



REVISTA
SINERGIAS

Publicación Semestral de la Secretaría de Posgrado de la
Universidad Nacional Guillermo Brown, en colaboración
con UNaB Editora.

Métodos computacionales aplicados a la Política Pública en Argentina

sinergias.unab.edu.ar

NÚMERO #01
JULIO DE 2024

Análisis de notificación de casos de tuberculosis antes y durante la pandemia de COVID-19 mediante aprendizaje automático

De León, Romina & Wettstein, Magali

Análisis de notificación de casos de tuberculosis antes y durante la pandemia de COVID-19 mediante aprendizaje automático

Resumen

La integración de la Inteligencia Artificial en bases de datos masivas, inabarcables al ojo humano, como aquellas provenientes de las notificaciones de casos de enfermedades infecciosas podría proporcionar nuevos conocimientos sobre la afección que resultó de los aislamientos sociales durante la pandemia de COVID-19, especialmente entre la declaración de la misma y el primer año de desarrollo en pacientes con tuberculosis. Para ello, se empleará aprendizaje automático supervisado en el procesamiento de datos con variables predictoras diversas. En este contexto, también se llevará a cabo un análisis de series temporales para evaluar si las medidas de dichas circunstancias afectaron los tiempos de demora en el diagnóstico y en el tratamiento de la tuberculosis. Este enfoque tendrá como objetivo identificar patrones, tendencias y posibles correlaciones que podrían proporcionar nuevos conocimientos sobre cómo la pandemia de COVID-19 impactó específicamente en la gestión de la tuberculosis, contribuyendo así a una comprensión más completa de las dinámicas de salud pública en situaciones de crisis.

Palabras clave

Tuberculosis, COVID-19, aprendizaje automático, inteligencia artificial, aislamiento social.

1. Introducción³

Los avances tecnológicos posteriores al período de las guerras mundiales no tienen precedentes. Incluso desde la mitad del siglo pasado, se han producido actualizaciones digitales a pasos colosales, y en todas las áreas de nuestra vida, por ende no son ajenos todos los ámbitos académicos. Eric Hobsbawm (2018) asegura que ningún otro período de la historia ha sido más impregnado ni ha dependido más de las Ciencias que el siglo XX (443).

La estructura tecnológica satura la vida humana, demostrando su indispensable omnipresencia, incluso en las Ciencias de la Salud. Por ello, en las últimas décadas, la aplicación y análisis con diversas metodologías computacionales ligadas a la Inteligencia Artificial (IA), en grandes cantidades de datos, ha permitido la toma de decisiones proactivas y la predicción de eventos de enfermedades a través de modelos de aprendizaje automático. La IA forma parte de la Informática y tiene como objetivo resolver problemas, en torno al aprendizaje y el reconocimiento de patrones.

Una subdisciplina de la IA es el aprendizaje automático o *machine learning* (en inglés, ML). Este tuvo sus primeros avances a mediados de siglo XX por científicos que trabajaban en la empresa IBM (International Business Machines Corporation), con el desarrollo de un programa para jugar damas, presentando el potencial que tendría esta metodología. Asimismo el ML utiliza algoritmos que aprenden de los datos masivos, estructurados o no estructurados, para predecir resultados y extraer nuevos conocimientos. Además, los algoritmos son los que toman decisiones de manera similar al humano después de un proceso de entrenamiento con datos. Es por ello que tanto la IA como el aprendizaje automático, al ser incorporados a las Ciencias de la Salud, en particular para trabajar con datos masivos de enfermedades infecciosas como la tuberculosis, permite agilizar y generar nuevos conocimientos de datos médicos, procurando mejorar resultados de salud y vivencias en tres niveles, en lo individual, de cada paciente, en las comunidades de pacientes y, en la sociedad en su conjunto (Martínez Sesmero, 2015).

La tuberculosis (TB), enfermedad infecciosa con antecedentes que datan de hace más de 10.000 años, es causada por la bacteria *Mycobacterium tuberculosis* y suele afectar a los pulmones. Es prevenible y curable,

¹ IIBICRIT-CONICET, rdeleon@conicet.gov.ar, [Romina Soledad De León \(0000-0003-2495-7213\) - ORCID](https://orcid.org/0000-0003-2495-7213).

² INER-ANLIS, mwettstein@anlis.gob.ar

³ El presente trabajo se desarrolló en el marco de la Diplomatura de Ciencias Sociales Computacionales de la Universidad Nacional Guillermo Brown (UNAB) cohorte 2023.

transmitiéndose de una persona a otra a través de gotitas de aerosol suspendidas en el aire expulsadas por personas con enfermedad pulmonar activa. El diagnóstico se basa en radiografías torácicas, pruebas cutáneas de tuberculina y análisis sanguíneo. Su tratamiento se realiza con antibióticos durante períodos de alrededor de seis meses. En 2022, la TB fue la segunda causa de muerte por un solo agente infeccioso en el mundo, después de la enfermedad por coronavirus (COVID-19), y causó casi el doble de muertes que el VIH/SIDA. Actualmente, la Organización Mundial de la Salud (OMS)⁴ estima que existen 2.000 millones de infectados por esta bacteria, y por año se contagian alrededor de 8 millones de personas.

En la actualidad esta infección continúa siendo una de las que posee tasas de morbilidad⁵ más alta a nivel mundial. Incluso, la Organización Panamericana de la Salud (OPS)⁶ considera que las muertes anuales en América han aumentado a más de 3.000 en 2020 debido a la interrupción de los servicios esenciales en el contexto de la pandemia mundial de la COVID-19⁷. Por lo anterior, tanto los Estados Miembros de las Naciones Unidas (ONU) como la Organización Mundial de la Salud (OMS)⁸ expresan la necesidad de medidas urgentes para finalizar la epidemia global de TB para 2030, destacando como requerimiento la inversión en recursos para intensificar la respuesta para los enfermos, sobre todo a los más vulnerables.

En conjunto, se puede entender que el ingreso de la IA al ámbito de la salud, con modelos de aprendizaje, algoritmos de ML, permitirán detectar patrones en los datos, facilitando el estudio predictivo y generando mayores conocimientos que impactarán en la calidad de vida de los pacientes. Por lo cual, el presente trabajo tiene por objetivo realizar una evaluación actualizada de la notificación de casos de tuberculosis en Argentina⁹, especialmente durante el período 2019-2021, para llevar a cabo una comparación en los sucesos de la pandemia de COVID-19. Se utilizará como base de datos el reporte sobre la comunicación de casos que genera el Sistema Nacional de Vigilancia de la Salud (SNVS 2.0)¹⁰, considerando, sólo, el período mencionado en busca de revelar la modificación o no de la notificación de afectados antes y durante la pandemia, en particular teniendo en cuenta las restricciones sociales que se llevaron a cabo en todo el mundo, durante varios meses¹¹.

2. Metodología

En este estudio se utilizó el informe del 31 de agosto de 2023 del SNVS¹², que contiene las notificaciones de casos registrados y validados de las 24 jurisdicciones argentinas entre los años 2019 y 2021. El dataset no estaba disponible directamente desde la URL, por lo tanto, después de descargarlo en formato CSV, se subió al repositorio GitHub. Esto se hizo con el objetivo de facilitar el trabajo colaborativo, que permite realizar

⁴ Para más información, véase [Tuberculosis](#).

⁵ Tasa de muertes por enfermedad en una población y en un tiempo determinados.

⁶ La Organización Panamericana de la Salud actúa como oficina regional de la Organización Mundial de la Salud. Su primer encuentro fue en diciembre de 1889 en Washington, EE. UU., pero se conformó como tal en 1902 en la misma ciudad.

⁷ La pandemia de COVID-19 se derivó de la enfermedad causada por el virus SARS-CoV-2.78. Los primeros casos identificados fueron a finales de noviembre de 2019 en Wuhan, ciudad de China. El 30 de enero de 2020 fue declarada la emergencia de salud pública internacional por la OMS, condición que mantuvo hasta el 5 de mayo de 2023. El 11 de marzo de 2020 fue declarada como pandemia por la alta cantidad de personas infectadas y muertes alrededor del mundo. En nuestro país, el 20 de marzo del mismo año, se declaró el Aislamiento social, preventivo y obligatorio (ASPO) con el decreto N° 297/2020, accesible desde el siguiente [link](#). El ASPO se extendió hasta el 31 de enero de 2021 reemplazado con el Distanciamiento social, preventivo y obligatorio (DIASPO) hasta junio de 2021. El proceso de vacunación comenzó el 29 de diciembre de 2020, el 76, 34% de la población argentina había recibido al menos una dosis para el 29 de octubre de 2021. Actualmente, según la OMS la enfermedad se encuentra en nivel endémico, debido a que ya no se reportan casos con tanta frecuencia como en sus primeros años.

⁸ Para acceder al informe mundial sobre tuberculosis 2023 generado por la OMS véase: [Global Tuberculosis Report 2023](#).

⁹ La ley nacional N°15.46 de Argentina promulgada en 1960 estableció la obligatoriedad, dentro del territorio nacional, de comunicar todos los casos de enfermedades infecciosas.

¹⁰ Desde SNVS se genera un reporte individual por parte de los servicios de salud a lo largo y ancho del país, de manera remota e inmediata. Accesible desde [SISA](#).

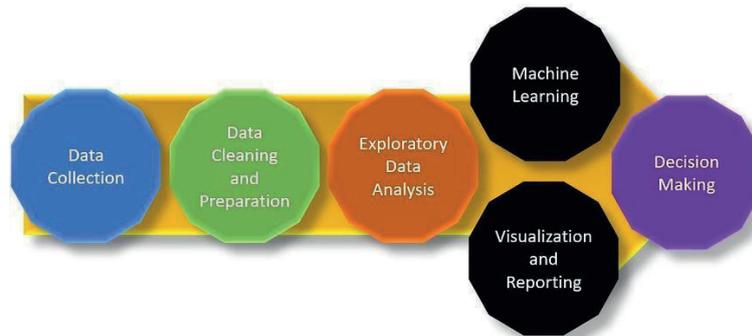
¹¹ Véase nota 6.

¹² Véase nota 8.

cambios, incluir comentarios y mantener actualizado el seguimiento de los progresos¹³.

El desarrollo práctico se inició con un análisis exploratorio de datos (EDA por sus siglas en inglés). Este conjunto de técnicas estadísticas busca acercarse a los datos y explorar las relaciones entre variables. Para ello, mediante métodos sistemáticos, se organiza, limpia y prepara los datos. Además se tratan y evalúan la presencia de valores ausentes (missing) y se detectan valores atípicos (outliers). Además, se generan visualizaciones que permiten comprender y revelar relaciones entre variables, así como seleccionar las herramientas a utilizar en el aprendizaje automático (Monterde i Bort y Perea Lara, 1991). A continuación se presenta un esquema que representa este análisis (Gráfico 1).

Gráfico 1. Flujo de proceso EDA desde la recopilación de datos hasta la toma de decisiones.



Fuente: [How to ace Exploratory Data Analysis | by Rahul Pandey | Medium | Analytics Vidhya](#).

Para continuar la exploración de datos se transformaron en dataframe¹⁴. Luego, se realizó una inspección estadística de las variables numéricas. Por lo cual, se convirtieron a formatos más adecuados, en el caso de las fechas. Se codificaron variables categóricas, como los grupos etarios, sexo, residencia, clasificación del caso. Se reajustaron las variables de *Provincias de residencia*, en particular los valores CABA por Ciudad Autónoma de Buenos Aires¹⁵, información relevante para las gráficas georreferenciadas. Se eliminaron columnas superfluas, como las semanas epidemiológicas¹⁶. Se determinó la correlación entre variables y se adecuaron valores vacíos, duplicados¹⁷ o faltantes. Se retuvieron únicamente pacientes por Documento Nacional de Identidad (DNI) y fecha de notificación. Además, se organizó de manera ascendente al dataframe. Se destaca respecto a la eliminación de valores faltantes que puede generar pérdida de información por la posible relevancia de los datos en las filas eliminadas. Sin embargo, luego de considerar el porcentaje que los valores vacíos implicaban, o de realizar un resumen del dataframe y de las filas a eliminar, se determinó que la opción más pertinente sería la de reemplazar dichos valores por una estimación basada en los datos existentes o eliminar filas que no aportaban trascendencia a los requerimientos del análisis. Los reemplazos para valores vacíos se hicieron por la media, la mediana o algún otro método de imputación.

La limpieza y análisis del dataframe se limitaron al contexto de la pandemia COVID-19 en Argentina y las

¹³ El código e indicaciones de implementación del presente trabajo se encuentran disponibles en el repositorio [GitHub](#) y en [Google Colab](#), pudiendo replicar la evaluación y llegando a los mismos resultados.

¹⁴ Es importante señalar la diferencia entre base de datos, dataset y dataframe, que durante todo este trabajo se utilizarán casi como sinónimos. Una base de datos recopila y organiza información, puede almacenar referencias sobre personas, productos, pedidos u otras cosas; por ello, puede entenderse a la misma como dataset. Este, por su parte, es un conjunto de datos, presentando la información en tablas o matrices, que son conformadas por columnas que representan variables particulares, y filas, como categorías de datos. Finalmente, un dataframe es una matriz de datos, que permite datos numéricos y alfanuméricos, además facilita el análisis de los objetos en una muestra de datos, y la información es estructurada en columnas identificadas respectivamente. Por todo lo anterior, se comprende que la consulta, modificación y análisis de información es más sencilla en dataframes.

¹⁵ Como datos de georreferencia se tomaron los datos censales, accesibles desde:

<https://infra.datos.gob.ar/catalog/modernizacion/dataset/7/distribucion/7.31/download/localidades-censales.geojson>.

¹⁶ Cabe destacar que la decisión de prescindir de esta información se basa en la premisa de que la fecha proporcionará de manera más efectiva la información relevante para el análisis.

¹⁷ Los datos duplicados refieren a las tomas de muestras, individuales enviadas para confirmar el evento, las de control del tratamiento, así como si el paciente tiene más de una comorbilidad (presencia de otras enfermedades en simultáneo).

fechas de inicio del ASPO¹⁸. El decreto que anunciaba dicho proceso presentaba en el segundo artículo que “durante la vigencia del ‘aislamiento social, preventivo y obligatorio’, las personas deberán permanecer en sus residencias habituales o en la residencia en que se encuentren a las 00:00 horas del día 20 de marzo de 2020, momento de inicio de la medida dispuesta”. Por ende, se creó una nueva variable clasificatoria de casos que los etiquetaba según **antes** y **después** de la fecha mencionada. El objetivo principal de dicha clasificación fue comparar grupos etarios y regionales respecto a las características de cada caso así como de los tiempos de demora (diagnóstico, tratamiento). Sumado a ello, también se recortó según la clasificación final del resultado de tratamiento de tuberculosis. Es decir, solamente se mantuvieron los pacientes que fueron clasificados como **curados** o **tratamiento completo**, ya que según las *Definiciones y marco de trabajo para la notificación de Tuberculosis* de la OMS (2013) son considerados casos con tratamiento **exitoso**. Entendemos que al filtrar únicamente los casos de éxito, se puede reducir el *ruido* en los datos. Este tipo de eliminaciones puede generar beneficios en la búsqueda de patrones claros o tendencias específicas asociadas a los casos exitosos, pues se elimina la variabilidad introducida por los casos con resultados diferentes.

De manera similar, se agruparon, para realizar visualizaciones más pertinentes, por grupos de edad. Para lo cual se generó una nueva columna con *grupo de edad* que permite catalogarlas de 0 a 18 años, de 19 a 30 años, de 31 a 45 años, de 46 a 60 años, y mayores de 61. Finalmente, se añadieron dos columnas: una que indicaba el tiempo de demora respecto al inicio de síntomas y fecha de diagnóstico, y otra con el tiempo de demora del tratamiento, que cuantifica el comienzo del tratamiento y la finalización del mismo. Los valores negativos de estas nuevas variables, que pueden deberse a errores de cálculo o en la carga de datos, se sustituyen por la media de los valores positivos.

La visualización gráfica y de datos tabulares es propicia para mejorar la interpretación de los datos, así como para limitar los pasos a seguir en el análisis, la toma de decisiones y seleccionar los mejores métodos para el aprendizaje automático supervisado puesto que permiten captar conceptos y patrones que no son relevados sencillamente en las grandes bases de datos (Wickham y Golemund, 2016). Asimismo, para analizar las relaciones entre las variables, se calculó la matriz de covarianza, que permite apreciar en su diagonal principal la varianza de cada variable, mientras que los elementos restantes arrojan las covarianzas entre las variables. Análogamente, se realizaron diagramas de dispersión evaluando los conjuntos de datos en relación a dos variables, una en cada eje cartesiano. Las representaciones gráficas por medio de histogramas evidencian variables cuantitativas continuas de un conjunto de valores, en forma de barras que proporcionan la frecuencia de valores observados (Murray y Spiegel, 2009). Con ellos, se pueden diferenciar parámetros, como posición, dispersión, asimetría, etc., que son de gran aporte dentro del análisis descriptivo de datos.

En lo que respecta al *machine learning* o aprendizaje automático, campo de estudio derivado de la inteligencia artificial, se ocupa del desarrollo y estudio de algoritmos estadísticos que pueden generalizar eficazmente, es decir, realizar tareas sin instrucciones explícitas. Este enfoque se ha aplicado en diferentes áreas, con desarrollo de diversos modelos como los lingüísticos, de visión por computadora, reconocimiento de voz, filtrado de correo electrónico, agricultura y medicina, etc. Dentro de este subcampo existen subdivisiones, las más destacadas y utilizadas son el aprendizaje supervisado y el no supervisado¹⁹. El supervisado es utilizado para datos donde las etiquetas son conocidas, ejemplo de ello podría ser la clasificación o la regresión. En este trabajo se utilizaron varios de estos modelos, pero fueron descartados o conservados según el ajuste más aceptable de los resultados. Incluso se consideraron modelos con menor sensibilidad a valores atípicos que otros, sin embargo, resultaron poco concluyentes y no se tuvieron en cuenta²⁰.

¹⁸ Véase nota 6.

¹⁹ Este tipo no fue utilizado para la predicción de los datos de este trabajo debido a que arrojaba segmentaciones poco concluyentes. El aprendizaje automático no supervisado se utiliza para indagar y hallar patrones en datos sin etiquetar. Una de las técnicas de aprendizaje no supervisado es la de clustering K-means. Este algoritmo es un método de cuantificación vectorial, cuyo origen proviene del procesamiento de señales. Tiene como objetivo particionar n observaciones en k clusters o grupos, donde cada observación pertenece al clúster con la media más cercana (es decir, los centros de clúster o clúster centroide), actuando como prototipo del grupo.

²⁰ Ejemplo de ello, son los bosques aleatorios para árboles de regresión, que si bien representan uno de los algoritmos más importantes y usados en machine learning. Realizan predicciones sobre nuevas observaciones combinando las predicciones de todos los árboles que conforman el modelo. Su potencial radica en que métodos estadísticos y de machine learning basados en estos engloban técnicas supervisadas no paramétricas que consiguen fraccionar el espacio de predictores en regiones simples, donde es más sencillo manejar las interacciones. Estos modelos son más robustos

Dentro de los supervisados, se consideró la regresión lineal multivariada²¹. Los modelos de regresión lineal simple, múltiple o multivariado buscan obtener la regresión entre variables independientes y una variable dependiente. Es decir, teniendo una serie de variables predictoras, obtiene la relación con una variable cuantitativa a predecir; la regresión lineal explica la variable y con las variables x , generando una función lineal que mejor se ajusta. Pero en la práctica generalmente tenemos n variables predictoras, por ello podríamos hacer n regresiones lineales simples; sin embargo, cada una ignoraría a las otras $n-1$ variables y no se sacaría ventaja de las relaciones entre variables. Por ende, será necesario considerar cómo dos o más variables independientes influyen sobre una variable dependiente (Hastie, Tibshirani y Friedman, 2009). Estos modelos pueden emplearse para predecir el valor de una variable dependiente o para evaluar la influencia que tienen los predictores sobre ella (esto último se debe analizar con cautela para no malinterpretar causa-efecto).

Adicionalmente, se incorporó un análisis estadístico de supervivencia utilizando la técnica de Kaplan-Meier (Kaplan y Meier, 1958). Este enfoque proporciona una comprensión detallada de la duración de eventos en función del tiempo, particularmente valioso para eventos de interés en el estudio. Se exploraron patrones de supervivencia a lo largo del tiempo, permitiendo identificar posibles variaciones en la duración de eventos relevantes para el análisis.

Finalmente, se tuvieron en consideración las localidades de los pacientes. Se puede entender que los sistemas de información geográficos (SIG) son una herramienta importante para el estudio, análisis y control epidemiológico, pues a la información médica, se suman datos geográficos, variables ambientales, socioeconómicas, entre otras, que en el contexto de la pandemia de COVID-19 podrían revelar los retrocesos en la prevención así como el empeoramiento de la calidad de vida de las personas por el aumento de la pobreza y la accesibilidad a los sistemas de salud (OPS, 2002; Cuello-Rüttler y Gudiño, 2017). Por lo anterior, se entiende que el contexto geográfico agrega información cualitativa a los datos, permitiendo revelar si existen o no patrones geográficos nacionales respecto a los enfermos de tuberculosis antes y después de la pandemia. Con este fin, se utilizaron las localidades censales, descargados en GeoJSON desde el Sistema de almacenamiento de archivos y catálogos de la Red de Nodos de Datos Abiertos de la Administración Pública Nacional²². Se contabilizaron los casos por cada localidad y se representaron en mapas con nodos proporcionales.

En resumen, el análisis abarcó desde la regresión, que modela las relaciones entre variables, hasta la georreferenciación, que explora la dimensión espacial de los datos, y culminó con un análisis de supervivencia que profundiza en la duración de eventos de interés a lo largo del tiempo. Esta combinación de enfoques proporciona una visión integral del conjunto de datos, permitiendo interpretaciones más robustas y contextualizadas.

3. Análisis de resultados

Los resultados se evaluaron comenzando con el recorte a la base de datos descargada, como se mencionó en la [metodología](#). Se examinaron 36.391 casos de tuberculosis notificados en Argentina durante el período comprendido entre los años 2019 y 2021, con el objetivo de analizar posibles patrones y variaciones en la incidencia de la enfermedad, intermediada por el aislamiento social de la pandemia de COVID-19. Analíticamente se ha visualizado la cronología de los casos, dividiéndolos en aquellos notificados antes del inicio de la pandemia ($n= 16.042$) y aquellos registrados después de la declaración del ASPO ($n=20.349$). Estos valores corresponden únicamente a los reportados, sin duplicados, entre 2019 y 2021. De esta manera, se puede apreciar en esta distinción temporal la posible influencia de la pandemia en la notificación y gestión de casos de tuberculosis, brindando una perspectiva más completa sobre la dinámica de la enfermedad en el contexto de eventos sanitarios imperantes. En la tabla siguiente (Tabla 1a) se muestra la cantidad total de casos discriminados por sexo; en la tabla 1b se presentan los datos con filtros mencionados aplicados, incluso con duplicados eliminados, y en la tabla 2 los tratamientos que figuran como curados o con tratamiento

ante valores atípicos, pues asignan menor peso a las observaciones individuales, por lo que los valores atípicos no influyen tanto en la toma de decisiones.

²¹ Puede verse más información sobre el tema en: [Regresión lineal múltiple](#).

²² Las localidades censales se encuentran disponibles desde:

<https://infra.datos.gob.ar/catalog/modernizacion/dataset/7/distribution/7.31/download/localidades-censales.geojson>.

completo contra la TB.

Tabla 1a. Casos totales según distinción por sexo.

Clasificación respecto a la pandemia	Antes	Después
Sexo		
A	2.979	1.467
F	36.877	46.395
M	62.260	70.942
X	0	12

Fuente: Elaboración propia.

Tabla 1b. Casos filtrados, sin duplicados, según distinción por sexo.

Clasificación respecto a la pandemia	Antes	Después
Sexo		
A	553	374
F	6.504	8.455
M	8.985	11.518
X	0	2

Fuente: Elaboración propia.

Tabla 2. Casos con tratamiento completo o curados.

Clasificación Pandemia	Antes	Después
Resultado Tratamiento		
Curado	1.015	1.285
Tratamiento completo	4.680	6.724

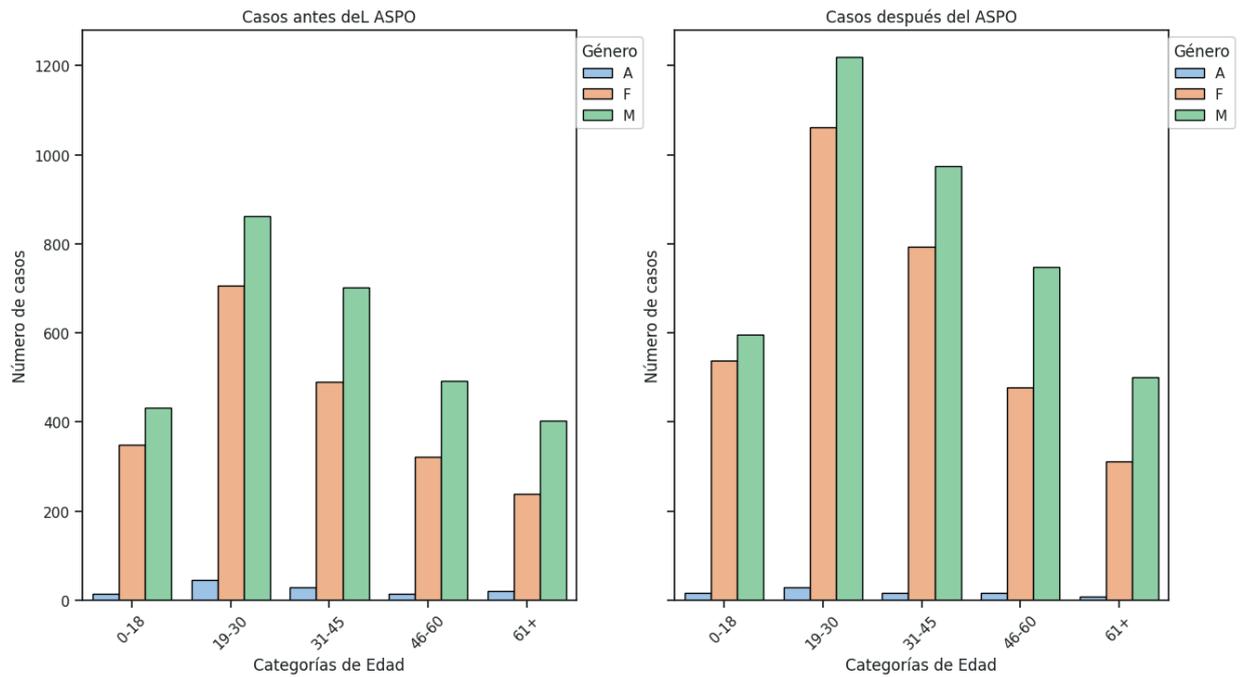
Fuente: Elaboración propia.

Es importante destacar que se optó por analizar los casos con resultado de tratamiento exitoso (curados y tratamiento completo; n=12.505), de los cuales 5.158 casos fueron notificados antes del ASPO y 7.347 casos después (Tabla 2). Al seleccionar casos con resultados exitosos, se buscó la variabilidad inherente en los resultados de tratamiento y reducir la influencia de factores confusos que podrían afectar la comparación, con el objetivo de mejorar la validez de las conclusiones derivadas del análisis.

3.1. Análisis descriptivo de los casos de tuberculosis notificados entre 2019 y 2021

La distribución de casos según grupos de edad se aprecia en el siguiente gráfico, dónde se destaca una mayor incidencia en el rango de 19 a 60 años durante ambos períodos analizados.

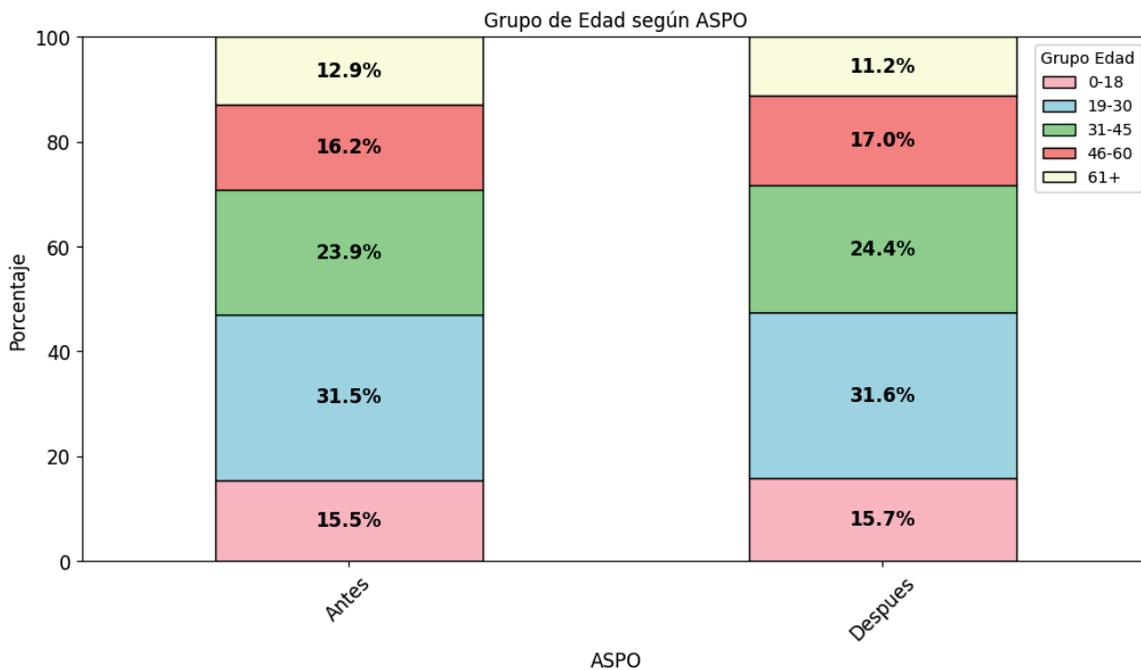
Gráfico 2. Notificación total de casos de tuberculosis notificados antes y después del ASPO según grupo edad, en Argentina.



Fuente: elaboración propia

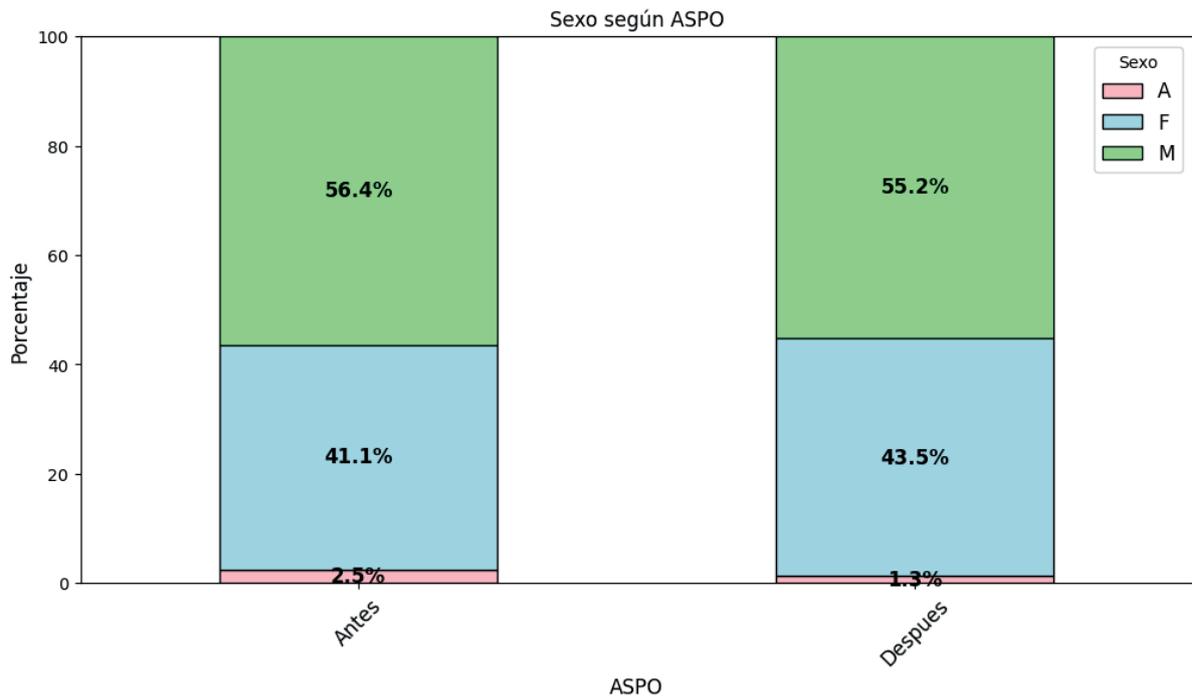
Asimismo, la partición porcentual para los grupos de edad (Gráfico 3) y género (Gráfico 4) evidencia una marcada semejanza.

Gráfico 3. Distribución de la proporción de casos notificados antes y después del ASPO según grupo edad.



Fuente: elaboración propia

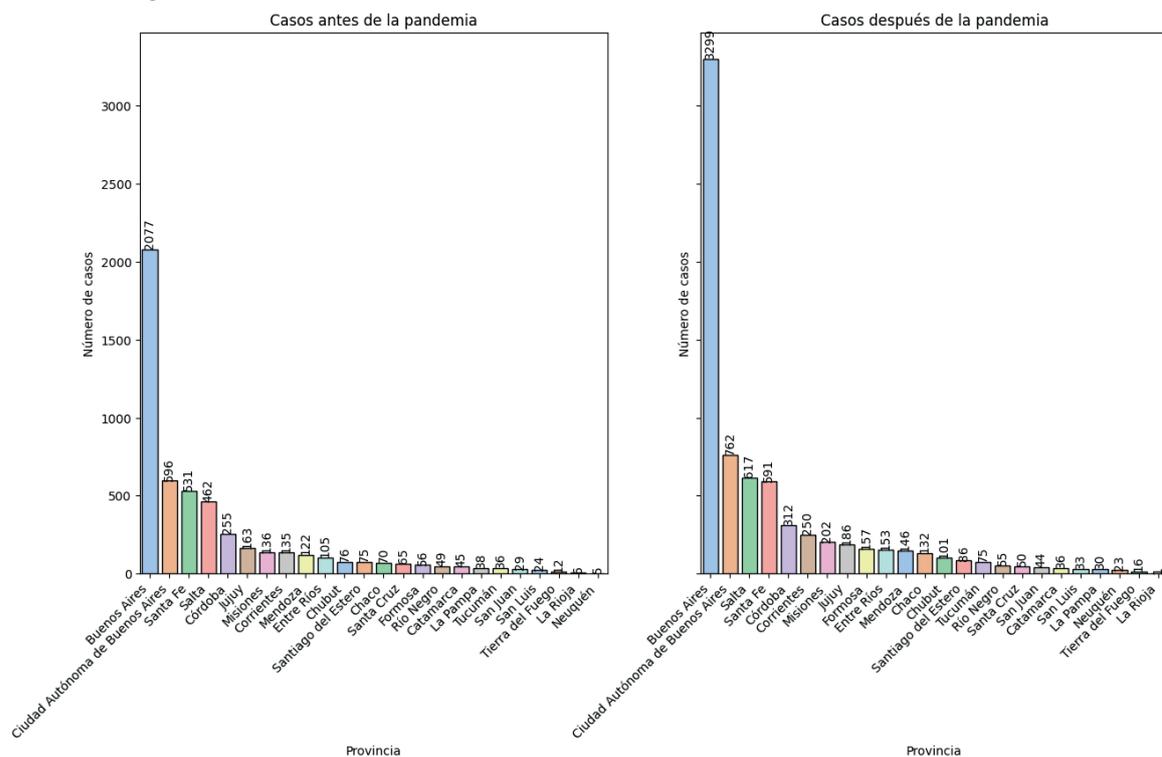
Gráfico 4. Distribución de la proporción de casos notificados antes y después del ASPO según género. (Argentina, 2019 - 2021).



Fuente: elaboración propia

Al considerar pacientes por provincia de residencia (Gráfico 5), la gráfica presenta un aumento en las notificaciones posteriores a la declaración de ASPO durante la pandemia. Las que presentaron mayor cantidad de casos continúan siendo, como antes del 2020, Buenos Aires, Ciudad Autónoma de Buenos Aires (CABA), Santa Fe, Salta y Córdoba.

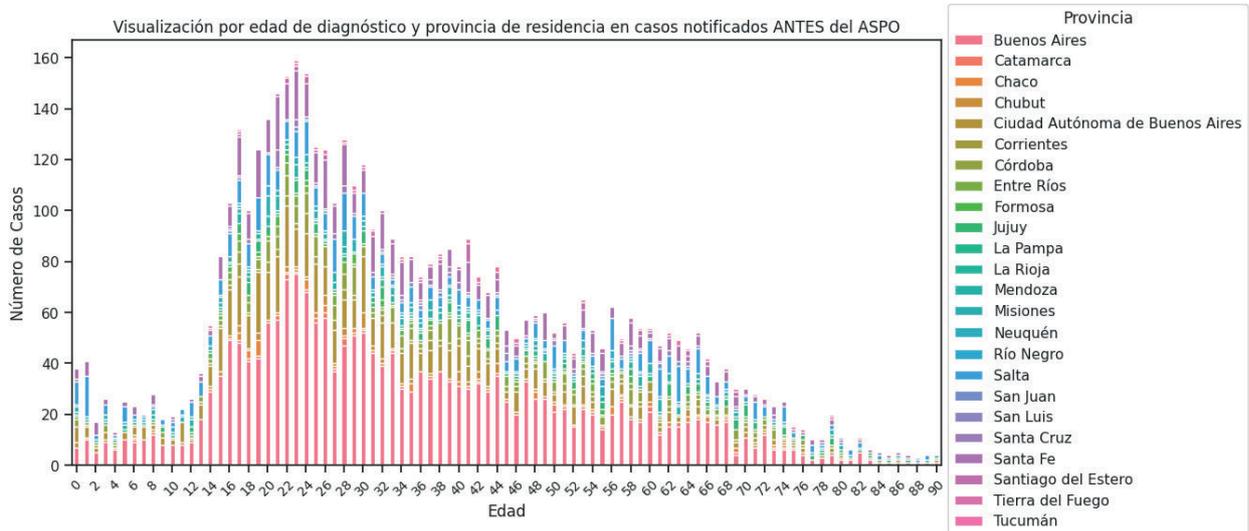
Gráfico 5. Distribución del número de casos notificados antes y después del ASPO según provincia de residencia. (Argentina, 2019 - 2021).



Fuente: elaboración propia

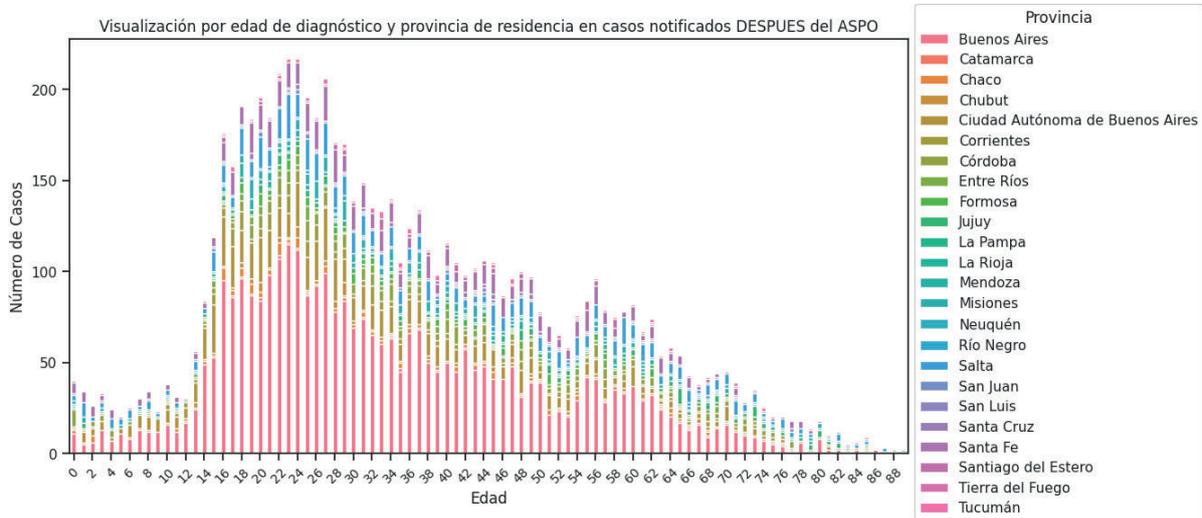
Por otro lado, la variaciones en edades y en provincia de residencia, antes de la pandemia y después del ASPO revela una notable similitud en ambas gráficas (Gráficos 6a y b), con una predominancia de casos notificados en el grupo de pacientes comprendido entre 15 y 40 años.

Gráficos 6a: Distribución por edad de diagnóstico según provincias de residencias en casos de tuberculosis notificados antes del ASPO.



Fuente: Elaboración propia.

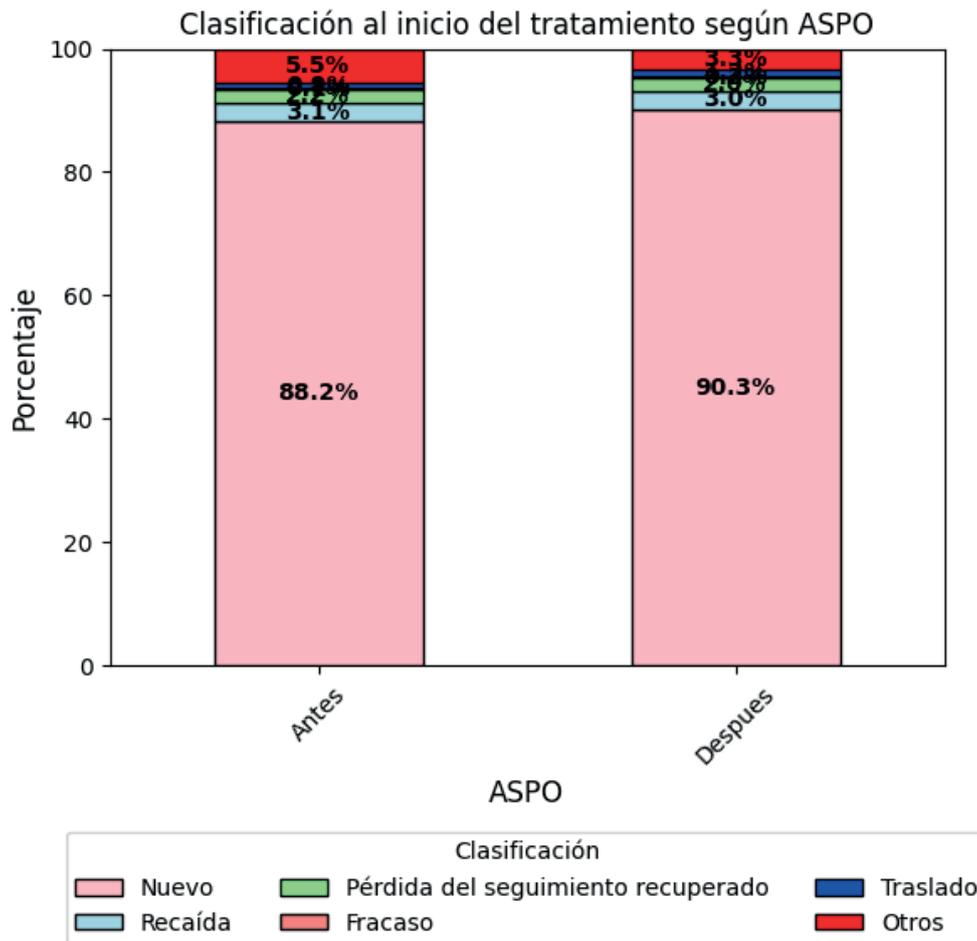
Gráficos 6b: Distribución por edad de diagnóstico según provincias de residencias en casos de tuberculosis notificados después del ASPO.



Fuente: elaboración propia.

En el siguiente gráfico (Gráfico 7) se visualiza que más del 80% de los casos en ambos grupos estudiados se clasificaron como nuevos, es decir que nunca habían sido tratados por tuberculosis o recibieron medicamentos anti-TB al menos por un mes.

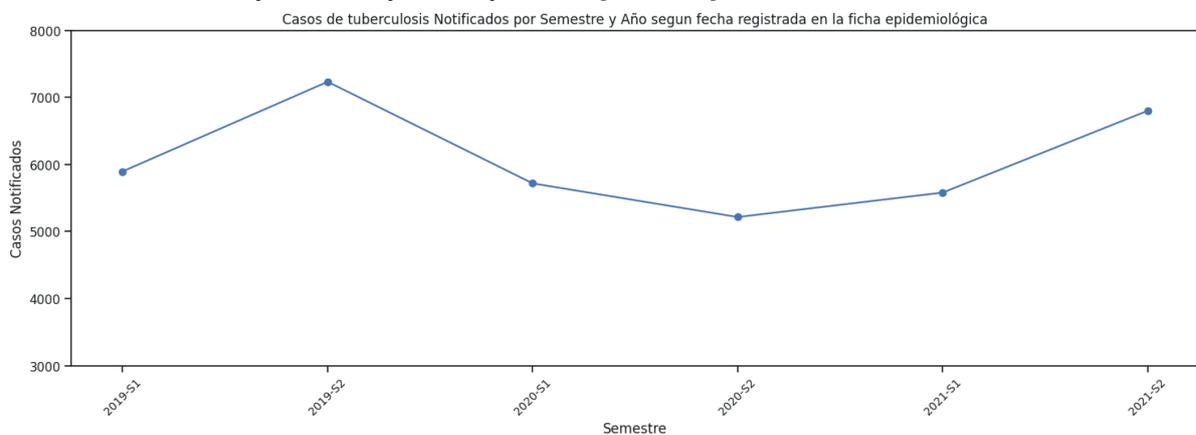
Gráfico 7. Distribución de casos notificados antes y después del ASPO según clasificación del caso al inicio del tratamiento. (Argentina, 2019 - 2021).



Fuente: elaboración propia

A continuación (Gráfico 8), en la evolución semestral de casos notificados, se observa una disminución en los registros a partir del primer trimestre de 2020 en comparación con el segundo semestre del año anterior. Mientras que desde el primer semestre de 2021, se percibe un aumento, sumado a que la tendencia se mantiene hasta el segundo semestre del mismo año.

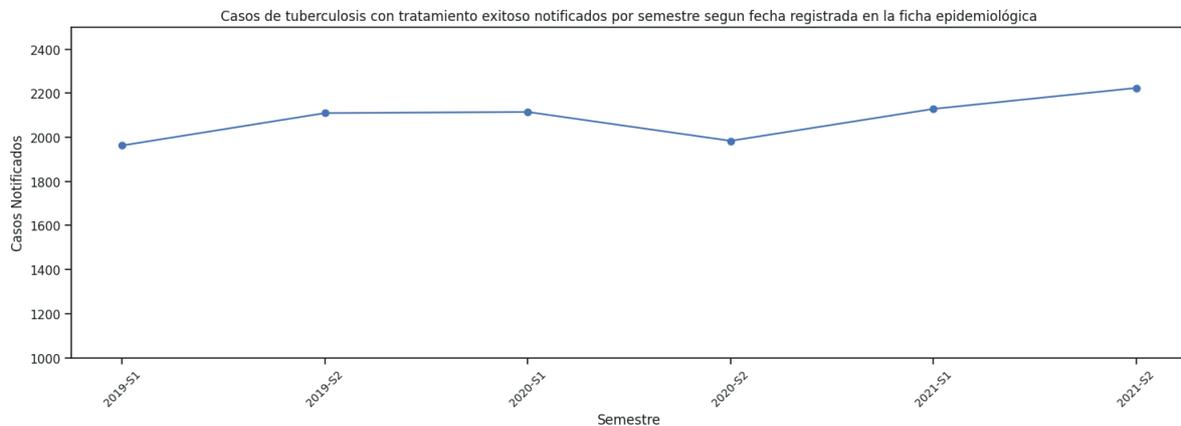
Gráficos 8. Evolución por semestre del total de casos de tuberculosis notificados, según fecha registrada en la ficha de notificación epidemiológica, en Argentina (2019-2021).



Fuente: elaboración propia

Cabe destacar que al analizar la evolución de las notificaciones de tratamientos exitosos, si bien se observa un patrón similar al del total de casos, esta disminución no es tan marcada como en el total de casos (Gráfico 9).

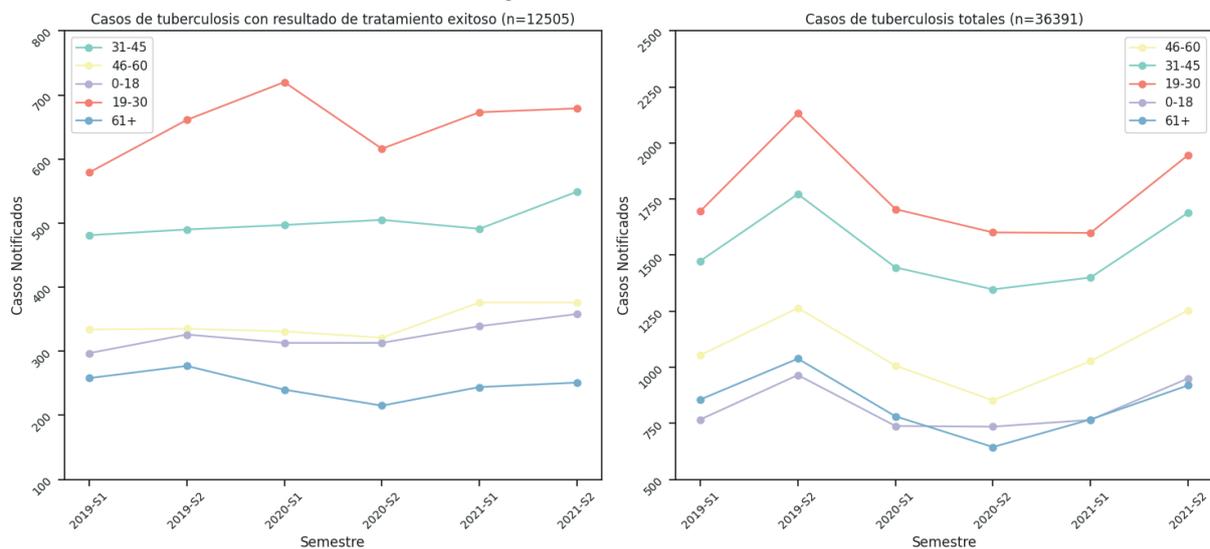
Gráficos 9. Evolución por semestre casos de tuberculosis notificados con resultado de tratamiento exitoso, según fecha registrada en la ficha de notificación epidemiológica, en Argentina (2019-2021).



Fuente: elaboración propia

Sin embargo, al examinar por grupo de edad (Gráfico 10), se destaca una variación consistente en la mayoría de los conjuntos durante el periodo considerado, con una excepción notable en los menores de 18 años. Este último, presenta un descenso durante el primer semestre de 2020, que persiste en el segundo semestre, y experimenta un aumento en el primero del año siguiente. De manera similar, al focalizar solo en casos con tratamiento exitoso existe una mayor variabilidad en cada agrupamiento etario. Sin embargo, es notable que dentro del grupo de 0 a 18 años y mayores de 31 años la variabilidad no es tan marcada.

Gráficos 10. Evolución semestral de la incidencia de casos de tuberculosis totales y casos con resultado de tratamiento exitoso, por grupo edad según fecha registrada en la ficha de notificación epidemiológica, en Argentina (2019-2021).



Fuente: elaboración propia

3.2. Análisis del tiempo de demora en el diagnóstico y tratamiento de la tuberculosis.

Para este análisis se han generado histogramas (Gráfico 11 y 12) que evalúan la distribución de tiempos de demora (calculado en días) en los casos notificados antes y después del ASPO. Estos gráficos revelan asimetría hacia la derecha, donde la media supera tanto a la mediana como a la moda en ambas categorías. Se puede entender que este patrón sugiere una distribución no simétrica. De manera adicional el histograma del

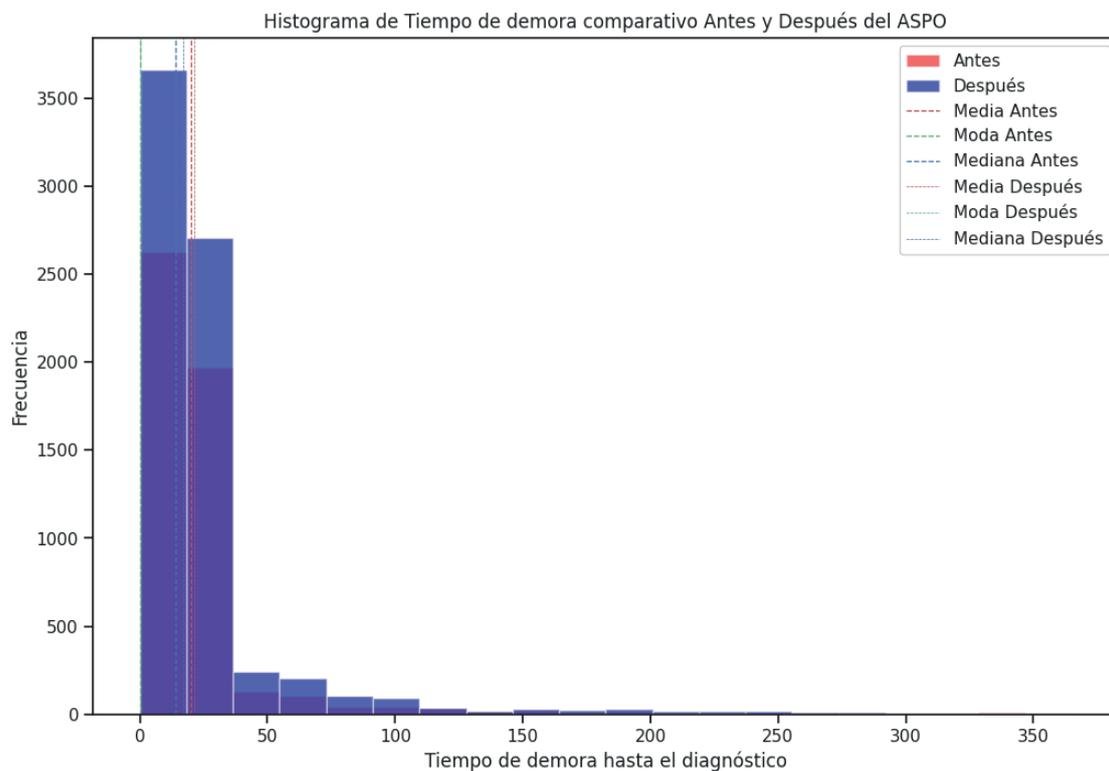
tiempo de demora en el tratamiento (TDT) destaca la presencia de dos modas, indicando una distribución bimodal. Por otro lado, en la Tabla 3, se evidencia una leve variabilidad en tiempos de demora, antes y después de la pandemia, análogamente ocurre en la media como en la mediana. Es decir, que la media aumentó aproximadamente día y medio, mientras que la mediana se incrementó en 3 días, tanto en la demora en el diagnóstico como en el tratamiento.

Tabla 3: Medidas de tendencia central del tiempo de demora hasta el diagnóstico y hasta completar el tratamiento de casos de Tuberculosis con Tratamiento Exitoso.

Clasificación pandemia	Tiempo de Demora (Diagnóstico) - Media	Tiempo de Demora (Diagnóstico) - Mediana	Tiempo de Demora (Tratamiento) - Media	Tiempo de Demora (Tratamiento) - Mediana
Antes	22.55	14.0	227.36	216.0
Después	23.81	17.0	229.86	219.0

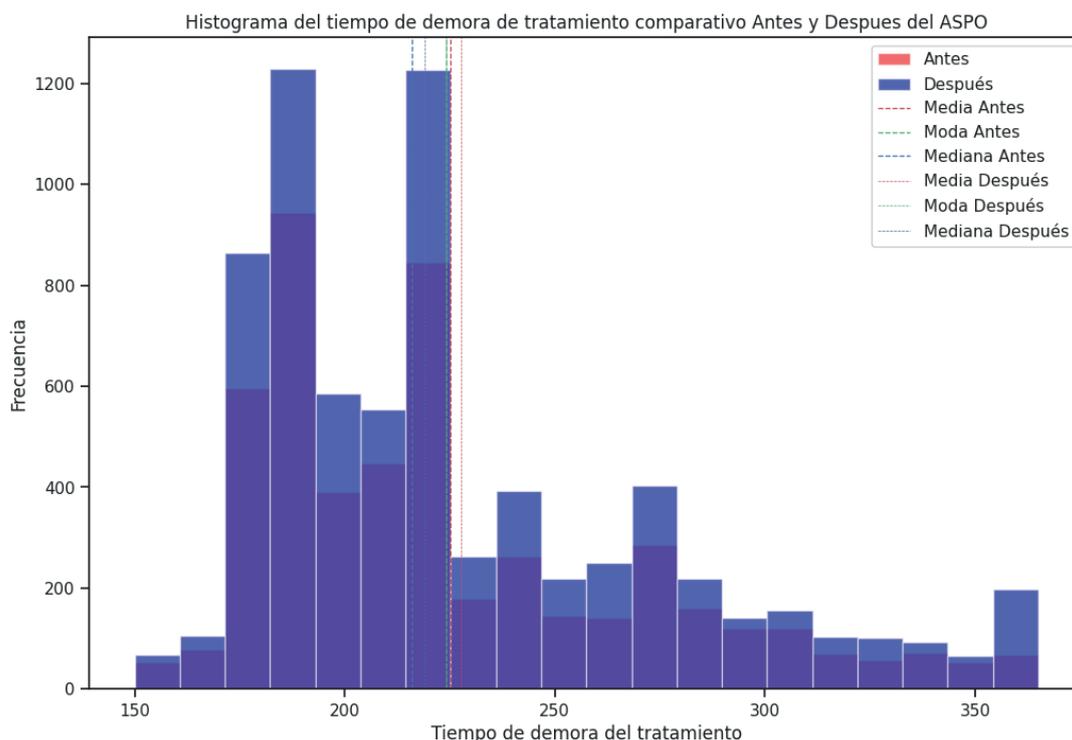
Fuente: elaboración propia.

Gráfico 11. Distribución del Tiempo de Demora en el Diagnóstico de Casos de Tuberculosis con Tratamiento Exitoso.



Fuente: elaboración propia

Gráfico 12. Distribución del Tiempo de Demora hasta completar el tratamiento en Casos de Tuberculosis con Tratamiento Exitoso.

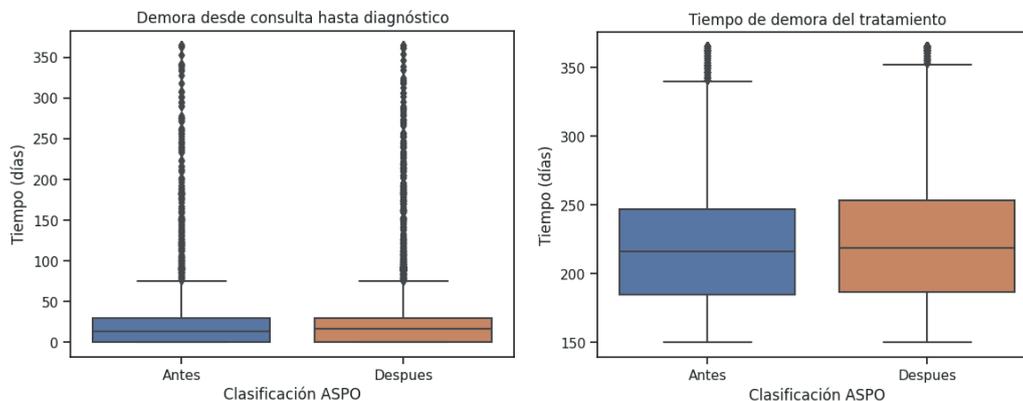


Fuente: elaboración propia

Cabe destacar que si la mediana es mayor a la media en un conjunto de datos, sugiere que la distribución de los datos está sesgada hacia la derecha, es decir, por ende refleja la existencia de valores atípicos o extremadamente altos. Estos son los que aumentan la mediana en comparación con la media. Por ello, en este tipo de situaciones podría ser más pertinente utilizar la mediana como medida de tendencia central en lugar de la media, pues la mediana es menos sensible a los valores atípicos.

Los valores atípicos se presentan en el gráfico 13, sin reflejar variaciones considerables en el tiempo de demora hasta el diagnóstico. Similarmente, la mediana como los cuartiles son similares en ambos periodos. No obstante, luego del ASPO se aprecia un ligero aumento en el tiempo de demora hasta finalizar el tratamiento, y un mayor intervalo de confianza, lo que indica una mayor variabilidad en los tiempos de demora.

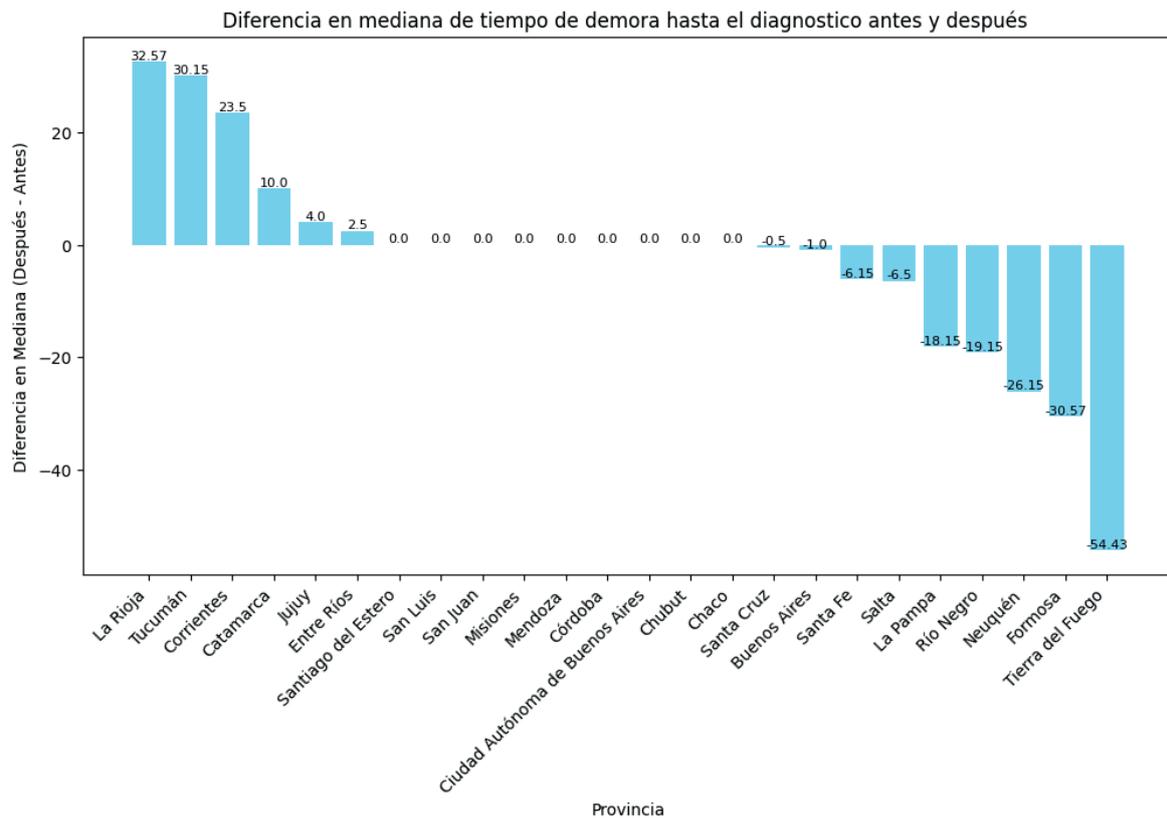
Gráfico 13: Tiempos de Demora en el Diagnóstico y Tratamiento de casos de Tuberculosis en Argentina según momento de notificación (2019-2021).



Fuente: elaboración propia.

En cambio, al analizar cada provincia (Gráfico 14), se visualiza que la diferencia en la mediana del tiempo hasta el diagnóstico no se comporta de la misma manera que en el total de los datos. Se aprecian importantes disimilitudes, por ejemplo en algunas provincias no presentan distinciones entre los periodos analizados, en tanto que en otras como La Rioja, Tucumán, Corrientes, Catamarca, Jujuy y Entre Ríos el tiempo de demora fue mayor después de declararse el ASPO. Por el contrario en Santa Fe, Salta, La Pampa, Río Negro, Neuquén, Formosa y Tierra del Fuego fue menor.

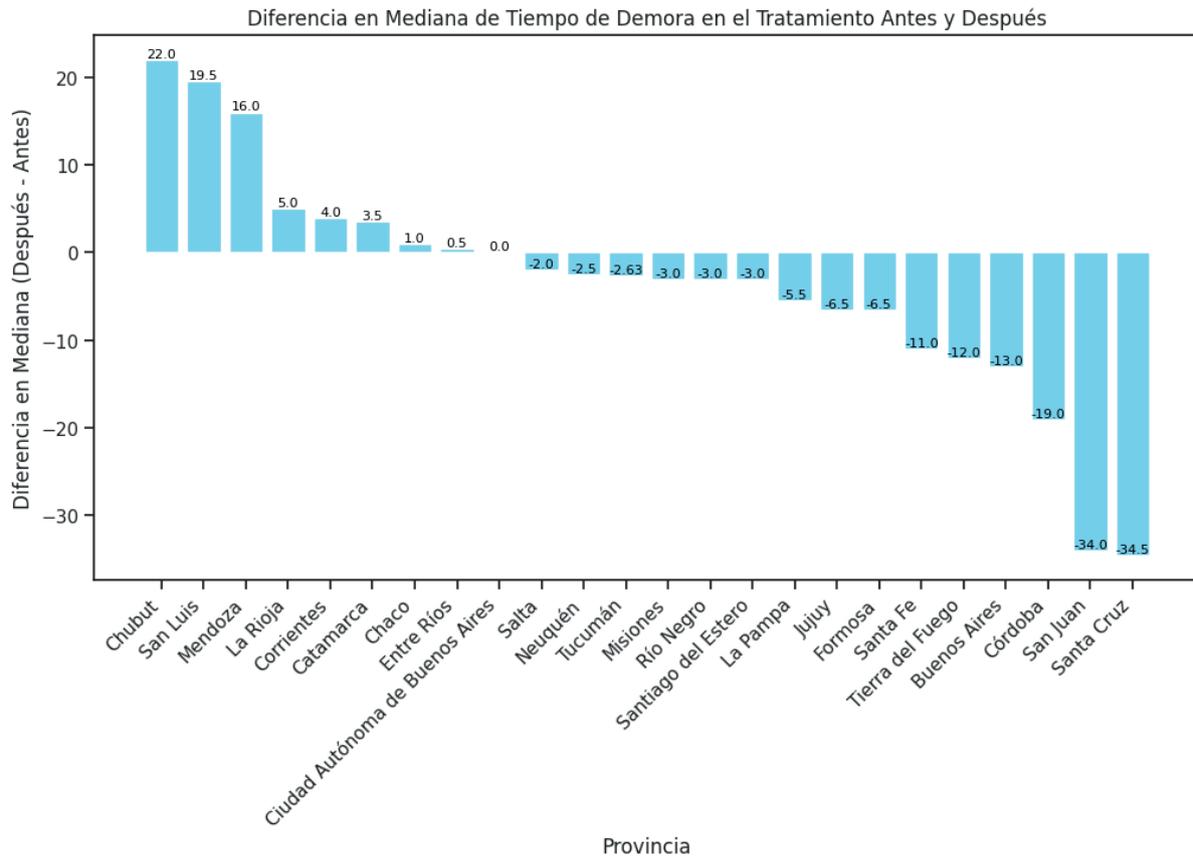
Gráfico 14: Diferencia de la mediana en el tiempo de demora hasta el diagnóstico de casos de Tuberculosis con Tratamiento Exitoso en según provincia en Argentina (2019-2021).



Fuente: elaboración propia.

La diferencia en los tiempos de demora hasta completar el tratamiento (Gráfico 15) se destaca una menor variación en las provincias de Entre Ríos y CABA, inferior a un día, tanto antes como después del ASPO. En contraste, Chubut, San Luis y Mendoza experimentaron un aumento, que superaron los 15 días de diferencia en el tiempo de tratamiento después del ASPO. No obstante, en San Juan y Santa Cruz, a pesar de presentar una mayor dispersión, reducen sus tiempos de tratamiento posterior al aislamiento, con discrepancias superiores a 30 días. Estos datos resaltan los antagonismos del impacto de la pandemia en los tiempos de tratamiento entre distintas provincias, con incrementos y reducciones en la demora. Lo anterior, acentúa lo fundamental de mejorar la eficiencia en la prestación de servicios de salud en cada región.

Gráfico 15: Diferencia de la mediana en el tiempo de demora hasta completar el tratamiento de casos de Tuberculosis con Tratamiento Exitoso en según provincia en Argentina (2019-2021).



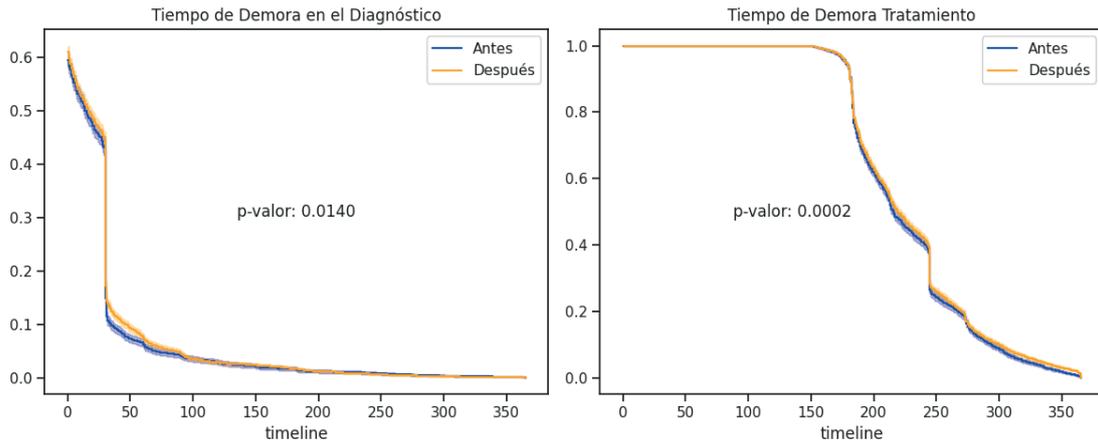
Fuente: elaboración propia.

3.3 Análisis estadístico de supervivencia

Los gráficos de supervivencia grafican la probabilidad de que los eventos, en este caso los tiempos de demora en el diagnóstico y de demora en el tratamiento, no hayan ocurrido hasta un tiempo dado. En el contexto del presente artículo se tienen en cuenta dos periodos distintos *antes* y *después* de la declaración del ASPO, en el gráfico 16 las líneas azul y naranja, respectivamente, representan la supervivencia estimada. En ambos conjuntos, las curvas difieren entre dichos lapsos, y sugieren cambios en la distribución de los tiempos de demora. Es por ello, que para respaldar estas observaciones, se utilizó la prueba de *log-rank*, herramienta estadística que evalúa la existencia de diferencias significativas en las funciones de supervivencia entre dos grupos. Por medio de este test se obtuvo que los *p-valores* son menores a 0.05²³ para ambos conjuntos de datos, e indicó que existen diferencias estadísticamente significativas entre los tiempos de demora en el diagnóstico y en la demora en el tratamiento respecto al antes y después del aislamiento. En consecuencia, se concluye que la clasificación mencionada está asociada a los cambios significativos en estos tiempos.

²³ Se señala que el p-valor = 0.014 en la demora en el diagnóstico y el p-valor = 0.0002 en la demora en el tratamiento.

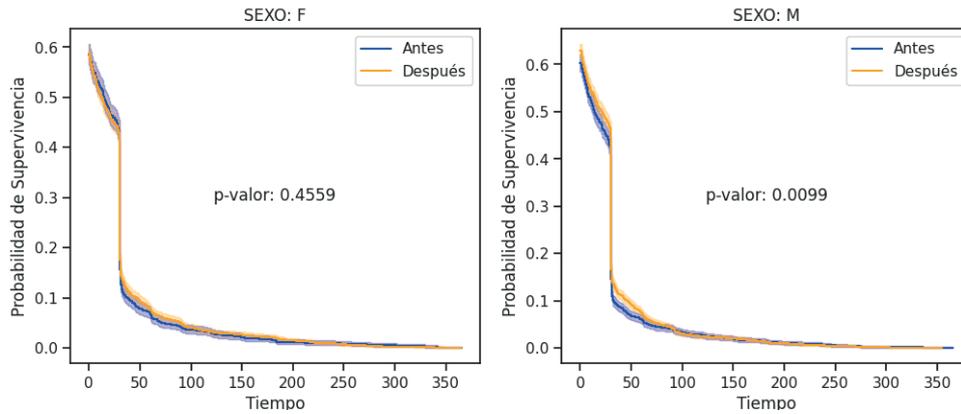
Gráfico 16. Análisis de Supervivencia: Tiempos de Demora en Casos Notificados antes y después del ASPO en Argentina (2019-2021)



Fuente: elaboración propia.

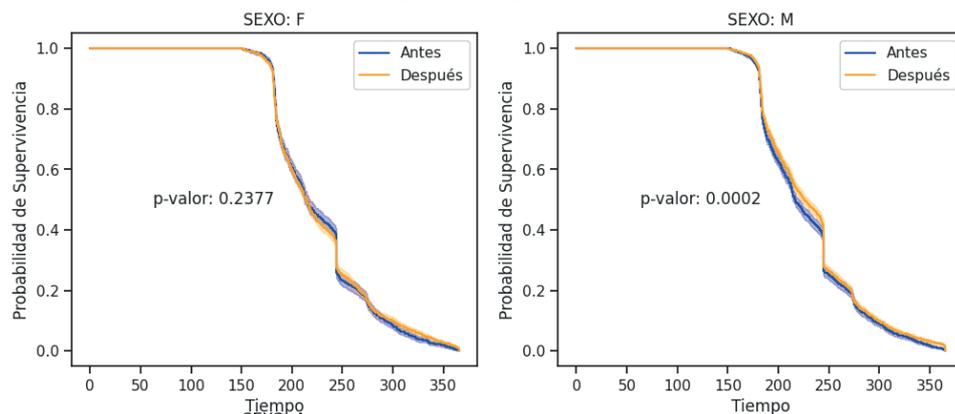
Al considerar los sexos en gráficos de supervivencia (Gráficos 17 y 18), en el femenino, no se observan diferencias significativas en ninguno de los tiempos evaluados. Esta consistencia sugiere que las dinámicas en torno al diagnóstico y tratamiento de la tuberculosis en dicho grupo permanecieron relativamente estables. En contraste, en el grupo masculino, la diferencia significativa se refleja antes y después del ASPO en la demora en el diagnóstico y en la duración del tratamiento. Estas evidencias indican que las medidas de aislamiento reflejaron un impacto específico en los patrones de la TB en hombres, introduciendo variaciones notables en la probabilidad de ser diagnosticado y en la progresión durante el tratamiento.

Gráfico 17. Análisis de Supervivencia: Tiempos de Demora en el Diagnóstico de Casos Notificados antes y después del ASPO en Argentina (2019-2021)



Fuente: Elaboración propia.

Gráfico 18. Análisis de Supervivencia: Tiempos de Demora en el Tratamiento de Casos Notificados antes y después del ASPO en Argentina según sexo (2019-2021)

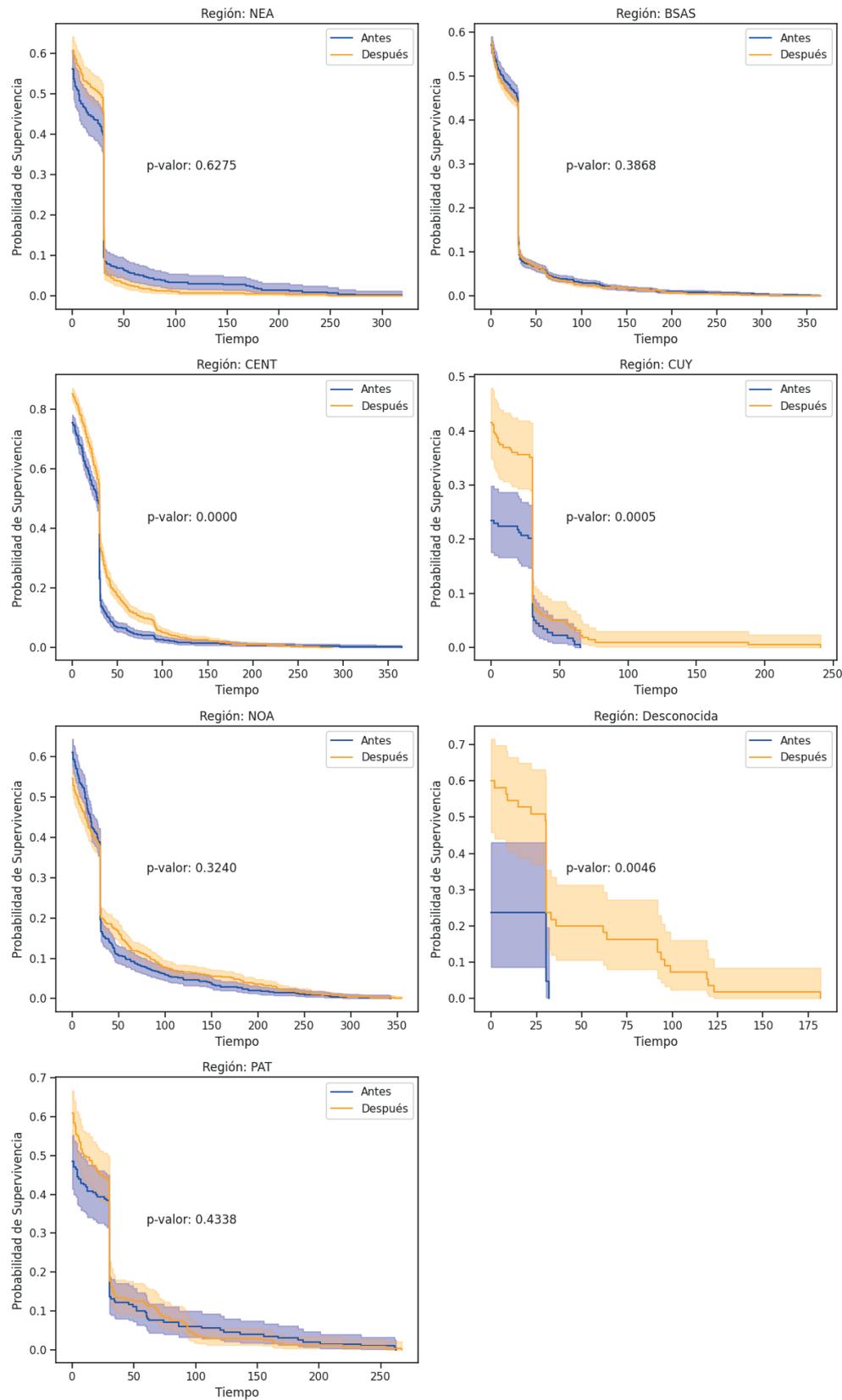


Fuente: elaboración propia.

En lo que respecta a tiempos de demora en el diagnóstico según las diferentes regiones de Argentina (Gráfico 19)²⁴, se contempla que no hay uniformidad en la probabilidad de supervivencia. En particular, las regiones Centro y Cuyo exhiben diferencias estadísticamente significativas en ambos periodos (p -valor < 0.05), y sugiere que las variaciones en los tiempos de demora antes y después del ASPO son significativas y no pueden atribuirse simplemente al azar. En cambio, en las regiones restantes, los p -valores no alcanzaron el umbral crítico de 0.05, es decir no hay diferencias significativas. Esto implica que las disparidades en tiempos de demora podrían deberse a variaciones aleatorias o no sistemáticas, en lugar de patrones discernibles.

²⁴ Se hizo una separación en las siguientes regiones. NOA: Catamarca, Jujuy, Salta, Santiago del Estero y Tucumán; NEA: Chaco, Corrientes, Formosa, Misiones; Patagonia: Chubut, Neuquén, Río Negro, Santa Cruz y Tierra del Fuego; Centro: Córdoba, Entre Ríos, Santa Fe, La Pampa; Cuyo: Mendoza, La Rioja, San Juan, San Luis; BSAS: Buenos Aires y CABA.

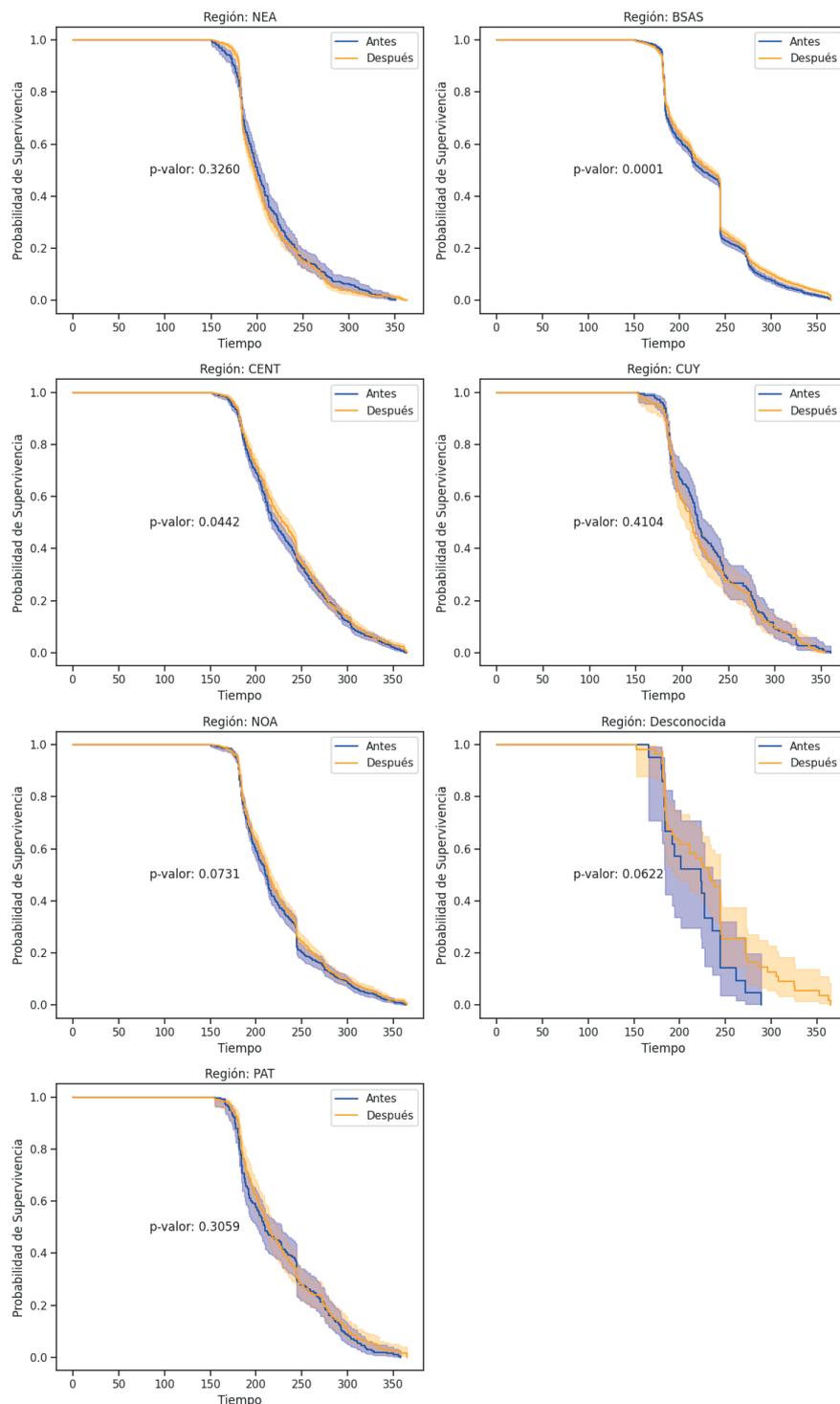
Gráfico 19. Análisis de Supervivencia: Tiempos de Demora en el Diagnóstico de Casos Notificados antes y después del ASPO en Argentina (2019-2021)



Fuente: elaboración propia.

En lo que respecta a los tiempos de demora en el tratamiento (Gráfico 20), se observa que las regiones de Buenos Aires y Centro, fueron las únicas que mostraron un p-valor significativamente inferior a 0.05. Estas diferencias podrían estar relacionadas con diversos factores, como la infraestructura de salud, los recursos disponibles o políticas regionales específicas que fueron tomadas durante el ASPO e impactaron en la eficiencia de los procesos de tratamiento.

Gráfico 20. Análisis de Supervivencia: Tiempos de Demora en el Tratamiento de Casos Notificados antes y después del ASPO en Argentina (2019-2021)



Fuente: elaboración propia.

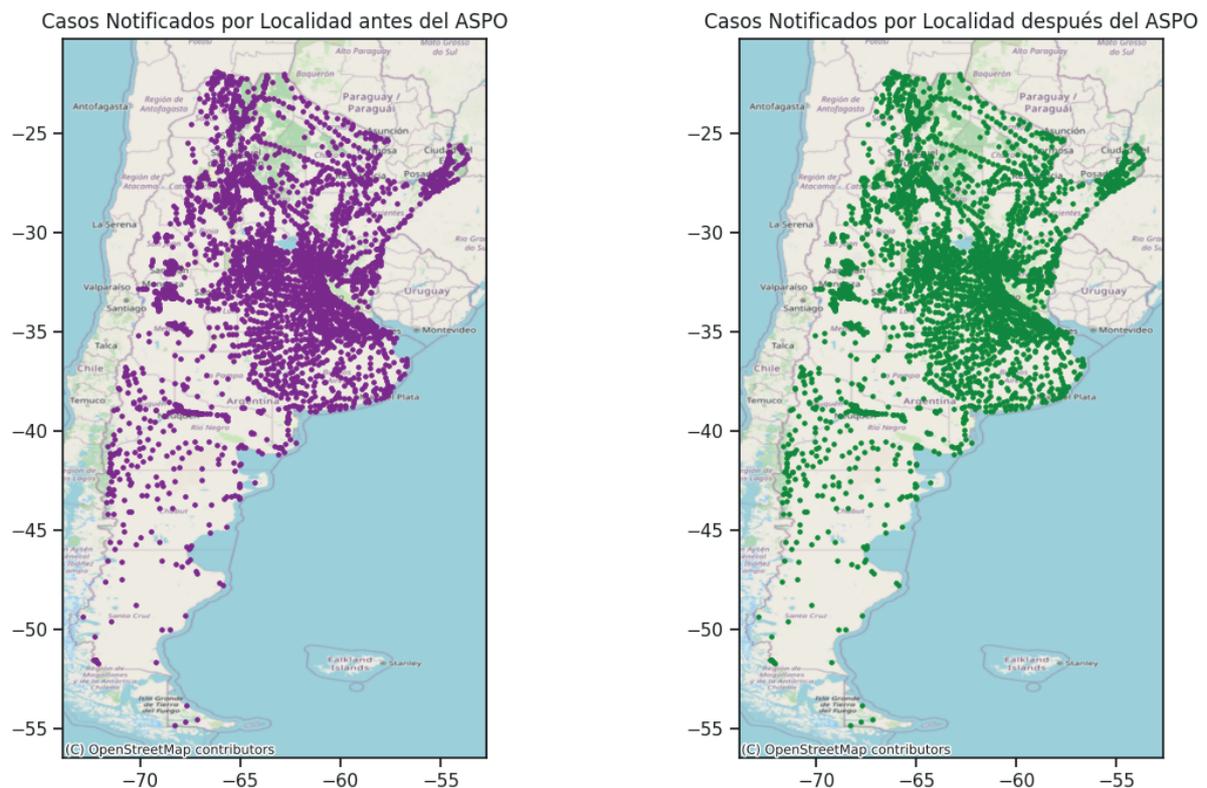
3.4 Georreferenciación de datos

En [metodología](#) se refirió a las herramientas y bases de datos utilizadas para realizar el solapamiento de datos georreferenciados y presentar las proporciones entre cantidad de casos y regiones, provincias o localidades.

3.4.1 Visualización de casos en localidades antes de la pandemia

Las georreferencias de localidades (Gráfico 23) reportaron los casos antes como después de la pandemia buscando evaluar posibles patrones geográficos. Asimismo, no se presentan diferencias sustanciales, presentando distribuciones uniformes en ambos periodos así como en regiones.

Gráfico 23: Georreferenciación de casos de tuberculosis según localidad de residencia de casos notificados antes y después del ASPO.

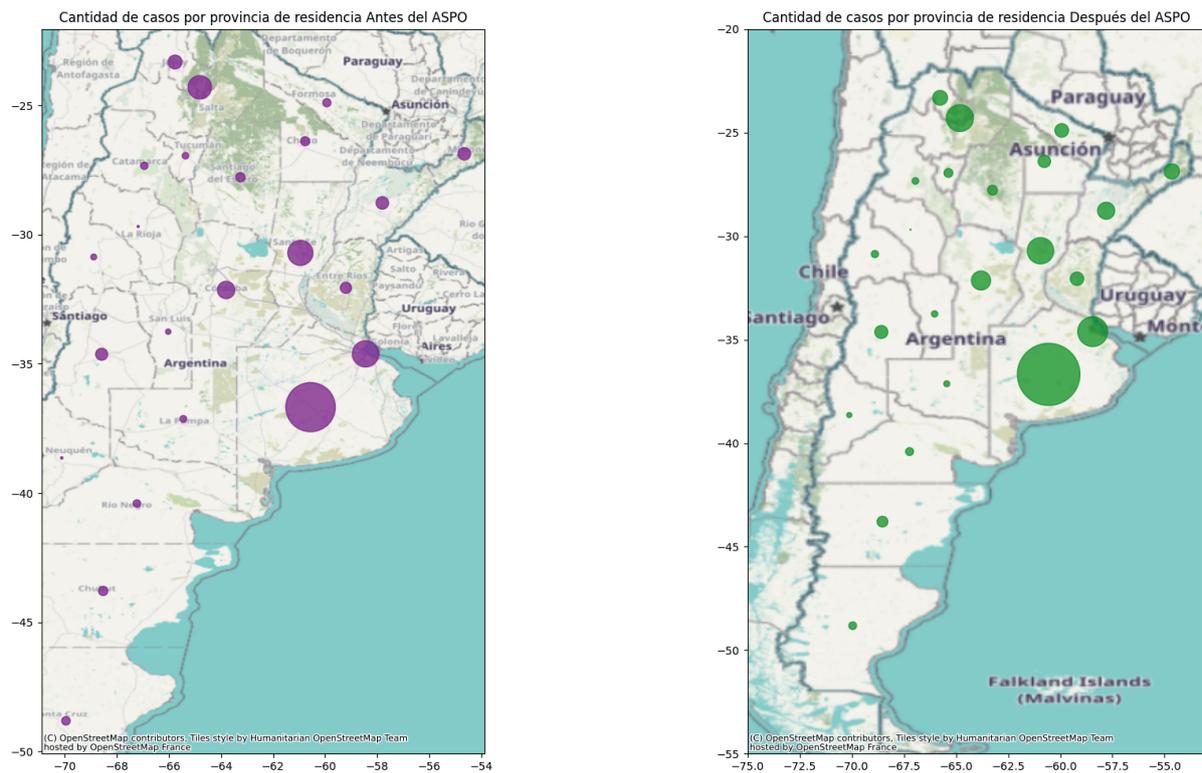


Fuente: elaboración propia.

3.4.2. Mapa con conteo de casos por provincias

Al momento de considerar la relación entre provincias y densidad de casos, antes y después de la pandemia, el patrón espacial refleja similitudes, aunque solo evidencia mayor cantidad de casos en algunas provincias, como la de Buenos Aires (Gráfico 24).

Gráfico 24. Georreferenciación según densidad de conteo de casos notificados por provincia antes y después del ASPO.



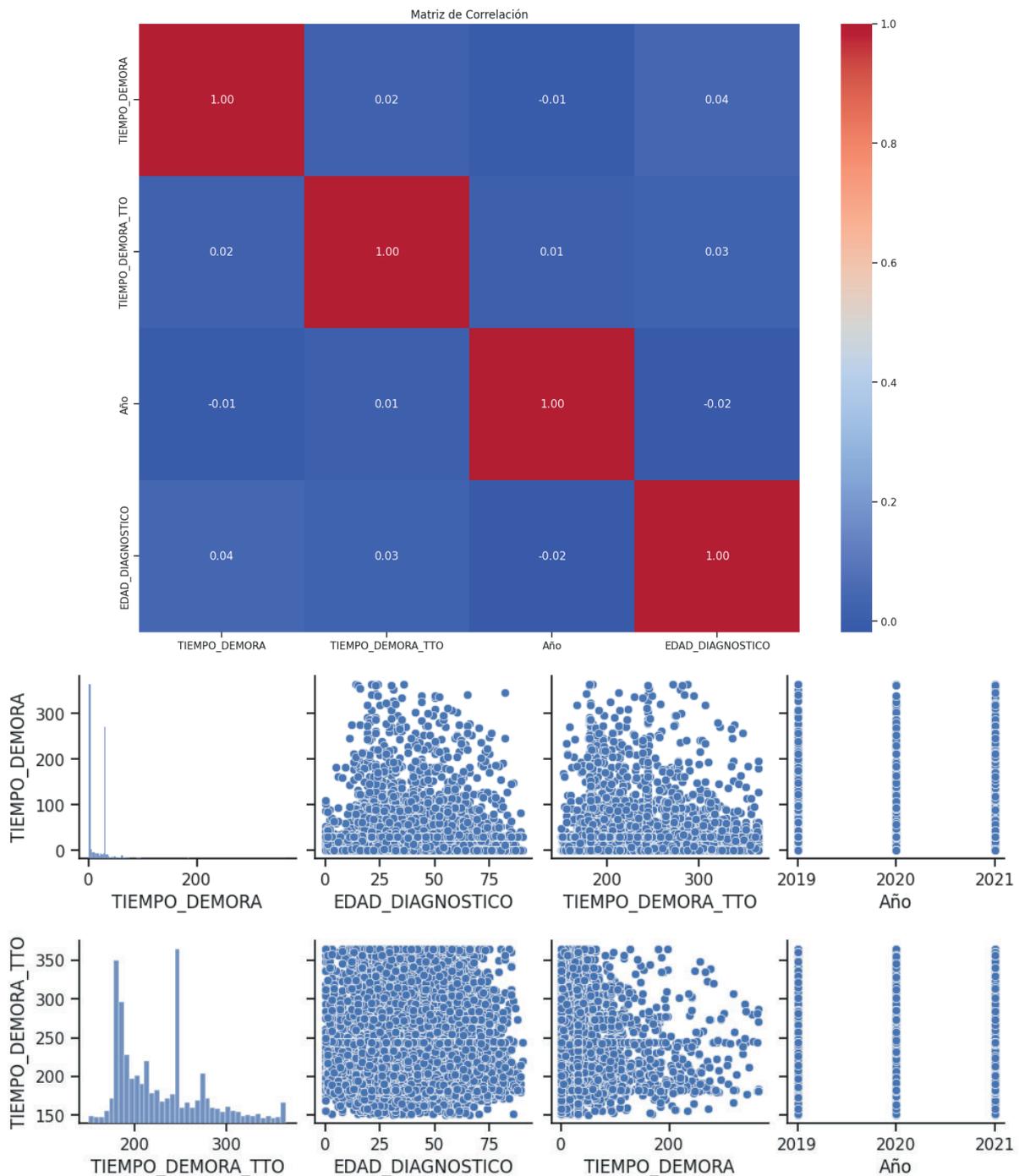
Fuente: elaboración propia.

3.4. Análisis de correlación

Antes de continuar con la regresión multivariada, debe presentarse cómo se relacionan entre sí las variables predictoras y la de respuesta. Si existen fuertes correlaciones entre las predictoras, podría hallarse multicolinealidad, y ello podría afectar la precisión e interpretación del modelo de regresión.

Por lo cual, en los tres gráficos siguientes (Gráfico 21a, b y c) se puede observar que, las cuatro variables analizadas, tiempo de demora en el diagnóstico (TDD), tratamiento, edad y año de notificación presentan un valor cercano a cero. Este valor indica una correlación muy débil entre las variables, es decir que no hay una relación lineal fuerte entre las mismas.

Gráfico 21a, b y c. Análisis de correlación entre variables cuantitativas de casos notificados antes y después del ASPO.

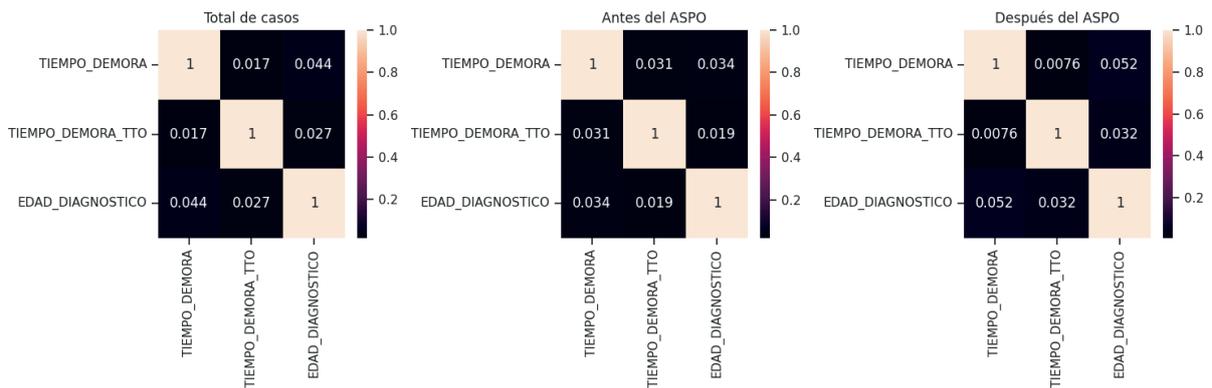


Fuente: elaboración propia.

Si se analiza por momento de diagnóstico (Gráfico 22) tampoco se aprecia correlación entre edad y tiempos de demora en ninguno de los períodos. Por ende, no se constata una relación lineal entre la edad de los pacientes y el TDD (tiempo de demora en el diagnóstico) y TDT (tiempo de demora en el tratamiento). En otras palabras, la edad de los pacientes no influiría en la cantidad de tiempo que transcurre antes de recibir atención médica.

Es importante destacar que, aunque no se observa una correlación evidente en el gráfico, esto no excluye la posibilidad de que otras variables o factores no representados en este análisis puedan estar influyendo en el TDD y TDT.

Gráfico 22: Análisis de covarianza entre las variables tiempo de demora en el diagnóstico, tratamiento y edad de diagnóstico de casos notificados antes y después del ASPO.



Fuente: elaboración propia.

3.5 Aprendizaje Supervisado

Por medio del aprendizaje automático se busca la predicción del tiempo de demora en el diagnóstico de individuos con diagnóstico de tuberculosis, para ello, se aplica el modelo de regresión lineal multivariado²⁵.

Tabla 4. Modelo de regresión multivariado para el total de casos notificados de tuberculosis del 2019 a 2021 en Argentina.

```

=====
Dep. Variable:      TIEMPO_DEMORA_TTO      R-squared:          0.009
Model:              OLS                    Adj. R-squared:     0.008
Method:             Least Squares          F-statistic:        12.52
Date:               Tue, 12 Dec 2023        Prob (F-statistic): 5.31e-20
Time:               16:29:19                Log-Likelihood:     -65402.
No. Observations:  12395                    AIC:                1.308e+05
Df Residuals:      12385                    BIC:                1.309e+05
Df Model:          9
Covariance Type:   nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
const	151.6785	2.329	65.129	0.000	147.114	156.244
CLASIFICACION_PANDEMIA_Antes	74.4909	1.232	60.461	0.000	72.076	76.906
CLASIFICACION_PANDEMIA_Despues	77.1876	1.253	61.625	0.000	74.732	79.643
TIEMPO_DEMORA	0.0170	0.011	1.538	0.124	-0.005	0.039
CLASIF_INICIO_TRAT_Recaida	7.4951	2.944	2.546	0.011	1.725	13.266
EDAD_DIAGNOSTICO	0.0612	0.023	2.605	0.009	0.015	0.107
SEXO_M	2.7222	3.141	0.867	0.386	-3.434	8.878
SEXO_F	-0.1833	3.157	-0.058	0.954	-6.371	6.004
CLASIF_INICIO_TRAT_Nuevo	-4.1122	1.629	-2.525	0.012	-7.305	-0.919
CLASIF_INICIO_TRAT_Pérdida del seguimiento recuperado	14.0136	3.359	4.172	0.000	7.430	20.597
CLASIFICACION_EXTRAPULMONAR_Diseminada	22.7289	5.157	4.408	0.000	12.621	32.837

```

=====
Omnibus:           1249.843    Durbin-Watson:      1.958
Prob(Omnibus):     0.000    Jarque-Bera (JB):   1662.644
Skew:              0.890    Prob(JB):           0.00
Kurtosis:          3.223    Cond. NO.           1.03e+16
=====

```

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 3.14e-25. This might indicate that there are strong multicollinearity problems or that the design matrix is singular.

Fuente: elaboración propia.

En la tabla anterior se aprecia que el coeficiente de determinación (R^2) es de 0.009. Este valor indica que aproximadamente el 0.9% de la variabilidad en el tiempo de demora en el tratamiento puede ser explicada por las variables predictoras incluidas en el modelo. No obstante, este valor es bajo, la prueba F arroja un p-valor significativamente pequeño ($p < 0.05$), por lo cual sugiere que al menos una de las variables predictoras es estadísticamente significativa en relación con la variable de respuesta.

Cuando se analizan los coeficientes de las variables predictoras, la que representa la clasificación de la pandemia, como se ha mencionado a lo largo del presente trabajo, posee dos niveles (antes y después). Ambos

²⁵ Véase [metodología](#).

evidencian diferencias significativas en lo que respecta al tiempo de demora en relación a un nivel de referencia no especificado. Además, el TDT posee un coeficiente positivo de 0.0170, ello implica que manteniendo todas las demás variables constantes, manifiesta un aumento esperado de 0.0170 unidades sobre el TDT por cada unidad de cambio en el tiempo de demora en el diagnóstico. Empero, este coeficiente no es estadísticamente significativo ($p=0.124$).

En otro orden, la variable de clasificación al inicio del tratamiento goza de varios niveles ('Recaída', 'Nuevo', 'Pérdida del seguimiento recuperado'), cada uno expone diferencias significativas respecto al tiempo de demora en relación al nivel de referencia no especificado. Asimismo, las variables de sexo (masculino y femenino) no son estadísticamente significativas, es decir que no exhibe un efecto significativo en el tiempo de demora. Por último, la variable de clasificación extrapulmonar revela un coeficiente positivo de 22.7289, esto sugiere que preservando las demás variables constantes, expone un aumento esperado de 22.7289 unidades en el TDT cuando la clasificación es 'Diseminada'.

Por todo lo anterior, se considera que si bien el modelo expone algunas relaciones estadísticamente significativas, el pequeño R^2 sugiere que las variables incluidas explican una pequeña proporción de la variabilidad en el tiempo de demora en el tratamiento.

Tabla 5. Modelo de regresión multivariado para el total de casos notificados de tuberculosis del 2019 a 2021 en Argentina.

OLS Regression Results						
Dep. Variable:	TIEMPO_DEMORA_TTO	R-squared:	0.001			
Model:	OLS	Adj. R-squared:	0.001			
Method:	Least Squares	F-statistic:	5.997			
Date:	Mon, 11 Dec 2023	Prob (F-statistic):	0.00249			
Time:	14:10:15	Log-Likelihood:	-65050.			
No. Observations:	12319	AIC:	1.301e+05			
Df Residuals:	12316	BIC:	1.301e+05			
Df Model:	2					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	225.9477	0.966	233.841	0.000	224.054	227.842
EDAD_DIAGNOSTICO	0.0679	0.024	2.883	0.004	0.022	0.114
TIEMPO_DEMORA	0.0199	0.011	1.791	0.073	-0.002	0.042
Omnibus:	1254.165	Durbin-Watson:	1.954			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1673.382			
Skew:	0.896	Prob(JB):	0.00			
Kurtosis:	3.216	Cond. No.	117.			

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Fuente: elaboración propia.

De manera análoga, el modelo de regresión lineal múltiple se aplicó para analizar la relación entre la variable dependiente, el tiempo de demora en el tratamiento, y dos variables independientes: la edad en el momento del diagnóstico y el tiempo de demora inicial. Este arrojó un bajo poder explicativo ($R^2=0.001$), que indica una pequeña variabilidad en el tiempo de demora en el tratamiento respecto a las variables independientes.

Por su parte, los coeficientes de regresión señalan que si se mantiene constante el tiempo de demora inicial, el aumento de una unidad en la edad de diagnóstico se asocia a un incremento de 0.0679 en el TDT. No obstante, un aumento de una unidad en el tiempo de demora inicial se asocia a un aumento de 0.0199 en el TDT. Es decir que ambos coeficientes fueron estadísticamente significativos, con p -valores de 0.004 y 0.073, respectivamente.

A pesar de que existe una relación estadísticamente significativa entre las variables independientes y la

demora en el tratamiento, esta es muy débil y explica solo una pequeña parte de la variabilidad observada en la demora del tratamiento. Por lo tanto, es posible que otros factores no incluidos en el modelo también influyan.

Cabe destacar que estos valores no mejoran al analizar las relaciones por grupo según momentos de notificación del caso. Pues, en ambos modelos, los coeficientes de regresión para las variables independientes edad de diagnóstico no son estadísticamente significativos, con *p-valores* de 0.189 y 0.189, respectivamente. Mientras que para la variable independiente tiempo de demora en el diagnóstico sin son estadísticamente significativos, con *p-valores* de 0.032, para los casos antes y después del ASPO. Sin embargo, al igual que para el total de los casos es importante destacar que, aunque los *p-valores* son estadísticamente significativos, la magnitud de los coeficientes es pequeña, lo que indica que las variables incluidas tienen un impacto limitado en el tiempo de demora en el tratamiento.

Tabla 6. Modelo de regresión multivariado para los casos notificados de tuberculosis antes del ASPO en Argentina.

OLS Regression Results						
Dep. Variable:	TIEMPO_DEMORA_TTO	R-squared:	0.001			
Model:	OLS	Adj. R-squared:	0.001			
Method:	Least Squares	F-statistic:	3.254			
Date:	Mon, 11 Dec 2023	Prob (F-statistic):	0.0387			
Time:	14:12:07	Log-Likelihood:	-26710.			
No. Observations:	5084	AIC:	5.343e+04			
Df Residuals:	5081	BIC:	5.345e+04			
Df Model:	2					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	224.9016	1.447	155.436	0.000	222.065	227.738
EDAD_DIAGNOSTICO	0.0457	0.035	1.313	0.189	-0.023	0.114
TIEMPO_DEMORA	0.0361	0.017	2.142	0.032	0.003	0.069
Omnibus:	486.786	Durbin-Watson:	2.006			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	638.523			
Skew:	0.866	Prob(JB):	2.22e-139			
Kurtosis:	3.130	Cond. No.	115.			

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Fuente: Elaboración propia.

Tabla 7. Modelo de regresión multivariado para los casos notificados de tuberculosis después del ASPO en Argentina.

OLS Regression Results						
Dep. Variable:	TIEMPO_DEMORA_TTO	R-squared:	0.001			
Model:	OLS	Adj. R-squared:	0.001			
Method:	Least Squares	F-statistic:	3.254			
Date:	Mon, 11 Dec 2023	Prob (F-statistic):	0.0387			
Time:	14:13:05	Log-Likelihood:	-26710.			
No. Observations:	5084	AIC:	5.343e+04			
Df Residuals:	5081	BIC:	5.345e+04			
Df Model:	2					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	224.9016	1.447	155.436	0.000	222.065	227.738
EDAD_DIAGNOSTICO	0.0457	0.035	1.313	0.189	-0.023	0.114
TIEMPO_DEMORA	0.0361	0.017	2.142	0.032	0.003	0.069
Omnibus:	486.786	Durbin-Watson:	2.006			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	638.523			
Skew:	0.866	Prob(JB):	2.22e-139			
Kurtosis:	3.130	Cond. No.	115.			

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Fuente: elaboración propia.

4. Conclusiones

Teniendo en cuenta la presencia de las Ciencias computacionales en los diversos ámbitos de la vida, en este trabajo se ha buscado relacionar el aprendizaje automático con una enfermedad infecciosa, como es la tuberculosis. Para ello, con los reportes obtenidos del Sistema Nacional de Vigilancia de la Salud de la República Argentina, se evaluaron las notificaciones de casos de TB en el contexto previo y durante la pandemia de COVID-19. El objetivo principal radicó en determinar las relaciones entre tiempos de diagnóstico y de tratamiento, edades de pacientes y ubicación de estos.

La notificación de casos por provincia evidencia un aumento después de la declaración del ASPO, siendo Buenos Aires, la Ciudad Autónoma de Buenos Aires, Santa Fe, Salta y Córdoba las que presentaron mayor incidencia. Asimismo, al considerar la distribución etaria y de sexo, se observó una notable similitud antes y después del ASPO, con un leve predominio de casos en el grupo de entre 15 y 40 años.

En lo que respecta a la temporalidad de las notificaciones se constató una disminución de estas a principios del año 2020, que coincidiría con el principio del aislamiento obligatorio en Argentina, seguida de un notable aumento en el primer semestre de 2021. Esta tendencia se mantiene en los casos tratados con éxito, aunque la disminución no es tan marcada. Al enfocarse en grupos de edad específicos, se destaca una variación en el grupo de 0 a 18 años, con un descenso, seguido de un aumento en el primer semestre de 2021.

El análisis de los tiempos de demora en el diagnóstico y tratamiento presenta distribuciones asimétricas hacia la derecha, sugiriendo una presencia de valores atípicos. Tras el ASPO, se observa un leve aumento en los tiempos de demora, tanto en diagnóstico como en tratamiento. Sin embargo, al desglosar por provincia, se revelan diferencias notables. Algunas provincias mantienen estabilidad, mientras que otras experimentan aumentos o reducciones notables en los tiempos de demora, evidenciando disparidades regionales.

Por otra parte, se pudo observar que la georreferenciación de los casos no mostró diferencias sustanciales en la distribución geográfica de la tuberculosis antes y después de la pandemia. Esto indica que la enfermedad mantuvo una distribución comparativamente uniforme en todo el país durante ambos períodos.

El análisis reveló que el valor de R-cuadrado para la relación entre las variables independientes y la demora en el tratamiento es extremadamente bajo, alrededor del 0.1%. Esto indica que solo una pequeña fracción de la variabilidad en la demora del tratamiento se puede atribuir a las variables independientes incluidas en el modelo.

En particular, no se observó una relación significativa entre la edad de los pacientes y la demora en el tratamiento, lo que sugiere que las variables analizadas tienen una relación muy débil con esta variable.

La prueba de *log-rank* respalda cambios significativos en los tiempos de demora asociados con la clasificación de la pandemia, sugiriendo la necesidad de adaptar estrategias de salud pública para abordar estas variaciones. En resumen, estos hallazgos subrayan la compleja interacción entre la pandemia de COVID-19 y el diagnóstico y tratamiento de la tuberculosis en Argentina, destacando la importancia de considerar contextos demográficos regionales y específicos en la eficiencia dispar en la atención de la salud de los casos de tuberculosis.

Referencias bibliográficas

Banco de recursos de comunicación del Ministerio de Salud de la Nación (2023). Boletín n° 6: Tuberculosis y lepra en la Argentina.

<https://bancos.salud.gob.ar/recurso/boletin-ndeg-6-tuberculosis-y-lepra-en-la-argentina>

Cuello-Rüttler, L., & Gudiño, M. E. (2017). Los sistemas de información geográfica (SIG) como herramienta para el ordenamiento territorial y la salud pública.

https://bdigital.uncu.edu.ar/objetos_digitales/10881/01e3cuello-gudio.pdf

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer.

Hobsbawm, E. J. (2018). *Historia del siglo XX: 1914-1991*. Crítica.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282), 457–481. <https://doi.org/10.1080/01621459.1958.10501452>

Martínez Sesmero, J. M. (2015). “Big Data”; aplicación y utilidad para el sistema sanitario. *Farmacia hospitalaria : organo oficial de expresion cientifica de la Sociedad Espanola de Farmacia Hospitalaria*, 39(2), 69-70. <https://doi.org/10.7399/fh.2015.39.2.8835>

Monterde i Bort, H., & Perea Lara, M. (1991). Capítulo 1: Introducción y principios básicos. *El enfoque del análisis exploratorio de datos: (Y su aplicación al campo de la psicología)* (pp. 9-39).

Organization, P. A. H., & Salud, O. P. de la. (2002). *Sistemas de información geográfica en salud: Conceptos básicos*. <https://iris.paho.org/handle/10665.2/40000>

Spiegel, M. R. y Stephens, L.J. (2009). *Estadística*. Serie Schaum. Mc Graw Hill.

Wickham, H. y Golemund, G. (2016). *R for Data Science*. <https://r4ds.hadley.nz/>

World Health Organization (2013). *Definiciones y marco de trabajo para la notificación de tuberculosis-Revisión 2013*. <https://www.who.int/es/publications/i/item/978924150534>

World Health Organization (2023). *Global Tuberculosis Report 2023*. <https://www.who.int/publications-detail-redirect/9789240083851>