los precios son muy altos, se genera *ineficiencia estática* como consecuencia de que el costo de acceder al tratamiento supera a los beneficios que otorga. En consecuencia, en la política de precios se debe balancear el objetivo del acceso sostenible a los

tratamientos existentes, por un lado, y el desarrollo de innovación, por el otro. (Chalkidou et al., 2020, 3)

Figura 12: Representación gráfica de los objetivos que se buscan balancear en la determinación del precio de un medicamento de alto precio



Elaboración propia

En la actualidad, a nivel global, para definir las decisiones sobre los precios de comercialización de las distintas tecnologías sanitarias, se promueven políticas públicas que establecen dos tipos de mecanismos para determinarlos. Por un lado, se promueve la evaluación de tecnologías sanitarias (HTA, según sus siglas en inglés – *Health Technology Assessment*-), una metodología muy difundida por varios países industrializados de la OCDE (Organización para la Cooperación y el Desarrollo Económicos) para definir la cobertura de la innovación. Consiste en evaluar toda la información clínica, económica y social relativa al uso de una tecnología. A partir de la definición de límites sobre el valor económico de la vida, se define si el uso del medicamento es costo-efectivo. Con esto, se garantiza que la decisión sobre la cobertura sea informada; es decir, se toman las decisiones contemplando las opciones disponibles de tratamiento, el costo de diagnóstico y la utilización de los tratamientos disponibles (evaluando el medicamento nuevo en comparación con el estándar de tratamiento ya establecido). En muchos países donde la decisión de compra del producto es centralizada, la evaluación de tecnologías sanitarias es un proceso efectivo para definir el precio que refleje correctamente el valor del producto para esa condición médica en particular. Otro mecanismo muy difundido para definir los precios es conocido como “precio de referencia interno”. Se establecen clústeres de áreas terapéuticas y los productos se reembolsan en relación al costo que tiene el tratamiento estándar de la patología para la que se busca incorporar a la nueva tecnología sanitaria. Éste último mecanismo, sin embargo, es muy cuestionado por no contemplar la relación precio-calidad.

A pesar de que los métodos presentados permiten de alguna manera reflejar el valor del producto (principalmente la metodología de la evaluación de tecnologías sanitarias – HTA), no se refleja la posibilidad de que el acceso a la tecnología sea equitativo para todos los potenciales usuarios. En este sentido, considerando en particular el caso argentino (donde como ya mencionamos existe gran fragmentación en la cobertura de la salud), se genera un esquema desigual en el acceso a los medicamentos.

Es como consecuencia de la presencia de inequidad en el acceso a los tratamientos innovadores que hace algunos años se viene promoviendo la “diferenciación de precios”. En tal sentido, se define que la diferenciación de precios consiste en establecer precios distintos para el mismo bien en distintos mercados (Danzon, 2018). Dicho recurso es muy utilizado en la teoría del *Revenue Management* (o,

en español, gestión del ingreso) donde, para ciertas industrias es muy frecuente el establecimiento de diversos mecanismos que permiten que el precio pagado por el acceso a un bien o servicio refleje la voluntad de pago del consumidor. Esto es así en la industria hotelera, aerocomercial, energética, y en muchos otros sectores que de a poco incorporan estas estrategias con el fin de obtener el mayor ingreso posible. En salud, dicho concepto podría redirigirse a la maximización del acceso a los tratamientos.

En esta línea, tal como sostiene Danzon (2018), *“una estrategia de diferenciación de precios funciona particularmente bien para los productos farmacéuticos por la relativa inversión alta en investigación que podría ser considerado un costo hundido global que beneficia a todos sus consumidores a nivel global”*. Asimismo, se argumenta, *“el establecimiento de precios diferenciales permite un mayor acceso a las tecnologías y un consecuente aumento en la investigación y desarrollo de otras tecnologías”*. En este sentido, tal como se plantea en Daems y otros (2011), *“para mejorar el acceso a los medicamentos a nivel global, es crítico que los fabricantes sean capaces de ofrecer a los países menos ricos, precios más bajos. Al mismo tiempo, los innovadores y fabricantes deben ser capaces de recuperar sus costos y a su vez obtener un retorno justo en su inversión que permita seguir desarrollando su negocio y escalar su capacidad productiva. Este balance delicado se puede alcanzar a través de la amortización de los gastos desigualmente entre los países de manera tal que los más pobres, paguen un precio relativamente menor, pero, sin embargo, que sea equitativo con respecto a la proporción pagada por los países más ricos”*. Más adelante indica que *“al permitir que las tecnologías estén disponibles para un número mayor de ciudadanos y teniendo en cuenta la inequidad en riqueza e ingreso, la firma en efecto está aplicando el principio de equidad en el precio y, a su vez, alcanzando sus objetivos de negocio y sociales”*. A partir de la propuesta de definición de precios diferenciales entre países, Daems y otros (2011), plantean que *“el efecto redistributivo del precio diferencial no es solo posible entre países, sino también internamente para cada país”* (Daems et al., 2011). Los autores plantean que en muchos casos dicha posibilidad es compleja como consecuencia de la gran heterogeneidad al interior de los países y los canales de distribución. Sin embargo, proponen dos ejes principales para definir una posible clusterización: una medición multidimensional de la pobreza (o MPI –*Multidimensional Poverty Index*-, en sus siglas en inglés) y una medición que expresa la proporción del gasto de bolsillo dentro del gasto total en salud (o HSCI –*Health Systems Channel Index*, en inglés). A partir de estas dos variables, se definen distintos clústeres para los que se podrían ofrecer descuentos comerciales diferenciales. El objetivo de máxima con este planteo es reflejar las diferencias entre la voluntad y disponibilidad de pago que cada país tiene para pagar por un producto. Más específicamente, la compañía farmacéutica cobra menos a países más sensibles al precio y más a aquellos con menor elasticidad y disponibilidad de pago.

En ese marco, la compañía farmacéutica para la que se pretende realizar el trabajo, tiene como objetivo promover el acceso equitativo a la innovación y a la vez, garantizar un negocio sustentable. Se plantea alcanzar ese resultado a través distintas estrategias que permitan garantizar la cobertura sustentable de las tecnologías sanitarias. Dentro de ese marco, los cuatro desafíos principales están enfocados a ser eficientes en el gasto presupuestario, garantizar incentivos para mejorar los resultados en salud, desarrollar capacidad para establecer la medicina personalizada para cada paciente y colaborar en la reducción de la burocracia en los procesos. Es lógico suponer que la capacidad de negociación de los distintos financiadores, como consecuencia de sus características, sea muy dispar y su acceso a descuentos comerciales también sea muy desigual. Sin embargo, en la actualidad, no se utiliza información epidemiológica o demográfica en la definición de los acuerdos comerciales que se ejecutan.

Se pretende incorporar estimaciones de riesgo epidemiológico, a partir de los datos demográficos, de requerir medicamentos de alto precio, en áreas terapéuticas que resultan de relevancia para la compañía farmacéutica. En este sentido, la compañía puede acompañar a la ejecución de políticas que mejoren el diagnóstico oportuno y, a su vez, definir precios óptimos que garanticen sustentabilidad y asequibilidad de los tratamientos. Si bien en la actualidad, la compañía utiliza un modelo que define la determinación de los descuentos a ofrecer para un financiador que garantiza sustentabilidad en los precios promedios para los productos, no se refleja un criterio de “equidad” en términos de los grupos poblacionales para los que se asignan y no garantiza acceso temprano a las nuevas tecnologías sanitarias. Asimismo, como hay gran diversidad de compradores de los productos, la empresa farmacéutica comercializa sus medicamentos mediante dos canales principales: venta directa (alrededor de 56 puntos porcentuales de la venta total) y venta a través de droguerías (restantes 44 puntos porcentuales). La venta a través de este último canal implica una gran dificultad en términos de información para la compañía dado que no puede identificar con exactitud los compradores finales de los productos que comercializa (pero históricamente ha dirigido los descuentos a las droguerías que luego definen sus propios negocios con los financiadores con los que se vincula).

Como consecuencia de las tendencias antes mencionadas y de la situación descrita, se están generando mecanismos alternativos de acceso a través de una plataforma propia donde se genera la posibilidad de impactar en el precio neto para todos los financiadores que demandan productos de la compañía farmacéutica. En esta línea, se posibilita la ejecución de una estrategia de precios diferenciales que agilice el acceso a las tecnologías innovadoras y garantice sustentabilidad y equidad en ese acceso.

b. Exploración y preparación de los datos

I. Descripción de variables iniciales

El primer gran desafío dentro de este trabajo está dado como consecuencia de que se involucra el uso de múltiples fuentes de datos para la elaboración de los distintos perfiles de financiadores. Asimismo, mucha de la información potencialmente relevante para esta investigación no está disponible en páginas web de datos abiertos para el acceso público de la población. En consecuencia, se han desarrollado múltiples pedidos de acceso a la información a organismos públicos que centralizan la información con el fin de incorporar esos datos en el análisis. Asimismo, se incurrieron en muchas horas donde hubo que hacer un trabajo que permitiera conectar los datos de distintas bases y estimar aquellos para los que no se cuenta con una fuente que permita obtenerlos.

A partir de las distintas fuentes e información se logró elaborar una base de datos con la siguiente información para la que se analizará la relevancia en el análisis:

- Información estadística de la población cubierta por los distintos subsectores de la salud (información provista a través de la Superintendencia de Servicios de Salud, INDEC, el Ministerio de Salud de la Nación, entre otras)
- Información sobre la disponibilidad de recursos económicos de cada financiador (estimados con información pública sobre salarios, datos provistos públicamente por AFIP y estimaciones de diversas fuentes).
- Información sobre la disponibilidad de centros propios de salud por parte de los financiadores. Dicho dato se obtiene a partir de la información pública de REFES (Registro federal de establecimientos de salud) y permite analizar el nivel de integración vertical de cada financiador de salud

- Información epidemiológica de los productos analizados en el trabajo. Estimaciones de incidencia por rango etario más afectado, género y tratamientos alternativos para cada grupo.
- Información de venta de la compañía: precios de venta promedio para cada financiador en particular, cantidades vendidas, mecanismo de compra, racionales de descuentos, etc. Esta información, sin embargo, refleja una imagen muy limitada del total del negocio ya que la demanda de la mayoría de los financiadores es información a la que la compañía farmacéutica no tiene acceso, ya que provee sus productos a través de intermediarios (droguerías) y por lo tanto, sólo hay información disponible para pocos financiadores que compran de forma directa y, adicionalmente, hay información parcial, obtenida a través de una plataforma mediante la cual la compañía otorga reembolsos y beneficios a los financiadores donde se reflejan datos para algunos financiadores asociados a la herramienta. En total, en consecuencia, se tiene información sobre los montos de descuentos y la venta para un total de 62 financiadores.

En particular, a partir de los orígenes de los datos, se detallan los distintos elementos que constituyen la base a partir de la cual se comenzó el análisis. Se incluye en el Anexo 1 y 2 el detalle de las fuentes, asunciones y mecanismos realizados con el fin de determinar ciertos valores.

INFORMACIÓN ESTADÍSTICA CONSTRUIDA DE FUENTES SECUNDARIAS PÚBLICAS DE INFORMACIÓN

- **Financiador:** se tiene información para 392 organismos, públicos y privados, que constituyen cerca del 100% del consumo institucional de productos farmacéuticos de alto precio (PAMI, empresas de medicina prepaga, obras sociales nacionales, obras sociales provinciales, organismos públicos -banco de drogas nacional y provinciales, Incluir Salud- y obras sociales con regímenes especiales -IOSFA, Universitarias, entre otras-)
- **Subsector de salud:** Público o cobertura exclusiva pública, privado o EMP, nacional OSN, provincial OSP, especial
- **Región principal:** se indica si se estima que tiene afiliados en todo el país o solo en alguna región en particular
- **Total de afiliados:** en total se tiene información para 54359350 personas en el país²⁹. Si bien no existe algún número explícito sobre doble cobertura de la población, se entiende que es de entre un 10 y un 15%, por lo que podríamos trabajar perfectamente con esa información
- **Edad promedio de los afiliados:** en promedio la edad para Argentina construida por los datos es de alrededor de 36 años. 7 financiadores tienen afiliados con edades en promedio mayores a los 55 años, 93 tienen afiliados con edades de en promedio entre 40 y 55 años, los restantes 320 tienen entre 24 y 39 años en promedio. A partir de esto, se busca definir el riesgo adicional asociado a la demografía de la población a enfrentar costos más elevados en los tratamientos de sus beneficiarios
- **Ingresos mensuales totales estimados:** construido a partir de la mejor fuente de información posible (presupuestos ejecutados, cálculos a través de estimaciones de salarios/retenciones para el sector, transferencias de organismos públicos -AFIP, SSS- a las obras sociales).
- **Edad de los beneficiarios:** agrupada en rangos de 5 años y por género, es decir, se cuenta con el número de afiliados ordenados en columnas del tipo: F_0; F_01a04; F_05a09;

²⁹ La diferencia significativa con respecto a la población real de Argentina se da como consecuencia de las diferencias en cortes temporales de las bases de las que se extrae la información demográfica y, por otro lado, la tasa de doble cobertura ya mencionada anteriormente en este trabajo.

F_10a14; F_15a19; F_20a24; F_25a29; F_30a34; F_35a39; F_40a44; F_45a49; F_50a54; F_55a59; F_60a64; F_65a69; F_70a74; F_75a79; F_>=80; M_0; M_01a04; M_05a09; M_10a14; M_15a19; M_20a24; M_25a29; M_30a34; M_35a39; M_40a44; M_45a49; M_50a54; M_55a59; M_60a64; M_65a69; M_70a74; M_75a79; M_>=80

- **Cantidad de centros propios de Salud:** Para cada financiador de salud se tiene la cantidad de centros propios que, según la base pública de REFES, dependen de esos organismos

Tabla 1. Muestra de la base con información para los distintos financiadores de salud de Argentina³⁰

Nombre	Ecosistema	Cantidad total de afiliados	Region Principal	Edad promedio afiliados	Ingresos mensuales totales	Ingresos por afiliado	F_0	F_01a04	F_05a09	F_10a14	F_15a19	F_20a24
PAMI_INSSJP	PAMI	5181414	ALL	72.13	40631445938	7841.77	4	673	4430	10584	12714	2002
PUBLICO_PBA	Público	6789349	Buenos Aires	27.18	22580691033	3325.90	50077	215576	369856	344391	306460	354406
OSDE	Privado-Sindical	2120518	ALL	36.22	17250333333	8134.96	2269	49031	75687	72422	65711	66393
PUBL_SALUD_NACION	Público	733398	ALL	27.52	13492963333	18397.87	5880	24374	37722	35870	32829	40536
PUBLICO_CABA	Público	1041185	CABA	29.53	11136135721	10695.64	5209	30983	47657	45598	39603	57591
IOMA	Provinciales	2028254	Buenos Aires	42.67	8762743971	4320.34	3166	6373	49997	59235	101101	74718
SWISS_MEDICAL	Privado	976038	AMBA	35.36	8032083333	8229.27	1975	19444	30169	28933	28117	29282
PUBLICO_SANTA_FE	Público	1676340	Santa Fe	28.83	6458217566	3852.57	14191	56960	79687	73119	70705	91285
GALENO_ARGENTINA	Privado	561259	AMBA	36.20	5214000000	9289.83	1274	11872	19796	19899	19066	17487
PUBLICO_CORDOBA	Público	1752946	Cordoba	28.73	4564957203	2604.16	15980	59331	79200	79529	76356	90099
INCLUIR_EXPROFE	Público	983291	ALL	32.85	4259172500	4331.55	0	0	66494	64022	32653	40118
SANCOR	Privado-Sindical	745435	Centro	31.03	4026094435	5401.00	2931	20048	26773	22350	18854	25339
OSECAC	Sindicales	1605251	ALL	34.90	4024857673	2507.31	3255	24282	46573	53895	58905	62383
PUBLICO_TUCUMAN	Público	783417	Tucuman	26.65	3944341840	5034.79	7515	32328	38498	33386	32927	46735
PUBLICO_SALTA	Público	715609	Salta	26.14	3481147066	4864.59	6799	24898	36492	33819	32658	45434
PUBLICO_NEUQUEN	Público	333583	Neuquen	31.52	3467888569	10395.88	906	11228	13315	15008	13061	16854

INFORMACIÓN ESTADÍSTICA DE INCIDENCIA O PREVALENCIA ESTIMADA PARA PATOLOGÍAS SELECCIONADAS

Con el objetivo de generar un análisis aplicado a un negocio de medicamentos en particular, y de poder generar un análisis de potencial relevancia para los distintos financiadores que podrían requerir medicamentos provistos por la compañía farmacéutica en análisis, se ha trabajado con una estimación, para cada financiador, utilizando datos del *Global Cancer Observatory* (GLOBOCAN, 2020), organismo dependiente de la ONU y del *Institute for Health Metrics and Evaluation* (IHME, 2019), y de relevamientos de bibliografía para ciertas patologías³¹, con el fin de estimar la incidencia ajustada por edad y género de las siguientes patologías estratégicas para el laboratorio farmacéutico:

- HER2 (cáncer de mama con ese subtipo)
- Cáncer de Hígado (en particular para una primera línea en cierto escenario)
- Hemofilia
- Atrofia muscular espinal - AME-
- Cáncer de vejiga (en una segunda línea de tratamiento)
- Cáncer de pulmón
- Esclerosis Múltiple
- Linfoma de *Non Hodgkin*

INFORMACIÓN DE LA VENTA DE LA COMPAÑÍA

- Se cuenta con información de la venta de la compañía desde febrero de 2017 hasta el mes de octubre de 2022 para los productos que estaban disponibles en cada momento del

³⁰ Los ingresos mensuales totales y los ingresos por afiliado son en pesos estimados para el año 2021.

³¹ Remitirse a Anexo 2 para más detalles sobre la estimación realizada

tiempo. No todos los productos estaban lanzados al mercado al momento del comienzo de la información

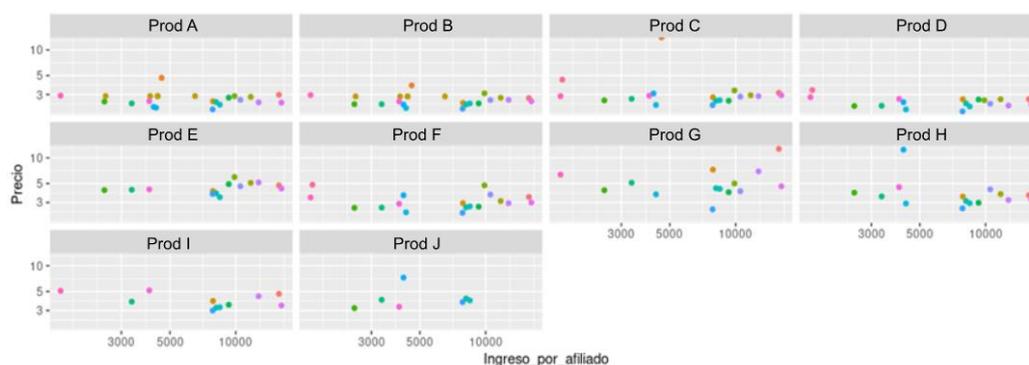
- Se utilizan **10 productos para hacer el análisis de los financiadores**. Se incluyen productos que se describen a continuación:
 - Se seleccionan medicamentos para oncología (para cáncer de pulmón, piel y mama, hematológicos, entre otros), tratamientos neurológicos y otros destinados a patologías de baja incidencia
 - Los productos se encuentran en fases distintas en términos de ciclo de vida: 4 se encuentran en fase de sostenibilidad del producto, 4 en etapa de desinversión y 2 en fase de expansión y crecimiento del producto.
 - 2 productos de los que se incluyen corresponden a patologías crónicas con uso permanente de los pacientes, 4 corresponde a tratamientos que tienen un uso temporal con un fin explícito (X cantidad de dosis totales por paciente) y 4 que son de uso temporal sin un final explícito (en particular corresponden a tratamientos para oncología donde la medicación es a veces hasta la progresión de la enfermedad o muerte y, por lo tanto, el periodo de uso del producto puede variar para cada paciente)
 - Cuatro de los productos del trabajo “no” son el estándar de tratamiento -selección recomendada por los médicos a priori- para la indicación que cubren y los 6 restantes sí lo son.
 - Se incluyen productos ya establecidos en la sociedad (1), estratégicos (3) y los productos que fueron muy relevantes para la compañía históricamente y que hoy tienen varios biosimilares, conocidas como copias (3).

II. Exploración de los datos de la venta

Inicialmente en este trabajo, se pretende entender cuáles variables definen en la actualidad el precio neto pagado por los distintos financiadores y se pretende entender si existe contemplación de una visión de equidad en la determinación de los precios.

En la figura 13, se busca entender la relación entre los ingresos promedios por afiliado y el precio al que efectivamente acceden a los tratamientos. Se observa que si bien existe gran diversidad en cuanto a los ingresos (por lo que posiblemente la capacidad de pago sea diferente para cada uno), el precio al que efectivamente acceden a los productos no responde a dicha variable.

Figura 13. Análisis de Precio pagado y de ingreso promedio por afiliado para el financiador

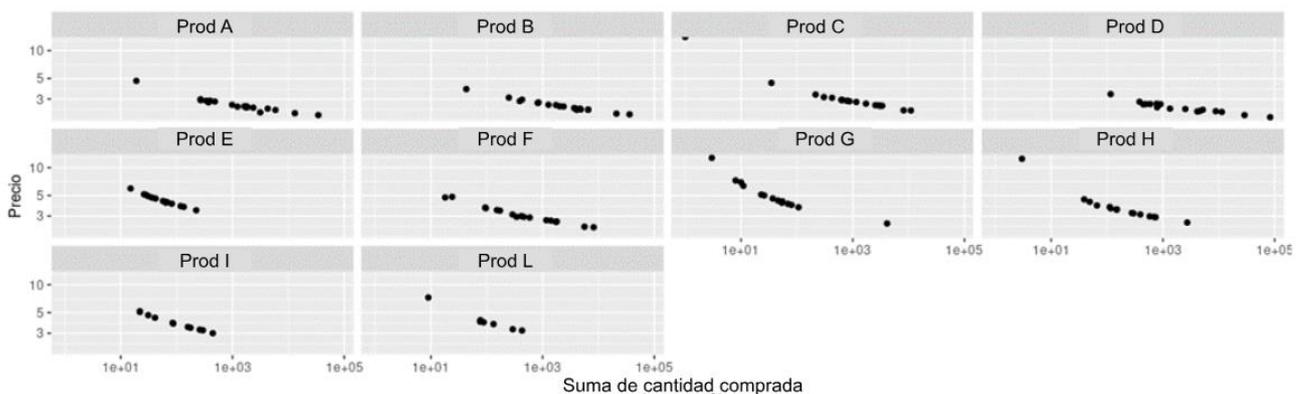


Fuente: Elaboración propia a partir de datos confidenciales

Por otro lado, en la figura 14, se busca representar la relación entre el precio pagado y la cantidad comprada por cada financiador. En este caso, se respeta la relación decreciente entre cantidad y precio pagado para los distintos productos. Es relevante observar que las pendientes de estas curvas decrecientes son bastante distintas entre sí para los distintos productos bajo análisis por lo que se podría intuir que es evidente que otras variables definen la forma en que se deberían establecer los escalones de descuentos. A priori, a partir del análisis gráfico, la cantidad de afiliados evidentemente es una variable que influye en la definición del precio neto de compra para cada financiador en particular. La pendiente de las curvas es una medida de la elasticidad precio específica para cada producto donde se refleja efectivamente la interacción de las características que definen el nivel de precios y la flexibilidad que se debe tener en cuanto a los descuentos para garantizar la comercialización de los productos que están bajo análisis.

Como se puede observar en términos generales, aquellas curvas que son más aplanadas, demuestran una situación de elasticidad grande por parte de los financiadores; si se realizaran pequeños cambios en los precios, la demanda reacciona fuertemente. Por otro lado, para ciertos productos se muestra que la demanda es muy inelástica, cuando la pendiente de la curva que se ve entre los puntos, es más empinada.

Figura 14. Análisis de Precio pagado y cantidad comprada por financiador directo



Fuente: Elaboración propia a partir de datos confidenciales

Cuando se intenta entender cuál es el racional que define los precios que efectivamente pagan los distintos financiadores se analiza, para aquellos que son clientes directos, cómo se comporta la variable de precio con respecto a algunos de las variables utilizadas en el trabajo. En la figura 15, se representa mediante un mapa de correlaciones el precio neto pagado por tratamientos para cáncer de mama, en función de distintos elementos descriptivos de los financiadores. Si bien se realizó en el trabajo un análisis de correlación con más variables, se ha decidido reflejar solo una parte de ellas para demostrar su falta de correlación con el precio. Tal como surge del análisis presentado más adelante en esta misma sección, cuando se analiza en detalle las variables demográficas, de estructura y de ingreso, se ha identificado alta correlación entre algunas de ellas y, por lo tanto, a los fines de que la figura resulte más ilustrativa, se ha hecho una selección limitada de variables en el análisis.

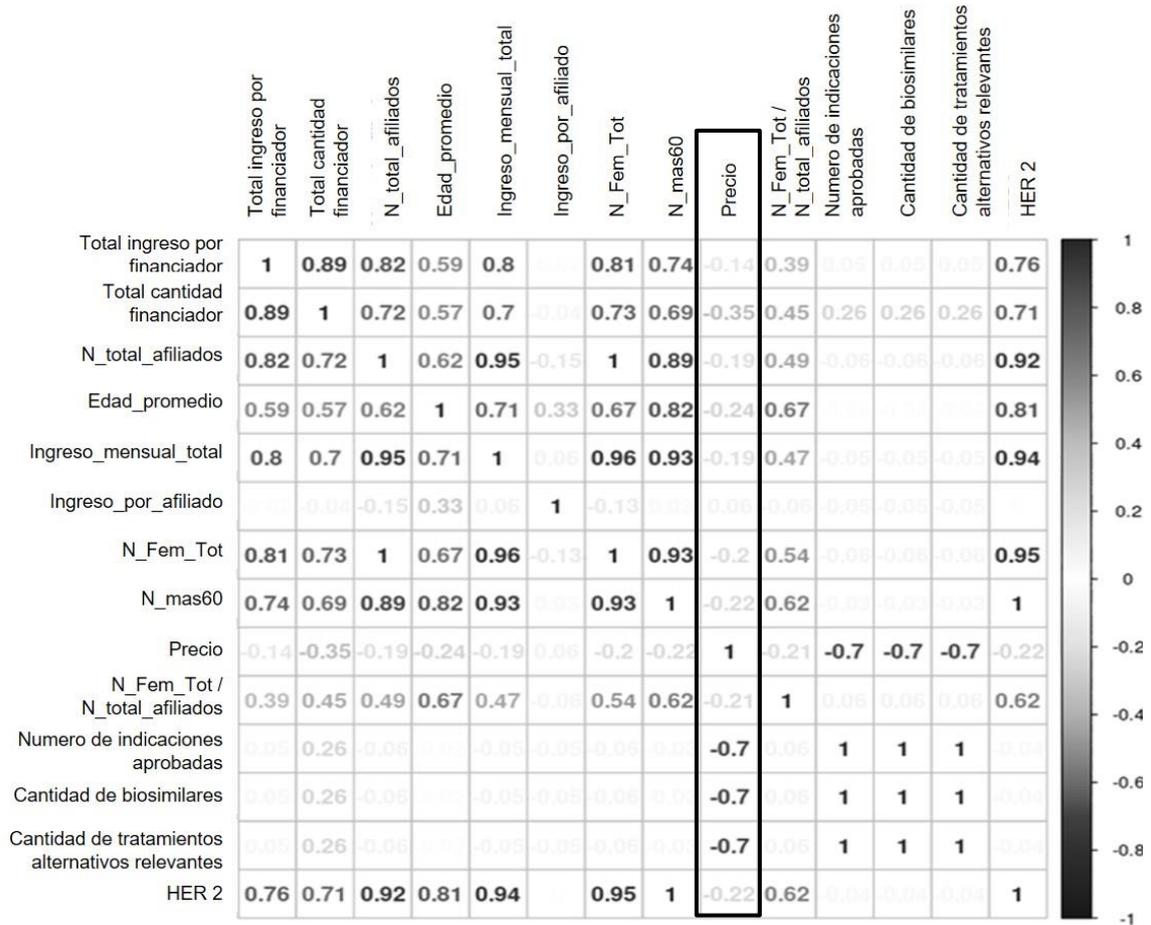
En este caso resulta llamativo que no exista relación entre la incidencia potencial de la patología y el precio efectivamente pagado. Asimismo, se observa que el precio es muy dependiente de la cantidad de copias, con lo que se demuestra que efectivamente el laboratorio recupera la rentabilidad de su inversión cuando aún conserva la exclusividad de los productos y que luego, en función de la cantidad

de copias, se disminuye en mayor o menor medida el precio neto pagado. Las variables cantidad de copias, cantidad de competidores y número de indicaciones están muy correlacionadas entre sí y hasta se podría argumentar que su comportamiento es lineal. Esto sucede lógicamente ya que mientras más copias haya, hay mayor competencia y, por otro lado, cuando una molécula tiene más tiempo en el mercado, va incorporando más indicaciones y confianza en su efectividad. En consecuencia, es razonable suponer que el tiempo en el mercado es la variable que principalmente impulsa a las otras dos que se mencionan. Otro hallazgo de relevancia es que se evidencia que, en la actualidad, la cantidad comprada no responde directamente a los ingresos disponibles para cada financiador de la salud. Por lo tanto, resulta evidente que no hay criterios de equidad en la asignación actual de los descuentos comerciales para los distintos financiadores.

Hay múltiples variables que también tienen gran correlación entre sí. Por ejemplo, se puede afirmar que cuando hay una mayor cantidad de beneficiarios, en general existe correlación positiva con la edad que en promedio tienen los beneficiarios. Asimismo, la cantidad de mujeres en general tiene una correlación positiva con la cantidad de afiliados y la edad promedio, lo cual refleja que la expectativa de vida de las mujeres es mayor.

Por otro lado, como es de suponer, los ingresos totales en general son superiores cuanta más población se cubra (sin embargo, esto no sucede con los ingresos promedios por afiliado).

Figura 15: Matriz de correlaciones de información de precios y ventas de la compañía en análisis en la actualidad en para el tratamiento del cáncer de mama³²



Fuente: Elaboración propia

III. Técnicas empleadas en el análisis de datos y selección de variables

La ingeniería de las variables es una etapa muy importante dentro del proceso de creación de un modelo de *machine learning*. Consiste en la comprensión, selección, transformación y creación de variables o características que sean útiles en la construcción de un modelo predictivo. Esta tarea es fundamental ya que los algoritmos requieren que los datos sigan patrones precisos y relevantes para el análisis. Como indican Johnson & Kuhn (2020) “estas tareas son esenciales para el éxito del

³² Total ingreso financiador (revela el ingreso generado por ese financiador, es decir el precio por la cantidad vendida); Total cantidad financiador (refleja la cantidad comercializada); número de indicaciones aprobadas (es un número indicativo de las indicaciones del medicamento, es decir, para cuántos estadios de la enfermedad o escenarios está aprobado el uso del medicamento); Cantidad de biosimilares (refleja la cantidad de copias directas del producto en el mercado); Cantidad de tratamientos alternativos relevantes (refleja la cantidad de alternativas de alto precio disponibles para el tratamiento de esa patología); Precio (refleja el precio neto del producto para el financiador en análisis).

El detalle sobre las variables: N total afiliados, Edad_promedio, Ingreso_mensual_total, Ingreso_por_afiliado, N_Fem_Tot, N_mas60, N_Fem_Tot/N_total_afiliados, HER 2 se encuentra en el anexo 2 de este trabajo. Reflejan datos demográficos, económicos y de incidencia estimada para el financiador.

aprendizaje automático [...]. La selección de características adecuadas puede hacer que incluso los algoritmos de aprendizaje más simples funcionen de manera efectiva".

En tal sentido, si bien más adelante en la sección de ingeniería de datos se detalla en particular en qué modelos y con qué variables se implementan estas técnicas, se cree relevante hacer una breve descripción sobre los objetivos y la metodología que asumen estos métodos. Se utiliza el entorno y lenguaje de programación de R, a través de *RStudio* como editor de datos, para desarrollar los distintos análisis. Las técnicas que son empleadas, con el objetivo de realizar análisis exploratorio de los datos (EDA, por las siglas en inglés de *Exploratory Data Analysis*) son las que se detallan a continuación:

- **Análisis gráfico de la distribución de las distintas variables:** El primer paso fundamental dentro de las tareas de la ingeniería de datos consiste en entender cómo se comportan para las distintas observaciones, las distintas variables que se tienen disponibles. Con esto se puede tener una primera visión sobre el tipo de comportamiento de los datos que se observan, las posibles distribuciones que se analizan, entender si es relevante encarar ciertos análisis adicionales, etcétera. Asimismo, tiene sentido entender ciertas métricas que se observan para cada variable, buscando entender los promedios, los valores máximos y mínimos, la distribución o desvío de los datos, etcétera. En el caso de este trabajo se pretende utilizar los siguientes comandos de R: *describe*, *ggplot* y otras, con el fin de ejecutar esta primera tarea.
- **Detección y tratamiento de valores atípicos en el *dataset*:** Consiste en identificar anomalías en una base de datos. Las anomalías pueden suceder como consecuencia de una captura incorrecta de los datos (podría pasar como consecuencia de tomar distintas unidades de medida para una misma variable erróneamente, por ejemplo) o en casos donde hay muy pocas observaciones con un comportamiento muy diferenciado al resto (en este caso podría pasar con ciertos financiadores que cubren una cantidad muy significativa de afiliados en comparación con la mayoría). En ese sentido, el análisis de los *outliers* es una etapa fundamental por el efecto que generan esas observaciones sobre la variabilidad de los datos.

Para ejecutar esta tarea se pretende utilizar el análisis de rangos intercuartiles. Consiste en determinar cuartiles con la distribución de una variable y luego, calcular la diferencia entre el primer (Q1) y el tercer (Q3) cuartil. Los cuartiles dividen los datos en cuatro partes iguales, de manera que Q1 se tiene el 25% de los datos y debajo de Q3 se encuentra el 75% de los datos. En consecuencia, es una medida de la dispersión de los datos en el rango medio.

$$IQR = Q_3 - Q_1 \quad [1]$$

Luego, con ese dato, y utilizando un factor conocido como coeficiente de distancia intercuartílica (CDI) que es definido como parámetro (en general se usa 1.5), se consideran valores atípicos a todos los datos que se encuentran debajo del valor $Q_1 - \frac{CDI}{IQR}$ y aquellas que están por encima de $Q_3 + \frac{CDI}{IQR}$

- **Normalización de variables:** La normalización es un proceso fundamental para la implementación de ciertos algoritmos. Se refiere al proceso de escalar los datos de entrada con el fin de aplicar una distribución uniforme. Permite ajustar los datos a una escala uniforme y con eso evitar que el peso de cierta variable, como consecuencia quizás de tener unidades de medida diferentes, genere sesgo en el análisis.

Para implementarlo se utilizará la función *scale()* en R, que calcula la media y la desviación estándar y luego escala los valores al restarle la media y dividirlo por la desviación estándar.

- **Log-normalización de variables:** esta técnica es útil cuando se identifica en los datos una distribución sesgada y donde la media aumenta con la media. Al aplicar esta técnica, los datos se transforman a una distribución más simétrica y con menor varianza. Con ello, se permite mejorar la precisión y la estabilidad del análisis que se está realizando. Luego, esta técnica se combina con la normalización con el fin de ajustar la escala de los datos. Según se ha descrito (Montgomery & Peck, 2018), “la transformación logarítmica puede mejorar la simetría de una distribución y reducir la varianza de los datos. Esto puede ser útil para mejorar la interpretación de los resultados y la aplicabilidad de modelos estadísticos. Además, la transformación logarítmica puede ayudar a linealizar una relación entre dos variables, lo que facilita su análisis e interpretación”. El efecto sobre la distribución de los datos resulta en mayor simetría de los datos y homocedasticidad

Se implementa la siguiente función:

$$X_{lognorm} = \log(x - \min(x) + 1) \quad [21]$$

- **Análisis de correlación entre variables:** En el proceso de selección de variables para ejecutar un modelo de clustering es fundamental retener sólo las variables relevantes. Con este fin, es necesario identificar cuáles son las variables “relevantes”, para lo que se establece algún criterio de cálculo que permita medir la correlación cruzada entre las distintas variables e implementar alguna técnica para trabajar con esas variables. De esta manera, se consigue hacer una selección de variables que reflejen correctamente el comportamiento de las observaciones. Se utiliza el coeficiente de correlación de Pearson que varía entre -1 y 1 donde -1 refleja una relación inversa perfecta y 1 indica una relación perfecta directa (es decir que cuando una variable aumenta, la otra aumenta en igual proporción). Un valor de 0 indica que no hay relación lineal entre dos variables. En el trabajo, se busca que las variables que sean consideradas, sean las que expresan características que separan a los datos y que permiten generar grupos de observaciones con características equivalentes. La ecuación que se aplica para ejecutar la tarea consiste en:

$$r = \frac{\Sigma(x-\underline{x}) * (y-\underline{y})}{\sqrt{\Sigma(x-\underline{x})^2 * \Sigma(y-\underline{y})^2}} \text{ donde } x \text{ e } y \text{ son dos variables continuas y } \underline{x} \text{ e } \underline{y} \text{ son sus medias respectivas}$$

- **Implementación de herramientas de filtering:** Esta técnica es utilizada para seleccionar o filtrar observaciones de un conjunto de datos en función de una o varias condiciones especificadas. En síntesis, permite seleccionar un subconjunto de datos de interés a partir de un conjunto de datos más grande. Se propone fijar algún umbral a partir del cual las variables con correlación superior sufran algún tipo de transformación (consolidar varias variables, eliminarlas, etc.). En el trabajo en particular se pretende utilizar esta herramienta para identificar las variables con alta correlación y luego se les realiza un análisis en particular.
- **Análisis mediante técnicas de componentes principales:** consiste en hacer un análisis de múltiples variables donde lo que se busca es reducir la dimensionalidad de un conjunto de datos, pero manteniendo la mayor cantidad posible de información. Es relevante decir que este método sólo permite variables numéricas en su construcción. El objetivo de este análisis, en otras palabras, es encontrar las combinaciones lineales de las variables que maximizan la varianza de los datos y con eso se reduce la cantidad de variables necesarias para describir

los datos. Las nuevas variables, que son conocidas como componentes principales (PC), son ortogonales entre sí y se ordenan según la cantidad de varianza que explican en los datos. La fórmula que se utiliza es:

$$PC_1 = a_{11} * x_1 + a_{21} * x_2 + \dots + a_{p1} * x_p \quad [3]$$

$$PC_2 = a_{12} * x_1 + a_{22} * x_2 + \dots + a_{p2} * x_p$$

...

$$PC_p = a_{1p} * x_1 + a_{2p} * x_2 + \dots + a_{pp} * x_p$$

donde: PC_1, PC_2, \dots, PC_p son los p componentes principales, x_1, x_2, \dots, x_p son las variables originales y $a_{11}, a_{21}, \dots, a_{p1}, a_{12}, \dots, a_{pp}$ son los coeficientes que definen las combinaciones lineales de las variables originales para cada componente principal.

Con el análisis de los componentes, se facilita la decisión sobre el posible comportamiento vectorial de ciertas variables y así definir cuántas de las dimensiones identificadas tendría sentido incluir en el trabajo.

- **Análisis mediante herramienta FAMD (*Factor Analysis of Mixed Data*):** Una de las limitaciones del análisis de componentes principales es la imposibilidad de incluir variables categóricas en el análisis. Para tal fin, tal como se explica en el trabajo de Chough y de Leon (2013), la función FAMD (*Factor Analysis of Mixed Data*) es una herramienta útil para realizar clustering en conjuntos de datos que contienen tanto variables categóricas como continuas. En el caso particular de este trabajo, tan solo se partió con dos variables categóricas, pero se cree que podría tener sentido incluirlas en un posible análisis. Los autores ya mencionados explican que esta función hace un análisis de factores en los datos, lo que permite una mejor representación de la estructura subyacente de los datos mixtos. Lo que hace es combinar las técnicas de CFA -*confirmatory factor analysis* -, una técnica de modelado estadístico que se utiliza para evaluar la adecuación de un modelo de factores a los datos para evaluar la estructura de datos mixtos y definir si existe una estructura de factores con AC - análisis de correspondencia -, que es una técnica estadística que se utiliza para analizar datos categóricos y visualizar la relación entre las categorías. En lugar de considerar solo la distancia euclidiana entre los puntos de datos, la FAMD utiliza una distancia basada en los componentes principales del análisis de factores. La distancia resultante es más apropiada para datos mixtos y puede mejorar la calidad del *clustering* (Chough & de Leon, 2013). En esta técnica se implementa el siguiente análisis matricial:

$$X = AWF + E \quad [4]$$

dónde: X es una matriz de datos observados, A es una matriz de pesos de las variables categóricas, W es una matriz de pesos de las variables numéricas, F es una matriz de factores latentes y E es una matriz de errores de medición. En este modelo, en primer lugar, a través de la matriz de pesos de las variables categóricas (A) y la matriz de pesos de las variables numéricas (W), se estandarizan las variables. Luego, a través de la matriz de factores latentes (F), se representan los factores subyacentes que explican la variabilidad en los datos. Por último, mediante la matriz de errores de medición (E), se representa el ruido o la variabilidad aleatoria en los datos.

IV. Análisis de variables disponibles y transformación de los datos

Variables iniciales:

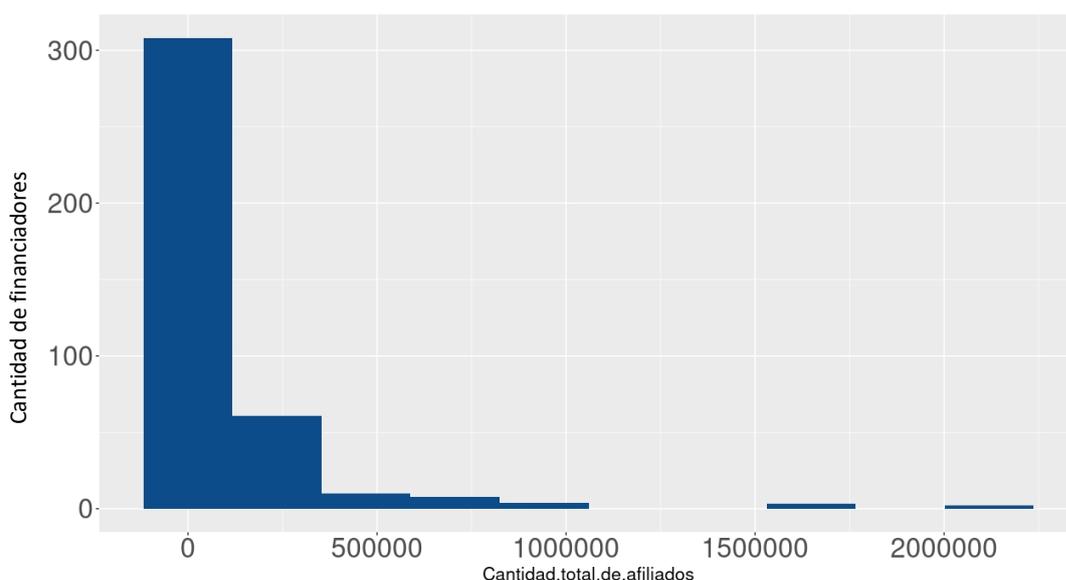
Partimos inicialmente con 68 variables con las que se fueron realizando diferentes correcciones y análisis específicos que resultaron en un diversos *datasets* que luego pretenden ser analizados en la etapa de análisis de resultados.

Se elimina la variable “subecosistema” ya que solo da información relevante para el responsable de ventas para cada financiador.

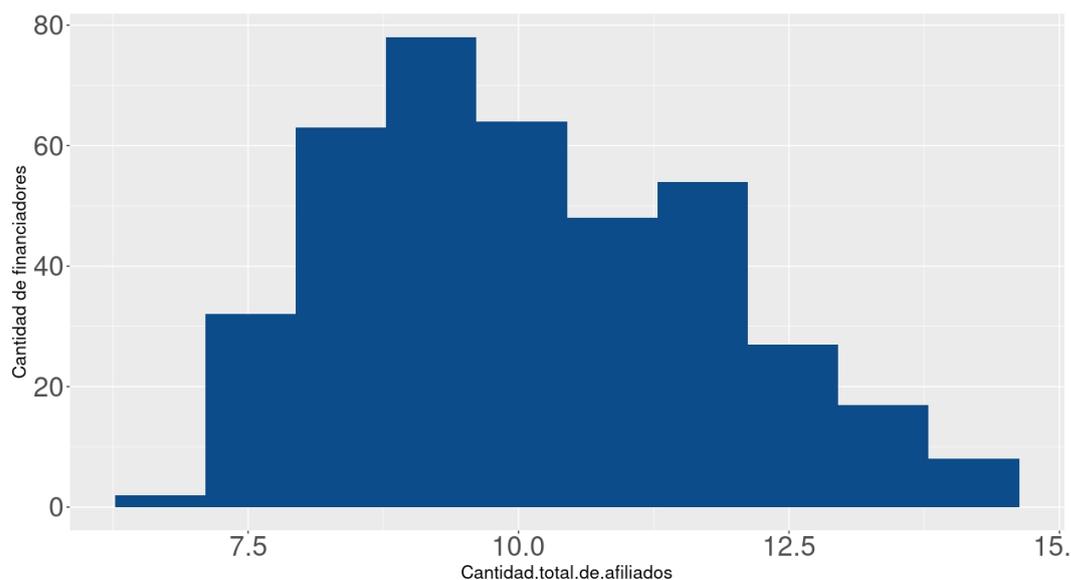
a. N total de afiliados:

Haciendo un análisis gráfico de esta variable, resulta evidente que existen algunos financiadores cuya cantidad total de afiliados es significativamente superior al resto. Estos son: PAMI con más de 5 millones de afiliados y el sector público de la Provincia de Buenos Aires, que brinda cobertura a casi 7 millones de personas. Si se eliminan estas dos observaciones del gráfico, se puede ver que, de todos modos, existe una concentración significativa de financiadores de pocos afiliados y que hay muy pocos que concentran grandes cantidades de beneficiarios. En particular, tal como se indica en la figura 16, más de 300 financiadores de salud brindan cobertura a una cantidad baja de beneficiarios (menos de 50000 beneficiarios). Por otro lado, se observa un grupo de alrededor de 50 financiadores con cobertura también de alrededor de 100000 beneficiarios. Por último, más hacia la derecha de la figura, cuando se refleja cobertura de una cantidad mayor de beneficiarios, se observa que la cantidad de financiadores con estas características es mucho menor. En consecuencia, se podría argumentar que la financiación de los servicios de salud y medicamentos está dispersa en su mayoría en financiadores muy chicos por su tamaño y que existen algunos pocos que cubren, y por lo tanto también negocian, por una cantidad más alta de beneficiarios.

Figura 16: Representación de la cantidad de afiliados por financiador y su transformación logarítmica



Fuente: Elaboración propia



Fuente: Elaboración propia

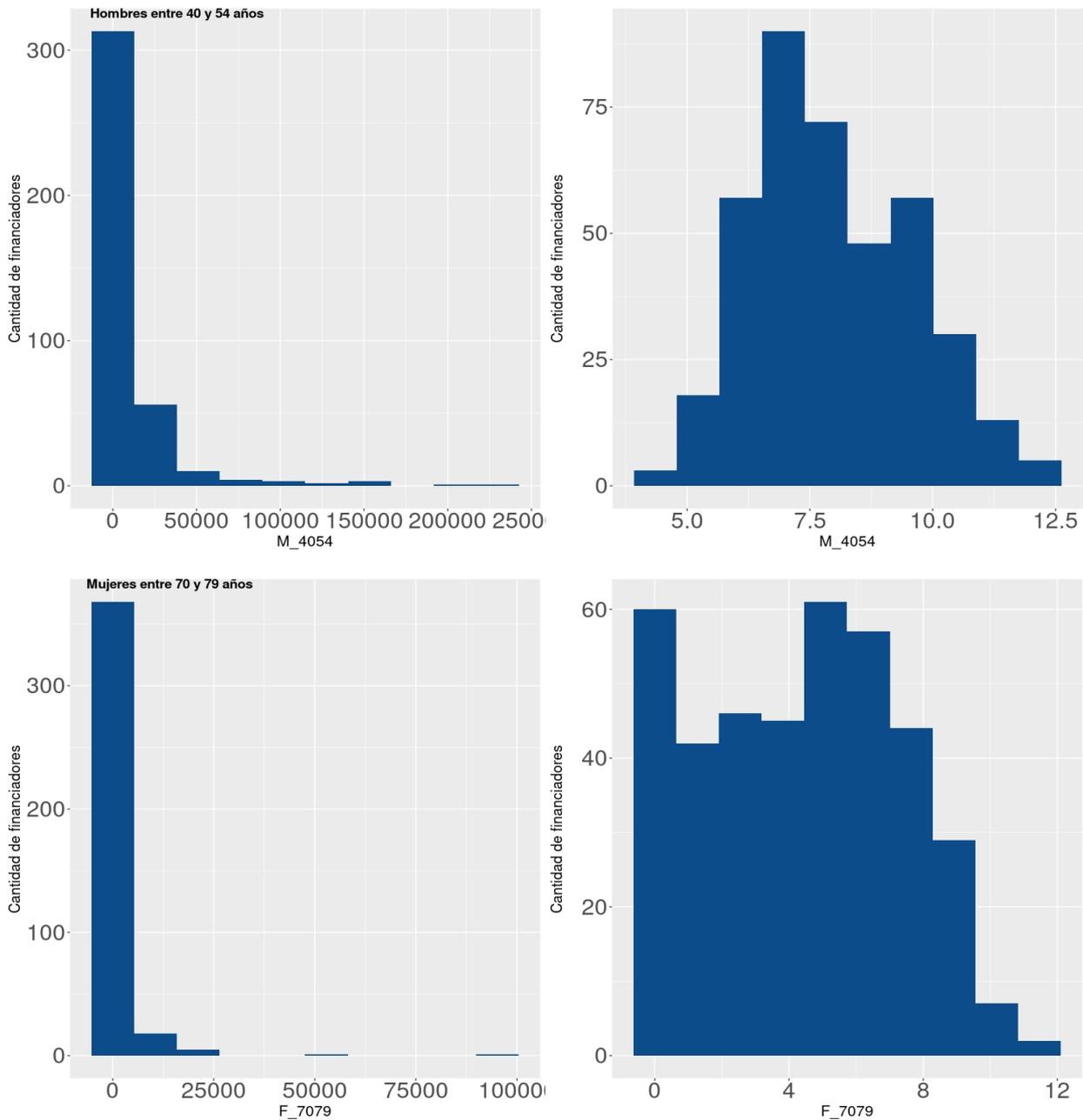
Se aplica una transformación mediante la función logaritmo a algunas de las variables de dimensión del tamaño del financiador. El criterio es considerar que dicha transformación genera varios efectos importantes en cuanto a la distribución y la relación con otras variables. En particular, el objetivo de esta transformación se enmarca dentro de la propuesta de la escalera de poderes de Tukey (1977), donde se busca transformar una distribución sesgada para que se vuelva normal o casi normal. Además, se consigue reducir la variabilidad del error en la variable (heterocedasticidad).

Como se observa en la figura 17, se representa que al aplicar la *log normalización* a las variables de distribución de cantidad de población por rango etario, se corrige la enorme variabilidad que tienen ciertos financiadores que concentran grandes poblaciones y, por lo tanto, se transforma el sesgo en la distribución entre financiadores. Con esta transformación de los datos se implica, en consecuencia, *una reducción del sesgo de esta variable*.

En particular, se presenta en la figura 17, la distribución que tienen los distintos financiadores en cuanto a la cobertura que brindan a hombres de entre 40 y 54 años. Tal como resulta evidente a partir de un primer análisis visual del comportamiento de la variable, la concentración de financiadores con poca población (y, por lo tanto, consecuente baja cantidad de beneficiarios con las características de edad indicadas) respeta a priori la misma situación que cuando se analiza el N total de afiliados. Es decir, hay pocos financiadores que concentran grandes grupos de población y, en general, la mayoría de financiadores da cobertura a grupos chicos de beneficiarios.

En consecuencia, al aplicar la transformación log normal, se observa que la distribución de este grupo población, muy relevante en cuanto al aporte que realizan en su contribución económica a los financiadores, por tratarse de población en edad activa laboral, sigue una distribución más normal. En conclusión, se puede observar que hay un sesgo hacia la derecha en cantidad de financiadores con este grupo, aunque la distribución es bastante normal.

Figura 17: Representación gráfica de transformación logarítmica en funciones que representan población total por rango etario



Fuente: Elaboración propia

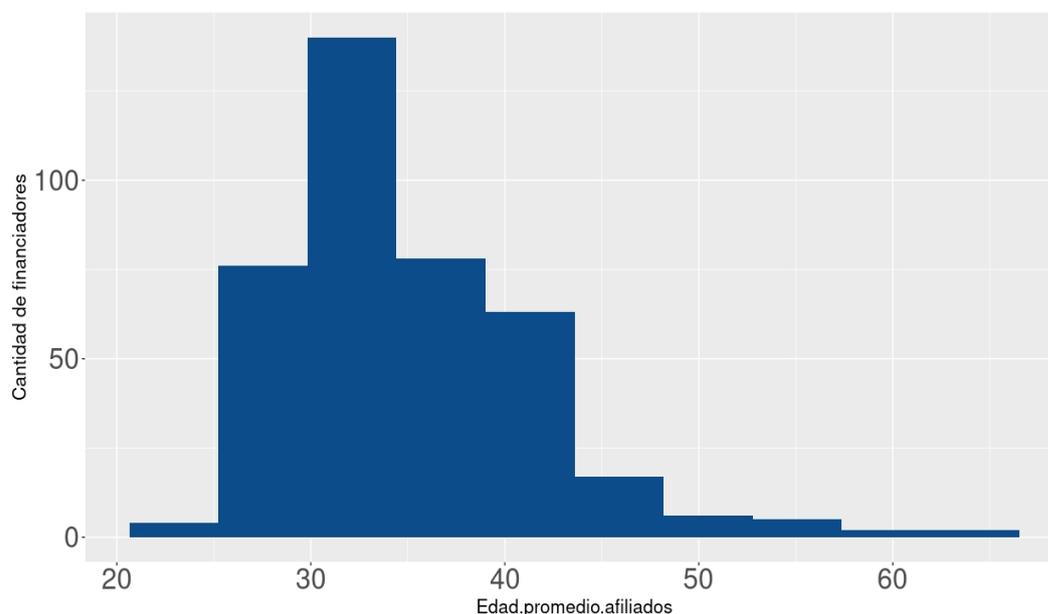
b. Edad promedio de los afiliados

Cuando se analiza la variable que refleja la edad promedio de los afiliados para los distintos financiadores, se observa que existe un leve sesgo a la izquierda donde pocos financiadores concentran a los grupos poblacionales de mayor edad. Esto sucede, en el caso argentino, como consecuencia de la existencia de PAMI donde un solo financiador de un tamaño muy grande, tiene en su padrón a casi el total de los jubilados y pensionados del país. Por lo tanto, existe lógicamente sesgo en la representación de ciertas patologías en esos grupos que concentran poblaciones en promedio de mayor edad. En este caso, solamente se decidió normalizar la variable. Se representa

en la figura 18 el comportamiento de la variable con respecto a la cantidad de financiadores para cada edad promedio.

La mayoría de los financiadores tienen una edad promedio en sus beneficiarios alrededor de los 30 años. Este dato es reflejo de la población económicamente activa y la concentración de la fuerza laboral, distribuida en los financiadores de la seguridad social, que son los más representativos en cuanto a cantidad de financiadores que brindan sus servicios. Por otro lado, podría interpretarse, a partir de suposiciones y de los datos epidemiológicos incluidos en el análisis (mayor detalle en otra parte de esta sección), que el riesgo epidemiológico de esa población es menor. Como mínimo se observa que el promedio de edad de los beneficiarios es superior a los 20 años. Luego, la gran mayoría de financiadores tienen en promedio entre 25, aproximadamente, y 45 años. Por último, algunos financiadores tienen un promedio de edad más alto. Sin embargo, la cantidad de financiadores con estas características es mucho menor.

Figura 18: Análisis de la edad promedio en financiadores argentinos



Fuente: Elaboración propia

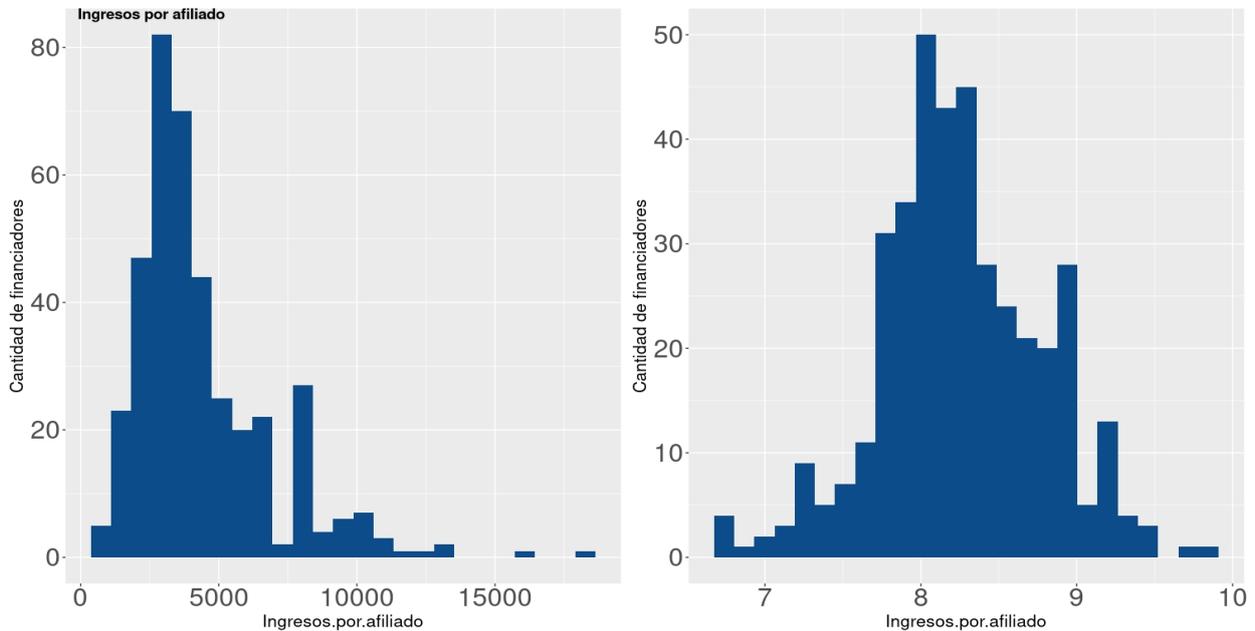
c. Ingresos mensuales promedio por afiliado

Se pretende trabajar con y sin la aplicación de logaritmos para la función de ingresos por afiliados, buscando corregir la presencia de sesgos en los datos y la heterocedestacidad, con gran variabilidad hacia la derecha; en el análisis visual de los datos, se observan algunos financiadores con ingresos por afiliado excepcionalmente altos. Por lo tanto, se busca corregir el sesgo de la transformación logarítmica. Se observa en la distribución original que hay muchos grupos con ingresos bajos en promedio para sus afiliados. Por otro lado, se observa concentración en cierto monto que equivale al aporte promedio de cuotas en empresas de medicina prepaga³³ y, por último, pocos financiadores que perciben ingresos más altos por parte de sus beneficiarios. Se representa en la figura 19 el

³³ Tal como se presenta en el anexo 1, la construcción de los datos es posible que genere concentración en cierto rango de ingresos por afiliado como consecuencia de la fuente disponible de datos que alimenta a ese análisis.

comportamiento de la variable ingreso y su comportamiento post aplicación de *log normalización* de la variable.

Figura 19: Representación gráfica de la variable ‘ingreso por afiliado’ para los distintos financiadores y su comportamiento post aplicación de log normalización de la variable

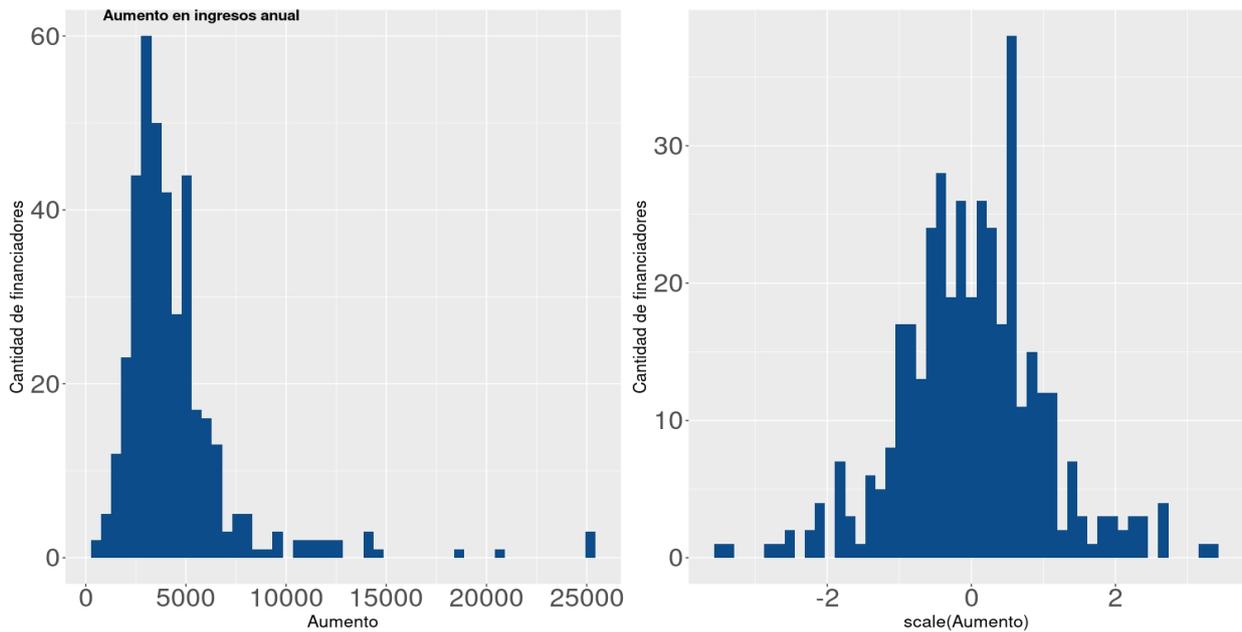


Fuente: Elaboración propia

Luego, en función de la tendencia de la evolución de los ingresos de los últimos 12 meses, se calcula un valor denominado “Aumento”, que refleja el ingreso proyectado por afiliado a un año del cálculo anterior. Con esto, se busca reflejar la tendencia de evolución de ingresos promedios por afiliado que estaba experimentando el financiador. Nuevamente se muestra la existencia de un grupo concentrado de financiadores con ingresos bajos y un subgrupo de financiadores con ingresos proyectados que resultan más altos. La concentración de la mayoría de los financiadores se da en un nivel bajo de ingresos (menos de 5000), luego, se concentra un grupo significativo de financiadores (en torno a los 8000); por último, hay un subgrupo menor de financiadores que tienen ingresos que son en promedio más altos.

Asimismo, se experimenta aplicando logaritmo a la función para controlar el sesgo en los datos iniciales, a partir de la técnica ya presentada con anterioridad (Tukey, 1977). Se presenta dicha información en la figura 20. Se observa que cuando se aplica la lognormalización, los ingresos promedios mensuales por el aumento en ingresos acumulado en un año reducen su sesgo, y se estandariza la variabilidad entre los distintos financiadores. Dicho resultado se puede observar por la distribución representada en la misma figura.

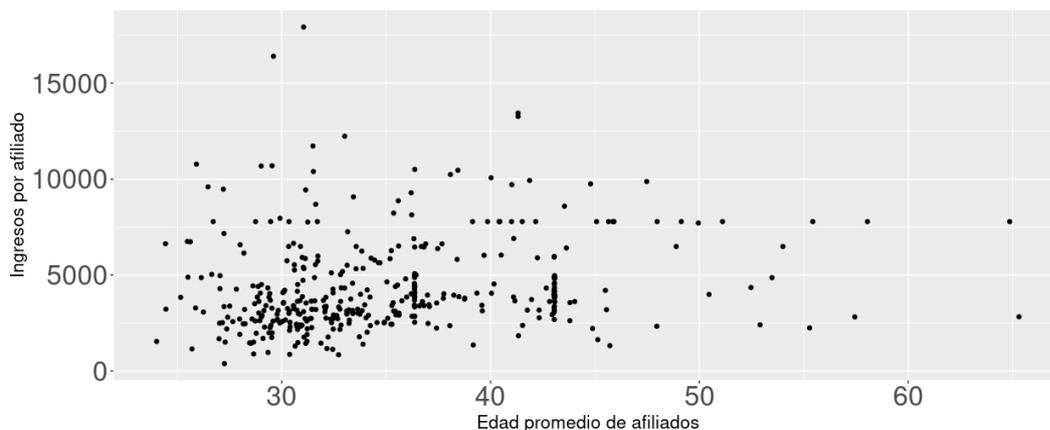
Figura 20: Representación gráfica de la variable Aumento y su comportamiento post aplicación de log normalización de la variable



Fuente: Elaboración propia

Un análisis que se explora para entender el nivel de correlación, es el de edad promedio de los afiliados y los ingresos promedio por afiliado. Como se puede observar en la figura 21, si bien existe claramente un grupo más concentrado con un rango etario más joven, existe diversidad significativa de ingresos para todos los financiadores, independientemente de las edades promedio de sus beneficiarios. Cuando se observa a aquellos financiadores con edades promedio más bajas, se observa que el promedio de ingreso por afiliado es menor. Sin embargo, alrededor de los 30 años se observa a aquellos financiadores que representan el mayor ingreso promedio por afiliado. Para aquellos financiadores con cobertura a toda la población de entre 30 y 45 años, aproximadamente, se observa que hay distintos niveles de promedios de ingresos por afiliado que luego, para las edades más altas, se traducen en prácticamente dos subgrupos en paralelo con niveles distintos de ingreso.

Figura 21: Representación gráfica de variable edad promedio e ingreso promedio por afiliado



Fuente: Elaboración propia

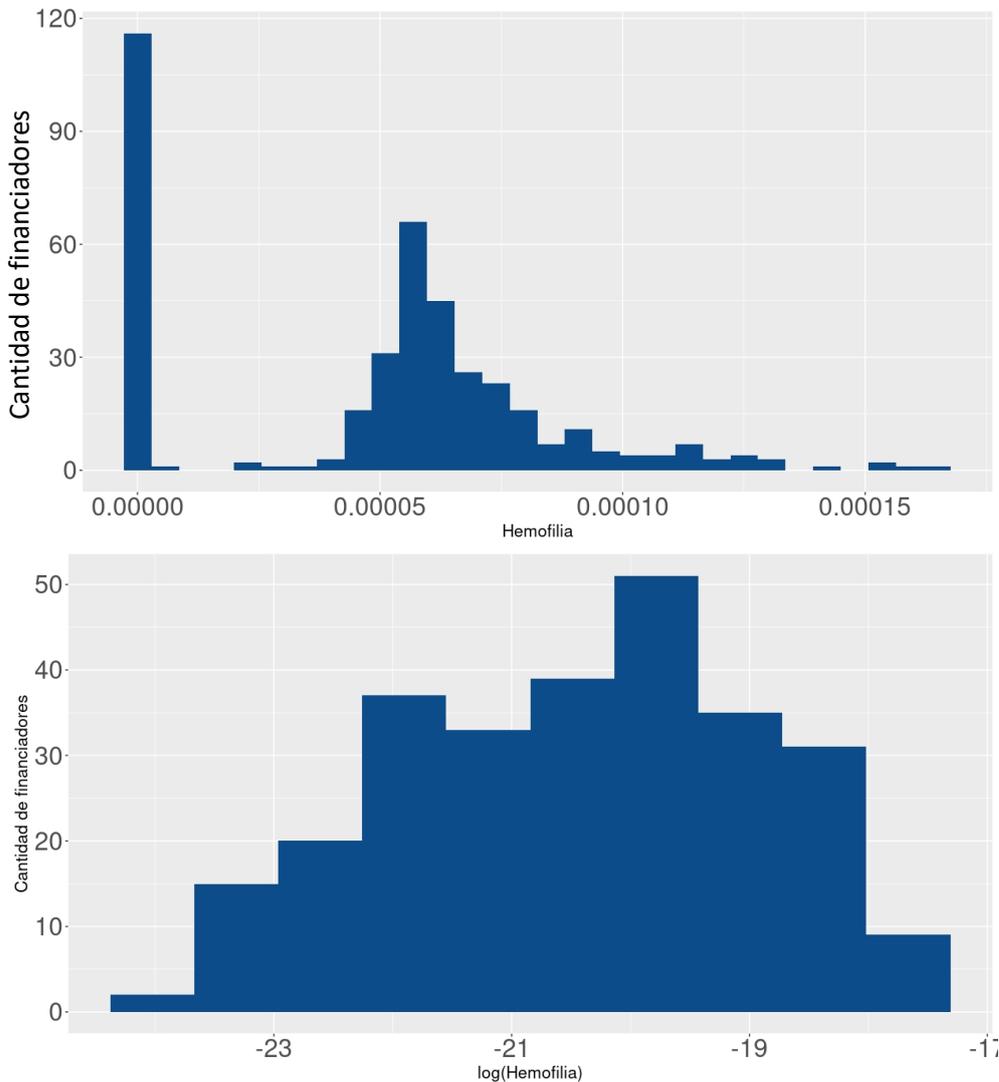
d. Incidencias y Prevalencias por patologías estratégicas

Como se presenta en el anexo 2, se realizan estimaciones a partir de la estandarización del análisis de incidencia y prevalencia específica de las distintas patologías analizadas para cada financiador y se pretende identificar cuáles son los principales costos asociados a una compañía farmacéutica con tratamientos para ciertas patologías, para cada financiador en particular.

En tal sentido, resulta interesante observar, en la figura 22, cómo el riesgo específico de tener ciertas patologías para los distintos financiadores de salud en argentina, como consecuencia del reparto desigual de población por edad y rango etario, implica que las patologías que implican gastos para ciertos financiadores, no necesariamente lo son para otros.

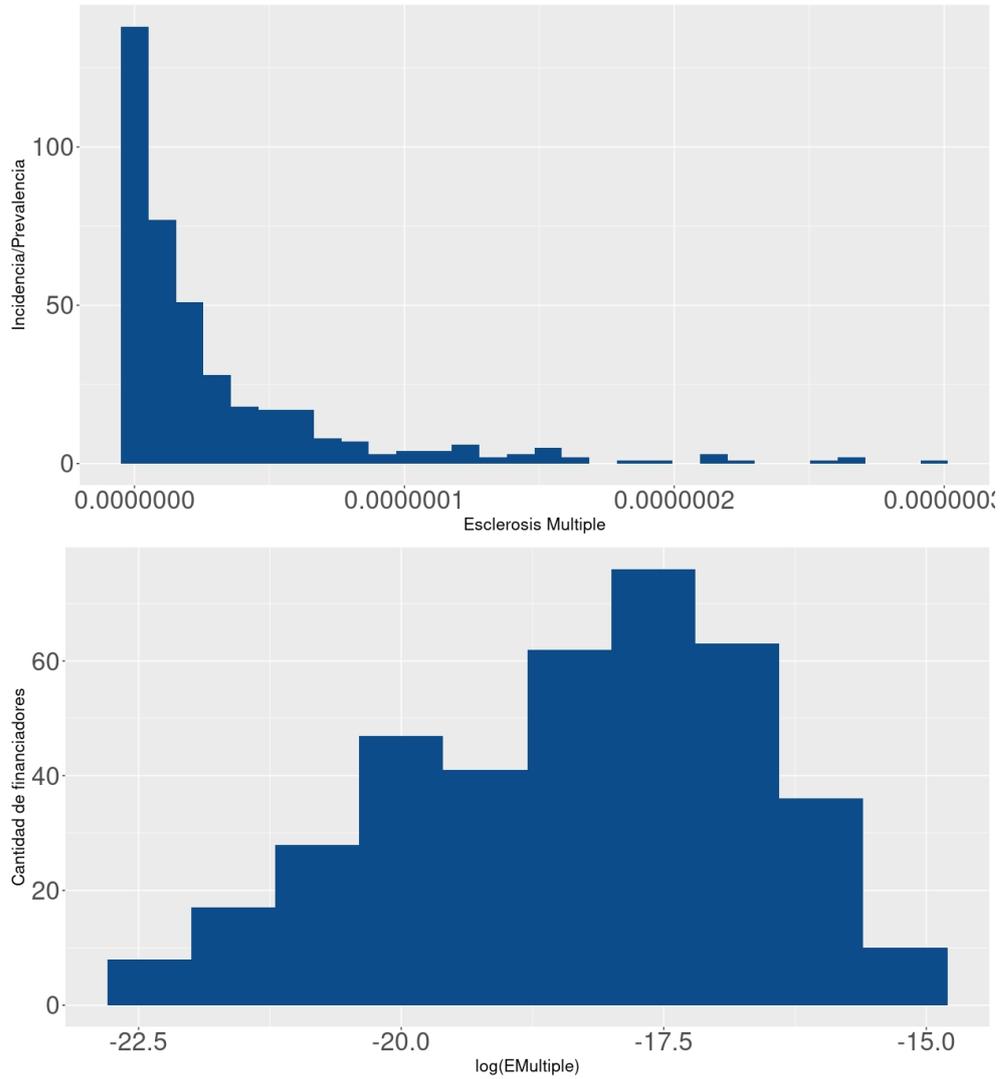
Figura 22: Representación de la incidencia o prevalencia específica por afiliado para los distintos financiadores de Argentina y su versión en función logarítmica

A. Hemofilia A: Se analiza la prevalencia específica cada 100000 habitantes de la patología en función de la tasa específica de hombres por financiador.



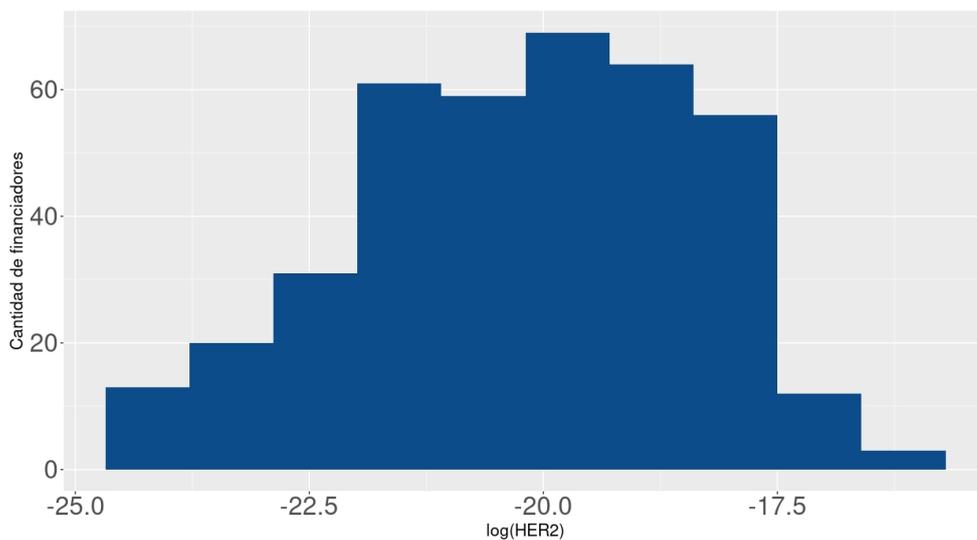
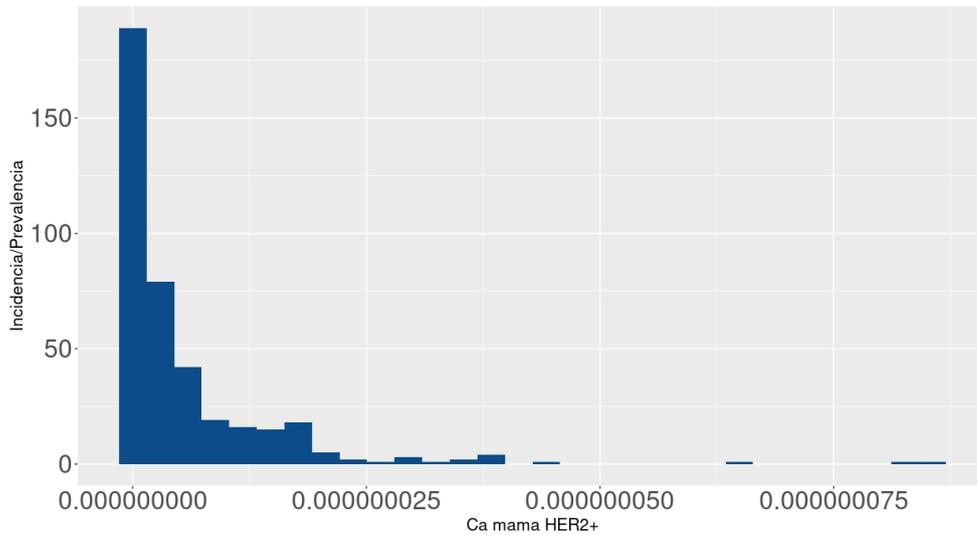
Fuente: Elaboración propia

B. Prevalencia específica de Esclerosis múltiple en función del rango etario de población cubierta por el financiador y género para cada rango etario



Fuente: Elaboración propia

C. Incidencia específica cada 100000 habitantes de cáncer de mama HER 2+ en función de rango etario y género cubierto por el financiador

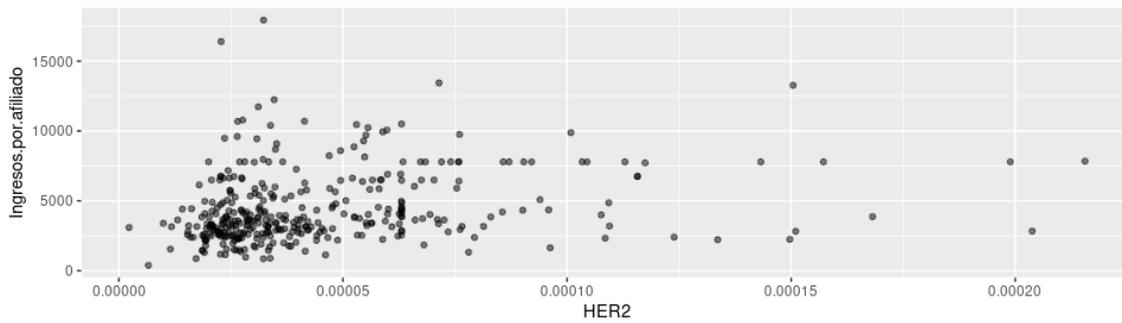


Fuente: Elaboración propia

e. Incidencia y capacidad económica

Luego, cuando se analiza la relación entre la incidencia específica de cierta patología, en relación con los ingresos promedios de cada financiador, se observa que, si bien existe una concentración de financiadores con pocos ingresos y poca incidencia, el reparto económico de la incidencia de HER2, en este primer caso representado en la figura 23, no demuestra gran relación. Se observa que, en función de la distribución de riesgo, la mayoría de la población con menor incidencia tiene ingresos más bajos. Luego, la población con mayor incidencia específica, se representan los mismos grupos identificados en el análisis de edad promedio. En consecuencia, podría esperarse una correlación alta entre esas dos variables: edad y cáncer de mama HER2+.

Figura 23: Ingreso promedio por afiliado contra incidencia de cáncer de mama HER2



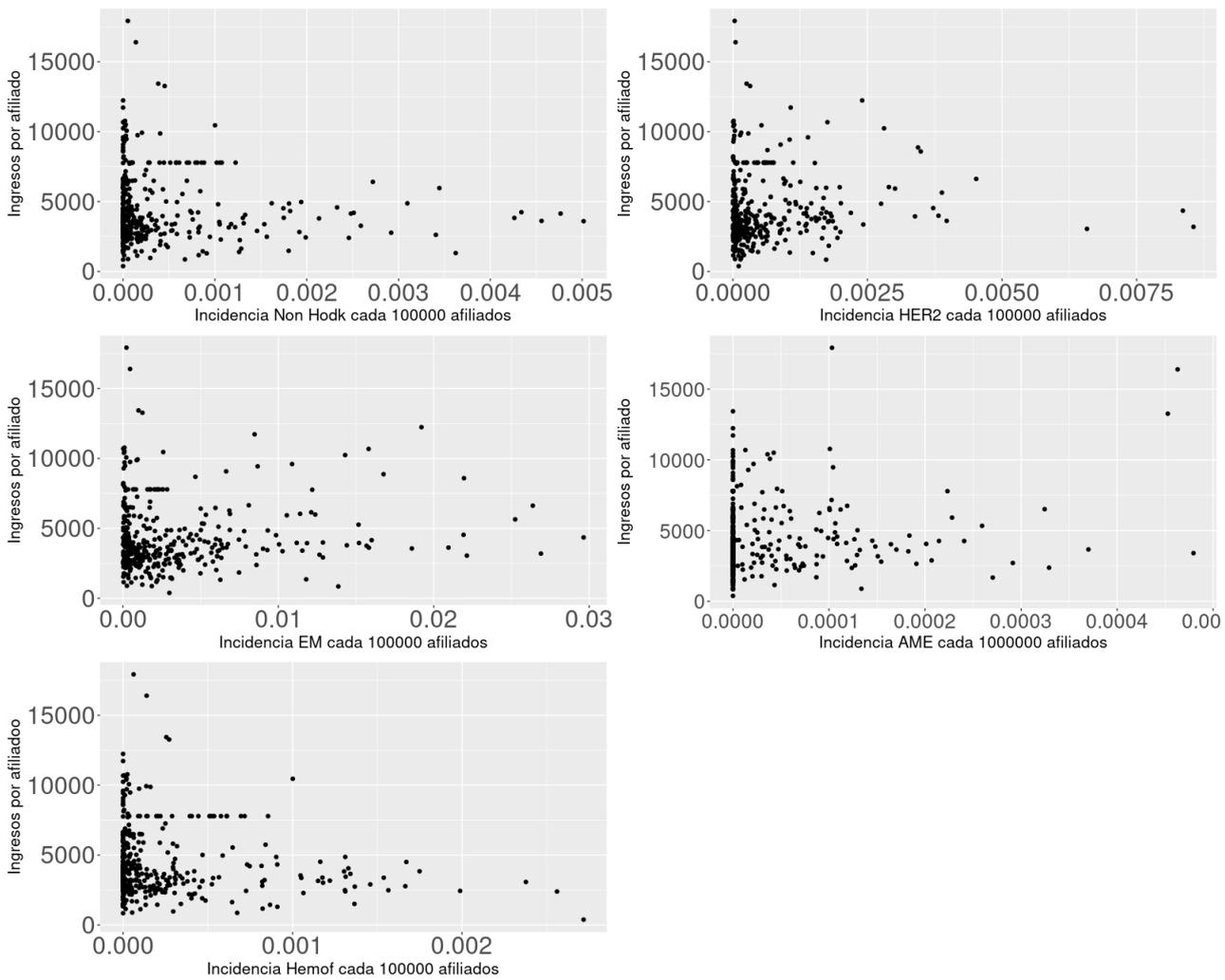
Fuente: elaboración propia

Luego, tal como es presentado en la figura 24, se representa que, para ser incidentes o prevalentes de ciertas patologías, los financiadores no tienen un reparto equitativo de ingresos. En consecuencia, se justifica que tendría sentido contemplar el ingreso promedio de cada financiador en el análisis oportuno de la estrategia para cada producto en particular, con su incidencia específica.

En el análisis de algunas patologías, se observan particularidades en torno a la relación entre estas variables. En el caso del análisis de linfoma de *Non Hodgking*, se observa que los grupos con mayor incidencia promedio tienen ingresos de alrededor de 5000, al igual que la mayoría de los financiadores. Luego, existe un subgrupo con ingresos promedios más altos y una tasa de incidencia más alta. Esto sucede de la misma manera en el análisis de la prevalencia de esclerosis múltiple. Sin embargo, en este segundo caso se identifica un subgrupo de financiadores que tienen ingresos promedios más altos y una tasa de prevalencia más alta por afiliado. Por otro lado, hay una gran cantidad de financiadores que tienen una tasa muy baja, a nivel teórico, de prevalencia esperada de esclerosis múltiple, en función de los rangos etarios y género de su población cubierta.

En el caso de las patologías de baja incidencia analizadas, Hemofilia y AME, se define realizar un análisis puntualizado aparte para identificar si siguen los mismos patrones de distribución en cuanto a la comparación entre las dos variables que se están analizando: la prevalencia y el ingreso por afiliado. Este análisis se detalla en la siguiente sección.

Figura 24: Ingresos económicos e incidencia de patologías estratégicas

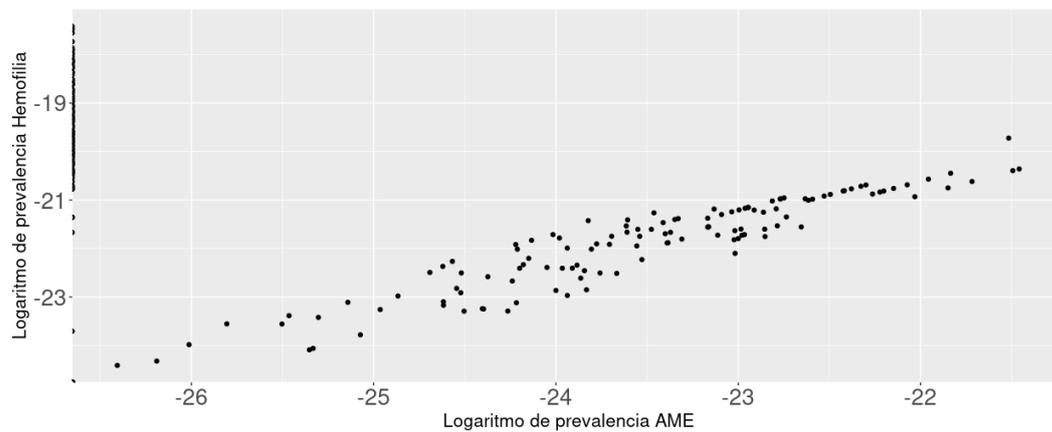


Fuente: Elaboración propia

f. Análisis de patologías de baja incidencia

Dentro de este análisis gráfico surge la necesidad de entender el comportamiento de las patologías de baja incidencia, y, por lo tanto, se busca representar en la figura 25, la relación existente entre dos patologías: AME y Hemofilia. Se observa en dicha figura que, para una gran cantidad de financiadores, la prevalencia de una u otra patología está muy correlacionada. Por otro lado, se observa que ambas variables podrían aportar información diferente al análisis y por lo tanto se conservan ambas. Sin embargo, se podría argumentar, a priori, que es esperable observar una alta correlación entre las dos variables.

Figura 25: Prevalencia específica de AME y Hemofilia para los distintos financiadores



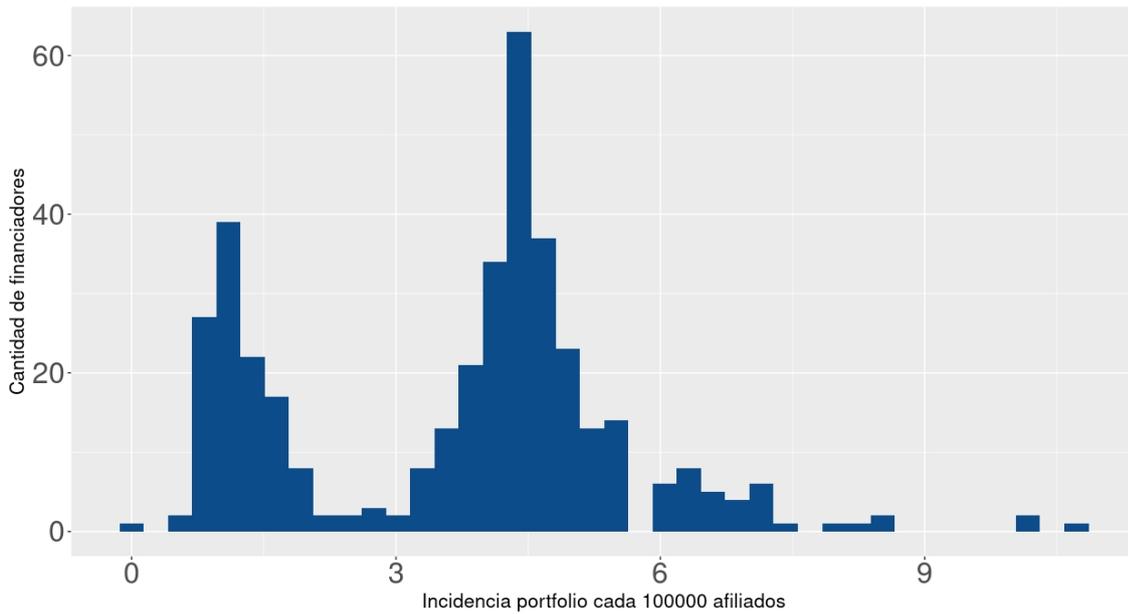
Fuente: elaboración propia

g. Análisis de incidencia de portfolio basado en costo de pacientes

Por último, haciendo un análisis de costos específicos asociados a ciertas patologías, se genera una variable que demuestre la incidencia total potencial del portafolio estratégico de la compañía para cada financiador. Como se observa en la figura 26, la incidencia específica de productos de la compañía, no se reparte equitativamente para los distintos financiadores del país. Si bien existe claramente un grupo significativo de financiadores cuya incidencia específica pareciera ser similar en promedio, hay varios financiadores cuya incidencia específica es muy alta y otros para los cuales la compañía farmacéutica en cuestión, no tiene mucho para ofrecer.

En esta curva es evidente el sesgo que existe hacia la derecha, con pocos financiadores con grandes indicadores de incidencia y una concentración significativa en torno a una incidencia de 5 cada 100000 beneficiarios.

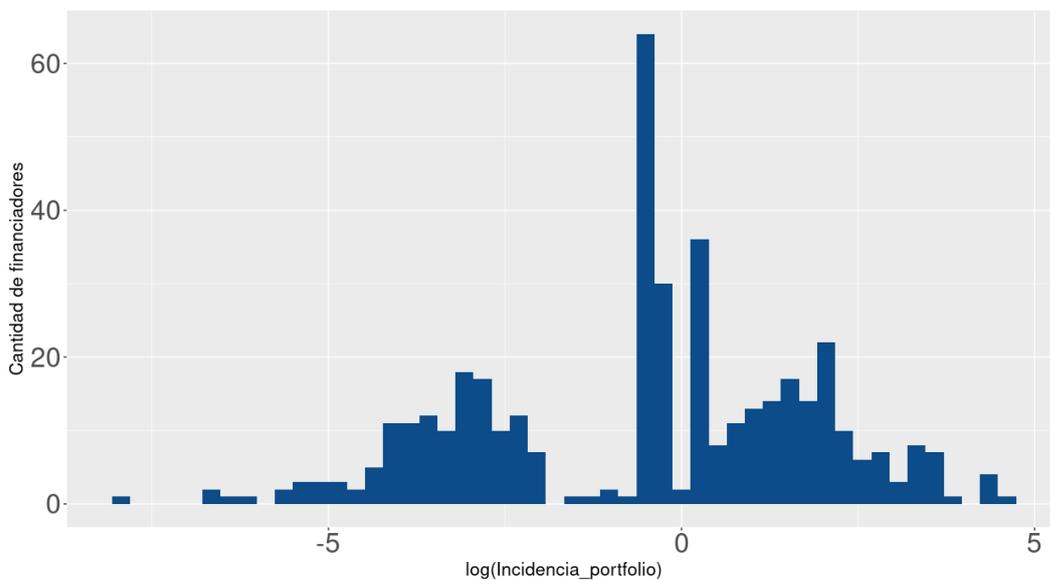
Figura 26: Incidencia potencial de pacientes del portfolio de la compañía analizada en los distintos financiadores



Fuente: elaboración propia

Luego, se propone la aplicación de *log normalización* como consecuencia de la identificación de sesgo en los datos. Se presenta en la figura 27, la representación gráfica del comportamiento luego de aplicar la corrección indicada. Se observan dos grupos principales de incidencia donde, evidentemente, se podría pensar que se reflejan los dos tipos de incidencia principal que se tratan por el laboratorio (patologías de baja incidencia que afectan a población más joven y patologías oncológicas muy asociadas a la edad).

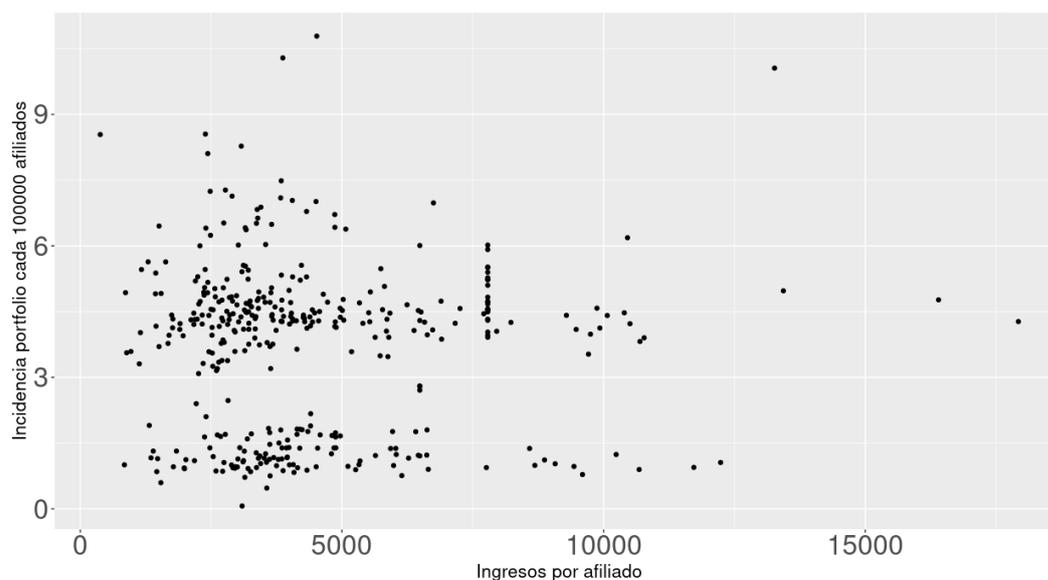
Figura 27: Incidencia del portfolio de la compañía con ajuste de *log normalización* de la variable para corregir el sesgo de los datos y estandarizar la variable



Fuente: Elaboración propia

Por otro lado, se pretende entender también si existen subgrupos posibles a simple vista cuando se grafican los ingresos promedios por afiliado y la incidencia potencial de los productos de la compañía. En tal sentido se observan varios con ingresos bajos y también relativa baja incidencia. Luego, se ven grupos con ingresos altos y baja incidencia y otros grupos con alta incidencia del portfolio de la compañía y también ingresos relativamente bajos. Es evidente que el análisis de los ingresos es fundamental para garantizar la sostenibilidad de los financiadores de este último grupo, por ejemplo. Se representa en la figura 28 el comportamiento de dichas variables.

Figura 28: Representación gráfica del ingreso promedio por afiliado y la incidencia potencial del portafolio de productos estratégicos de la compañía en análisis



Fuente: Elaboración propia

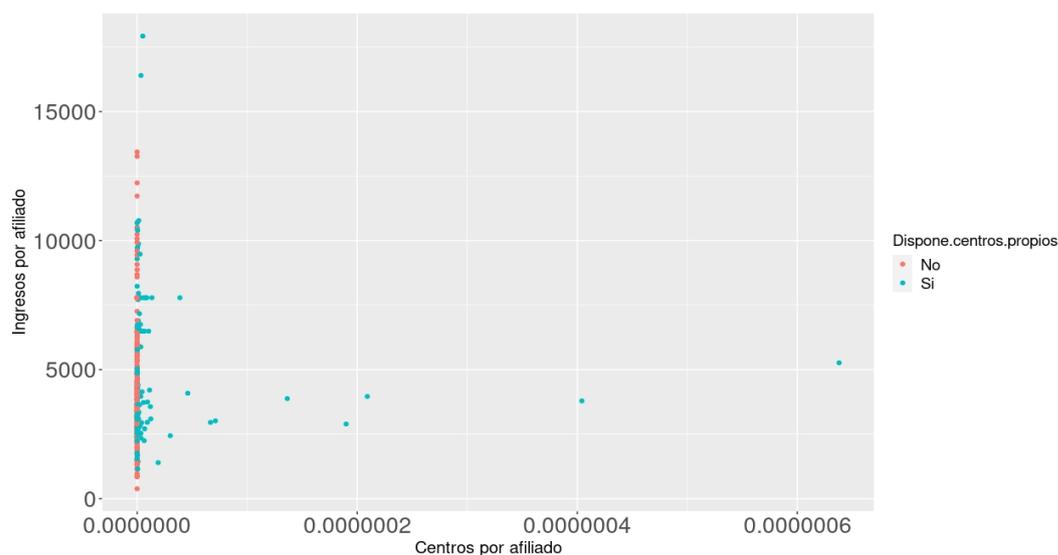
h. Centros propios de salud

Por último, el último elemento que se explora en este trabajo, es el de la disponibilidad de centros propios de salud. Dicha variable permite identificar el nivel de integración vertical que podría tener el financiador. Si se disponen una gran cantidad de centros propios para cubrir a los afiliados, el financiador tiene mayor posibilidad de elaborar guías de tratamiento específicas para sus centros y con esto, hacer un mayor control de los gastos potenciales. Por el contrario, financiadores que no tienen centros, dependen del trabajo de distintos centros sobre los que tienen menor control en el diagnóstico y la decisión de tratamiento. Como se observa en la figura 29, hay una gran proporción de financiadores para cada subgrupo, es decir que disponen o no disponen de centros. Por otro lado, hay algunos pocos que tienen una gran cantidad de centros. Estos financiadores podrían ser muy claves para la compañía ya que posiblemente atienden en esos centros a beneficiarios de otros financiadores que no los disponen. En definitiva, aquellos que son dueños de centros de salud podrían ser estratégicos.

Para la construcción de esta variable, se ha hecho un análisis pormenorizado de los datos disponibles en REFES (Registro Federal de Efectores de Salud). Por un lado, se asignaron las instituciones públicas de Salud al distrito al que principalmente prestan servicios sanitarios

(siguiendo un criterio federal³⁴). En la mayoría de las provincias se consideraron todas las instituciones que tienen disponibles y, en el caso de las instituciones que se encuentran en el AMBA, se hizo una asignación estimada a distintos distritos (estimado a partir de informes disponibles en páginas *web* o estimados por el autor). Luego, las instituciones privadas fueron analizadas una por una asignándolas a algún financiador en particular en aquellos casos en que, a partir de la información contenida en distintas páginas *web*, se pudiera asociar a algún financiador en particular.

Figura 29: Disponibilidad de Centros propios y Cantidad de centros propios por beneficiario



Fuente: Elaboración propia

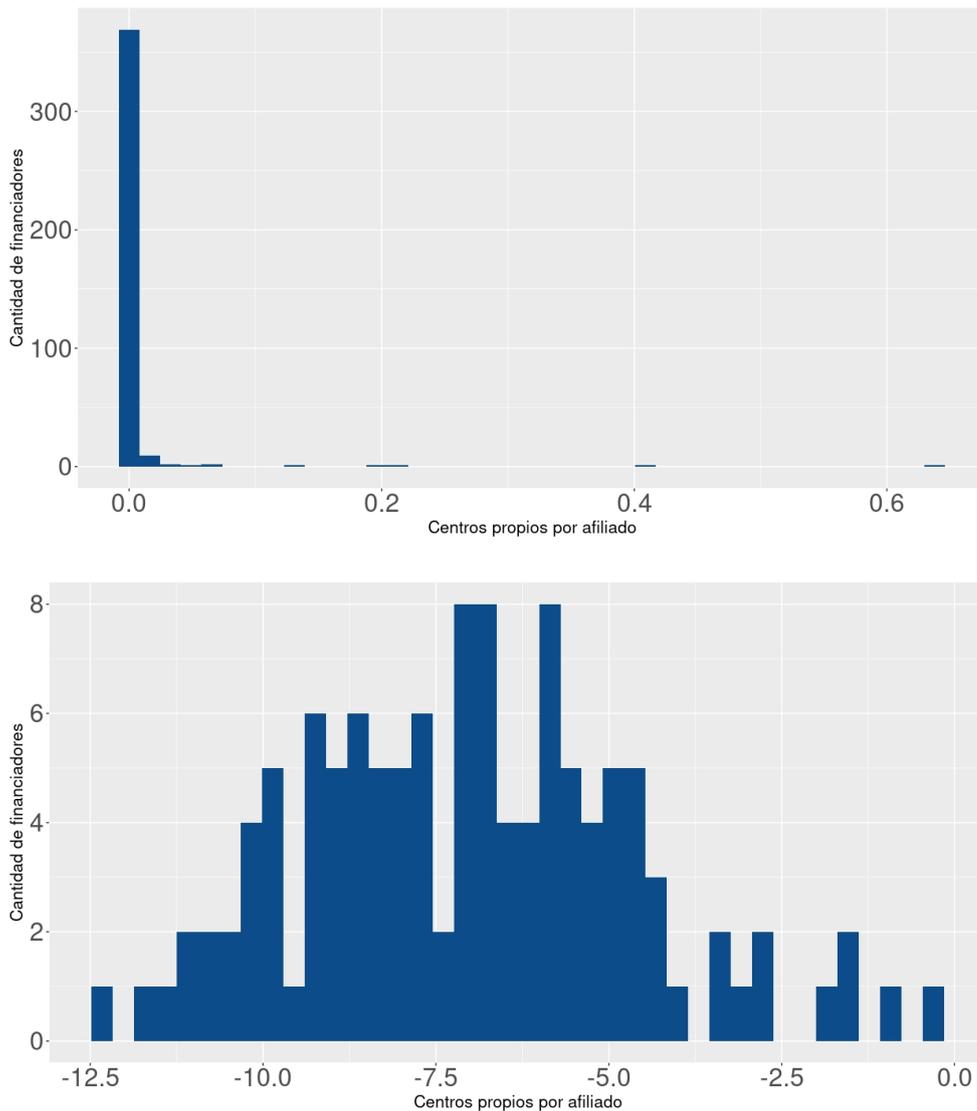
Luego, a partir del comportamiento de los datos, se genera un ajuste poblacional de esta información donde se divide el número de afiliados cubiertos por el financiador por la tasa de consultas promedios anuales por afiliado publicado por información estadística del Ministerio de Salud de la Nación³⁵ (0.66) y se divide la cantidad de centros propios por la cantidad de afiliados * 0.66. Se adiciona un análisis aplicando logaritmo para entender mejor el comportamiento de la variabilidad de este dato.

³⁴ Cuando se indica que se sigue un criterio federal se refiere a que se asume que, si cierta institución de salud está instalada en cierta provincia, entonces los beneficiarios que obtendrán atención en esa institución son los de esa provincia. Es importante mencionar que dicha asunción se considera una simplificación en el análisis como consecuencia de que, para muchas patologías de alto precio, la población se moviliza en el territorio con el objetivo de obtener un tratamiento, pero es una propuesta inicial para la inclusión de la variable.

Además, se asume que los hospitales cubren solamente a la población con cobertura pública exclusiva, lo que también es un supuesto que se podría corregir con nuevas fuentes de información.

³⁵ Informe de Indicadores básicos en salud del año 2020. Ministerio de Salud Argentina (2021).

Figura 30: Representación de cantidad de centros propios corregido por la cantidad de millones de afiliados cubiertos por financiador y representación logarítmica para aquellos con disponibilidad de centros propios



Fuente: Elaboración propia

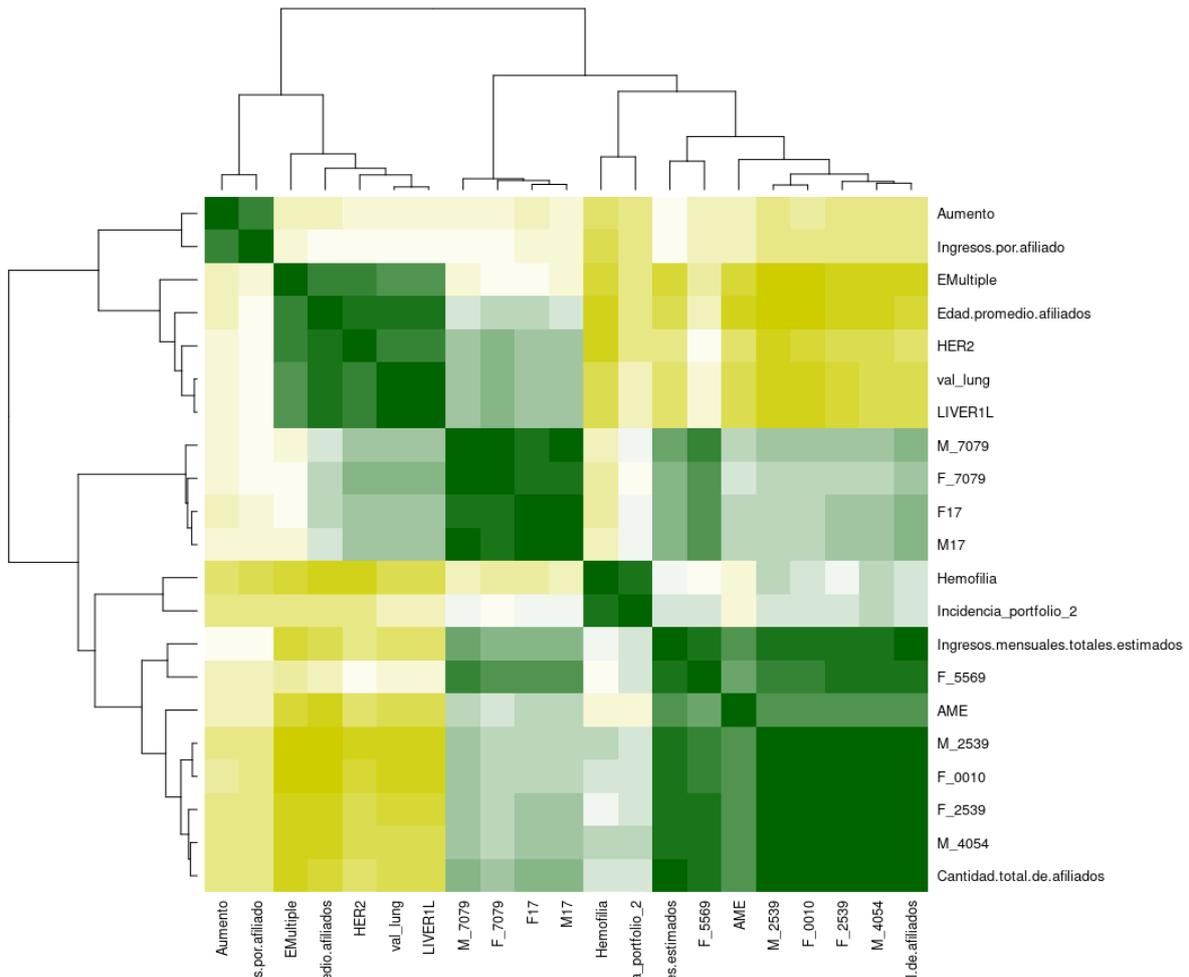
La segunda etapa, a partir del análisis específico de las distintas variables que se tienen para los financiadores y la ejecución de las tareas de estandarización, *log normalización* y construcción de nuevas variables, es ejecutar un análisis de correlaciones cruzadas con el fin de entender si tiene sentido incluir en el análisis, a tal cantidad de variables estimadas.

Con el objetivo de entender en profundidad el tipo de aporte de variabilidad en el comportamiento de los datos aportados por cada variable incorporada en el modelo, se ejecuta un análisis con la técnica FAMD (*Factor Analysis of mixed data*). En el anexo 3, se incluye un detalle de la explicación de las distintas variables en su contribución a las dimensiones identificadas. Se realiza el análisis con las variables normalizadas y luego se hace el mismo análisis con las variables log normalizadas, obteniendo resultados similares en ambos modelos. Se presentan los resultados de ese análisis en el mismo anexo. Con el resultado obtenido del análisis específico, se observa que el aporte en la

variabilidad de los datos que dan las variables categóricas, al menos con aquellas que se tienen en este *dataframe*, no tiene sentido. Por lo tanto, son descartadas del análisis.

Luego, a partir del análisis gráfico de correlaciones entre variables, se identifica que muchas de las variables para estos datos, reflejan alta correlación (Figura 31). Esto sucede cuando la variación del comportamiento de una variable está muy alineada al comportamiento de otra variable. En tal sentido, en un primer modelo se pretende aplicar la técnica de *filtering* para eliminar variables con mucha correlación entre sí, combinando este análisis con componentes principales.

Figura 31: Análisis gráfico de mapa de correlaciones de variables



Fuente: Elaboración propia

En la figura 31, se representa el mapa de correlaciones donde mientras más intenso y oscuro es el color verde, hay correlación más cercana a 1 y cuando el color del verde es más claro, correlación opuesta.

Este mapa de correlaciones, se vuelve muy relevante en la definición de la estrategia a aplicar en el análisis ya que se observan las distintas variables que efectivamente permitirían diferenciar los datos. Por un lado, se observa que las variables de ingreso por afiliado y aumento están correlacionadas. En tal sentido, se evidencia que aquellos financiadores que inicialmente tuvieron ciertos ingresos, los incrementaron de manera tal que no se modificaron (alta correlación).

Luego, se evidencia que ciertas patologías están muy correlacionadas entre sí. Este es el caso de la esclerosis múltiple, cáncer de mama HER2, incidencia de cáncer de pulmón, incidencia de cáncer de

hígado; en particular, el análisis permite identificar que la edad promedio está muy relacionada con la incidencia de las mencionadas patologías.

Otro de los grupos con gran correlación entre sí, son los que representan a la población de más edad. En la información del mapa de correlaciones, se observa que aquellos financiadores que cubren a los hombres y mujeres de más de 70 años (M_7079, M17, F_7079, F17), en general, son los mismos. Esto se intuye por la alta correlación identificada entre dichas variables.

La correlación alta entre las variables Hemofilia y de Incidencia_portfolio_2, que refleja un promedio ponderado del portfolio en relación a la cantidad de viales y de ingresos, y se observa que la correlación es alta. Se concluye que la ponderación otorgada a la patología en la construcción de la variable es alta.

En el último grupo con correlación alta entre sí, se evidencia que las variables de ingresos mensuales totales estimados, cantidad de afiliados, poblaciones más jóvenes (M_2539, F_0010, F_2529, M_4054), tienen alta correlación entre sí. Asimismo, se evidencia alta correlación con la patología AME, cuya incidencia esperada está asociada a población más joven. Es evidente que la cobertura de la población más joven está a cargo de financiadores de más tamaño en cuanto a población cubierta y, linealmente, en cuanto al presupuesto total que se estima, en este modelo, que manejan.

A partir del análisis visual, se podrían identificar tres grupos de variables: un grupo con más ingresos en promedio y con más prevalencia e incidencia de patologías como el cáncer y la esclerosis múltiple; otro con la población mayor, que claramente está cubierta a través de pocos organismos (en realidad, PAMI, que según el informe de Coberturas de Salud en Argentina³⁶, representa a 75 puntos porcentuales de la población de más de 65 años); el tercer grupo está representado por financiadores que cubren a grupos más grandes de población y que en general tienen entre sus beneficiarios a la población más joven, a los niños y, a su vez, a la prevalencia esperada de patologías de baja incidencia como AME.

Con las técnicas implementadas, se busca entender el comportamiento vectorial de las distintas variables disponibles y determinar la inclusión de las dimensiones identificadas como relevantes dentro de ese análisis. Con esto se pretende construir una nueva base a considerar como modelo para ejecutar el *clustering*. En definitiva, este análisis es utilizado en el primer modelo a ser testeado en este trabajo para realizar la segmentación de financiadores.

V. Modelos seleccionados para ser testeados en el análisis de *clustering*

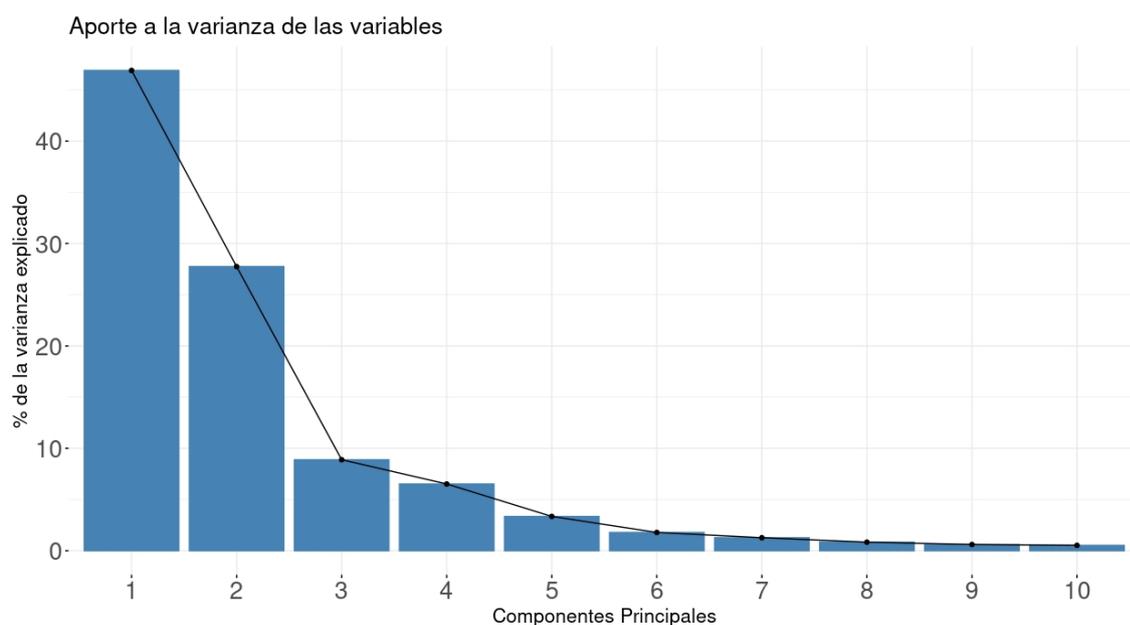
- **Modelo 1: Técnica *Filtering* combinada de análisis de componentes principales. Estrategia única de segmentación**

Como se menciona previamente, en la figura 31, se presenta el mapa de correlaciones identificado inicialmente entre las variables que se incorporan al análisis. Se observa que existen 5 subgrupos de variables claramente identificadas que se comportan, a priori, alta correlación.

Luego, a partir de la selección de las variables con alta correlación, se construye un *dataframe* al que se le aplica el análisis de componentes principales. En la figura 32, se presenta el peso específico que implican las distintas dimensiones en la explicación de la variabilidad de los datos.

³⁶ Informe de Coberturas de Salud en Argentina. Secretaría de Equidad. Ministerio de Salud, 2022.

Figura 32: Análisis de componentes principales de variables con alta correlación entre sí

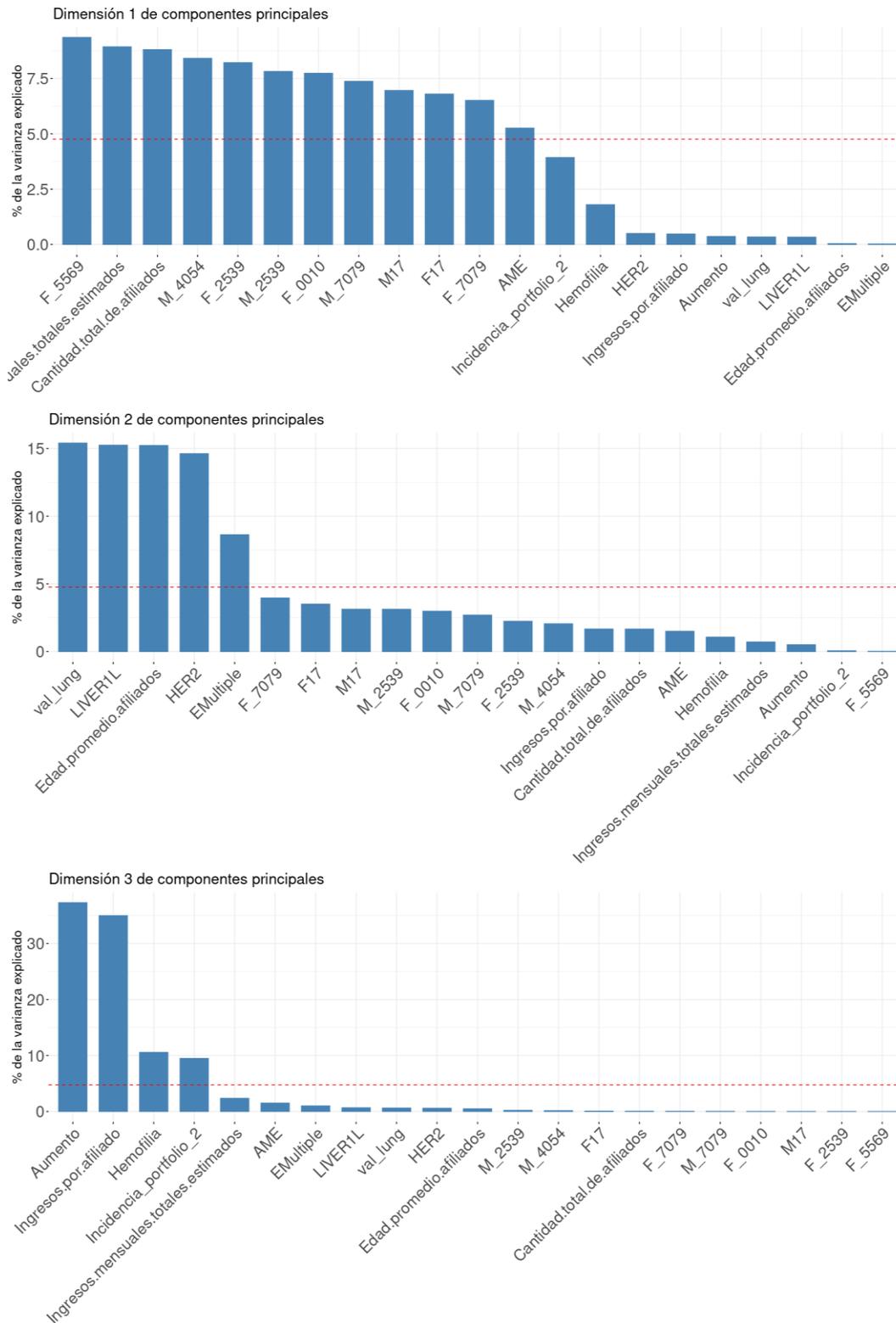


Fuente: Elaboración propia

Cuando se analizan en profundidad las distintas dimensiones y la representación que tienen dentro de cada dimensión, se observa que los tres ejes de diferenciación más significativos, en línea con las conclusiones del análisis del mapa de correlaciones, están dados por las siguientes dimensiones (representadas en la figura 33):

- **Dimensión 1:** Representa variables relacionadas con el tamaño del financiador. En general, los financiadores más grandes en comparación con los más chicos, tienen su gran diferencial en ciertos rangos etarios y concentran gran volumen de afiliados de esos rangos específicos.
- **Dimensión 2:** Refleja la asociación con la edad y las patologías más vinculadas con ciertos rangos etarios. Asimismo, vemos que dentro de este grupo se incorpora la variable de ingreso promedio por afiliado.
- **Dimensión 3:** Pondera las variables económicas del financiador. Refleja el ingreso promedio del financiador y a su vez contempla la ponderación de los incrementos económicos sobre el ingreso del financiador. A su vez refleja el impacto económico que reflejan ciertas patologías de baja incidencia sobre la población específica del financiador.

Figura 33: Análisis de la contribución de las distintas variables con alta correlación en la dimensión 1, 2 y 3

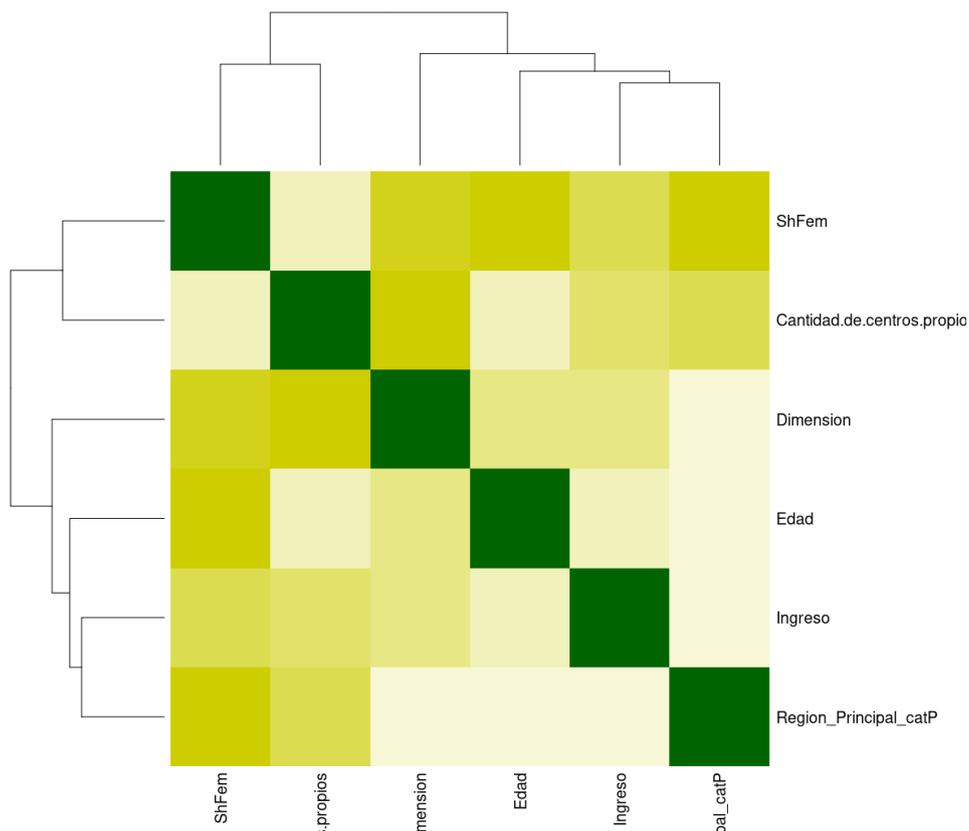


Fuente: Elaboración propia

El primer modelo, en consecuencia, está generado por una base con 6 variables donde se reflejan las distintas dimensiones identificadas mediante el análisis de componentes principales y aquellas que no tenían alta correlación a priori con el análisis de *filtering*. Definimos a las variables de este modelo

como: **Cantidad de centros** (refleja la concentración en la prestación de sus servicios para la población que cubre), **Share Femenino** (refleja proporción de población femenina), **dimensión** (variable que refleja al componente principal 1 en el análisis previo), **Región principal** (se refleja presencia de mayor o menor intensidad de cobertura para la región en particular. La variable graficada es Region_Principal_CatP), **Edad** (corresponde al componente principal 2 y refleja poblaciones añosas y mayor presencia de patologías oncológicas), **Ingreso** (refleja disponibilidad de recursos y el impacto que generan los medicamentos de más alto precio que cubren patologías de baja incidencia). Se presenta la matriz de correlaciones de este modelo en la figura 34.

Figura 34: Representación de matriz de correlaciones de primer modelo a ser testeado para la segmentación de financiadores



Fuente: Elaboración propia

Modelo 2: Modelo Análisis con variables por subgrupos de patologías- Estrategia de múltiples segmentaciones

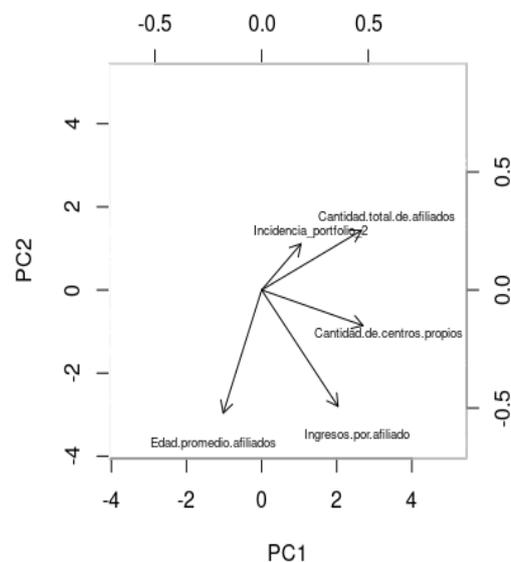
Como alternativa al análisis, se propone establecer una clusterización que divide a los financiadores según el subtipo de producto que esté bajo análisis. Consiste en reformular la segmentación a nivel agregado y ejecutar tres alternativas adicionales al modelo: para el análisis del portfolio en general, sin incluir variables asociadas a patologías; para patologías agrupadas según: edad y baja incidencia. Se generan tres *dataframes* adicionales a ser testeados:

- **Portfolio:** donde solo se incluirá la variable que pondera la población incidente en función del peso de la compañía sobre la tasa de incidencia del financiador, ajustando dicho cálculo a la venta total de la compañía en concepto de un paciente (en función de los precios de lista y las tasas de uso de los productos).

- **Baja incidencia:** Se incluyen patologías de baja incidencia principalmente dado que en general están más presentes en financiadores grandes (al menos en consecuencia del cálculo de la tasa de incidencia utilizada por esta estimación)
- **Edad y género:** Se incluyen patologías más asociadas a la edad de la población: principalmente patologías oncológicas. Asimismo, se incluyen patologías que son más incidentes en población femenina (esclerosis múltiple)

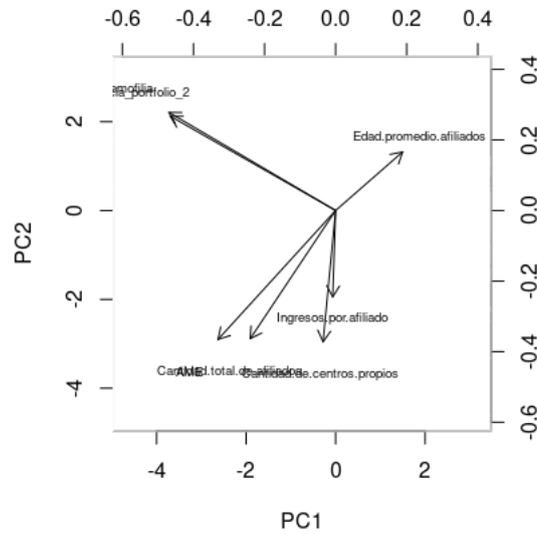
En este modelo, se incluyen sólo las variables numéricas (y en consecuencia no se incluye la Región Principal). Asimismo, se incorporan a las variables ya identificadas las siguientes: Cantidad total de afiliados, Edad promedio de afiliados, Incidencia portfolio, Cantidad de centros propios, Ingresos por afiliado. Además, se ajustan algunas de las variables a escala logarítmica. Se representan en las figuras 35, 36 y 37 los resultados del análisis de componentes principales para cada uno de los *dataframes* analizados alternativamente al modelo 1. Se observa que existe aporte equitativo en el nivel de variabilidad de los datos.

Figura 35. Dataframe Portfolio. Representación de análisis de componentes principales



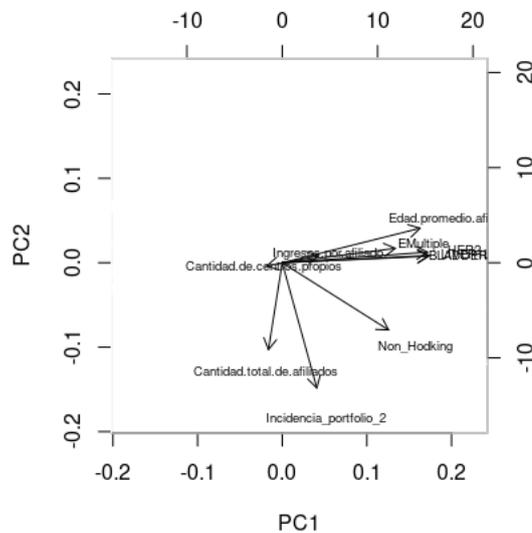
Fuente: Elaboración propia

Figura 36. Dataframe baja incidencia: Representación del análisis de componentes principales



Fuente: Elaboración propia

Figura 37. Dataframe Edad: Representación del análisis de componentes principales



Fuente: Elaboración propia

Se busca analizar la precisión de los *dataframes* definidos en la etapa de Resultados de este trabajo. Por otro lado, en función de los subtipos de financiadores que resulten de los distintos clústeres, se busca construir el modelo de optimización lineal que permite definir la decisión de precios para cada segmento de financiadores.

c. Modelado: Selección de modelos a testear

I. Modelos de clustering

Como se presenta en la introducción, los métodos de *clustering* son muy relevantes dentro de las técnicas de minería de datos. Los algoritmos de *clustering* son muy utilizados para la segmentación de clientes, el procesamiento de imágenes, entre otros. El objetivo primordial de este método es el de dividir un conjunto de objetos de datos en grupos de manera tal que los objetos en un mismo grupo sean lo más similares posibles entre sí y que sean lo más distintos posibles a los otros grupos de datos.

Se pretende testear cuatro enfoques distintos para abordar la tarea de clusterización de financiadores. Inicialmente se pretende analizar la efectividad de los algoritmos de partición. Dentro de esta familia, se utiliza la técnica de ***K-means clustering*** y el mecanismo de ***Agglomerative Hierarchical Clustering***. Luego, se prueba la efectividad para determinar los posibles grupos de datos con una técnica de *clustering* basada en un **modelo que asume una distribución Gaussiana de los datos de maximización de la expectativa**. A continuación, se presentan brevemente las técnicas que luego se evalúan para los modelos de datos ya propuestos.

K means Clustering

Los algoritmos de partición son técnicas que subdividen los datos en k grupos donde k representa al número de grupos pre especificado en el análisis. La técnica de *K means* es la más frecuente dentro de las técnicas de partición y consiste en dividir los datos en k clústeres distintos y que no sean superpuestos entre sí.

Para realizar la agrupación, en primer lugar, se debe elegir un número de clústeres que debe estar entre 1 y k, a donde se asignan las distintas observaciones. Estas son las asignaciones iniciales de las observaciones. Luego, el proceso de reasignación corresponde a un mecanismo matemático donde la idea detrás del algoritmo es buscar que la variación interna del clúster sea la menor posible. Este coeficiente de variación que se define dentro de cada clúster C_k es una medida que en general es definida con la distancia euclidiana cuadrada, es decir la $W(C_k)$. El problema que se busca resolver es la minimización de la medida de distancia en cada clúster k-ésimo. Es decir que el problema consiste en resolver:

$$\min_{C_1, \dots, C_k} \left\{ W(C_k) = \frac{1}{|C_k|} \sum_{i, i' \in C_k} \sum_{j=1}^p (x_{ik} - x_{i'k})^2 \right\} \quad [5]$$

Luego, el algoritmo, tal como se describió, trabaja con el siguiente mecanismo:

1. Asignar aleatoriamente un número, de 1 a k para cada una de las observaciones.
 - a. Para cada k clúster, calcule el centroide. El k-ésimo centroide del clúster es el vector de las p medias de las características para las observaciones del k-ésimo clúster
 - b. Asignar cada observación al clúster cuyo centroide sea el más cercano a partir de la cercanía definida por la distancia euclidiana
2. Iterar hasta que las asignaciones de clústeres dejen de cambiar

Agglomerative Hierarchical Clustering o clustering jerárquico

Este es un método aglomerativo o de abajo para arriba. El número de clústeres no es necesario que haya sido seleccionado previamente. El principio que define este algoritmo parte de la construcción de un dendograma que es una representación de las observaciones en forma de un árbol. Cada hoja

de ese dendograma constituye una de las n observaciones. A medida que se sube desde las hojas por el árbol, algunas hojas comienzan a fusionarse en ramas en función de su similitud. Mientras más arriba en el árbol iremos llegando a asociaciones entre diversas ramas. Mientras más arriba en el árbol sea la fusión entre las diversas ramas, las diferencias serán mayores entre sí. Es relevante indicar que la similitud entre observaciones solo puede ser observada a nivel vertical. Es decir, cuando son fusionadas a través de ramas. Sin embargo, si se observa en un eje horizontal, no se puede sacar conclusiones sobre la similitud entre las observaciones.

En función del número de clústeres que se quiera obtener, se pueden ir haciendo cortes horizontales a través del dendograma. Visto de arriba hacia abajo, si no se hacen cortes, entonces se obtiene un único clúster, luego cuando se baja por las ramas y se hace un corte a la altura n , se obtiene una cantidad n de clústeres de observaciones.

Se puede resumir al algoritmo de *clustering* aglomerativo como:

1. Empezar con n observaciones y una medida de distancia o disimilitud (puede ser la distancia Euclidiana) para todas las combinaciones entre n y $2 = n(n - 1)/2$ de pares desiguales. Tratar cada observación como un propio clúster inicialmente.
2. Para cada elemento $i = n, n - 1, \dots, 2$:
 - a. Examinar todos los pares de disimilitudes entre los i clústeres e identificar el par de clústeres que tienen el menor nivel de disimilitud entre sí (es decir, los que son más similares entre sí). Fusionar estos dos clústeres. La disimilitud entre esos dos clústeres indica la altura del dendograma en el cual la fusión será fijada.
 - b. Computar nuevamente la similitud entre los $i - 1$ clústeres remanentes

Modelos de maximización de expectativa utilizando distribución Gaussiana

El modelo conocido como *Gaussian Mixture Model* (GMM) brinda más flexibilidad que los modelos por particiones ya que hacen un uso ingenuo del valor medio.

Los modelos gaussianos implican asunciones menos restrictivas que tan solo la media ya que utilizan dos parámetros, por un lado, la media y por el otro el desvío estándar. Luego, a partir de asumir una cantidad k de clústeres, se inicializa una asignación aleatoria y se evalúa cuál es la probabilidad de que una observación específica pertenezca a un clúster en particular. Mientras más alta sea la medida de probabilidad de que una cierta observación sea asignada a un clúster k , entonces más sentido tendrá la asignación de esa observación a ese clúster específico. Dicho cálculo de probabilidades se repite hasta obtener la máxima probabilidad logarítmica resultante de la asignación de dichas observaciones.

La probabilidad de que una observación pertenezca a un clúster k se define con la fórmula:

$$r_{ik} = \frac{\pi_k N(x_i; \mu_k, \Sigma_k)}{\sum_k \pi_k N(x_i; \mu_k, \Sigma_k)} \quad [6]$$

Definimos al algoritmo a utilizar como:

1. Inicializar aleatoriamente los valores de μ_k y de Σ_k
2. Repetir el siguiente proceso hasta maximizar la función logarítmica:
 - a. **E-Step:** Para cada observación x_i , se calcula la probabilidad de pertenecer al clúster C_k . Cuando se observa una probabilidad muy alta de pertenencia al clúster, es cuando más seguros se estará de que hay una buena asignación a ese grupo

- b. **M-Step:** Se actualizan los valores de μ_k y de $\sum k$ con la nueva asignación y se vuelve al paso anterior hasta alcanzar el valor máximo de cierta asignación:

$$\mu_k = \frac{\sum r_{ik} * x_{ij}}{\sum w_j} \quad [7]$$

$$\sum k = \frac{\sum r_{ik} * (x_i - \mu_k)(x_i - \mu_k)^T}{\sum r_{ik}} \quad [8]$$

II. Métricas de Rendimiento

Para evaluar el trabajo realizado en la asignación de clústeres para las distintas observaciones, existen distintas medidas que permiten evaluar la calidad de esas asignaciones. Todas tienen el objetivo de detectar la similitud entre los elementos de cada clúster y la diferencia entre los datos de diferentes clústeres. La mayoría de los algoritmos de clusterización producen resultados más allá de la existencia real de clústeres en los datos. Por lo tanto, se vuelve una tarea fundamental la implementación de métricas que permitan entender si la asignación realizada tiene sentido o no.

En un nivel general, la validación de clústeres se puede implementar a través de herramientas externas, internas y relativas (Jain y Dubes, 1988). Dentro de estas familias, los criterios externos se emplean principalmente cuando se quieren evaluar los resultados de clústeres contra una solución deseada y conocida de antemano. Éstos criterios son muy útiles para la evaluación y comparación del algoritmo en experimentos controlados (Färber et al., 2010). Luego, los criterios de validez interna basan su evaluación sólo en las tareas de *clustering* y los datos propios que se utilizaron para entrenar el modelo. Evalúan qué tan buena es la estructura del método de clusterización sin necesidad de información ajena al propio algoritmo y su resultado. Por último, los criterios internos, que también son relativos, son utilizados para evaluar y comparar la calidad de las diferentes particiones de manera relativa. En este trabajo, por la forma de los datos, se ha decidido implementar el criterio de la validez interna.

Dentro de los criterios de validez interna, se utilizan herramientas basadas en los criterios de cohesión intra clúster y separación inter clúster. Para ejecutar estas tareas, el cálculo de la conectividad, el índice de Dunn y la anchura de la silueta son tres medidas que se utilizan para evaluar la calidad de los resultados del *clustering*. Se definen brevemente estas métricas que en la sección Resultados son analizados para los distintos modelos:

- **Conectividad:** Es una medida que evalúa la distancia promedio entre los puntos de un clúster. Cuanto menor sea la distancia promedio, más “conectados” estarán los puntos en el clúster. Esta herramienta es utilizada para evaluar la cohesión del clúster. Una conectividad alta indica que los puntos del clúster están muy cerca unos de otros y, por lo tanto, tienen una alta cohesión. Se utiliza la siguiente fórmula para realizar su cálculo:

$$C = \left(\frac{1}{N}\right) * \sum_{i, j \in C, i \neq j} \frac{d(i, j)}{\min\{k \in C, k \neq i\}} d(i, k) \quad [9]$$

dónde: N representa el tamaño del conjunto de datos, $d(i, j)$ es la distancia entre los elementos i y j dentro de las variables, luego la función $\min\{k \in C, k \neq i\}$ indicada, representa la distancia mínima entre el objeto i y cualquier otro objeto dentro del clúster C . Por último, $\sum_{i, j \in C, i \neq j}$ representa la suma entre todas las posibles parejas de objetos distintos en el clúster C . En conclusión, se calcula la distancia mínima de i a cualquier otro elemento en C y cuando se calcula eso para todos los objetos que pertenecen a ese mismo clúster C y la media de dichas relaciones se convierte en la métrica representativa de la

conectividad. Al tener que realizar ese proceso iterativo para cada elemento i en comparación con cada elemento dentro del mismo clúster, esta técnica es costosa a nivel computacional. Sin embargo, se considera que al tener una base de datos que no es tan grande en tamaño, podría tener sentido evaluar el comportamiento de esta métrica en los modelos analizados.

- **Índice de Dunn:** Es una métrica que evalúa la separación entre los clústeres. Cuanto mayor sea el índice de Dunn, mayor será la separación entre los clústeres. En otras palabras, un índice de Dunn alto indica que los clústeres están bien separados y, por lo tanto, son distintos entre sí.

$$D = \frac{\min_{1 \leq i \leq j \leq n} d(i, j)}{\max_{1 \leq k \leq n} d'(k)} \quad [10]$$

dónde: $d(i, j)$ representa la distancia entre grupos i y j , $d'(k)$, mide la distancia intra-grupos del cluster k . La distancia entre los dos grupos puede ser cualquier número que se utilice de distancia entre los centroides de los grupos. El criterio busca grupos con alta semejanza intra-grupo y baja semejanza inter-grupo.

- **Anchura de la silueta:** esta medida evalúa la similitud entre los puntos de un clúster y los puntos de otros clústeres. Cuanto mayor sea la anchura de la silueta, más similares serán los puntos dentro del clúster y menos similares serán con los puntos de otros clústeres. Con esta métrica se puede evaluar la homogeneidad del clúster donde un valor de anchura alto indica que los puntos dentro del clúster son muy similares y, por lo tanto, el clúster es homogéneo. Para una observación i dentro de la base, se obtiene:

$$S = \frac{(b - a)}{\max(a, b)} \in (-1, 1) \quad [11]$$

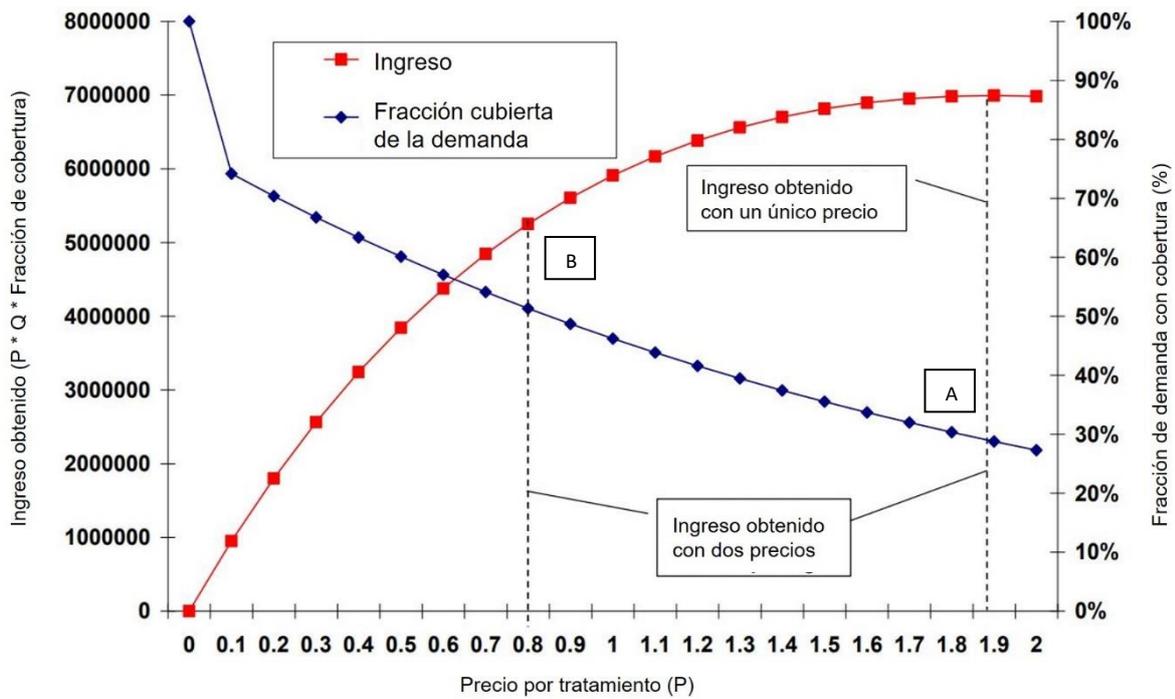
dónde: a es el promedio de la distancia de la observación i con las demás observaciones del clúster al que pertenece esa observación y b es la distancia mínima a otro clúster que no es el mismo de esa observación. Es decir, ese clúster con que se compara es la segunda mejor opción para esa observación. En consecuencia, cuánto más alto sea b , más individualizado estará ese clúster y mientras más alto sea a , más cohesión interna en el clúster habrá. Luego, se estandariza para que el valor no supere -1 y 1 y pueda ser analizado.

III. Modelos de determinación de precios diferenciales

Como ha sido brevemente presentado en la introducción, existen diversas publicaciones que recomiendan la implementación de modelos de precios diferenciales para medicamentos de alto precio como una herramienta que potencialmente podría generar mayor acceso a esos tratamientos, y a la vez obteniendo sustentabilidad para el sistema de salud. Como indica Yadav (2010), en su trabajo sobre precios diferenciales para medicamentos, el análisis económico similar al del nivel entre países también podría ser utilizado en una segmentación a nivel intra país con discriminación de precios entre los financiadores de alto ingreso que podrían incrementar la posibilidad de pago de las poblaciones de menor ingreso y, de hecho, se podría generar ingresos incrementales para las compañías farmacéuticas. En tal sentido, Yadav (2010), presenta una figura (adaptada en Figura 38), donde se muestra cómo la falta en la posibilidad de determinar precios diferenciales implica que un porcentaje significativo de la población no tenga acceso al tratamiento (Yadav, 2010). En la figura del autor, se evidencia que, si se ofertara un único precio, la cobertura de la población sería tan solo de 30 puntos porcentuales como consecuencia de la inequidad en los segmentos de ingresos dentro del país. Por el contrario, se refleja que podría alcanzarse, con dos precios en el mismo territorio, 90

puntos porcentuales de la población total y, de hecho, se incrementarían los ingresos percibidos por el financiador en el mercado. En la figura 38, se representa que en el punto de decisión A, sólo se define un único precio que es alto y el porcentaje de población cubierta es de 30 puntos porcentuales. En un escenario de dos precios, se podrían determinar la estrategia A y, adicionalmente, la estrategia B. Con esto, la población efectivamente cubierta implica la suma de los dos grupos de demanda identificados.

Figura 38: Diferenciación de precios a nivel país con efectos positivos en el ingreso y la cobertura de la población



Fuente: Adaptación de Yadav, 2010³⁷

Para resolver el problema de la definición del porcentaje de descuento, se pretenden utilizar diversas herramientas de soporte a la toma de decisiones.

Se busca reflejar un concepto muy difundido en la teoría de la economía de la salud, conocido como precios diferenciales basados en valor (VBDP, por las siglas en inglés de *Value Based Differential Pricing*) que ha sido presentado en el trabajo de Danzon et al (2015), donde se presenta un marco de análisis y asunciones tendientes a la definición de precios eficientes entre y dentro de los distintos países. Como definen, “se trata de designar un sistema en que los precios de los productos se relacionen con el valor clínico con respecto a los tratamientos existentes”. Se incorpora, además, la relación entre el valor, pero también la maximización de la ganancia en salud dada a partir de un presupuesto en cada país y que permita alcanzar eficiencia dinámica y estática a nivel global (Danzon et al., 2015). En tal sentido, en este trabajo se pretende reflejar, a través de un modelo de optimización lineal, el equilibrio entre esos dos elementos. Se incorpora en el análisis la utilidad de los consumidores y las presiones presupuestarias a la vez que se definen precios que permiten hacer transferencias de ganancias a las compañías farmacéuticas para incentivar el desarrollo de investigación de nuevas tecnologías superadoras. Se busca perseguir conceptualmente el planteo realizado por Danzon et al. (2015), donde se define una función de medida de la elasticidad E donde

³⁷ Yadav, P. (2010). Differential Pricing for Pharmaceuticals. MIT Zaragoza International Logistics Program.

los precios varían inversamente con el nivel de elasticidad. La elasticidad para el producto j es igual a la elasticidad precio compensada ϵ_j sumado a la elasticidad ingreso (η) de la demanda multiplicada por la cuota específica (s) del producto j :

$$E_j = \epsilon_j + s_j \eta_j \quad [12]$$

Luego, se propone la diferenciación de dicha ecuación:

$$\frac{dE_j}{dY} = \frac{ds_j}{dY} \eta_j + \frac{d\eta_j}{dY} s_j \quad [13]$$

donde la cuota s es positiva y, para un cierto precio, desciende con respecto al ingreso. Por otro lado, la relación entre la elasticidad ingreso y el ingreso es positiva; es decir, cuanto más ingreso, mayor elasticidad ingreso.

Luego, a partir de las variables que se determinen a partir del análisis exploratorio de variables y el proceso de ingeniería de datos, se pretende construir un modelo de asignación de descuentos mediante programación lineal.

Se pretende, conceptualmente, maximizar el acceso a los tratamientos sujeto a las variables que determinan la capacidad total de pago por productos farmacéuticos y la voluntad de pago por esos productos farmacéuticos. En tal sentido, se pretende incorporar el análisis de elasticidad específica para cada producto, donde se observa el nivel de sensibilidad en la cantidad comprada con respecto a la tasa de descuento fijada y, por otro lado, se busca determinar un umbral de costos específico para cada grupo donde se permita un acceso sostenible a la tecnología y la garantía del precio mínimo necesario por parte de la compañía farmacéutica.

Los modelos de optimización lineal implican la maximización de una función objetivo (en este caso el acceso a los medicamentos) sujeto a las restricciones económicas, en la voluntad de pago, poblacionales, etcétera que sean determinadas a partir de la solución de análisis no supervisado implementado. En pocas palabras, estos métodos facilitan la decisión de la asignación de recursos (en este caso a través de descuentos) sujeto a las restricciones que pudieran existir.

3. Resultados

d. Evaluación de resultados

IV. Modelos de clustering

Se abordan los 2 modelos distintos que se propusieron en la etapa de análisis de los datos para entender cuál respondía mejor al comportamiento de los financiadores y con el fin de identificar grupos de financiadores de entre 3 y 6 buscando maximizar el nivel de *silhouette score* y a la vez, producir resultados que resulten razonables como insumos para el modelo de precios diferenciales. Asimismo, se decide incorporar, tal como se definió previamente, el análisis del índice de *Dunn* y de Conectividad como herramientas alternativas que permitan que la selección del modelo a definir sea el ideal para representar la estrategia a ser implementada. El objetivo se aborda, como ya ha sido mencionado, a partir de entender primero si valdría la pena una estrategia integrada de segmentación de financiadores a nivel general (modelo 1) y, por el otro, buscando entender si valdría la pena una segmentación a nivel producto - financiador, con variables específicas para cada producto en el caso específico que se esté analizando según se trate de un medicamento para una patología de baja incidencia, más asociada a la edad o considerando simplemente el peso en el

portfolio de la compañía (Modelo 2). En tal caso, se muestran en orden los resultados de los distintos modelos que se fueron experimentando en este trabajo.

Modelo 1: Técnica *Filtering* combinada de análisis de componentes principales. Estrategia única de segmentación

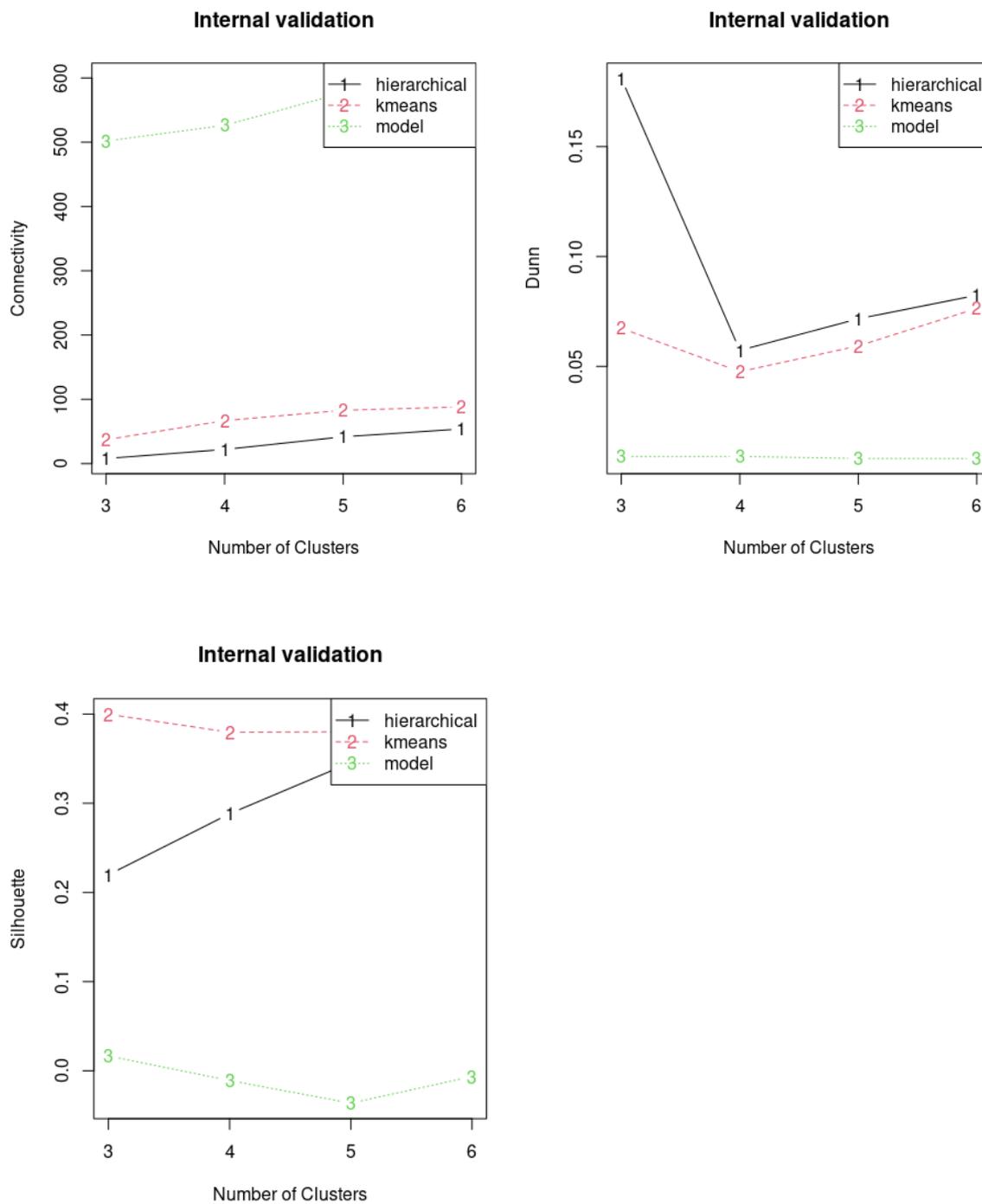
Se analizan los resultados de implementar tres modelos de *clustering*: *Hierarchical*, *k means* y *model based clustering* y se exploran distintos escenarios de cantidad de clústeres k a ser analizados (k=3,4,5 y 6).

En la tabla 2, se presentan los resultados de los indicadores *de Dunn*, *Connectivity* y *Silhouette* obteniéndose que la estrategia ideal es implementar tres clústeres con la estrategia *k means*. En consecuencia, aplicando ese modelo, se obtendría un indicador de Dunn de 0.06 y un indicador de Connectivity de 37.16. En la figura 39, se representa la performance de las distintas métricas analizadas gráficamente.

Tabla 2: Resultados de experimentación del modelo 1 con variables seleccionadas previamente. Selección del modelo de *k means* para generar k clústeres de financiadores

Modelo	Análisis de k-clústeres	Anchura de silueta	Índice de Dunn	Conectividad
<i>K Means</i>	k= 3	0.3998	0.0673	37.16
	k= 4	0.3797	0.0476	66.80
	k= 5	0.3802	0.0593	82.98
<i>Clustering jerárquico</i>	k= 3	0.2198	0.1810	7.9536
	k= 4	0.2883	0.0573	22.009
	k= 5	0.3457	0.0717	41.74
<i>Clustering basado en modelo normalizado</i>	k= 3	0.016	0.009	501.82
	k= 4	-0.011	0.009	526.24
	k= 5	.0.036	0.008	576.76

Figura 39: Representación gráfica de métricas de rendimiento analizadas en el primer modelo de clústeres



Fuente: Elaboración propia

En el modelo seleccionado, se establecen 3 clústeres que contienen cada uno a 123, 181 y 89 financiadores en cada clúster determinado. En consecuencia, el primero agrupa a 31 puntos porcentuales de los financiadores, el segundo a 46 puntos porcentuales y el tercero, a 22 puntos porcentuales.

A partir de un análisis de los perfiles que selecciona dicho modelo de financiadores, se muestran a continuación la representación de cada clúster en la tabla 3.

Tabla 3: Representación de valores promedios por clúster determinado

Clúster	Financiador de ejemplo	Cantidad de centros propios	Share Femenino	Región Principal	Dimensión	Edad	Ingreso
1	OSMATA Mecánicos	0.00003	0.4315	0.5368	-3.778	0.9625	0.2600
2	OS Federación Trabajadores Obras Sanitarias	0.00000	0.4300	0.5368	1.8689	1.3204	0.8178
3	OS Hospital Italiano de Córdoba	0.0004	0.5476	0.1857	-0.2742	-2.6310	-0.0319

En el clúster 1, se encuentran financiadores con algo de disponibilidad en promedio de centros propios, con mayor presencia masculina entre sus beneficiarios, que habitan en su mayoría dispersos en el territorio nacional, con un tamaño muy grande en relación al resto de los grupos de financiadores, con un ingreso intermedio y un promedio de edad relativamente intermedio. Se podría afirmar que este subgrupo de financiadores son los más del estilo de clase media, en edad económicamente activa, pero con ingresos bajos o medios. Se encuentran dentro de este grupo desde organismos públicos de algunas provincias, empresas de medicina prepaga, algunas obras sociales universitarias, etc. Es un subgrupo que tiene cierta comodidad a nivel económico y son los que en mayor medida tienen riesgo significativo de patologías de baja incidencia porque cubre a grandes grupos de poblaciones relativamente jóvenes. Se encuentran la obra social de amas de casa, prepagas como OSDE y Galeno, el público de La Rioja y Catamarca, entre otros, Swiss Medical, OSECAC, entre otros.

El clúster 2 está muy asociado a obras sociales nacionales u organismos públicos con mayores ingresos por beneficiario cubierto donde se agrupa a la población que pertenece a obras sociales nacionales, en su mayoría asociados a sociedades de trabajadores, los gobiernos provinciales de la región patagónica, trabajadores de Shell, Ford, de la industria aerocomercial y otras compañías argentinas. Se incluyen dentro de este grupo a financiadores sin centros propios, con mayoría de género masculino entre sus cubiertos, dispersos en el territorio nacional, con un menor tamaño en general y con edades relativamente intermedias.

Por último, el tercer clúster identificado agrupa a financiadores que dan cobertura a grupos de población en general con mayor edad al resto, con mucha mayor prevalencia de cobertura a mujeres, con baja recaudación de ingresos en promedio por afiliado, en general de trabajo más local a través de hospitales de comunidad o prestadores propios. Se encuentra dentro de este grupo al Hospital Italiano, DAMSU Mendoza, OSPLAD (personal docente), DOSUBA, OSEP Mendoza, entre otros. Son los financiadores que en general tienen a sus beneficiarios a lo largo de toda su vida y por lo tanto tienen mayor prevalencia de patologías asociadas a la vejez (oncológicas, principalmente).

- **Modelo 2: Modelo Análisis con variables por subgrupos de patologías- Estrategia de múltiples segmentaciones**

En este caso, como ya ha sido indicado en la etapa de selección de modelos dentro de la sección Datos, se procede a analizar dos modelos alternativos. Uno que contemple en el análisis sólo de las variables de incidencia o prevalencia específica de patologías de baja incidencia. Y otro modelo que contemple patologías más asociadas a la edad donde las variables consideradas sean las de incidencia de productos oncológicos.

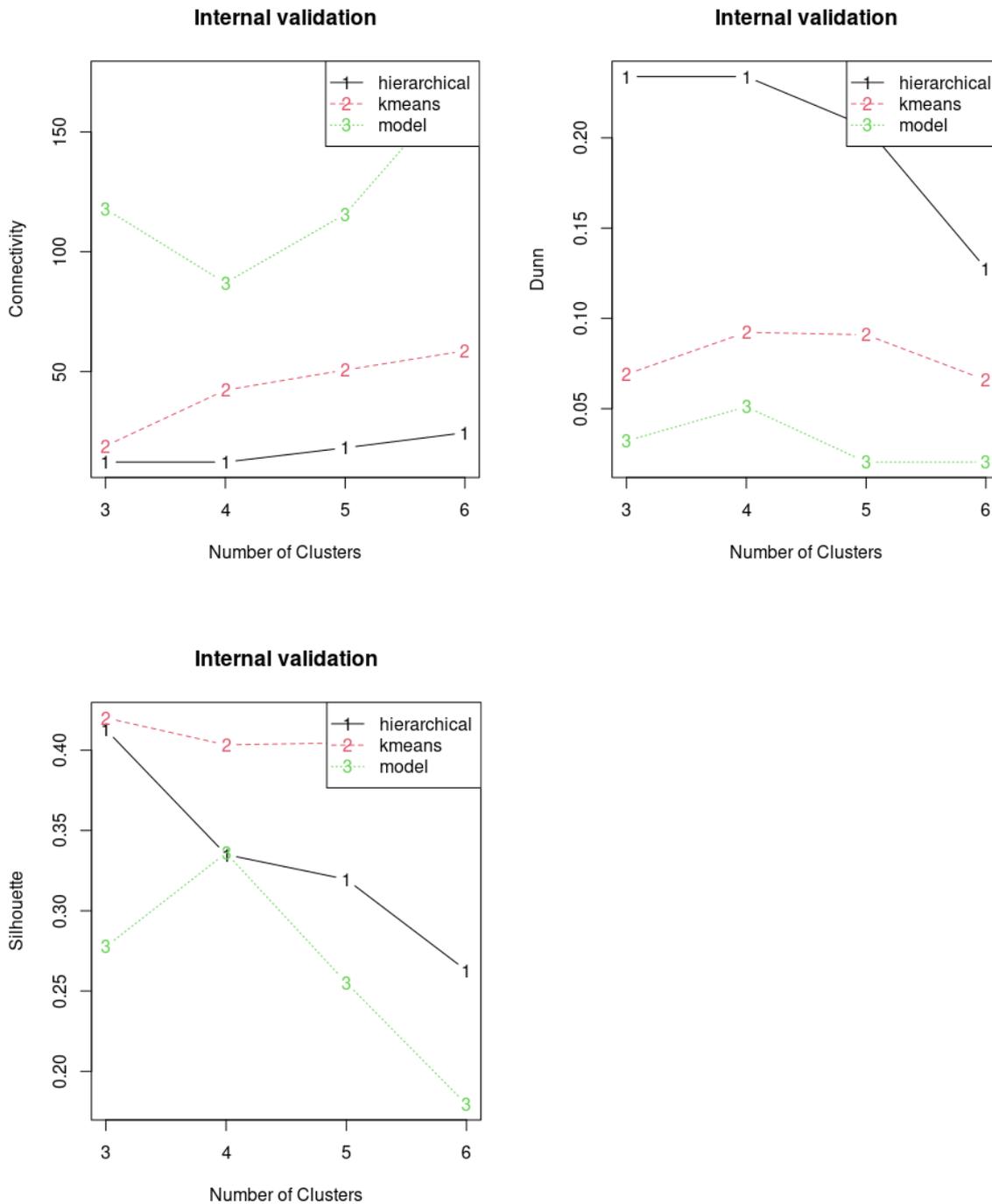
Como se indica previamente, en este modelo se incorporan las variables: "AME", "Hemofilia", "Incidencia portfolio", "Cantidad de centros propios", "Ingresos por afiliado", "Edad promedio afiliados", "Cantidad total de afiliados"

Tabla 4: Análisis de segmentación 1. Análisis de la Incidencia enfermedades de baja incidencia

Modelo	Análisis de k-clústeres	Anchura de silueta	Índice de Dunn	Conectividad
K Means	k= 3	0.42	0.069	19.01
	k= 4	0.40	0.092	42.29
	k= 5	0.40	0.090	50.69
Clustering jerárquico	k= 3	0.41	0.233	12.28
	k= 4	0.33	0.233	12.28
	k= 5	0.32	0.206	18.24
Clustering basado en modelo normalizado	k= 3	0.27	0.03	117.87
	k= 4	0.33	0.05	87.03
	k= 5	0.25	0.02	115.68

Por otro lado, cuando se presentan gráficamente los resultados, se observa por el comportamiento de la métrica de la anchura de silueta, que el mejor modelo es el de k= 3 clústeres de financiadores. En la figura 40, se representa la performance de las métricas.

Figura 40: Representación gráfica de métricas de rendimiento analizadas en el modelo de análisis específico para enfermedades de baja incidencia



Fuente: Elaboración propia

Luego, analizando en particular el modelo de esos tres clústeres, se obtienen tres familias de 116, 155 y 122 financiadores, respectivamente, donde las distintas dimensiones se representan en la figura 42, con sus valores promedio.

Tabla 5: Representación de grupos de modelo para patologías de baja incidencia en k=3 clústeres

Clúster	Financiador de ejemplo	Cantidad de afiliados	Edad promedio afiliados	Ingresos por Afiliado	Hemofilia	AME	Cantidad de centros propios	Incidencia Portfolio
1	OSMATA Mecánicos	0.0708	0.19561	0.1625	0.4006	0.5426	0.0305	0.4163
2	OS Industria Lechera	0.0184	0.2474	0.1977	0.1977	0.000	0.0000	0.4279
3	OS Personal Prensa Mar del Plata	0.0012	0.2990	0.1889	0.1889	0.000	0.0000	0.1083

En este caso se observa que, al analizar los valores promedios de la tabla 5, si se prioriza el portfolio de asignación de descuentos para patologías de baja incidencia, existen tan solo dos grupos significativos ya que como se observa en la figura 42, la incidencia y prevalencia de este grupo de patologías en los financiadores que integran al grupo 3, es prácticamente nula. Son aquellos financiadores que dan cobertura a grupos muy pequeños de población. Tienen ingresos medios con respecto al resto de los grupos de financiadores, edad en promedio más alta y, por lo tanto, menor cantidad de nacimientos (y niños) pasibles de mayor prevalencia específica de estas patologías. La incidencia del portfolio para esos financiadores es baja. Integran este grupo algunas obras sociales universitarias chicas, obras sociales de grupos muy pequeños de trabajadores como escribanos, personal de ceras Johnson, entre otros de ese estilo de cobertura más de mutualidad.

El grupo más relevante en cuanto a incidencia y prevalencia de estas patologías es el número 1 donde se ve gran cobertura a poblaciones con riesgo de tener estas patologías. Este grupo está fuertemente afectado en sus gastos presupuestarios por el financiamiento a este grupo de patologías. Se encuentran dentro de este grupo APROSS, ASE, otras obras sociales provinciales, el programa del estado de INCLUIR Salud, lógicamente, IOMA, IOSFA, OSDE y todos los financiadores que brindan cobertura a grandes grupos poblacionales con niños y que por lo tanto tienen más prevalencia de patologías de baja incidencia. Asimismo, se encuentran dentro de este grupo a los gobiernos provinciales. Es el grupo que menor ingreso en promedio tiene por afiliado y en general tienen mayor integración vertical en su prestación.

Luego, se identifica también el clúster número 2 donde si bien hay mayor probabilidad de tener prevalencia de estas patologías, se cubre a grupos intermedios en cantidad de afiliados y con promedios de edad más altos, en general. Por lo tanto, existe menor riesgo de que estas patologías sean muy prevalentes en estos financiadores. Se encuentran acá las obras sociales universitarias como DAMSU Mendoza, el Hospital Italiano de Buenos Aires, Obras sociales nacionales como la de Techint, Luz y Fuerza, entre otras. Son financiadores de ingresos económicos que en promedio son intermedios (pero superiores a los de mayor prevalencia que integran el grupo 1).

Por último, dentro de esta estrategia de segmentos de productos, se busca analizar la prevalencia de patologías oncológicas más asociadas a poblaciones de mayor edad. Se establece así el segundo criterio de segmentación.

Como se indicó previamente, en este modelo se incorporan las variables: "E Múltiple", "Non Hodking", "BLADDER 2L", "HER2", "LIVER1L", "Incidencia portfolio", "Cantidad de centros propios", "Ingresos por afiliado", "Edad promedio afiliados", "Cantidad total de afiliados"

Tabla 6: Análisis de segmentación 2: Patologías más asociadas a grupos de mayor edad (oncología, principalmente)

Modelo	Análisis de k-clústeres	Anchura de silueta	Índice de Dunn	Conectividad
K Means	k= 3	0.3284	0.074	35.13
	k= 4	0.3028	0.050	67.74
	k= 5	0.3365	0.085	76.84
Clustering jerárquico	k= 3	0.3883	0.1559	14.35
	k= 4	0.3581	0.1559	18.208
	k= 5	0.3466	0.1559	20.90
Clustering basado en modelo normalizado	k= 3	0.177	0.0118	99.11
	k= 4	0.2587	0.0276	105.73
	k= 5	0.1045	0.0168	240.77

Como se observa en la tabla 6, el resultado óptimo de clusterización es la estrategia de aplicar *clustering* jerárquico con un corte para determinar 3 grupos de financiadores. Asimismo, se presenta en la figura 41 las métricas de performance ejecutadas para definir el número óptimo de clústeres y el criterio de identificación de dicha estrategia.

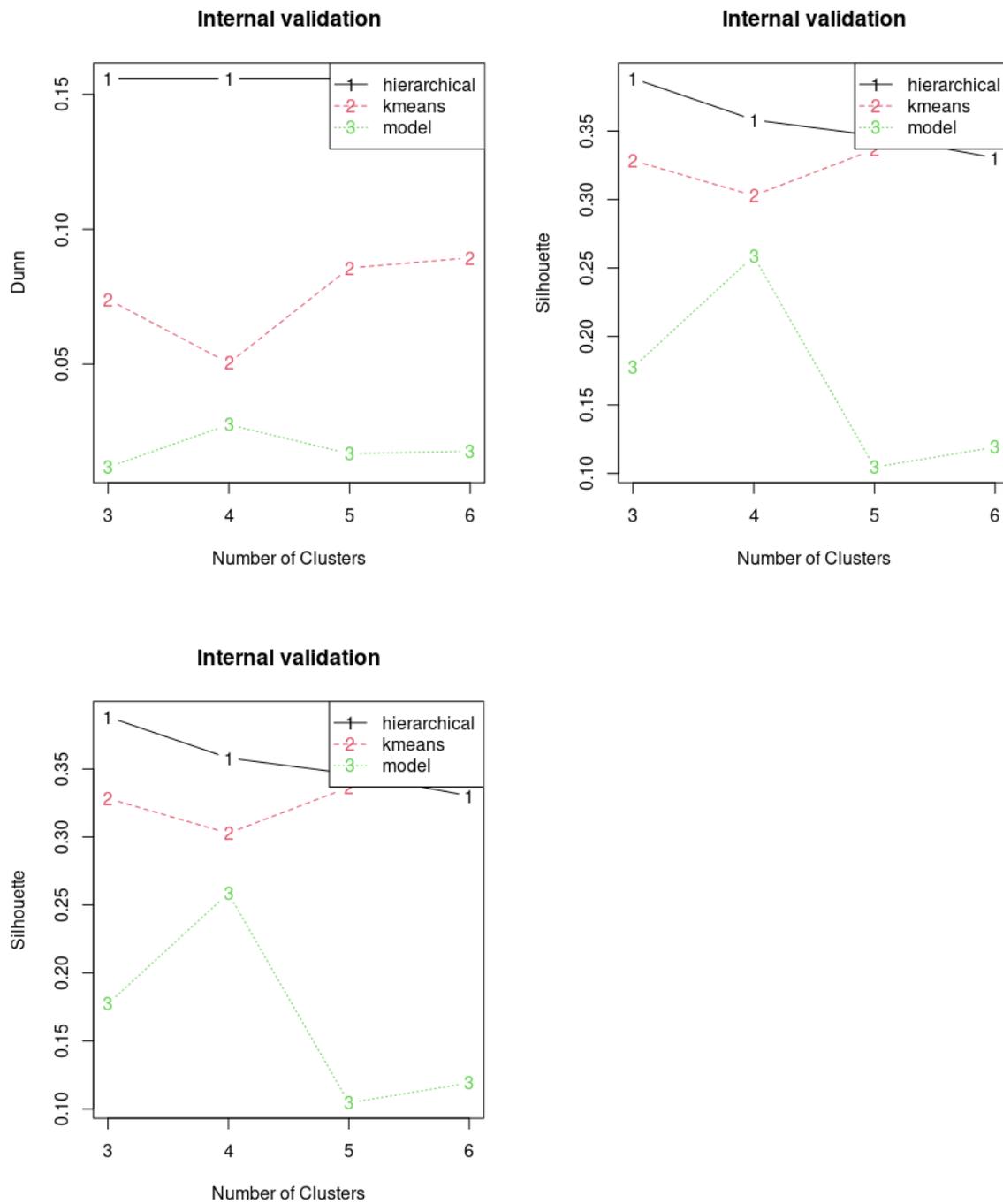
Cuando se analiza el resultado de la asignación de esa estrategia en tres grupos, se observa que se identifican tres clústeres con tamaño 272, 109 y 12, respectivamente.

En el primer grupo, se encuentran todos los financiadores con poblaciones más jóvenes donde hay menor prevalencia e incidencia esperada de las patologías más asociadas a la edad de la población. Se encuentran financiadores con poblaciones jóvenes, de diverso tamaño que dan cobertura a población más joven y que no tienen, a partir de lo presentado anteriormente, tanto riesgo de padecer las patologías en análisis. Se observan en este grupo la obra social APROSS, Swiss Medical, las obras sociales de edades jóvenes como OSECAC, la obra social de peluqueros, la de trabajadores en centros de contacto, entre otras. Tienen diversos ingresos económicos (los más altos y los más bajos del análisis).

En el grupo 2, se encuentran financiadores grandes como OSDE, IOMA y otros que tienen una gran cantidad de población de diversos rangos etarios y que por lo tanto son más significativos en esos financiadores, la presencia de las patologías bajo análisis. En general, tienen ingresos medios en promedio por afiliado. La característica que agrupa a este grupo de financiadores es la cantidad de afiliados cubiertos.

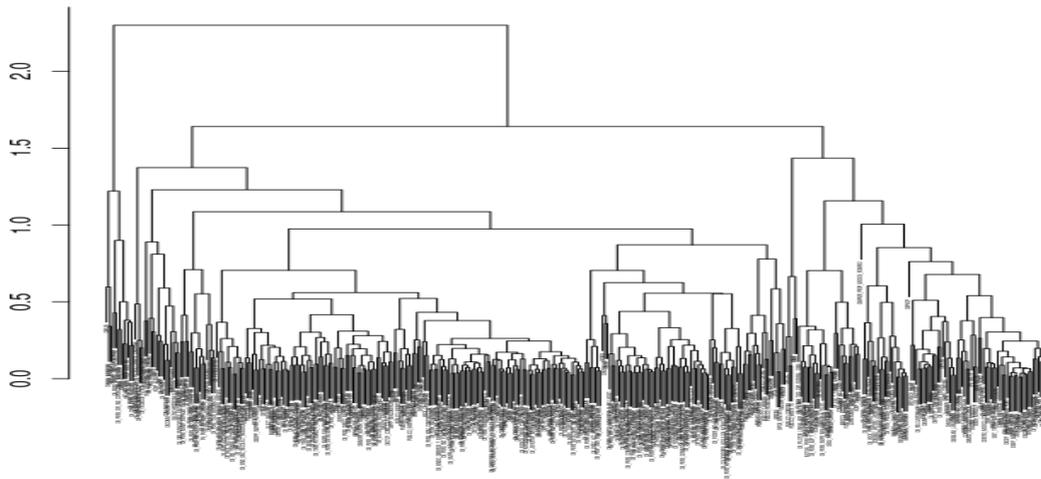
Por último, el grupo 3, está integrado por financiadores que se organizan de manera más mutual y, por lo tanto, cubren en mayor medida a poblaciones con más edad. En definitiva, esto implica que la incidencia de las patologías oncológicas es muy significativa dentro de su población. Se encuentran en este grupo OSEP Mendoza, la OSPJN, la obra social de los trabajadores estatales, entre otras que dan cobertura a grupos significativos de poblaciones con más edad, en promedio.

Figura 41: Representación gráfica de métricas de performance en análisis modelo portfolio oncología y medicamentos para mayor prevalencia por edad



Fuente: Elaboración propia

Figura 42: Dendograma de la distribución de los datos a partir de la definición de la estrategia de segmentación para patologías más asociadas a la edad



Fuente: Elaboración propia

Como resulta evidente a partir del análisis de las métricas de performance, se corta el orden de jerarquías para definir $k = 3$ grupos dentro de ese dendograma

El resultado final del análisis del modelo 2 ejecutado nos ofrece una alternativa de *k means* con 3 clústeres para el análisis de la estrategia para productos del portfolio de enfermedades de baja incidencia y una estrategia de clustering jerárquico de $k = 3$ niveles para la asignación de grupos para definir la estrategia de productos más asociados con el portfolio de oncología y neurología.

V. Análisis de resultados y selección del modelo definitivo

Cuando se contrastan los resultados de las dos estrategias alternativas (análisis agregado y análisis segmentado al negocio) se obtiene que la segunda alternativa refleja un mayor nivel de anchura de silueta que el primer modelo, donde se obtiene una performance de 0.3998. Luego, en la estrategia de portfolios separados, se obtiene una métrica de 0.42 en el modelo para patologías de baja incidencia y de 0.388 en el modelo de oncología y neurología. En promedio se obtiene una performance de 0.40 con esta segunda alternativa. En conclusión, y entendiendo también el comportamiento del mercado y del negocio, se opta por el segundo criterio, ya que tiene una performance levemente mejor y refleja mejor los datos de patologías muy estratégicas en la comercialización de productos de la compañía. En ese sentido, se concluye que dicha opción garantiza una mejor decisión para cada familia de productos.

Las variables que explican esta distribución son: Cantidad de centros propios, Share Femenino, Región principal, Dimensión (tamaño), Edad, Ingreso. Con este criterio se buscan generar las variables para implementar la decisión de asignación de descuentos.

e. Plan de Implementación

VI. Modelos de determinación de precios diferenciales

Con el objetivo de presentar una posible implementación del modelo de segmentación para los financiadores, se propone realizar un sencillo ejercicio que permita entender una modelización del negocio con el objetivo de definir la política óptima de asignación de descuentos a los distintos grupos identificados a partir de la segmentación. Para tomar esa decisión, se propone un modelo de

optimización lineal entera donde se contemplen las distintas variables identificadas como relevantes en la estrategia de segmentación y, además, se busquen reflejar las distintas circunstancias presentadas como relevantes en el tratamiento y la cobertura de los medicamentos de alto precio.

Para plantear un modelo de optimización lineal entera, en primer lugar, es relevante identificar las variables relevantes para el modelo. En este sentido, se proponen distintas dimensiones a contemplar que luego se ven reflejadas en las distintas variables y restricciones incorporadas.

En tal sentido, se establecen los siguientes pasos para implementar una propuesta de modelo teórica a ser implementada en la compañía.

En primer lugar, se identifican posibles variables a incorporar en un primer posible modelo de optimización lineal entera:

- $j \in (1, m)$: donde j es un producto y existen m productos farmacéuticos disponibles: Los m productos farmacéuticos están asociados a las dos posibles familias de productos que se proponen en el modelo de segmentación elegido previamente: patologías de baja incidencia y patologías más asociadas a la edad de los beneficiarios
- $i \in (1, n)$: donde i es un financiador o grupo de financiadores y existen n grupos. En relación al planteo de los distintos segmentos de financiadores determinados en la etapa anterior de este trabajo, se propone trabajar con $n=5$ grupos de financiadores. Es decir, se consideran los 3 clústeres obtenidos mediante el análisis de segmentación realizado en la etapa anterior y se consideran además a 2 que habían sido eliminados por su dimensión significativamente más alta que los otros financiadores: PAMI y la población con Cobertura Pública Exclusiva de la Provincia de Buenos Aires.
- p_{ij} : precio neto del producto j para el financiador i (en el caso de PAMI y PBA con CPE o clúster de financiadores $i= 1,2,3$ en función de la categoría del producto j).
- x_{ij} : cantidad consumida (representada en pacientes cubiertos) por el financiador i para el producto j
- x_{ij}^{MAX} = Incidencia máxima estimada para el producto j en el financiador i : En este caso, se utiliza la variable de prevalencia e incidencia estimada para el modelo de segmentación. Se identifican la incidencia esperada de cada uno de los cinco segmentos de financiadores a considerar (PAMI, PBA Público, Clúster 1, Clúster 2, Clúster 3). En función del producto j del que se trate, se utiliza el modelo para enfermedades de baja incidencia o el de enfermedades más asociadas a grupos con promedio de edad más alto.
- $FLOOR_j$ = Precio piso promedio que se asume para cada producto j . El laboratorio debe satisfacer en el agregado que, en promedio, el precio de un producto j sea superior al $FLOOR$.
- c_j = costo variable del medicamento j resultante de tenerlo en el país y disponible para la venta
- C_{fix} = Costo fijo de mantenimiento del laboratorio farmacéutico a cubrir en el agregado de la venta
- K_{ij} = Representa un umbral para cada producto j en cada financiador i donde se define un porcentaje máximo para ser invertido en la cobertura del total de la población pasible
- Se podría trabajar en una función que refleje la cobertura por parte de los financiadores (si lo aprueban o no) y otra que muestre la intención de prescripción de esos medicamentos para un paciente en particular

Luego, se definen las siguientes reglas de negocio que buscan reflejar el problema planteado en el trabajo:

1. Se establece un modelo donde el laboratorio ofrece $j \in (1, m)$ productos farmacéuticos a un precio específico para cada financiador $i \in (1, n = 5)$
2. El laboratorio farmacéutico tiene como objetivo maximizar el ingreso esperado de la comercialización de sus productos $j \in (1, m)$ y, para tal fin, puede decidir el precio al que vende cada producto para cada financiador i (o subgrupo de financiadores, en este caso)

$$\max \sum_{i=(1,n)}^n \sum_{j=(1,m)}^m p_{ij} * x_{ij} \quad [14]$$

3. Cada financiador i consume una cantidad x_{ij} de cada producto j
4. Cada financiador decide si aprueba el uso del producto en función del precio al que se le ofrece. En definitiva, la decisión sobre la cobertura o rechazo del producto se podría definir como una función de dos tramos donde por simplificación asumimos que consume 0 en caso de rechazos³⁸ y luego enfrentan la curva de demanda propia:

$$x_{ij} = \begin{cases} 0, & wtc_{ij} < p_{ij} \\ D(x_{ij}), & wtc_{ij} \geq p_{ij} \end{cases} \quad [15]$$

En este escenario, se define una función que busca reflejar la decisión de la cobertura o del rechazo del tratamiento por parte del financiador (la definimos como WTC, por las siglas en inglés de *willingness to cover* o intención de cobertura). En la compañía se tiene una herramienta conocida como semáforo de acceso donde para cada financiador del país se tiene 1 o 0 en caso de que se cubra o no se cubra el medicamento para una indicación específica.

5. En caso de aprobar el producto, cada financiador consume una cantidad $x_{ij} \geq 0$ y $x_{ij} \leq x_{ij}^{MAX}$, donde x_{ij}^{MAX} representa la incidencia máxima estimada del producto j para el financiador i

$$0 \leq x_{ij} \leq x_{ij}^{MAX} \quad [16]$$

6. El laboratorio farmacéutico debe satisfacer un precio promedio de comercialización de sus productos tal que se cumpla la siguiente restricción:

³⁸ En Argentina, los financiadores podrían verse obligados a cubrir a ciertos pacientes mediante recursos judiciales que los obliguen al precio full. Estas excepciones no se consideran a los fines del análisis aunque se podría trabajar en definir una referencia del riesgo de amparo para el financiador i para la patología tratada por el medicamento j donde, cuando no hay aprobación de parte del financiador, podría darse la cobertura del tratamiento a cierta probabilidad r_i o r_{ij} (dependiente del producto j o del producto-financiador ij) que refleja la obligatoriedad judicial de la cobertura. El precio pagado por parte del financiador es el precio de lista del producto sin descuentos asociados. Se esboza un posible planteo de dicha solución que, sin embargo, por no tener fuentes disponibles en la compañía que permitan identificarlo, sería de difícil utilización en un modelo inicial.

$$x_{ij} = \begin{cases} p_{ij}^{LIST} * r_j * x_{ij}^{MAX}, & wtc_{ij} < p_{ij} \\ D(x_{ij}), & wtc_{ij} \geq p_{ij} \end{cases}$$

$$\frac{\sum_{i=1}^n x_{ij} * p_{ij}}{\sum_{i=1}^n x_{ij}} \geq FLOOR_j \forall j \in (1, m) \quad [17]$$

7. El laboratorio farmacéutico debe satisfacer un precio mínimo de comercialización de sus productos que implique no negatividad del precio y cumplimiento de leyes de competencia como consecuencia de no incurrir en ventas que sean inferiores al costo asociado de proveer un producto j para el laboratorio.

$$p_{ij} \geq p_j^{MIN} \forall j \in (1, m) \quad [18]$$

8. El laboratorio farmacéutico debe satisfacer sus costos variables y fijos en el agregado de los productos. En ese sentido, se debe satisfacer que el ingreso obtenido por el financiador, resultado de la cantidad vendida a cada precio específico debe ser igual o mayor que la satisfacción del costo variable de cada uno de los productos (c_j) y del costo fijo del laboratorio (C_{fix}):

$$\sum_{i=1, j=1}^{n, m} x_{ij} p_{ij} \geq \sum_{i=1, j=1}^{n, m} c_j * x_{ij} + C_{fix} \quad [19]$$

9. Se propone establecer un precio máximo para los medicamentos que sea compatible con el presupuesto disponible del financiador. Para cada producto bajo análisis se establece que el consumo agregado del producto para el total de pacientes pasibles, no puede superar cierto umbral de ingreso para el financiador en análisis. Para respetarlo, se establece que cada financiador tiene un límite máximo de gasto presupuestario por producto que equivale a K_{ij} . Dicho monto es equivalente, en función al producto j , a k veces el ingreso promedio por afiliado en caso de querer cubrir al 100% de su población potencial. La variable k_{ij} depende del producto j en particular para el que se realiza el análisis y del financiador en particular. Se define dicho valor k_j como el diferencial entre el costo actual estimado para el financiador que incurre en los cuidados vinculados a la patología j sumado a un valor incremental definido a partir de la percepción del producto en la patología (construido internamente por el equipo en función de estrategia y de información de la compañía).

$$x_{ij} * p_{ij} \leq K_{ij} \quad [20]$$

$$K_{ij} = x_{ij}^{MAX} * k_j * ingreso.afiliado_{ij} \quad [21]$$

10. La función de demanda para cada subgrupo de financiadores identificados es una función que depende del precio p_{ij} , de la elasticidad y de una variable que no forma parte del problema de optimización de este modelo donde se refleja la intención de prescripción del médico y la intención de ser tratado por el paciente con ese medicamento (que en general no influye como consecuencia de las mencionadas fallas de mercado en etapas previas).

En esta función de demanda, se debería reflejar que la demanda es decreciente en función del precio, pero que ese decrecimiento es más alto para aquellos casos donde los financiadores tienen centros propios que en aquellos que no lo tienen. Asimismo, que dicha demanda es decreciente en función de la edad promedio de los afiliados (el beneficio esperado para los pacientes es menor como consecuencia de tener menor cantidad de años

de vida estimados). A su vez, se podría reflejar que existe un componente que no forma parte de lo que el laboratorio puede controlar con el precio que es la prescripción del médico. En este caso, se podría utilizar una función aleatoria de probabilidad que sea inicialmente similar para todos los financiadores, pero que con el correr del tiempo y la utilización del modelo diferencial de precios y otras estrategias, permita entender en mejor medida distintas variables que definen la probabilidad de prescripción. Se propone un principio de solución a continuación.

Se plantean dos funciones que implican la cobertura y la demanda del medicamento por parte del financiador para cubrir a los pacientes. Por un lado, tal como ha sido descrito en el punto 4, se construye una función de demanda de tipo logística donde se define a 1 como la decisión de cobertura del tratamiento y a 0 como definición de rechazo. Dicha función se construye a partir de las distintas variables analizadas en el modelo de segmentación (Dimensión, Edad, Share Femenino, Ingresos por afiliado, Centros propios).

Luego, se construye una segunda y subsecuente función aleatoria que define la decisión de prescripción de parte del médico de la receta que permita la cobertura del tratamiento para un paciente en particular. Se asume en este trabajo que la definición en la posible prescripción del médico es solo posible cuando existe acceso al tratamiento por parte del financiador (indicación 1 en semáforo de acceso).

Luego, en caso de existir el acceso posible en el financiador, se construye una variable donde hay una probabilidad de prescripción del médico:

$$prob.presc_{ij} = f(p_{ij}, n_{comp}, años.merc, centros.prop, diagnóstico, Región, \dots) \quad [22]$$

Esa es una variable que con el status de información actual no es posible de construir fehacientemente y que, por otro lado, no es parte del objetivo del trabajo. Se podría plantear dicha función como un posible próximo paso en el desarrollo del modelo para definir los precios óptimos. En este caso, se propone incorporar en un primer modelo, la cuota de mercado promedio para el financiador o segmento de financiador i para el que se esté analizando. Dicha variable es estimada como una probabilidad que es dato (corresponde a un análisis estimado en función de la información propia de la compañía sobre la venta y algunos informes de mercado).

En definitiva, una posible propuesta para reflejar la función de demanda que enfrenta el laboratorio, se podría reflejar la existencia de una parte de la demanda que no está vinculada a la decisión del precio y está relacionada con la prescripción y, por otro lado, se define una función logística para reflejar la probabilidad de la definición de cobertura por parte del financiador i cuando enfrenta cierto precio específico para el producto j (p_{ij})

$$D_{ij}(p_{ij}) = WTC_{ij}(p_{ij}) * prob.presc_{ij} * x_{ij}^{MAX} \quad [23]$$

$$WTC_{ij}(p_{ij}) = \frac{K e^{-(a+bx+\dots+jx)}}{(1 + e^{-(a+bx+\dots+jx)})^2} \quad [24]$$

Estas funciones de demanda permiten reflejar de un modo alternativo, la idea esbozada por Danzon et al. (2015). Se define una función de medida de la elasticidad E donde los precios varían inversamente con el nivel de elasticidad. La elasticidad para el producto j es igual a la elasticidad precio compensada ϵ_j sumado a la elasticidad ingreso (η) de la demanda multiplicada, expresada en la variable de WTC, por la cuota específica (s) del producto j que refleja la variable de probabilidad de prescripción indicada:

$$E_j = \epsilon_j + s_j \eta_j \quad [12]$$

11. Por simplificación, se ha optado por obviar en el modelo de optimización lineal presentado a la restricción de stock que tiene el laboratorio, generada como consecuencia de que, en su decisión, el laboratorio farmacéutico debe definir la cantidad esperada de venta de viales en el próximo periodo como consecuencia de que el periodo de importación de los productos puede demorar hasta 6 meses. Se podría plantear una mejora al modelo incorporando la posible limitación de stock del laboratorio farmacéutico como consecuencia de la generación de una proyección de demanda errónea en el periodo previo a la efectiva tenencia del producto para ser comercializado o a la falta de ventas en el periodo de vigencia del medicamento. En tal sentido, el error en el cálculo esperado de venta del producto para cada financiador, podría implicar que los costos de importación no se vean satisfechos (generando una transformación en la ecuación planteada en la ecuación 15). Asimismo, los productos farmacéuticos tienen una fecha de vencimiento. Por lo tanto, se podría incluir en la restricción, la posible pérdida económica que se generaría en caso de acumular stock y que deban destruirse.

Para reflejar dichas restricciones, se podrían incluir variables de temporalidad en el modelo. Es decir, en el periodo 0, la compañía define cuántas unidades espera vender. En el periodo 1, en función a la demanda (que depende de la cobertura y por lo tanto de los precios) vende los productos. Por último, el stock residual que le queda, en caso de que eso suceda, debe destruirse y tiene aparejado un costo. Asimismo, en caso de que la previsión de unidades sea insuficiente en el periodo 0, el laboratorio no puede abastecer a toda su demanda.

En conclusión, se presenta el siguiente modelo a optimizar.

Función objetivo

$$\max \sum_{i=(1,n)}^n \sum_{j=(1,m)}^m p_{ij} * x_{ij} \quad [14]$$

Sujeto a:

Restricción del negocio del laboratorio farmacéutico

$$\frac{\sum_{i=1}^n x_{ij} * p_{ij}}{\sum_{i=1}^n x_{ij}} \geq FLOOR_j \quad \forall j \in (1, m) \quad [17]$$

$$p_{ij} \geq p_j^{MIN} \quad \forall j \in (1, m) \quad [18]$$

$$\sum_{i=1, j=1}^{n, m} x_{ij} p_{ij} \geq \sum_{i=1, j=1}^{n, m} c_j * x_{ij} + C_{fix} \quad [19]$$

Restricción de cobertura y demanda

$$x_{ij} = \begin{cases} 0, & wtc_{ij} < p_{ij} \\ D(x_{ij}), & wtc_{ij} \geq p_{ij} \end{cases} \quad [15]$$

$$D_{ij}(p_{ij}) = WTC_{ij}(p_{ij}) * prob. presc_{ij} * x_{ij}^{MAX} \quad [23]$$

En la ecuación 10 se incorpora a la ecuación 9 que modela la probabilidad de prescripción y en la variable de WTC se incorpora la variable de elasticidad 12.

$$WTC_{ij}(p_{ij}) = \frac{K e^{-(a+bx+\dots+jx)}}{(1 + e^{-(a+bx+\dots+jx)})^2} \quad [24]$$

Restricción de capacidad de pago

$$x_{ij} * p_{ij} \leq K_{ij} \quad [20]$$

$$K_{ij} = x_{ij}^{MAX} * k_j * ingreso.afiliado_{ij} \quad [21]$$

Restricción de cantidad de pacientes potenciales por segmento

$$0 \leq x_{ij} \leq x_{ij}^{MAX} \quad [16]$$

Restricción de no negatividad

En este caso las dos variables de decisión ya se encuentra aseguradas como valores positivos ya que se indicó que $p_{ij} \geq p_j^{MIN} \forall j \in (1, m)$ y que $0 \leq x_{ij}$

En tal caso, según de qué patología j se esté tratando, se analiza la prevalencia esperada para ese subgrupo de financiadores (basado en el segmento de mercado asignado en la etapa anterior), por otro lado, en la función de WTC se propone incorporar las variables de centros propios, edad promedio afiliados, ingresos promedios afiliados.

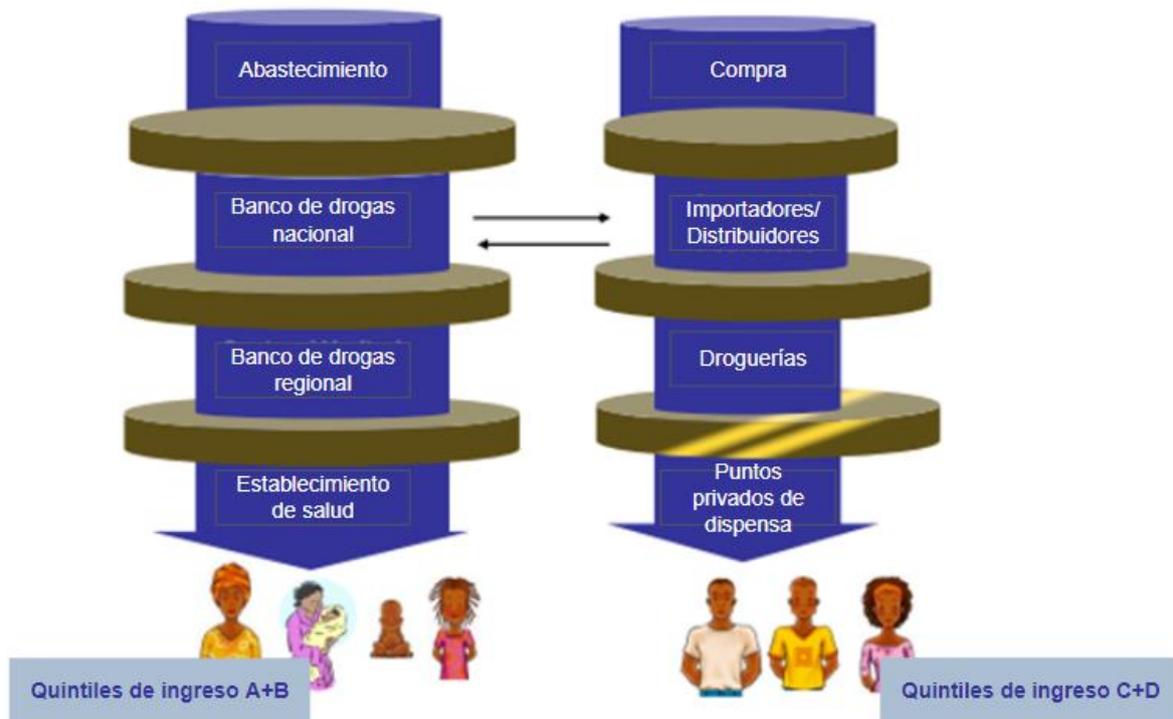
VII. Recomendaciones de implementación del modelo

En este trabajo se ha buscado establecer una estrategia de segmentación de financiadores en función de ciertas variables que permiten diferenciarlos y con esto se ha esbozado una propuesta de implementación de un modelo de precios diferenciales que contemple dichas variables. Se ha propuesto establecer, para esos grupos en particular, un modelo que estime el precio óptimo a enfrentar por cada subgrupo de financiadores.

Resulta relevante hacer breves recomendaciones sobre la posible implementación de esta estrategia. En primer lugar, es importante entender si podrían existir potencialmente dificultades para implementar, en términos legales, un modelo como el que se propone. En este sentido, se ha identificado que la ley Robinson-Patman, esboza distintas cuestiones en torno a la implementación de una diferenciación de precios. Por otro lado, tiene algunas particularidades en lo que respecta a su ámbito de aplicación. La ley prohíbe la discriminación de precios en mercados de bienes tangibles, cuando dicha discriminación involucra productos de similar grado y calidad. Además, la discriminación que esta ley reprime es aquella que limita sustancialmente la competencia o tiende a crear un monopolio, y no aquella que representa de por sí un abuso o da una señal sobre la existencia de una posición dominante. Esto es así porque la ley sólo admite explícitamente tres defensas: el cambio en las condiciones de mercado (que justifica las diferencias de precios entre distintos momentos del tiempo), las diferencias en el costo de provisión de las distintas unidades, y las reducciones de precios destinadas a igualar los precios cobrados por los competidores. En tal sentido, en relación al modelo propuesto, no se identifican conflictos legales en cuanto a la posible implementación de este modelo. En principio, se identifica que las condiciones de mercado son diferentes para los distintos financiadores, el riesgo de padecer ciertas patologías y de enfrentar ciertos costos asociados a su población, es diferente y, por lo tanto, en términos de buscar garantizar acceso sostenible a los tratamientos, tendría sentido proponer este tipo de implementación de esta estrategia.

En términos de otros riesgos que podrían surgir en torno a la implementación de un modelo de precios diferenciales, consiste en la dificultad de diferenciar los canales a los que se proveen los productos (Yadav, 2010). En la figura 43, se representa una imagen que el autor señala en su trabajo y donde se indica que es fundamental que un producto no pueda fluir de un canal al otro.

Figura 43: Gráfico extraído de Yadav (2010) donde se muestra la posibilidad de tener canales individualizables en la implementación de una estrategia de precios diferenciales para distintos niveles socioeconómicos.



Fuente: Extraído y traducido de Yadav (2010) simplemente para ilustrar propuesta de diferenciación de estrategia en canales³⁹

En tal sentido, en la implementación en particular de este modelo para el caso a analizar, se considera que es posible identificar correctamente el canal mediante el que se entregan los beneficios a cada financiador ya que dichos beneficios son entregados o bien directamente para los clientes directos, o a través de la herramienta identificada donde es fundamental la carga de información fehaciente del financiador y del usuario final del tratamiento (anonimizado, de todos modos).

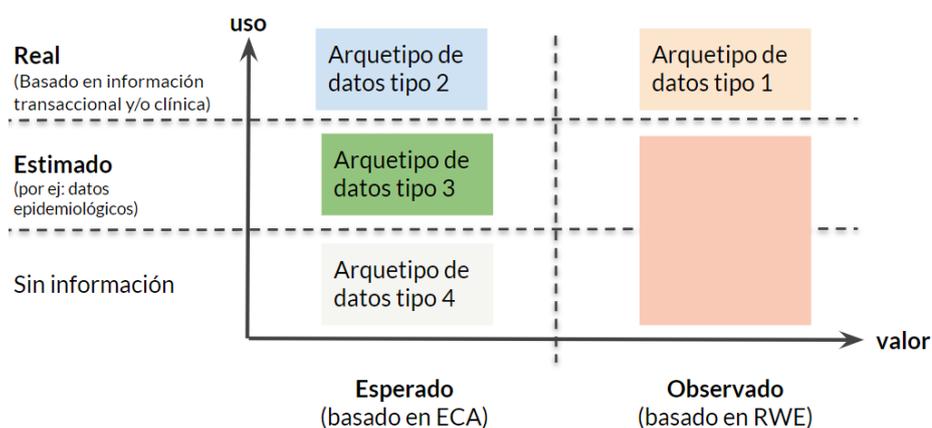
Otra consideración relevante, está vinculada a si efectivamente a través de la estrategia recomendada en este trabajo, se podría llegar a un modelo diferencial basado en valor para la determinación de los precios de los medicamentos para cada financiador. La respuesta frente a ese cuestionamiento es que muy difícilmente se llegue a tal situación. Sin embargo, el planteo de este trabajo permite esbozar un camino hacia tal situación. En tal sentido, tal como muestran en su trabajo Cole et al (2021), existen distintos arquetipos de datos frente a la definición de un modelo de precios diferenciales basados en valor para cada indicación en particular de un tratamiento (y, lógicamente, para cada financiador). En la figura 44 se presentan esos arquetipos. Conceptualmente,

³⁹ Yadav, P. (2010). Differential Pricing for Pharmaceuticals. MIT Zaragoza International Logistics Program.

en el arquetipo 1 el nivel de información es nulo para la toma de decisiones y, cuando se incorporan distintas capas de información se avanza hacia niveles donde se refleja en mejor medida el valor en la definición de los precios. En consecuencia, en función del nivel de información disponible de la utilización de un producto para un paciente por parte de un prestador de salud, se pueden ejecutar distintos modelos de financiamiento que reflejen el valor en la definición del precio pagado según diversos arquetipos de datos (Cole et al., 2021).

En el planteo realizado en este trabajo, podría argumentarse que, gracias a la construcción y sistematización de los datos, se podría ejecutar un arquetipo de datos 3 donde la información teórica de cada financiador es incluida en la definición de los descuentos asignados para cada producto. Podría trabajarse en un modelo que refleje para poblaciones donde haya mayor información en la vida real sobre los pacientes (arquetipos 2 y 1) y con ellos se podrían definir estrategias más precisas aún para cada financiador y que permitan garantizar los conceptos de sustentabilidad presentados previamente en el trabajo.

Figura 44: Arquetipos de información para la implementación de una estrategia diferencial de precios para productos farmacéuticos⁴⁰



Fuente: Traducción y reinterpretación de Cole et al. (2021)⁴¹

Es relevante considerar que, pensando en una posible implementación de este modelo, es importante contar con información actualizada de los financiadores de salud. En tal sentido, se debe trabajar en enriquecer las bases de afiliados con el fin de entender la evolución en la cantidad de afiliados por financiador, así como también la evolución de los ingresos económicos entre distintos periodos. Se debe considerar que el tiempo incurrido en un recurso en la compañía es una variable a considerar para implementar una actualización a este modelo.

Por otro lado, existen mejoras en la estimación de los precios óptimos a ofrecer que podría construirse en modelos de cálculo de la probabilidad de que un paciente realice un amparo para acceder al tratamiento, un modelo que permita entender las variables que definen y que influyen en la prescripción por parte del médico de cierto tratamiento. Por último, podría trabajarse en relación a otros modelos alternativos de demanda para analizar la decisión por parte de los financiadores que incorporen, por ejemplo, costos asociados a la cobertura de la población en mayor profundidad.

⁴⁰ ECA: Estudio clínico aleatorizado y RWE: Evidencia del mundo real (por las siglas en inglés de *real world evidence*)

⁴¹ Cole, A., Neri, M., & Cookson, G. (2021). Expert Consensus Programme: Payment Models for Multi-Indication Therapies. OHE Consulting Report. Office of Health Economics.

En relación al plan de monitoreo y mantenimiento de la información utilizada en este modelo, se pretende plantear una propuesta de actualización cada 6 meses donde se ejecuten los pedidos de información de datos para actualizarlo. Asimismo, se pretende actualizar la estimación de los segmentos con el objetivo de analizar si el número total de pacientes que reciben el tratamiento efectivamente se ve incrementado, tal como busca este proyecto. Con la disponibilidad continua de los datos del sistema de salud, se busca incorporar análisis de series temporales para entender la posible evolución de los financiadores en el tiempo y si eso influye o no en las decisiones de cobertura.

En tal sentido, se proponen en las conclusiones y próximos pasos, una posible mejora de las variables que se puedan incorporar a futuro en este análisis.

4. Conclusiones y próximos pasos

A partir de entender en profundidad el negocio, las herramientas disponibles y las distintas metodologías de análisis de variables, se han esbozado posibles estrategias de desarrollo de un modelo que refleje una transformación en la toma de decisiones sobre el precio de comercialización de productos farmacéuticos de alto precio para una compañía farmacéutica, a los distintos financiadores de salud en Argentina.

El principal desafío a resolver es la inequidad en el acceso a los tratamientos innovadores. En este sentido, se propuso implementar una segmentación de financiadores de salud con el fin de, oportunamente, establecer una diferenciación de precios que refleje la capacidad y voluntad de pago a partir de distintas variables (cantidad de centros propios, Share Femenino, Región principal, Dimensión -tamaño-, Edad, Ingreso, entre otras). Se han propuesto dos estrategias posibles de segmentación que generan una anchura de silueta prometedor de 0.40, en promedio.

Luego, para presentar una posible estrategia de implementación del modelo de segmentación para los financiadores, se propuso realizar un sencillo ejercicio a nivel teórico de optimización lineal entera que permita entender una posible modelización del negocio con el objetivo de definir la política óptima de asignación de descuentos (establecidos para reflejar una política de fijación de precios diferenciales) a los distintos grupos identificados a partir de la segmentación.

Con el objetivo de abordar el problema de negocio, se ha utilizado un marco para estructurar el trabajo conocido como CRISP DM, que es muy utilizado en la minería de datos en las industrias. En tal sentido, se ha estructurado este trabajo siguiendo la mencionada metodología en la estructura de los distintos capítulos: Comprensión del negocio, Exploración y Preparación de los datos, Modelado, Evaluación de resultados y Plan de implementación.

En la etapa de comprensión del negocio, se reflejaron los objetivos específicos del negocio farmacéutico de medicamentos de alto precio y se presentaron sus fallas de mercado. Luego, se enmarcó al sistema de salud argentino como mercado relevante y se presentó brevemente la evolución en el último tiempo del mercado representativo impulsada por la transición demográfica y epidemiológica de la población y la evolución y actualidad del negocio de la innovación y los medicamentos de alto precio. Asimismo, se presentó el proceso resumido de incorporación de este tipo de tecnologías en Argentina. Con este análisis, se considera que se ha generado un mapa con las posibles variables de decisión. Asimismo, dicha etapa fue fundamental para definir la recolección y sistematización de datos disponibles en múltiples fuentes que se utilizan en las etapas siguientes.

En el capítulo de Exploración y Preparación de datos, se buscó cumplir las tareas que se proponen en el modelo de minería de datos utilizado: se reportan datos iniciales (resumen en anexo 2), se hace análisis exploratorio de los datos (EDA por las siglas en inglés de *Exploratory data analysis*), y se hace un análisis y depuración de los datos con el objetivo de garantizar calidad en la información (se incluye un resumen de los datos iniciales con correcciones implementadas por el autor en el Anexo 2). Por otro lado, se construyen derivaciones de las variables, a través de técnicas como componentes principales. Estas variables se integran y se aplican transformaciones a través de la escalera de Tukey (1977), mediante función logarítmica, a múltiples variables.

En el capítulo siguiente, Modelado, se presentan los modelos a testear y se definen las métricas de rendimiento y selección de parámetros. En este caso, en el trabajo se desarrolla una construcción analítica de la posible implementación, a través de un modelo de optimización lineal, de los grupos de segmentos obtenidos a partir del trabajo de clusterización. Se presentan brevemente en esta sección los objetivos de acceso que se buscan generar y la disponibilidad de información en la literatura sobre modelos que permitan la diferenciación de precios de medicamentos para garantizar un acceso a más población. Se argumenta que, a través de dichos precios diferenciales, se podría alcanzar la definición de un precio basado en el valor que le aporta a cierto afiliado que está asociado a cierto financiador (o segmento de financiadores, en función del análisis de *clustering* realizado).

En la siguiente sección, de evaluación de resultados, se ponen en práctica los distintos modelos y se definió determinar dos estrategias de segmentación de financiadores: una para patologías más asociadas a la edad de la población y la otra para patologías de baja incidencia. En ambos casos, se utilizan tres clústeres de financiadores y se opta por dejar por fuera de estos grupos a: PAMI y la población con cobertura pública exclusiva de la Provincia de Buenos Aires⁴². En el primer modelo, se utiliza *clustering* jerárquico y en el segundo, *kmeans*.

Por último, en la etapa de Implementación, se ha buscado proponer un posible escenario de implementación de la estrategia, a través de un modelo de optimización lineal entera. Su resolución se considera un próximo paso de este proyecto. Su objetivo, como se refleja en las funciones propuestas, es garantizar que se permita el mayor acceso posible a los medicamentos, con cuidado del gasto económico de cada subgrupo de financiadores y la maximización de la rentabilidad del laboratorio farmacéutico. Entre otras variables orientadas a garantizar un concepto de equidad y máximo acceso, se incorporan: potencial máximo de pacientes a cubrir por financiador, precio máximo determinado como un límite porcentual del gasto presupuestario de un financiador⁴³, la necesidad de satisfacer, idealmente, la variable de *WTC* (por las siglas en inglés de *Willigness to cover*, es decir, voluntad de cobertura), que es la que define la probabilidad de cobertura por parte del financiador y, por lo tanto, la probabilidad de acceso al tratamiento en caso de que exista la prescripción; por último, se refleja la incertidumbre asociada a la otra gran etapa relevante en la definición del acceso al tratamiento, que corresponde a la prescripción médica donde se propone una función de demanda del producto que no necesariamente depende netamente del precio.

Es importante mencionar algunas de las limitaciones principales que se identifican de este trabajo y que podrían ser posibles caminos para mejorar las decisiones estratégicas. Por un lado, como ha sido indicado, es muy difícil determinar efectivamente la voluntad de pago por tener un truncamiento significativo de la información efectiva sobre cobertura de tratamientos. En tal sentido, los datos

⁴² Estimado en función de datos provistos por Ministerio de Salud en base a padrón de beneficiarios del Programa Nacional SUMAR (2022). Datos obtenidos mediante consulta de acceso a la información pública. Ver anexo 1.

⁴³ Que podría reflejar criterios de equidad en términos de costo, riesgo de tener población prevalente, ingreso total del financiador, otros costos relevantes del financiador estimados en función de la población cubierta.

sobre la venta de la que se tiene información es muy parcial y podría ser trabajada para entender si efectivamente refleja lo que se analiza en la etapa de exploración de esa variable.

Por otro lado, es relevante indicar que muchos de los datos fueron construidos en base a supuestos y estimaciones, con fuentes de dispar fecha de origen⁴⁴. En consecuencia, es importante que, en caso de querer utilizar los datos obtenidos, se suavice su uso a través de herramientas que permitan mejorar las estimaciones. Asimismo, es clave garantizar su actualización.

Adicionalmente, hay diversas bases de datos que podrían sumarse al análisis con el fin de entender si permiten mejorar las conclusiones. En este sentido, algunas posibles son: base de diagnóstico de VIH por provincia y nivel de ingresos, tasas de diagnóstico por provincia, encuestas de nutrición nacional, bases de recetas de medicamentos gestionados por el sector público, información sobre amparos en salud. Por otro lado, se podría profundizar en el análisis geográfico de cobertura, buscando identificar tasas de diagnóstico, y que, con ello, se permita relevar las chances de encontrar casos de pacientes en ciertos financiadores.

Por otro lado, podría tener sentido incorporar información que permita reflejar de una mejor manera la evolución en las capacidades de los financiadores, la volatilidad en la cobertura de sus poblaciones posibles, entre otras cuestiones. Por último, es sumamente relevante para la implementación de esta estrategia, pensar en la necesidad de entender las variables que definen la probabilidad de prescripción para un paciente potencial y, dicho modelo, podría colaborar en la generación de estrategias encaminadas a la mayor penetración de las tecnologías innovadoras, en condiciones de sustentabilidad para los financiadores. Asimismo, se podría incluir un análisis sobre la visita a médicos por parte de la compañía y la selección de médicos estratégicos a visitar para cada financiador.

Si bien se considera que la estrategia sugerida por este trabajo es insuficiente para garantizar un modelo diferencial basado en valor para la determinación de los precios de los medicamentos para cada financiador, se ha pretendido esbozar un camino hacia esa situación. La generación de un mecanismo innovador de asignación de descuentos podría implicar un acceso temprano y universal a los productos farmacéuticos innovadores y a su vez garantizar un nivel sustentable de retornos para la compañía farmacéutica. En conclusión, el objetivo que se ha buscado entender implicaría, potencialmente, un beneficio muy significativo tanto para la sociedad como para la compañía en análisis. Por lo tanto, comprende un primer paso en el camino de garantizar ese norte que está reflejado en la visión de la compañía para la que se realiza este trabajo. Es relevante aclarar, de todos modos, que los modelos desarrollados consisten en prototipos prescriptivos preliminares (*MVP*, por las siglas de *minimum viable product*) que dan una primera medida respecto a su implementación práctica. En este sentido, el proceso de sistematización y análisis de la información constituye una herramienta fundamental en la construcción de soluciones sostenibles para el sistema de gestión de la salud y la toma de decisión de cobertura.

Con el fin de implementar el proyecto detallado, los precios obtenidos se buscan aplicar a través de descuentos relevados como valores promedio a implementar para cada financiador. Éstos pretenden reflejar otro elemento que no ha sido profundizado en este trabajo que son las soluciones innovadoras de acceso a los medicamentos de alto precio. Gracias a la sistematización de la información realizada para los distintos tipos de financiadores identificados, se podrían generar modelos alternativos de pago que permitan generar mecanismos de capitado de pacientes, acuerdos de riesgo compartido, pagos por resultado, entre otros, donde se puedan prever con exactitud la cantidad de pacientes potenciales para cada financiador. En este sentido, tal como se argumenta en la etapa de implementación, la información construida permite obtener un arquetipo 3 que permite generar soluciones de ingreso gestionado que garanticen sustentabilidad en el acceso. Por otro lado,

⁴⁴ Remitirse al anexo 1 para un detalle sobre las fechas de corte utilizadas para el trabajo.

la información también permite colaborar en la previsibilidad en el gasto presupuestario que pudiera reflejarse en cada financiador, gracias a la estimación más precisa de la cantidad potencial de pacientes. Evidentemente el modelo de datos también es pasible de gran capacidad de mejora, pero se considera un progreso en cuanto a la sistematización de los datos en un sistema tan fragmentado, como es el de Argentina.

5. Referencias

Agencia Nacional de Discapacidad (2022). Padrón de beneficiarios del programa nacional INCLUIR Salud. Accedido mediante pedido de acceso a la información con datos a agosto de 2022.

Antonio, Y., Indratno, S., & Saputro, S. (2021). Pricing of cyber insurance premiums using a Markov-based dynamic model with clustering structure. *PLoS One*. 10.1371/journal.pone.0258867

Área Técnica de CA.DI.ME. (2022). Informe sectorial: Servicios de salud. 19. <https://www.cadime.com.ar/newsletter/Informe%20Sectorial%20CADIME%20N%2019%20-%20Junio%202022.pdf>

Bisang, R., Cetrángolo, O. (1997). Descentralización de los servicios de salud en la Argentina. Proyecto Regional sobre Financiamiento Público y Provisión de los Servicios Sociales: El rol del Estado en Educación básica y atención primaria de salud en America Latina y el Caribe. CEPAL. Chile.

Breidert, C. (2007). Estimation of Willingness-to-Pay: Theory, Measurement, Application. Deutscher Universitätsverlag.

Casado, D. (2000). El envejecimiento demográfico y la demanda futura de servicios sanitarios: ¿Más personas mayores? *Gestión Clínica y Sanitaria*, 2, 106. <http://iiss.es/gcs/gestion%20clinica%205.pdf>

Chalkidou, K., Claxton, K., & Yadav, P. (2020). Value-based tiered pricing for universal health coverage: an idea worth revisiting. *Gates open research*, 4(16). <https://doi.org/10.12688/gatesopenres.13110.3>

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R. (2000). CRISP-DM 1.0: Step-by-Step Data Mining Guide. Copenhagen: SPSS.

Chough, K. C., & de Leon, A. R. (Eds.). (2013). *Analysis of Mixed Data: Methods & Applications*. Taylor & Francis.

Cole, A., Neri, M., & Cookson, G. (2021). Expert Consensus Programme: Payment Models for Multi-Indication Therapies. OHE Consulting Report. Office of Health Economics.

Comisión Nacional de Defensa de la Competencia Argentina. (2019). "INVESTIGACIÓN DE MERCADO SOBRE LAS CONDICIONES DE COMPETENCIA EN EL MERCADO DE MEDICAMENTOS. Retrieved 2023, from https://www.argentina.gob.ar/sites/default/files/im._4_-_medicamentos_-_disposicion_y_anexo.pdf

CONADU. Escalas salariales de paritarias. Consultado en: <https://conadu.org.ar/grilla-salarial-paritaria-2021/> consultado el 5/04/2023

Daems, R., Ramani, S. V., & Maes, E. (2011). Global Framework for Differential Pricing of Pharmaceuticals (Vol. 54). United Nations University - Maastricht Economic and Social Research Institute on Innovation and Technology (MERIT). <https://ideas.repec.org/p/unm/unumer/2011054.html>

Danzon, P. M. (2018, Dec). Differential Pricing of Pharmaceuticals: Theory, Evidence and Emerging Issues. *Pharmacoeconomics*, 1395-1405. 10.1007/s40273-018-0696-4

Danzon, P. M., Towse, A., & Mestre Ferrandiz, J. (2015). Value-Based Differential Pricing: Efficient Prices for Drugs in a Global Context. *Health Economics*, 24(3), 294-301. <http://dx.doi.org/10.1002/hec.3021>

Färber, I., Günemann, S., Kriegel, H. P., & Kröger, P. (2010). On Using Class-Labels in Evaluation of Clusterings.

Farres, R. (2012). Optimal pricing models in B2B organizations. *Journal of Revenue and Pricing Management*, 11, 35–39. 10.1057/rpm.2011.36

Glanc M, & Del Prete S. (2021). Gestión del alto precio en salud en Argentina: un dilema no resuelto. Universidad ISALUD y FAMSA.

Global Cancer Observatory (Globocan) depende de la Organización Mundial de la Salud y publica datos de incidencia de la patología ajustada a rango etario para Argentina para el año 2020.

Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Reference Life Table. Seattle, United States of America: Institute for Health Metrics and Evaluation (IHME), 2021.

Haddix, A. C., Shaffer, P. A., & Duñet, D. O. (Eds.). (1996). *Prevention Effectiveness: A Guide to Decision Analysis and Economic Evaluation*. Oxford University Press.

Hadian, M., & Fattahi, M. R. (2019). Application of cluster analysis for healthcare financing: case of Iran. *BMC Health Services Research*, 143.

Hileman, G., Muzayan Mehmud, S., & Rosenberg, M. (2016). *Risk Scoring in Health Insurance: A primer*.

Informe de afiliados IOSFA (2022). Informe de afiliados de IOSFA elaborado con datos al 2 de junio de 2022.

IQVIA. (2022). *Global Trends in R&D: Overview through 2021*". IQVIA. <https://www.iqvia.com/insights/the-iqvia-institute/reports/global-trends-in-r-and-d-2022>

Jain, A. K, Dubes, R. (1948). *Algorithms for Clustering Data*. Michigan State University. Prentice Hall.

Johnson, K., & Kuhn, M. (2020). *Feature Engineering and Selection: A Practical Approach for Predictive Models*. CRC Press, Taylor & Francis Group.

Ley 23661 Argentina. Disponible en <http://servicios.infoleg.gob.ar/infolegInternet/anexos/0-4999/63/norma.htm>

Marin, G., Cañas, M., Marin, G, Marin, L. Nucher, D., Diaz P, D., Urtasun, M. (2022), Impacto económico de medicamentos de alto precio/costo en la seguridad social de Argentina. El caso del Instituto de Obra Social para las Fuerzas Armadas y de Seguridad. *Medicina*. Buenos Aires.

Ministerio de Hacienda (2018). *Industria Farmacéutica. Serie Informes de cadenas de valor, Año 3, N° 21, agosto 2018.* Disponible en: https://www.argentina.gob.ar/sites/default/files/sspmicro_cadenas_de_valor_farmacia_0.pdf

Ministerio de Salud Argentina (2021). *Informe de Indicadores básicos 2020 en Salud*. Disponible en: https://www.argentina.gob.ar/sites/default/files/indicadores_basicos_2020.pdf

Ministerio de Salud Argentina (2022). Mortalidad en personas mayores durante el periodo 2010-2019. Disponible en: <https://www.argentina.gob.ar/noticias/salud-presento-analisis-sobre-mortalidad-en-personas-mayores-durante-el-periodo-2010-2019>

Ministerio de Salud Argentina (2023). Informe de Coberturas de Salud en Argentina. Año 2022. Secretaría de Equidad. Ministerio de Salud.

Ministerio de Salud Argentina (2022). PUCO (Padrón Único Consolidado Operativo). Accedido mediante consulta de acceso a la información pública.

Ministerio de Salud (2022). Consulta sistema REFES. Registro Federal de Establecimientos de Salud. https://sisa.msal.gov.ar/sisadoc/docs/050101/refes_canales.jsp

Ministerio de Salud y Desarrollo Social Argentina, CONETEC Manual Operativo de Estructura y Funcionamiento: Documento Técnico número 1. disponible en <https://www.argentina.gob.ar/sites/default/files/1-manual-operativo-conetec.pdf>

Mobula, L., Sarfo, S., Arthur, L., Burnham, G., Plange-Rhule, J., Ansong, D., Gavor, E., & Ofori-Adjei, D. (2018). A multi-center prospective cohort study to evaluate the effect of differential pricing and health systems strengthening on access to medicines and management of hypertension and diabetes in Ghana: A study protocol. *Gates Open Res*, 2(2018). 10.12688/gatesopenres.12797.2.

Montgomery, D. C., & Peck, E. A. (2018). *Introduction to linear regression analysis*. John Wiley & Sons.

Mwachofi, A., & Al-Assaf, A. F. (2011). Health care market deviations from the ideal market. *Sultan Qaboos University medical journal*, 11(3), 328–337

Nnoaham, K.E., & Cann, K.F. (2020). Can cluster analyses of linked healthcare data identify unique population segments in a general practice-registered population? *BMC Public Health*, 20(798). <https://doi.org/10.1186/s12889-020-08930-z>

Organización Panamericana de Salud (2009). El acceso a los medicamentos de alto costo en las Américas: contexto, desafíos y perspectivas. Serie técnica número 1. Washington.

Paluzzi, J. E., Leach, B., Munderi, P., & UN Millennium Project. Task Force on HIV/AIDS, Malaria, TB, and Access to Essential Medicines. Working Group on Access to Essential Medicines. (2005). Prescription for healthy development : increasing access to medicines / lead authors, Beryl Leach, Joan E. Paluzzi, Paula Munderi ; UN Millennium Project, Task Force on HIV/AIDS, Malaria, TB, and Access to Essential Medicines. Working Group on Access to Es. United Nations Development Programme. Retrieved 2023, from <https://digitallibrary.un.org/record/570036?ln=en>

Panageas, K. S. (2015). Clinical trial design for rare cancers - why a less conventional route may be required. NCBI. Retrieved March 31, 2023, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4724195/>

Pastorino, R., De Vito, C., Migliara, G., & Glocker, K. (2019). Benefits and challenges of Big Data in healthcare: an overview of the European initiatives. In *European journal of public health* (pp. 23-27). European Public Health Association. 10.1093/eurpub/ckz168

Patel, N. (2017). Fee-for-value in the pharmaceutical industry: a policy framework applying data science to negotiate drug prices. *Journal of law and the biosciences*, 4, 205-215. 10.1093/jlb/lx002

Portal de Datos Abiertos PAMI. Disponible en <https://datos.pami.org.ar/>

Roberts, J. (1998). Primary care: core values. Primary care in an imperfect market. BMJ. Jul 18; doi: 10.1136/bmj.317.7152.186. PMID: 9665903; PMCID: PMC1113538., 317(7152):186-9.

Secretaría de Derechos Humanos y pluralismo cultural (2016), “Pacto Internacional De Derechos Económicos, Sociales Y Culturales”. Disponible en: https://www.argentina.gob.ar/sites/default/files/derechoshumanos_publicaciones_colecciondebolsillo_07_derechos_economicos_sociales_culturales.pdf

Superintendencia de Servicios de Salud (2022). Padrón de afiliados a obras sociales nacionales. Accedido por un pedido de acceso a la información

Superintendencia de Servicios de Salud (2022). Padrón de afiliados de empresas de medicina prepaga. Accedido por un pedido de acceso a la información con fecha diciembre de 2021

Tobar, F. (2014). Módulo 1: Salud y cuestión social. En Chiara, M. (Compiladora). Cuaderno de Trabajo: Gestión Territorial Integrada para el Sector Salud. Instituto del Conurbano. Universidad Nacional de General Sarmiento.

Torres, R., Jorgensen, N., Robba, M. (2020). Mitos y realidades de las obras sociales. Universidad ISalud. Buenos Aires. Argentina.

Tukey, J. W. (1977). Exploratory Data Analysis. Reading, Massachusetts: Addison-Wesley

Van Ryzin, G., van Ryzin, G. J., & Talluri, K. T. (2005). The Theory and Practice of Revenue Management. Springer.

Yadav, P. (2010). Differential Pricing for Pharmaceuticals. MIT Zaragoza International Logistics Program.

Anexo 1: Construcción de datos de Financiadores

Fecha estimada de datos utilizados:

- Obras Sociales Nacionales. Consulta acceso información pública a marzo 2022.
- Empresas de Medicina Prepaga. Consulta acceso información pública a diciembre 2021.
- INSSJP. Publicado en página web del organismo. Marzo 2022.
- Obras Sociales Provinciales. Construido a partir de datos PUCO (agosto 2022) y utilizando distribución demográfica de diversos cortes temporales a partir de investigación en páginas web donde se utilizaron informes institucionales de gestión o publicaciones en diarios locales, nacionales, COSSPRA, entre otros. Por último, a partir de algunos modelos estimados, se extrapólo esa distribución al resto.
- Obras Sociales Especiales: Se utiliza información Cobertura Salud en Argentina obtenido de la Dirección de Estrategias de Financiamiento y Dirección de Estrategias de Integración de Coberturas (2022). Se relevó informe institucional de IOSFA a diciembre 2021, algunas obras sociales universitarias con informes dispares de población cubierta, el análisis de Población universitaria publicado en el año 2010, entre otras fuentes con el fin de estimar la posible distribución demográfica de esos financiadores.
- Población con cobertura pública exclusiva: Se utiliza información provista por la Secretaría de Equidad en Salud, a través de la Dirección Nacional de Fortalecimiento de los Sistemas

Provinciales, que corresponde a la población inscripta en el Programa Nacional SUMAR. Se accede mediante consulta de acceso a la información pública y los datos corresponden a agosto del año 2022.

- Población beneficiaria INCLUIR Salud. Se accede a padrón informado por la Agencia Nacional de Discapacidad mediante un pedido de acceso a la información con datos actualizados a agosto de 2022 (EX-2022-84660323- -APN-DNAIP#AAIP).

Se explica brevemente la metodología de construcción de las bases de datos utilizadas en el modelo.

Entidades reguladas por la Superintendencia de Servicios de Salud (Obras Sociales Nacionales y Empresas de Medicina Prepaga):

Fuentes de datos y supuestos:

- La población por rango etario se alimenta de una base provista por la Superintendencia de Servicios de Salud en una Consulta de acceso a la información pública de expediente: EX - 2022 - 71779229 - DNAIP#AAIP con notificación de respuesta 20/07/2022.
- Dicha información está alimentada a partir de dos fuentes distintas:
 - Por un lado, la información población de las Obras Sociales Nacionales se construye a partir del Padrón de Opciones de Beneficiarios en Relación de Dependencia. Dicho padrón agrupa a todas las personas que trabajan en relación de dependencia y que aportan a una Obra Social Nacional. Es importante mencionar que en muchos casos las obras sociales nacionales derivan los aportes de sus afiliados a Empresas de Medicina Prepaga
 - La información de las Empresas de Medicina Prepaga se alimenta de una base que se construye a partir de la información presentada por las Entidades de Medicina Prepaga inscriptas en forma definitiva o provisoria en el RNEMP. Las entidades deben informar a la Superintendencia sus afiliados, pero, tal como se ha observado con diversos El conflicto que se ha observado con estos datos es que las distintas empresas no presentan de una manera uniforme los datos propios como consecuencia de la coexistencia de dos fuentes de financiamiento principales para estas entidades. Por un lado, existen algunas que informan el total a quienes proveen sus servicios (ya sea a través de desregulados de Obras Sociales Nacionales o afiliados directos a la Empresa de Medicina Prepaga) y por otro, hay algunos que sólo informan los afiliados directos a la entidad. En tal caso, se definió un tipo de financiadores Privado-Sindical donde se sumó a los afiliados por obra social que forman parte de ese financiador. Sin embargo, existe cierta incertidumbre de que haya otros financiadores que no se identificaron oportunamente y que por lo tanto podría implicar sub representación o sobrerrepresentación de algún financiador. Para corregir los posibles desbalances en dicha información se utilizó la información provista por una consultoría comprada por la compañía a una consultora que, sin embargo, tiene información sólo para un número limitado de financiadores. Sin embargo, asumimos que, en el caso de aquellos prestadores grandes, se pudieron identificar y corregir algunas de las disparidades.
- **Para la construcción de los ingresos anuales de los financiadores se utilizó el siguiente criterio:**

En caso de Obras Sociales Nacionales:

- Se sumó el total de dinero transferido por la AFIP que está informado en el Boletín Estadístico Anual que publica el organismo. En particular se utilizó la información correspondiente al año 2021 en las transferencias a las distintas Obras Sociales Nacionales, en concepto de lo que dicha entidad recolecta a través de los aportes y contribuciones salariales de los empleados en relación de dependencia. Esta información se encuentra disponible para el público acceso a través del Boletín Anual de Seguridad Social, en el cuadro 10 de la publicación, donde se indica el Monto Distribuido al Sistema Nacional de Salud para cada Obra Social publicado anualmente por el Organismo AFIP
- Se sumó el monto total transferido en el año 2021 a través del S.U.R. -Sistema Único de Reintegros- para el reconocimiento de las prestaciones médicas de baja incidencia y alto impacto económico y las de tratamiento prolongado. La información está disponible en la página web de la Superintendencia de Servicios de Salud). Para percibir este reintegro, el financiador debe contar con la necesidad de prestar algún tratamiento incluido en el listado de patologías incluidas. Por otro lado, los medicamentos deben estar incluidos en un Anexo que elabora la Superintendencia y que actualiza periódicamente (los medicamentos demoran entre 6 meses y varios años en ingresar en dicho padrón)
- Se sumaron los montos transferidos en concepto de la matriz SANO⁴⁵ (Subsidio Automático Nominativo de Obras Sociales). Esta información se encuentra accesible para la población en el portal de datos de la Superintendencia de Servicios de Salud:
- Se suman además los ingresos percibidos por los distintos financiadores en concepto del Programa SUMA (Subsidio de Mitigación de Asimetrías) y SUMARTE (Subsidio de Mitigación de Asimetrías para el Régimen de Trabajo Especial), SUMA 65 (Subsidio de Mitigación de Asimetrías para Mayores de 65 Años). Esta información está disponible en la página web de la Superintendencia de Servicios de Salud. Todos estos fondos están destinados a complementar el financiamiento de la cobertura de salud de la población beneficiaria incluida como consecuencia de las disparidades en cuanto a ingresos de la población cubierta o los grupos poblacionales a que dan cobertura a través del: Régimen de Trabajo Especial (toda vez que actualmente el aporte con destino a obra social), el Régimen Simplificado para Pequeños Contribuyentes y el Régimen para Trabajadores de Casas Particulares.

⁴⁵ Según la Superintendencia de Servicios de Salud, “el SANO garantiza a las entidades beneficiarias del sistema, un valor mínimo de aportes por afiliado, destinado a cubrir las prestaciones medico asistenciales. Esta asignación tiene en cuenta, la estimación del gasto según grupos demográficos con características epidemiológicas similares, evitando la selección de riesgo, ya que la fijación de valores relacionados con el gasto prestacional desalienta la discriminación y garantiza la accesibilidad al Sistema. Este modelo de distribución contempla, entre otros parámetros, la cantidad de beneficiarios totales de cada uno de los Agentes del Seguro de Salud, incluyendo los grupos familiares de los trabajadores.

Figura 45: Cuadro de asignaciones en concepto de matriz de riesgo SANO para obras sociales nacionales

VALOR ASIGNADO		
Grupo Etario	Masculino	Femenino
0 a 14	1.556,17	1.556,17
15 a 49	2.446,81	2.877,45
50 a 64	2.877,45	2.877,45
65 o más	6.361,70	6.361,70

Fuente: Superintendencia de Servicios de Salud (2022)⁴⁶

Figura 46: Representación de modelo de base de cálculo de datos de financiamiento de Obras Sociales Nacionales

Denom	RNOS_sii	Total_Transferido2021AFIP	Transf_2021_SUR	Transf2021_SANC	Transf2021_SUMA	Transf2021_SUMATE	Acum2022AFIP	TOTAL 2021
OBRA SOCIAL DE EJECUTIVOS Y DEL PERSONAL DE DIRECCION DE EMPRESAS	400800	\$59.863.454.359	\$1.740.064.563	\$417.991.069	\$806.024.289	\$12.516.232	\$7.074.510.905	\$62.840.050.513
OBRA SOCIAL DE LOS EMPLEADOS DE COMERCIO Y ACTIVIDADES CIVILES	126205	\$42.119.501.787	\$3.928.732.632	\$621.469.572	\$1.588.412.635	\$40.175.450	\$4.924.860.436	\$48.298.292.076
OBRA SOCIAL DEL PERSONAL DE DIRECCION ACCION SOCIAL DE EMPRESARIOS	400909	\$20.255.016.620	\$902.652.330	\$223.868.428	\$296.499.266	\$1.970.478	\$2.461.993.412	\$21.680.007.121
OBRA SOCIAL UNION PERSONAL DE LA UNION DEL PERSONAL CIVIL DE LA NACIO	125707	\$20.047.962.504	\$1.255.631.255	\$358.091.724	\$484.792.679	\$7.663.360	\$2.268.062.919	\$22.154.141.522
OBRA SOCIAL DE COMISARIOS NAVALES	113809	\$18.689.634.507	\$396.217.660	\$133.411.790	\$260.966.813	\$47.582	\$2.071.000.343	\$19.480.278.352
OBRA SOCIAL DE PETROLEROS	115300	\$14.476.506.765	\$592.317.315	\$138.075.869	\$414.103.785	\$8.833.373	\$1.695.828.066	\$15.629.837.108

Fuente: Elaboración propia

- **En caso de las Empresas de Medicina Prepaga:**
 - En los casos en que las empresas de medicina prepaga ofrezcan también servicios como obra social, se ha utilizado la información disponible en el boletín anual estadístico de la AFIP, donde se reporta información acerca de las transferencias a obras sociales nacionales. Además, se utilizó información sobre la estimación de facturación de las distintas compañías privadas que ha sido obtenida a través de una consultoría comprada por la compañía farmacéutica que lo proveen a modo informativo. Corresponden a una estimación aproximada pero que se construye en función de Balances de los financiadores, páginas web de empresas de medicina prepaga, informes institucionales que publican algunas empresas, balances públicos en casos de empresas grandes, entre otras. En particular, se decidió utilizar el siguiente criterio:
 - En los casos de las Empresas de Medicina Prepagas grandes (primeras 30) se utilizó la información de ingresos provista por un trabajo de consultoría de marketing.
 - Se han explorado múltiples páginas web, informes institucionales, entre otros, con el fin de buscar reflejar de la mejor manera posible los ingresos por afiliado de estas empresas de medicina prepaga.

⁴⁶ Consultado en página web de la Superintendencia de Servicios de Salud. Fecha de datos: Diciembre 2021. <https://www.argentina.gob.ar/sssalud/transparencia/subsidios/sano>

- En los casos en que no se tuviera información sobre facturación de parte de las herramientas antes mencionadas, se utilizó el valor promedio de cápita para empresas de medicina prepaga para todo el país provisto por el mismo trabajo de consultoría indicado. En tal sentido, se multiplicó el total de beneficiarios afiliados a la EMP por la cápita promedio indicada por la consultoría.

Obras Sociales Provinciales:

- En este subgrupo de financiadores se tuvieron que realizar varios supuestos con el objetivo de construir la información utilizada. Lamentablemente, no existe para este grupo de financiadores alguna entidad como la SSS que centralice la información y que realice publicaciones al respecto para los distintos financiadores. Están agrupadas a través de la COSSPRA, pero su gestión es a nivel provincial y cada una posee mayor o menor desarrollo y publicación de estadísticas propias.
- En primer lugar, se consultaron las distintas páginas web de los financiadores y se relevó la existencia de “Informes de Gestión” o de información estadística pública sobre los beneficiarios. Para la estimación de las poblaciones por rango etario se fueron utilizando distintos criterios que se indican a continuación:
 - En todos los casos se observa que ninguna fuente permite identificar la proporción de nacimientos anuales que tiene cada financiador. Dicha estadística se considera relevante dado que permite hacer cálculos de población incidente para ciertas patologías congénitas o genéticas. Es por ello que se buscó hacer una estimación de los nacimientos anuales por financiador a partir de la tasa provincial de natalidad que, si bien no refleja necesariamente a la población cubierta por tales financiadores, permite hacer una estimación sobre esta variable.
 - En el caso de IOMA, obra social de la Provincia de Buenos Aires, la información sobre población fue brindada a partir de una consulta realizada a la institución, pero se desconoce exactamente la fecha de los datos. Por lo tanto, se han utilizado estimaciones que permiten reflejar la población de un modo aproximado ya que, lamentablemente, no se encontraron publicaciones de parte del organismo con esa información. Es por ello que se debieron hacer algunas correcciones en ciertos rangos etarios dado que no eran similares a los utilizados en este trabajo. Para ello se utilizaron las distribuciones que se habían obtenido para otros financiadores nacionales buscando respetar la distribución proporcional de afiliados, pero conservando el total de afiliados para ese rango. En consecuencia, por ejemplo, se contaba con la información de beneficiarios entre 5 y 14 años y para poder dividirlo en los rangos 5 a 9 y 10 a 14, se utilizó un promedio de los datos de otros financiadores provinciales para los que sí había información y, además, se analizó si dicho número difería en gran medida con la proporción de población identificada a partir del último Censo del año 2010. Se asume que por dar cobertura a poblaciones similares

- (empleados públicos, docentes y policías) se podría asumir comportamientos similares al analizar diferentes provincias.
- Neuquén, Tierra del Fuego, Tucumán, La Rioja y Mendoza si tenían la población publicada por rango etario y género en sus páginas web. En tal sentido, esta información fue clave para construir distribuciones similares para los demás financiadores para los que no había información poblacional distribuida por rango etario. El ejercicio consistió en analizar gráficamente si la distribución en los distintos rangos etarios y de género para estas provincias eran similares entre sí y a partir de observar que su comportamiento era similar, asumimos que el resto de las Obras Sociales Provinciales tendrían una forma similar. Por otro lado, se analizó la información agregada publicada por COSSPRA a nivel nacional y también se observó un comportamiento similar. Por último, se validaron estas distribuciones (o al menos las características más relevantes) con quienes interactúan con dichos financiadores, identificándose que efectivamente tienen un comportamiento tal como el que se representó en los datos. Sin embargo, es evidente que podría haber errores como consecuencia de haber sido construidos bajo estos supuestos indicados.
 - Para la estimación de los ingresos económicos de cada financiador, también se tuvieron que asumir distintas cuestiones como consecuencia de la gran disparidad territorial y la falta de un ente nacional que consolide la información y la publique. El mecanismo de estimación utilizado fue realizado a partir de los siguientes criterios:
 - Tal como indica Torres et al. (2020, 104), “Para el financiamiento de las OSP, los trabajadores aportan recursos en proporción a su salario, y el Estado Provincial aporta contribuciones patronales, y a su vez, es el agente de retención de los aportes personales deducidos de los salarios. De aquí surge una relación especial entre el Estado Provincial y las OSP, que puede permitirle a aquel, en algunas situaciones especiales, financiarse a través de la generación de deuda con las OSP. Otra fuente de ingresos corresponde a la percepción de coseguros y copagos”. En consecuencia, se asumieron dos fuentes principales de ingresos para los financiadores: aportes y contribuciones de los empleados públicos de las distintas provincias y aportes y contribuciones del personal jubilado que continúa con cobertura de su obra social provincial. Para el primero se utilizó una base publicada por la Dirección Nacional de Asuntos Provinciales que funciona dentro de la Secretaría de Hacienda del Ministerio de Economía de la Nación. Dicho organismo, publica la ocupación y gastos salariales provinciales. Con esta información, se utilizó la tabla provista por la consultoría antes indicada para determinar el valor de la cápita en concepto del personal activo (contemplando sólo el personal del estado para entender el monto económico asumiendo que no se hacen aportes adicionales por dar cobertura al grupo familiar, por ejemplo). Para los ingresos en concepto de aportes y contribuciones del personal jubilado que pertenece a las Obras Sociales Provinciales, se utilizó un cuadro que publica ANSES

donde se indica el Haber Medio Previsional de ANSES según tipo de beneficio, sexo y jurisdicción. Con esta información y asumiendo que los empleados del estado no ingresaron en moratorias previsionales (para contemplar el haber medio), aplicamos los mismos porcentajes en concepto de aportes y contribuciones a la obra social provincial y buscamos construir el monto total percibido en ese concepto.

- Es evidente que, como consecuencia de estas asunciones, es posible que existieran diferencias con respecto a la situación particular de cada financiador, pero fue el mejor enfoque que pensamos que se podría utilizar como consecuencia de la falta de información disponible para la consulta pública.

Sector Público Nacional y Provincial:

- Para la estimación de los ingresos totales por beneficiario del sector, se buscó hacer un análisis renglón por renglón dentro del detalle del Presupuesto Nacional ejecutado para el periodo 2021 con destino a la satisfacción de la salud. En este sentido, se contemplaron sólo aquellos recursos efectivamente destinados para la población con cobertura exclusiva (se descuentan los montos destinados a COVID, vacunas, capacitaciones de salud sexual y reproductiva)
- Para entender el gasto presupuestario que ejecutan los gobiernos provinciales en concepto de salud, se analizó la información de la Ejecución Presupuestaria de la Administración Pública no financiera provincial en concepto de Salud. En este caso, si bien sabemos que es posible que los gobiernos provinciales ejecuten políticas provinciales de salud pública que no están tan solo dirigidas a la población con cobertura exclusiva, es prácticamente imposible identificar en este caso el detalle de lo que está destinado al 100% de la población y lo que está destinado a la población con cobertura exclusiva.
- Para identificar la población beneficiaria de la cobertura pública de salud, se utilizó información provista a través de la información del Programa SUMAR donde se tiene relevada la población que tiene cobertura exclusiva a nivel nacional. Se distribuyó esta población por provincia y luego se calculó el ingreso total de los financiadores por provincia contemplando la suma de los gastos ejecutados por el gobierno nacional para la población con cobertura exclusiva más el gasto presupuestario provincial en salud dividido en la cantidad de beneficiarios de esa provincia.

INCLUIR SALUD:

- Se realizó un pedido de acceso a la información pública solicitando el detalle de la población por edad y género que tienen cobertura del programa
- Se obtuvo a partir de la ejecución presupuestaria nacional el detalle de los gastos ejecutados para el programa

Obras sociales especiales

En el caso de este ecosistema, al no existir una fuente centralizada que controle la pertenencia de sus beneficiarios a estos financiadores, no hay fuentes confiables que permitan entender la población cubierta, el financiamiento y demás aspectos que caracterizan a un sector que brinda cobertura en total alrededor de 1.000.000 de personas.

Por considerarse relevante su inclusión, se hicieron varias asunciones y estimaciones de la población incluida

- Para calcular la población por rango etario, lamentablemente no fue posible encontrar información para esa población en particular. Por lo tanto, con el fin de poder incluir dichos financiadores en el modelo, se llevaron a cabo muchas asunciones.
- En primer lugar, para determinar una estimación de la cantidad total de beneficiarios para cada financiador, en el caso de las Obras Sociales Universitarias, se distribuyó un valor de 300.000 afiliados que son los tomados por organismos públicos entre las universidades públicas más relevantes del país en términos de recursos humanos contratados. Para definirlos, se hizo uso del Portal de Estadísticas Universitarias del Ministerio de Educación de la Nación. En el resto de los financiadores involucrados se tenía al menos alguna publicación en medios sobre la población cubierta. Fueron consultados balances, páginas sindicales, papers, entre otras fuentes que dieron contenido a la información finalmente incorporada.
- La distribución de la población por rango etario y género para las Obras Sociales Universitarias y de la Obra Social de la Policía Federal, fue estimado con los datos promedio de la Superintendencia de Servicios de Salud para las obras sociales nacionales
- Para IOSFA, se cuenta con información provista por el organismo acerca de su población
- Para la OSPJN, que brinda cobertura al personal del Poder Judicial de la Nación, se asumió una distribución similar a la de la información de D.A.S., Obra Social del personal del Congreso, cuya fuente de datos fue un informe publicado por el mismo organismo
- Para calcular los ingresos económicos, para IOSFA y la Policía Federal se obtuvo a partir de la ejecución presupuestaria nacional el detalle de los gastos ejecutados para los programas que derivan aportes a algunos de los financiadores implicados. Luego, el presupuesto de la OSPJN fue extraído de un informe de peritaje a la gestión de la obra social.
- Para la OSPJN y el DAS fue extraída de informes de los organismos y publicaciones de prensa y/o sindicales.
- Para las Obras Sociales Universitarias, se utilizó la información estadística sobre escalafón docente promedio para cada universidad (% Exclusivo promedio de 10 años de experiencia, % Semi exclusivo de 7 años de antigüedad, % Simple de 5 años de antigüedad, % Docentes Preuniversitarios de 7 años de antigüedad, % No docentes - La información sobre el salario promedio que se utilizó para cada financiador, fue extraído a partir de información publicada en la página de Conadu⁴⁷ (Federación Nacional de Docentes Universitarios) asumiendo una retención para el financiador de un total de 9% (6% de contribución por parte de la Universidad y 3% por parte del trabajador). En el caso del personal no docente, se utilizó la información de retención salarial de una obra social nacional asumiendo que los valores podrían tener sentido (la Obra Social De La Federación Argentina Del Trabajador De Las Universidades Nacionales). Todos estos resultados fueron contrastados con los del resto de los ecosistemas de salud nacionales demostrando que los valores asumidos son “sensatos” con respecto a los datos del resto del ecosistema sindical.

⁴⁷ <https://conadu.org.ar/grilla-salarial-paritaria-2021/> consultado el 5/04/2023

Anexo 2: Construcción de datos para modelos. Descripción de procesamiento de los datos

Inicialmente se utilizaron para definir los subgrupos de financiadores a dos familias de datos principales para los distintos financiadores. Por un lado, se trabajó con la información demográfica de los financiadores a través de la cobertura por género y grupo etario cubierto. Asimismo, se incorporaron variables descriptivas de la estructura del financiador (a través del análisis del nivel de integración vertical del financiador, de la distribución territorial en Argentina y del ecosistema de salud al que pertenece (y, por lo tanto, su mecanismo de financiamiento asociado). Por último, se sumó información acerca del mecanismo de compra de medicamentos de alto precio a la compañía en análisis, de los ingresos totales, del porcentaje en que incrementaron sus ingresos en el periodo de análisis en comparación con la inflación. Los financiadores se tienen identificados con un nombre único.

Con cada familia de variables y utilizando diferentes fórmulas que permitieran ajustar la prevalencia o incidencia de las distintas patologías estratégicas para la compañía en análisis en los distintos financiadores, se exploraron análisis que permitieran derivar la información demográfica en números relevantes para la estimación de la población objetivo de la tecnología sanitaria.

En esta sección se describen las variables significativas para el análisis que efectivamente se utilizó en la generación de clústeres.

Tabla 7: Base inicial de población estimada para cada financiador por rango de edad y género

Variable	Descripción	Clase
F0	Cantidad de mujeres nacidas	Entero
F1	Cantidad de mujeres entre 1 y 4 años	Entero
F2	Cantidad de mujeres entre 5 y 9 años	Entero
F3	Cantidad de mujeres entre 10 y 14 años	Entero
F4	Cantidad de mujeres entre 15 y 19 años	Entero
F5	Cantidad de mujeres entre 20 y 24 años	Entero
F6	Cantidad de mujeres entre 25 y 29 años	Entero
F7	Cantidad de mujeres entre 30 y 34 años	Entero
F8	Cantidad de mujeres entre 35 y 39 años	Entero
F9	Cantidad de mujeres entre 40 y 44 años	Entero
F10	Cantidad de mujeres entre 45 y 49 años	Entero
F11	Cantidad de mujeres entre 50 y 54 años	Entero
F12	Cantidad de mujeres entre 55 y 59 años	Entero
F13	Cantidad de mujeres entre 60 y 64 años	Entero
F14	Cantidad de mujeres entre 65 y 69 años	Entero

F15	Cantidad de mujeres entre 70 y 74 años	Entero
F16	Cantidad de mujeres entre 75 y 79 años	Entero
F17	Cantidad de mujeres mayores a 80 años	Entero
M0	Cantidad de hombres nacidos	Entero
M1	Cantidad de hombres entre 1 y 4 años	Entero
M2	Cantidad de hombres entre 5 y 9 años	Entero
M3	Cantidad de hombres entre 10 y 14 años	Entero
M4	Cantidad de hombres entre 15 y 19 años	Entero
M5	Cantidad de hombres entre 20 y 24 años	Entero
M6	Cantidad de hombres entre 25 y 29 años	Entero
M7	Cantidad de hombres entre 30 y 34 años	Entero
M8	Cantidad de hombres entre 35 y 39 años	Entero
M9	Cantidad de hombres entre 40 y 44 años	Entero
M10	Cantidad de hombres entre 45 y 49 años	Entero
M11	Cantidad de hombres entre 50 y 54 años	Entero
M12	Cantidad de hombres entre 55 y 59 años	Entero
M13	Cantidad de hombres entre 60 y 64 años	Entero
M14	Cantidad de hombres entre 65 y 69 años	Entero
M15	Cantidad de hombres entre 70 y 74 años	Entero
M16	Cantidad de hombres entre 75 y 79 años	Entero
M17	Cantidad de hombres mayores a 80 años	Entero

Tabla 8: Variables construidas a partir de información base de la tabla 7 que fueron incluidas en el análisis del modelo

Variable	Descripción	Clase	N	Promedio	Máximo	Mínimo
N_Masc_tot	Cantidad de hombres cubiertos por el financiador	Entero	Suma de hombres (M) estimados del financiador	70582	3525594	3
N_nac	Cantidad de nacimientos estimados por año para el financiador	Entero	Suma de hombres (M) y mujeres (F) con nivel 0	1267	101484	0
N_mas60	Cantidad de beneficiarios mayores a 60 años	Entero	Suma de hombres (M) y mujeres (F) desde el nivel 13 en adelante estimados del financiador	26124	4878025	0
N_Fem_Tot	Cantidad de mujeres cubiertas por el financiador	Entero	Suma de mujeres (F) estimadas del financiador	74830	3263754	2

Y, por otro lado, con distintas fuentes estadísticas se hicieron estimaciones de incidencia y prevalencia de patologías estratégicas para la compañía farmacéutica que resultaron en una tabla con la siguiente información relevante

Tabla 9: Variables construidas a partir de información base de la tabla 7 combinadas con fuentes de incidencia y prevalencia extraídas de diversas fuentes de información (Globocan⁴⁹, Global Burden of Disease⁵⁰ y a partir de opinión de expertos⁵¹)

Variable	Descripción	N	Promedio	Máximo	Mínimo
N_total_afiliados	Número total de beneficiarios estimados	Suma total de beneficiarios contemplados en el análisis	145412	6789349	5
Edad	Edad promedio de los beneficiarios	Promedio de edad calculado para el financiador	35	72	24
Hemofilia	Cantidad estimada de población prevalente de Hemofilia en el financiador	Prevalencia por rango etario y género de la patología	8	423	0
AME	Cantidad estimada de población prevalente de Atrofia Muscular Espinal en el financiador	Prevalencia por rango etario y género de la patología	2	110	0
PP_Prevalence ⁵²	Cantidad estimada de población prevalente de esclerosis múltiple primaria progresiva en el financiador	Prevalencia por rango etario y género de la patología	5	334	0
RR_Prevalence ⁵³	Cantidad estimada de población prevalente de esclerosis múltiple remitente recurrente en el financiador	Prevalencia por rango etario y género de la patología	22	1606	0
Non_Hodk	Cantidad estimada de población incidente de Linfoma de Non Hodking en el financiador	Incidencia por rango etario y género de la patología			0
HER2	Cantidad estimada de población incidente de cáncer de mama HER2+ en el financiador	Incidencia por rango etario y género de la patología	72	11171	10

⁴⁸ Utilizando estas variables se genera la variable ShFem o ShareFemint donde se calcula la cantidad total de mujeres sobre el total de beneficiarios contemplados

⁴⁹ Global Cancer Observatory (Globocan) depende de la Organización Mundial de la Salud y publica datos de incidencia de la patología ajustada a rango etario para Argentina para el año 2020.

⁵⁰ Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Reference Life Table. Seattle, United States of America: Institute for Health Metrics and Evaluation (IHME), 2021.

⁵¹ Se consultan médicos expertos en el tema en casos de patologías que no tienen datos estimados para Argentina en particular.

⁵² Se suma a RR_Prevalence en EMultiple

⁵³ Se suma a PP_Prevalence en EMultiple

ALK ⁵⁴	Cantidad estimada de población incidente de cáncer de pulmón de células no pequeñas ALK en el financiador	Incidencia por rango etario y género de la patología	1	180	3
SCLC ⁵⁵	Cantidad estimada de población incidente de cáncer de pulmón de células pequeñas pasibles de tratamiento en el financiador	Incidencia por rango etario y género de la patología	4	655	0
NSCLC ⁵⁶	Cantidad estimada de población incidente de cáncer de pulmón de células no pequeñas pasibles de tratamiento en el financiador	Incidencia por rango etario y género de la patología	37	6,446	0
val_lung	Cantidad estimada de población incidente de cáncer de pulmón en el financiador	Incidencia por rango etario y género de la patología	43	7,584	0
BLADDER1L ₅₇	Cantidad estimada de población incidente de cáncer de vejiga pasibles de tratamiento en primera línea en el financiador	Incidencia por rango etario y género de la patología	0	84	0
BLADDER2L ₅₈	Cantidad estimada de población incidente de cáncer de vejiga pasibles de tratamiento en segunda línea en el financiador	Incidencia por rango etario y género de la patología	11	2140	0
LIVER1L	Cantidad estimada de población incidente de cáncer de hígado en el financiador	Incidencia por rango etario y género de la patología	6	1134	0
Incidencia_portfolio_2	Incidencia ponderada de las diversas patologías analizadas. Se pondera el uso esperado del paciente del producto, el costo del tratamiento en conjunto con la incidencia potencial de varias patologías incluidas en esta tabla pero que luego son descartadas del análisis	Se utiliza un parámetro α_i que ha sido construido por el autor con ponderaciones de patologías considerando: mediana de tratamiento, costo de tratamiento, prevalencia específica (ponderando todas las variables: ALK, BLADDER 1L y BLADDER 2L, LIVER1L, NSCLC, SCLC, PP_Prevalence, RR_Prevalence)			

Tabla 10: Base Financiadores y estructura de costos y de gestión (por disponibilidad o no de centros propios de salud que brindan cobertura exclusivamente a los afiliados y que, en su mayoría cubren a otros financiadores de salud)

Variable	Descripción	Clase	N	Promedio	Máximo	Mínimo
Cantidad.de.centros.propios	Cantidad de centros propios de salud para provisión de servicios de alta complejidad con terapia intensiva	Numérico	31	6.26	496	0
Ingreso_por_afiliado	Estimación de ingreso promedio mensual por beneficiario para el financiador	Numérico	Ingresos anuales totales / Cantidad	4385	21538	381

⁵⁴ Se descarta por su correlación con val_lung, al significar simplemente la aplicación de filtros en esa variable. Se deja solo esa variable.

⁵⁵ Igual a indicación referencia 54

⁵⁶ Igual a indicación referencia 54

⁵⁷ Se descarta del análisis por el tamaño mínimo y correlación con otros tipos de cáncer

⁵⁸ Igual a indicación referencia 54

			total de beneficiarios /12 meses			
Aumento ⁵⁹	Porcentaje de actualización de los ingresos mensuales percibidos por el financiador en relación a evolución IPC. Puede ser consecuencia de: incremento de beneficiarios del financiador o incremento en los ingresos percibidos para cada beneficiario (según su vía de financiamiento).	Numérico	Ingresos periodo t+1 disponible/ ingreso periodo t -IPC t+1 /IPC t (en ARS del año 2021 en comparación con año 2022).	65%	9515%	-0.81%
Ingreso_Mensual_Total	Estimación de cantidad de pesos totales de ingresos estimados para el año 2021 dividido por los 12 meses del año	Numérico	Se estimó de gasto de cobertura en salud de \$233119919421 para la población representada (en ARS del año 2021)	\$550M	40631M	\$16273

Tabla 11: Variables descriptivas de los financiadores

Variable	Descripción	Clase	N
Compradirecta_1	Información sobre si es financiador que es cliente o no directo (Si)	Factor	2
Dispone.de.centros.propios.de.salud	Disposición de centros propios de salud para provisión de servicios de alta complejidad con terapia intensiva	Factor	2
Ecosistema	Ecosistema de salud al que pertenece el financiador	Factor	7
Region_principal ⁶⁰	Regiones o áreas específicas o todo el territorio	Factor	8
Nombre_financiador	Nombre de referencia del financiador	Factor	366

Anexo 3: Análisis de variables categóricas mediante técnica

FAMD. Profundización en el análisis

Análisis de Factores para datos mixtos con variables normalizadas y log normalizadas en herramienta FAMD

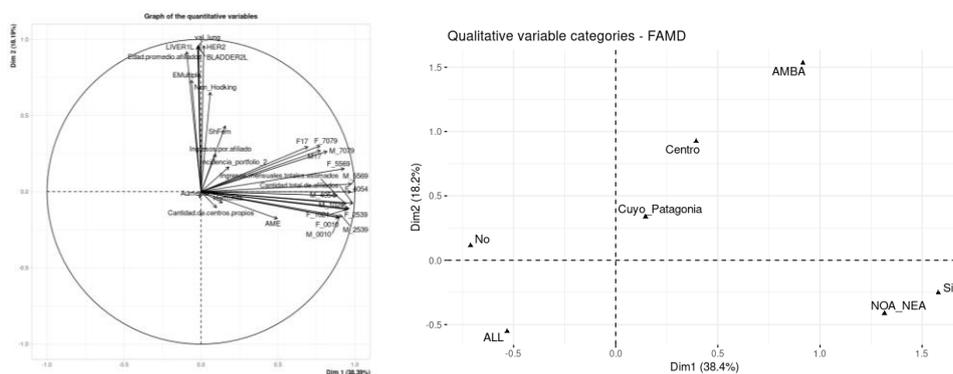
⁵⁹ En el modelo se pondera el aumento con el ingreso de base ya que se releva aquellos que tienen aumentos más grandes en términos porcentuales, partían de ingresos muy bajos por lo que no se quería sobreestimar su efecto en términos de análisis del ingreso

⁶⁰ En la salida de R, se utiliza Región_Principal_CatP donde se transforma el factor en el porcentaje representativo al que pertenece cierto financiador. Es decir, por ejemplo, si se encuentra a lo largo de todo el país, se agrupa en conjunto con el 53% de los financiadores

De inicio, se normalizaron las 32 variables del *dataframe* construido inicialmente. Se dejaron algunas variables con ciertos desvíos poblacionales (OSDE y IOMA) por su tamaño, principalmente. Las variables iniciales, fueron trabajadas a través de, en primer lugar, una técnica de normalización y estandarización. En segundo término, se hizo un análisis a través de la técnica FAMD que permite realizar clustering en conjuntos de datos que contienen tanto variables categóricas como continuas. En el caso de este análisis, conservamos dos variables categóricas: Dispone Centros de Salud y Región Principal (ALL, NOA-NEA, Patagonia-Cuyo, Centro, AMBA). La función FAMD permite realizar un análisis de factores en los datos, posibilitando una mejor representación de la estructura subyacente de los datos mixtos. Esta técnica se basa en el análisis de factores clásica (CFA) pero fue diseñada para manejar datos mixtos que contienen tanto variables categóricas como continuas. El método combina las técnicas de análisis de componentes principales para las variables continuas y de correspondencia (MCA) para las variables categóricas. En lugar de considerar solo la distancia euclidiana entre los puntos de datos, FAMD utiliza una distancia basada en los componentes principales del análisis de factores. La distancia resultante es más apropiada para datos mixtos y puede mejorar la calidad del clustering. Asimismo, se ha indicado que se permite analizar los datos respetando su naturaleza original.

En la siguiente figura, 47, se representa en un gráfico, la dirección vectorial en la que se comportan las variables con respecto a las dimensiones 1 y 2 seleccionadas por el modelo.

Figura 47: Representación de las dimensiones 1 y 2 de los componentes principales en forma vectorial y de las variables categóricas



Fuente: Elaboración propia

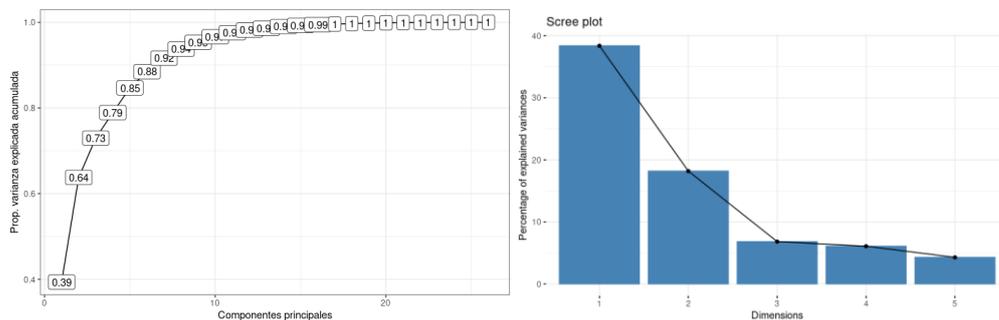
Luego, cuando se analiza el aporte que realiza cada dimensión identificada dentro del total se puede identificar que, tal como se representa en la siguiente figura, el aporte que el modelo considera significativo es a través de 5 dimensiones principales. Se puede observar en la mencionada figura, el aporte en la variabilidad de los datos identificados a través del modelo.

Es a partir de este análisis, que se buscó entender cuáles eran esas variables significativas dentro de esas primeras 5 dimensiones. En consecuencia, se pudo llegar a la conclusión, a partir del análisis de la figura 48, que las dimensiones que explican fundamentalmente la variabilidad de los datos son:

- Dimensión 1: Tamaño del financiador, se ven todas variables que indican cantidades más significativas de afiliados
- Dimensión 2: Edad y patologías incidentes para poblaciones de edad más alta en promedio

- Dimensión 3: Portfolio de la compañía ponderado y medición de incidencia de enfermedades de baja incidencia
- Dimensión 4: Portfolio de la compañía ponderado y medición de incidencia de enfermedades de baja incidencia. Esta dimensión será descontada por entender que no aporta información significativa adicional al análisis
- Dimensión 5: Estructura económica. Esta dimensión tiene aportes de ingresos promedio por afiliado, centros propios y región principal

Figura 48: Contribución en la variabilidad de los datos de los distintos factores identificados por el modelo FAMD



Fuente: Elaboración propia

Como ha sido reflejado en el detalle en la sección ingeniería de variables, se experimenta la ejecución de un análisis mediante la técnica de Factor Analysis of Mixed Data para entender si tiene sentido incluir de algún modo las variables categóricas con que se cuenta en el *dataframe* original. En la siguiente figura, 49, se refleja el aporte de variabilidad de las distintas variables incluidas en la técnica FAMD. Luego, en la misma figura se analiza el aporte de las distintas variables en la explicación de la variabilidad total de los datos.

Se observa que las variables categóricas que potencialmente se podrían incluir tan solo se ven reflejadas en la quinta dimensión del componente de variabilidad, explicando muy poco de la variabilidad de los datos.

Por lo tanto, se descartan las variables categóricas del análisis.

Además de este análisis, se buscó contemplar este análisis de FAMD una vez más, con los datos en logaritmo. En la figura 50, se puede observar la representación vectorial de los datos, la contribución de las dimensiones determinadas por el modelo y la descripción de las variables que contribuyen en mayor y menor medida en cada una de esas dimensiones determinadas.

Figura 51: Modelo de tipo FAMD con variables logarítmicas - representación vectorial de los datos, contribución de las dimensiones determinadas por el modelo y la descripción de las variables que contribuyen en mayor y menor medida en cada una de esas dimensiones determinadas

