Universidad Rey Juan Carlos

TESIS DOCTORAL

Epidemiological and Clinical Analysis of COVID-19: Impacts of Comorbidities, Vaccination, and Population-specific Outcomes

Autor:
Rafael García Carretero

Director:
Ángel Gil de Miguel

Programa de Doctorado en Epidemiología y Salud Pública

Escuela Internacional de Doctorado

2024

A mis padres, que me lo han dado todo

Agradecimientos

La escritura de una tesis doctoral no es un destino, es un viaje en el que se va construyendo, ladrillo a ladrillo, artículo tras artículo, una investigación. No es mi primera tesis. Era consciente del viaje cuando me matriculé en 2022. Fue en 2022 cuando, tras la insistencia de mi amiga Blanca San José, gran bibliotecaria y documentalista, una de las personas más inteligentes y sensatas que he conocido, conocí al Dr. Ángel Gil, catedrático de la URJC. Ambos me animaron a volver a emprender el viaje que supone escribir una tesis. Por eso quisiera comenzar expresando mi más sincero agradecimiento a Ángel, cuya experiencia, conocimiento y apoyo constante fueron fundamentales para la realización de este trabajo. Su guía no solo me proporcionó claridad académica, sino también motivación en momentos de duda. Su confianza en mí me impulsó a seguir adelante y superar los desafíos. Ángel me brindó la oportunidad de crecer académica y profesionalmente. Mi gratitud también va dirigida a la Dra. Ruth Gil, cuyo apoyo y disposición fueron esenciales para la culminación de esta tesis. Aprecié profundamente su confianza en mi trabajo y el ambiente de confianza que ambos me ofrecieron.

Mi familia tiene un lugar especial en esta dedicatoria. He de agradecer especialmente a mis padres su profundo e incondicional amor y su constante apoyo. Su fe en mí ha sido el motor que me permitió completar este camino.

Óscar Vázquez, Belén Rodríguez y Gabriel Martínez han sido, son y serán, mis compañeros y mis amigos. Con ellos se puede decir aquello de «todo mal, pero las risas, increíbles».

Magadalena Valverde ha sido mi mejor amiga de todos los tiempos, a la que quiero con locura, incluso a pesar de la distancia. Hablando de enfermedad mental, nos hemos entendido bien porque compartimos el mismo nivel de locura según DSM-VI (ella es psiquiatra y sabe de lo que habla).

Hay una mención especial (y quizá un rincocito en el infierno) a los revisores anónimos, cuyas detalladas y punzantes observaciones me hicieron replantear más de una vez la validez de mi investigación y de mi vida académica. Agradezco también a los errores de código y análisis fallidos que, aunque me llevaron al borde del abismo, también me enseñaron más de lo que esperaba.

También agradezco a todas las series de ciencia ficción que he consumido en este tiempo, en especial a Star Trek, por recordarme que incluso en el espacio, las misiones no siempre van según lo planeado, pero siempre hay una solución si tienes una tripulación de confianza. Por eso, agradezco a todos los que me han acompañado en mis viajes, incluso cuando el camino parecía perdido, os estaré eternamente agradecido. No podría haber llegado hasta aquí sin la ayuda de los tripulantes del 74656.

Finalmente, me gustaría agradecerme a mí mismo, por no abandonar el proyecto en los momentos más oscuros, por levantarme cada vez que me caí y, sobre todo, por haber aprendido que no hay clave del éxito, que triunfar, rendirse, ganar o perder no cambia nuestra valía como médicos, como investigadores o como personas. A veces, incluso rendirse es una opción. *Ka'pla.*

Bien plus que des mots, tu soutien a illuminé mon chemin.
Les moments de doute étaient plus légers grâce à toi.
Avec toi, chaque étape semblait plus facile.
Ne jamais oublier que tu es une lumière dans l'obscurité.
Chaque sourire partagé m'a donné la force de continuer.
Hors du commun, ton amitié a fait toute la différence.
Elle est précieuse, comme toi.

– Ambroise-Paul-Toussaint-Jules Valéry (1871-1945)

Min styrka har vuxit med ditt stöd.
Alla dagar blev ljusare med dig vid min sida.
Runt dig känns världen tryggare.
Ingen kan ersätta det du betyder för mig.
Att vara med dig har gjort resan lättare.

– Antiguo canto marinero nórdico, anónimo, siglo X

Abstract

The COVID-19 pandemic has posed unprecedented challenges worldwide, significantly impacting public health, healthcare systems, and socioeconomic structures. This thesis examines the multifaceted effects of COVID-19, focusing on the epidemiological, clinical, and healthcare responses in Spain. The research encompasses six interconnected studies that investigate various aspects of the pandemic, utilizing advanced analytical methods, including machine learning.

The first study explores the relationship between metabolic syndrome and COVID-19 severity, identifying key risk factors for hospitalization using machine learning models. It highlights the significant role of metabolic syndrome, obesity, and hypertension in increasing the likelihood of severe outcomes. The second and third studies provide a comprehensive epidemiological analysis of COVID-19 trends in Madrid and across Spain, respectively. These studies delineate the variations in admissions, critical care requirements, and mortality rates over different waves of the pandemic, underscoring the regional disparities and temporal patterns in disease burden.

The fourth study assesses the impact and effectiveness of COVID-19 vaccination programs in Spain. By comparing vaccination and non-vaccination scenarios through machine learning models, the research quantifies the reduction in hospitalizations and deaths attributable to vaccination, emphasizing its critical role in mitigating the pandemic's severity. The fifth study investigates the outcomes and evolution of COVID-19 in patients with hematological malignancies, identifying these patients as a high-risk group with significantly elevated mortality rates. The sixth study focuses on patients living with HIV, analyzing their clinical outcomes and identifying HIV infection as a risk factor for increased mortality due to COVID-19.

Collectively, these studies provide a detailed understanding of the epidemiological and clinical impacts of COVID-19 in Spain, highlighting the importance of targeted public health interventions, the efficacy of vaccination, and the need for specialized care for vulnerable populations. The findings underscore the utility of machine learning in epidemiological research, offering valuable insights for future pandemic preparedness and response strategies.

Resumen

La pandemia de COVID-19 ha planteado desafíos sin precedentes en todo el mundo. Ha afectado e manra significativa la salud pública, los sistemas de atención sanitaria y la estructura socioeconómica de la población. Esta tesis examina y describe el comportamiento de la pandemia de COVID-19. Nos hemos centrado en los aspectos epidemiológicos y clínicos en España y cómo ha afectado a la capacidad de asistencia sanitaria del país. La investigación abarca seis estudios interconectados que investigan varios aspectos de la pandemia, utilizando métodos analíticos avanzados, incluyendo el *machine learning*.

El primer estudio explora, a nivel local y como un estudio piloto, la relación entre el síndrome metabólico y la gravedad del COVID-19. Hemos tratado de identificar los factores de riesgo clave para la hospitalización utilizando modelos de *machine learning*. En este estudio evidenciamos el papel significativo del síndrome metabólico, la obesidad y la hipertensión en el aumento de la probabilidad de desenlace fatal. El segundo y tercer estudios proporcionan un análisis epidemiológico integral de las tendencias del COVID-19 en un hospital de segundo nivel, primero, y en toda España, después. Estos estudios delinean las variaciones en las hospitalizaciones, la necesidad de cuidados críticos y las tasas de mortalidad durante diferentes olas de la pandemia, subrayando las disparidades regionales y los patrones temporales en la carga asistencial.

El cuarto estudio evalúa el impacto y la efectividad de los programas de vacunación contra el COVID-19 en España. Al comparar escenarios de vacunación y no vacunación a través de modelos de *machine learning*, la investigación cuantifica la reducción de hospitalizaciones y muertes atribuibles a la vacunación, enfatizando su papel crítico en mitigar la gravedad de la pandemia. El quinto estudio investiga los resultados y la evolución del COVID-19 en pacientes con neoplasias hematológicas, identificando a estos pacientes como un grupo de alto riesgo con tasas de mortalidad significativamente elevadas. El sexto estudio se centra en los pacientes que viven con VIH, analizando sus resultados clínicos e identificando la infección por VIH como un factor de riesgo para el aumento de la mortalidad debido al COVID-19.

En conjunto, estos estudios proporcionan una comprensión detallada de los impactos epidemiológicos y clínicos del COVID-19 en España, destacando la importancia de las intervenciones de salud pública dirigidas, la eficacia de la vacunación y la necesidad de

atención especializada para poblaciones vulnerables. Los hallazgos subrayan la utilidad de *machine learning* en la investigación epidemiológica, ofreciendo valiosas ideas para la preparación y las estrategias de respuesta ante futuras pandemias.

Resumen Ejecutivo

Este trabajo aborda un análisis estadistico sobre el impacto de la pandemia de COVID-19 a nivel epidemiológico y clínico en los pacientes que precisaron ingreso hospitalario debido a la infección aguda. También aborda el impacto a nivel asistencial que el sistema sanitario español experimentó. Por ello, esta tesis doctoral se centra principalmente en las hospitalizaciones, mortalidad, datos demográficos, el efecto de las estrategias de vacunación y el uso de métodos avanzados de análisis para predecir resultados en pacientes con COVID-19. Esta tesis incluye seis estudios que recopilan datos de diferentes periodos y aspectos de la pandemia. De este modo, se explora cómo las olas sucesivas de COVID-19 afectaron al sistema sanitario español, partiendo de un caso particular, un hospital de la periferia de Madrid, hasta llegar a examinar el impacto sobre el territorio nacional y cómo evolucionó la pandemia, la presión sobre las unidades de cuidados intensivos (UCI), y las diferencias entre olas epidémicas de características demográficas y clínicas de los pacientes.

Antecedentes

La pandemia de COVID-19, causada por el virus SARS-CoV-2, ha representado una de las mayores crisis de salud pública a nivel mundial. En España, uno de los países europeos con mayor incidencia y mortalidad, la pandemia impuso una presión sin precedentes sobre el sistema sanitario y sobre sus trabajadores, particularmente durante las primeras olas de contagio. A lo largo de los meses cubiertos en los estudios incluidos en esta tesis, se observó una evolución significativa en la forma en que los hospitales manejaron la pandemia, desde la saturación inicial de recursos hasta la implementación de estrategias de vacunación masiva, que cambiaron la dinámica de las hospitalizaciones y la mortalidad.

El primer caso de COVID-19 en España fue confirmado el 31 de enero de 2020, y desde entonces, el sistema de salud español se enfrentó a múltiples olas de infecciones que variaron en intensidad, tanto por la respuesta del sistema de salud como por la aparición de nuevas variantes del virus. La respuesta hospitalaria incluyó la expansión de las UCI y la adopción de medidas de emergencia, como el uso de quirófanos y áreas de recuperación como unidades críticas. Sin embargo, la primera ola resultó en una elevada mortalidad, con casi un tercio de los pacientes fallecidos en el

hospital, lo que evidenció la vulnerabilidad inicial del sistema ante el impacto de la pandemia.

Objetivos principales de la investigación

Toda la investigación presentada tiene como objetivo general analizar el impacto de la pandemia de COVID-19 en el sistema sanitario español en términos de carga asistencial, es decir, cómo evolucionaron los ingresos hospitalarios y la mortalidad. Inicialmente se planteó como un estudio descriptivo de carga asistencial. Posteriormente se analizó el perfil demográfico de los pacientes. Con la evolución en el tiempo de la pandemia, se pudieron establecer patrones epidemiológicos y demográficos que alteraron la propia evolución. Estos análisis se basaron en el análisis de datos epidemiológicos, el uso de modelos predictivos avanzados y la evaluación de factores de riesgo en distintas poblaciones vulnerables.

1. **Evaluar el impacto epidemiológico del COVID-19 en los hospitales**: Uno de los objetivos centrales era cuantificar la carga asistencial que la pandemia supuso para los hospitales, especialmente durante las primeras olas, cuando el sistema sanitario experimentó una saturación crítica. Este objetivo incluye el análisis del número de ingresos hospitalarios, la ocupación de las UCI, duración de las estancias hospitalarias y las tasas de mortalidad.
2. **Identificar los factores de riesgo asociados a la severidad de la COVID-19**: Otro objetivo clave fue determinar los factores demográficos y clínicos que predisponen a los pacientes a desarrollar formas graves de COVID-19. Las publicaciones incluidas en esta tesis analizan cómo el síndrome metabólico, las enfermedades cardiovasculares, las neoplasias hematológicas y la coinfección por VIH afectan los resultados clínicos de los pacientes hospitalizados.
3. **Evaluar el impacto de la vacunación en la reducción de hospitalizaciones y muertes**: Un objetivo fundamental fue medir la efectividad de las vacunas en la reducción de hospitalizaciones y mortalidad por COVID-19. Utilizando técnicas de aprendizaje automático (*machine learning*), esta tesis busca estimar cuántas hospitalizaciones y muertes se pudieron prevenir gracias a la vacunación.
4. **Aplicar modelos predictivos para mejorar la gestión clínica y de recursos**: Como objetivo secundario, esta tesis buscaba desarrollar y aplicar modelos predictivos basados en *machine learning* para predecir la gravedad de la COVID-19 en pacientes infectados y la carga futura sobre

el sistema hospitalario. Estos modelos permiten a los profesionales sanitarios priorizar a los pacientes de mayor riesgo y optimizar la asignación de recursos sanitarios.

5. **Proporcionar recomendaciones para la planificación de respuestas ante futuras pandemias**: Finalmente, otro objetivo secundario, pero no menos importante, fue proporcionar recomendaciones basadas en la experiencia clínica adquirida durante la pandemia de COVID-19 con el fin de mejorar la preparación y respuesta del sistema de salud ante futuras emergencias sanitarias. Esto incluye el diseño de estrategias de vacunación más eficientes, el refuerzo de la capacidad hospitalaria y la creación de planes de contingencia para mitigar los efectos de futuras crisis de salud pública.

Resultados

Resultados globales.

Los estudios realizados primero en un hospital de la periferia de Madrid y más tarde aplicados a todo el territorio nacional recopilaron datos de medio millón de pacientes hasta diciembre de 2021, que se convirtió en 1,2 millones de pacientes para final de diciembre de 2022. Se cubrieron varias olas de la pandemia. Los resultados reflejan la evolución de la situación epidemiológica en función de las estrategias de salud pública, el comportamiento del virus y la implementación de campañas de vacunación.

Uno de los hallazgos más importantes fue la significativa disminución de la mortalidad a medida que se avanzaba en la pandemia. Durante la primera ola, la tasa de mortalidad fue del 16.6%, con una mayor prevalencia de fallecimientos en hombres (19.5% frente al 13.2% en mujeres). No obstante, para la quinta ola, la mortalidad se redujo al 4%, lo que coincide con la implementación de la vacunación masiva entre los grupos más vulnerables. A partir de la cuarta ola, se observó una reducción en la edad media de los hospitalizados, que cayó a 47 años, en contraste con los 70 años observados durante la primera ola. Este cambio en la edad de los hospitalizados se atribuye a la protección otorgada por la vacunación en las personas mayores.

Los análisis también revelaron que la duración de la estancia hospitalaria disminuyó con el tiempo, con una mediana de 7 días en total, que fue consistentemente más corta en las mujeres que en los hombres. Los ingresos en la UCI mostraron un patrón similar, con un pico durante las primeras olas y una tendencia decreciente en las olas posteriores. La mortalidad en la UCI fue alta durante la primera ola, pero mostró una mejora considerable a medida que avanzaba la pandemia y se mejoraban las estrategias de manejo clínico.

En cuanto a los factores demográficos, se observó una mayor afectación en hombres que en mujeres, no solo en cuanto a hospitalizaciones, sino también en términos de mortalidad. Los hombres representaron el 55% de los hospitalizados y tuvieron una mayor probabilidad de fallecer a causa del COVID-19. Además, los pacientes mayores de 70 años fueron los más vulnerables, particularmente durante las primeras olas. Sin embargo, a partir de la cuarta ola, el número de hospitalizaciones en personas mayores disminuyó significativamente, lo que evidencia la efectividad de las vacunas en este grupo de alto riesgo.

Pacientes con VIH

Los pacientes que viven con VIH han mostrado una mayor susceptibilidad a desarrollar formas graves de COVID-19 y experimentar peores resultados clínicos en comparación con la población general. En los estudios incluidos en esta investigación, se observó que los pacientes con VIH presentaron tasas más elevadas de hospitalización y mortalidad. La tasa de mortalidad entre estos pacientes fue un 25% mayor que la observada en individuos sin VIH, lo que sugiere que la inmunosupresión crónica asociada al VIH, junto con otras comorbilidades frecuentes en estos pacientes, los coloca en un grupo de alto riesgo frente al SARS-CoV-2.

Un hallazgo destacado es que estos pacientes presentan comorbilidades poco frecuentes en la población general, como enfermedades hepáticas y pulmonares, y que contribuyeron significativamente a los malos resultados clínicos. No obstante, se especula que probablemente la vacunación jugó un papel clave en reducir el impacto de la COVID-19 en estos pacientes, aunque se ha documentado que la respuesta inmunitaria a las vacunas puede ser menos efectiva en pacientes con VIH, especialmente aquellos con bajo nivel de CD4+ o sin supresión viral completa.

Pacientes con neoplasias hematológicas

Los pacientes con neoplasias hematológicas, como linfomas, leucemias y trastornos de células plasmáticas, se identificaron como uno de los grupos más vulnerables durante la pandemia de COVID-19. Los estudios incluidos en esta investigación muestran que estos pacientes presentaron tasas significativamente más altas de hospitalización y mortalidad en comparación con la población general. En concreto, la tasa de mortalidad global en pacientes con neoplasias hematológicas fue del 19.8%, lo que refleja el alto riesgo, probablemente en relación con la inmunosupresión inherente a su enfermedad y los tratamientos agresivos como la quimioterapia y la inmunoterapia.

Entre los subgrupos de neoplasias hematológicas, los pacientes con linfomas no Hodgkin y leucemias fueron los más afectados, con un riesgo de mortalidad hasta 1.7 veces mayor que otros pacientes hematológicos con COVID-19. Estos hallazgos subrayan la vulnerabilidad de estos pacientes ante infecciones graves, agravada por la disminución de la respuesta inmunitaria. A pesar de la reducción de la mortalidad en las olas posteriores de la pandemia, en parte debido a la vacunación, los pacientes con neoplasias hematológicas continuaron presentando un riesgo elevado debido a su respuesta inmunitaria limitada a las vacunas.

Los resultados también indican que la vacunación, aunque crucial, no fue tan efectiva en este grupo como en la población general, lo que resalta la necesidad de considerar estrategias adicionales, como dosis de refuerzo más frecuentes o tratamientos preventivos específicos. La investigación concluye que es fundamental seguir desarrollando intervenciones personalizadas para mejorar los resultados clínicos en pacientes con neoplasias hematológicas, especialmente durante futuras pandemias o brotes de enfermedades infecciosas.

Estimación de la efectividad de las vacunas usando machine learning

El estudio dedicado a la estimación de la efectivdad de las vacunas mostró que la vacunación contra el COVID-19 ha tenido un impacto significativo en la reducción de las hospitalizaciones y la mortalidad en España. A medida que la vacunación se implementó de manera masiva a partir de finales de 2020 y principios de 2021, se observó una disminución notable en las tasas de hospitalización y muerte en todas las olas subsiguientes. Las olas cuarta y quinta fueron particularmente reveladoras, ya que la edad media de los pacientes hospitalizados disminuyó drásticamente, lo que indica que las personas mayores, quienes inicialmente fueron las más afectadas, estaban mejor protegidas gracias a las vacunas.

Los modelos predictivos utilizados en el análisis estimaron que se evitaron entre 115,000 y 170,000 hospitalizaciones, así como entre 24,000 y 25,000 muertes, en probable relación con la campaña de vacunación. Estos hallazgos subrayan la importancia de las vacunas no sólo para reducir la gravedad de la enfermedad, sino también para aliviar la presión sobre el sistema de salud, permitiendo a los hospitales manejar la carga asistencial de manera más efectiva.

El uso de técnicas avanzadas de análisis estadístico desempeñó un papel crucial en la estimación precisa del impacto de la vacunación en la reducción de hospitalizaciones y muertes. Modelos como ElasticNet y RandomForest permitieron simular escenarios hipotéticos en los que no se hubieran implementado las vacunas, comparándolos con los datos reales observados, lo que permitió estimar el número de hospitalizaciones y fallecimientos evitados. Estas técnicas permitieron analizar grandes cantidades de datos clínicos y epidemiológicos de manera eficiente, identificando las variables clave que influyen en los resultados, como la edad, las comorbilidades y el estado de vacunación.

Machine learning no solo ayudó a mejorar la precisión de las predicciones, sino que también permitió una mayor comprensión de cómo diferentes factores de riesgo interactúan para influir en la gravedad de la COVID-19. Estas herramientas analíticas avanzadas son fundamentales para la planificación de recursos en salud pública y para la toma de decisiones informadas, ya que proporcionan una base sólida para evaluar el impacto de intervenciones como la vacunación y para predecir el curso de futuras pandemias.

Discusión

Los estudios incluidos en esta investigación destacan varias conclusiones importantes sobre la dinámica de la pandemia y la efectividad de las respuestas implementadas. La primera ola de COVID-19 fue la más devastadora para los hospitales, con altos niveles de saturación en las UCI y una tasa de mortalidad alarmante, especialmente en personas mayores y en aquellos con comorbilidades como la hipertensión, la diabetes y las enfermedades cardíacas. La falta inicial de preparación, junto con la alta transmisibilidad del virus y la limitada capacidad de pruebas diagnósticas, contribuyeron a la gravedad de esta ola.

Con la introducción de las vacunas a finales de 2020, se observó un cambio drástico en la evolución de la pandemia. Las vacunas redujeron significativamente las hospitalizaciones y las muertes, lo que permitió al sistema de salud aliviar la carga sobre sus recursos críticos. Los datos de este estudio confirman que la vacunación fue el factor determinante en la disminución de la mortalidad, incluso cuando surgieron nuevas variantes más transmisibles del virus, como Delta y Omicron.

Además, el uso de técnicas avanzadas de análisis, como el aprendizaje automático, permitió predecir con precisión los resultados en los pacientes hospitalizados. Estas herramientas proporcionaron estimaciones robustas del número de hospitalizaciones y muertes evitadas gracias a la vacunación. Nuestra investigación estimó que entre 115,000 y 170,000 hospitalizaciones y entre 24,000 y 25,000 muertes fueron prevenidas por las vacunas en España durante el periodo de estudio. Estos hallazgos subrayan la importancia del uso de modelos predictivos para la planificación de recursos en futuras pandemias y para mejorar la toma de decisiones en salud pública.

La variabilidad regional en el impacto de la pandemia también es un tema destacado en esta investigación. Las regiones más densamente pobladas y con mayores desafíos socioeconómicos experimentaron mayores tasas de hospitalización y mortalidad. Este hallazgo refuerza la necesidad de políticas de salud pública más equitativas y una mejor distribución de recursos para enfrentar crisis sanitarias, asegurando que los sistemas de salud puedan responder eficazmente en todas las regiones del país.

Conclusiones

Conclusiones globales

En conjunto, toda esta investigación examina y analiza la carga asistencial que sufrieron los hospitales españoles como respuesta inicial a la pandemia. Se demuestra el impacto clínico en los hospitales. También se demuestra que las medidas de salud pública implementadas posteriormente, especialmente la vacunación, fueron fundamentales para reducir la carga de la enfermedad. La introducción de las vacunas no solo disminuyó la mortalidad y las hospitalizaciones, sino que también cambió el perfil demográfico de los pacientes hospitalizados, con una menor afectación en las personas mayores y una mayor concentración de hospitalizaciones en personas más jóvenes a medida que avanzaba la pandemia.

El uso de modelos predictivos avanzados permitió una mejor comprensión del impacto potencial de la pandemia en diferentes escenarios, lo que subraya la importancia de utilizar técnicas de análisis de datos en tiempo real para mejorar las respuestas ante futuras crisis de salud pública. Estos modelos proporcionaron información clave para la planificación de recursos y la priorización de intervenciones, como las campañas de vacunación.

Machine learning

Machine learning ha desempeñado un papel clave en los recursos estadísticos de esta investigación por su capacidad para analizar, predecir y comprender los complejos patrones epidemiológicos del COVID-19. En esta tesis se resalta su importancia tanto para el diagnóstico temprano como para la predicción de la gravedad de la COVID-19, así como para estimar el impacto de las medidas preventivas como la vacunación. Una de las contribuciones más significativas de *machine learning* es su capacidad para manejar grandes volúmenes de datos de múltiples fuentes, como historiales clínicos electrónicos, bases de datos epidemiológicas y datos demográficos, lo que permite una evaluación precisa de los factores de riesgo que influyen en la progresión de la enfermedad y los resultados clínicos.

Los modelos predictivos basados en *machine learning*, particularmente ElasticNet y RandomForest, fueron utilizados en varios estudios incluidos en esta disertación para identificar los factores más relevantes que influyen en la gravedad del COVID-19, incluyendo edad, comorbilidades, sexo y otros parámetros clínicos. Estos modelos demostraron ser herramientas valiosas para estimar la probabilidad de hospitalización, ingreso en UCI y mortalidad, lo que permite a los profesionales de la salud priorizar el manejo de los pacientes de mayor riesgo. Además, el uso de técnicas avanzadas como el análisis de características y la selección de variables permitió afinar los modelos para centrarse en los factores clave, mejorando la interpretabilidad y aplicabilidad clínica de los resultados.

Una aplicación destacada de *machine learning* en esta investigación fue la estimación del número de hospitalizaciones y muertes evitadas gracias a la vacunación contra el COVID-19. Utilizando modelos predictivos, se estimó que entre 115,000 y 170,000 hospitalizaciones, así como entre 24,000 y 25,000 muertes, fueron prevenidas en España gracias a la campaña de vacunación. Estos resultados fueron posibles gracias a la capacidad del *machine learning* para comparar escenarios hipotéticos sin vacunación con los datos reales observados, lo que permitió cuantificar el impacto directo de las vacunas en la mitigación de la carga del COVID-19 en el sistema de salud.

Machine learning no solo ha permitido una mejor comprensión de la dinámica de la pandemia, sino que también ha proporcionado herramientas cruciales para la toma de decisiones en salud pública. La capacidad de estos modelos para simular diferentes escenarios epidemiológicos y predecir futuros brotes ha sido clave para planificar la asignación de recursos hospitalarios y diseñar estrategias de vacunación más efectivas. En conjunto, el uso de *machine learning* en el estudio del COVID-19 ha demostrado ser un avance revolucionario en la investigación de enfermedades infecciosas, proporcionando una mayor precisión en los diagnósticos y en la predicción de resultados clínicos.

Impacto en pacientes con síndrome metabólico y comorbilidades cardiovasculares

El síndrome metabólico, que incluye condiciones como la obesidad, la hipertensión, la resistencia insulínica y la dislipemia, se identificó como un factor de riesgo importante para la hospitalización y la mortalidad en pacientes con COVID-19. A lo largo de los estudios presentados en esta investigación, se observó que los pacientes con síndrome metabólico tenían un mayor riesgo de desarrollar formas graves de la enfermedad, lo que resultó en una alta tasa de hospitalización y un aumento significativo en la mortalidad. El estado inflamatorio crónico asociado con el síndrome metabólico, sumado a la disfunción endotelial y la activación del sistema inmune, probablemente exacerba la gravedad del COVID-19, llevando a complicaciones como el síndrome de dificultad respiratoria aguda (SDRA) y el fallo multiorgánico.

Entre las comorbilidades cardiovasculares, la hipertensión y las enfermedades cardíacas preexistentes se destacaron como los principales predictores de malos resultados en pacientes hospitalizados por COVID-19. Los estudios incluyeron análisis multivariantes asociaban su prevalencia a la necesidad de cuidados intensivos y el riesgo de mortalidad. Los pacientes con antecedentes de insuficiencia cardíaca y enfermedad coronaria experimentaron tasas más altas de mortalidad, lo que resalta la necesidad de un manejo clínico riguroso de las comorbilidades cardiovasculares durante la infección por COVID-19. Estos hallazgos subrayan la importancia de identificar y tratar agresivamente los factores de riesgo cardiovascular en pacientes infectados, dado el impacto potencial en la evolución clínica de la enfermedad.

En conclusión, los pacientes con síndrome metabólico y otras comorbilidades cardiovasculares constituyen un grupo de alto riesgo que requiere un enfoque de atención integral. La vacunación, junto con el control estricto de los factores de riesgo cardiovascular, podría ser crucial para reducir la carga de la enfermedad y mejorar los resultados en estas poblaciones vulnerables. Esta tesis recomienda continuar investigando estrategias terapéuticas específicas que puedan abordar tanto el control de la infección viral como la gestión de las comorbilidades subyacentes.

Impacto en pacientes con VIH

Los pacientes con VIH representaron una población particularmente vulnerable durante la pandemia de COVID-19. En el estudio incluido en esta investigación se observó que los pacientes con VIH tenían un mayor riesgo de desarrollar complicaciones graves al contraer SARS-CoV-2 en comparación con la población general. Las hospitalizaciones y la mortalidad fueron significativamente más altas en este grupo, con un aumento del 25% en el riesgo de mortalidad ajustado por comorbilidades. Esto se debe en parte a la inmunosupresión crónica que experimentan los pacientes con VIH, lo que limita su capacidad para desarrollar una respuesta inmunitaria adecuada frente a infecciones virales graves como COVID-19. La disertación destaca la importancia de priorizar a las personas con VIH en las campañas de vacunación y en los tratamientos preventivos, así como la necesidad de un monitoreo estrecho y un manejo clínico proactivo de las comorbilidades asociadas, como las enfermedades hepáticas y pulmonares, que agravan el pronóstico en este grupo de pacientes.

A pesar de estos riesgos elevados, los resultados también sugieren que el acceso oportuno a la vacunación y la implementación de terapias antivirales y corticoesteroides ayudaron a mitigar la gravedad de la enfermedad en algunos casos. Sin embargo, se concluye que es necesario continuar investigando el efecto a largo plazo de la vacunación y la inmunosupresión en pacientes con VIH, así como explorar la posibilidad de administrar dosis de refuerzo adicionales o vacunas específicas para optimizar la respuesta inmunitaria en esta población.

Impacto en pacientes con neoplasias hematológicas

Los pacientes con neoplasias hematológicas, como los linfomas, leucemias y trastornos de células plasmáticas, también enfrentaron un riesgo desproporcionadamente alto de complicaciones graves y mortalidad durante la pandemia. La disertación destaca que los pacientes con neoplasias hematológicas tuvieron una tasa de mortalidad significativamente más alta que aquellos sin estas afecciones, con una mortalidad general del 19.8%. La disfunción inmunitaria asociada a estas neoplasias, exacerbada por los tratamientos inmunosupresores como la quimioterapia y los agentes biológicos, contribuyó a este aumento del riesgo. Los datos de los estudios sugieren que subgrupos específicos, como los pacientes con linfomas no Hodgkin y leucemias, presentaron un riesgo de muerte hasta 1.7 veces mayor en comparación con otros pacientes hospitalizados por COVID-19.

La vacunación también tuvo un impacto positivo en esta población, aunque su respuesta inmunitaria a las vacunas fue más débil en comparación con la población general. A pesar de la reducción general de la mortalidad en las olas posteriores, estos pacientes continuaron presentando desafíos significativos debido a la reducción de la efectividad de la vacuna y el riesgo persistente de infecciones graves. La investigación concluye que es fundamental seguir investigando estrategias de manejo personalizado, incluyendo el uso de vacunas de refuerzo y la implementación de terapias dirigidas para mejorar los resultados en pacientes con neoplasias hematológicas durante brotes de enfermedades infecciosas.

Finalmente, los estudios incluidos en esta disertación refuerzan la importancia de estar preparados para futuras pandemias mediante la implementación de sistemas de salud flexibles, capaces de adaptarse rápidamente a aumentos en la demanda de cuidados críticos. La experiencia adquirida durante la pandemia de COVID-19 será crucial para enfrentar futuras emergencias de salud pública y para desarrollar estrategias que protejan a las poblaciones más vulnerables.

Contents

Introduction

Introduction | 1

1.1 Background on COVID-19 and its Global Impact

Brief history and emergence of COVID-19. Global spread and pandemic declaration

COVID-19 stands for Coronavirus Disease 2019. It is caused by the novel coronavirus SARS-CoV-2. The COVID-19 pandemic resulted in an unprecedented global disruption. Also, it challenged public health systems worldwide [1–3]. The novel coronavirus was first identified in December 2019 in Wuhan, China [4], and rapidly spread worldwide. Actually, the World Health Organization (WHO) declared it a pandemic in March 2020. After the first three years of the pandemic, over 645 million people have been infected with the virus globally, and more than 6.5 million deaths have been reported [5, 6]. In Spain alone, the pandemic has resulted in over 13 million confirmed cases and more than 118,000 deaths, severely challenging its healthcare system [7, 8].

Impact on healthcare systems and economies

The impact of COVID-19 extends beyond health, affecting economies, education, and everyday life. Lockdowns and social distancing measures, although essential for controlling the spread of the virus, have led to economic disadvantages and disruptions to education systems worldwide [9, 10]. The pandemic has highlighted the need for robust healthcare systems and the importance of epidemiological research in guiding public health responses.

COVID-19 is characterized by a wide range of clinical manifestations, from asymptomatic infection to severe respiratory distress, multi-organ failure, and death [11, 12]. The severity of the disease is influenced by several factors, including age, sex, and underlying health conditions [13]. Notably, individuals with pre-existing comorbidities such as diabetes, cardiovascular diseases, and chronic respiratory conditions are at a higher risk of developing severe illness and requiring hospitalization. Among these comorbidities, metabolic syndrome—a cluster of conditions including obesity, hypertension, insulin resistance, and dyslipidemia—has emerged as a significant predictor of adverse COVID-19 outcomes [14, 15].

Understanding the epidemiological trends and clinical characteristics of COVID-19 is crucial for developing effective public health strategies and clinical interventions. Epidemiological analysis helps in identifying high-risk populations, understanding the spread of the virus, and evaluating the effectiveness of measures taken to control the pandemic. Moreover, analyzing clinical data can provide insights into the pathophysiology of the disease, aiding in the development of targeted treatments and management strategies for affected patients.

The global scientific community has mobilized rapidly in response to the pandemic, leading to an unprecedented volume of research [16, 17]. This body of work encompasses studies on virus transmission, vaccine development, treatment protocols, and the socio-economic impacts of the pandemic. Despite these efforts, there remain significant gaps in our understanding of the disease, particularly regarding the interactions between COVID-19 and various comorbid conditions [18, 19]. This thesis aims to contribute to this growing field of knowledge by providing a detailed epidemiological and clinical analysis of COVID-19, with a particular focus on the impacts of comorbidities, vaccination, and population-specific outcomes.

1.2 Understanding Epidemiological Trends and Outcomes

Why epidemiological analysis is critical for managing pandemics.

Understanding trends and outcomes is key in Epidemiology and Public Health for several reasons, which are discussed as follows.

Early Detection and Surveillance.
Understanding Disease Dynamics.

Epidemiological studies help in the early detection of disease outbreaks, allowing for timely interventions. By analyzing trends over time, epidemiologists can identify emerging threats before they become widespread, enabling public health officials to implement preventive measures and mobilize resources effectively, allowing for timely interventions that can prevent widespread transmission. For example, during the COVID-19 pandemic, tracking infection rates helped identify hotspots and implement targeted lockdowns and quarantine measures to curb the spread of the virus [20].

Epidemiological studies help understand how a disease spreads, its incubation period, modes of transmission, and risk factors. This knowledge is essential for developing effective public health strategies, such as social distancing, mask mandates, and vaccination campaigns, to mitigate the impact of the pandemic [21].

Public Health Policy Making.

Epidemiological data is foundational for informed policymaking. It supports the development of strategies that mitigate the impact of diseases, such as social distancing measures, vaccination programs, and travel restrictions. Throughout the COVID-19 pandemic, epidemiological findings have been pivotal in shaping public health policies and responses globally [22].

Data is essential for evaluating the effectiveness of public health interventions. By examining the outcomes of health interventions, epidemiology helps determine the effectiveness of different strategies. For COVID-19, evaluating the outcomes of various interventions, like lockdowns and vaccination drives, has been essential for understanding their effectiveness in reducing transmission and mortality rates [23].

Evaluation of Interventions.

Data-driven models can forecast potential future outbreaks and healthcare needs, enabling better preparedness and response planning. Predictive modeling, based on epidemiological data, was widely used during the COVID-19 pandemic to anticipate hospital capacity needs and vaccine distribution logistics [23, 24].

Forecasting and Predictions.

Epidemiological data is crucial for public health comunication and education. Clear communication about the risks, transmission dynamics, and protective measures based on epidemiological evidence can promote better public compliance and engagement with health guidelines [21].

Public Health Education and Communication.

Epidemiology does not only serve local or national contexts but also informs global health decisions. The international spread of diseases like COVID-19 requires a coordinated global response, and understanding epidemiological trends across countries can aid in developing collective strategies for disease control and prevention [20].

Global Health.

1.3 Overview of Thesis Structure

The dissertation is structured to comprehensively discuss the impact of COVID-19 in Spain between 2020 and 2022, utilizing a multidisciplinary approach that integrates epidemiological data, clinical outcomes, the effectiveness of public health interventions, and the advantage of advanced statistial analyses, such as *machine learning*. The outline of the structure and content of this dissertation is as follows.

Outline of chapters and their objectives

Introduction in Chapter 1 provides an overview of the COVID-19 pandemic, including its origin, spread, and global impact. It also introduces the specific objectives of the thesis and establishes the research context. In this chapter I briefly discuss the state-of-the-art on health care topics related with COVID-19 regarding this dissertation (history, health public topics, and comorbidities).

Literature Review in Chapter 2 synthesizes relevant research on COVID-19, focusing on epidemiological studies, clinical outcomes associated with the virus, and the impact of public health interventions like vaccinations and social distancing measures.

I describe the methodology of the research in **Methods** in Chapter 3. This chapter details the research methods used in the included manuscripts, including data collection, analytical tools, and the rationale behind the use of *machine learning* algorithms. It also addresses ethical considerations.

Results shows the six studies that support this dissertation. I have summarized the main scientific articles in Table 1.1.

1. Study with short title **Metabolic syndrome and COVID-19** in Chapter **??** examines the association between metabolic syndrome and the severity of COVID-19.
2. **Trends and Outcomes in a Secondary Hospital** is the short title of the next publication in Chapter **??**. It analyzes trends in hospital admissions and outcomes over different waves of the pandemic in a secondary hospital in Madrid. It explores visualization techniques that will be eventually useful in the next publications.
3. **COVID-19 pandemic in Spain** is the main core paper of this dissertation. Chapter **??** shows a nationwide perspective on COVID-19 hospitalizations, ICU admissions, and mortality rates. It analyzes the impact of the pandemic in Spain in 2020 and 2021.
4. The next research was **Effectiveness of Vaccination in Spain**, in Chapter **??**, and it assesses the impact of vaccination on the trends of hospitalizations and mortality. I used *machine learning* algorithms to simulate a non-vaccination scenario, that is, what will be the behaviour or the pandemic in absence of vaccines.
5. Then I explore the impact of COVID-19 in some susceptible population. The first paper is, in short, **COVID-19 in Patients with HIV**. Chapter **??** focuses on the COVID-19 outcomes for patients living with HIV.
6. Finally, to investigate the outcomes for patients with hematological malignancies, I included the paper **COVID-19 and Hematological Malignancies** in Chapter **??**.

The **Discussion** in Chapter 10 synthesizes findings from the individual studies, discussing the clinical implications for daily practice. It compares the results with existing literature and explores the contribution of the research to understanding the COVID-19 pandemic.

The **Conclusions** in Chapter 11 summarize the key findings, highlighting the impact of COVID-19 on different population groups and the effectiveness of public health interventions. It also discusses the broader implications for future pandemic preparedness and response strategies.

Table 1.1: Core publications supporting the current dissertation.

Title	Journal	Indexation	Category	Status
Insulin Resistance and Metabolic Syndrome as Risk Factors for Hospitalization in Patients with COVID-19: Pilot Study on the Use of Machine Learning	Metabolic Syndrome and Related Disorders	JCR, Q4	Medicine, Research & Experimental	Published in 2023
Differences in Trends in Admissions and Outcomes among Patients from a Secondary Hospital in Madrid during the COVID-19 Pandemic: A Hospital-Based Epidemiological Analysis (2020–2022)	Viruses	JCR, Q1	Virology	Published in 2023
Hospitalization burden and epidemiology of the COVID-19 pandemic in Spain (2020–2021)	BMC Infectious Diseases	JCR, Q3	Infectious Diseases	Published in 2023
Impact and effectiveness of COVID-19 vaccines using machine learning to analyze time series: a population-based study	Journal of Clinical Medicine	JCR, Q1	Medicine, General & Internal	Published in 2024
Outcomes and Patterns of Evolution of Patients with Hematological Malignancies during the COVID-19 Pandemic: A Nationwide Study (2020–2022)	Journal of Clinical Medicine	JCR, Q1	Medicine, General & Internal	Published in 2024
Outcomes of Patients Living with HIV Hospitalized due to COVID-19: A 3-Year Nationwide Study (2020–2022)	AIDS and Behavior	JCR, Q1	Social Sciences, Biomedical; Public, Environmental & Occupation Health	Published in 2024

Appendices It includes the first page of each published article supporting the research and the findings of this thesis.

This structured approach ensures a thorough exploration of the various dimensions of the COVID-19 pandemic in Spain, providing valuable insights and practical recommendations for managing current and future public health challenges.

1.4 Objectives

1. **Analyze Epidemiological Trends of COVID-19 in Spain**. The main objective was to investigate the patterns of COVID-19 spread, hospital admissions, ICU usage, and mortality across different waves of the pandemic, including regional variations and the factors influencing these trends. Specifically, in Chapter **??** I analyzed the impact of the COVID-19 pandemic on a secondary hospital in Madrid by examining trends in patient admissions, critical care needs, and mortality rates across different pandemic waves. The study seeks to highlight the hospital's response and the variations in outcomes over time. In Chapter **??** I described the epidemiological characteristics and trends in hospitalizations, ICU admissions, and mortality rates due to COVID-19 across Spain during 2020 and 2021. This study aims to identify regional differences and the overall burden of COVID-19 on the Spanish healthcare system.
2. **Examine the Impact of Comorbidities on COVID-19 Outcomes**. Another objective was to evaluate how pre-existing conditions such as metabolic syndrome, hematological malignancies, and HIV affect the severity, hospitalization rates, and mortality of COVID-19 patients, identifying high-risk groups that require targeted interventions. Regarding this topic, the main specific objective was to investigate the relationship between metabolic syndrome (including conditions such as obesity, hypertension, insulin resistance, and dyslipidemia) and the severity of COVID-19, particularly in terms of hospitalization risk. Chapter **??** also aims to utilize *machine learning* models to identify key predictors of hospitalization among COVID-19 patients. Also, I explored the clinical characteristics, outcomes, and risk factors for mortality among hospitalized COVID-19 patients with hematological malignancies in Spain. The study aims to identify high-risk subgroups within this population and assess their vulnerability to severe outcomes from COVID-19. Finally, I investigated the demographic and clinical characteristics of people living with HIV (PLWH) who were hospitalized due to COVID-19, and to determine

whether HIV infection is a risk factor for increased mortality in this population. The study aims to compare the outcomes of PLWH with those of the general population hospitalized with COVID-19, using advanced statistical and machine learning methods.

3. **Assess the Effectiveness of COVID-19 Vaccination Programs**. The main objective was to determine the impact of vaccination campaigns in Spain on reducing COVID-19-related hospitalizations and deaths, using machine learning models to estimate the outcomes averted due to vaccination efforts. Specifically, I tried to assess the effectiveness of COVID-19 vaccination programs in reducing the severity and mortality associated with COVID-19 in Spain. The study employs machine learning models to estimate the number of hospitalizations and deaths prevented due to vaccination efforts and to analyze the relationship between vaccination rollout and declining COVID-19 cases and deaths.
4. **Provide Evidence-Based Recommendations for Future Pandemics**. To offer actionable insights and recommendations for healthcare providers, policymakers, and public health authorities based on the findings, focusing on improving pandemic preparedness, response strategies, and managing high-risk populations.

Literature Review | 2

2.1 Existing Research on COVID-19 Epidemiology: a Summary

Understanding the Spread and Dynamics of COVID-19.

The COVID-19 pandemic, caused by the novel coronavirus SARS-CoV-2, has spurred extensive epidemiological research to understand its transmission dynamics, spread, and impact on different populations. Since its emergence in December 2019, the virus has demonstrated a high transmission rate facilitated by respiratory droplets, aerosols, and surface contamination. Early studies identified that asymptomatic carriers and pre-symptomatic individuals significantly contribute to the virus's spread, complicating efforts to control the pandemic through traditional public health measures like contact tracing and quarantine [21].

Key Epidemiological Parameters

Research has focused on several key epidemiological parameters, including the basic reproduction number (R0), which indicates the average number of secondary infections produced by a single infected individual. Initial estimates of R0 for SARS-CoV-2 ranged from 2.2 to 3.9, suggesting a highly contagious virus [25]. The variability in R0 estimates across different studies reflects differences in population density, social behaviors, and the effectiveness of public health interventions.

Impact of Public Health Measures

Numerous studies have evaluated the effectiveness of public health measures such as lockdowns, social distancing, mask mandates, and travel restrictions in reducing transmission rates. For instance, evidence from several countries indicates that stringent lockdowns significantly reduced the spread of the virus, particularly when implemented early in the outbreak [22]. However, the relaxation of these measures often led to resurgences in cases, highlighting the need for balanced strategies that consider both public health and socioeconomic impacts.

Variants of Concern and Epidemiological Shifts

The emergence of SARS-CoV-2 variants of concern (VOCs) has added complexity to the epidemiological landscape. Variants such as Alpha, Beta, Delta, and Omicron have shown increased transmissibility, partial resistance to neutralization by antibodies, and, in some cases, altered pathogenicity [23]. The rapid spread of these variants has necessitated ongoing adjustments to public health strategies, including the updating of vaccines and booster recommendations to ensure continued effectiveness against evolving viral strains.

Demographic and Geographic Variability

Research has also highlighted significant demographic and geographic variability in COVID-19 epidemiology. Older adults, males, and individuals with underlying health conditions such as obesity, diabetes, and cardiovascular diseases have been identified as being at higher risk for severe outcomes [24]. Additionally, disparities in healthcare access and quality have led to disproportionate impacts on marginalized communities and low- and middle-income countries, underscoring the importance of equity-focused responses [20].

Future Research on Epidemiological Studies

As the pandemic continues to evolve, ongoing epidemiological research is crucial for monitoring the effectiveness of public health interventions and understanding the long-term impacts of COVID-19. Future studies should focus on the interplay between vaccination coverage, waning immunity, and the emergence of new variants to inform adaptive strategies for pandemic control and prevention.

2.2 Comorbidities and COVID-19 Outcomes

The presence of comorbidities has been identified as a critical factor influencing the severity and outcomes of COVID-19. Comorbidities are pre-existing medical conditions or chronic diseases that can exacerbate the effects of COVID-19, increasing the risk of severe illness, complications, and mortality. Understanding how these comorbidities affect COVID-19 outcomes is essential for developing targeted strategies to protect vulnerable populations and manage healthcare resources effectively.

Obesity and Metabolic Syndrome

Obesity has emerged as one of the most significant comorbidities affecting COVID-19 outcomes. Studies have shown that individuals with obesity are at higher risk for severe COVID-19 due to various factors, including impaired respiratory function, chronic inflammation, and a higher likelihood of other comorbidities such as hypertension and diabetes. Metabolic syndrome, which includes conditions such as obesity, insulin resistance, hypertension, and dyslipidemia, further compounds this risk. Research indicates that these factors can lead to an exaggerated immune response, increasing the likelihood of severe respiratory distress and multi-organ failure [26, 27].

Diabetes

Diabetes, particularly type 2 diabetes, has been consistently associated with worse COVID-19 outcomes. Patients with diabetes are more prone to infections due to immune system dysregulation. Hyperglycemia, a hallmark of diabetes, can impair the immune response, increase inflammation, and contribute to a pro-thrombotic state, all of which exacerbate the severity of COVID-19. Furthermore, diabetic patients often have other related comorbidities, such as cardiovascular disease and obesity, which also contribute to the increased risk of severe disease and mortality [28].

Cardiovascular Disease

Cardiovascular diseases (CVD), including hypertension, coronary artery disease, and heart failure, have been identified as major risk factors for severe COVID-19 outcomes. The virus can cause direct myocardial injury and exacerbate pre-existing heart conditions, leading to complications such as myocarditis, arrhythmias, and heart failure. Additionally, the inflammatory response triggered by COVID-19 can worsen atherosclerosis and increase the risk of acute coronary events. Patients with CVD are also at higher risk for thrombotic complications, which have been commonly observed in severe COVID-19 cases [29].

Chronic Respiratory Diseases

Chronic respiratory conditions such as chronic obstructive pulmonary disease (COPD) and asthma have been shown to increase the risk of severe outcomes in COVID-19 patients. These conditions are characterized by reduced lung function and chronic inflammation, which can impair the body's ability to clear infections and respond effectively to viral pathogens. COPD, in particular, is associated with a higher risk of hospitalization, intensive care unit (ICU) admission, and mortality due to COVID-19. However, some studies suggest that asthma, especially when well-managed, may not significantly increase the risk of severe COVID-19 [30, 31].

Immunocompromised Conditions

Individuals with immunocompromised conditions, including those with HIV/AIDS, cancer, or on immunosuppressive therapy (e.g., organ transplant recipients), are at increased risk for severe COVID-19 outcomes. The impaired immune response in these individuals can lead to prolonged viral shedding, increased risk of secondary infections, and a higher likelihood of severe disease progression. Additionally, patients undergoing cancer treatment may have additional risk factors such as neutropenia and organ dysfunction, further complicating COVID-19 management [32].

Chronic Kidney Disease (CKD)

Chronic kidney disease has been identified as an independent risk factor for severe COVID-19. Patients with CKD, especially those on dialysis, are more susceptible to severe complications due to their compromised immune systems, underlying chronic inflammation, and comorbidities such as hypertension and diabetes. Moreover, COVID-19 can directly affect renal function, leading to acute kidney injury, which significantly increases the risk of mortality in affected patients [33, 34].

Other Comorbidities

Several other conditions, such as liver disease, neurological disorders, and psychiatric conditions, have also been associated with increased risk of severe outcomes in COVID-19 patients. These comorbidities can complicate the clinical management of COVID-19, increase the risk of multi-organ failure, and necessitate more intensive care and monitoring.

2.3 Vaccination Impact on COVID-19

Vaccination has been one of the most critical tools in the global response to the COVID-19 pandemic. Since the development and rollout of COVID-19 vaccines, there has been a significant impact on the dynamics of the pandemic, including reductions in infection rates, severity of disease, hospitalizations, and mortality. This section provides an overview of the impacts of vaccination on COVID-19 based on global studies and data.

Reduction in Infection and Transmission Rates

COVID-19 vaccines have been highly effective in reducing the rates of infection and transmission of the virus. Studies from around the world have demonstrated that fully vaccinated individuals have a significantly lower risk of contracting COVID-19 compared to unvaccinated individuals. For instance, mRNA vaccines, such as those developed by Pfizer-BioNTech and Moderna, have shown efficacy rates of about 90-95% in preventing symptomatic COVID-19 in clinical trials and real-world studies. Additionally, vaccines have been shown to reduce the viral load in breakthrough infections, which may decrease the likelihood of transmission [35].

Prevention of Severe Disease and Hospitalization

One of the most important impacts of COVID-19 vaccination has been the prevention of severe disease and hospitalization. Vaccinated individuals who do contract COVID-19 are far less likely to experience severe outcomes, such as hospitalization, intensive care unit (ICU) admission, or death, compared to those who are unvaccinated. For example, data from various health agencies and studies have consistently shown that the majority of severe cases and hospitalizations occur in unvaccinated populations. The vaccines' ability to prevent severe outcomes is particularly crucial in protecting vulnerable populations, including the elderly and those with underlying health conditions [36].

Impact on Mortality Rates

Vaccination has also had a profound effect on reducing COVID-19-related mortality. Several studies have indicated a significant decline in death rates among vaccinated individuals, even in the presence of new variants of concern. This reduction in mortality is attributed to the vaccines' ability to elicit strong immune responses that prevent severe disease and complications. For example, in countries with high vaccination coverage, the case fatality rate of COVID-19 has dramatically decreased compared to earlier stages of the pandemic when vaccines were not available [37].

Influence on Pandemic Dynamics and Health Systems

The widespread rollout of vaccines has fundamentally altered the dynamics of the COVID-19 pandemic. In countries with high vaccination rates, there has been a noticeable shift from widespread transmission with high mortality and severe outcomes to a pattern where most cases are mild or asymptomatic, and healthcare systems are less overwhelmed. This has allowed health systems to shift focus from emergency response to managing the pandemic's long-term effects and dealing with other non-COVID-related health issues that were neglected during the pandemic's peak [38].

Challenges in Achieving Global Immunity

Despite the successes of vaccination campaigns, there have been significant challenges in achieving global immunity. Vaccine distribution has been highly unequal, with wealthier nations securing the majority of doses early on, leaving low- and middle-income countries with limited access. This disparity has hindered global efforts to control the virus and has led to prolonged waves of infection in under-vaccinated regions. Furthermore, vaccine hesitancy and misinformation have also posed significant barriers to achieving high vaccination rates in some communities, affecting overall herd immunity goals [39, 40].

Effectiveness Against Variants

The emergence of SARS-CoV-2 variants, such as Delta and Omicron, has tested the effectiveness of COVID-19 vaccines. Although vaccines remain highly effective at preventing severe disease and death, their efficacy against mild or asymptomatic infection has been reduced against some variants. This has led to recommendations for booster doses to enhance and prolong immunity, especially in the face of these more transmissible and partially vaccine-resistant variants. Booster programs have shown effectiveness in increasing antibody titers and providing better protection against these variants [41].

Considerations

Looking forward, the impact of COVID-19 vaccination will continue to evolve as new variants emerge and vaccine technologies advance. There is ongoing research into developing vaccines that provide broader protection against a range of coronaviruses and that are more easily distributed and administered. Understanding the duration of immunity provided by vaccines and the effectiveness of booster doses will be critical in shaping future public health strategies.

Overall, COVID-19 vaccination has significantly reduced the impact of the pandemic by lowering infection rates, preventing severe disease, and reducing mortality. However, challenges remain in achieving global vaccination coverage and managing the ongoing emergence of new variants. Continued efforts to improve vaccine distribution, address hesitancy, and adapt to new variants will be crucial in controlling the pandemic and preventing future outbreaks.

Methods

Methods 3

3.1 Data Collection

3.1.1 National Databases and Patient Records

This thesis relies on comprehensive data collected from various national databases and patient records across Spain, providing a robust foundation for analyzing the epidemiological and clinical impacts of COVID-19. The primary data sources include:

The Minimum Basic Data Set (MBDS) The National Hospital Data Information System provided us with the Minimum Basic Data Set at Hospitalization (MBDS-H). This administrative database includes 95-97% of discharge reports for patients hospitalized across Spain. It provides detailed information on patient demographics, diagnoses, comorbidities, treatments, and outcomes. Managed by the Spanish Ministry of Health, this system offers electronic data on hospital admissions related to COVID-19, covering a wide range of hospitals nationwide.

Regional Health Service Databases Data was also extracted from regional health services, particularly focusing on secondary hospitals in Madrid, to examine local trends and outcomes.

COVID-19 vaccination data Vaccination data were downloaded from the European Union/European Economic Area (EU/EEA) from the European Centre for Disease Prevention and Control. These data were collected through The European Surveillance System. All EU/EEA Member States are requested to report basic indicators on vaccination (vaccines categorized by manufacturers, number of doses administered, vaccinated population, among others). Data are categorized by target and age groups at a national level.

3.1.2 Criteria for Data Inclusion and Exclusion

The data inclusion and exclusion criteria were rigorously defined to ensure the reliability and relevance of the analyses:

Inclusion Criteria

- Patients with a confirmed diagnosis of COVID-19, as indicated by a positive PCR test or antigen test.
- Hospitalized patients during the specified study period (February 2020 to December 2022).

- Data records with complete information on key variables, including age, sex, comorbidities, and hospitalization outcomes.
- Specific subgroups, such as patients with hematological malignancies or those living with HIV, were identified based on relevant diagnostic codes.

Exclusion Criteria

- Records with missing or incomplete critical data (e.g., missing outcomes or demographic information).
- Patients without confirmed COVID-19 diagnosis.
- Duplicate records or entries related to transfers between healthcare facilities.

3.2 Statistical and Machine Learning Methods

A variety of statistical tools were employed to analyze the data, focusing on both descriptive and inferential statistics:

1. **Descriptive Statistics**. Used to summarize the basic features of the data, including patient demographics, hospitalization rates, ICU admissions, and mortality rates. Measures of central tendency (mean, median) and dispersion (standard deviation, interquartile range) were calculated for continuous variables, while frequencies and percentages were reported for categorical variables.
2. **Logistic Regression**. Applied to assess the association between comorbidities (e.g., obesity, diabetes, cardiovascular disease) and COVID-19 outcomes, such as the need for hospitalization and mortality. Odds ratios (ORs) were calculated to quantify the strength of these associations, with adjustments made for potential confounders.
3. **Time Series Analysis**. Utilized to examine trends in hospital admissions and outcomes over time, particularly across different waves of the pandemic. Machine Learning Techniques for Predictive Modeling To enhance the predictive accuracy and interpretability of the models, advanced machine learning techniques were integrated into the analysis:
4. **ElasticNet Regression**. This regularization method, combining the properties of Lasso and Ridge regression, was used to predict outcomes such as hospitalization and mortality. ElasticNet helps in feature selection by reducing the impact of less significant variables while retaining those with the most predictive power.

5. **Random Forests**. An ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random forests were particularly useful for handling large datasets with numerous variables and interactions, such as predicting outcomes based on a combination of demographic factors and comorbidities.
6. **LIME (Local Interpretable Model-gnostic Explanations)**. Employed to easily interpret the machine learning models by quantifying the contribution of each feature to the predictions. This method provided insights into which variables, such as age, BMI, or specific comorbidities, were most influential in predicting the outcome of COVID-19 in a given patient.

3.3 Ethical Considerations

3.3.1 Ethical Approval and Consent Procedures

Given the sensitive nature of health data, the study adhered to strict ethical standards:

Ethical Approval The research protocol was reviewed and approved by the Institutional Review Board (IRB) of the relevant institutions, including the Spanish Ministry of Health. The study complied with the Declaration of Helsinki and the regulations set forth by the European Union's General Data Protection Regulation (GDPR).

Informed Consent As the study involved retrospective analysis of anonymized patient data, individual informed consent was not required. However, the use of data was governed by agreements with healthcare providers and databases to ensure ethical compliance.

3.3.2 Data Anonymization and Privacy Measures

To protect patient confidentiality and comply with data protection laws:

Data Anonymization All patient data were anonymized before analysis. Identifiable information, such as names, addresses, and specific hospital identifiers, was removed or encrypted to prevent re-identification.

Secure Data Storage The data were stored in secure, encrypted databases with access limited to authorized personnel only. Regular audits were conducted to ensure compliance with data protection protocols.

Compliance with GDPR The study followed the guidelines of the General Data Protection Regulation (GDPR), ensuring that all data handling, storage, and processing were in line with European data protection standards.

Results

4 Metabolic syndrome and COVID-19

METABOLIC SYNDROME AND RELATED DISORDERS
Volume 21, Number 8, 2023

Pp. 443–452
DOI: 10.1089/met.2023.0083

Open camera or QR reader and scan code to access this article and other resources online.

Insulin Resistance and Metabolic Syndrome as Risk Factors for Hospitalization in Patients with COVID-19: Pilot Study on the Use of Machine Learning

Rafael Garcia-Carretero, MD, PhD,[1] Oscar Vazquez-Gomez, MD,[1] Marta Lopez-Lomba, MD,[2]
Ruth Gil-Prieto, PhD,[3] and Angel Gil-de-Miguel, MD, PhD[3]

Abstract

Aim: Conditions linked to metabolic syndrome, such as obesity, hypertension, insulin resistance, and dyslipidemia, are common in patients with severe coronavirus disease 2019 (COVID-19). These conditions can act synergistically to contribute to negative outcomes. We describe and analyze the relationship between metabolic syndrome and COVID-19 severity in terms of risk of hospitalization.
Methods: We designed a retrospective, cross-sectional study, including patients with confirmed COVID-19 diagnosis. Clinical and laboratory parameters regarding metabolic syndrome were collected. The Homeostatic Model Assessment of Insulin Resistance (HOMA-IR) was used to assess insulin resistance. The outcome was needed for hospitalization. Logistic regression was used to calculate odds ratios, and to determine the association between variables and risk of hospitalization. Advanced approaches using machine learning were also used to identify and interpret the effects of predictors on the proposed outcome.
Results: We included 2716 COVID-19 patients with a mean age of 61.8 years. Of these, 48.9% were women, 28.9% had diabetes, and 50.6% were diagnosed with metabolic syndrome. Overall, 212 patients required hospitalization. Patients with metabolic syndrome had a 58% greater chance of hospitalization if they were men, 32% if they had metabolic syndrome, and 23% if they were obese. Machine learning methods identified body mass index, metabolic syndrome, systolic blood pressure, and HOMA-IR as the most relevant features for our predictive model.
Conclusion: Metabolic syndrome and its related biomarkers increase the odds for a severe clinical course of COVID-19 and the need for hospitalization. Machine learning methods can aid understanding of the effects of single features when assessing risks for a given outcome.

Keywords: insulin resistance, metabolic syndrome, COVID-19, hospitalization, machine learning

Introduction

As of the end of 2022, >645 million people have been infected with coronavirus disease 2019 (COVID-19).[1] In Spain, >13 million confirmed cases have been identified, leading to >118,000 deaths.[2] The novel coronavirus severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) responsible for the disease was first identified in Wuhan (Hubei, China). By January 2020, the World Health Organization had declared the outbreak a public health emergency and then a pandemic in March 2020.[3]

Among its more severe consequences is respiratory illness that causes respiratory failure and distress.[4–6] The typical early clinical course of patients who test positive for

[1]Department of Internal Medicine, Mostoles University Hospital, Rey Juan Carlos University (Madrid), Mostoles, Spain.
[2]Department of Microbiology, Mostoles University Hospital, Mostoles, Spain.
[3]Department of Preventive Medicine and Public Health, Rey Juan Carlos University (Madrid), Mostoles, Spain.

SARS-CoV-2 includes mainly mild symptoms (<5–7 days after infection). However, in some cases, a subset of patients can develop a more severe disease (8–15 days after infection) that can cause severe acute respiratory syndrome (SARS), leading to respiratory failure and need for hospitalization.[7]

At the beginning of the pandemic, some studies found that risk factors related to metabolic syndrome were also associated with high risk of severe COVID-19.[8–11] Ghoneim et al. reported an increased cumulative incidence of COVID-19 in patients with pre-existing hypertension, obesity, dyslipidemia, and diabetes.[12] In addition, hypertension, insulin resistance,[13] dyslipidemia,[14] and obesity[15] are more prevalent in patients with worse outcomes than in patients without these conditions. Metabolic syndrome is a cluster of comorbidities related to insulin resistance. To identify the condition of metabolic syndrome, some criteria are used: obesity, hyperglycemia, dyslipidemia, and hypertension. Metabolic syndrome identifies individuals at high risk of cardiovascular disease (CVD). It also combines and links the risk factors for CVD.[16,17]

From a pathogenic point of view, metabolic syndrome predisposes to low-grade proinflammatory status, endothelial dysfunction, and dysregulation of the immune response.[18,19] All of these conditions can compromise the immune system and contribute to the activation of the cytokine storm associated with SARS.[20,21] Metabolic syndrome can therefore potentiate the effects of immune dysregulation and predispose patients to a worse prognosis than in patients without metabolic syndrome.

Establishing the effects of both metabolic syndrome and its related biomarkers in the risk for worse outcomes in terms of hospitalization, hospital stay, or death may be key to understanding the clinical course of COVID-19. We hypothesize that conditions linked to metabolic syndrome are associated with a more severe course of COVID-19, and those conditions can act synergistically to predispose patients to severe COVID-19.

We designed this study as a pilot to analyze data about patients with COVID-19. Rather than testing a hypothesis on the predictive value of certain risk factors, we assessed the feasibility of our approach using machine learning, for use in a larger scale study. Our ultimate objective was to analyze the relationship between metabolic syndrome, insulin resistance, and other related biomarkers, and the risk of hospitalization in patients diagnosed with COVID-19, using machine learning.

Methods

Study design and patient selection

We performed a retrospective, cross-sectional, cohort study of individuals who tested positive for COVID-19 in reverse-transcription polymerase chain reaction (RT-PCR) between March 22, 2020, and April 5, 2020 (2 weeks). The participants were treated at hospital (a tertiary, university hospital) during the observation period, and all of them were diagnosed with COVID-19. They presented to our institution with symptoms compatible with SARS-CoV-2 infection.

Using historical, electronic records, we recovered results of all RT-PCRs performed in that period by our institution for this pilot study. Inclusion criteria included adult patients with COVID-19 confirmed through RT-PCR tests during the observation period. We removed participants with suspicion of COVID-19 but no RT-PCR test and those with missing data related to metabolic syndrome or insulin resistance. Supplementary Fig. S1 shows the workflow we applied to our data. A final data set with 2716 observations was produced.

Data collection

Clinical and laboratory data on cardiovascular risk were recovered from discharge reports and retrospectively from the electronic health records of our institution. Data included demographic, clinical, and laboratory parameters. At baseline, we collected demographic data including age and sex. We also collected clinical parameters such as body mass index (BMI) and blood pressure. Hypertension was defined as systolic blood pressure (SBP) >139 mmHg or diastolic blood pressure >89 mmHg. Individuals on antihypertensive drugs were also included as hypertensive patients.

Laboratory data were collected within a month before the participants presented with COVID-19 symptoms. COVID-19–related laboratory parameters such as C-reactive protein, interleukin-6, leukocytes, and lymphocytes at the time of COVID-19 diagnosis were not recovered because they were outside the scope of our study. As noted, laboratory parameters and biomarkers related to CVD were uric acid, cholesterol, triglycerides, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, and ferritin. Creatinine and estimated creatinine-based glomerular filtration rate, as assessed using the Chronic Kidney Disease Epidemiology Collaboration equations, were used to assess kidney function.[22]

For glycemic status, glycosylated hemoglobin (HbA1c), fasting plasma glucose, and insulin were collected, so that diabetes mellitus, metabolic syndrome, and insulin resistance could be diagnosed. Insulin resistance was assessed through the Homeostatic Model Assessment of Insulin Resistance (HOMA-IR), using the following equation: [fasting plasma insulin (mU/L) $\times$ fasting plasma glucose (mg/dL)]/22.5. All laboratory variables were obtained using a Cobas 8000 e602 analyzer (Roche Diagnostics, Mannheim, Germany). RT-PCR tests were performed using an m2000 Real Time System (Abbott Molecular, Inc., Des Plaines, IL).

Diagnosis of Type 1 and Type 2 diabetes mellitus was based on clinical reports or based on the American Diabetes Association (ADA) criteria[23]; that is, fasting plasma glucose levels >126 mg/dL (7.0 mM) or HbA1c >6.5% (>48 mmol/mol). Metabolic syndrome was defined according to the Adult Treatment Panel III (ATP-III) criteria.[24] Individuals on antidiabetic drugs were also considered as having diabetes mellitus.

The primary outcome, that is, the response variable of our study, was hospitalization related to severe COVID-19. This outcome was collected from the electronic records of our hospital, and included patients presenting with respiratory failure, respiratory distress, pneumonia, or the need for oxygen support. This primary outcome was compared between the case group (hospitalized patients) and the control group (nonhospitalized individuals). Because this was a pilot study, we did not define subgroups according to the number of criteria present for the diagnosis of metabolic syndrome. Therefore, the additive of the synergistic effect of these criteria was not explored.

Descriptive and univariate correlation analyses

All statistical analyses, both standard and based on machine learning, were performed with R version 4.2.2. We first performed exploratory data analyses to plot each variable to check for normality, although this condition was mathematically checked using the Shapiro–Wilk test. Therefore, variables with normal-shaped distributions were reported with means and standard deviations, whereas nonnormally distributed variables were reported as medians and interquartile ranges. Categorical variables such as sex, diabetes, metabolic syndrome, and hospitalization were reported as percentages. For continuous and categorical variables, correlation analyses were performed using either the Mann–Whitney–Wilcoxon or the chi-square test, respectively. Statistical significance was set at $P < 0.05$.

Multivariate analyses

We combined a classical approach (binary logistic regression) and machine learning [recursive feature elimination (RFE), and penalized logistic regression] to calculate the beta coefficients of the variables and their odds ratios (ORs). Binary logistic regression is the most frequently used statistical approach in biomedical sciences with binary outcomes, that is, yes/no.

Logistic regression is simple and straightforward, and it allows easy interpretation of the effects of explanatory variables on a response variable. It allows researchers to easily calculate ORs to evaluate the final model. However, the model sometimes has several features selected as explanatory variables, making it too complex. The rationale for using machine learning at this stage was that it allowed us to select a set of features, that is, create a more parsimonious model—without losing accuracy or reliability. As noted, we used two primary approaches: RFE and penalized logistic regression.

RFE is a machine learning algorithm used for feature selection. It selects a subset of the most relevant variables (features) for a given data set. As noted, the fewer features, the less complex the data and the more interpretable is the final model. This method fits the model and ranks variables by importance, dropping the less-relevant variables and refitting the model. RFE repeats this process until a specified number of features remain. However, the most suitable number of variables cannot be known, so the algorithm tests different values and reports the mean accuracy for different numbers of input variables.

We plotted the results with the distribution of accuracy scores for different sets of variables. To score the different sets of variables and select the best-scoring collection, we used RFE crossvalidated (RFE-CV) for variable selection, although other algorithms were possible.[25–27] To use the RFE function of the caret package in R, we first selected the control options as follows. We used random forest (RF) as the algorithm to control RFE because it has an interesting built-in mechanism for computing feature importance.

In addition, we will use repeated 10-fold crossvalidation with five repeats to improve the performance of feature selection with RFE. We randomly split the data set into a training set (80% of observations) and a testing set (20% of observations). We included all features so that the algorithm could try all possible solutions (*i.e.,* only one variable, two variables, or all variables) to find the optimal features.

RFE-CV reports the number of optimal variables so that a more parsimonious model can be built with penalized multivariate logistic regression using an L1-penalized regularization. This approach is also known as least absolute shrinkage and selection operator (LASSO), and it is also based on machine learning.[28,29] It discards variables that do not contribute to the final model. It forces beta coefficients to very small values or to exactly zero. All beta coefficients thus shrink, but those with smaller effects are dropped as they reach zero.

Again, we used crossvalidation for the internal validation of the LASSO algorithm. We plotted the average evaluation score to select the set of variables to maximize the model's predictive accuracy. Penalization was determined according to the lambda value, which was used to select the subset of variables. LASSO is recommended by the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) checklist for developing and validating risk and diagnostic models.[30]

Interpretability of the results

Once a subset of features had been selected, the beta coefficients obtained, and the ORs calculated, we used a combination of approaches to better interpret the results. First, we used RF, another machine learning algorithm, which generates many independent decision trees that are then combined to get a single output.[31] However, our aim was to improve the interpretability for decision making concerning a given patient.

ORs and the outputs obtained from logistic regression, RFE, LASSO, and RF allow the model to be interpreted at a global scale. However, it is also important to focus the implementation of the results on single patients to derive local explanations, which are key for decision making. While RF was used to obtain a global interpretation, the local interpretable model-agnostic explanations (LIME) technique was used to obtain the local interpretation.[32]

Rationale for a pilot study

Many aspects of feasibility were raised to establish the rationale for designing this research study as a pilot study. Our aim was to have a cohort of patients with enough cases and controls to obtain reliable results within a given observation period. We chose 2 weeks in March and April 2020 to assess both the number of nonhospitalized individuals and the proportion of eligible hospitalized patients who could be enrolled. Further, we needed access to electronic data and discharge reports to properly collect variables and outcomes. Finally, we used advanced tools when fitting the model to evaluate the impact of the variables on the primary outcome. Thus, we assessed whether our design could provide a realistic examination of recruitment, data collection tools, and machine learning-based analysis methods.[33]

Results

Descriptive analyses and association with hospitalization

Table 1 shows the descriptive data and first correlation analyses. In all, 2716 patients were included, and almost

TABLE 1. BASELINE FEATURES OF OUR COHORT AND UNIVARIATE ANALYSES

	Total (n = *2716*)	*Nonhospitalized patients* (n = *2504*)	*Hospitalized patients* (n = *212*)	P
Sex (women, %)	48.9	48.2	56.6	0.023
Age (years)	61.8 ± 14.9	61.0 ± 14.8	71.0 ± 12.5	<0.001
BMI	24.6 ± 5.6	24.1 ± 5.3	30.6 ± 5.7	<0.001
Diabetes (%)	28.9	28.7	31.1	0.505
Metabolic syndrome (%)	50.6	49.7	60.4	0.004
SBP (mmHg)	139.0 (24.0)	139.0 (23.0)	143.5 (18.0)	<0.001
DBP (mmHg)	84.2 ± 11.3	84.2 ± 11.4	84.2 ± 9.4	0.582
Uric acid (mg/dL)	6.6 ± 2.2	6.6 ± 2.2	6.7 ± 1.8	0.242
Cholesterol (mg/dL)	212.2 ± 40.9	212.0 ± 41.6	214.8 ± 31.5	0.042
Triglycerides (mg/dL)	146 (105)	144.0 (103)	156 (105)	0.003
LDL-C (mg/dL)	130.1 ± 35.8	129.9 ± 36.5	131.4 ± 27.6	0.125
HDL-C (mg/dL)	62.3 ± 17.7	62.2 ± 17.8	63.1 ± 17.4	0.234
Fasting plasma glucose (mg/dL)	103 (30)	102 (30)	115 (32)	<0.001
HbA1c (%)	5.9 (0.9)	5.9 (1.0)	6.0 (0.7)	0.164
Insulin (mU/L)	14.9 (12.0)	14.7 (11.8)	16.6 (13.2)	0.003
HOMA-IR	3.7 (3.6)	3.6 (3.5)	4.2 (4.4)	<0.001
Ferritin (mg/dL)	158 (171)	158 (171)	146 (182)	0.636
Creatinine (mg/dL)	0.8 (0.2)	0.8 (0.2)	0.8 (0.2)	0.975
CKD-EPI-eGFR-creat (mL/min/1.72 m^2)	94.5 (23.3)	95.1 (23.9)	91.6 (21.3)	<0.001

BMI, body mass index; CKD-EPI-eGFR-creat, Chronic Kidney Disease Epidemiology Collaboration equation to estimate the creatinine-based glomerular filtration rate; DBP, diastolic blood pressure; HbA1c, glycosylated hemoglobin; HDL-C, high-density lipoprotein cholesterol; HOMA-IR, Homeostatic Model Assessment of Insulin Resistance; LDL-C, low-density lipoprotein cholesterol; SBP, systolic blood pressure.

half of them were women (48.9%), with a mean age of 61.8 years. More than a quarter (28.9%) had diabetes mellitus, and half were diagnosed with metabolic syndrome. In total, 2504 individuals were discharged from the hospital with mild or moderate symptoms that did not require hospitalization. However, 212 patients were hospitalized with symptoms of pneumonia and respiratory failure.

Metabolic syndrome and several of its related biomarkers, such as SBP, obesity, triglycerides, or fasting plasma glucose, were associated with the risk of hospitalization. Regarding glycemic status, HbA1c, insulin, and HOMA-IR were positively associated with the risk of hospitalization. Some of the variables identified as being associated with hospitalization are shown in Fig. 1.

Logistic regression and penalized logistic regression for hospitalization

The results of the multivariate analysis using binary logistic regression are shown in Table 2. This list of variables is smaller than that for the bivariate analyses shown in Table 1, because nonrelevant variables were dropped when the model was fitted. However, the model is too complex, as 12 variables were selected as relevant. Being male (OR = 1.58), having metabolic syndrome (OR = 1.32), and having obesity (OR = 1.23) were identified as the most relevant variables associated with hospitalization. This means that patients who have metabolic syndrome have a 58% greater risk of hospitalization if they are men, 32% if they have metabolic syndrome, and 23% if they are obese.

We also performed penalized logistic regression using machine learning. The results of RFE are shown in Fig. 2 along with the selection of features by LASSO. RFE proposes that 13 variables are the optimal solution, but accuracy was good in a range between 5 and 14 variables. LASSO proposes a more constrained range between three and six variables. Because the selection of three variables could decrease performance, according to RFE, we chose the value of lambda that allowed us to select six variables. Then, beta coefficients were mathematically computed by LASSO (Table 2). The remaining variables were considered nonrelevant, and they were dropped from the model. Sex (male), age, BMI, metabolic syndrome, SBP, and HOMA-IR were selected as the most important features associated with the risk of hospitalization.

LASSO only provides beta coefficients so that ORs can be calculated, but it does not report confidence intervals. Table 3 shows the performance parameters. Overall, in terms of classification accuracy, sensitivity, specificity, and concordance (C-Index), the differences among the three models were not significant. That is, performance accuracy remains, but with a more parsimonious model (6 features instead of 12).

Local interpretation of results from the machine learning models

Finally, to improve the interpretation of the previous results, we used RF to rank variables in order of importance. Variable importance for the selected features can be visually examined to enable observation of which ones were the most important for predicting the response variable. Therefore, using all of the variables, Fig. 3 shows the results provided by the RF algorithm, but we only display the top 10 features. The remaining variables had no impact on the model. Age, sex, metabolic syndrome, HOMA-IR, BMI, and SBP were the most relevant variables, as indicated by penalized logistic regression.

As noted, we used LIME to divide the behavior of the model into two sets of observations to justify the interpretations of our results. Figure 4 shows two plots: one for a hospitalized patient (panel A) and the other for a nonhospitalized

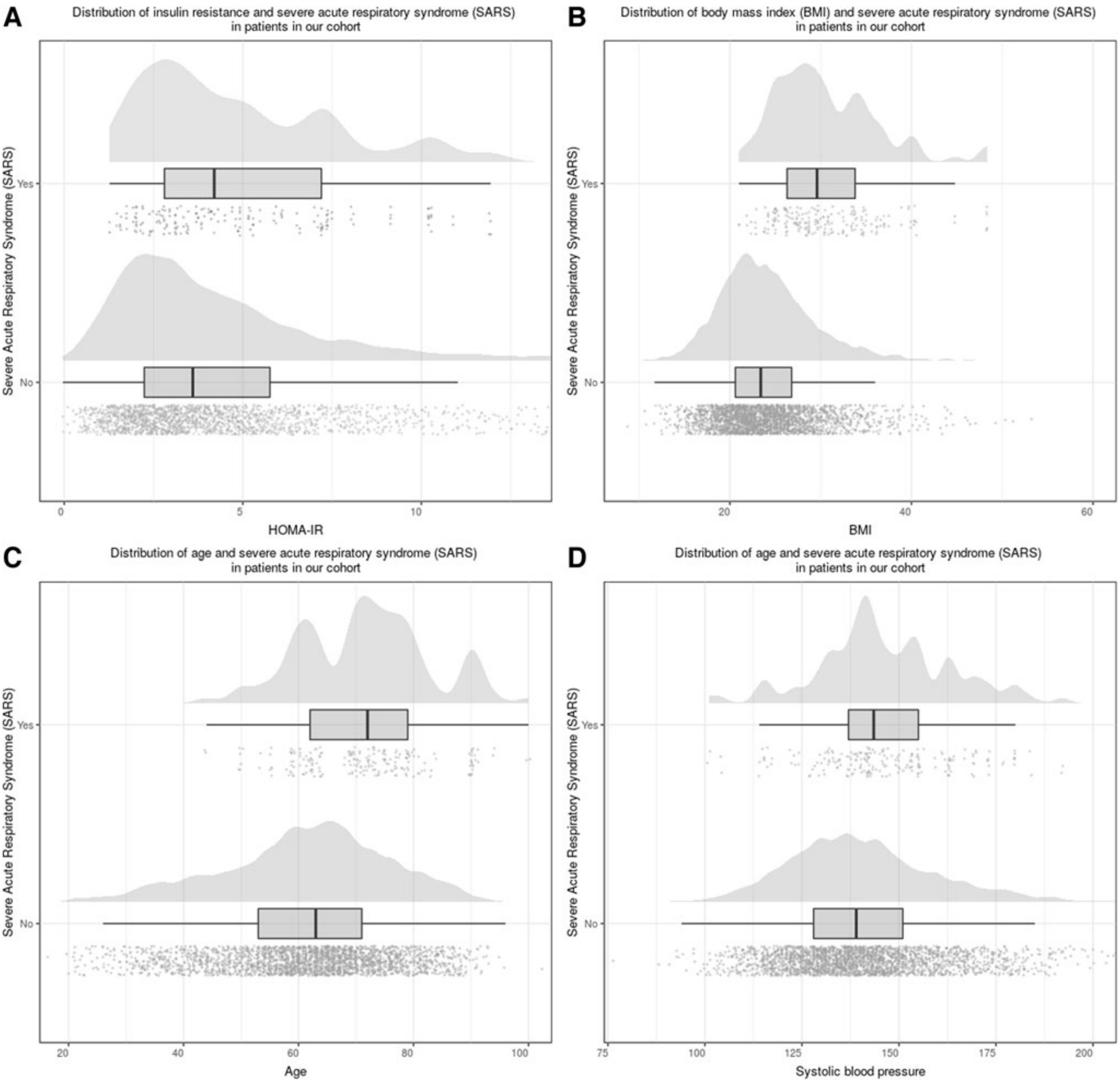

FIG. 1. Plots of some variables associated with the risk of hospitalization due to COVID-19. They show the distribution of HOMA-IR **(A)**, BMI **(B)**, age **(C)**, and SBP **(D)**. BMI, body mass index; COVID-19, coronavirus disease 2019; HOMA-IR, Homeostatic Model Assessment of Insulin Resistance; SBP, systolic blood pressure.

patient (panel B). The length of each bar represents the magnitudes or the effects of single variables on the patients. The color indicates whether a given effect is negative (red) or positive (blue). Only the 10 most influential variables are displayed. Blue bars indicate an increased probability of having the response variable (*i.e.,* supports the probability of being hospitalized), while red ones decrease the probability (*i.e.,* contradicts the probability of being hospitalized). These plots also indicate how well the model explains the need for hospitalization in a given patient.

For the first patient, our results predicted a probability of hospitalization of 85% accuracy, and an explanation fit of 76%. BMI, metabolic syndrome, and HOMA-IR were the most influential variables for this patient. Being older (>72 years) and hypertensive contradict the probability, but these variables had a small effect on this particular patient. However, the patient who required no hospitalization was predicted to have 91% of probability of not being hospitalized. Age <72 years, HOMA-IR <3.7, and not having metabolic syndrome were the most influential variables in this patient. Being preobese or obese (BMI >27.7), being hypertensive, and having chronic kidney disease contradicted this probability. Nevertheless, the explanation fit was 83%.

Discussion

We investigated whether metabolic syndrome and related conditions were associated with an increased risk of

TABLE 2. MULTIVARIATE ANALYSIS USING LOGISTIC REGRESSION AND PENALIZED LOGISTIC REGRESSION

	Binary LR			*LASSO (penalized regression)*	
	Beta coefficient	*OR (95% CI)*	P	*Beta coefficient*	*OR*
Male	0.462	1.58 (1.08–2.32)	0.018	0.223	1.250
Age	0.082	1.08 (1.06–1.10)	<0.001	0.062	1.064
BMI	0.212	1.23 (1.19–1.27)	<0.001	0.197	1.217
Diabetes	–0.837	0.43 (0.25–0.74)	0.003	—	—
Metabolic syndrome	0.283	1.32 (1.19–1.71)	<0.001	0.164	1.178
SBP	0.009	1.02 (1.01–1.03)	0.033	0.003	1.002
Uric acid	–0.092	0.91 (0.81–1.02)	0.127	—	—
Triglycerides	0.003	1.04 (1.02–1.05)	<0.001	—	—
HbA1c	–0.175	0.84 (0.65–1.07)	0.167	—	—
Insulin	–0.048	0.95 (0.92–0.98)	0.007	—	—
HOMA-IR	0.112	1.12 (1.02–1.22)	0.014	0.082	1.085
CKD-EPI-eGFR-creat	0.019	1.02 (1.01–1.03)	0.009	—	—

Log-likelihood = –456.7, chi-square = 147.7, *P* value = 0.001. Excluded variables were dropped from the final LR model.
CI, confidence interval; LASSO, least absolute shrinkage and selection operator; LR, logistic regression; OR, odds ratio.

hospitalization in patients with COVID-19. We found that metabolic syndrome, obesity, hypertension, and insulin resistance (determined according to high HOMA-IR) were independent risk factors after adjusting for sex and age. Our analyses using machine learning were key for developing a predictive model but also for interpretation of the effects of each condition in a single patient.

Our findings support the results of several previous studies in which metabolic syndrome was associated with hospitalization.[11,34] Regarding other biomarkers, our findings are in line with Wu et al.,[35] who explored the prognostic value of several biomarkers related to metabolic syndrome, including BMI, hypertension, and dyslipidemia. They found that patients with metabolic syndrome had worse hospitalization rates than those without it.

Metabolic syndrome and its biomarkers are common comorbidities associated with severity in patients with SARS-CoV-2 in terms of both hospitalization and death. The nature of the association between metabolic syndrome and severity of COVID-19 remains unclear. The clustering of obesity, dyslipidemia, insulin resistance, and hypertension is related to poor outcomes in several disorders,[16] so it is hypothesized that these comorbidities can also increase the risk of severe symptoms in COVID-19. Another hypothesis is that metabolic syndrome is associated with low-grade systemic inflammation and endothelial dysfunction,

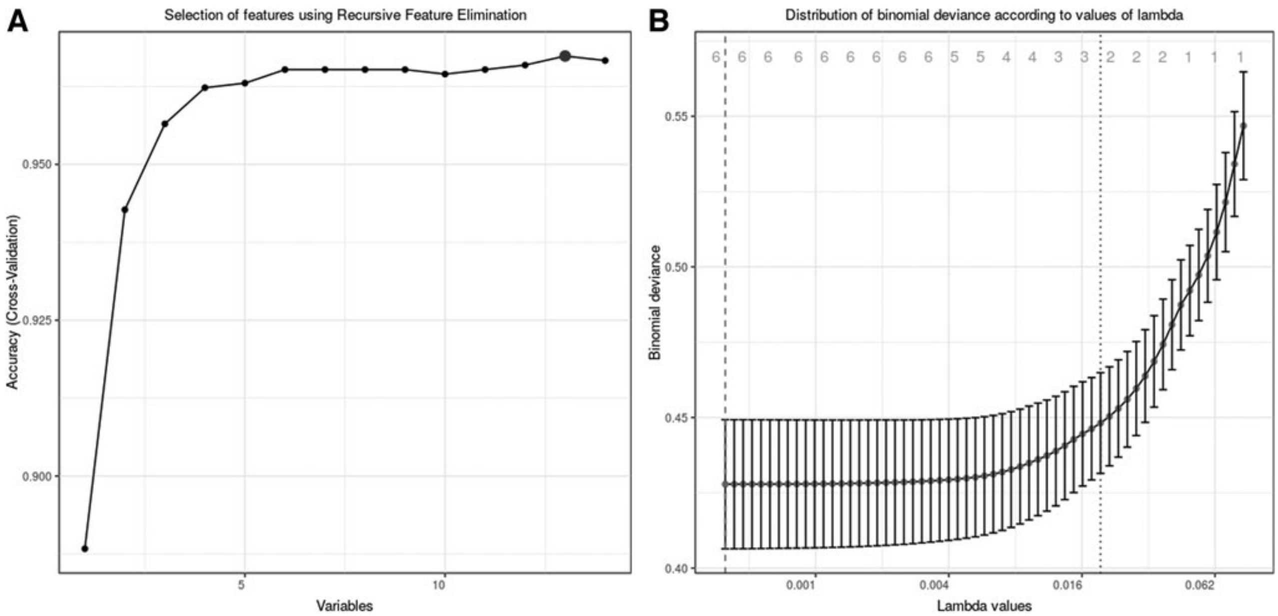

FIG. 2. **(A)** The performance of different sets of variables. From sets of four variables or greater, accuracy of the model is acceptable. The *thicker dot* represents the optimal solution (13 variables). **(B)** The results for penalized regression (LASSO). The lambda values are the penalization of the beta coefficients. Vertical lines represent the range of lambda in which accuracy is not adversely affected. The numbers at the *top* of the plot represent the number of variables of the model if penalized regression is used. A range between three and six variables was selected. LASSO, least absolute shrinkage and selection operator.

Table 3. Predictive Performance of the Risk Models

	LR	*RFE-CV*	*LASSO*
Accuracy	0.76 (0.71–0.82)	0.75 (0.71–0.80)	0.76 (0.70–0.81)
Sensitivity	0.77	0.75	0.77
Specificity	0.79	0.75	0.78
C-Index	0.69	0.65	0.65
RMSE	0.79	0.51	0.49

LASSO, penalized logistic regression using the least absolute shrinkage and selection operator approach; RFE-CV, recursive feature elimination with 10-fold crossvalidation; RMSE, root mean square error.

which have also been proposed as key pathogenic mechanisms in the development of severe COVID-19.[18,19,36]

A recent study of >29,000 hospitalized patients with COVID-19 analyzed the association between metabolic syndrome and its related biomarkers and mortality. Not only did metabolic syndrome predispose patients to an increased risk of severe disease and mortality but also each additional criterion (obesity, insulin resistance, hypertension, dyslipidemia) had a cumulative effect on the risk of severe COVID-19.[37]

A recent review highlighted the role of obesity and insulin resistance in the risk of complications due to COVID-19, such as hospitalizations, intensive care unit (ICU) admissions, and mortality.[38] COVID-19 is more severe in obese patients requiring hospitalization because they are at considerably increased risk of ICU admission, invasive mechanical ventilation, or death. Concerning insulin resistance, some studies have reported a high risk of hospitalization, ICU admission, and mortality in patients with impaired glucose metabolism, regardless of the diagnosis of diabetes; that is, even in the nondiabetic range. Our results are in line with some publications,[39,40] which indicate that obesity and insulin resistance are predictors of severity in COVID-19, and therefore can be considered independent risk factors for severe COVID-19.

In agreement with previous studies, we found that older age and male sex were risk factors for severe COVID-19; in some cohorts of previous studies,[41,42] the proportion of men was >60%. In this study, the main risk factor for hospitalization was age, a nonmodifiable condition. Age is strongly related to hospitalization and complications due to SARS-CoV-2, and patients >80 years old have a >20-fold increased risk of hospitalization. In addition, our results are in line

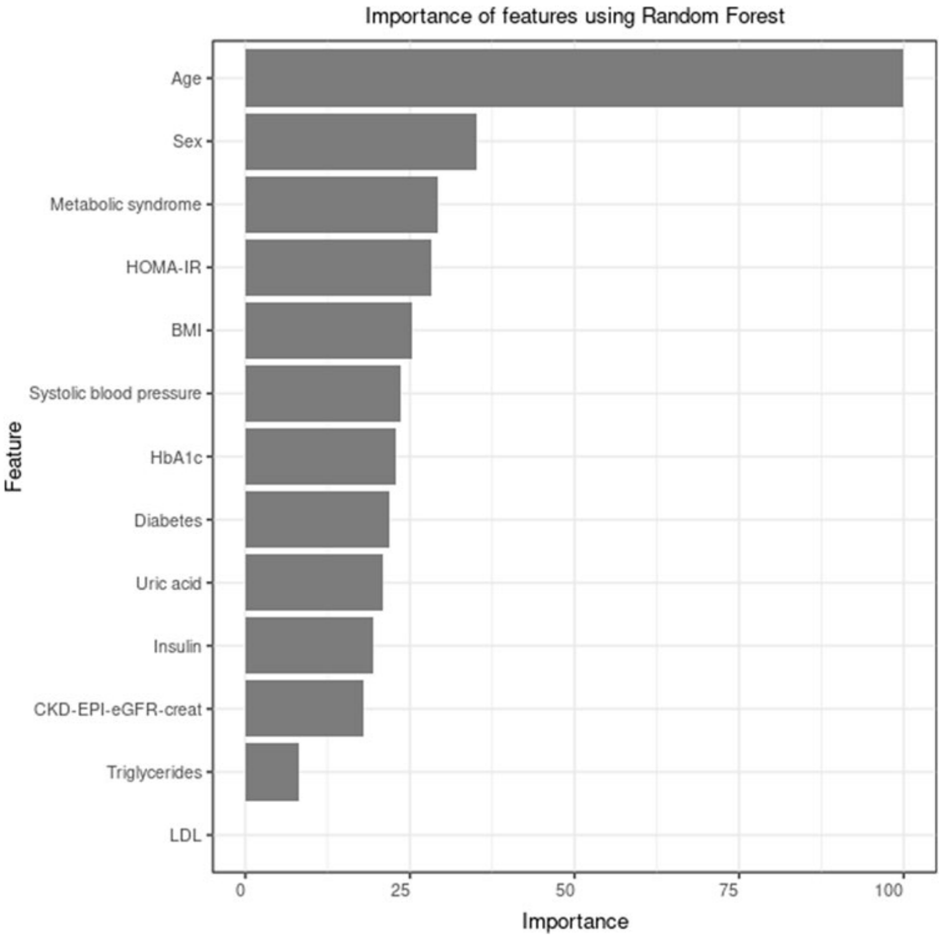

FIG. 3. Importance of features of patients with SARS-Cov-2 being hospitalized based on RF algorithm. RF, random forest; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

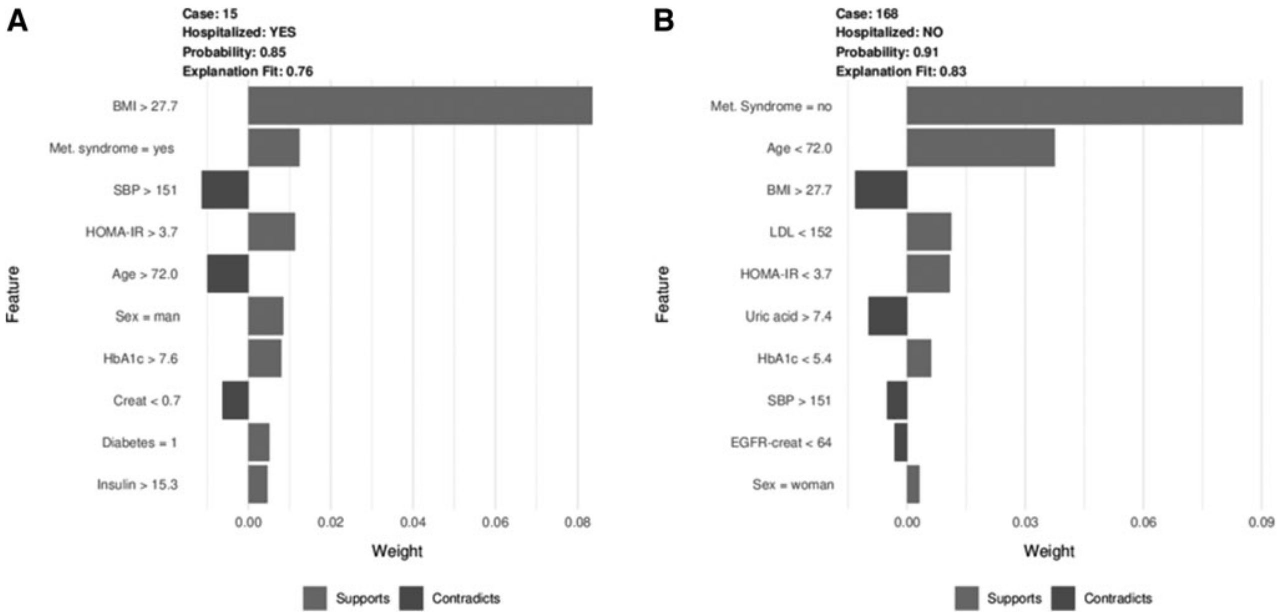

FIG. 4. Local interpretation for two single patients (labeled **A** and **B**) using the RF-based model.

with some studies that have reported that male sex is associated with hospitalization and increased mortality.[43] We consider our results important in terms of treatment and preventive measures. While age and sex are nonmodifiable predictors, obesity, hypertension, and insulin resistance can be modified with lifestyle interventions.

Our first statistical approach was to use logistic regression, a classical, reliable method, but we had too many variables. Among the proposed alternatives to address this issue, more advanced methods, such as penalized logistic regression (LASSO), have been suggested to produce a more parsimonious model; that is, fewer variables to explain the outcome[30] while retaining performance. Accuracy and reliability were retained from logistic regression but with only six variables instead of 12. Our proposed model combining RFE and LASSO is simple, parsimonious, and comprehensive.

We ran algorithms that involved feature selection, that is, for each model only a subset of relevant variables were selected, and nonrelevant, noisy variables were dropped. A simpler and easier-to-understand model could be built. Advanced machine learning techniques can help clinicians recognize patterns in observations that otherwise may be overlooked. These techniques can help us understand and better interpret the results. Data can be clarified to produce superior insights.[44–46]

Limitations

Our study had a number of limitations. First, it was designed as a cross-sectional, pilot study at a single institution. Although the patient sample was large, it could not be considered representative of the general population. Thus, our study was limited by the size and the time range of the sample. A multicenter or nationwide study could develop or challenge our results. Pilot studies such as this one are designed to be more descriptive and exploratory rather than powered to assess the effect of a variable. However, we consider our results consistent across several approaches (logistic regression, LASSO, RFE, RF, LIME), and they were in line with previous studies, so they may provide preliminary, reliable evidence for estimating the impact of metabolic syndrome and its related biomarkers.

It is worth noting that the sample was selected from March to April, 2020, when populations were not vaccinated. In further research, vaccination should be added as a new variable for analyzing the severity of the SARS-CoV-2 infection, as vaccination decreases the risk of hospitalization in patients with COVID-19.

Our study was not designed to identify causal relationships between related conditions of metabolic syndrome and the severity of COVID-19 but to assess the effect of these biomarkers on the risk of hospitalization, both globally and locally. Included variables were collected within the month before the infection. Because some laboratory parameters, such as leukocytes, lymphocytes, or C-reactive protein, may change in acute inflammatory episodes and can act as confounding factors, some data were intentionally not included as variables in our analyses. In addition, some laboratory parameters could not be included, such as markers of liver function, so we cannot exclude their role as potential confounding factors in our predictive models.

Another limitation was that we used metabolic syndrome as a unique condition—that is, we did not analyze the additive association between related biomarkers because we did not group patients as having three, four, or five of five criteria for metabolic syndrome. Thus, we did not analyze whether the risk was increased if a single criterion was added to the diagnosis of metabolic syndrome. Finally, we did not explore the effects of losing weight, controlling blood pressure, or managing insulin resistance on lowering the risk of hospitalization. This topic requires further investigation.

Conclusion

In this pilot study, we assessed whether metabolic syndrome and its related biomarkers are associated with increased odds for hospitalization due to COVID-19. Although our findings may require further research, our data suggest that metabolic syndrome, hypertension, insulin resistance, and obesity are linked to the development of SARS and the need for hospitalization in patients with confirmed COVID-19. They can be considered independent predictors for hospitalization. The novelty of our research lies on the use of machine learning in assessing individual risk for every single patient based on his/her conditions.

Authors' Contributions

Dr. R.G-.C. contributed to conceptualization (lead), software (lead), methodology (lead), writing—original draft (lead), formal analysis (lead), writing—review and editing (equal). Dr. O.V-.G., Dr. L-.L, and Dr. A.G-.d-.M. assisted with review and editing (equal). Dr. Lopez-.LombaL. supported review and editing (equal). Dr. G-.P. presented with methodology (supporting), review and editing (equal).

Availability of Data and Materials

A contract signed with Mostoles University Hospital, which provided the data set, prohibits the authors from providing their data to any other researcher. Furthermore, the authors were required to destroy the data upon the conclusion of their investigation. The data cannot be uploaded to any public repository.

Ethics Approval and Consent to Participate

This study was approved by the Ethical Board of Mostoles University Hospital (CEIC 2020/025). No identifying information was included in the article. Because the authors used historical data, informed consent was not necessary. All procedures involving human participants were conducted in accordance with the ethical standards of the responsible institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Author Disclosure Statement

No conflicting financial interests exist.

Funding Information

No funding was received for this article.

Supplementary Material

Supplementary Figure S1

References

1. Weekly Epidemiological Update on COVID-19-14 2022. n.d. Available from: https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19—14-december-2022 [Last accessed: January 16, 2023].
2. Ministerio de Sanidad—Profesionales—Situación Actual Coronavirus. n.d. Available from: https://www.sanidad.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov/situacionActual.htm [Last accessed: January 24, 2023].
3. World Health Organization. Statement on the second meeting of the International Health Regulations (2005). Emergency Committee Regarding the Outbreak of Novel Coronavirus (2019-nCoV). WHO Press: Geneva; 2020.
4. Machhi J, Herskovitz J, Senan AM, et al. The natural history, pathobiology, and clinical manifestations of SARS-CoV-2 infections. J Neuroimmune Pharmacol 2020;15(3):359–386; doi: 10.1007/s11481-020-09944-5
5. Lovato A, de Filippis C. Clinical presentation of COVID-19: A systematic review focusing on upper airway symptoms. Ear Nose Throat J 2020;99(9):569–576; doi: 10.1177/0145561320920762
6. Gorbalenya AE, Baker SC, Baric RS, et al. The species Severe acute respiratory syndrome-related coronavirus: Classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol 2020;5(4):536–544; doi: 10.1038/s41564-020-0695-z
7. Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: A retrospective cohort study. Lancet (London, England) 2020;395(10229):1054–1062; doi: 10.1016/S0140-6736(20)30566-3
8. Richardson S, Hirsch JS, Narasimhan M, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City Area. JAMA 2020;323(20):2052–2059; doi: 10.1001/jama.2020.6775
9. Petrilli CM, Jones SA, Yang J, et al. Factors associated with hospital admission and critical illness among 5279 people with coronavirus disease 2019 in New York City: Prospective cohort study. BMJ 2020;369:m1966; doi: 10.1136/bmj.m1966
10. Chen Y, Yang D, Cheng B, et al. Clinical characteristics and outcomes of patients with diabetes and COVID-19 in association with glucose-lowering medication. Diabetes Care 2020;43(7):1399–1407; doi: 10.2337/dc20-0660
11. Xie J, Zu Y, Alkhatib A, et al. Metabolic syndrome and COVID-19 mortality among adult black patients in New Orleans. Diabetes Care 2020;44(1):188–193; doi: 10.2337/dc20-1714
12. Ghoneim S, Butt MU, Hamid O, et al. The incidence of COVID-19 in patients with metabolic syndrome and non-alcoholic steatohepatitis: A population-based study. Metabol Open 2020;8:100057; doi: 10.1016/j.metop.2020.100057
13. Sanyaolu A, Okorie C, Marinkovic A, et al. Comorbidity and its impact on patients with COVID-19. SN Compr Clin Med 2020;2(8):1069–1076; doi: 10.1007/s42399-020-00363-4
14. Atmosudigdo IS, Lim MA, Radi B, et al. Dyslipidemia increases the risk of severe COVID-19: A systematic review, meta-analysis, and meta-regression. Clin Med Insights Endocrinol Diabetes 2021;14:1179551421990675; doi: 10.1177/1179551421990675
15. Yang J, Hu J, Zhu C. Obesity aggravates COVID-19: A systematic review and meta-analysis. J Med Virol 2021;93(1):257–261; doi: 10.1002/jmv.26237
16. Huang PL. A comprehensive definition for metabolic syndrome. Dis Model Mech 2009;2(5–6):231–237; doi: 10.1242/dmm.001180
17. American Diabetes Association. 4. Comprehensive Medical Evaluation and Assessment of Comorbidities: Standards of Medical Care in Diabetes—2021. Diabetes Care 2021;44(Supplement 1):S40—S52.
18. Chen G, Wu D, Guo W, et al. Clinical and immunological features of severe and moderate coronavirus disease 2019. J Clin Invest 2020;130(5):2620–2629; doi: 10.1172/JCI137244

19. Bonaventura A, Vecchié A, Dagna L, et al. Endothelial dysfunction and immunothrombosis as key pathogenic mechanisms in COVID-19. Nat Rev Immunol 2021;21(5): 319–329; doi: 10.1038/s41577-021-00536-9
20. Hu B, Huang S, Yin L. The cytokine storm and COVID-19. J Med Virol 2021;93(1):250–256; doi: 10.1002/jmv.26232
21. Montazersaheb S, Hosseiniyan Khatibi SM, Hejazi MS, et al. COVID-19 infection: An overview on cytokine storm and related interventions. Virol J 2022;19(1):92; doi: 10.1186/s12985-022-01814-1
22. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. Ann Intern Med 2009; 150(9):604–612.
23. Anonymous. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes—2018. Diabetes Care 2018;41(Supplement 1):S13 LP-S27.
24. Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults EP on. Executive Summary of The Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, And Treatment of High Blood Cholesterol In Adults (Adult Treatment Panel III). JAMA 2001;285(19): 2486–2497.
25. Guyon I, Elisseeff A. An introduction to variable and feature selection. J Machine Learn Res 2003;3:1157–1182.
26. Guyon I, Weston J, Barnhill S, et al. Gene selection for cancer classification using support vector machines. Machine Learn 2002;46(1):389–422.
27. Darst BF, Malecki KC, Engelman CD. Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. BMC Genet 2018; 19(1):65; doi: 10.1186/s12863-018-0633-8
28. Tibshirani R. Regression Shrinkage and Selection via the Lasso. J Royal Stat Soc Series B (Methodological) 1996; 58(1):267–288.
29. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010;33(1):1–22; doi: 10.18637/jss.v033.i01
30. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. BMC Med 2015;13(1):1; doi: 10.1186/s12916-014-0241-z
31. Breiman L. Random forests. Machine Learn 2001;45(1):5–32; doi: 10.1023/A:1010933404324
32. Ribeiro MT, Singh S, Guestrin C. ''Why Should I Trust You?'': Explaining the Predictions of Any Classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2016; 1135–1144; doi: 10.48550/arXiv.1602.04938
33. In J. Introduction of a pilot study. Korean J Anesthesiol 2017;70(6):601–605; doi: 10.4097/kjae.2017.70.6.601
34. Lohia P, Kapur S, Benjaram S, et al. Metabolic syndrome and clinical outcomes in patients infected with COVID-19: Does age, sex, and race of the patient with metabolic syndrome matter? J Diabetes 2021;13(5):420–429; doi: 10.1111/1753-0407.13157
35. Wu S, Zhou K, Misra-Hebert A, et al. Impact of metabolic syndrome on severity of COVID-19 illness. Metab Syndr Relat Disord 2022;20(4):191–198; doi: 10.1089/met.2021.0102
36. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020;395(10223):497–506; doi: 10.1016/S0140-6736(20)30183-5
37. Denson JL, Gillet AS, Zu Y, et al. Metabolic syndrome and acute respiratory distress syndrome in hospitalized patients with COVID-19. JAMA Netw Open 2021;4(12):e2140568; doi: 10.1001/jamanetworkopen.2021.40568
38. Stefan N, Birkenfeld AL, Schulze MB. Global pandemics interconnected—Obesity, impaired metabolic health and COVID-19. Nat Rev Endocrinol 2021;17(3):135–149; doi: 10.1038/s41574-020-00462-1
39. Wu J, Huang J, Zhu G, et al. Elevation of blood glucose level predicts worse outcomes in hospitalized patients with COVID-19: A retrospective cohort study. BMJ Open Diabetes Res Care 2020;8(1):e001476; doi: 10.1136/bmjdrc-2020-001476
40. Fei F, Smith JA, Cao L. Clinical laboratory characteristics in patients with suspected COVID-19: One single-institution experience. J Med Virol 2021;93(3):1665–1671; doi: 10.1002/jmv.26527
41. Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: Summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. JAMA 2020;323(13):1239–1242; doi: 10.1001/jama.2020.2648
42. Karagiannidis C, Mostert C, Hentschker C, et al. Case characteristics, resource use, and outcomes of 10,021 patients with COVID-19 admitted to 920 German hospitals: An observational study. Lancet Respir Med 2020;8(9):853–862; doi: 10.1016/S2213-2600(20)30316-7
43. Williamson EJ, Walker AJ, Bhaskaran K, et al. Factors associated with COVID-19-related death using OpenSAFELY. Nature 2020;584(7821):430–436; doi: 10.1038/s41586-020-2521-4
44. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med 2019;380(14):1347–1358; doi: 10.1056/NEJMra1814259
45. Obermeyer Z, Emanuel EJ. Predicting the future—Big data, machine learning, and clinical medicine. N Engl J Med 2016;375(13):1216–1219; doi: 10.1056/NEJMp1606181
46. Beam AL, Kohane IS. Big data and machine learning in health care. JAMA 2018;319(13):1317–1318; doi: 10.1001/jama.2017.18391

Address correspondence to:
Rafael Garcia-Carretero, MD, PhD
Department of Internal Medicine
Mostoles University Hospital
Rey Juan Carlos University (Madrid)
Calle Rio Jucar, s/n
Mostoles 28935
Spain

E-mail: rgcarretero@salud.madrid.org

5 Trends and Outcomes in a Secondary Hospital

viruses

MDPI

Article

Differences in Trends in Admissions and Outcomes among Patients from a Secondary Hospital in Madrid during the COVID-19 Pandemic: A Hospital-Based Epidemiological Analysis (2020–2022)

Rafael Garcia-Carretero [1,*,†], Oscar Vazquez-Gomez [1,†], María Ordoñez-Garcia [2,†], Noelia Garrido-Peño [3,†], Ruth Gil-Prieto [4,†] and Angel Gil-de-Miguel [4,†]

[1] Department of Internal Medicine, Mostoles University Hospital, 28935 Móstoles, Madrid, Spain; govaz5@hotmail.com
[2] Department of Hematology, Mostoles University Hospital, 28935 Móstoles, Madrid, Spain; maria.ordonez@hotmail.com
[3] Department of Pharmacy, Mostoles University Hospital, 28935 Móstoles, Madrid, Spain; noelia.garrido@salud.madrid.org
[4] Department of Preventive Medicine and Public Health, Rey Juan Carlos University, 28922 Alcorcón, Madrid, Spain; ruth.gil@urjc.es (R.G.-P.); angel.gil@urjc.es (A.G.-d.-M.)
* Correspondence: rgcarretero@salud.madrid.org
† These authors contributed equally to this work.

Citation: Garcia-Carretero, R.; Vazquez-Gomez, O.; Ordoñez-Garcia, M.; Garrido-Peño, N.; Gil-Prieto, R.; Gil-de-Miguel, A. Differences in Trends in Admissions and Outcomes among Patients from a Secondary Hospital in Madrid during the COVID-19 Pandemic: A Hospital-Based Epidemiological Analysis (2020–2022). *Viruses* **2023**, *15*, 1616. https://doi.org/10.3390/v15071616

Academic Editor: Pablo Barreiro

Received: 4 July 2023
Revised: 17 July 2023
Accepted: 21 July 2023
Published: 24 July 2023

Abstract: Spain had some of Europe's highest incidence and mortality rates for coronavirus disease 2019 (COVID-19). This study highlights the impact of the COVID-19 pandemic on daily health care in terms of incidence, critical patients, and mortality. We describe the characteristics and clinical outcomes of patients, comparing variables over the different waves. We performed a descriptive, retrospective study using the historical records of patients hospitalized with COVID-19. We describe demographic characteristics, admissions, and occupancy. Time series allowed us to visualize and analyze trends and patterns, and identify several waves during the 27-month period. A total of 3315 patients had been hospitalized with confirmed COVID-19. One-third of these patients were hospitalized during the first weeks of the pandemic. We observed that 4.6% of all hospitalizations had been admitted to the intensive care unit, and we identified a mortality rate of 9.4% among hospitalized patients. Arithmetic- and semi-logarithmic-scale charts showed how admissions and deaths rose sharply during the first weeks, increasing by 10 every few days. We described a single hospital's response and experiences during the pandemic. This research highlights certain demographic profiles in a population and emphasizes the importance of identifying waves when performing research on COVID-19. Our results can extend the analysis of the impact of COVID-19 and can be applied in other contexts, and can be considered when further analyzing the clinical, epidemiological, or demographic characteristics of populations with COVID-19. Our findings suggest that the pandemic should be analyzed not as a whole but rather in different waves.

Keywords: COVID-19; health care; SARS-CoV-2; time series visualization

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the coronavirus responsible for the coronavirus disease 2019 (COVID-19) pandemic and its more serious consequence, a severe respiratory illness called SARS [1,2]. It was first identified in the city of Wuhan (Hubei, China). The World Health Organization declared the outbreak a public health emergency on 30 January 2020, and eventually a pandemic on 11 March 2020 [3,4]. SARS-CoV-2 is a single-stranded RNA virus that often affects humans [5]. According to the U.S. Department of Health and Human Services, this virus is related to SARS-CoV-1, which caused an outbreak of SARS between 2002 and 2004 [6,7], and with the Middle East Respiratory Syndrome coronavirus (MERS-CoV), which first occurred in 2012 and has been causing persistent endemics in the countries of the Middle East [8,9].

Outbreaks of pandemics typically spread in regular patterns, usually as logarithmic increases in the number of confirmed cases, which are also called exponential curves. However, waves of COVID-19 have varied widely among countries and even regions within a single country depending on the intensity of government and public measures and interventions, along with other factors such as the use of lockdowns, social distancing measures, vaccination, or border policies [10,11]. For example, there were several differences in incidence and mortality rates between countries such as Spain or Italy (with mortality rates up to 15%) and countries such as Germany or Canada (with mortality rates less than 5%) in the first wave of COVID-19 [12–14]. The reasons for these differences remain unclear, although some authors have proposed differences in characteristics of the population, government strategies, heterogeneous health systems, or public health interventions [15,16].

Several drugs have been introduced for the treatment of COVID-19. Corticosteroids, such as dexamethasone, have demonstrated efficacy in reducing inflammation and improving outcomes in severe cases. Plasmapheresis, a procedure that removes and replaces blood plasma, has been explored as a potential treatment option to remove harmful antibodies in critically ill patients. Anticoagulants, such as heparin, are administered to prevent blood clotting complications associated with COVID-19. Immunomodulators, such as tocilizumab, act to regulate the immune response and are utilized in severe cases with cytokine release syndrome. Antiviral drugs, including remdesivir, target the replication of the SARS-CoV-2 virus. These drugs, used in various combinations and based on disease severity, have shown promise in improving outcomes and reducing the severity of COVID-19 [17–24].

1.1. The Epidemiological Situation in Spain: A Timeline

Clinical and demographic data on the first wave of COVID-19 in Spain were published early on and offered an overall view of the pandemic [25]. Spain's public health system and intensive care units (ICUs) were overwhelmed with the excessive workload, the high incidence of hospital admissions, and deaths due to COVID-19 [26]. Hospitalizations due to common illnesses and programmed, noncritical surgical interventions decreased. Although some research has been performed in Spain regarding demographic characteristics [25,27,28], to the best of our knowledge no study has compared clinical data and outcomes for different waves. It is key to highlight the importance of identifying waves when performing research on COVID-19 at a given time and in a given location to assess evidence-based decision-making and the impact of COVID-19.

The first case of COVID-19 in Spain was confirmed on 31 January 2020. When the World Health Organization declared the existence of a pandemic on 11 March 2020, Italy and Spain had the highest incidence in Europe, and Spain declared a state of emergency on 14 March 2020. In addition to the lockdown, several measures were instituted, such as mobility restrictions, border closings, and mandatory masking [29]. Spain had one of the highest incidences of COVID-19 in Europe, accounting for 172,541 confirmed cases and 18,056 deaths in the first wave (14 April 2020) [30–32]. As of 16 June 2023, a total of 3,905,048 confirmed cases and 121,622 deaths had been reported in Spain [31]. The global incidence was 767,984,989 confirmed cases and 6,943,390 deaths.

1.2. The Importance of Data Visualization

Understanding the global pandemic of COVID-19, with its vast amount of data and statistics, can be overwhelming. Data visualization plays a crucial role in identifying trends and gaining insights into the pandemic. However, it is important to note that not all statistics are reliable, and the way data are presented can influence our perception of the situation. Therefore, researchers need to effectively display and represent data to comprehend the outbreak better; although a simple daily count of new cases is the easiest way to present pandemic data, it can be misleading without proper context. To grasp the evolving nature of the pandemic, graphs like histograms, scatterplots, and line plots provide more meaningful trends at a glance. Additionally, a cumulative graph, which shows the total number of confirmed cases per day since the beginning of the pandemic, can be helpful. However, it is essential to exercise caution with cumulative graphs, as they might not clearly indicate if the growth rate is slowing. Researchers must identify a plateau in the curve to demonstrate a slowdown, as cumulative charts always show increasing cases [33].

With arithmetic-scale graphs, researchers can easily identify patterns or trends. The distance along any axis always represents the same quantity. In our research, the space between tick marks along the y-axis (vertical axis) is the same, as the y-axis shows a continuous variable (admissions, cases, deaths). As a result, the distance from 1 to 10 is the same as the distance from 11 to 20. Ticks represent absolute values. If the same data were displayed using a logarithmic scale for the y-axis, we would obtain a semi-logarithmic-scale line graph. In this chart, the distance from 1 to 10 is the same as the distance from 10 to 100. This means that the y-axis is ranked in order of magnitude (10^0, 10^1, 10^2, 10^3). We use a semi-logarithmic scale in certain cases that are especially useful for understanding the impact of the pandemic, in particular if the disease is growing exponentially.

1.3. Objectives of This Research Study

Improved knowledge of the distribution of confirmed cases, admissions, and deaths due to COVID-19 throughout subsequent phases of the pandemic shed light on the behavior of the virus and its impact on the health care system. New insights will help public health authorities make appropriate decisions and design interventions to manage the pandemic. In light of recent events, our research endeavors to utilize the most up-to-date data from our hospital to examine and investigate the implications and repercussions of the pandemic. Our primary objective is to elucidate the temporal progression of various key variables, such as hospitalizations, occupancy rates, ICU admissions, and deaths, in order to uncover any discernible patterns and trends that may be underlying. Additionally, by presenting these data, we aim to quantify the extent of the health care impact caused by COVID-19 within a specific hospital setting. This valuable information has the potential to contribute to the overall understanding of individual hospital experiences and even serve as a reference for health care systems on a national scale, aiding in their assessment of the pandemic's effect on their own systems.

2. Materials and Methods

2.1. Data Collection

We conducted a retrospective, descriptive, epidemiological study in which we determined the frequency and distribution of cases of the COVID-19 pandemic. We included individuals admitted to the hospital whose cause of hospitalization was COVID-19. Therefore, hospitalization due to COVID-19 was defined as having a confirmed infection with SARS-CoV-2 (usually, a positive polymerase chain reaction (PCR) test result). Patients whose admission criteria or discharge report included severe acute respiratory infection as the cause of hospitalization were included. We analyzed trends in newly confirmed cases admitted to our hospital, occupancy time, and mortality rates. Because this study was merely a descriptive investigation, no hypotheses were made. Data were collected from electronic records of Mostoles University Hospital (Spain). Age, sex, admission, discharge dates, comorbidities, drug therapy, status at discharge (alive or dead), and ICU admission were collected. Hospital and ICU stay refer to the length of stay; that is, the duration (in days) of a single episode of hospitalization or ICU admission. Furthermore, by incorporating sex-disaggregated data into statistical presentations, we promote a more inclusive and accurate understanding of the clinical presentation of COVID-19, which would allow us to uncover gender-based patterns, disparities, and trends that may remain hidden in aggregated data. Each patient was given a unique identification number to ensure anonymity. This study was approved by the Ethical Board of Mostoles University Hospital (CEIC 2020/025). According to official sources, our hospital attends to a population of 168,000, with more than 500 hospital beds and 12 critical care beds available [34]. Mostoles University Hospital can be considered a secondary hospital (also known as an intermediate complexity hospital).

Given that the COVID-19 pandemic is still ongoing, we established a research window from the beginning of the pandemic in our hospital (25 February 2020) to the end of the observation period on 12 May 2022. This observation window covered six waves (almost 27 months) of the pandemic. We are aware that every country, and even every region within a single country, has had different waves of COVID-19, and the distribution of confirmed cases varies with the implementation of control strategies. Given these regional differences, we describe the experiences and the distribution of the pandemic at our own institution; our splitting of the pandemic into six waves is utterly idiosyncratic and cannot be extrapolated to other settings.

2.2. Statistical Analyses

We plotted continuous variables (age, hospital stay, ICU stay) to check for normality. However, because visual inspection can be unreliable, we used the Shapiro–Wilk test. That is, we combined visual inspection and significance testing to ensure that the assumptions of the statistical tests were met. We also performed several tests of independence. Continuous variables were tested with the Mann–Whitney–Wilcoxon test as a nonparametric alternative to the one-sample t-test when the data could not be assumed to be normally distributed. Data were then expressed as means and standard deviations or as medians and interquartile ranges. We also used the one-proportion Z-test with Yates continuity correction to compare observed proportions of patients in each wave, given that there were only two categories (men and women), to determine whether the proportion of men with COVID-19 differed significantly from the proportion of women with the disease. Categorical data (deaths, ICU admissions) were tested with the chi-square test. Log-linear analysis was used to examine the relationship between more than two categorical variables, such as comparing age and sex throughout the different waves. We used this technique exclusively for hypothesis testing (i.e., as a test of independence). Although we could have used Pearson's chi-square test instead of log-linear analysis, chi-square only allows for a two-way contingency table analysis (i.e., only two variables can be compared at a time) [35,36]. In contrast, log-linear analysis is a form of categorical data analysis used mostly with three-way contingency tables and can be considered an extension of Poisson regression. This is why they are called Poisson log-linear models.

We set the significance level at $p = 0.05$. We used R language (version 4.2.0) and Python (version 3.7.3 with scikit-learn libraries). The use of either language was not exclusive but rather complementary, depending on the ease of producing statistical metrics or visualization of the data.

2.3. Data Visualization

As noted previously, the importance of data representation and visualization for exploratory analyses should not be overlooked because the way in which data are represented affects how researchers interpret them and thus what conclusions are drawn from them. Most research focuses on the visualization of data as time series; that is, groups of observations of a single entity ordered in time. Here, we considered several entities or variables: hospital admissions, ICU and hospital occupancy, and deaths. Observations were conducted daily, and we plotted them with an aim to describe (i.e., we tried to interpret their distribution over time and extract basic useful insights). Because our objective was not to forecast the future, we did not analyze factors such as seasonality, autocorrelation, or stationarity.

Apart from visualizing the data, we also produced tables with useful demographic characteristics, which we split into waves. Waves were set according to a specific signal so that we could extract the position and intensity of multiple peaks and valleys. It is important to note that, as mentioned before, we identified the dates of peaks and valleys for our institution, but we are aware that dates vary from hospital to hospital and even among regions of the same country. We calculated peaks and valleys inside our time series using the function find_peaks from scipy.signal (Python language). We then plotted the resultant figures to show the different segments we considered waves.

The time series were plotted as a continuous line, but given the daily variations, this line turned out to be sharp and rough (i.e., very noisy). We then chose to plot the time series as single points (i.e., a scatterplot) but with a smooth line to identify interesting trends in the data. The simplest method of reading data is to use the moving average (i.e., a model that states that an observation is the mean of a window of past observations). We defined a window to apply the moving average model to smooth the time series and highlight different trends. Whereas a scatterplot represents real data, a moving average line represents trends and different waves. A moving average is less sensitive to abrupt changes, outliers, or missing values and corrects the trend of the time series. It thus smooths fluctuations in the short term. We chose a window of 7 days (i.e., we calculated the mean of the previous seven values). We also used population pyramids for both admissions and deaths. A population pyramid shows the percentage or count of the population by age and sex using two histograms. For other representations of data, we followed Allen et al. [37] and introduced rainclouds, an alternative to bar plots and boxplots, to display probability density plots, raw data points, and boxplots, which show complex, heterogeneous data at a glance.

Finally, as mentioned previously, we used arithmetic-scale cumulative and semi-logarithmic-scale cumulative graphs to better represent the impact of the pandemic. We used some of these charts or a combination of them to obtain a better understanding of the data.

3. Results

3.1. General Characteristics and Waves

We collected data from 3315 hospitalized patients from 25 February 2020, to 12 May 2022. We identified six waves based on the peaks and valleys in the time series (see Figure 1 showing the calculated peaks and valleys). Based on Figure 1, we plotted Figure A1 in the Appendix A, which splits the time series into six different waves, with plateaus of different lengths between each wave. This method of organization allowed us to group patients in waves and analyze their characteristics. Table 1 summarizes our findings, with clinical characteristics, associated comorbidities, and drug therapy. The number of patients decreased in each wave, except for the sixth wave, in which we found 513 patients. COVID-19 affected more men than women, both globally (55% men vs. 45% women) and in each individual wave. Hypertension, obesity, and type 2 diabetes were the most common comorbidities. We observed a decrease in the frequency of all comorbidities over time, especially in the fourth and fifth waves. The most common drug was dexamethasone, which began to be used at the end of the first wave. Since then, almost all patients were on corticosteroids. Patients on immunomodulatory drugs such as baricitinib, tocilizumab, or anakinra decreased over time. Of note, the standard treatment at the beginning of the pandemic was the combination of lopinariv/ritonavir, hydroxychloroquine, and azithromycin, but they were no longer used after the second wave, as they were replaced by improved drug therapies.

Figure A2 summarizes the evolution of the COVID-19 pandemic during the 27 months since the beginning of the outbreak. We plotted daily admissions, ICU/hospitalization ward occupancy, and deaths over time. The median stay was 7 days (interquartile range: 8). Men tended to stay longer than women, both in the hospitalization ward and ICU (Table 2). Figure A3, panel A shows the distribution of hospitalization stays by wave.

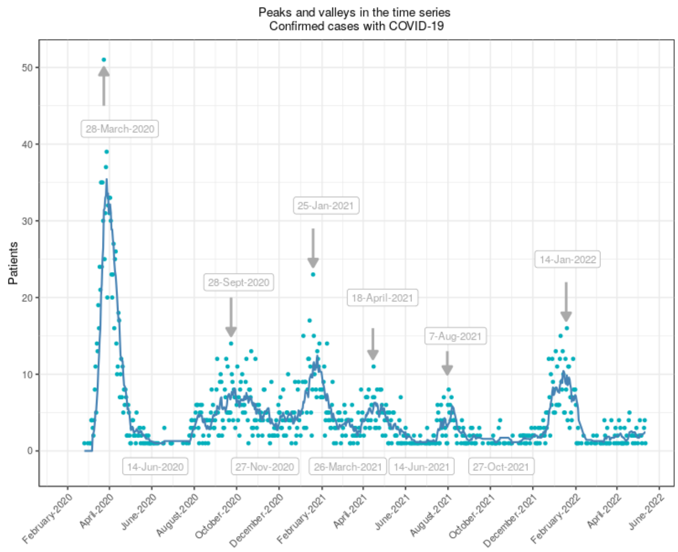

Figure 1. Calculated peaks and valleys inside the time series. Dots represent raw data, whereas the blue line represents a moving average of the time series.

Table 1. Epidemiological and demographic characteristics of patients admitted to our hospital between 2020 and 2022.

	First Wave	Second Wave	Third Wave	Fourth Wave	Fifth Wave	Sixth Wave	All Waves
Patients, n (%)	1024 (30.9%)	652 (19.7%)	646 (19.4%)	278 (8.3%)	202 (6.2%)	513 (15.5%)	3315
Sex							
Men	553	363	359	164	105	279	1823 (55%)
Women	471	289	287	114	97	234	1492 (45%)
Age, median (IQR)	70 (22.2)	65 (26)	66 (23)	60 (21)	47 (32.8)	70 (26)	67 (25)
Age ranges							
<20	3	8	10	4	12	31	68
21–40	43	68	48	24	51	32	266
41–60	246	181	186	112	64	96	885
61–80	476	273	284	119	47	223	1422
>80	256	122	118	19	28	131	674
Comorbidities							
Type 2 diabetes	19.3%	21.8%	21.7%	16.4%	18.4%	21.2%	20%
Hypertension	33.7%	33.2%	34.7%	29.7%	25.5%	32.4%	32.7%
Obesity	8.4%	0.13%	13.6%	15.6%	14.1%	12.4%	11.8%
AMI	6.8%	0.07%	7.5%	4.8%	0.06%	6.6%	6.7%
CHF	6.1%	0.08%	0.08%	3.8%	7.8%	7.3%	6.9%
Dementia	5.2%	4.6%	4.4%	1.9%	4.6%	4.3%	4.5%
Kidney disease	8.9%	9.4%	0.1%	5.3%	9.4%	0.1%	8.9%
Liver disease	0.5%	0.5%	0.5%	0.4%	0.4%	0.6%	0.5%
Malignancy	5.3%	5.9%	0.06%	3.9%	5.6%	7.5%	5.5%
COPD	7.1%	7.3%	8.1%	0.06%	7.4%	8.9%	7.3%
CEVD	0.7%	0.8%	0.8%	0.5%	0.7%	0.01%	0.7%
Drug therapy, n (%)							
Dexamethasone	131 (12.8%)	614 (94.2%)	641 (99.2%)	275 (98.9%)	200 (99%)	502 (97.9%)	2763 (83.3%)
Remdesivir	0 (0%)	230 (35.3%)	156 (24.1%)	96 (34.5%)	41 (20.3%)	79 (15.4%)	602 (18.2%)
Baricitinib	44 (4.3%)	95 (14.6 %)	130 (20.1%)	63 (22.7%)	39 (19.3%)	46 (9%)	417 (12.6%)
Tocilizumab	137 (13.4%)	222 (34%)	90 (13.9%)	61 (21.9%)	21 (10.4%)	27 (5.3%)	558 (4.1%)
Anakinra	12 (1.2%)	19 (2.9%)	13 (2%)	1 (0.4%)	0 (0%)	3 (0.6%)	48 (1.4%)
LPV/r, HCQ, AZM	830 (81.1%)	57 (8.7%)	0 (0%	0 (0%)	0 (0%)	0 (0%)	887 (26.8%)

AMI: acute myocardial infarction. CHF: congestive heart failure. CEVD: cerebrovascular disease. COPD: chronic obstructive pulmonary disease. IQR: interquartile range. LPV/r: lopinariv/ritonavir. HCQ: hydroxychloroquine. AZM: azithromycin.

Table 2. Outcomes in terms of ICU admissions and mortality of patients admitted to our hospital (2020–2022).

	Total Patients	Men	Women	*p* Value
Age	67.0 (25.0)	66.0 (24.0)	68.0 (25.0)	0.001 *
First wave	70.0 (22.2)	68.0 (20.0)	72.0 (24.0)	0.001
Second wave	65.0 (26.0)	64.0 (24.0)	67.0 (27.0)	0.159
Third wave	66.0 (23.0)	65.0 (25.0)	68.0 (21.5)	0.004
Fourth wave	60.0 (21.0)	59.0 (20.0)	65.5 (22.8)	0.211
Fifth wave	47.0 (32.8)	47.0 (32.0)	47.0 (34.0)	0.566
Sixth wave	70.0 (26.0)	72.0 (25.5)	69.0 (26.0)	0.2
Hospital stay (days)	7.0 (8.0)	7.0 (8.0)	6.0 (7.0)	0.001 *
First wave	8.0 (9.0)	8.0 (10.0)	7.0 (8.0)	0.103
Second wave	7.0 (9.0)	7.0 (8.0)	7.0 (8.0)	0.068
Third wave	6.0 (7.0)	6.0 (8.0)	6.0 (5.0)	0.028
Fourth wave	7.0 (7.8)	8.0 (7.2)	7.0 (7.8)	0.586
Fifth wave	5.0 (6.0)	5.0 (6.0)	5.0 (5.0)	0.353
Sixth wave	5.0 (6.0)	5.0 (6.0)	5.0 (6.0)	0.588
ICU admissions	154 (4.6%)	108 (5.9%)	46 (3.1%)	0.001 **
First wave	59 (5.8%)	46 (8.3%)	13 (2.8%)	0.001
Second wave	27 (4.1%)	16 (4.4%)	11 (3.8%)	0.853
Third wave	27 (4.2%)	18 (5.0%)	9 (3.1%)	0.323
Fourth wave	17 (6.1%)	11 (6.7%)	6 (5.3%)	0.81
Fifth wave	8 (4.0%)	7 (6.7%)	1 (1.0%)	0.067
Sixth wave	16 (3.1%)	10 (3.6%)	6 (2.6%)	0.614
ICU stay (days)	19.0 (27.0)	18.0 (24.5)	21.5 (35.5)	0.492 *
First wave	7.0 (6.8)	7.0 (7.2)	2.0 (5.0)	0.023
Second wave	6.0 (4.0)	7.0 (4.0)	8.0 (4.8)	0.347
Third wave	5.0 (5.0)	5.0 (5.0)	5.0 (5.8)	0.998
Fourth wave	7.0 (6.0)	10.5 (5.5)	8.5 (10.2)	0.263
Fifth wave	5.0 (5.0)	14.0 (4.8)	49 (5.8)	0.001
Sixth wave	5.0 (5.0)	15.5 (5.0)	17.5 (5.8)	0.625
Deaths	310 (9.4%)	197 (10.8%)	113 (7.6%)	0.002 **
First wave	170 (16.6%)	108 (19.5%)	62 (13.2%)	0.008
Second wave	40 (6.1%)	27 (7.4%)	13 (4.5%)	0.165
Third wave	53 (8.2%)	33 (9.2%)	20 (6.9%)	0.379
Fourth wave	12 (4.3%)	9 (5.5%)	3 (2.6%)	0.37
Fifth wave	8 (4.0%)	4 (3.8%)	4 (4.1%)	1
Sixth wave	27 (5.3%)	16 (5.7%)	11 (4.7%)	0.746

*: Mann–Whitney–Wilcoxon test. **: Chi-square test. ICU: Intensive Care Unit. Hospital and ICU stay refer to the length of stay; that is, the duration (in days) of a single episode of hospitalization or ICU admission. Age and hospital stay are expressed as the median (interquartile range). The rest of the variables are expressed as frequencies and percentages. Percentages regarding ICU admissions and deaths refer to the ICU admission rate and mortality rate, respectively, in each wave.

3.2. Analyses by Sex and Age

Regarding sex and age, we found that globally women admitted to the hospital because of COVID-19 were older than men, except for those impacted during the third wave. It is worth noting that the average age of the entire population dropped to 47 (interquartile range: 32.8) at the end of the fifth wave. Population pyramids for both admissions and deaths are plotted in Figure 2, respectively. Most hospitalized patients were older than 40 years old. Figure 2 also shows that men predominated in the cohort studied, and that mortality was predominant in men older than 60. However, a closer look at the raincloud computed to visualize those data (Figure A3) shows that the distribution of the age was not normal in any of the waves. In fact, the distribution of age is often multimodal, and utilizing the mean or median as a metric can be misleading. Thus, we disaggregated the data by age, sex, and wave and performed a log-linear analysis, which can be understood as a Poisson regression applied to multiway contingency tables, as mentioned earlier. We found there were no significant differences by wave in those aged 0 to 30 years old. We studied mutual independence among age, wave, and sex and joint independence among these three variables (i.e., interactions among variables during the observation period). Admissions in the population younger than 40 remained steady from the first wave until the sixth wave. We also found that admissions of young patients (<61), regardless of sex, decreased over time. However, an interesting phenomenon occurred in admissions of those 61 to >80 years old: the number of hospitalizations decreased steadily until the fifth wave but increased in the sixth wave, with no differences by the sex of the patient.

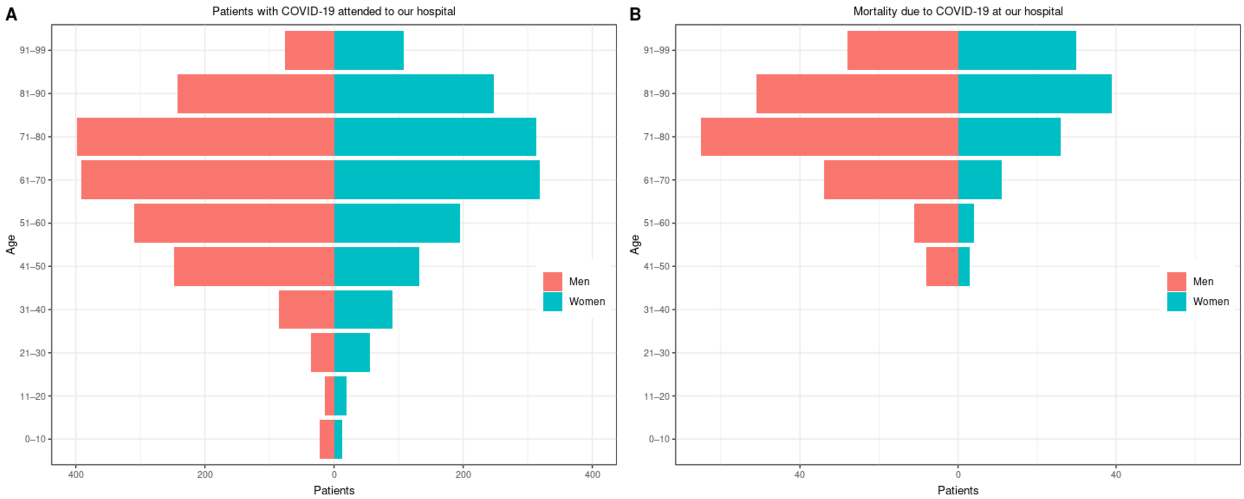

Figure 2. Distributions of patients in the population admitted to the hospital (**A**) and mortality (**B**), both in age–sex pyramids. The plots were computed to display the distribution of the population attended to in our hospital affected with COVID-19. The pyramids depict the impact of the illness by age and sex on admission (**A**) and mortality (**B**).

3.3. ICU Admissions

Panel B in Figure A2 shows the impact of the pandemic on the ICU in our hospital. The horizontal line shows the baseline of 12 critical beds, which were often outnumbered by ICU admissions. This forced hospital authorities to create new spaces for critical beds, such as surgery rooms and postoperative beds. We recorded 154 patients admitted to the ICU; similar to ward admissions, the number of ICU patients decreased steadily until the fifth wave but increased in the sixth wave, with no differences in the sex of the patient. The median ICU stay was 19 days in contrast to 7 days for the median ward stay. Men stayed longer than women in the ICU. The distribution of the median ICU stay was more heterogeneous, often multimodal, and had frequent outliers; a single median likely would not convey the real context of the situation, so we plotted the results in Figure A3, panel B, by means of a raincloud chart.

3.4. Mortality

We recorded a global mortality rate of 9.4% among hospitalized patients (310 deaths). The data showed that men were more prone to die than women (197 men vs. 113 women, $p = 0.002$), as seen in Figure 2. The first wave had a higher mortality rate (16.6%) than the other waves for both men and women. The fourth and fifth waves had lower mortality rates. A slight increase could be seen in the sixth wave. It is worth noting that we observed this increase globally. Table 3 shows the distribution of mortality by groups of age and sex. Men had a higher mortality rate than women, but when we split the data into waves, we could not find differences in terms of sex.

We plotted several charts to visualize the impact of mortality on our hospital compared to admissions. Figure A4 shows these graphs. We first plotted the cumulative incidence of admissions and deaths among confirmed cases of COVID-19 on an arithmetic scale (panels A and B in Figure A4). However, the semi-logarithmic-scale plots showed the true impact and burden on the hospital (panels C and D in Figure A4). In the first 2 months, there were approximately 1000 admissions and more than 100 deaths, almost one-third of all admissions and deaths in the observation period. In May 2020, the curves tended to flatten and remain steady until January 2021 (third wave). After a stable plateau, an increased incidence was observed in January 2022 (sixth wave).

Table 3. Mortality (in absolute values) due to COVID-19 according to age range and sex.

Age Group	1st Wave	2nd Wave	3rd Wave	4th Wave	5th Wave	6th Wave
Men						
<20	0	0	0	0	0	0
21–40	0	0	0	0	0	0
41–60	11	1	2	1	2	2
61–80	58	11	17	7	1	5
>80	39	15	14	1	1	9
Women						
<20	0	0	0	0	0	0
21–40	0	0	0	0	0	0
41–60	5	0	0	0	0	2
61–80	19	4	5	3	1	5
>80	38	9	15	0	3	4
Total						
<20	0	0	0	0	0	0
21–40	0	0	0	0	0	0
41–60	16	1	2	1	2	4
61–80	77	15	22	10	2	10
>80	77	24	29	1	4	13

We also created some plots to visualize mortality over time compared to all discharges. Figure A5 summarizes the data from Table 2. Up to 25% of all persons discharged in the first month of the pandemic (March 2020) died, which emphasizes the great impact of the pandemic on daily work.

3.5. The Pandemic in the Area Attended to by Our Hospital

We plotted the confirmed cases in the area attended to by our hospital (population: 168,000) to put our data into context (Figure A6, online data source: see [38]). Here, our aim was to compare public, official data on confirmed cases among the general population to data on those who had been admitted to our hospital. We found differences in the heights of the peaks of incidence.

Finally, we also plotted the ratio of population in the region of Madrid with a complete vaccination schedule (Figure A7, online data source: see [34]), which raised to 55% by August 2021.

4. Discussion

The main objective of our research was to analyze and visualize admissions, occupancy, and mortality due to COVID-19 to evaluate the impact of the pandemic on daily work and pressure on the health care system in a peripheral hospital. The outbreak of COVID-19 overwhelmed our ICUs and the capacity of our hospital. We split the pandemic into waves and analyzed each one separately. The waves turned out to be heterogeneous and dissimilar. Patients in each wave had different epidemiological and demographic profiles. The reasons for these differences remain unclear, but we have some hypotheses: age, sex, social events, new treatments, vaccination, and SARS-CoV-2 variants. Although here we discuss these hypotheses, we can only establish the association, but not causation, between certain events and peaks of hospitalizations in our geographic area.

Regarding age- and sex-specific analyses, women were less prone to have bad outcomes (ICU admission, death), except in the elderly. The interpretation of such differences was challenging, and inconsistent findings can justify a necessity for a more precise analysis that would elucidate the impact of sex and age in the outcomes of COVID-19. Regarding the diagnosis of COVID-19 cases, no apparent sex or gender bias has been observed, although this may vary across countries. However, a noteworthy finding emerges when considering disease progression to severe conditions and mortality, as male individuals exhibit a significant disadvantage. A hypothesis would be that men tend to die earlier than women globally, so it could be that COVID-19 is exacerbating underlying mortality differences. The existence of biological differences in immune systems between men and women can influence their ability to combat infections, including SARS-CoV-2. Generally, females exhibit greater resistance to infections compared to males. Furthermore, lifestyle choices, such as higher rates of smoking and alcohol consumption among men, may contribute to this disparity. Moreover, it is noteworthy that women tend to display a more responsible attitude towards the COVID-19 pandemic when compared to men [39,40].

The use of new drugs, beginning in summer of 2020, can explain the lower mortality after the first wave, shown in Table 1 and Figure A4. This can be explained by the use of new treatments, such as corticosteroids [41,42], antivirals, and immunomodulatory drugs, such as dexamethasone, remdesivir, anakinra, tocilizumab, or baricitinib [19–24]. Corticosteroids are beneficial in treating severe COVID-19 cases by reducing lung inflammation and preventing complications. Dexamethasone, studied in the RECOVERY trial, has been shown to reduce mortality in hospitalized patients requiring supplemental oxygen or mechanical ventilation. Remdesivir can shorten recovery time, especially in severe cases, but its impact on mortality reduction remains inconclusive. Combining the immunomodulatory drug baricitinib with remdesivir has demonstrated faster recovery and improved outcomes in hospitalized patients, particularly those needing supplemental oxygen or high-flow therapy. Regarding ventilation, the choice between invasive and non-invasive ventilation plays a critical role in the outcome of patients with COVID-19 in ICUs. It is typically employed in patients with severe respiratory failure or acute respiratory distress syndrome (ARDS). In contrast, non-invasive ventilation provides respiratory support through a mask or nasal prongs without the need for intubation, and although both ventilation strategies aim to support breathing, invasive ventilation is associated with higher levels of respiratory support and is often used in more critically ill patients. The choice of ventilation mode can significantly impact patient outcomes, with invasive ventilation generally being associated with higher mortality rates compared to non-invasive ventilation. However, the decision regarding the appropriate ventilation strategy should be individualized, taking into account factors such as disease severity, patient characteristics, and careful assessment of risks and benefits.

Our cohort had prevalent conditions such as type 2 diabetes, metabolic syndrome, or cardiovascular disease. These conditions, along with advanced age, have been associated with worse outcomes in individuals infected with SARS-CoV-2. Older patients with pre-existing conditions are particularly vulnerable, as age can weaken the immune system and make individuals more susceptible to severe illness. Moreover, comorbidities can further increase the severity of COVID-19 symptoms, contribute to a higher risk of complications, and lead to a higher mortality rate. These underlying health conditions and the aging process can exacerbate the inflammatory response triggered by the virus, resulting in complications such as acute respiratory distress syndrome and multiorgan dysfunction, which can explain the high mortality among the elderly in our cohort.

The first wave was associated with the initial outbreak and was restrained by strict public health measures, such as confinement and lockdown. The second wave began in the summer of 2020, when those restrictions ended and social distancing measures were relaxed. This wave reached a peak in the autumn of 2020, probably because of the return to work and school. The third peak began in December 2020, probably as a result of holiday events and Christmas gatherings, and continued until January 2021. The next waves showed a rapid fall in hospitalizations, probably because of vaccination, and its peak might have been associated with the Easter holidays. Since then, waves showed the beneficial effect of vaccination. Regardless of vaccination, it seems that waves and peaks were related to social events: holidays, gatherings, and the relaxation of public health measures such as social distancing.

Vaccination began in the European Union (and also in Spain) in December 2020, and its protective effects can explain the lower admissions and lower mortality among the elderly since [43,44]. Beginning in April 2021, there was a rapid decline in admissions due to COVID-19. Although we are aware that interpretation can be challenging, some authors demonstrated the beneficial impact of vaccination [45]. The use of vaccines resulted in a decrease in hospitalizations, probably because their protection was based on achieving a mild clinical presentation of COVID-19. They had a great impact in terms of hospital admissions and mortality. As mentioned, the effect of vaccination since the summer of 2021 was studied by Barandalla et al. [45], who developed simulated curves of hospitalizations in the absence of vaccines and then compared those curves with the real incidence. By showing the decrease in incidence, they demonstrated the beneficial impact of the vaccination rollout on hospitalizations. With our study, we hypothesize that this impact was steady in the fourth and fifth waves; that is, the steady vaccination of the elderly consolidated the decline in admissions due to COVID-19. The authors state that new COVID-19 hospitalizations occurred in younger, non-vaccinated individuals. They demonstrated that the elderly were not the most frequently hospitalized group since mid-May 2021, but individuals <50 not yet vaccinated.

An interesting point arises regarding vaccination and the subsequent drop in mortality among elderly individuals. Some recent publications have focused on choosing the best vaccination strategy based on certain criteria [46,47]. A protocol should be applied depending on the population structure when the aim is to prevent the spread of an illness, limit the number of deaths, and reduce the impact on health care. In those studies, the authors focus on priority based on population structure (i.e., first vaccinating elderly individuals, who accounted for the most vulnerable group in Spain). They demonstrate that different disease characteristics and different population structures may play an important role in the choice of certain vaccination protocols. In the case of Spain, a country included in one of the publications [46], vaccinating the elderly resulted in a reduction in overall mortality and was probably responsible for the increased number of confirmed cases and admissions among the unvaccinated, younger population.

In addition, changes in SARS-CoV-2 variants can explain the beginnings of several waves and the different patient profiles. It is beyond the scope of this research to provide extensive background on the variants or different viral lineages of SARS-CoV-2 [48,49]. Nevertheless, we can hypothesize that some waves were related to changes in the predominant variant. Several strains were detected in Spain during the first wave [50], probably due to several genetic variants. By September 2020, a new variant (B.1.1.7, also known as the alpha variant) had been described in Europe and was spreading rapidly in several countries [51]. This variant, which had increased transmissibility, virulence, and lethality, may have been responsible for several waves up to summer 2021 (i.e., up to the fifth wave). However, the spread of the alpha variant coincided with the beginning of vaccination in Europe. We hypothesize that, thanks to the vaccines, both admissions and deaths dropped from the second wave to the fifth wave. In August 2021, the delta variant (B.1.617.2) replaced alpha as the predominant variant in Spain [48,52], affecting a younger, unvaccinated population in the fifth wave (summer 2021). This may explain why hospitalization was more frequent among young people, as we mentioned previously (with a median age of 47, according to Table 1; see also Figure A3).

As we also mentioned earlier, changes in variants played a key role in the distribution of the pandemic. The omicron variant and its descendants (B.1.1.529) were identified in November 2021 [53]. This variant was more transmissible than previous variants, and by January 2022 (sixth wave: from November 2021 to March 2022) it was predominant in Spain [46,47]. This variant affected the population even if they were vaccinated because it was more infectious and could evade their immune responses. Specifically, with regard to the sixth wave, we found more admissions and an older hospitalized population than in the fourth and fifth waves. We also found that mortality increased slightly with respect to the previous waves. We hypothesize that this is because of both the omicron variant and the older age of patients. Fortunately, the number of admissions and deaths was lower than in the first wave, which highlights the effectiveness of widespread vaccination [54]. Although it should be properly analyzed in further studies, we hypothesize that the omicron variant was responsible for the sixth wave.

In terms of impact on health care, the semi-logarithmic-scale charts in Figure A4 are worthy of discussion. In the first wave, the initial outbreak phase, SARS-CoV-2 spread exponentially rather than arithmetically, so a log scale is the natural way to track the spread. As we have shown, vertical distances represent multiplicative differences (i.e., 10^0, 10^1, 10^2, 10^3, etc.). Cases increased by 10 every few days. The chart emphasizes the growth and progression of the outbreak in the first wave and its impact on health care in our hospital in terms of admissions and mortality. As mentioned earlier, almost one-third of all admissions and deaths in the observation period occurred during the first weeks of the pandemic. This emphasizes the great impact of the pandemic on daily work. The growth is less pronounced over time (i.e., the curve flattens).

The impact of the COVID-19 pandemic on health care systems in Spain has also been studied up to the third wave, before vaccination, in December 2020 [55]. Analyses of the first, second, and third waves revealed several differences, such as a rise in the number of confirmed cases in the general population due to the less restrictive testing policy (the use of rapid antigen tests), a lower number of severe cases requiring admission to the ICU, and decreased mortality rates. The lower mortality and smaller number of patients requiring ICU admission were studied by Taboada et al. [56], who proposed that corticosteroids and new immunomodulatory drugs were responsible for this phenomenon. In international studies, it was also found that demographic and clinical features of patients with confirmed COVID-19 differed between the third wave and previous waves [57–59].

4.1. Epidemiological Modeling

Regarding the distribution and patterns of the time series, it is possible to rely on hospital admissions to fit predictive models rather than on confirmed cases in the general population. Although the fitting of predictive models is beyond the scope of this research, it is worth noting that analyses of time series based on confirmed cases in the general population may not represent the real state of the pandemic [60]. The number of confirmed cases depends on the testing policy: the more tests performed, the higher the incidence in the general population and the lower the hospitalization rate. For example, according to official data (Figure A6, see data source in [34]), there was a decoupling between the first wave in the area near our hospital and the first wave in terms of admissions, ICU occupancy, and mortality, mainly because of the restrictive testing policy (only patients strongly suspected of having COVID-19 were tested). We hypothesize that the reason why the first wave is decoupled with respect to admissions is because of such a testing strategy. In contrast, it can be observed in Figure A3 that the rest of the lines are coupled from the second to the fifth waves in terms of admissions, occupancy, and mortality, probably because the testing policy was less restrictive [61,62]. A similar phenomenon can be observed in the sixth wave, as a decoupling between Figures A3 and A6 can be seen. There were more than 5000 confirmed cases among our population, but we recorded only 513 hospitalized patients. We hypothesize that a much more permissive testing policy allowed the detection of even mild cases of COVID-19 that did not require hospitalization. We would like to emphasize the role of the testing policy, which probably was a key factor in detecting confirmed cases of COVID-19. In the first waves, the testing policy was very restrictive (not extensive), and the peaks of the time series were decoupled with respect to the curve of the admissions in the same period. In contrast, in the third, fourth, and fifth waves admissions and hospital/ICU occupancy matched, as expected. Finally, the sixth wave in the area nearby showed a higher peak than expected, probably due to extensive testing.

It is worth mentioning an epidemiological study by Red Nacional de Vigilancia Epidemiológica (Epidemiological Surveillance National System (RENAVE in Spanish)) with analyses of data very similar to ours [32]. Confirmed cases, admissions, and deaths were analyzed up to 10 May 2022, with different visualizations, focused specifically on analyses of age ranges. In that research, data from Spain were reported. The authors split the outbreak into periods and established a turning point for each wave based on the 14-day cumulative incidence. The waves in that research are coupled with those in our study.

4.2. Public Health Measures

A question that may arise in the future is what relevant public health measures should be taken by countries. The results, the distribution of waves, and the impact of the pandemic analyzed here were based on several variables, such as population structure, the health care system, testing policies, social interventions, non-pharmacological measures, and vaccination strategies. Spain adopted a suppression strategy as an immediate response to the pandemic based on the aforementioned variables. The aim of this strategy was to reduce the spread of the virus and mortality. In contrast, Barat et al. [63] described the interesting approach adopted by Sweden. Instead of a suppression strategy, Sweden chose a mitigation policy based on its own priorities and legal system. This mitigation strategy was applied in the first wave (March–April 2020) and consisted of risk-tailored measures to protect elderly individuals. In choosing a mitigation policy, Swedish authorities were trying to avoid the potential socioeconomic inequities that are often associated with the massive lockdowns seen with suppression strategies.

Regarding outcomes, Sweden made specific recommendations to prioritize the protection of elderly individuals. Seroprevalence among the elderly was lower in Sweden than in Spain [64]. Mortality in Sweden was higher than in other Scandinavian countries but lower than in other European countries [63]. It is beyond the scope of our study to discuss whether countries should have adopted a mitigation strategy (such as Sweden) or a suppressive one (such as Spain), and the ability to make comparisons is limited by affected populations, testing policies, the timeline of the pandemic, and socioeconomic determinants. Other countries, such as Denmark and Norway, pursued a suppression strategy [65]. Outside Europe and North America, public health strategies in six Asian countries were also analyzed [66]. It is the responsibility of national experts to assess the different approaches and dominant ideas on public health measures regarding the pandemic.

Concerning testing policy, a recent publication by Zhang et al. [67] demonstrated that mass testing was associated with 25% cut admissions due to COVID-19. The city of Liverpool (United Kingdom) was selected for a pilot study. The intervention was to test asymptomatic individuals to identify infected people in order to protect vulnerable individuals, to quarantine the contacts, and ultimately to improve public health. This intervention reduced COVID-19-related admissions because promoting effective isolation of confirmed individuals and their contacts resulted in reduced onward transmission. The study estimated a 32% reduction in admissions compared with the expected admissions with no intervention.

Finally, an intriguing study highlights a recurring pattern of panic and neglect in funding pandemic preparedness. Resources tend to increase in the aftermath of crises but subsequently decline. The high economic costs incurred by the COVID-19 pandemic further highlight the urgent need for investment in preparedness. Estimates for the annual funding required for pandemic preparedness vary, but they remain relatively small compared to the projected costs associated with events like COVID-19. Sustainable funding for pandemic preparedness necessitates effective collaboration between global health stakeholders and national health system leaders, as demonstrated by the importance of timely health responses when political commitment is present [68].

4.3. Limitations

Our aim was to describe trends and distributions of the pandemic in our hospital and to determine the impact on our health care system. Consequently, our research focused on data on hospitalization, not on the total number of confirmed patients in our region.

The hospitalization rate (i.e., the proportion of admissions among confirmed cases) varies over time, among countries, and even among regions within the same country, depending on the testing policy [32,61]. Therefore, we propose that widespread testing would improve estimates of the true admissions rate [60,62]. As mentioned previously, when interpreting the impact of COVID-19 on hospitals, researchers cannot rely solely on confirmed cases, because sometimes it is difficult to know what proportion of the population has been tested. Likewise, it can be difficult to estimate the mortality rate among the infected population. Another limitation is the local design of the study, as we are aware that every country, and even every region within a single country, has had different waves of COVID-19. Therefore, the distribution of confirmed cases varies. Given these regional differences, we have described the experiences and the distribution of the pandemic at our own institution. Our splitting of the pandemic into six waves is utterly idiosyncratic and cannot be extrapolated to other settings. Finally, it is important to note that this study is limited by the absence of socioeconomic information, which could have provided valuable insights into the potential influence of socioeconomic factors on disease severity, access to health care resources, and treatment outcomes. In addition, we were unable to examine the impact of physical activity as a potential factor influencing the clinical progression of COVID-19. However, an intriguing study involving 131 individuals demonstrated that patients with sufficient or high levels of physical activity were more likely to experience recovery, whereas those with insufficient activity had an increased risk of death. This suggests a correlation between physical activity and a less severe course of the disease [69].

Some previous publications have conducted studies on the effects of COVID-19, focusing on specific secondary hospitals or limited time periods. These studies provided information about the demographic characteristics and outcomes of the patients included in their respective cohorts [70–73]. Although we recognize that there may be regional variations in the characteristics of COVID-19, and we acknowledge that our findings may not directly apply to other settings, it is important to note that our study encompassed a period of 27 months during the pandemic and involved over 3000 hospitalized patients. However, the strength of this study lies in its meticulous emphasis on complete follow-up for all patients throughout their entire hospital stay which enhances the reliability and validity of the results. This comprehensive monitoring allows for a detailed understanding of the disease progression, treatment outcomes, and potential complications, providing valuable insights for medical practitioners and researchers alike. This approach ensures a representative sample, minimizing potential biases and increasing the generalization of the findings.

5. Conclusions

In this study, our objective was to present a comprehensive overview of trends and patterns observed during various stages of the COVID-19 pandemic. We aimed to highlight differences in demographic data, clinical information, and healthcare indicators. Throughout the course of the pandemic, we witnessed advancements in patient management, such as the development of new drugs and the rollout of vaccination programs. However, we also observed the emergence of different variants and lineages, which ultimately had a significant impact on hospitalization rates and mortality trends. When formulating non-medical strategies to address future waves of COVID-19 or outbreaks of new infectious diseases, it is crucial to consider these factors, in addition to the clinical experience gained during the pandemic. To gain insights into the diverse patient profiles, we analyzed administrative and demographic data across different waves of the pandemic. By dividing the pandemic into distinct waves, we were able to identify variations in patient demographics, likely influenced by key milestones preceding each wave. These milestones included factors such as the initiation of vaccination programs, changes in SARS-CoV-2 variants, and the implementation of social distancing measures. Our findings suggest that analyzing the pandemic as a whole may not capture the complete picture, and it is more informative to examine individual waves. Factors specific to each wave, such as the timing of COVID-19 infections, vaccination rates, and predominant virus variants, should be considered when designing future clinical studies on the pandemic. As a result, our research provides a general analytical framework that can be applied to other settings.

Author Contributions: R.G.-C. conceived and designed the study, wrote the first draft of the manuscript, and preprocessed and analyzed the data. O.V.-G. and M.O.-G. made substantial contributions to the interpretation of the results, critically reviewed the first draft of the manuscript, and made valuable suggestions. N.G.-P. collected and processed the database and contributed to the visualization of the data and time series processing. R.G.-P. and A.G.-d.-M. supervised the project and critically reviewed and edited the final draft of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was approved by the Ethical Board of Mostoles University Hospital (CEIC 2020/025). No identifying information was included in the manuscript. Because the authors used historical data, informed consent was not necessary. All procedures involving human participants were conducted in accordance with the ethical standards of the responsible institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Data Availability Statement: A contract signed with Mostoles University Hospital, which provided the dataset, prohibits the authors from providing their data to any other researcher. Furthermore, the authors must destroy the data upon the conclusion of their investigation. The data cannot be uploaded to any public repository.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
COVID-19	Coronavirus disease 2019
ICUs	Intensive care units
RENAVE	Red Nacional de Vigilancia Epidemiologica (Epidemiological Surveillance National System)

Appendix A

The appendix contains details and data supplemental to the main text, mainly figures we have produced to better understand the pandemic. We believe these figures can be representative of the impact of the pandemic on health care.

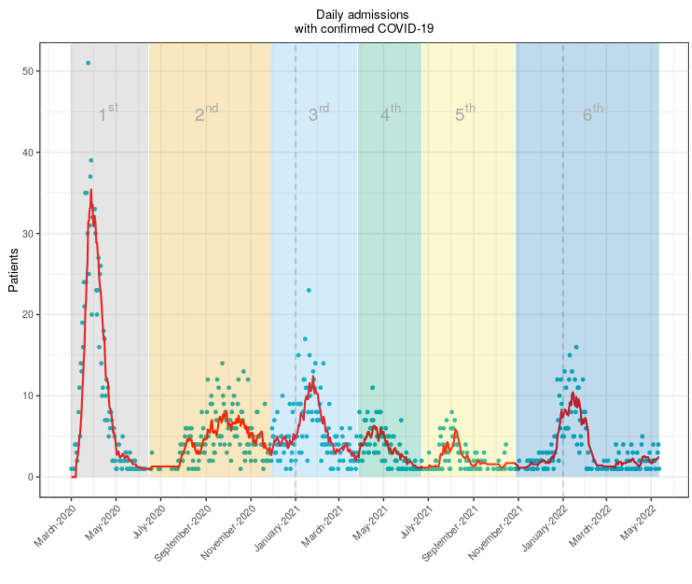

Figure A1. The six waves, according to peaks and valleys calculated inside our time series.

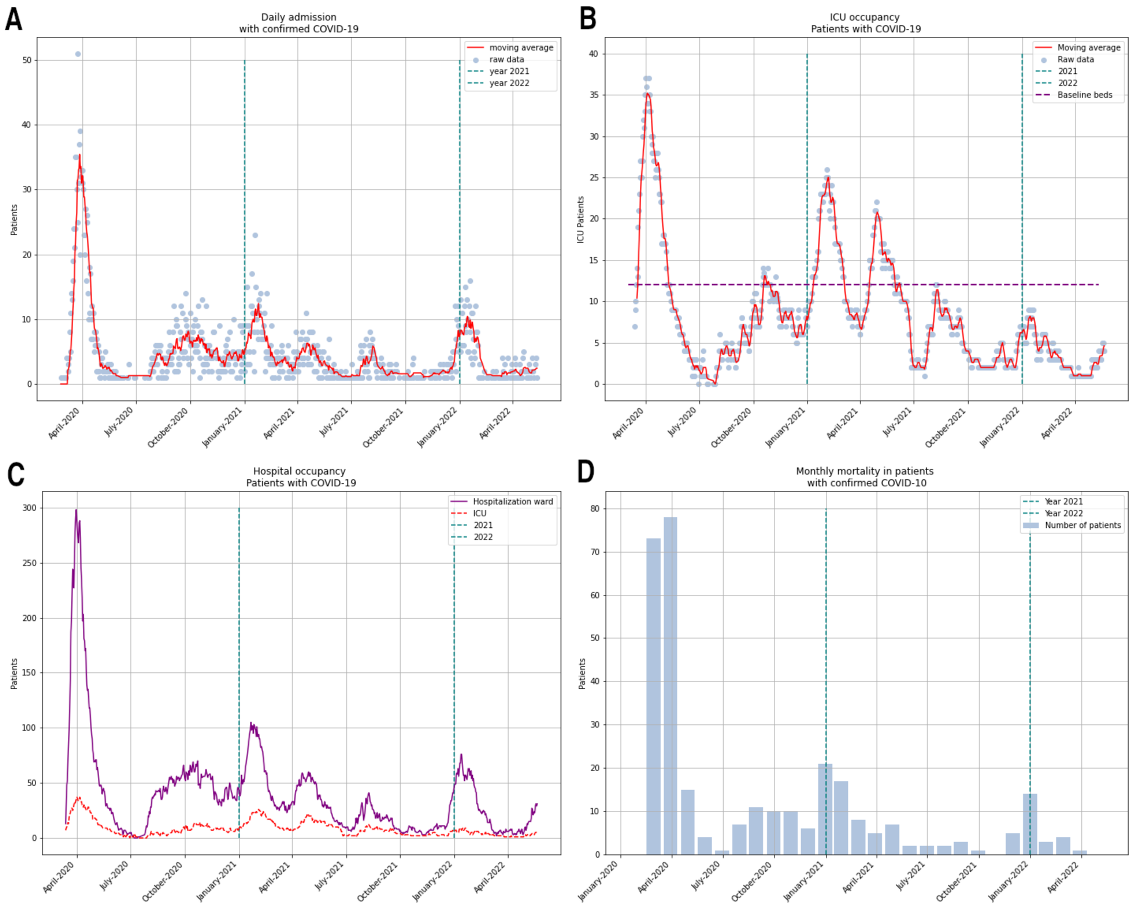

Figure A2. Evolution of daily admission (**A**), ICU/hospitalization ward occupancy (**B**,**C**), and mortality (**D**). The four graphs are coupled, which shows the concordance in the evolution of the pandemic during the observation period. Plots of daily admissions (**A**) and ICU occupancy (**B**) show both blue dots (representing raw data) and a red line (moving average). Moving average lines allowed us to identify trends, softened the curve, and avoided the creation of sharp lines. Sharp lines are usually the result of local daily fluctuations and can be misleading.

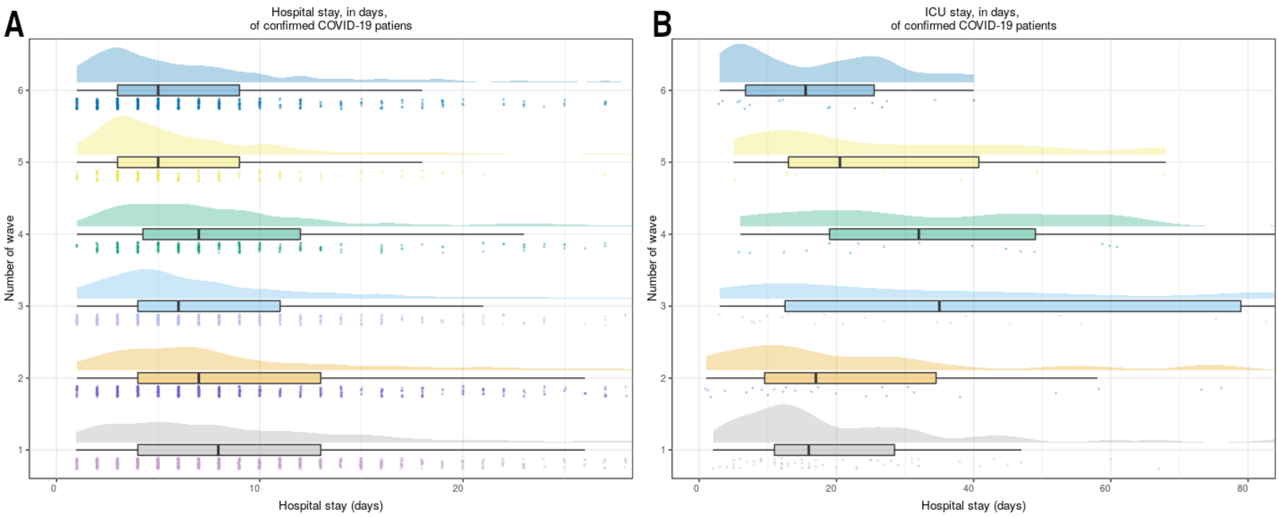

Figure A3. *Cont.*

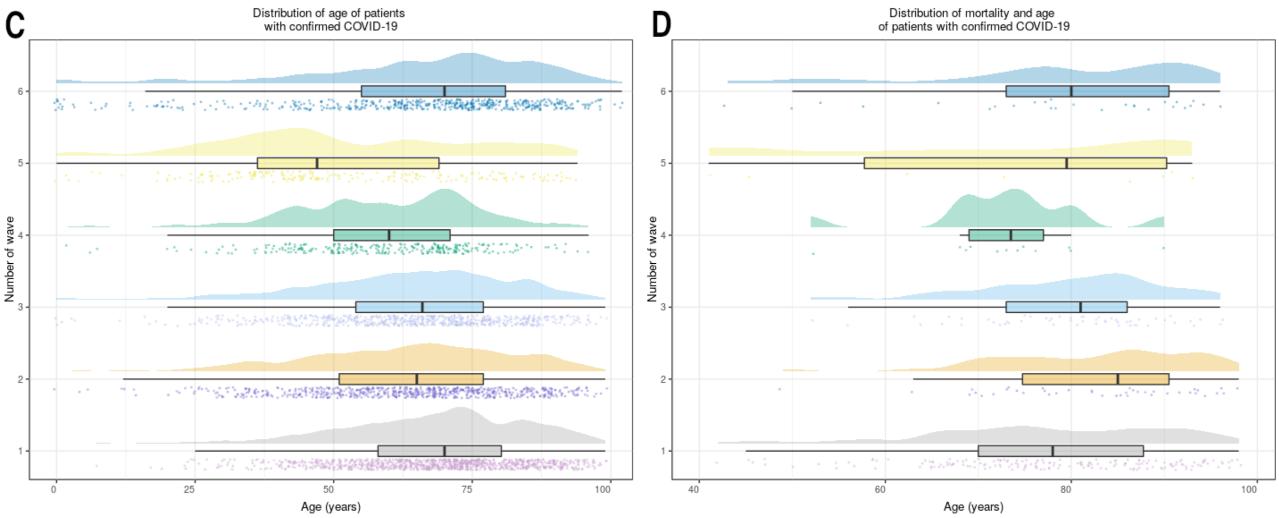

Figure A3. Rainclouds showing the distribution along the different waves of hospital stays of patients admitted to our hospital with confirmed COVID-19 (**A**), hospital stays of patients admitted to the ICU (**B**), patient age (**C**), and the number of deaths during the study period (**D**).

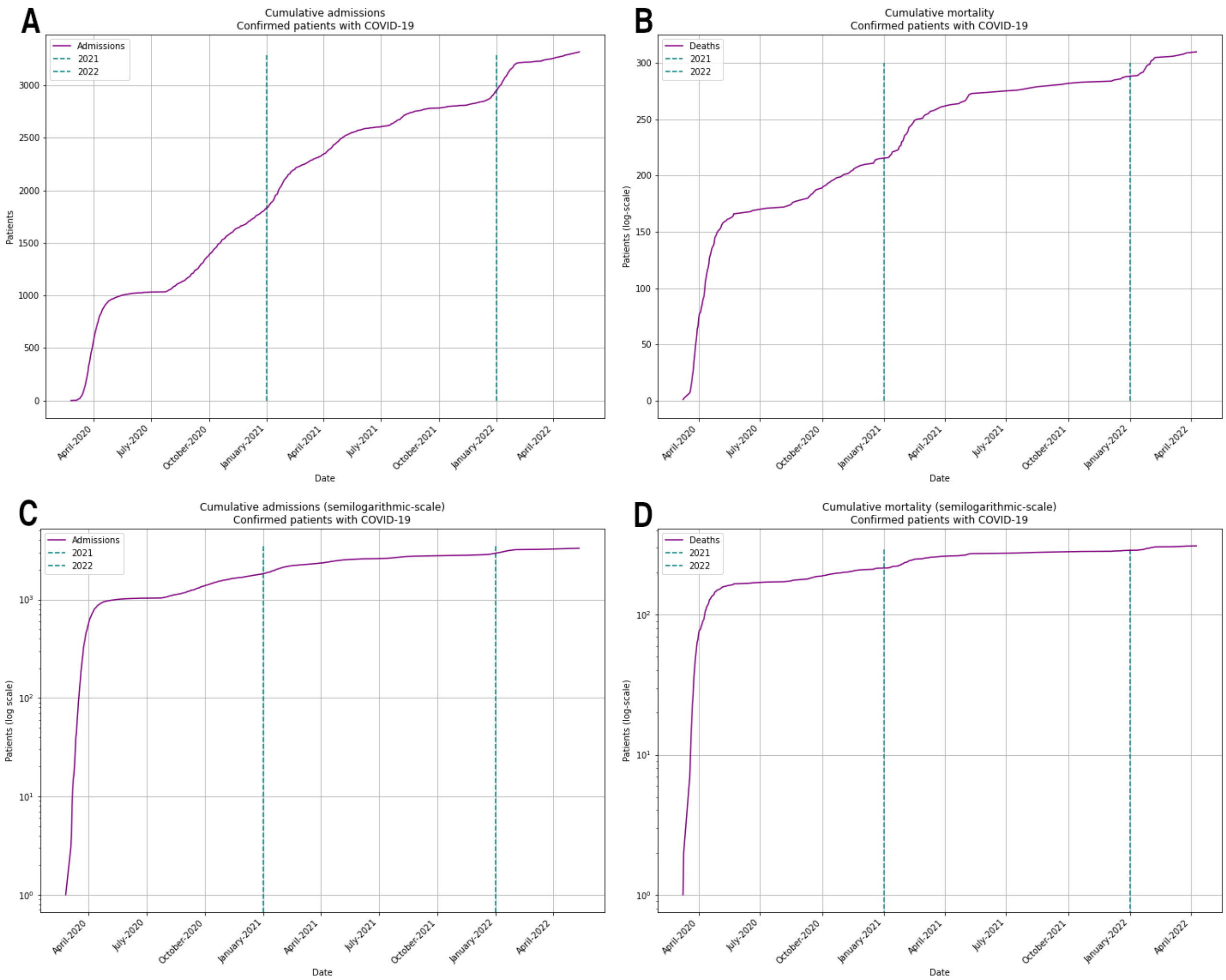

Figure A4. The cumulative sum of daily admissions and mortality. Arithmetic-scale plots (**A**,**B**) are self-explanatory. Semi-logarithmic-scale plots (**C**,**D**) highlight the impact on health care: the almost vertical initial trend shows that one-third of all admissions and all deaths occurred in the first 6 weeks of the pandemic.

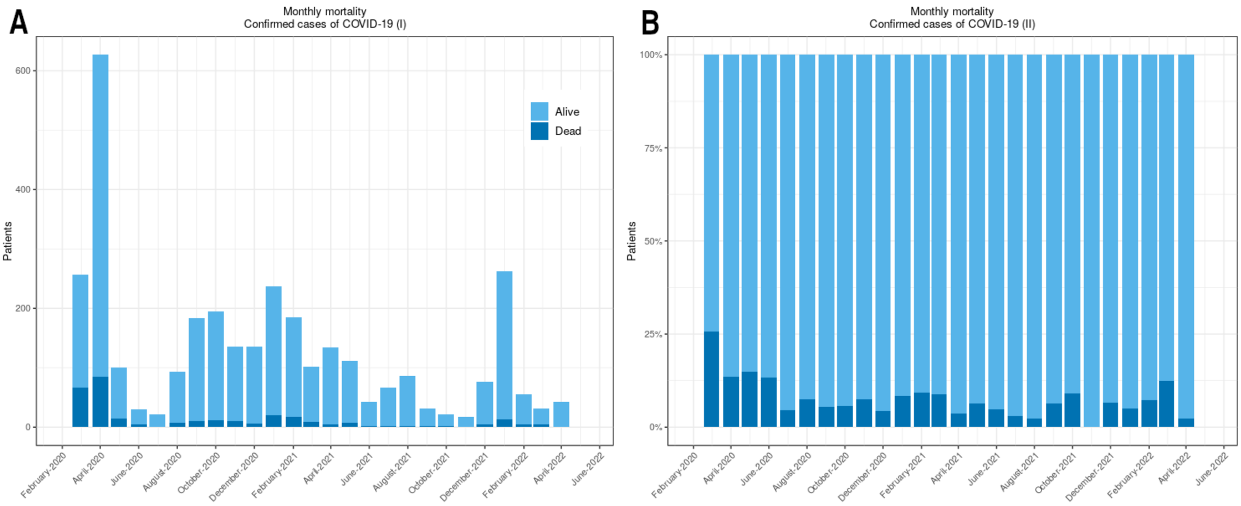

Figure A5. Discharge status over time (alive vs. dead) in absolute values (**A**) and percentages (**B**).

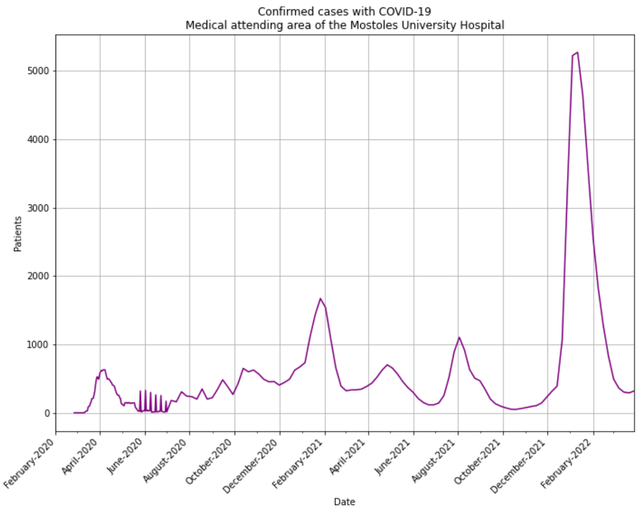

Figure A6. Confirmed cases of COVID-19 in the general population in the area near our hospital, which has a population of 168,000. The first wave is decoupled with respect to the first wave of admissions.

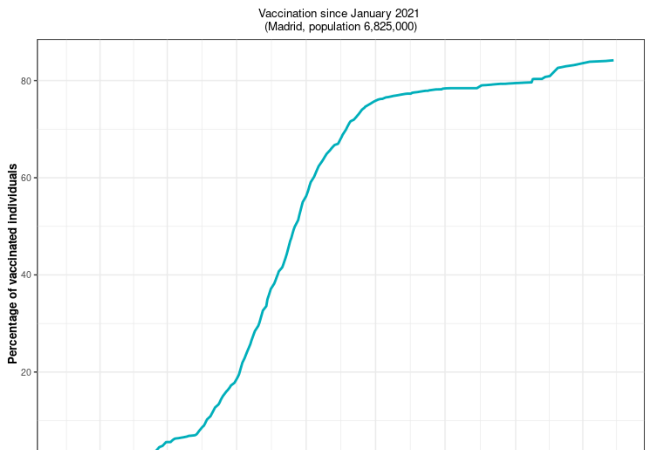

Figure A7. Ratio of vaccinated population in the region of Madrid with complete vaccination schedule.

References

1. Lovato, A.; De Filippis, C. Clinical presentation of COVID-19: A systematic review focusing on upper airway symptoms. *Ear Nose Throat J.* **2020**, *99*, 569–576. [CrossRef]
2. Gorbalenya, A.E.; Baker, S.C.; Baric, R.S.; de Groot, R.J.; Drosten, C.; Gulyaeva, A.A.; Haagmans, B.L.; Lauber, C.; Leontovich, A.M.; Neuman, B.W.; et al. The species Severe acute respiratory syndrome-related coronavirus: Classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.* **2020**, *5*, 536–544.
3. WHO Director-General's Opening Remarks at the Media Briefing on COVID-19—11 March 2020—who.int. Available online: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-COVID-19---11-march-2020 (accessed on 28 June 2023).
4. Statement on the Second Meeting of the International Health Regulations (2005) Emergency Committee Regarding the Outbreak of Novel Coronavirus (2019-nCoV)—who.int. Available online: https://www.who.int/news/item/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov) (accessed on 28 June 2023).
5. Machhi, J.; Herskovitz, J.; Senan, A.M.; Dutta, D.; Nath, B.; Oleynikov, M.D.; Blomberg, W.R.; Meigs, D.D.; Hasan, M.; Patel, M.; et al. The natural history, pathobiology, and clinical manifestations of SARS-CoV-2 infections. *J. Neuroimmune Pharmacol.* **2020**, *15*, 359–386. [CrossRef]
6. Van Doremalen, N.; Bushmaker, T.; Morris, D.H.; Holbrook, M.G.; Gamble, A.; Williamson, B.N.; Tamin, A.; Harcourt, J.L.; Thornburg, N.J.; Gerber, S.I.; et al. Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. *N. Engl. J. Med.* **2020**, *382*, 1564–1567. [CrossRef]
7. Chan, J.F.W.; Yuan, S.; Kok, K.H.; To, K.K.W.; Chu, H.; Yang, J.; Xing, F.; Liu, J.; Yip, C.C.Y.; Poon, R.W.S.; et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: A study of a family cluster. *Lancet* **2020**, *395*, 514–523. [CrossRef] [PubMed]
8. Pustake, M.; Tambolkar, I.; Giri, P.; Gandhi, C. SARS, MERS and COVID-19: An overview and comparison of clinical, laboratory and radiological features. *J. Fam. Med. Prim. Care* **2022**, *11*, 10. [CrossRef]
9. Zhu, Z.; Lian, X.; Su, X.; Wu, W.; Marraro, G.A.; Zeng, Y. From SARS and MERS to COVID-19: A brief summary and comparison of severe acute respiratory infections caused by three highly pathogenic human coronaviruses. *Respir. Res.* **2020**, *21*, 224. [CrossRef] [PubMed]
10. Onder, G.; Rezza, G.; Brusaferro, S. Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. *JAMA* **2020**, *323*, 1775–1776. [CrossRef]
11. Lazzerini, M.; Putoto, G. COVID-19 in Italy: Momentous decisions and many uncertainties. *Lancet Glob. Health* **2020**, *8*, e641–e642. [CrossRef]
12. Dowd, J.B.; Andriano, L.; Brazel, D.M.; Rotondi, V.; Block, P.; Ding, X.; Liu, Y.; Mills, M.C. Demographic science aids in understanding the spread and fatality rates of COVID-19. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 9696–9698. [CrossRef]
13. Bignami-Van Assche, S.; Ghio, D. A demographic adjustment to improve measurement of COVID-19 severity at the developing stage of the pandemic. *medRxiv* **2020**, *27*, e03. [CrossRef]
14. Ghio, D.; Acosta, E.; Fisman, D.; Noymer, A.; Stilianakis, N.I.; Assche, S.B.V. Population health and COVID-19 in Canada: A demographic comparative perspective. *Can. Stud. Popul.* **2021**, *48*, 131–137. [CrossRef]
15. Aburto, J.M.; Schöley, J.; Kashnitsky, I.; Zhang, L.; Rahal, C.; Missov, T.I.; Mills, M.C.; Dowd, J.B.; Kashyap, R. Quantifying impacts of the COVID-19 pandemic through life-expectancy losses: A population-level study of 29 countries. *Int. J. Epidemiol.* **2022**, *51*, 63–74. [CrossRef] [PubMed]
16. Kashnitsky, I.; Aburto, J.M. COVID-19 in unequally ageing European regions. *World Dev.* **2020**, *136*, 105170. [CrossRef] [PubMed]
17. Jamil, Z.; Khan, A.A.; Yousuf, H.; Khalid, K.; Abbasi, S.M.; Waheed, Y. Role of therapeutic plasmapheresis in SARS-CoV-2 induced cytokine release syndrome: A retrospective cohort study on COVID-19 patients. *Int. J. Gen. Med.* **2022**, *15*, 4907–4916. [CrossRef]
18. Jamil, Z.; Khan, A.A.; Khalid, S.; Asghar, M.; Muhammad, K.; Waheed, Y. Beneficial effects of anticoagulants on the clinical outcomes of COVID-19 patients. *Antibiotics* **2021**, *10*, 1394. [CrossRef]
19. Kyriazopoulou, E.; Huet, T.; Cavalli, G.; Gori, A.; Kyprianou, M.; Pickkers, P.; Eugen-Olsen, J.; Clerici, M.; Veas, F.; Chatellier, G.; et al. Effect of anakinra on mortality in patients with COVID-19: A systematic review and patient-level meta-analysis. *Lancet Rheumatol.* **2021**, *3*, e690–e697. [CrossRef] [PubMed]
20. Kalil, A.C.; Patterson, T.F.; Mehta, A.K.; Tomashek, K.M.; Wolfe, C.R.; Ghazaryan, V.; Marconi, V.C.; Ruiz-Palacios, G.M.; Hsieh, L.; Kline, S.; et al. Baricitinib plus remdesivir for hospitalized adults with COVID-19. *N. Engl. J. Med.* **2021**, *384*, 795–807. [CrossRef]
21. Wolfe, C.R.; Tomashek, K.M.; Patterson, T.F.; Gomez, C.A.; Marconi, V.C.; Jain, M.K.; Yang, O.O.; Paules, C.I.; Palacios, G.M.R.; Grossberg, R.; et al. Baricitinib versus dexamethasone for adults hospitalised with COVID-19 (ACTT-4): A randomised, double-blind, double placebo-controlled trial. *Lancet Respir. Med.* **2022**, *10*, 888–899. [CrossRef]
22. Marconi, V.C.; Ramanan, A.V.; de Bono, S.; Kartman, C.E.; Krishnan, V.; Liao, R.; Piruzeli, M.L.B.; Goldman, J.D.; Alatorre-Alexander, J.; de Cassia Pellegrini, R.; et al. Efficacy and safety of baricitinib for the treatment of hospitalised adults with COVID-19 (COV-BARRIER): A randomised, double-blind, parallel-group, placebo-controlled phase 3 trial. *Lancet Respir. Med.* **2021**, *9*, 1407–1418. [CrossRef]
23. Rosas, I.O.; Bräu, N.; Waters, M.; Go, R.C.; Hunter, B.D.; Bhagani, S.; Skiest, D.; Aziz, M.S.; Cooper, N.; Douglas, I.S.; et al. Tocilizumab in hospitalized patients with severe COVID-19 pneumonia. *N. Engl. J. Med.* **2021**, *384*, 1503–1516. [CrossRef]

24. Salama, C.; Han, J.; Yau, L.; Reiss, W.G.; Kramer, B.; Neidhart, J.D.; Criner, G.J.; Kaplan-Lewis, E.; Baden, R.; Pandit, L.; et al. Tocilizumab in patients hospitalized with COVID-19 pneumonia. *N. Engl. J. Med.* **2021**, *384*, 20–30. [CrossRef]
25. Casas-Rojo, J.; Antón-Santos, J.; Millán-Núñez-Cortés, J.; Lumbreras-Bermejo, C.; Ramos-Rincón, J.; Roy-Vallejo, E.; Artero-Mora, A.; Arnalich-Fernández, F.; García-Bruñén, J.; Vargas-Núñez, J.; et al. Clinical characteristics of patients hospitalized with COVID-19 in Spain: Results from the SEMI-COVID-19 Registry. *Rev. Clínica Española (Engl. Ed.)* **2020**, *220*, 480–494. [CrossRef]
26. Vidal-Cortés, P.; Martín, M.C.; Díaz, E.; Bodí, M.; Igeño, J.C.; Garnacho-Montero, J. Impact of one year of pandemic on Spanish Intensive Care Units. 2022. *Rev. Esp. Quimioter.* **2022**, *35*, 392–400. [CrossRef]
27. Pastor-Barriuso, R.; Perez-Gomez, B.; Hernan, M.A.; Perez-Olmeda, M.; Yotti, R.; Oteo-Iglesias, J.; Sanmartín, J.L.; Leon-Gomez, I.; Fernandez-Garcia, A.; Fernández-Navarro, P.; et al. Infection fatality risk for SARS-CoV-2 in community dwelling population of Spain: Nationwide seroepidemiological study. *BMJ* **2020**, *371*, m4509. [CrossRef] [PubMed]
28. Dahal, S.; Mizumoto, K.; Rothenberg, R.; Chowell, G. Investigating spatial variability in COVID-19 pandemic severity across 19 geographic areas, Spain, 2020. *medRxiv* **2020**, *17*. [CrossRef]
29. BOE-A-2020-3692 Real Decreto 463/2020, de 14 de Marzo, por el que se Declara el Estado de Alarma para la Qestión de la Situación de Crisis Sanitaria Ocasionada por el COVID-19—boe.es. Available online: https://www.boe.es/eli/es/rd/2020/03/14/463 (accessed on 29 June 2023).
30. Briz-Redón, Á.; Serrano-Aroca, Á. A spatio-temporal analysis for exploring the effect of temperature on COVID-19 early evolution in Spain. *Sci. Total Environ.* **2020**, *728*, 138811. [CrossRef]
31. Frias, L. ISCIII CNE-RENAVE—cnecovid.isciii.es. Available online: https://cnecovid.isciii.es/ (accessed on 29 June 2023).
32. Informes COVID-19—isciii.es. Available online: https://www.isciii.es/QueHacemos/Servicios/VigilanciaSaludPublicaRENAVE/EnfermedadesTransmisibles/Paginas/InformesCOVID-19.aspx (accessed on 29 June 2023).
33. Epi, A. The Epidemiologist R Handbook—Appliedepi.org. Available online: https://appliedepi.org/epirhandbook/ (accessed on 29 June 2023).
34. Ministerio de Sanidad-Ciudadanos-Prestaciones y centro sanitarios-Centros y servicios del SNS-Instituto de Informacion Sanitaria-Centros y servicios del SNS-Hospitales—sanidad.gob.es. Available online: https://www.sanidad.gob.es/en/ciudadanos/centrosCA.do (accessed on 29 June 2023).
35. Agresti, A. *An Introduction to Categorical Data Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2018; Volume 656.
36. Howell, D.C. *Statistical Methods for Psychology*; Cengage Learning: Belmont, CA, USA, 2012.
37. Allen, M.; Poggiali, D.; Whitaker, K.; Marshall, T.R.; Kievit, R.A. Raincloud plots: A multi-platform tool for robust data visualization. *Wellcome Open Res.* **2019**, *4*, 63. [CrossRef]
38. COVID 19-TIA Zonas Básicas de Salud-Datos Abiertos Comunidad de Madrid—datos.comunidad.madrid. Available online: https://datos.comunidad.madrid/catalogo/dataset/covid19_tia_zonas_basicas_salud (accessed on 29 June 2023).
39. Burki, T. The indirect impact of COVID-19 on women. *Lancet Infect. Dis.* **2020**, *20*, 904–905. [CrossRef] [PubMed]
40. Bwire, G.M. Coronavirus: Why men are more vulnerable to COVID-19 than women? *SN Compr. Clin. Med.* **2020**, *2*, 874–876. [CrossRef] [PubMed]
41. Group, R.C. Dexamethasone in hospitalized patients with COVID-19. *N. Engl. J. Med.* **2021**, *384*, 693–704.
42. Sterne, J.A.; Murthy, S.; Diaz, J.V.; Slutsky, A.S.; Villar, J.; Angus, D.C.; Annane, D.; Azevedo, L.C.P.; Berwanger, O.; Cavalcanti, A.B.; et al. Association between administration of systemic corticosteroids and mortality among critically ill patients with COVID-19: A meta-analysis. *JAMA* **2020**, *324*, 1330–1341.
43. Sentís, A.; Kislaya, I.; Nicolay, N.; Meijerink, H.; Starrfelt, J.; Martínez-Baz, I.; Castilla, J.; Nielsen, K.F.; Hansen, C.H.; Emborg, H.D.; et al. Estimation of COVID-19 vaccine effectiveness against hospitalisation in individuals aged $\geq$ 65 years using electronic health registries; a pilot study in four EU/EEA countries, October 2021 to March 2022. *Eurosurveillance* **2022**, *27*, 2200551. [CrossRef]
44. Monge, S.; Mazagatos, C.; Olmedo, C.; Rojas-Benedicto, A.; Simón, F.; Vega-Piris, L.; Sierra, M.J.; Limia, A.; Larrauri, A. Effectiveness of vaccines against SARS-CoV-2 used in Spain: infection, hospitalization and mortality in people aged fifty/fifty-nine. *Rev. Esp. Salud Publica* **2022**, *96*, e202209060.
45. Barandalla, I.; Alvarez, C.; Barreiro, P.; de Mendoza, C.; González-Crespo, R.; Soriano, V. Impact of scaling up SARS-CoV-2 vaccination on COVID-19 hospitalizations in Spain. *Int. J. Infect. Dis.* **2021**, *112*, 81–88. [CrossRef]
46. Rodriguez-Maroto, G.; Atienza-Diez, I.; Ares, S.; Manrubia, S. Vaccination strategies in structured populations under partial immunity and reinfection. *J. Phys. A Math. Theor.* **2023**, *56*, 204003. [CrossRef]
47. Bubar, K.M.; Reinholt, K.; Kissler, S.M.; Lipsitch, M.; Cobey, S.; Grad, Y.H.; Larremore, D.B. Model-informed COVID-19 vaccine prioritization strategies by age and serostatus. *Science* **2021**, *371*, 916–921. [CrossRef] [PubMed]
48. Tracking SARS-CoV-2 Variants—who.int. Available online: https://www.who.int/activities/tracking-SARS-CoV-2-variants (accessed on 29 June 2023).
49. Thakur, V.; Bhola, S.; Thakur, P.; Patel, S.K.S.; Kulshrestha, S.; Ratho, R.K.; Kumar, P. Waves and variants of SARS-CoV-2: Understanding the causes and effect of the COVID-19 catastrophe. *Infection* **2022**, *50*, 309–325. [CrossRef] [PubMed]
50. López, M.G.; Chiner-Oms, Á.; García de Viedma, D.; Ruiz-Rodriguez, P.; Bracho, M.A.; Cancino-Muñoz, I.; D'Auria, G.; de Marco, G.; García-González, N.; Goig, G.A.; et al. The first wave of the COVID-19 epidemic in Spain was associated with early introductions and fast spread of a dominating genetic variant. *Nat. Genet.* **2021**, *53*, 1405–1414. [CrossRef]
51. Hodcroft, E.B.; Zuber, M.; Nadeau, S.; Vaughan, T.G.; Crawford, K.H.; Althaus, C.L.; Reichmuth, M.L.; Bowen, J.E.; Walls, A.C.; Corti, D.; et al. Spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *Nature* **2021**, *595*, 707–712. [CrossRef] [PubMed]

52. Rashedi, R.; Samieefar, N.; Akhlaghdoust, M.; Mashhadi, M.; Darzi, P.; Rezaei, N. Delta variant: The new challenge of COVID-19 pandemic, an overview of epidemiological, clinical, and immune characteristics. *Acta Bio Medica Atenei Parm.* **2022**, *93*, e2022179.
53. Papanikolaou, V.; Chrysovergis, A.; Ragos, V.; Tsiambas, E.; Katsinis, S.; Manoli, A.; Papouliakos, S.; Roukas, D.; Mastronikolis, S.; Peschos, D.; et al. From delta to Omicron: S1-RBD/S2 mutation/deletion equilibrium in SARS-CoV-2 defined variants. *Gene* **2022**, *814*, 146134. [CrossRef]
54. Andrews, N.; Stowe, J.; Kirsebom, F.; Toffa, S.; Rickeard, T.; Gallagher, E.; Gower, C.; Kall, M.; Groves, N.; O'Connell, A.M.; et al. COVID-19 vaccine effectiveness against the Omicron (B. 1.1. 529) variant. *N. Engl. J. Med.* **2022**, *386*, 1532–1546. [CrossRef] [PubMed]
55. Soriano, V.; de Mendoza, C.; Gómez-Gallego, F.; Corral, O.; Barreiro, P. Third wave of COVID-19 in Madrid, Spain. *Int. J. Infect. Dis.* **2021**, *107*, 212–214. [CrossRef]
56. Taboada, M.; González, M.; Alvarez, A.; Eiras, M.; Costa, J.; Álvarez, J.; Seoane-Pillado, T. First, second and third wave of COVID-19. What have we changed in the ICU management of these patients? *J. Infect.* **2021**, *82*, e14–e15. [CrossRef]
57. Tandon, P.; Leibner, E.S.; Hackett, A.; Maguire, K.; Mashriqi, N.; Kohli-Seth, R. The third wave: Comparing seasonal trends in COVID-19 patient data at a large hospital system in New York City. *Crit. Care Explor.* **2022**, *4*, e0653. [CrossRef]
58. Tandon, P.; Leibner, E.; Ahmed, S.; Acquah, S.O.; Kohli-Seth, R. Comparing seasonal trends in coronavirus disease 2019 patient data at a quaternary hospital in New York City. *Crit. Care Explor.* **2021**, *3*, e0381. [CrossRef]
59. Matsunaga, N.; Hayakawa, K.; Asai, Y.; Tsuzuki, S.; Terada, M.; Suzuki, S.; Ohtsu, H.; Kitajima, K.; Toyoda, A.; Suzuki, K.; et al. Clinical characteristics of the first three waves of hospitalised patients with COVID-19 in Japan prior to the widespread use of vaccination: A nationwide observational study. *Lancet Reg. Health-West. Pac.* **2022**, *22*, 100421. [CrossRef]
60. Fox, S.J.; Lachmann, M.; Tec, M.; Pasco, R.; Woody, S.; Du, Z.; Wang, X.; Ingle, T.A.; Javan, E.; Dahan, M.; et al. Real-time pandemic surveillance using hospital admissions and mobility data. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2111870119. [CrossRef]
61. Du, Z.; Wang, L.; Bai, Y.; Wang, X.; Pandey, A.; Fitzpatrick, M.C.; Chinazzi, M.; y Piontti, A.P.; Hupert, N.; Lachmann, M.; et al. Cost-effective proactive testing strategies during COVID-19 mass vaccination: A modelling study. *Lancet Reg. Health-Am.* **2022**, *8*, 100182. [CrossRef]
62. Mercer, T.R.; Salit, M. Testing at scale during the COVID-19 pandemic. *Nat. Rev. Genet.* **2021**, *22*, 415–426. [CrossRef]
63. Baral, S.; Chandler, R.; Prieto, R.G.; Gupta, S.; Mishra, S.; Kulldorff, M. Leveraging epidemiological principles to evaluate Sweden's COVID-19 response. *Ann. Epidemiol.* **2021**, *54*, 21–26. [CrossRef]
64. Pollán, M.; Pérez-Gómez, B.; Pastor-Barriuso, R.; Oteo, J.; Hernán, M.A.; Pérez-Olmeda, M.; Sanmartín, J.L.; Fernández-García, A.; Cruz, I.; de Larrea, N.F.; et al. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): A nationwide, population-based seroepidemiological study. *Lancet* **2020**, *396*, 535–544. [CrossRef] [PubMed]
65. Laage-Thomsen, J.; Frandsen, S.L. Pandemic preparedness systems and diverging COVID-19 responses within similar public health regimes: A comparative study of expert perceptions of pandemic response in Denmark, Norway, and Sweden. *Glob. Health* **2022**, *18*, 3. [CrossRef] [PubMed]
66. De Foo, C.; Verma, M.; Tan, S.M.; Haldane, V.; Reyes, K.A.; Garcia, F.; Canila, C.; Orano, J.; Ballesteros, A.J.; Marthias, T.; et al. COVID-19 public health and social measures: A comprehensive picture of six Asian countries. *BMJ Glob. Health* **2022**, *7*, e009863. [CrossRef] [PubMed]
67. Zhang, X.; Barr, B.; Green, M.; Hughes, D.; Ashton, M.; Charalampopoulos, D.; García-Fiñana, M.; Buchan, I. Impact of community asymptomatic rapid antigen testing on COVID-19 related hospital admissions: Synthetic control study. *BMJ* **2022**, *379*, e071374. [CrossRef] [PubMed]
68. Micah, A.E.; Bhangdia, K.; Cogswell, I.E.; Lasher, D.; Lidral-Porter, B.; Maddison, E.R.; Nguyen, T.N.N.; Patel, N.; Pedroza, P.; Solorio, J.; et al. Global investments in pandemic preparedness and COVID-19: Development assistance and domestic spending on health between 1990 and 2026. *Lancet Glob. Health* **2023**, *11*, e385–e413. [CrossRef]
69. Sutkowska, E.; Stanek, A.; Madziarska, K.; Jakubiak, G.K.; Sokołowski, J.; Madziarski, M.; Sutkowska-Stępień, K.; Biernat, K.; Mazurek, J.; Borowkow-Bulek, A.; et al. Physical Activity Modifies the Severity of COVID-19 in Hospitalized Patients—Observational Study. *J. Clin. Med.* **2023**, *12*, 4046. [CrossRef]
70. Babar, I.; Ekenna, O.; Clarkson, M.R.; Boudreaux, D.; Bennett, W.; Roth, R. COVID-19 in a Mississippi Community Hospital. *South. Med. J.* **2021**, *114*, 305–310. [CrossRef]
71. Abdulah, D.M.; Mirza, A.M.S. Receiving COVID-19 vaccine, hospitalization, and outcomes of patients with COVID-19: A prospective study. *Monaldi Arch. Chest Dis.* **2023**, *93*, e2194. [CrossRef]
72. Darwish, I.; Harrison, L.B.; Passos-Castilho, A.M.; Labbé, A.C.; Barkati, S.; Luong, M.L.; Kong, L.Y.; Tutt-Guérette, M.A.; Kierans, J.; Rousseau, C.; et al. In-hospital outcomes of SARS-CoV-2-infected health care workers in the COVID-19 pandemic first wave, Quebec, Canada. *PLoS ONE* **2022**, *17*, e0272953. [CrossRef] [PubMed]
73. Abeysuriya, V.; Seneviratne, S.L.; De Silva, A.P.; Mowjood, R.; Mowjood, S.; de Silva, T.; de Mel, P.; de Mel, C.; Wijesinha, R.; Fernando, A.; et al. Postdischarge outcomes of COVID-19 patients from South Asia: A prospective study. *Trans. R. Soc. Trop. Med. Hyg.* **2022**, *116*, 1129–1137. [CrossRef] [PubMed]

6 The COVID-19 Pandemic in Spain (2020–2021)

BMC Infectious Diseases

RESEARCH

Open Access

Hospitalization burden and epidemiology of the COVID-19 pandemic in Spain (2020–2021)

Rafael Garcia-Carretero[1*], Oscar Vazquez-Gomez[1], Ruth Gil-Prieto[2] and Angel Gil-de-Miguel[2]

Abstract

Background Spain had some of Europe's highest incidence and mortality rates for coronavirus disease 2019 (COVID-19). Here we describe the epidemiology and trends in hospitalizations, the number of critical patients, and deaths in Spain in 2020 and 2021.

Methods We performed a descriptive, retrospective, nationwide study using an administrative database, the Minimum Basic Data Set at Hospitalization, which includes 95–97% of discharge reports for patients hospitalized in Spain in 2020 and 2021. We analyzed the number of hospitalizations, admissions to intensive care units, and deaths and their geographic distribution across regions of Spain.

Results As of December 31, 2021, a total of 498,789 patients (1.04% of the entire Spanish population) had needed hospitalization. At least six waves of illness were identified. Men were more prone to hospitalization than women. The median age was 66. A total of 54,340 patients (10.9% of all hospitalizations) had been admitted to the intensive care unit. We identified 71,437 deaths (mortality rate of 14.3% among hospitalized patients). We also observed important differences among regions, with Madrid being the epicenter of hospitalizations and mortality.

Conclusions We analyzed Spain's response to COVID-19 and describe here its experiences during the pandemic in terms of hospitalizations, critical illness, and deaths. This research highlights changes over several months and waves and the importance of factors such as vaccination, the predominant variant of the virus, and public health interventions in the rise and fall of the outbreaks.

Keywords COVID-19, Health care, SARS-CoV-2

*Correspondence:
Rafael Garcia-Carretero
rgcarretero@salud.madrid.org
[1]Department of Internal Medicine, Mostoles University Hospital, Rey Juan Carlos University, Madrid, Spain
[2]Department of Preventive Medicine and Public Health, Rey Juan Carlos University, Madrid, Spain

Introduction

A novel coronavirus disease was first described in 2019 and named 2019-nCoV. Its denomination was later replaced by severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2). This virus is the causative agent of coronavirus disease 2019 (COVID-19) [1]. This disease can range from mild symptoms to severe viral pneumonia with acute respiratory failure and distress syndrome, also known as severe acute respiratory syndrome [1, 2]. Although the World Health Organization reported cases of pneumonia in Wuhan City, Hubei Province, China, as early as 2019, the outbreak was finally declared a public health emergency in January 2020 and a pandemic in March 2020 [3].

Outbreaks of COVID-19 usually spread in regular patterns, but waves of the disease vary widely among different countries and regions within a single country. During the first weeks of the pandemic mortality rates in Spain and Italy rose to 15%, in contrast to countries such as Canada and Germany, whose mortality rates were less than 5% in the first wave [4–8]. The impact and the distribution of these waves depend on several factors, such as public health measures, interventions, lockdowns, and vaccination policies [9, 10]. SARS-CoV-2 variants can also have a great effect on mortality, because they can increase the transmissibility of the virus and clinical severity of the disease or decrease the effectiveness of public health interventions [11, 12].

Demographic data on the first weeks of the pandemic in Spain were reported early and provided a global overview of COVID-19 [13]. Spain reported its first confirmed case in January 2020. The epidemic escalated quickly and spread across all regions of the country. A state of emergency was finally declared in March 2020. Spain reported not only one of the highest incidence rates in Europe but also one of the highest mortality rates, accounting for 172,541 confirmed cases and 18,056 deaths by the end of April 2020 [14], corresponding to the first wave of the pandemic.

Describing and analyzing trends in the pandemic at a national level can help researchers and public health authorities obtain new insights into the pandemic. We collected all available data from the first 2 years of the pandemic to explore trends and the impact of the pandemic in Spain. Our aim was to analyze the behavior of the pandemic throughout its first 2 years at a national level. Here we describe the distribution of hospitalizations and admissions to intensive care units (ICUs) and the mortality rate by region in Spain.

Methods

Study design and data collection

We performed a nationwide, population-based, epidemiological study of all hospitalizations due to COVID-19 from February 2020 to December 2021. We used the Minimum Basic Data Set at Hospitalization (MBDS-H), which is an administrative database based on hospital discharge reports. The registry is mandatory for all public and private hospitals and so covers 95–97% of hospitalizations in Spain. This database is built from hospital discharge reports [15]. The MBDS-H includes not only demographic information, such as age, sex, and the province where hospitalization occurred, but also the date of admission to the hospital, the date of discharge, mortality, and ICU admission. There are two variables of interest: primary and secondary diagnosis. Data are available on reasonable request from the statistical portal of the Spanish National Health System [16] and are provided without personal data that could be used to identify patients to ensure their privacy.

Diagnoses in the data set are coded according to the International Classification of Diseases, 10th Revision, Clinical Modification (ICD-10-CM) [17]. We used codes U07.1 and B97.29 to extract the data. Because the data are based on hospital admissions, patients who either were discharged from the emergency room or were seen in outpatient settings are excluded. We included data from individuals whose primary or secondary diagnoses included codes U07.1 and/or B97.29 (both referring to COVID-19, according to the ICD-10-CM) in 2020 or 2021. As of this writing, data from 2022 are not yet available.

The Spanish National Health System updated the national coding recommendations, aligning them with the ICD-10-CM Guidelines for coding and reporting. Specifically, the code U07.1 COVID-19 has been recommended by the ICD-10-CM for encoding COVID-19 infections. Consequently, all records from April 1st, 2020, onwards should be encoded using this new code. In Spain, considering the overwhelmed state of Spanish hospitals, it was decided, after assessment by the Encoding Technical Unit (Unidad Técnica de Codificación), to implement the use of this new code starting from July 2020. This ensured the consistent use of data across the country by professional coders. This encoding change was taken into account during the preprocessing of our database.

During the initial months of the pandemic, the coding procedure involved considering the primary diagnosis along with the first secondary diagnosis. For instance, pneumonia due to COVID-19 was identified using codes J12.89 (other viral pneumonia) and B97.29 (other coronavirus as the cause of diseases classified elsewhere). However, since July 2020, the code U07.1 COVID-19 should be used only when a positive test result for SARS-CoV-2 has been confirmed. It should not be used if the testing is suspected, likely, or inconclusive. According to the recommendations, U07.1 should be encoded as the primary

Table 1 Overview of the main characteristics of hospitalized patients

	Total	2020	2021	p value
Patients	498,789	252,176	246,613	NA
Sex (male, %)	56.1	55.6	56.5	< 0.001
Age (years)	66 (28)	68 (27)	65 (28)	< 0.001
Hospital stay (days)	8 (9)	11.2 (9)	8 (9)	< 0.001
ICU (patients)	54,354	22,949	31,405	< 0.001
ICU (%)	10.9	9.1	12.7	< 0.001
ICU stay (days)	10.0 (21)	10 (16.7)	11 (23.5)	< 0.001
Deaths	71,437	40,512	30,925	< 0.001
Mortality rate (%)	14.3	16.1	12.5	< 0.001

NA: not applicable. ICU: intensive care unit. Data are expressed as percentages for categorical variables and as medians (interquartile ranges) for continuous variables

diagnosis, while the secondary diagnosis should be used to describe the clinical manifestations. For example, "U07.1 COVID-19+J12.89" should be used to describe pneumonia due to COVID-19. Our database underwent rigorous quality control at the Encoding Technical Unit (Spanish Ministry of Health) before being provided to investigators. We have taken all these aspects into consideration in order to conduct our analyses effectively.

Statistical analyses

We designed our research to be a descriptive, retrospective study. In order to represent the data from the data set, various statistical methods were employed. Firstly, absolute values were extracted to gain a clear understanding of the actual numbers involved, such as the total hospital admissions and ICU admissions. This allowed for a comprehensive overview of the data set. Additionally, percentages were calculated to provide a relative perspective, enabling comparisons between different variables. These percentages were computed for variables such as age groups, gender distribution, and geographic regions, shedding light on the proportional representation within each category. Although no correlational analysis was conducted, this descriptive statistical approach successfully facilitated the presentation and interpretation of the data set.

We performed the Anderson-Darling normality test to check the distribution of continuous variables, such as age, and report data as means and standard deviations or medians and interquartile ranges. Categorical variables are reported as absolute values and percentages. Differences between the years were assessed with the chi-square test for categorical variables and the Mann-Whitney U test for continuous variables. All statistical analyses were performed in R 4.2.2 (2022-10-31) on a GNU/Linux computer. For statistical significance, we used a p value of 0.05.

Results

Global overview

A total of 498,789 patients (1.04% of the Spanish population) were admitted to hospitals with a diagnosis of COVID-19 in the first 2 years of the pandemic. Table 1 shows the results by year. Men were more prone to hospitalization than women in this period. The median age was 66, but we found that patients were slightly older in 2020 than in 2021 (68 vs. 65, respectively). Hospital stay was longer in 2020 (median, 11.2 days) than in 2021 (median, 8 days). A total of 54,340 patients, or 10.9% of all hospitalizations, were admitted to the ICU. We also observed a global mortality rate of 14.3% among hospitalized patients, with 71,437 total deaths. The mortality rate was higher in 2020 (16.1%) than in 2021 (12.5%). Table 2 shows ICU admissions (as a surrogate of severity) and in-hospital mortality by different age groups. Figure 1 shows population pyramids for both hospital admissions and mortality. Men were more represented than women among both admissions and mortalities in almost all age ranges, except among patients older than 90 years old. Mortality was most frequent in patients ages 70 to 90 years old.

Evolution of the pandemic

In the period observed, we identified at least six waves of hospitalizations (Fig. 2). It is worth noting that one third

Table 2 Admissions, ICU admissions and mortality by age group

	2020					2021				
Age groups	**Total Admissions**	**ICU Admissions**	**ICU (%)**	**Deaths**	**Mortality rate (%)**	**Total Admissions**	**ICU Admissions**	**ICU (%)**	**Deaths**	**Mortality rate (%)**
< 14	1,978	177	8.9	6	0.3	3,475	205	5.9	14	0.4
15–44	29,618	2,279	7.7	343	1.2	41,106	4,400	10.7	355	0.9
45–64	77,489	9,711	12.5	3,764	4.9	78,126	13,221	16.9	3,631	4.6
65–74	47,373	7,151	15.1	6,688	14.1	45,818	9,287	20.3	5,784	12.6
> 74	95,821	3,659	3.8	29,743	31	78,098	4,292	5.5	21,147	27.1

ICU: intensive care unit. Data are expressed as percentages and absolute values

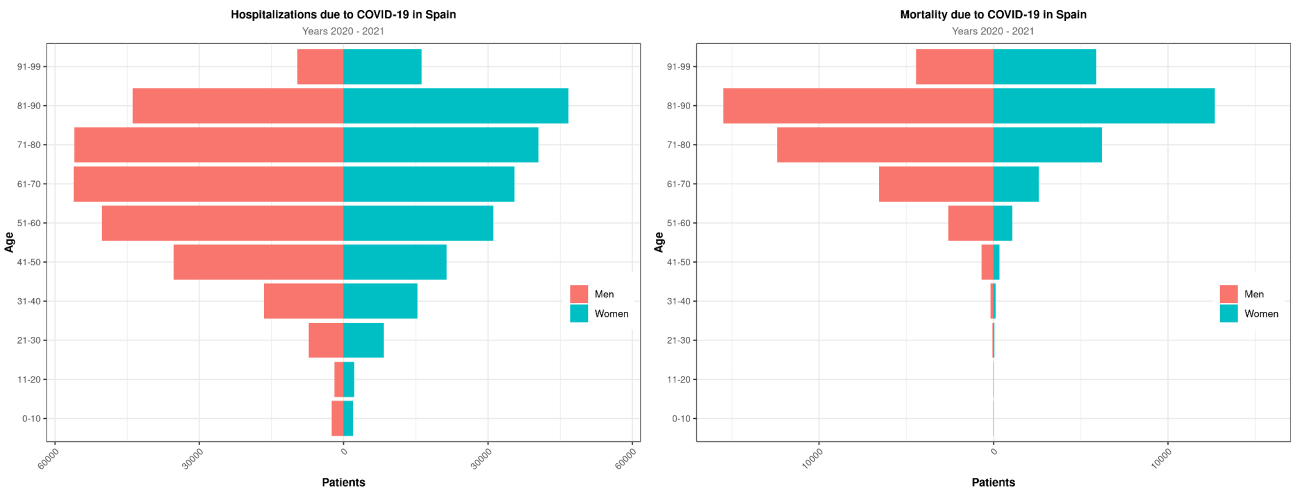

Fig. 1 Population pyramids for hospital admissions and mortality

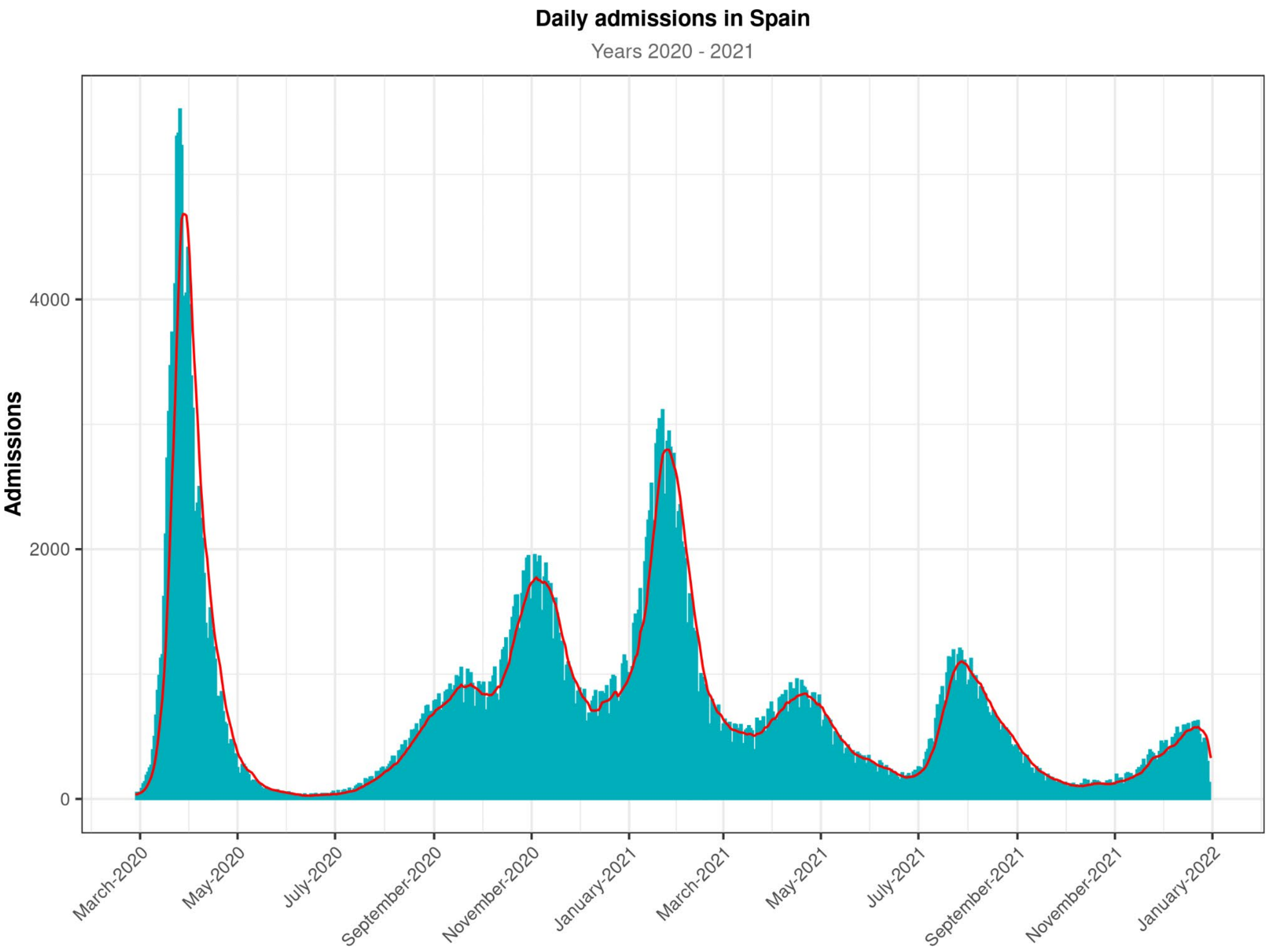

Fig. 2 Time series showing the evolution of COVID-19-related hospitalizations in Spain in 2020 and 2021. The red line denotes the 7-day moving average, whereas the gray bars denote absolute numbers of hospital admissions

Table 3 Distribution of hospital admissions among different regions of Spain

	2020				2021			
	Admissions	**Sex (male, %)**	**Age**	**Hospital stay (days)**	**Admissions**	**Sex (male, %)**	**Age**	**Hospital stay (days)**
Andalucía	22,999	55.59	68 (26)	8 (9)	35,648	56.22	63 (28)	8 (9)
Aragón	10,117	53.72	73 (27)	9 (9)	9,075	55.44	66 (29)	8 (8)
Asturias	5,804	52.41	76 (24)	9 (9)	5,588	55.53	68 (26)	7 (8)
Illes Balears	3,080	58.31	63 (29)	8 (9)	4,008	56.91	60 (29)	8 (10)
Islas Canarias	2,627	57.52	66 (25)	11 (12)	5,476	56.46	60 (29)	10 (11.2)
Cantabria	2,558	53.28	71 (27)	8 (7)	2,895	53.78	63 (30)	7 (6)
Castilla-León	18,880	56.19	74 (25)	8.5 (9)	15,463	58.2	68 (28)	8 (9)
C. La Mancha	12,313	57.7	71 (24)	8 (8)	7,709	56.14	69 (27)	8 (9)
Catalunya	51,237	56.17	67 (26)	7 (8)	47,943	57.12	65 (29)	7 (9)
C. Valenciana	17,448	55.69	66 (27)	8 (8)	28,531	56.76	66 (27)	7 (8)
Extremadura	3,674	52.86	73 (24)	9 (8)	4,419	55.4	71 (27)	8 (8)
Galicia	6,620	54.32	72 (25)	9 (10)	9,775	55.81	68 (28)	8 (11)
C. Madrid	70,105	55.44	66 (28)	7 (9)	45,407	55.93	63 (28)	8 (9)
R. Murcia	5,016	55.3	60 (29)	7 (8)	6,206	56.69	62 (28)	7 (8)
Navarra	3,329	54.73	69 (27)	7 (7)	2,220	55.95	65 (30)	7 (9)
Euskadi	13,273	56.42	70 (25)	7 (8)	13,432	57.65	65 (27)	7 (8)
La Rioja	2,532	53.36	70 (27)	8 (7)	2,007	56.15	67 (28)	7 (8)
Ceuta	239	46.3	65 (22)	9 (10)	293	59.73	58 (25)	11 (11)
Melilla	325	54.15	60 (25)	8 (8)	518	50.77	59 (27)	8 (8)
Total	252,176	55.59	68 (28)	8 (9)	246,613	56.52	65 (28)	8 (9)

Sex is expressed as a percentage, whereas age and hospital stay are expressed as medians (interquartile ranges)

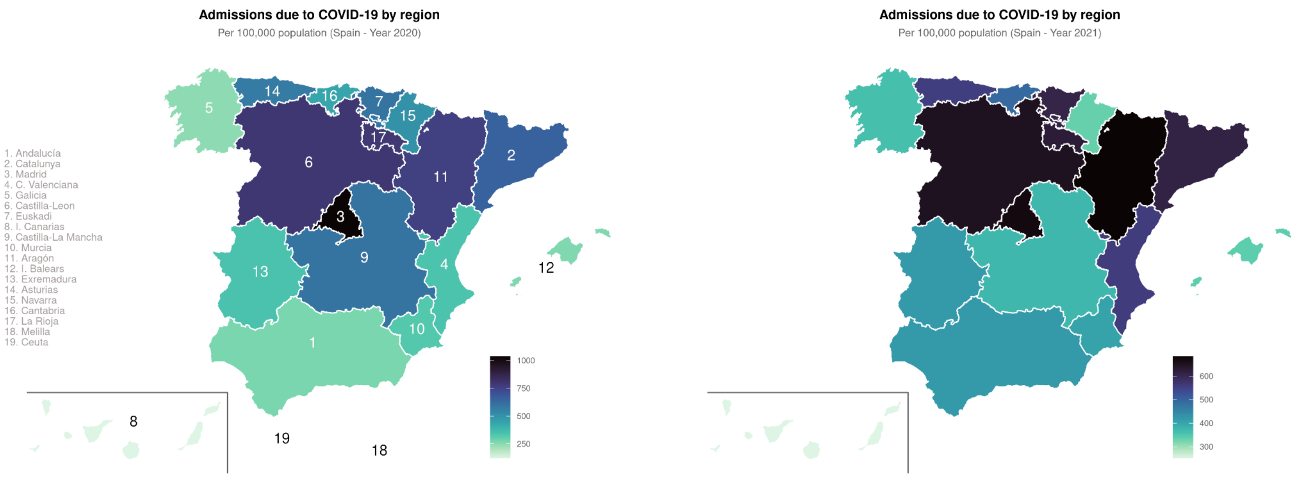

Fig. 3 Geographic distribution of hospitalizations by region, attending to the population of each region. Data are expressed as hospitalizations per 100,000 population

of all hospital admissions occurred within the first 12 weeks of the pandemic.

Distribution of the pandemic by region

The geographic distribution of the pandemic in the different regions of Spain can be seen in Table 3, which shows geographic variability among regions. During the first year of the pandemic, most hospitalizations were in Madrid (70,105 admissions) and Catalunya (51,237 admissions). Overall, in all regions, men were more likely than women to be hospitalized in both 2020 and 2021. The median age was above 60 in all regions (68 globally) in 2020, but in 2021 patients tended to be younger, with a global median age of 65 (in only one region above 70). Overall, sex, age, and hospital stay were consistent among regions of Spain and between years in a given region.

The number of hospital admissions per region in Table 3 is however expressed in absolute values and might be misleading. In 2021 some regions, such as Andalucía and C. Valenciana, had more admissions than in the previous year. We thus plotted the same data, attending to the population of each region. Figure 3 shows the geographic distribution of admissions per 100,000 population. This new distribution is more homogeneous than that shown in Table 3, with less pronounced differences among regions. Nevertheless, regions such as Andalucía, Extremadura, Murcia, Galicia, and Navarra had moderately low admissions per 100,000 population compared to regions such as Madrid and Catalunya.

We calculated the crude and the age-standardized (adjusted) mortality rates per region using the direct method, showed in Table 4. Figure 4 shows the geographic distribution of age-standardized mortality.

Severity of COVID-19: ICU admissions and mortality

We also assessed ICU admissions and mortality and their distribution by region. Table 5 shows the number of ICU admissions, the length of ICU stay, and mortality in the period studied. ICU admissions and mortality are reported as both absolute values and ratios. These ratios were calculated based on the number of hospitalized patients. Globally, there were 40,512 registered deaths in 2020 (mortality rate of 16%) and 30,925 deaths in 2021 (mortality rate of 12.5%). There were more ICU admissions in 2021 than in 2020 (Fig. 5). We also found differences among regions. In terms of mortality, Ceuta, Castilla-La Mancha, Extremadura, and Castilla-León reported the highest ratios of deaths per hospitalized patients (above 20%) in 2020. The global ratio decreased in 2021, but regions such as Ceuta and Extremadura registered the highest mortality rates in Spain. It is important to consider that when the number of events is extremely small, the calculation method for confidence intervals may result in a negative value for the lower confidence limit, as observed in the case of Ceuta. This occurrence is often attributed to the limited number of events, specifically 28 ICU admissions. Consequently, researchers should exercise caution and interpret these results with care.

Table 4 Age-adjusted (standardized) mortality rates per region, per 100,000 population

	Crude rate	Adjusted rate (95% CI)
Andalucía	115.9	121 (118.6–123.4)
Aragón	231.2	189.6 (182.9–196.5)
Asturias	169.4	117 (111.5–122.8)
Illes Balears	54.2	62.5 (57.7–67.5)
Canarias	41.9	45.8 (42.9–48.9)
Cantabria	83	67.9 (62–74.3)
Castilla-León	267.8	183 (178.5–187.7)
Castilla-La Mancha	184.8	173.4 (167.9–179.1)
Cataluña	153.4	148.2 (145.5–150.9)
C. Valenciana	134.9	127.6 (124.6–130.7)
Extremadura	164.1	139.2 (132.7–146)
Galicia	90.3	62.8 (60.3–65.4)
Madrid	232.5	234.1 (230.5–237.8)
Murcia	79.8	90.9 (85.9–96.2)
Navarra	116.1	104.9 (97.6–112.6)
Euskadi	143.7	114.9 (110.9–119)
La Rioja	183.5	155.7 (143.3–168.9)
Ceuta	138.4	198.1 (162.9–238.9)
Melilla	132.7	225 (184.7–271.8)

CI: Confidence interval

Discussion

The main objective of our nationwide study was to analyze two available years of data on hospitalizations, ICU admissions, and mortality due to COVID-19 provided by the Spanish National Health System. The COVID-19 outbreak overwhelmed the capacity of hospitals, and our aim was to assess its impact on the Spanish population. To the best of our knowledge, this is the first Spanish nationwide research on this topic.

Spain was one of the epicenters of the pandemic in Europe, with high rates of both hospitalizations and mortality, especially among the elderly. Specifically, Madrid had the highest rate of confirmed cases and hospitalizations [14, 18]. We found that in the first 2 years of the pandemic, almost 500,000 patients in Spain needed hospitalization, and more than 70,000 died. The COVID-19 pandemic overwhelmed hospital wards and ICUs and subsequently the capacity of the health care system.

The burden in terms of hospitalization, ICU admissions, death, and health care capacity varied by region. We found great variation in hospitalizations and deaths across regions, which highlights the different impacts on each region, even if hospital admissions are adjusted by population. The reasons for such differences are unclear, and a comparison of clinical outcomes and economic factors or sociodemographics by region is beyond the scope of this research; however, we can raise a hypothesis of economic disparities among regions [19]. Researchers studied geographic variability in the relationships between mortality rate and social, demographic, and health care factors in Spain during the first wave of the pandemic [20]. The unemployment rate, population density, and population size were the main underlying factors related to geographic variation in risk of death. The unemployment rate is an indicator of socioeconomic status, and regions with high unemployment invest less in education and health care. Also, population density is associated with social distancing, which emphasizes the importance of restrictive measures. Although the researchers encouraged further research, their data provide critical information for mitigating the impact of COVID-19 and future pandemics. This raises the question of how different regions fare during public health emergencies and can identify areas that can be improved for future outbreaks.

Several researchers have reported that older age is associated with hospitalization [13, 21]. Likewise, we found that 77.4% of hospitalizations were among individuals

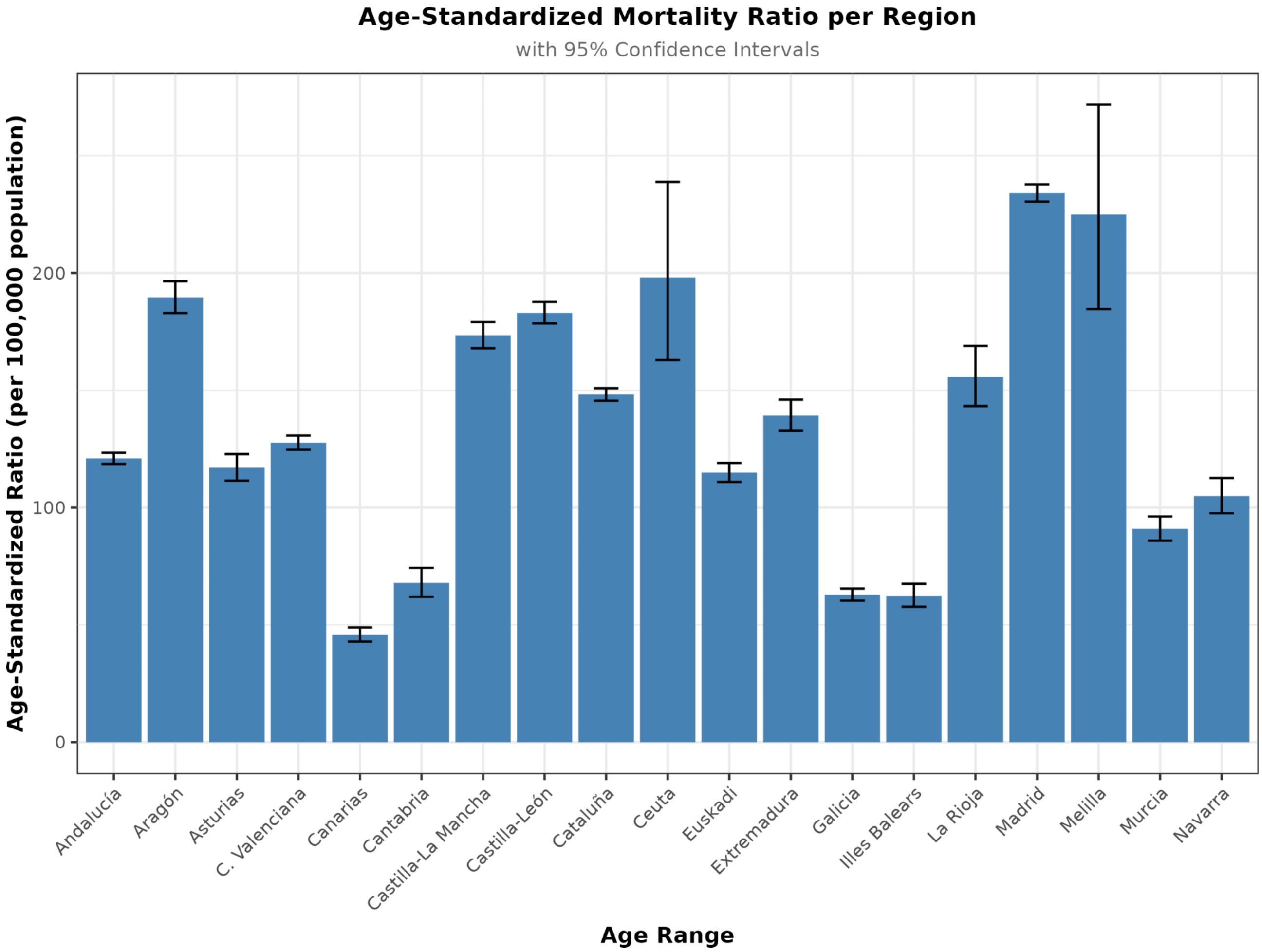

Fig. 4 Age-standardized (adjusted) mortality rate per region, using the direct method

older than 50, 42.7% were among those older than 70, and 23.4% of all hospital admissions were among patients older than 80. Moreover, for each additional decade of life, the risk increased by 71.5%. Although this was not a correlational study, we found that age was probably the major factor associated with hospitalization due to COVID-19.

Regarding the distribution of waves over time, we can hypothesize only about the association—not causation—between certain events and peaks of hospitalizations in Spain. The first wave was associated with the initial outbreak and was restrained by strict public health measures, such as confinement and lockdown. The second wave began in the summer of 2020, when those restrictions ended and social distancing measures were relaxed. This wave reached a peak in the autumn of 2020, probably because of the return to work and school. The third peak began in December 2020, probably as a result of holiday events and Christmas gatherings, and continued until January 2021. The fourth wave showed a rapid fall in hospitalizations, probably because of vaccination, and its peak might have been associated with the Easter holidays. The fifth and the sixth waves (September 2021 and December 2021, respectively) were similar to the fourth wave in terms of hospitalizations, and they showed the beneficial effect of vaccination. Regardless of vaccination, it seems that waves and peaks were related to social events: holidays, gatherings, and the relaxation of public health measures such as social distancing.

In the summer of 2020 new drugs began to be used to treat COVID-19 in hospitalized patients, such as remdesivir; corticosteroids [22, 23]; and immunomodulatory drugs such anakinra, tozilizumab, and baricitinib [24–29]. The use of these new drugs can explain the lower mortality in 2021 [30].

Other Spanish studies have also identified several waves, characterized as peaks in the incidence of confirmed cases of COVID-19. It is worth noting an epidemiological study by Red Nacional de Vigilancia Epidemiologica (RENAVE) in which data from Spain were reported [18]. In this study, confirmed cases, hospitalizations, and deaths were analyzed through May

Table 5 Distribution of ICU admissions and mortality among different regions of Spain

	2020					2021				
	ICU admissions	ICU (%)	ICU stay	Deaths	Mortality rate (%)	ICU admissions	ICU (%)	ICU stay	Deaths	Mortality rate (%)
Andalucía	2,314	10.06	10 (17)	4,287	18.64	4,115	11.54	10 (16)	5,563	15.61
Aragón	706	6.98	14 (19)	1,917	18.95	625	6.89	13 (22)	1,150	12.67
Asturias	541	9.32	10 (15)	1,048	18.06	799	14.3	14 (24)	653	11.69
Illes Balears	356	11.56	7 (19)	295	9.58	649	16.19	1 (10)	343	8.56
I. Canarias	470	17.89	11 (19)	331	12.6	1,150	21	11 (19)	579	10.57
Cantabria	213	8.33	9 (12)	294	11.49	453	15.65	9 (13)	192	6.63
Castilla-León	1,611	8.53	11 (16)	4,070	21.56	1,722	11.14	13 (24)	2,281	14.75
C. La Mancha	1,061	8.62	11 (19)	2,602	21.13	931	12.08	12 (22)	1,193	15.48
Catalunya	6,313	12.32	8 (14)	7,156	13.97	8,527	17.79	8 (12)	4,795	10
C. Valenciana	1,815	10.4	9 (14)	2,572	14.74	2,845	9.97	11 (17)	4,303	15.08
Extremadura	291	7.92	13 (16)	870	23.68	370	8.37	13.5 (16)	861	19.48
Galicia	470	7.1	9 (13)	1,103	16.66	1,008	10.31	10 (15)	1,326	13.57
C. Madrid	4,608	6.57	11 (17)	10,676	15.23	5,105	11.24	22 (31)	4,999	11.01
R. Murcia	654	13.04	9 (13)	474	9.45	901	14.52	8 (15)	749	12.07
Navarra	227	6.82	17 (25)	513	15.41	272	12.25	10 (18)	258	11.62
Euskadi	1,052	7.93	11 (17)	1,842	13.88	1,522	11.33	11 (18)	1,327	9.88
La Rioja	171	6.75	11 (13)	366	14.45	256	12.76	15 (21)	221	11.01
Ceuta	28	11.72	15 (25)	54	22.59	51	17.41	13 (19)	61	20.82
Melilla	48	14.77	7.5 (9)	42	12.92	104	20.08	6 (11)	71	13.71
Total	22,949	9.1	10 (16)	40,512	16.06	31,405	12.73	11 (19)	30,925	12.54

ICU: intensive care unit. ICU and mortality rates are expressed as percentages, whereas ICU stay is expressed as a median (interquartile range)

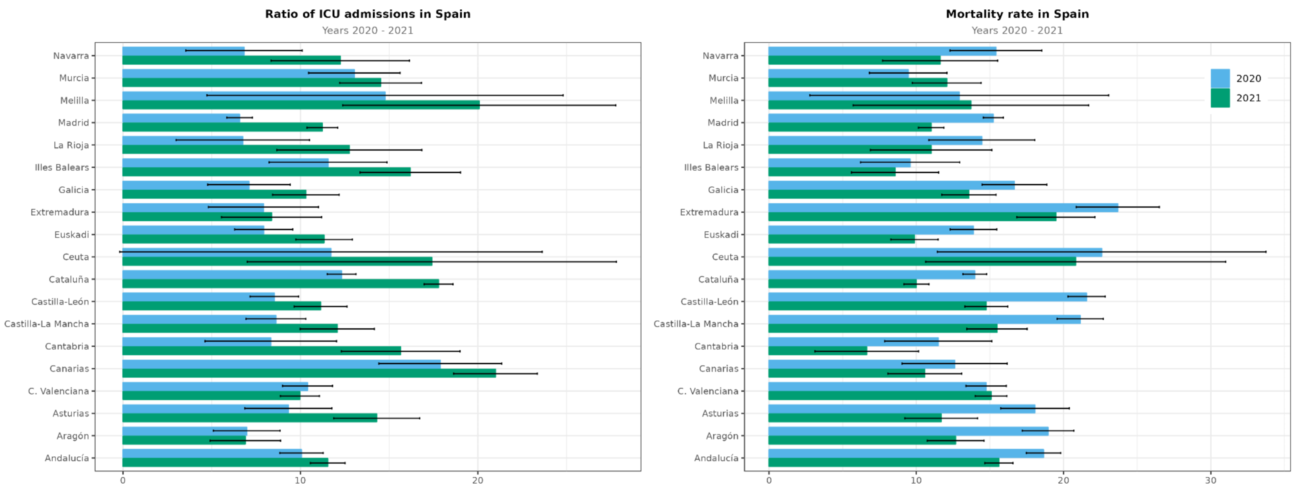

Fig. 5 Ratios of both admissions to the intensive care unit (ICU) and deaths among hospitalized patients

10, 2022. The researchers split the pandemic into periods and established a turning point for each wave based on the 14-day cumulative incidence. Waves in that research are in line with the ones in our study (i.e., their peaks occurred simultaneously with the waves of hospitalizations and deaths identified in our research). We found that after the first outbreak of COVID-19, hospitalizations increased over the first three waves of the pandemic. After that, hospitalizations underwent a significant decline. Early mass vaccination probably delayed and alleviated the fourth wave, which finally occurred in the spring of 2021.

The most important public health measure during the pandemic was vaccination. Vaccination had begun in the European Union by January 2021. Its effects, in terms of admission and mortality, were not evident until April 2021. Beginning in December 2020, vaccines were administered to the entire Spanish population, and they proved significant protection against severe COVID-19. The first peaks before vaccination occurred in March

2020, November 2020, and January 2021. Beginning in the spring of 2021, the use of vaccines resulted in a decrease in hospitalizations, probably because their protection was based on achieving a mild clinical presentation of COVID-19. They had a great impact in terms of hospital admissions and mortality. The effect of vaccination on the fourth wave was studied elsewhere [21]. Moreover, we demonstrated that this impact was steady in the fifth and sixth waves. It is of interest to mention a study by Barandalla et al. [21], who developed simulated curves of hospitalizations in the absence of vaccines and then compared those curves with the real incidence. By showing the decrease in incidence, they demonstrated the beneficial impact of the vaccination rollout on hospitalizations.

Several noteworthy caveats should be mentioned. For one thing, several variants of the SARS-CoV-2 virus were identified throughout the pandemic. The changes in SARS-CoV-2 variants can also explain the beginning of certain waves. In the summer of 2020 the alpha variant (B.1.1.7) was described [12], and this variant spread rapidly in several European countries. It had greater transmissibility, virulence, and lethality than previous variants and could have been responsible for the second and third waves. Alpha was replaced by delta (B.1.617.2) in Spain in the summer of 2021 and could have been one of the causes of the peak of the fifth wave. Nevertheless, there is no evidence that delta or other new variants had a negative impact on hospitalizations after the summer of 2021, as clinical presentations were often mild [31] and vaccination against these two variants was effective [32, 33].

Herd immunity is another caveat. An interesting Spanish study (Estudio Nacional de sero-Epidemiología de la Infección por SARS-CoV-2 en España, or ENE-COVID) analyzed seroprevalence after the first wave [34]. Only 5% of the entire Spanish population had antibodies against SARS-CoV-2, but measurements varied across different regions. Madrid, Castilla-La Mancha, and Castilla-León (central regions) had seroprevalence above 10% compared to approximately 1% in regions such as Galicia and C. Valenciana (peripheral regions along the coast). Herd immunity was not achieved in Spain after the first wave, which clearly rendered vaccination a key factor in achieving protection against SARS-CoV-2.

ENE-COVID not only demonstrated that there were several epicenters during the pandemic, such as Madrid, but also highlighted the proportion of symptomatic individuals who were not diagnosed by a proper polymerase chain reaction test. Concerning testing policy, a recent publication by Zhang et al. [35] demonstrated that mass testing was associated with a 25% decrease in hospitalizations due to COVID-19. The city of Liverpool (United Kingdom) was selected for a pilot study. The intervention involved testing asymptomatic individuals to identify infected people to protect vulnerable individuals, quarantine contacts, and ultimately improve public health. This intervention reduced COVID-19-related admissions because promoting effective isolation of confirmed individuals and their contacts resulted in reduced onward transmission. The study estimated a 32% reduction in admissions compared to the expected admissions with no intervention. ENE-COVID demonstrated that underdiagnosis was common during the first weeks of the pandemic. Testing policy and their impact on the COVID-19 pandemic in Spain before vaccines were available were also studied by several researchers [30, 36]. The first three waves were analyzed, and several differences were found between the first wave (with its more restrictive testing policy) and the second and third ones (with their less restrictive policies). These researchers demonstrated not only an increase in the number of confirmed cases in the general population but also decreases in the number of severe cases requiring treatment in the ICU and in mortality rates during the second and third waves compared to the first wave. In the summer of 2020 a less restrictive testing policy began to be effectively implemented in Spain, ensuring the detection of cases with mild symptoms and even asymptomatic individuals. These interventions contributed to addressing the outbreaks, as pointed out by another study that demonstrated that underreporting cases of COVID-19 can lead to poor outcomes [37]. The more cases are reported, the more able experts are to trace confirmed cases and to maintain quarantine. Thus, nonpharmaceutical interventions such as restrictive testing policies, confinement, masks, restricted mobility, and social distancing can also have a great impact on COVID-19.

Limitations: the reliability of administrative data

The main limitation of our study was our use of an administrative database. Although electronic health records help researchers collect data and make them available for research, some clinical data are not readily available. Information about vaccination or drugs were not recorded, so we could not assess the effect of vaccination or certain treatments. This issue is inherent to the MBDS-H because of the nature of its coding process. Also, we only had access to data from 2020 to 2021, because data from 2022 had not yet been processed. This limitation causes a delay in reporting epidemiological data such as ours. Another limitation our of study has to do with the reliability of the data. Although it is based on discharge reports, the MBDS-H was primarily designed as an administrative data set focused on economical management rather than clinical relevance [38–40]. Diagnoses are recorded and coded, but data regarding diagnostic procedures, treatments, and outcomes are not registered. Moreover, the validity of the collected data relies on the accuracy of the medical discharge reports and the recording of the variables. It is worth noting that diagnoses may have been provisional or redundant, and some of them may have been excluded if administrative personnel could not find an appropriate code in the ICD-10-CM. The administrative personnel may not have had any medical training. On the whole, we can state that the reliability of such databases may be not guaranteed [41]. However, although a certain amount of misclassification is expected, reliability can be improved with strategies such as examining secondary diagnoses [42]. In contrast, some have argued that the MBDS-H provides sufficiently valid information and can be a useful tool in epidemiological and clinical studies [43, 44].

Conclusions

Here we offered a picture of SARS-CoV-2-related hospitalizations in Spain in the first 2 years of the pandemic. Our aim was to describe trends and the distribution of the pandemic in Spain and to determine the impact of the pandemic on the Spanish health care system and population. We identified at least six waves over the 2 years. We also found differences between the years in terms of the number of hospitalizations, ICU admissions, and deaths. We proposed explanations for the rise and fall in hospitalizations, such as public health measures or social distancing, but we are aware that certain factors may have played an important role in the waxing and waning of each wave, such as vaccination rates or the predominant variant of the virus. Vaccination was probably the most important public health intervention to mitigate the pandemic, as it gave individuals protection against severe COVID-19. The distributions of both hospital admissions and mortality showed great differences across regions of Spain, probably because of underlying socioeconomic factors.

Acknowledgements
Not applicable.

Authors' contributions
Dr. Garcia-Carretero designed and conceived the study, preprocessed the data, analyzed and interpreted the data, and wrote the first draft of the manuscript. Dr. Vazquez-Gomez made substantial contributions to the interpretation of the results, critically reviewed the first draft of the manuscript, and made valuable suggestions. Drs. Gil-Prieto and Gil-de-Miguel supervised the project and critically reviewed and edited the final draft of the manuscript. All authors read and approved the final manuscript.

Funding
The authors received no financial support for this research or the authorship or publication of this article.

Data Availability
According to the terms of a contract signed with the Spanish Ministry of Health, which provided the data, the authors cannot provide the data set to any other researcher. Furthermore, the data must be destroyed at the conclusion of the research. Data can be obtained at https://www.sanidad.gob.es/estadEstudios/portada/home.htm.

Declarations

Ethics approval and Consent to Participate
Our retrospective study was approved by the Research and Ethics Committee of Mostoles University Hospital. All procedures involving human participants were conducted in accordance with the ethical standards of the responsible institutional and/or National Research Committee and with the tenets of the 1964 Helsinki Declaration and its later amendments and comparable ethical standards. The authors obtained consent to publish from their institution and its Research and Ethics Committee. The informed consent was waived by the Research and Ethics Committee of our institution because of the retrospective nature of the study and the analysis used anonymous clinical data.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 22 April 2023 / Accepted: 10 July 2023
Published online: 18 July 2023

References

1. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol. 2020 Apr;5(4):536–44.
2. Lovato A, de Filippis C. Clinical presentation of COVID-19: a systematic review focusing on Upper Airway symptoms. Ear Nose Throat J. 2020 Nov;99(9):569–76.
3. World Health Organization. WHO Director-General's opening remarks at the media briefing on COVID-19–11 March 2020 [Internet]. Available from: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020.
4. Dowd JB, Andriano L, Brazel DM, Rotondi V, Block P, Ding X, et al. Demographic science aids in understanding the spread and fatality rates of COVID-19. Proc Natl Acad Sci U S A. 2020 May;5(18):9696–8.
5. Ghio D, Acosta E, Fisman D, Noymer A, Stilianakis NI, Assche SBV. Population Health and COVID-19 in Canada: a demographic comparative perspective. Can Stud Popul. 2021;48(2–3):131–7.
6. Assche SBV, Ghio D. A demographic adjustment to improve measurement of COVID-19 severity at the developing stage of the pandemic [Internet]. medRxiv; 2020 [cited 2022 Jun 9]. p. 2020.03.23.20040998. Available from: https://www.medrxiv.org/content/https://doi.org/10.1101/2020.03.23.20040998v1.
7. Aburto JM, Schöley J, Kashnitsky I, Zhang L, Rahal C, Missov TI, et al. Quantifying impacts of the COVID-19 pandemic through life-expectancy losses: a population-level study of 29 countries. Int J Epidemiol. 2022 Feb;18(1):63–74.
8. Kashnitsky I, Aburto JM. COVID-19 in unequally ageing european regions. World Dev. 2020 Dec;1:136:105170.
9. Tracking SARS-. CoV-2 variants [Internet]. [cited 2022 Jun 9]. Available from: https://www.who.int/activities/tracking-SARS-CoV-2-variants.
10. Thakur V, Bhola S, Thakur P, Patel SKS, Kulshrestha S, Ratho RK et al. Waves and variants of SARS-CoV-2: understanding the causes and effect of the COVID-19 catastrophe. Infection. 2022 Apr 1;50(2):309–25.
11. Andrews N, Stowe J, Kirsebom F, Toffa S, Rickeard T, Gallagher E et al. Covid-19 Vaccine Effectiveness against the Omicron (B.1.1.529) Variant. New England Journal of Medicine. 2022 Apr 21;386(16):1532–46.
12. Hodcroft EB, Zuber M, Nadeau S, Vaughan TG, Crawford KHD, Althaus CL et al. Spread of a SARS-CoV-2 variant through Europe in the summer of 2020. Nat 2021 Jul;595(7869):707–12.
13. Casas-Rojo JM, Antón-Santos JM, Millán-Núñez-Cortés J, Lumbreras-Bermejo C, Ramos-Rincón JM, Roy-Vallejo E, et al. Clinical characteristics of patients hospitalized with COVID-19 in Spain: results from the SEMI-COVID-19 Registry. Rev Clin Esp (Barc). 2020 Nov;220(8):480–94.
14. Pastor-Barriuso R, Pérez-Gómez B, Hernán MA, Pérez-Olmeda M, Yotti R, Oteo-Iglesias J, et al. Infection fatality risk for SARS-CoV-2 in community dwelling

population of Spain: nationwide seroepidemiological study. BMJ. 2020 Nov;27:371:m4509.
15. España. Real Decreto 69/2015, de 6 de febrero, por el que se regula el Registro de Actividad de Atención Sanitaria Especializada. BOE; 2015.
16. Ministerio de Sanidad Consumo y Bienestar Social. Portal Estadistico. Area de Inteligencia de Gestion [Internet]. [cited 2019 Jul 6]. Available from: https://pestadistico.inteligenciadegestion.mscbs.es/publicoSNS/comun/ArbolNodos.aspx?idNodo=23525.
17. Ministerio de Sanidad Consumo y Bienestar Social. eCIEMaps - CIE-10-ES Diagnosticos [Internet]. [cited 2019 Jul 6]. Available from: https://eciemaps.mscbs.gob.es/ecieMaps/browser/index_10_mc.html.
18. Informe no 128 Situación de COVID. -19 en España a 10 de mayo de 2022.pdf [Internet]. [cited 2022 Jun 9]. Available from: https://www.isciii.es/QueHacemos/Servicios/VigilanciaSaludPublicaRENAVE/EnfermedadesTransmisibles/Documents/INFORMES/Informes%20COVID-19/INFORMES%20COVID-19%202022/Informe%20n%C2%BA%20128%20Situaci%C3%B3n%20de%20COVID-19%20en%20Espa%C3%B1a%20a%2010%20de%20mayo%20de%202022.pdf.
19. Palomino JC, Rodríguez JG, Sebastian R. The COVID-19 shock on the labour market: poverty and inequality effects across Spanish regions. Regional Studies. 2022 Sep 23;0(0):1–15.
20. Dahal S, Mizumoto K, Rothenberg R, Chowell G. Investigating spatial variability in COVID-19 pandemic severity across 19 geographic areas, Spain, 2020 [Internet]. medRxiv; 2020 [cited 2023 Apr 1]. p. 2020.04.14.20065524. Available from: https://www.medrxiv.org/content/https://doi.org/10.1101/2020.04.14.20065524v3.
21. Barandalla I, Alvarez C, Barreiro P, de Mendoza C, González-Crespo R, Soriano V. Impact of scaling up SARS-CoV-2 vaccination on COVID-19 hospitalizations in Spain. Int J Infect Dis 2021 Nov 1;112:81–8.
22. Horby P, Lim WS, Emberson JR, Mafham M, Bell JL, Linsell L et al. Dexamethasone in Hospitalized Patients with Covid-19. The New England journal of medicine. 2021 Feb;384(8):693–704.
23. WHO Rapid Evidence Appraisal for COVID-19 Therapies (REACT) Working Group, Sterne JAC, Murthy S, Diaz JV, Slutsky AS, Villar J, et al. Association between Administration of systemic Corticosteroids and Mortality among critically ill patients with COVID-19: a Meta-analysis. JAMA. 2020 Oct;6(13):1330–41.
24. Kyriazopoulou E, Huet T, Cavalli G, Gori A, Kyprianou M, Pickkers P et al. Effect of anakinra on mortality in patients with COVID-19: a systematic review and patient-level meta-analysis. The Lancet Rheumatology. 2021 Oct 1;3(10):e690–7.
25. Wolfe CR, Tomashek KM, Patterson TF, Gomez CA, Marconi VC, Jain MK et al. Baricitinib versus dexamethasone for adults hospitalised with COVID-19 (ACTT-4): a randomised, double-blind, double placebo-controlled trial. The Lancet Respiratory Medicine [Internet]. 2022 May 23 [cited 2022 Jun 9];0(0). Available from: https://www.thelancet.com/journals/lanres/article/PIIS22132600(22)00088-1/fulltext.
26. Kalil AC, Patterson TF, Mehta AK, Tomashek KM, Wolfe CR, Ghazaryan V et al. Baricitinib plus Remdesivir for hospitalized adults with Covid-19. New Engl J Med 2021 Mar 4;384(9):795–807.
27. Marconi VC, Ramanan AV, de Bono S, Kartman CE, Krishnan V, Liao R et al. Efficacy and safety of baricitinib for the treatment of hospitalised adults with COVID-19 (COV-BARRIER): a randomised, double-blind, parallel-group, placebo-controlled phase 3 trial. The Lancet Respiratory Medicine. 2021 Dec 1;9(12):1407–18.
28. Salama C, Han J, Yau L, Reiss WG, Kramer B, Neidhart JD et al. Tocilizumab in Patients hospitalized with Covid-19 pneumonia. New Engl J Med 2021 Jan 7;384(1):20–30.
29. Rosas IO, Bräu N, Waters M, Go RC, Hunter BD, Bhagani S et al. Tocilizumab in Hospitalized patients with severe Covid-19 pneumonia. New Engl J Med 2021 Apr 22;384(16):1503–16.
30. Taboada M, González M, Alvarez A, Eiras M, Costa J, Álvarez J, et al. First, second and third wave of COVID-19. What have we changed in the ICU management of these patients? J Infect. 2021 Jun;82(6):e14–5.
31. Hacisuleyman E, Hale C, Saito Y, Blachere NE, Bergh M, Conlon EG, et al. Vaccine breakthrough infections with SARS-CoV-2 variants. N Engl J Med. 2021 Jun;10(23):2212–8.
32. Veneti L, Valcarcel Salamanca B, Seppälä E, Starrfelt J, Storm ML, Bragstad K, et al. No difference in risk of hospitalization between reported cases of the SARS-CoV-2 Delta variant and alpha variant in Norway. Int J Infect Dis. 2022 Feb;1:115:178–84.
33. Seppälä E, Veneti L, Starrfelt J, Danielsen AS, Bragstad K, Hungnes O et al. Vaccine effectiveness against infection with the Delta (B.1.617.2) variant, Norway, April to August 2021. Euro Surveill. 2021 Sep 2;26(35):2100793.
34. Pollán M, Pérez-Gómez B, Pastor-Barriuso R, Oteo J, Hernán MA, Pérez-Olmeda M, et al. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. The Lancet. 2020 Aug;22(10250):535–44.
35. Zhang X, Barr B, Green M, Hughes D, Ashton M, Charalampopoulos D, et al. Impact of community asymptomatic rapid antigen testing on covid-19 related hospital admissions: synthetic control study. BMJ. 2022 Nov;23:379:e071374.
36. Soriano V, de Mendoza C, Gómez-Gallego F, Corral O, Barreiro P. Third wave of COVID-19 in Madrid, Spain. Int J Infect Dis. 2021 Jun;107:212–4.
37. Moriña D, Fernández-Fontelo A, Cabaña A, Arratia A, Ávalos G, Puig P. Cumulated burden of COVID-19 in Spain from a Bayesian perspective. European Journal of Public Health. 2021 Aug 1;31(4):917–20.
38. Howe JL, Adams KT, Hettinger AZ, Ratwani RM. Electronic Health Record Usability Issues and potential contribution to patient harm. JAMA. 2018 Mar;319(12):1276–8.
39. Erickson SM, Rockwern B, Koltov M, McLean RM. Putting Patients First by Reducing Administrative Tasks in Health Care: A Position Paper of the American College of Physicians. Annals of internal medicine. 2017 May;166(9):659–61.
40. Sinsky C, Tutty M, Colligan L. Allocation of Physician Time in Ambulatory practice. Annals of internal medicine. 2017 May;166(9):683–4.
41. Calle JE, Saturno PJ, Parra P, Rodenas J, Perez MJ, Eustaquio FS, et al. Quality of the information contained in the minimum basic data set: results from an evaluation in eight hospitals. Eur J Epidemiol. 2000;16(11):1073–80.
42. Redondo-Gonzalez O, Tenias-Burillo JM. A multifactorial regression analysis of the features of community-acquired rotavirus requiring hospitalization in Spain as represented in the Minimum Basic Data Set. Epidemiol Infect. 2016 Sep;144(12):2509–16.
43. Fernandez-Navarro P, Lopez-Abente G, Salido-Campos C, Sanz-Anquela JM. The Minimum Basic Data Set (MBDS) as a tool for cancer epidemiological surveillance. Eur J Intern Med 2016 Oct;34:94–7.
44. Hernandez Medrano I, Guillan M, Masjuan J, Alonso Canovas A, Gogorcena MA. Reliability of the minimum basic dataset for diagnoses of cerebrovascular disease. Neurologia (Barcelona, Spain). 2017 Mar;32(2):74–80.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

7 Effectiveness of Vaccination in Spain

Journal of
Clinical Medicine

MDPI

Article

Impact and Effectiveness of COVID-19 Vaccines Based on Machine Learning Analysis of a Time Series: A Population-Based Study

Rafael Garcia-Carretero [1,*,†], Maria Ordoñez-Garcia [2,*,†], Oscar Vazquez-Gomez [1], Belen Rodriguez-Maya [1], Ruth Gil-Prieto [3] and Angel Gil-de-Miguel [3]

1 Internal Medicine Department, Mostoles University Hospital, Rey Juan Carlos University, 29835 Mostoles, Spain; govaz55@gmail.com (O.V.-G.); belenrmaya@gmail.com (B.R.-M.)
2 Hematology Department, Mostoles University Hospital, Rey Juan Carlos University, 29835 Mostoles, Spain
3 Department of Preventive Medicine and Public Health, Rey Juan Carlos University, 28933 Madrid, Spain; ruth.gil@urjc.es (R.G.-P.); angel.gil@urjc.es (A.G.-d.-M.)
* Correspondence: rgcarretero@salud.madrid.org (R.G.-C.); maria.ordonez@salud.madrid.org (M.O.-G.)
† These authors contributed equally to this work.

Abstract: Background: Although confirmed cases of infection with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) have been declining since late 2020 due to general vaccination, little research has been performed regarding the impact of vaccines against SARS-CoV-2 in Spain in terms of hospitalizations and deaths. **Objective**: Our aim was to identify the reduction in severity and mortality of coronavirus disease 2019 (COVID-19) at a nationwide level due to vaccination. **Methods**: We designed a retrospective, population-based study to define waves of infection and to describe the characteristics of the hospitalized population. We also studied the rollout of vaccination and its relationship with the decline in hospitalizations and deaths. Finally, we developed two mathematical models to estimate non-vaccination scenarios using machine learning modeling (with the ElasticNet and RandomForest algorithms). The vaccination and non-vaccination scenarios were eventually compared to estimate the number of averted hospitalizations and deaths. **Results**: In total, 498,789 patients were included, with a global mortality of 14.3%. We identified six waves or epidemic outbreaks during the observed period. We established a strong relationship between the beginning of vaccination and the decline in both hospitalizations and deaths due to COVID-19 in all age groups. We also estimated that vaccination prevented 170,959 hospitalizations (CI 95% 77,844–264,075) and 24,546 deaths (CI 95% 2548–46,543) in Spain between March 2021 and December 2021. We estimated a global reduction of 9.19% in total deaths during the first year of COVID-19 vaccination. **Conclusions**: Demographic and clinical profiles changed over the first months of the pandemic. In Spain, patients over 80 years old and other age groups obtained clinical benefit from early vaccination. The severity of COVID-19, in terms of hospitalizations and deaths, decreased due to vaccination. Our use of machine learning models provided a detailed estimation of the averted burden of the pandemic, demonstrating the effectiveness of vaccination at a population-wide level.

Keywords: COVID-19; vaccines; SARS-CoV-2; hospitalizations; mortality; machine learning

check for updates

Citation: Garcia-Carretero, R.; Ordoñez-Garcia, M.; Vazquez-Gomez, O.; Rodriguez-Maya, B.; Gil-Prieto, R.; Gil-de-Miguel, A. Impact and Effectiveness of COVID-19 Vaccines Based on Machine Learning Analysis of a Time Series: A Population-Based Study. *J. Clin. Med.* **2024**, *13*, 5890. https://doi.org/10.3390/jcm13195890

Academic Editor: Venerando Rapisarda

Received: 13 September 2024
Revised: 28 September 2024
Accepted: 30 September 2024
Published: 2 October 2024

1. Introduction

The coronavirus disease 2019 (COVID-19) pandemic has had a significant impact on the health of the population, as well as significant implications in all sectors of society and the daily lives of citizens [1–4]. It is claimed that high levels of vaccination coverage, the characteristics of the omicron variant, and increased diagnostic testing likely contributed to the observed impact of the pandemic in the last months of 2021. In addition, there was a very high incidence of confirmed cases, but a majority of these had mild symptoms or were asymptomatic. This placed a significant strain on primary health care rather than hospitals. Therefore, the occupancy percentage of hospital and intensive care unit (ICU) beds was much lower than expected relative to what occurred over the remainder of the pandemic [5–8].

By February 2022, more than 92% of the Spanish population over the age of 12 was fully vaccinated [9]. Current evidence indicates that the various COVID-19 vaccines have achieved high levels of effectiveness in restricting moderate and severe forms of the disease and reducing lethality. Vaccines, despite reducing the probability of infection, are less effective at completely preventing virus replication in the upper respiratory mucosa of a vaccinated individual, which means that transmission is possible from vaccinated individuals who have been infected, even if the disease is mild or asymptomatic [10–15]. This makes it infeasible to aim for the virus's eradication at present. Therefore, researchers should focus their efforts toward reducing the severity of infections while maintaining a level of transmission that is manageable and does not generate an excessive burden on the healthcare system.

As noted, due to the increase in vaccination coverage and the immunity generated from natural infections, the majority of the population is protected against severe COVID-19 [16]. Data show that protection has been maintained, even against a variant antigenically different enough from the previous ones to produce very high incidence rates in the population that previously had immunity.

Observational studies, such as case–control or cohort studies, are not always feasible, so several studies using alternative approaches have been conducted to demonstrate the effectiveness of vaccination [8,17]. Likewise, several meta-analyses have studied the effectiveness of vaccination from the following three perspectives: efficacy against infection, efficacy against severe disease (and, hence, reduction in risk of hospital admission), and ability to reduce the transmissibility of vaccinated individuals who become infected [13,15,16,18]. However, the impact of vaccination in terms of decreasing hospitalizations and deaths has not yet been investigated in a nationwide, population-based, epidemiological study in Spain.

Given the unique characteristics of the Spanish healthcare system and the country's age-stratified vaccination strategy, studying Spain offers an opportunity to understand the differential impact of vaccination across diverse demographic groups, contributing insights that are not directly generalizable to other populations.

Hypothesis and Objectives of Our Research

We designed a population-based study to assess vaccination as a major public health intervention. By this means, we investigated whether vaccines have been beneficial to the Spanish population. Our research objective was to determine whether vaccination reduced the number of hospitalizations and deaths. We conducted our study in three stages. First, we described the differences between two periods, namely the first months of the pandemic, during which no vaccination was present, and the last months of 2021, when a high proportion of the Spanish population was vaccinated. Secondly, we compared trends in hospitalizations and deaths with the vaccination rate. Finally, we assessed the effectiveness of vaccines against severe disease in terms of the reduction in hospitalizations and mortality due to COVID-19. We estimated the number of averted hospitalizations and deaths. We also compared the evolution of the pandemic across the following two scenarios: vaccination (the observed scenario) and non-vaccination (an estimated scenario). The estimated scenario was fitted using time series and machine learning analyses.

2. Materials and Methods

2.1. Data Collection and Study Design

We designed a retrospective, population-based study using data collected from electronic health records. We collected data from the Spanish Minimum Basic Data Set at Hospitalization (MBDS-H), provided by the Spanish Ministry of Health [19]. We also collected data related to COVID-19 vaccination in the European Union/European Economic Area (EU/EEA) from the European Centre for Disease Prevention and Control [20]. Figure A1 shows a flow chart of the study.

MBDS-H is a mandatory administrative registry of hospital discharges that covers more than 95% of Spanish hospitals, including public centers in the National Spanish Health System and private hospitals. Nearly 97% of total hospital discharges are covered in the database. The MBDS-H is exclusively built from discharge reports. Microdata from patients include information on sex, age, dates of admission and discharge, type of discharge, primary and secondary diagnoses at discharge, length of stay, and surgical or obstetric procedures, among other data. Other administrative data are recorded by default, including the province where the hospitalization occurred, place of residence, and cost of hospitalization. By default, the Ministry of Health provides de-identified data to ensure patient privacy; thus, no names or personal information were recorded. The purpose of the MBDS-H is to facilitate the development of retrospective studies for the calculation of the burden of hospitalization and assessment of risk factors from thousands of patients, i.e., enabling population-based studies. From 2016 onward, MBDS-H has used the coding system of the International Classification of Diseases, 10th edition. MBDS-H is considered a valuable system for the epidemiological analysis of any coded disease.

Vaccination data were downloaded from the European Centre for Disease Prevention and Control [20]. These data were collected through the European Surveillance System. All EU/EEA Member States are requested to report basic indicators on vaccination (vaccines categorized by manufacturer, number of doses administered, vaccinated population, etc.). Data are categorized by target and age group at a national level.

2.2. Inclusion and Exclusion Criteria

In this retrospective study, cases were collected from the MBDS-H from the Spanish Ministry of Health. We included all patients with the code for COVID-19 (U07.1) in any diagnostic position (either primary or secondary diagnosis) from 1 January 2020 to 31 December 2021.

All age groups were studied, with special emphasis on those older than 60 years of age. We analyzed the healthcare impacts in terms of mortality and ICU admission by dividing the population into age groups. Patients with incomplete data regarding ICU admission, mortality, length of stay, or COVID-19 disease were excluded. We excluded patients with unknown data to ensure the accuracy and completeness of the dataset. As length of stay is a key outcome variable in the analysis of disease severity and healthcare utilization, including patients with missing values could bias the results and reduce the robustness of our conclusions.

2.3. Definition of Waves

We categorized the pandemic following a previous epidemiological study [21]. Using only data from Spain, we split the entire pandemic period into outbreaks or epidemiological waves based on the 14-day cumulative incidence, which marked the turning point for each wave. Every turning point indicated the end of one wave and the beginning of the next one, similar to the methodology used in previous epidemiological studies [22].

As mentioned in the introduction, the first and second stages of our study were descriptive. We analyzed the evolution of the pandemic and its outbreaks, comparing the first waves, during which time vaccination was absent, with the last waves of 2021, when vaccination was present. Herein, we describe the demographic and epidemiological differences between the two periods and their relationships with vaccination.

2.4. Vaccination Rollout

The Vaccination Strategy Against COVID-19 in Spain was developed by the Spanish Ministry of Health [23]. The working group prioritized certain age groups to receive the vaccine based on the supply of doses and the availability of current evidence, taking into consideration the demographic characteristics of the Spanish population [24]. Assessments of ethical concerns and risk factors were also considered to prioritize certain age groups over others. The elderly and healthcare workers were the first groups to receive the vaccine, and the rollout moved forward through the rest of the age groups over the course of 2021. We assessed the trends of vaccination over time using data from the European Centre for Disease Prevention and Control.

We split the population into age groups to assess the evolution of the pandemic in terms of hospitalizations and deaths. Then, we compared the vaccination rates to those trends by age group.

2.5. Estimated Scenarios of the Unvaccinated Population

We developed a population-based, epidemiological study to compare the following two scenarios: observed hospitalizations and deaths before and after vaccination and an estimated non-vaccination scenario using time series and machine learning models.

2.6. Mathematical Modeling of Hospitalizations and Mortality

We utilized ElasticNet and random forest models to forecast the impact of vaccination by fitting the models to a training dataset from July 2020 to February 2021 and validated them using a testing set. Each model was evaluated using cross-validation metrics (see Appendix A).

2.7. Machine Learning Algorithms

The models included ElasticNet for linear predictions and random forest to capture nonlinear patterns. Each algorithm was tuned to achieve optimal performance. The former assumes linearity, and the latter makes no assumptions on linearity. Researchers and data scientists apply machine learning algorithms in various fields, including health care, finance, and natural language processing [25–27].

2.8. Fitting the Models

To fit the models, we first split our time-series dataset into a training and a testing set. The training set covers the period between 1 July 2020 and 29 February 2021. We excluded the first wave (March to June 2020) because we considered it an outlier that could add noise to the final model. The testing set was not used to develop the models but for comparison purposes only. We made no assumptions on the likelihood, normality, or linearity of the training set. We fit the models to the training set, tuning the hyperparameters of each algorithm to achieve the best accuracy. For EN, we set alpha, and for RF, we set mtry and the numbers of trees. A key mathematical condition when tuning the models was that they should fit accurately with the observed data, i.e., the training dataset.

We used R package *randomforest* for the RF model and R package *glmnet* for the EN model. We fitted the two models to time series of both hospitalizations and deaths. Thus, we computed four models. Once they were developed, we forecasted data for the next months, namely 1 March 2021 to 31 December 2021. Finally, by comparing the estimated hospitalizations and deaths (had vaccination never been implemented) with the observed data, we could explore the number of events that were averted in the last 10 months of 2021 due to vaccination.

2.9. Statistical Analyses

All statistical and machine learning-based analyses were conducted using R language version 4.3.2 (Vienna, Austria) [28]. Statistical significance was set at 0.05.

3. Results

3.1. Nationwide Overview of the Pandemic

We included data from 498,789 hospital admissions (Table 1) and excluded 113 patients due to inconsistent or incomplete data. We split the observation period into waves, as described above (Figure 1). The first waves included more than 115,000 hospitalizations, and this number dropped up to 50,000 in the fourth and fifth waves. Men were admitted at higher rates than women (56.1%, $p = 0.001$), with no changes in the distribution during the pandemic. The median age was 66, but this tended to decrease across the fourth and fifth waves (59 and 57, respectively). Length of stay in both the standard hospitalization ward and in the ICU was more heterogeneous, and no clear trend could be established. Although the nationwide mortality ratio was 14.3% in Spain, we observed a decreasing trend from the first wave (18.2%) to the fourth and fifth waves (7.4% and 10.1%, respectively). Comorbidities such as type 2 diabetes, hypertension, coronary disease, dementia, kidney disease, malignancy (either solid tumor or hematological malignancy), and chronic respiratory disease showed a decreasing trend starting with the fourth wave. Other comorbidities, such as liver and cerebrovascular diseases, showed no changes. Obesity and heart failure showed a more heterogeneous trend.

Table 1. Epidemiological and demographic characteristics of patients hospitalized in Spain between 2020 and 2021.

	All Waves	First	Second	Third	Fourth	Fifth	Sixth	*p* Value
Hospital admissions	498,789	115,356	127,114	126,623	51,006	54,570	24,120	0.001
Sex (men)	56.1%	56.2%	55.5%	56.2%	57.5%	55.3%	56.2%	0.001
Age, median (IQR)	66 (28)	69 (25)	68 (28)	69 (25)	59 (24)	57 (38)	65 (28)	0.001
Hospital stay, median (IQR)	8 (9)	11.9 (9)	8 (9)	12.1 (9)	7 (7)	10.4 (8)	6 (7)	0.001
ICU								
Admissions	54,354	10,218	13,302	14,441	7194	6941	2258	0.001
ICU (%)	10.9	8.9	10.5	11.4	14.1	12.7	9.4	0.001
ICU stay, median (IQR)	10 (21)	11 (22.3)	10 (21.4)	11 (21.5)	11 (21.4)	10 (18.7)	6 (9.1)	0.001
Mortality								
Deaths	71,437	21,037	18,229	20,400	3793	5487	2491	0.001
Mortality rate (%)	14.3	18.2	14.3	16.1	7.4	10.1	10.3	0.001
Comorbidities								
Type 2 diabetes	21.7	20.6	23.2	24.3	18.3	18.3	21.4	0.001
Hypertension	33.8	35.2	34.7	37.2	31.1	25.0	31.6	0.001
AMI	7.1	7.3	7.4	7.9	5.3	5.7	7.5	0.001
CHF	7.2	6.5	7.8	8.4	4.2	6.8	7.9	0.001
Dementia	4.8	5.5	5.4	5.2	2.1	3.9	3.6	0.001
Kidney disease	10.5	10.3	11.2	11.7	6.5	9.7	11.1	0.001
Liver disease	0.4	0.4	0.4	0.5	0.3	0.4	0.5	0.001
Malignancy	5.6	5.4	5.8	6.0	4.2	5.3	7.4	0.001
Obesity	12.9	9.2	12.6	14.0	17.0	14.4	13.8	0.001
COPD	7.4	7.6	7.4	7.9	5.9	6.6	8.8	0.001
CEVD	0.7	0.7	0.8	0.9	0.5	0.6	0.8	0.001

ICU: intensive care unit; AMI: acute myocardial infarction; CHF: congestive heart failure; CEVD: cerebrovascular disease; COPD: chronic obstructive pulmonary disease; IQR: interquartile range. Age is expressed in years. Hospital/ICU stay is expressed in days.

Table 2 shows disaggregated data of hospital admissions, ICU admissions, and mortality by age group. These data are also represented in Figure A2.

Among hospitalizations, the predominant age group were >60 years old in the second and third waves. The >80-year-old age group dropped dramatically in the fourth and fifth waves, and the group of 18- to 49-year-olds was predominant in the fifth wave. Regarding mortality, deaths in all ages dropped quite evenly, although the patients who were more affected were >60 years old (Figure 2). Figure 2 displays data beginning with the second wave, as details of the following waves are of interest to compare the second and third waves on one side with the fourth and fifth waves on the other.

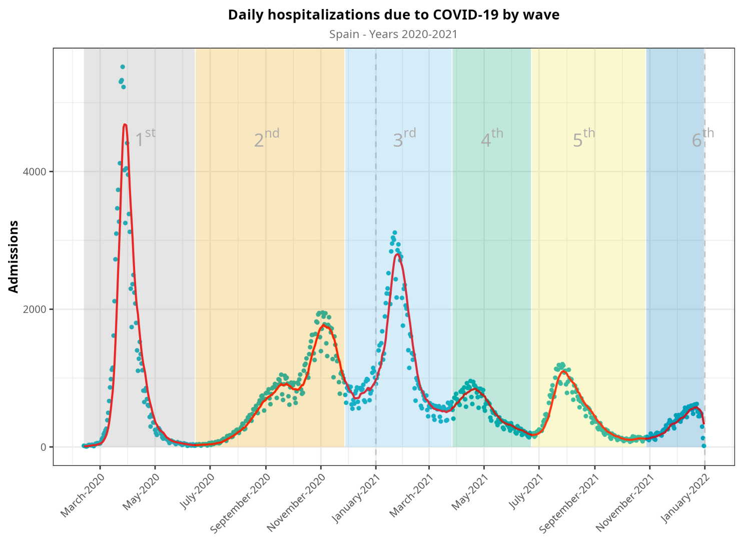

Figure 1. Evolution of hospitalizations during the COVID-19 pandemic in Spain from March 2020 to December 2021, split into waves. Blue dots represent raw data, whereas the red line represents a 7-day moving average of the time series. Vertical dash line at the end of the figure denotes that there were no available data for the sixth epidemic wave (it would continue, otherwise).

Table 2. Outcomes in terms of total admissions, ICU admissions, and mortality of hospitalized patients disaggregated by age group.

	Total	First Wave		Second Wave		Third Wave		Fourth Wave		Fifth Wave		Sixth Wave		*p*-Value
Admissions														
≤17	6568	583	(8.9%)	1622	(24.7%)	1021	(15.5%)	657	(10%)	1764	(26.9%)	921	(14%)	0.001
18–49	99,570	18,223	(18.3%)	23,525	(23.6%)	18,542	(18.6%)	13,669	(13.7%)	20,660	(20.7%)	4951	(5%)	0.001
50–59	78,535	18,524	(23.6%)	19,534	(24.9%)	19,432	(24.7%)	11,560	(14.7%)	6262	(8%)	3223	(4.1%)	0.001
60–79	178,667	45,100	(25.2%)	44,052	(24.7%)	48,808	(27.3%)	18,198	(10.2%)	13,267	(7.4%)	9242	(5.2%)	0.001
≥80	125,834	30,642	(24.4%)	35,998	(28.6%)	36,365	(28.9%)	5806	(4.6%)	11,707	(9.3%)	5316	(4.2%)	0.001
ICU														
≤17	490	95	(19.4%)	109	(22.2%)	74	(15.1%)	62	(12.7%)	111	(22.7%)	39	(8%)	0.001
18–49	10,227	1493	(14.6%)	2186	(21.4%)	1,957	(19.1%)	1513	(14.8%)	2633	(25.7%)	445	(4,4%)	0.001
50–59	11,187	2117	(18.9%)	2733	(24.4%)	2994	(26.8%)	1705	(15.2%)	1240	(11.1%)	398	(3.6%)	0.001
60–79	28,548	5862	(20.5%)	7159	(25.1%)	8300	(29.1%)	3537	(12.4%)	2526	(8.8%)	1164	(4.1%)	0.001
≥80	2372	342	(14.4%)	754	(31.8%)	687	(29%)	185	(7.8%)	248	(10.5%)	156	(6.6%)	0.001
Deaths														
≤17	27	4	(14.8%)	7	(25.9%)	6	(22.2%)	2	(7.4%)	6	(22.2%)	2	(7.4%)	0.001
18–49	1282	345	(26.9%)	300	(23.4%)	253	(19.7%)	92	(7.2%)	222	(17.3%)	70	(5.5%)	0.001
59–59	3274	905	(27.6%)	760	(23.2%)	861	(26.3%)	277	(8.5%)	343	(10.5%)	128	(3.9%)	0.001
60–79	25,427	7937	(31.2%)	5,914	(23.3%)	7074	(27.8%)	1871	(7.4%)	1733	(6.8%)	898	(3.5%)	0.001
≥80	40,855	11,707	(28.7%)	11,111	(27.2%)	12,041	(29.5%)	1499	(3.7%)	3127	(7.7%)	1370	(3.4%)	0.001

Mortality ratios were calculated by dividing the the presented events in a given age group in a wave by the total population in each group.

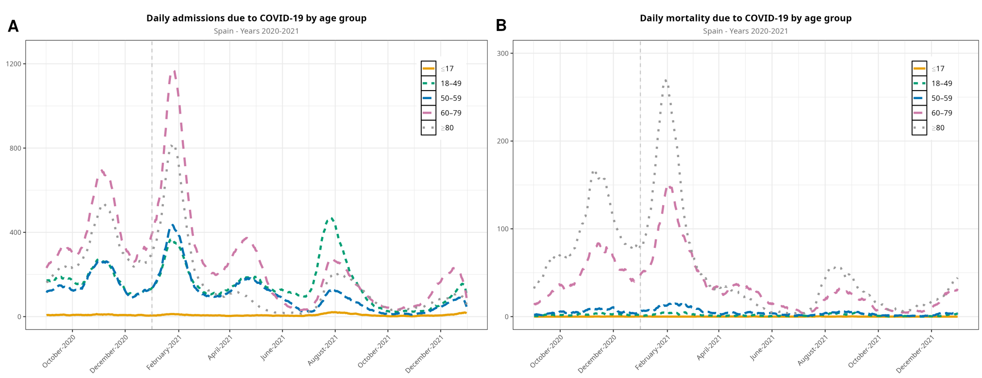

Figure 2. Evolution of the COVID-19 pandemic in terms of hospitalizations (**A**) and in-hospital deaths (**B**) between September 2020 and December 2021 (the first wave is not included in the visualization), disaggregated by age group. The vertical dashed line denotes the new year.

3.2. Vaccination Rollout

Figure 3 plots the vaccination rollout in Spain, both globally and by age group. Vaccination began in December 2020 with the elderly and healthcare workers. By 31 December 2021, 80.3% of the whole Spanish population was fully vaccinated, i.e., having received the complete regimen, and 97.2% had received at least one dose. By April 2021, 75% of >80-year-olds and 48% of >60-year-olds were fully vaccinated. Figure A3 provides more detail on age groups regarding vaccination coverage over time.

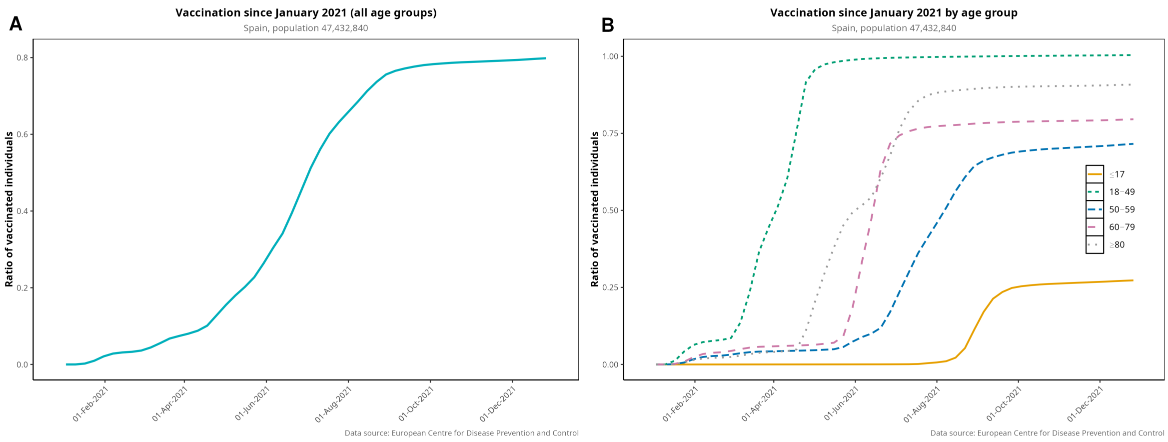

Figure 3. Vaccination rollout in Spain for the entire population (i.e., fully vaccinated individuals (**A**)) disaggregated by age group (**B**). Elderly patients were prioritized to receive the first dose of vaccine.

3.3. Hospitalizations and Deaths in an Estimated Scenario

As noted, our approach involved estimating both hospital admissions and mortality by parsing the time series using machine learning algorithms. We developed four models—one for hospitalizations and another for deaths—using each algorithm. Figure 4 shows the observed and estimated scenarios. Figure A4 shows the models with confidence intervals. Table 3 shows the estimates of hospitalizations and deaths in the absence of vaccination. Using the RF model, we estimated that 251,830 hospitalizations and 37,673 deaths would have occurred in a non-vaccination scenario during the period between March and December 2021. According to the EN model, the estimated numbers of hospitalizations and deaths were 307,617 and 37,141, respectively. Compared to the observed data, we estimated that vaccination prevented 115,172 hospitalizations and 25,078 deaths with the RF model and 170,959 hospitalizations and 24,546 deaths with the EN model. Finally, we plotted Figure 5, showing the cumulative hospitalizations and deaths, with both the RF and EN models.

Table 3. Observed, estimated, and averted events in the first year of vaccination.

	Hospitalizations		**Deaths**	
	Events	**(95% CI)**	**Events**	**(95% CI)**
RandomForest				
Observed	136,658	(NA)	12,595	(NA)
Estimated	251,830	(216,99–286,663)	37,673	(317,13–43,633)
Averted	115,172	(80,339–150,005)	25,078	(191,18–31,038)
ElasticNet				
Observed	136,658	(NA)	12,595	(NA)
Estimated	307,617	(214,502–400,733)	37,141	(15,143–59,138)
Averted	170,959	(77,844–264,075)	24,546	(2,548–46,543)

Estimation computed for the period between March, 2021 and December, 2021. We estimated data using the following two models: ElasticNet and random forest. NA: non-applicable; CI: confidence interval.

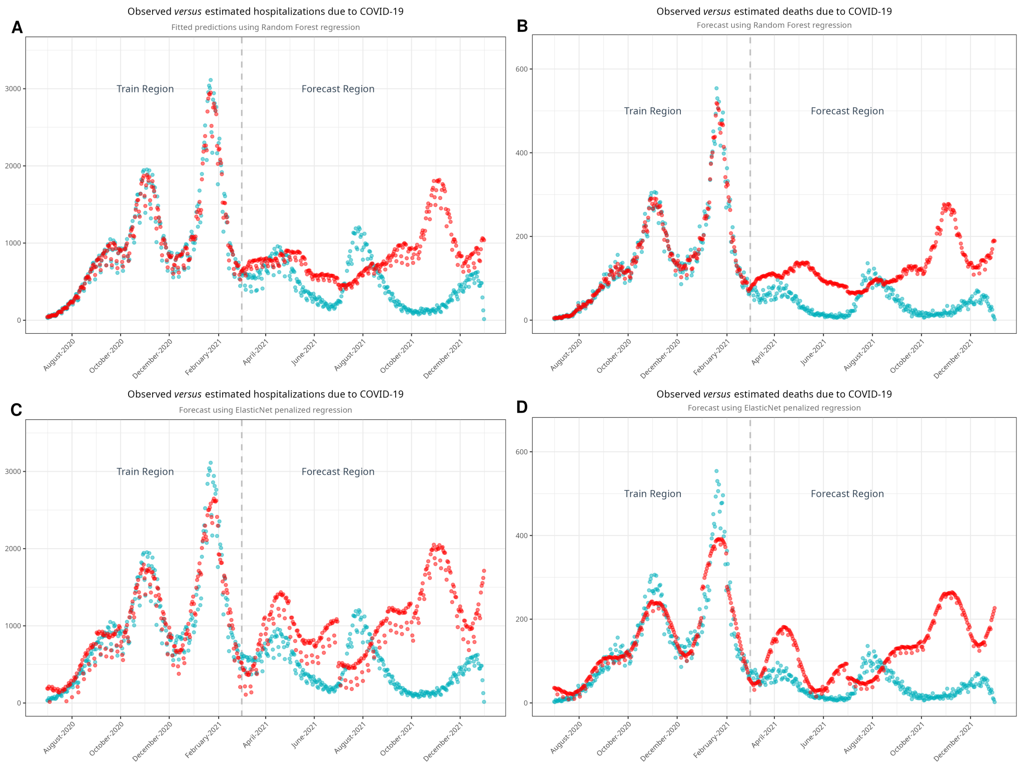

Figure 4. Models developed with random forest (**A**,**B**) and with ElasticNet (**C**,**D**) to estimate non-vaccination scenarios. Turquoise dots represent the observed values, while red dots represent estimated values. Note the good fit of the model in the train region before it estimates the values in the forecast region.

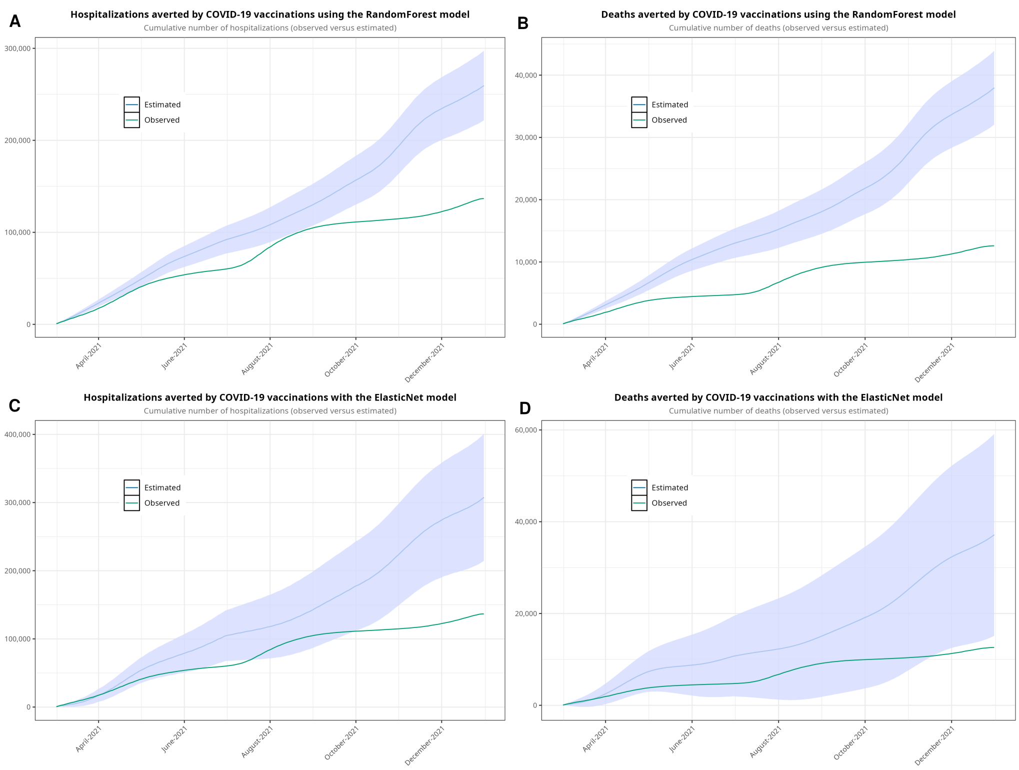

Figure 5. Cumulative sum estimated with random forest (**A**,**B**), showing the observed values and the estimates with 95% confidence intervals, and with ElasticNet (**C**,**D**). We forecasted values from 1 March to 31 December 2021.

4. Discussion

4.1. Descriptive Analyses

We have described the high number of hospitalizations and deaths during the first waves of the pandemic in Spain. We have demonstrated how this trend began to decrease in March to April 2021 as a result of vaccination, which was the major public health intervention during the COVID-19 pandemic.

Overall, the first wave showed the highest number of hospitalizations, the highest mortality rate, the longest hospital and ICU lengths of stay, and the oldest patients. The fourth and the fifth waves showed a decreasing trend in terms of hospitalizations and mortality. In addition, these last waves showed an overall younger, healthier population.

While the sixth wave was included in our analyses, the results might not be reliable, as this wave ended mid-February 2022, and its results are not fully represented in tables and figures. However, the fourth and fifth waves showed that the demographic profile of hospitalized individuals changed with respect to the previous waves, showing a turning point in the evolution of the pandemic.

With respect to admissions by age group, the group of patients under 17 contributed only marginally during the observed period of the pandemic. However, patients over 60 years old were the largest group of those admitted to the hospital due to COVID-19. Patients between 18 and 59 years old were hospitalized at a lower rate. Additionally, most of the deaths occurred in patients >60 years old, particularly in patients over 80 years old, whereas mortality in the rest of the age groups was marginal, as seen in Figure 3.

4.2. Vaccination Rollout

Vaccination in Spain began in late December 2020, as soon as vaccines were proven to be safe and to offer significant protection against severe forms of COVID-19, as part of a European initiative [29]. Within only a few weeks of the beginning of vaccination (2.2% of total population by March and 10% by May 2021), we observed a rapid decline in both hospitalizations and deaths beginning in March and April 2021. We also observed a strong temporal correlation between decreasing hospitalizations and deaths on one side and the evolving vaccination rollout on the other (Figures 2 and 3). The decline in hospitalizations and deaths was first observed in patients over 80 years old, showing a relationship between vaccination and protection against both outcomes. This relationship can also be seen in the remaining age groups after the beginning of vaccination. This steady pace of vaccination consolidated the decline in the severity of the pandemic. Our data are also in line with other studies that have investigated the benefit of vaccination and its protective effects [30–32]. We can state that in Spain, vaccination led to a significant reduction in the severity of COVID-19 across all age groups, with particularly marked benefits observed in the elderly population. While the overall reduction in hospitalizations and deaths due to vaccination is consistent with global findings, the timing and magnitude of these changes in Spain were influenced by the country's specific vaccination rollout strategy and healthcare infrastructure.

4.3. Modeling and Estimating Data in a Non-Vaccination Scenario

It can be challenging to quantify the impact of vaccination if an incomplete picture of the pandemic is obtained. Infections and confirmed cases are either often under-reported or underestimated [21,33]. For this reason, we relied on reported hospitalizations and deaths to determine this impact instead of trends of non-hospitalized, confirmed cases.

Our reference publication was that of Barandalla et al. [34], who developed simulated curves of hospitalization in the absence of vaccines and compared those curves with the observed incidence. That study investigated hospitalizations in Spain between February 2020 and June 2021. The authors estimated the expected hospitalizations during 2021 in the absence of vaccination, extrapolating data from the second wave. The scenario of an unvaccinated population was estimated to create a statistical model as follows. The authors disaggregated the entire population curve across age groups and took the proportion of hospitalization of age groups of unvaccinated or less vaccinated population as a reference. These proportions of hospitalizations were extrapolated to the remaining groups, yielding curves of the real incidence of hospitalization and curves of expected hospitalization in the absence of vaccines for each age group. Finally, these curves were compared. Showing a decrease in incidence, they demonstrated the beneficial impact of vaccination on hospitalizations. Likewise, vaccine effectiveness against hospitalization in $\geq$65-year-old age groups was estimated from October 2021 to March 2022 in a European study [35]. The reference group was the unvaccinated population. The authors performed a survival analysis using the Cox proportional hazards regression model to estimate the hazard ratios of hospitalization.

It is beyond the scope of this manuscript to discuss all studies that have used mathematical models to estimate mortality in the absence of vaccination, but it is worth mentioning some of them. A mathematical model reported by Watson et al. [17] estimated that 14.4 million deaths were prevented in 185 countries in 2021. The authors used a framework based on a "susceptible–exposed–infections–recovered–susceptible" model to estimate a non-vaccination scenario. This model was fitted using MCMC, and the authors calculated the time-varying reproductive number to determine the estimated number of contagions. Havers et al. [8] conducted a cross-sectional study that included adults hospitalized with COVID-19, comparing vaccinated versus unvaccinated individuals. Both studies demonstrated the effectiveness of vaccination and its impact on the evolution of the COVID-19 pandemic using different mathematical approaches. Auto-regressive time-series modeling was assessed in other studies [36,37].

In summary, previous studies conducted in Spain [34] have reported a significant reduction in hospitalizations following vaccination rollout using different modeling approaches. Our study adds to these findings by incorporating machine learning methods and providing a more granular age-stratified analysis, which is lacking in other reports. International studies, such as that by Watson et al. [17], have demonstrated similar impacts of vaccination on a nationwide scale, supporting the observed trends in our Spanish cohort.

Machine learning has also been used to estimate the evolution of the COVID-19 pandemic in terms of confirmed cases. Kırbaş et al. [38] conducted a comparative study using different approaches, including ARIMA, neural networks, and long short-term memory (LSTM), to forecast the evolution of the pandemic. LSTM provided predictions with the best accuracy. Neural networks were used by Nabi et al. to study the dynamics of confirmed cases of COVID-19 [39]. Although deep learning (i.e., neural networks) seems to have better prediction accuracy than standard statistical methods (ARIMA) or machine learning, it entails high costs in terms of computational resources and time [40].

Having discussed some of the more relevant publications in this field, we consider this study to stand out due to the use of advanced machine learning algorithms such as ElasticNet and random forest, which allowed us to create accurate non-vaccination scenarios. This approach enabled us to quantify the impact of vaccination with high precision. Furthermore, by analyzing hospitalization and mortality trends across multiple age groups, we have provided a more granular understanding of how vaccination influenced disease severity in different demographic subpopulations within Spain. Such detailed insights have not been previously reported in similar population-based studies.

Vaccination conferred sufficient protection against severe disease and altered the course of the COVID-19 pandemic. Given the conditions of the pandemic, measuring the impact of vaccination directly by comparing a vaccination scenario with a non-vaccination scenario was not possible at a nationwide level. This is why mathematical models are useful for estimating non-vaccination scenarios to achieve such comparisons. Thanks to our estimated scenarios, we could assess the impact of vaccination in Spain. Our approaches generated estimations of hospitalizations and deaths averted as a result of vaccination against SARS-CoV-2.

4.4. Limitations

We estimated how the pandemic would have evolved if no vaccines had been available by estimating a new scenario, but we did not include non-pharmaceutical interventions, viral variants, or limitations on mobility that could have altered the viral evolution in the absence of vaccination. It is of interest to mention that the last waves of 2021 in Spain, which were primarily caused by the omicron variant and its descendants (B.1.1.529), presented different characteristics than the previous waves [21,33], but its impact was not included in our model. In addition, forecasting using the time-series signature can be very accurate, particularly when time-based patterns are present in the underlying data. As with most uses of machine learning, the prediction is only as good as the patterns in the data. Forecasting using this approach may not be suitable when patterns are not present or when the future is highly uncertain (i.e., past results are not a suitable predictor of future performance). We could not use ARIMA or MCMC to create the estimated scenario, so we did not compare different approaches. Although it has been found that mortality due to COVID-19 may have been under-reported [41], in Spain, almost all deaths occurred in hospital, so our data can be considered reliable. This is key when modeling and fitting a machine learning algorithm because the final model can only be as good as the provided data. The wide confidence intervals, particularly for the ElasticNet model, reflect the inherent uncertainty in modeling complex phenomena such as pandemic outcomes. This uncertainty arises from potential changes in transmission dynamics, population behavior, and viral variants. While the confidence intervals indicate variability, the consistency in trend direction across models (ElasticNet and RandomForest) suggests that the main conclusions remain robust despite this uncertainty.

5. Conclusions

We fit mathematical models to estimate both hospitalizations and deaths due to COVID-19 in a non-vaccination scenario. We determined the impact of vaccination by estimating the hospitalizations and deaths that, otherwise, could have occurred if vaccines had not been administered. In Spain, demographic and clinical profiles shifted significantly during the first months of the pandemic, reflecting the differential impact of early vaccination efforts. Vaccination altered the evolution of the COVID-19 pandemic and prevented up to 24,546 deaths in Spain in 2021. Vaccination reduced not only mortality but also the number of hospitalizations and the burden of the pandemic. Its protective effect was observable shortly after the beginning of vaccination for each age group. Machine learning approaches can be useful in uncertain contexts because a time-series signature can provide accurate forecasts. By integrating machine learning models and age-stratified analyses, our study provides a comprehensive view of the pandemic's evolution in Spain, demonstrating how targeted vaccination strategies can alter disease trajectories at a national level.

Author Contributions: Conceptualization, R.G.-C. and M.O.-G.; methodology, R.G.-C. and O.V.-G.; software, R.G.-C.; validation, R.G.-C. and M.O.-G.; formal analysis, R.G.-C. and B.R.-M.; investigation, R.G.-C. and M.O.-G.; data curation, R.G.-C. and B.R-M.; writing—original draft preparation, R.G.-C. and O.V.-G.; writing—review and editing, R.G.-P. and A.G.-d.-M.; visualization, R.G.-C.; supervision, R.G.-P. and A.G.-d.-M.; project administration, R.G.-P. and A.G.-d.-M.; funding acquisition, R.G.-C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was approved by the Ethical Board of Universidad Rey Juan Carlos (ID number 2610202334423). No identifying information was included in the manuscript. Because the authors used historical data, informed consent was not necessary. All procedures involving human participants were conducted in accordance with the ethical standards of the responsible institutional and/or national research committee and the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards.

Informed Consent Statement: Not applicable.

Data Availability Statement: A contract signed with the Spanish Health Ministry, which provided the dataset, prohibits the authors from providing their data to any other researcher. Furthermore, the authors must destroy the database upon the conclusion of their investigation. The database cannot be uploaded to any public repository.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ARIMA	Auto-Regressive Integrated Moving Average
COVID-2019	Coronavirus Disease 2019
EN	ElasticNet
EU/EEA	European Union/European Economic Area
LSTM	Long Short-Term Memory
MBDS-H	Minimum Basic Data Set at Hospitalization
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
MCMC	Markov Chain with Monte Carlo
RF	Random Forest

Appendix A. Descriptive Analyses

Appendix A.1. Estimated Scenarios of the Unvaccinated Population

Designing research to investigate the effectiveness of vaccination in terms of reducing hospitalizations is a valuable endeavor but can be challenging. As noted, our objective was to determine whether vaccination reduced the number of hospitalizations and deaths. There are two common options for the design of clinical research on vaccines. The first approach is a retrospective cohort study, in which patients are divided into two groups (e.g., vaccinated and unvaccinated). The other design is a matched case–control study, in which a subset of vaccinated patients are matched with an equal number of unvaccinated patients based on relevant characteristics. In either design, the aim is to compare the hospitalization and death rates and clinical characteristics between the two groups. However, due to a lack of data, traditional research designs such as cohort and case–control studies would not be feasible in our context.

Thus, we designed a population-based, epidemiological, nationwide study to compare the following two scenarios: the observed scenario describing actual hospitalizations and deaths before and after vaccination and an estimated scenario simulating the trends of the pandemic had vaccination not occurred in 2021. The first scenario indicates that the outcomes in the first months of the pandemic were widely different from those of the last months of 2021 (with vaccination). Finally, we compared the two scenarios to extract the estimated effect of vaccination.

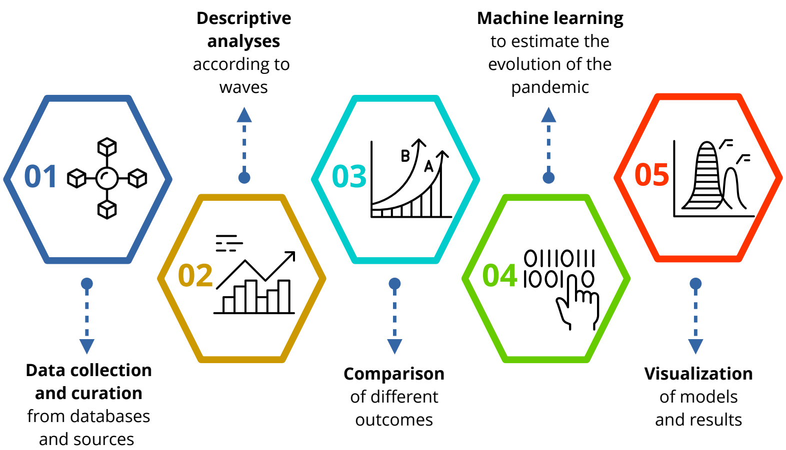

Figure A1. Flow chart and study design.

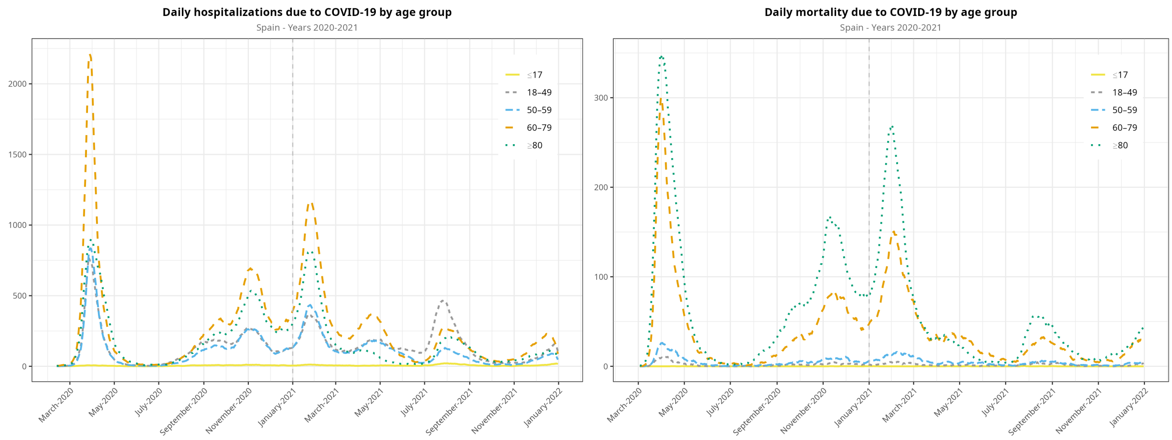

Figure A2. Evolution of the COVID-19 pandemic in terms of hospitalizations and deaths. All waves from the observation period are included.

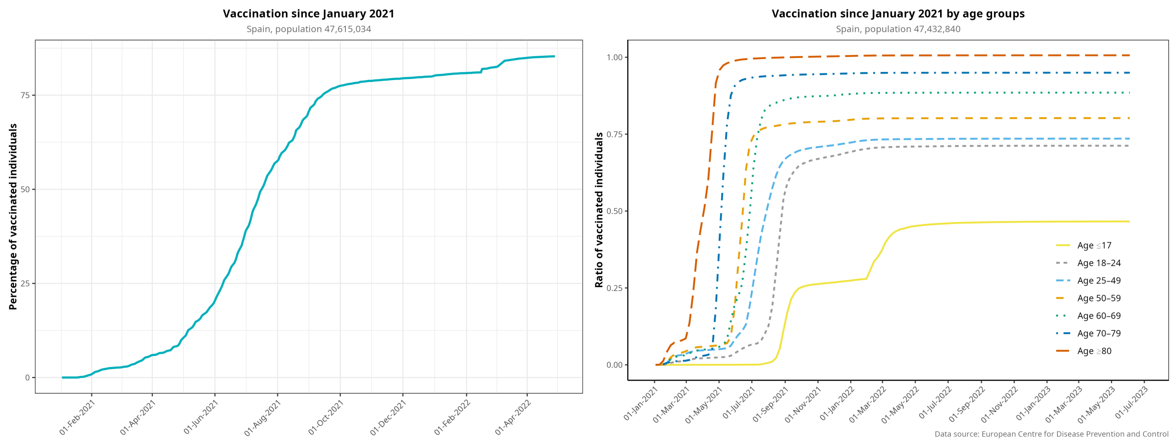

Figure A3. Vaccination rollout in Spain for the entire population (fully vaccinated individuals), disaggregated by age group. Elderly patients were prioritized for the first dose of the vaccine.

Appendix A.2. Mathematical Modeling of Hospitalizations and Mortality

First, we transformed our dataset into a time series. Among the methods used to analyze time series, traditional statistical models such as auto-regressive (AR) models can be specified as linear regressions on the lags of the time series. For example, an AR model only looks at the relationship between lags of a series and its future values. Seasonality and trend are key components of a given time series. While the trend represents a gradual change in the data, depicting long-term growth or decline, seasonality describes the short-term patterns that occur within a single unit of time and repeats indefinitely. Another useful technique is Markov Chain with Monte Carlo (MCMC) simulation, which describes dynamic changes based on the state of a given value and the chance of its transition. The MCMC algorithm explores the parameter space to find values that maximize the likelihood of the observed data.

Once historical data on the number of deaths from COVID-19 were collected, we prepared the data for analysis by checking for missing values, outliers, and inconsistencies. Then, we aggregated data on daily hospitalizations and daily deaths into appropriate time intervals obtain a time series. Exploratory analyses were used to understand the trends and patterns in the historical data. This involved creating visualizations and summary statistics to identify any seasonality or trends in the occurrence of COVID-19.

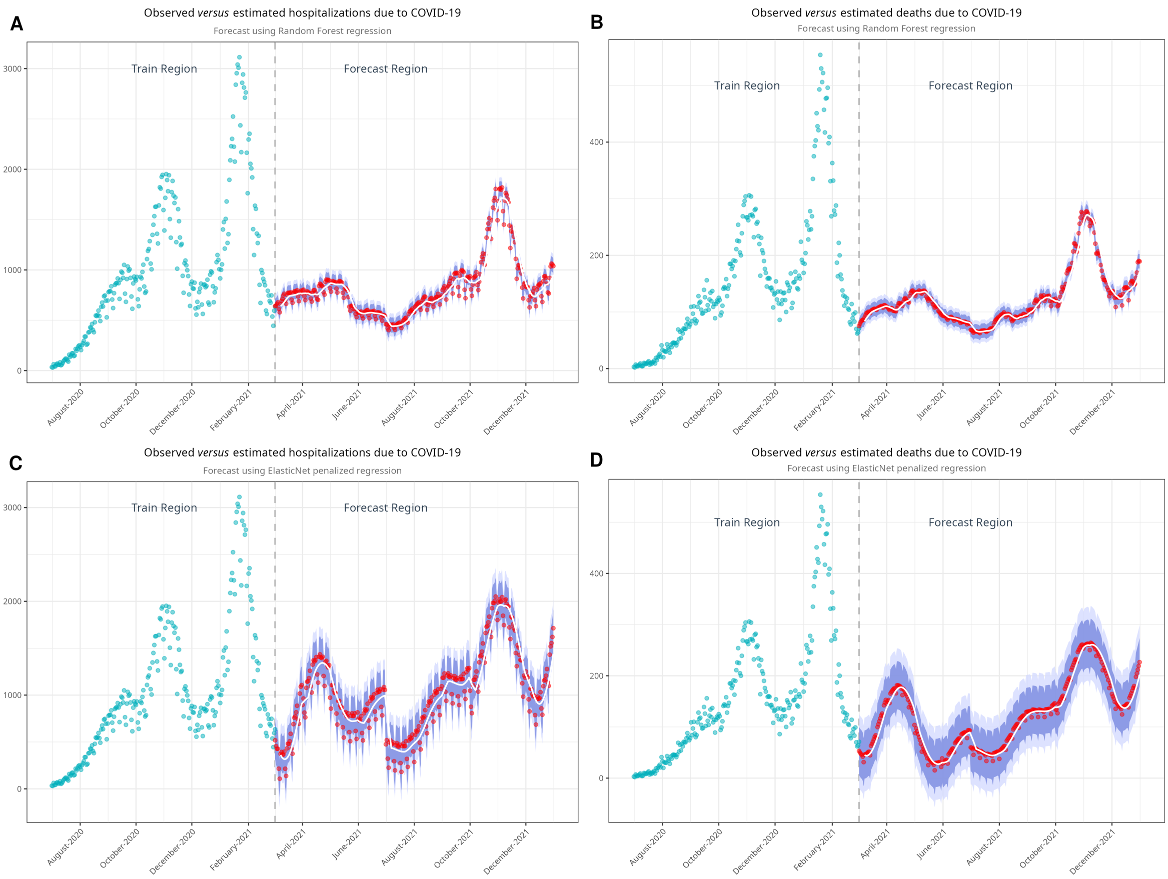

Figure A4. Models developed with random forest (**A**,**B**) and with ElasticNet (**C**,**D**) to estimate non-vaccination scenarios. Blue dots represent the observed values, while the smooth curves represent the estimated values and prediction intervals to account for variance between the model predictions and the observed data.

With a clean dataset, the next stage involved selecting a forecasting method. Estimating population data, specifically the number of events due to COVID-19, is a common but challenging task in epidemiology and clinical research. To make such predictions, statistical methods and models can be used to forecast future trends, but choosing an appropriate forecasting method depends on the characteristics of the data. More specifically, the choice of method depends on the complexity of the data and the availability of relevant predictor variables. As noted, common methods include time-series analysis, regression analysis, and machine learning techniques. We discarded time-series analysis methods such as Auto-Regressive Integrated Moving Average (ARIMA) because we were unable to capture and project time-dependent patterns in the data, given their nature. Specifically, ARIMA estimations did not converge in our dataset, likely because the developed model was not a good fit for the data. Likewise, MCMC required some assumptions that could not be fulfilled. The characteristics of our data and the underlying dynamics of COVID-19 hospitalizations did not justify the choice of the normal distribution and the assumption of independence in time steps. In addition, adjustments and the fine tuning of some parameters based on the likelihood of our data and the prior distributions were either too complex or unavailable.

Appendix A.3. Machine Learning Algorithms

We employed two algorithms, namely ElasticNet (EN) and random forest (RF). The former assumes linearity, and the latter makes no assumptions on linearity. We used two machine learning algorithms because these can capture non-obvious (both linear and nonlinear) patterns in data. We evaluated each model's performance using appropriate metrics, such as mean absolute percentage error, through cross-validation to ensure reliability for the training period (July 2020 to February 2021). Then, we used each model to forecast future hospitalizations and deaths in the population for the desired time period (March to December 2021). We created point forecasts (single estimates) and prediction intervals (confidence intervals) to quantify the uncertainty in our predictions.

EN is a machine learning technique used for linear regression and feature selection. It combines two regularization methods, namely L1 and L2 regularization [42]. L1 encourages some feature coefficients to be exactly zero, effectively performing feature selection by eliminating less important features. L2, on the other hand, penalizes large coefficients and prevents overfitting. EN strikes a balance between these two regularization techniques by introducing a hyperparameter that controls the mix of L1 and L2 penalties. This hyperparameter, often denoted as alpha, allows one to adjust the level of feature selection and regularization. A value of alpha equal to 0 corresponds to L2 regularization, while a value of 1 corresponds to L1 regularization. Any value in between blends the characteristics of both methods. EN is valuable when dealing with datasets containing many features, as it helps prevent overfitting and can automatically select the most relevant features. It is commonly used in predictive modeling, where the goal is to create accurate models that generalize well to new data while optimizing feature usage. Researchers and data scientists apply EN in various fields, including health care, finance, and natural language processing [25–27].

RF is a powerful machine learning technique used in various fields, including clinical research and engineering. It is essentially a collection of decision trees, where each tree is a simple predictive model [43,44]. It uses different subsets of the available data and features to create each decision tree. What sets RF apart is its random nature. This randomness injects diversity into models. By combining predictions from multiple trees, RF models become robust and less prone to overfitting, which makes them excellent at making generalizations from data. For researchers, RF can be used to make predictions based on complex, multidimensional data. It is well-suited to handle both numerical and categorical data, which is key in fields such as health care, where patient information can include a mix of variables. Clinicians and engineers find RF useful for various applications [45,46], such as disease prediction, image analysis, and quality control in manufacturing. RF models are known for their versatility, reliability, and ability to produce accurate and interpretable results, which makes them a valuable tool for decision support and pattern recognition.

References

1. Hosseinzadeh, P.; Zareipour, M.; Baljani, E.; Moradali, M. Social Consequences of the COVID-19 Pandemic. A Systematic Review. *Investig. Educ. Enferm.* **2022**, *40*, e10. [CrossRef] [PubMed]
2. Mofijur, M.; Rizwanul Fattah, I.M.; Alam, M.A.; Islam, A.B.M.S.; Ong, H.C.; Rahman, S.M.A.; Najafi, G.; Ahmed, S.F.; Uddin, M.A.; Mahlia, T.M.I. Impact of COVID-19 on the social, economic, environmental and energy domains: Lessons learnt from a global pandemic. *Sustain. Prod. Consum.* **2021**, *26*, 343–359. [CrossRef] [PubMed]
3. Osofsky, J.; Osofsky, H.; Mamon, L. Psychological and social impact of COVID-19. *Psychol. Trauma Theory Res. Pract. Policy* **2020**, *12*, 468–469. [CrossRef] [PubMed]
4. Saladino, V.; Algeri, D.; Auriemma, V. The Psychological and Social Impact of COVID-19: New Perspectives of Well-Being. *Front. Psychol.* **2020**, *11*, 577684. [CrossRef] [PubMed]
5. Patel, B.; Murphy, R.E.; Karanth, S.; Shiffaraw, S.; Peters, R.M., Jr.; Hohmann, S.F.; Greenberg, R.S. Surge in Incidence and Coronavirus Disease 2019 Hospital Risk of Death, United States, September 2020 to March 2021. *Open Forum Infect Dis* **2022**, *9*, ofac424. [CrossRef]
6. Delahoy, M.; Ujamaa, D.; Whitaker, M.; O'Halloran, A.; Anglin, O.; Burns, E.; Cummings, C.; Holstein, R.; Kambhampati, A.K.; Milucky, J.; et al. Hospitalizations Associated with COVID-19 Among Children and Adolescents—COVID-NET, 14 States, March 1, 2020–August 14, 2021. *MMWR Morb. Mortal. Wkly. Rep.* **2021**, *70*, 1255–1260. [CrossRef]
7. Taylor, C. COVID-19–Associated Hospitalizations Among Adults During SARS-CoV-2 Delta and Omicron Variant Predominance, by Race/Ethnicity and Vaccination Status—COVID-NET, 14 States, July 2021–January 2022. *MMWR Morb. Mortal. Wkly. Rep.* **2022**, *71*, 466–473. [CrossRef]
8. Havers, F.P.; Pham, H.; Taylor, C.A.; Whitaker, M.; Patel, K.; Anglin, O.; Kambhampati, A.K.; Milucky, J.; Zell, E.; Moline, H.L.; et al. COVID-19-Associated Hospitalizations Among Vaccinated and Unvaccinated Adults 18 Years or Older in 13 US States, January 2021 to April 2022. *JAMA Intern. Med.* **2022**, *182*, 1071–1081. [CrossRef]

9. de Sanidad, M. Situación actual Coronavirus. Available online: https://www.sanidad.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov/situacionActual.htm (accessed on 24 January 2024).
10. Stefanelli, P.; Trentini, F.; Petrone, D.; Mammone, A.; Ambrosio, L.; Manica, M.; Guzzetta, G.; d'Andrea, V.; Marziano, V.; Zardini, A.; et al. Tracking the progressive spread of the SARS-CoV-2 Omicron variant in Italy, December 2021 to January 2022. *Euro Surveill.* **2022**, *27*, 2200125. [CrossRef]
11. Assessment of the Further Spread and Potential Impact of the SARS-CoV-2 Omicron Variant of Concern in the EU/EEA, 19th Update. 2022. Available online: https://www.ecdc.europa.eu/en/publications-data/covid-19-omicron-risk-assessment-further-emergence-and-potential-impact (accessed on 15 October 2023).
12. Markov, P.; Ghafari, M.; Beer, M.; Lythgoe, K.; Simmonds, P.; Stilianakis, N.; Katzourakis, A. The evolution of SARS-CoV-2. *Nat. Rev. Microbiol.* **2023**, *21*, 361–379. [CrossRef]
13. Soheili, M.; Khateri, S.; Moradpour, F.; Mohammadzedeh, P.; Zareie, M.; Mortazavi, S.M.M.; Manifar, S.; Kohan, H.G.; Moradi, Y. The efficacy and effectiveness of COVID-19 vaccines around the world: A mini-review and meta-analysis. *Ann. Clin. Microbiol. Antimicrob.* **2023**, *22*, 42. [CrossRef] [PubMed]
14. Harder, T.; Koch, J.; Vygen-Bonnet, S.; Külper-Schiek, W.; Pilic, A.; Reda, S.; Scholz, S.; Wichmann, O. Efficacy and effectiveness of COVID-19 vaccines against SARS-CoV-2 infection: Interim results of a living systematic review, 1 January to 14 May 2021. *Eurosurveillance* **2021**, *26*, 2100563. [CrossRef] [PubMed]
15. Graña, C.; Ghosn, L.; Evrenoglou, T.; Jarde, A.; Minozzi, S.; Bergman, H.; Buckley, B.S.; Probyn, K.; Villanueva, G.; Henschke, N.; et al. Efficacy and safety of COVID-19 vaccines. *Cochrane Database Syst. Rev.* **2022**, *12*, CD015477. [PubMed]
16. Yang, Z.R.; Jiang, Y.W.; Li, F.X.; Liu, D.; Lin, T.F.; Zhao, Z.Y.; Wei, C.; Jin, Q.Y.; Li, X.M.; Jia, Y.X.; et al. Efficacy of SARS-CoV-2 vaccines and the dose–response relationship with three major antibodies: A systematic review and meta-analysis of randomised controlled trials. *Lancet Microbe* **2023**, *4*, e236–e246. [CrossRef] [PubMed]
17. Watson, O.; Barnsley, G.; Toor, J.; Hogan, A.; Winskill, P.; Ghani, A. Global impact of the first year of COVID-19 vaccination: A mathematical modelling study. *Lancet Infect. Dis.* **2022**, *22*, 1293–1302. [CrossRef]
18. Zeng, B.; Gao, L.; Zhou, Q.; Yu, K.; Sun, F. Effectiveness of COVID-19 vaccines against SARS-CoV-2 variants of concern: A systematic review and meta-analysis. *BMC Med.* **2022**, *20*, 200. [CrossRef]
19. Ministerio de Sanidad, Consumo y Bienestar Social. Portal Estadistico. Area de Inteligencia de Gestion. Available online: https://pestadistico.inteligenciadegestion.mscbs.es/publicoSNS/comun/ArbolNodos.aspx?idNodo=23525 (accessed on 6 July 2019).
20. Data on COVID-19 vaccination in the EU/EEA. 2023. Available online: https://www.ecdc.europa.eu/en/publications-data/data-covid-19-vaccination-eu-eea (accessed on 15 October 2023).
21. Garcia-Carretero, R.; Vazquez-Gomez, O.; Gil-Prieto, R.; Gil-de Miguel, A. Hospitalization burden and epidemiology of the COVID-19 pandemic in Spain (2020–2021). *BMC Infect. Dis.* **2023**, *23*, 476. [CrossRef]
22. Instituto de Salud Carlos III. Informe no 128 Situación de COVID-19 en España a 10 de mayo de 2022.pdf. 2022. Available online: https://www.isciii.es/QueHacemos/Servicios/VigilanciaSaludPublicaRENAVE/EnfermedadesTransmisibles/Documents/INFORMES/Informes%20COVID-19/INFORMES%20COVID-19%202022/Informe%20n%C2%BA%20128%20Situaci%C3%B3n%20de%20COVID-19%20en%20Espa%C3%B1a%20a%2010%20de%20mayo%20de%202022.pdf (accessed on 9 June 2022).
23. Gobierno de España. Estrategia de vacunación COVID-19. Available online: https://www.vacunacovid.gob.es/ (accessed on 12 September 2024).
24. Rodriguez-Maroto, G.; Atienza-Diez, I.; Ares, S.; Manrubia, S. Vaccination strategies in structured populations under partial immunity and reinfection. *J. Phys. A Math. Theor.* **2023**, *56*, 204003. [CrossRef]
25. Garcia-Carretero, R.; Vigil-Medina, L.; Barquero-Perez, O.; Ramos-Lopez, J. Relevant Features in Nonalcoholic Steatohepatitis Determined Using Machine Learning for Feature Selection. *Metab. Syndr. Relat. Disord.* **2019**, *17*, 444–451. [CrossRef]
26. Garcia-Carretero, R.; Barquero-Perez, O.; Mora-Jimenez, I.; Soguero-Ruiz, C.; Goya-Esteban, R.; Ramos-Lopez, J. Identification of clinically relevant features in hypertensive patients using penalized regression: A case study of cardiovascular events. *Med. Biol. Eng. Comput.* **2019**, *57*, 2011–2026. [CrossRef]
27. Garcia-Carretero, R.; Vigil-Medina, L.; Barquero-Perez, O.; Mora-Jimenez, I.; Soguero-Ruiz, C.; Goya-Esteban, R.; Ramos-Lopez, J. Logistic LASSO and Elastic Net to Characterize Vitamin D Deficiency in a Hypertensive Obese Population. *Metab. Syndr. Relat. Disord.* **2020**, *18*, 79–85. [CrossRef] [PubMed]
28. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020. Available online: https://www.r-project.org/ (accessed on 12 September 2024).
29. Farsalinos, K.; Poulas, K.; Vantarakis, A.; Leotsinidis, M.; Kouvelas, D.; Docea, A.O.; Kostoff, R.; Gerotziafas, G.T.; Antoniou, M.N.; Polosa, R.; et al. Improved strategies to counter the COVID-19 pandemic: Lockdowns vs. primary and community healthcare. *Toxicol. Rep.* **2021**, *8*, 1–9. [CrossRef] [PubMed]
30. Connors, M.; Graham, B.; Lane, H.; Fauci, A. SARS-CoV-2 Vaccines: Much Accomplished, Much to Learn. *Ann. Intern. Med.* **2021**, *174*, 687–690. [CrossRef]
31. Kustin, T.; Harel, N.; Finkel, U.; Perchik, S.; Harari, S.; Tahor, M.; Caspi, I.; Levy, R.; Leshchinsky, M.; Ken, Dror, S.; et al. Evidence for increased breakthrough rates of SARS-CoV-2 variants of concern in BNT162b2-mRNA-vaccinated individuals. *Nat. Med.* **2021**, *27*, 1379–1384. [CrossRef] [PubMed]

32. Wang, Z.; Muecksch, F.; Schaefer-Babajew, D.; Finkin, S.; Viant, C.; Gaebler, C.; Hoffmann, H.H.; Barnes, C.O.; Cipolla, M.; Ramos, V.; et al. Naturally enhanced neutralizing breadth against SARS-CoV-2 one year after infection. *Nature* **2021**, *595*, 426–431. [CrossRef]
33. Garcia-Carretero, R.; Vazquez-Gomez, O.; Ordoñez-Garcia, M.; Garrido-Peño, N.; Gil-Prieto, R.; Gil-de Miguel, A. Differences in Trends in Admissions and Outcomes among Patients from a Secondary Hospital in Madrid during the COVID-19 Pandemic: A Hospital-Based Epidemiological Analysis (2020–2022). *Viruses* **2023**, *15*, 1616. [CrossRef]
34. Barandalla, I.; Alvarez, C.; Barreiro, P.; de Mendoza, C.; González-Crespo, R.; Soriano, V. Impact of scaling up SARS-CoV-2 vaccination on COVID-19 hospitalizations in Spain. *Int. J. Infect. Dis.* **2021**, *112*, 81–88. [CrossRef]
35. Sentís, A.; Kislaya, I.; Nicolay, N.; Meijerink, H.; Starrfelt, J.; Martínez-Baz, I.; Castilla, J.; Nielsen, K.F.; Hansen, C.H.; Emborg, H.-D.; et al. Estimation of COVID-19 vaccine effectiveness against hospitalisation in individuals aged ≥65 years using electronic health registries; a pilot study in four EU/EEA countries, October 2021 to March 2022. *Eurosurveillance* **2022**, *27*, 2200551. [CrossRef]
36. Chyon, F.; Suman, M.; Fahim, M.; Ahmmed, M. Time series analysis and predicting COVID-19 affected patients by ARIMA model using machine learning. *J. Virol. Methods* **2022**, *301*, 114433. [CrossRef]
37. Maleki, M.; Mahmoudi, M.; Wraith, D.; Pho, K. Time series modelling to forecast the confirmed and recovered cases of COVID-19. *Travel Med. Infect. Dis.* **2020**, *37*, 101742. [CrossRef]
38. Kırbaş, İ.; Sözen, A.; Tuncer, A.; Kazancıoğlu, F. Comparative analysis and forecasting of COVID-19 cases in various European countries with ARIMA, NARNN and LSTM approaches. *Chaos Solitons Fractals* **2020**, *138*, 110015. [CrossRef] [PubMed]
39. Nabi, K.; Tahmid, M.; Rafi, A.; Kader, M.; Haider, M. Forecasting COVID-19 cases: A comparative analysis between recurrent and convolutional neural networks. *Results Phys.* **2021**, *24*, 104137. [CrossRef] [PubMed]
40. Makridakis, S.; Spiliotis, E.; Assimakopoulos, V.; Semenoglou, A.; Mulder, G.; Nikolopoulos, K. Statistical, machine learning and deep learning forecasting methods: Comparisons and ways forward. *J. Oper. Res. Soc.* **2023**, *74*, 840–859. [CrossRef]
41. Whittaker, C.; Walker, P.G.; Alhaffar, M.; Hamlet, A.; Djaafara, B.A.; Ghani, A.; Ferguson, N.; Dahab, M.; Checchi, F.; Watson, O.J. Under-reporting of deaths limits our understanding of true burden of COVID-19. *BMJ* **2021**, *375*, n2239. [CrossRef] [PubMed]
42. Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *JOurnal R. Stat. Soc. Ser. B (Methodol.)* **1996**, *58*, 267–288. [CrossRef]
43. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
44. Cutler, D.; Edwards, T., Jr.; Beard, K.; Cutler, A.; Hess, K.; Gibson, J.; Lawler, J.J. Random forests for classification in ecology. *Ecology* **2007**, *88*, 2783–2792. [CrossRef]
45. Garcia-Carretero, R.; Vazquez-Gomez, O.; Lopez-Lomba, M.; Gil-Prieto, R.; Gil-de Miguel, A. Insulin Resistance and Metabolic Syndrome as Risk Factors for Hospitalization in Patients with COVID-19: Pilot Study on the Use of Machine Learning. *Metab. Syndr. Relat. Disord.* **2023**, *21*, 443–452. [CrossRef]
46. Garcia-Carretero, R.; Vigil-Medina, L.; Mora-Jimenez, I.; Soguero-Ruiz, C.; Barquero-Perez, O.; Ramos-Lopez, J. Use of a K-nearest neighbors model to predict the development of type 2 diabetes within 2 years in an obese, hypertensive population. *Med. Biol. Eng. Comput.* **2020**, *58*, 991–1002. [CrossRef]

8 COVID-19 in Patients with HIV

AIDS and Behavior (2024) 28:3093–3102
https://doi.org/10.1007/s10461-024-04394-z

ORIGINAL PAPER

Outcomes of Patients Living with HIV Hospitalized due to COVID-19: A 3-Year Nationwide Study (2020–2022)

Rafael Garcia-Carretero[1] · Oscar Vazquez-Gomez[1] · Belen Rodriguez-Maya[1] · Ruth Gil-Prieto[2] · Angel Gil-de-Miguel[2]

Accepted: 24 May 2024 / Published online: 4 July 2024

Abstract

Scientific reports on the association between human immunodeficiency virus (HIV) in patients with COVID-19 and mortality have not been in agreement. In this nationwide study, we described and analyzed the demographic and clinical characteristics of people living with HIV (PLWH) and established that HIV infection is a risk factor for mortality in patients hospitalized due to COVID-19. We collected data from the National Hospital Data Information System at Hospitalization between 2020 and 2022. We included patients admitted to the hospital with a diagnosis of COVID-19. We established a cohort of patients with PLWH and compared them to patients without HIV (non-PLWH). For multivariate analyses, we performed binary logistic regression, using mortality as the dependent variable. To improve the interpretability of the results we also applied penalized regression and random forest, two well-known machine-learning algorithms. A broad range of comorbidities, as well as sex and age data, were included in the final model as adjusted estimators. Our data of 1,188,160 patients included 6,973 PLWH. The estimated hospitalization rate in this set was between 1.43% and 1.70%, while the rate among the general population was 0.83%. Among patients with COVID-19, HIV infection was a risk factor for mortality with an odds ratio (OR) of 1.25 (95% CI, 1.14–1.37, $p < 0.001$). PLWH are more likely to be hospitalized due to COVID-19 than are non-PLWH. PLWH are 25% more likely to die due to COVID-19 than non-PLWH. Our results highlight that PLWH should be considered a population at risk for both hospitalization and mortality.

Keywords COVID.19 · Mortality · HIV · Population-based

Introduction

The coronavirus disease 2019 (COVID-19), caused by the SARS-CoV-2 virus, imposed a great burden of illness worldwide, from both socioeconomic and healthcare points of view. As of July 5, 2023, a total of 13,914,811 confirmed cases of COVID-19 had been reported in Spain, with 682,216 hospitalizations and 122,057 deaths [1]. Some comorbidities are considered risk factors for adverse effects and mortality. Chronic conditions, such as diabetes, hypertension, obesity, immunosuppression, malignancies, and HIV infection, have been associated with worse outcomes in COVID-19 [2].

People living with HIV (PLWH) may be proportionally more affected by infection with SARS-CoV-2 than people without HIV (non-PLWH), although conflicting results have been reported [3]. If coinfection with HIV and SARS-CoV-2 puts patients at high risk for mortality, health systems should engage in more aggressive preventive measures

✉ Rafael Garcia-Carretero
rgcarretero@salud.madrid.org

Oscar Vazquez-Gomez
govaz5@hotmail.com

Belen Rodriguez-Maya
belenrmaya@gmail.com

Ruth Gil-Prieto
ruth.gil@urjc.es

Angel Gil-de-Miguel
angel.gil@urjc.es

[1] Department of Internal Medicine, Mostoles University Hospital, Rey Juan Carlos University (Madrid), Madrid, Spain

[2] Department of Preventive Medicine and Public Health, Rey Juan Carlos University (Madrid), Madrid, Spain

and therapeutic efforts to avoid adverse outcomes in PLWH when infected by SARS-CoV-2.

In this study, we analyzed the impact of coinfection with HIV and SARS-CoV-2 in PLWH. We describe the incidence of hospitalization among PLWH and related outcomes in terms of comorbidities, severity, and mortality.

Methods

Study Design

We conducted a retrospective, population-based study using data drawn from the Spanish National Hospital Data Information System at Hospitalization (MBDS-H), a valuable database for epidemiological analyses of included conditions created by the Spanish Ministry of Health. The MBDS-H is an administrative registry of discharge reports. Nearly 95% of hospitals in Spain, both public and private, are covered by the database. It is estimated that 97% of all discharge reports are registered in this database. The data are exclusively drawn from hospital discharges, including information on age, sex, date of admission/discharge, type of hospital, place of residence, and diagnoses. MBDS-H includes diseases encoded using the 10th Clinical Revision of the International Classification of Diseases (ICD-10-CM). A new dataset is generated in January of each year. However, due to the high volume of data, the data only become available after a delay of 1 year. The Spanish Ministry of Health provided us with data up to December 31, 2022.

Data Collection

We used data from populations covered by hospitals included in the MBDS-H information system, as noted. We were provided with the microdata extracted from the MBDS-H from the Ministry of Health between 2020 and 2022 using the code for COVID-19 (U07.1) in any diagnostic position. That is, we collected data for patients presenting with a diagnosis of COVID-19 from January 1, 2020, to December 31, 2022. For HIV infection, we used the codes Z21 and B20 to B24. No data on treatment or immunovirological status were provided. For each hospitalized patient, we collected data on age, sex, dates of admission and discharge, ICU admission, and type of discharge. Main and secondary diagnoses were also gathered to identify HIV infection, diabetes, hypertension, and other chronic conditions. Patients who had incomplete data regarding ICU admission, mortality, length of hospitalization, or diagnosed conditions of interest were excluded. No names or personal identifying details were recorded. Data were anonymized and de-identified to ensure patients' privacy.

Definition of Waves

We categorized the pandemic into waves based on the classification of the Epidemiological National Surveillance Net study, which exclusively used data from Spain. The observation periods were split into outbreaks based on the 14-day cumulative incidence and on a turning point for each wave, such that every turning point indicated the end of one wave and the beginning of the next [4].

Univariate Analysis

We performed descriptive and correlational analyses. We used means or medians with continuous variables as appropriate, as well as percentages with categorical variables. Average hospital length of stay is defined as the total number of days of stay, divided by the total number of hospitalizations. Mortality and the need for ICU admissions are considered clinical severity criteria. Deaths and ICU admissions, as numerators, are divided by the total number of hospitalizations to calculate the mortality rate and ICU admission rate, respectively. Both parameters are expressed as percentages. The chi-square test and the Wilcoxon signed-rank test were performed as tests of independence when appropriate.

Multivariate Analyses

Logistic regression was used to analyze mortality in our cohort and hence to estimate the impact of the included variables. We used a combination of a classical approach (with binary logistic regression) and a machine learning approach (with penalized logistic regression) to calculate beta coefficients for variables as well as odds ratios (ORs). Binary logistic regression is the most frequently used statistical approach in biomedical sciences with binary outcomes, i.e., yes/no. Logistic regression is simple and straightforward, and it provides easy interpretation of the effects of explanatory variables on response variables. However, a model may have too many features selected as explanatory variables, making it too complex for use. The rationale for the use of machine learning at this stage was that it allowed us to select a set of features, that is, a parsimonious model, without loss of accuracy or reliability.

As noted, our machine learning approach adopted logistic regression with L1-penalized regularization. This approach is also known as the least absolute shrinkage and selection operator (LASSO) [5, 6]. It discards variables that do not contribute to the fit of the final model. It forces beta coefficients to a range from very small values to exactly zero. All beta coefficients shrink, but those with weak effects are dropped. We used cross-validation to internally validate the LASSO algorithm. We plotted the average model evaluation scores to select the set of variables that maximized the model's predictive accuracy. Penalization is determined by the lambda value, which was used to select the subset of variables. LASSO is recommended by the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis checklist for developing and validating risk and diagnostic models [7].

Interpretability of Results

Once a subset of features was selected, the beta coefficients were obtained, and the ORs were calculated, we used random forest (RF), another machine learning algorithm based on decision trees, to better interpret the results. This generated many independent decision trees, which then were combined to obtain a single output [8, 9]. RF allowed the model to be interpreted at a global scale.

For all tests, the level of statistical significance was set at $P<0.05$. We used R language version 4.3.2 (Vienna, Austria) running on a Debian 12 GNU/Linux workstation for both standard and machine learning-based analyses.

Results

We assessed data on 1,188,160 hospitalized patients, including 6,973 patients living with HIV (0.58% of all hospitalized patients) between 2020 and 2022. The main characteristics are presented in Table 1. Men among PLWH were more likely to be hospitalized than men among non-PLWH. PLWH were also significantly younger than non-PLWH. They also had a lower prevalence of diabetes, hypertension, coronary disease, heart failure, and other cardiovascular risk factors. However, chronic liver disease, malignancy, and chronic pulmonary diseases were significantly more common in PLWH.

By December 31, 2021, it was estimated that there were between 136.436 and 162.307 PLWH in Spain. That is, the prevalence of HIV in Spain was between 0.28% and 0.31%. The rate of hospitalization of patients among the general population was 0.83% per year between 2020 and 2022. Regarding PLWH, the ratio of hospitalized patients was between 1.43% and 1.70%, that is, greater than the general population.

Overall, the prevalence of HIV and hepatitis B virus (HBV) coinfection was 0.7%, but we found that among PLWH the prevalence was 4.2%. Regarding hepatitis C virus, the prevalence among PLWH was even higher

Table 1 Baseline features of our cohort and univariate analyses

	Total	HIV infection		Test statistic	*P* value
		PLWH	No PLWH		
Patients	1,188,160	6,973	1,181,187	NA	NA
Sex (men)	54.9%	75.3%	54.8%	$X^2=1.18\times10^3$	0.001
Age (IQR)	73 (27)	54 (12)	73 (27)	$W=6.28\times10^9$	0.001
Hospital length of stay in days (IQR)	10.9 (8)	12.1 (9)	10.9 (8)	$W=5.9\times10^0$	0.002
ICU admissions (rate, %)	100,578 (8.5%)	745 (10.7%)	99,833 (8.5%)	$X^2=5.51\times10^1$	0.001
ICU length of stay in days (IQR)	8 (20.2)	7 (17.7)	8 (20.2)	$W=3.93\times10^7$	0.006
Deaths (rate, %)	153,144 (12.9%)	540 (7.7%)	152,604 (12.9%)	$X^2=2.04\times10^3$	0.001
Comorbidities					
Diabetes	24.4%	11.4%	24.5%	$X^2=6.46\times10^2$	0.001
Hypertension	31.9%	16.7%	32%	$X^2=7.44\times10^2$	0.001
Coronary disease	9.7%	7%	9.7%	$X^2=5.97\times10^1$	0.001
Heart failure	17%	4.9%	17.1%	$X^2=7.26\times10^2$	0.001
Dementia	6.5%	1.1%	6.6%	$X^2=3.42\times10^2$	0.001
Chronic kidney disease	15.6%	9.7%	15.6%	$X^2=1.83\times10^2$	0.001
Chronic liver disease	0.6%	5.8%	0.6%	$X^2=3.26\times10^3$	0.001
Malignancy	10.1%	12.1%	10.1%	$X^2=3.09\times10^1$	0.001
Obesity	11.8%	5.7%	11.9%	$X^2=2.53\times10^2$	0.001
Chronic pulmonary disease	14.8%	26.8%	14.7%	$X^2=8.07\times10^2$	0.001
Cerebrovascular disease	1.4%	0.6%	1.4%	$X^2=3.30\times10^1$	0.001
Coinfection					
HBV	0.3%	4.2%	0.3%	$X^2=2.99\times10^3$	0.001
HCV	0.7%	25.5%	0.6%	$X^2=5.77\times10^4$	0.001
HTLV-I/II	0%	0%	0%	NA	NA

PLWH: people living with HIV. HBV: Hepatitis B virus. HCV: Hepatitis C virus. HTLV-I/II: Human T-lymphotropic virus types I and II. NA: non-applicable. ICU: intensive care unit. IQR: interquartile range. W: Wilcoxon signed-rank test. X^2: Chi-square test

(25.5%). We did not found infection caused by human T-lymphotropic virus types I and II.

Due to the predominance of men among PLWH, we decided to plot the relationship between age and sex in a population pyramid (Fig. 1) and to show the main characteristics related to COVID-19 in Table 2. It can be seen that the mortality rate in men was higher than it was in women (8.5% vs. 5.6%, $X^2 = 25.7$, $p < 0.001$). In addition, cardiovascular risk factors, such as diabetes and coronary diseases, were more prevalent in men. Malignancy was also more prevalent in male PLWH. However, women tended to be more obese and to have a higher prevalence of chronic pulmonary disease. There were no differences regarding hypertension, heart failure, dementia, chronic kidney disease, chronic liver disease, or cerebrovascular disease.

Figure 2; Table 3 present the evolution of the pandemic among PLWH, splitting the observation period into epidemiological waves. The first wave included 1,303 hospitalizations, and this number steadily dropped until the fourth and fifth waves. More men were admitted than women, with no changes in the distribution along the pandemic. The median age was 54 years, with almost no changes for the entire period. While the mortality rate was 7.7% globally, we observed a decreasing trend from the third wave onward.

The first multivariate analyses were performed with all patients included. Table 4 presents the results. Mortality was used as the dependent variable. We included coinfection with SARS-CoV-2 and HIV as an independent variable, alongside sex, age, and the remaining comorbidities in the final model of logistic regression. HIV infection was a risk factor for hospitalizations due to COVID-19, that is, a patient had a 25% greater chance if a PLWH than a non-PLWH to be hospitalized for COVID-19.

The results of the multivariate analysis of the data for PLWH using binary logistic regression are shown in Table 5. Sex, age, malignancies, heart failure, hypertension, and chronic liver disease were the main risk factors for a PLWH to be admitted to a hospital. To identify the most relevant variables, we also performed penalized logistic regression for mortality. Compared to the standard

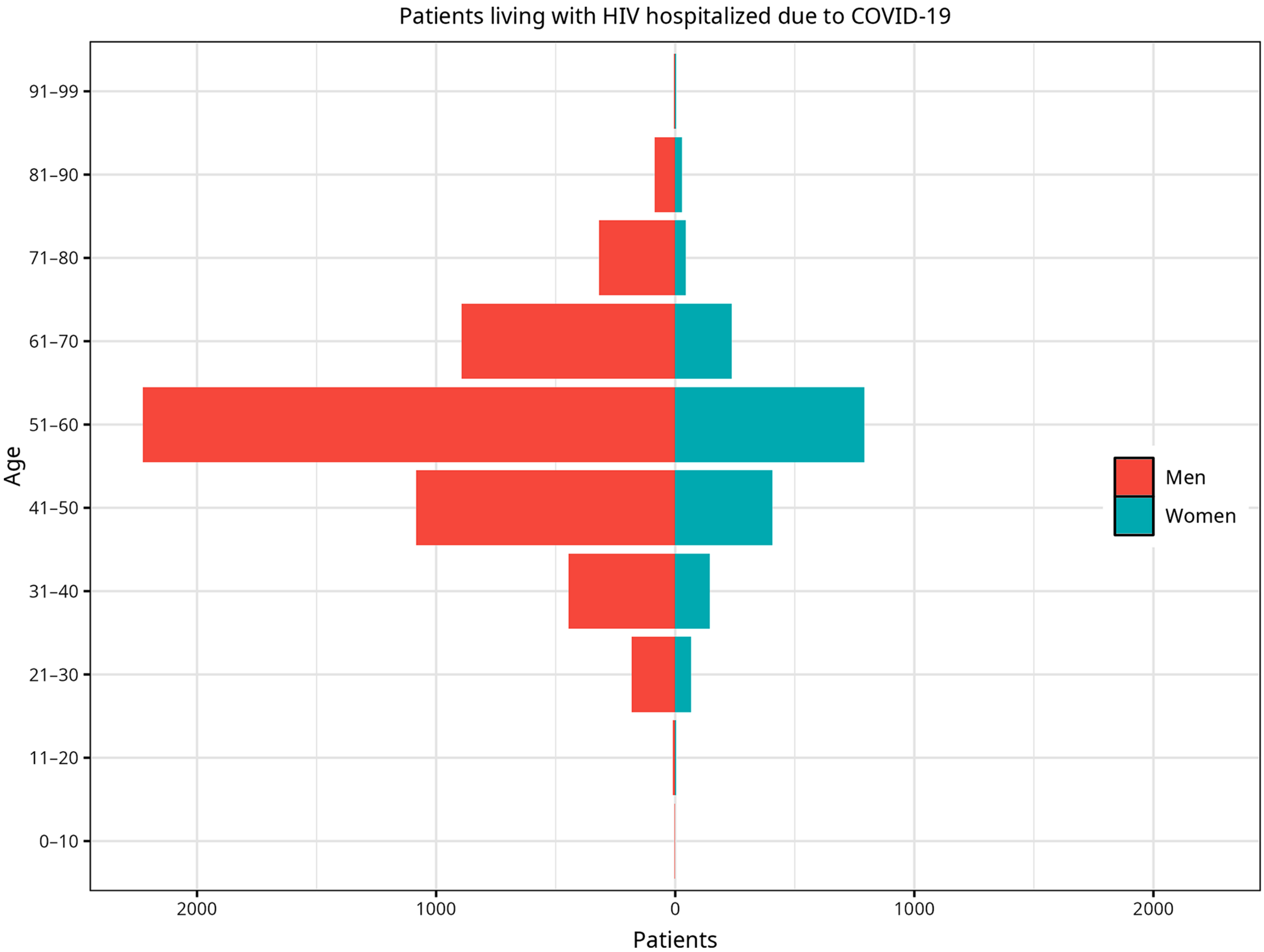

Fig. 1 Population pyramid by age range among hospitalized people living with HIV between 2020 and 2022

Table 2 Differences between men and women in the hospitalized Spanish population living with HIV

	Men	Women	Test statistic	*P* value
Patients	5,253	1,720	NA	NA
Age (IQR)	55 (12)	53 (12)	$W=4.92\times10^{6}$	0.062
Hospital length of stay in days (IQR)	12.3 (9)	11.7 (8)	$W=1.10\times10^{0}$	0.274
ICU admissions (rate, %)	586 (11.2%)	159 (9.2%)	$X^2=5.18\times10^{0}$	0.073
ICU length of stay in days (IQR)	8 (18.1)	6 (16.1)	$W=5.17\times10^{4}$	0.031
Deaths (rate, %)	444 (8.5%)	96 (5.6%)	$X^2=2.57\times10^{1}$	0.001
Comorbidities				
Diabetes	12.3%	8.7%	$X^2=1.68\times10^{1}$	0.001
Hypertension	17%	15.9%	$X^2=1.18\times10^{0}$	0.289
Coronary disease	8.1%	3.5%	$X^2=4.15\times10^{1}$	0.001
Heart failure	5%	4.5%	$X^2=6.70\times10^{-1}$	0.43
Dementia	1.1%	1.1%	$X^2=2.00\times10^{-2}$	0.894
Chronic kidney disease	10%	8.8%	$X^2=1.92\times10^{0}$	0.17
Chronic liver disease	5.8%	5.8%	$X^2=0.0\times10^{0}$	1
Malignancy	13.1%	9%	$X^2=2.09\times10^{1}$	0.001
Obesity	4.8%	8.4%	$X^2=3.22\times10^{1}$	0.001
Chronic pulmonary disease	25.7%	30.2%	$X^2=1.35\times10^{1}$	0.001
Cerebrovascular disease	0.6%	0.5%	$X^2=5.90\times10^{-1}$	0.495
Coinfection				
HBV	4.4%	3.4%	$X^2=3.55\times10^{0}$	0.55
HCV	25.6%	25.1%	$X^2=1.90\times10^{-1}$	0.676
HTLV-I/II	0%	0%	NA	NA

HBV: Hepatitis B virus. HCV: Hepatitis C virus. HTLV-I/II: Human T-lymphotropic virus types I and II. NA: non-applicable. ICU: Intensive care unit. IQR: interquartile range. W: Wilcoxon signed-rank test. X^2: Chi-square test

binary logistic regression, the list of variables was small, as LASSO dropped non-relevant variables as the model was being fitted. LASSO only identified age, malignancies, and heart failure as the most relevant variables associated with the risk of death. LASSO only provides beta coefficients so that ORs can be calculated, it does not report confidence intervals. In addition, these ORs tend to be lower than those in standard logistic regression, due to the characteristics of the algorithm. LASSO proposed a more constrained model using a lambda value that chose only three variables (Fig. 3).

Finally, to improve the interpretation of the results, we used RF to rank variables in order of importance. The variable importance for the selected features can be examined visually to allow us to observe which were the most important for predicting the response variable. Using all variables, therefore, Fig. 4 shows the results provided by the RF algorithm, and the 15 features included are displayed. Age, malignancy, and acute heart failure were the most relevant variables, as identified by penalized logistic regression.

Discussion

The main aim of our study was to determine the impact of COVID-19 in hospitalized PLWH. We found that PWLH with COVID-19 were at high risk for both hospitalization and mortality. Coinfection with HIV and SARS-CoV-2 was determined to be an independent risk factor for mortality when it was adjusted by age, sex, and other comorbidities.

More specifically, we found that hospitalized PLWH were younger than hospitalized non-PLWH. PLWH tended to have a lower prevalence of diabetes, obesity, and hypertension but a higher prevalence of chronic liver or pulmonary diseases. Malignancy was also higher in PLWH. This suggests that the demographic and clinical profiles among hospitalized PLWH are different than those in the general population. Below, we further discuss the relationship between these variables and the risk of mortality.

In a recent meta-analysis, the pooled prevalence of PLWH in European countries was 0.73% (95%CI 0.24–1.22), although there was risk of bias due to the small number of patients included [3]. Still, our results among hospitalized patients are similar to that prevalence. In the included European studies, the median age was 50, in line with the age of 54 in our study. In addition, in that meta-analysis, men represented 74% of all patients (the value was 75% in our study). Overall, our results are in line with those of Danwang et al. [3], with the exception of the mortality risk. Those authors did not find that HIV

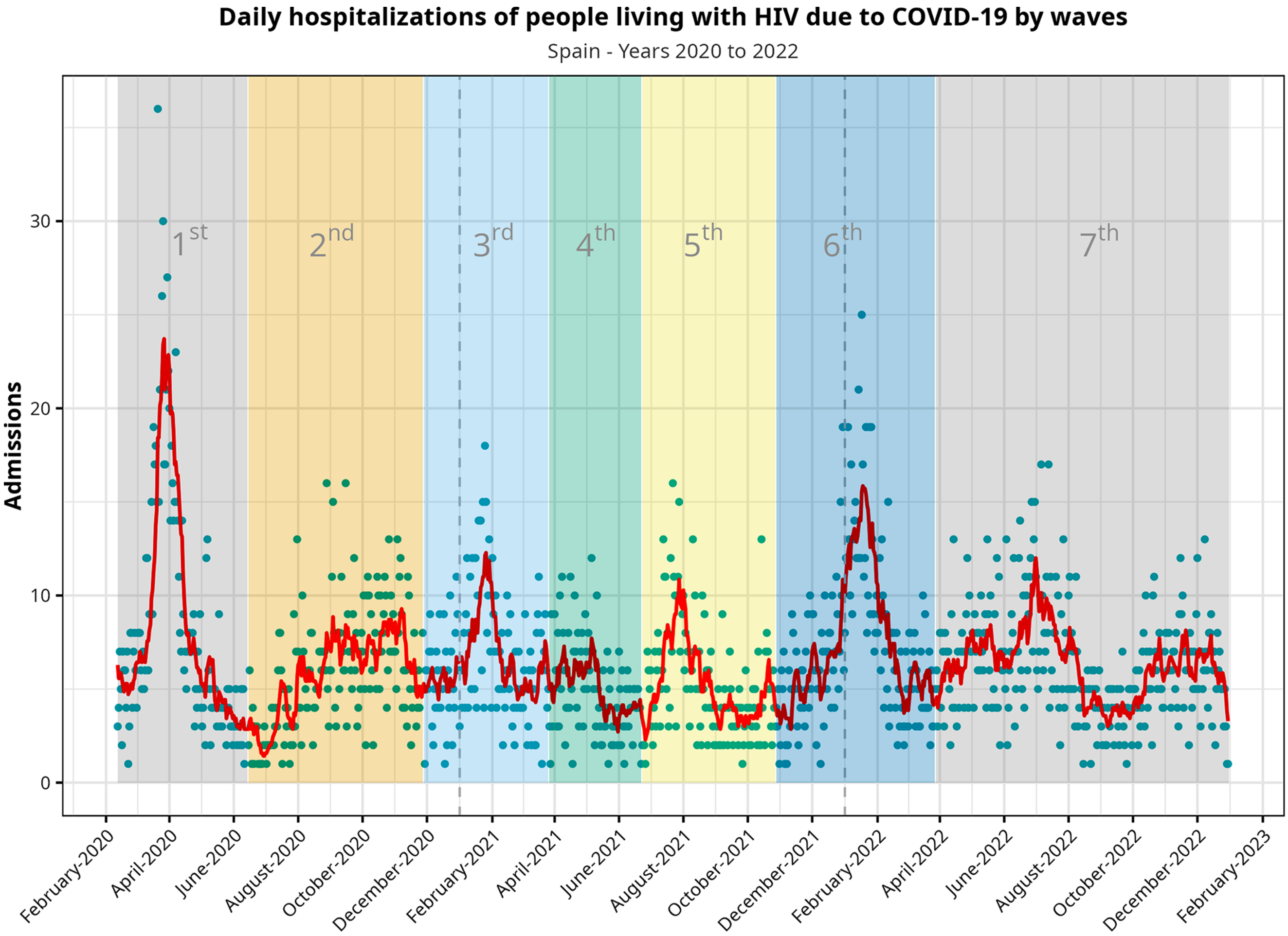

Fig. 2 Evolution of the pandemic during the observation period regarding people living with HIV

Table 3 Clinical characteristics of the hospitalized population living with HIV (2020–2021), by epidemic wave

	Total	First	Second	Third	Fourth	Fifth	Sixth	Seventh
Patients	6,973	1,303	1,004	649	491	598	1,181	1,747
Sex (men)	75.3%	76.7%	75.5%	78.6%	76.2%	74.9%	74.2%	73.7%
Age (IQR)	54 (12)	53 (11)	53 (13)	55 (12)	56 (11)	53 (16)	55 (12)	55 (12)
Hospital length of stay in days (IQR)	7 (9)	11.1 (8)	7 (10)	13.8 (9)	8 (9)	12.5 (9)	7 (9)	7 (9)
ICU admissions (rate, %)	745 (10.7%)	118 (9.1%)	101 (10.1%)	85 (13.1%)	56 (11.4%)	79 (13.2%)	130 (11%)	176 (10.1%)
Deaths (rate, %)	540 (7.7%)	117 (9%)	71 (7.1%)	65 (10%)	36 (7.3%)	47 (7.9%)	91 (7.7%)	113 (6.5%)

ICU: Intensive care unit. IQR: Interquartile range

infection increases the likelihood of severe COVID-19 outcomes.

Our results are also in line with a large retrospective study performed in the UK [10] that found that PLWH were more likely to be men, with a median age of 48. That study also found that, after adjusting for age, sex, and other comorbidities, PLWH were at higher risk for death.

In our multivariate analyses, age and sex were risk factors for mortality, not only in the general population but also in PLWH [11, 12]. HIV infection was a risk factor for mortality once sex and age were included in adjusted analyses, as reported previously [13–17].

We found that the mortality rate was lower among PLWH in unadjusted univariate analyses (Table 1). However, when adjusted multivariate analyses were performed, we found that HIV infection was associated with increased risk of mortality (OR 1.25, 95%CI 1.14–1.37); that is, that PLWH are 25% more likely to die due to COVID-19 than non-PLWH. As noted, the meta-analysis by Danwang et al. [3] did not find evidence for a link between HIV infection and mortality risk in COVID-19 patients, although the authors identified two studies that

Table 4 Multivariate analysis using logistic regression with mortality as a dependent variable. All patients were included in the analysis

	OR	95%CI	*P* value
HIV infection	1.25	1.14–1.37	0.001
Sex (man)	1.35	1.34–1.37	0.001
Age	1.05	1.05–1.05	0.001
Comorbidities			
Malignancy	2.21	2.18–2.25	0.001
Diabetes	1.01	1.00–1.02	0.12
Coronary disease	1.09	1.07–1.11	0.001
Heart failure	1.07	1.05–1.08	0.001
Hypertension	0.93	0.92–0.95	0.001
Obesity	0.96	0.94–0.98	0.001
Dementia	1.24	1.22–1.27	0.001
Cerebrovascular disease	2.07	1.99–2.14	0.001
Chronic liver disease	1.59	1.5–1.7	0.001
Chronic kidney disease	1.14	1.12–1.16	0.001
Chronic pulmonary disease	0.72	0.71–0.73	0.001
Coinfection			
HBV	0.95	0.85–1.05	0.291
HCV	1.06	0.99–1.13	0.111

CI: Confidence Interval. OR: Odds Ratio. HBV: Hepatitis B virus. HCV: Hepatitis C virus

Table 5 Multivariate analysis using logistic regression and penalized logistic regression among PLWH only, using mortality as a dependent variable

	Binary logistic regression		LASSO (penalized regression)	
	OR (95%CI)	*P* value	Beta coefficient	OR
Sex (man)	1.35 (1.07–1.72)	0.01	.	.
Age	1.04 (1.03–1.05)	0.001	0.01447651	1.01
Malignancy	4.21 (3.44–5.15)	0.001	1.03839103	2.82
Heart failure	2.11 (1.53–2.89)	0.001	0.08649972	1.09
Hypertension	0.72 (0.56–0.93)	0.01	.	.
Obesity	1.44 (0.99–2.06)	0.05	.	.
Dementia	0.41 (0.1–1.12)	0.13	.	.
Cerebrovascular disease	2.2 (0.9–4.81)	0.06	.	.
Chronic liver disease	1.51 (1.07–2.08)	0.01	.	.
Chronic pulmonary disease	0.65 (0.52–0.81)	0.001	.	.

OR: odds ratio. CI: confidence interval. LASSO: least absolute shrinkage and selection operator

suggested this association. Our results may seem controversial in light of this difference between univariate and multivariate outcomes with respect to the risk of mortality among PLWH. However, this phenomenon is well documented and can be explained by the fact that univariate analyses can miss some variables that are deemed relevant in multivariate analyses; for this reason, it can produce biased estimates of effects of other variables on the response [18].

Another controversy is that we also found that PLWH who were coinfected with SARS-CoV-2 were 75–110% more likely to be admitted to a hospital. Conditions such as HIV infection can be considered a risk factor, along with cardiovascular comorbidities, in patients with COVID-19. A preliminary report of a case-control study suggested that SARS-CoV-2 coinfection does not have an extraordinarily great impact on PLWH [19]. However, the authors emphasized limitations of their study and reported certain trends on severity and mortality that could be worse in PLWH than in non-PLWH. A later study suggested that PLWH are at increased risk for hospitalization [3] did not find an increased risk for adverse outcomes, including death. The authors referred to the role of immunodepression and immunovirological status in PLWH to explain their results. They hypothesized that the cytokine storm could be averted if immunodepression is present. They also proposed further research that would stratify immunovirological status, including CD4 + T lymphocytes counts and viral load, to identify patients that are most likely to present with severe forms of COVID-19. Bhaskaran et al. [10] demonstrated that HIV infection is a risk factor for mortality. They also stratified risk based on age and comorbidities.

Although well-controlled HIV infection has been associated with cardiovascular disease [17, 20, 21], the comorbidities analyzed in our cohort did not have a significant effect on mortality. Hypertension, obesity, chronic kidney disease, and coronary disease were risk factors in the general population but not in PLWH. We found that heart failure was associated with a higher risk of mortality in our cohort. PLWH may be at high cardiovascular risk, not only due to aging but also because anti-retroviral therapy (ART) may predispose PLWH to the development of cardiovascular diseases. Heart failure has been noted as an important comorbidity in PLWH, despite ART [22, 23].

It should be noted that we found not only a higher prevalence of malignancy among PLWH than among non-PLWH but also that cancer was a risk factor for mortality in the case of coinfection with SARS-CoV-2. In a recent multicenter study, Suarez et al. [24] investigated the relationship between malignancy and HIV in 17,978 PLWH in Spain. The authors found that mortality due to cancer was higher among PLWH than among the general population. Malignancy was split into several categories, including viral, nonviral, and non-AIDS-defining cancer, and they concluded that cancer was a risk factor for mortality in all categories analyzed.

Our study had several strengths. First, we used machine learning for data analyses, which gave us better insight into the results. In the first step, we used LASSO as an

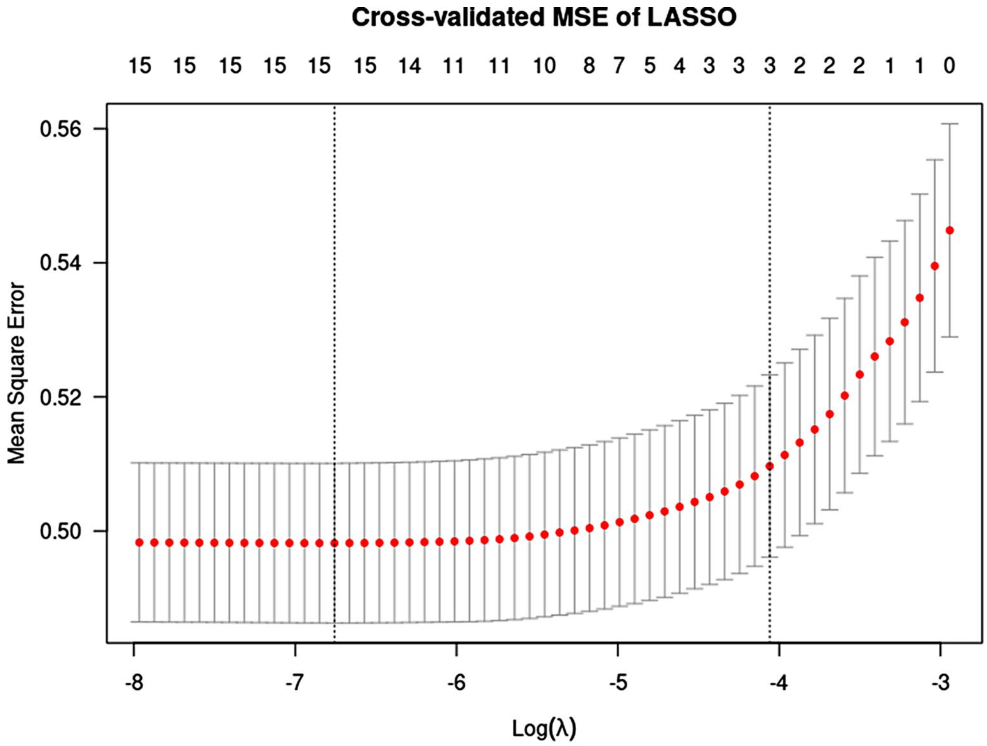

Fig. 3 Lambda values for penalized regression (LASSO). Lambda values represent the penalization of the beta coefficients. Vertical lines represent the range of lambda for which accuracy is not adversely affected. Numbers across the top of the plot represent the number of variables in the model when a certain value of lambda was used (15 versus 3 features). We chose the lambda values that received the minimum number of features without losing accuracy (determined according to mean square error)

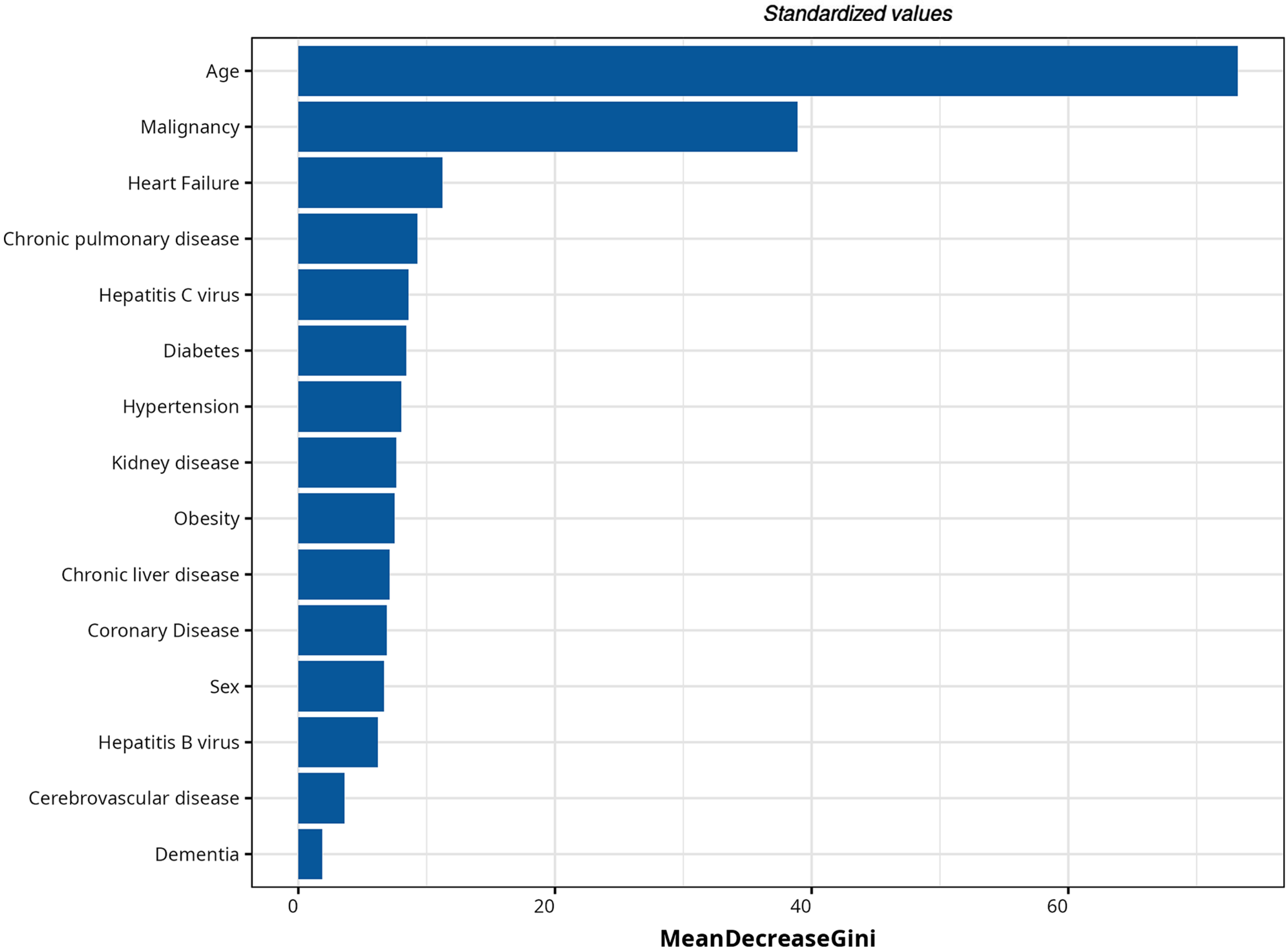

Fig. 4 Importance of features based on the random forest algorithm

alternative to standard logistic regression. LASSO provides more parsimonious models through feature selection. It selects only a subset of relevant variables while irrelevant or noisy variables are dropped, with no effect on the accuracy of the resulting model. LASSO is simple and easy to understand. In the second step, to better interpret the effect of each variable in the resulting model, we used RF to rank the variables, ordered by their importance in the model. Overall, machine learning allows for better explanation of results, which can help clinicians obtain better insight from them. For these reasons, we believe that our results are robust and provide important implications.

Another strength of our study was its inclusion of almost all patients hospitalized due to COVID-19 in Spain within the observation period and therefore our ability to analyze almost all PLWH coinfected and hospitalized with SARS-CoV-2. Furthermore, this research depicts the situation of hospitalized PLWH in Spain over the first 3 years of the pandemic. In addition, we included demographic and clinical estimators to adjust the risk (sex, age, cardiovascular disease). However, we are aware that further studies with stratified CD4 + T lymphocytes counts, viral loads, and ART would help shed light on the risk for adverse outcomes in PLWH.

A major limitation of this study is the lack of information on immunovirological status and on the prevalence of ART. However, Spain reached the aim of the United Nations Programme on HIV, that is, the 90–90–90 target, in 2021 [25, 26], that is, 90% of all PLWH will know their HIV status, 90% of PLWH will receive ART, and 90% of people receiving ART will have viral suppression. It is therefore plausible to assume that at least 80% of inpatients have received antiretroviral therapy and exhibit viral suppression [26]. Bhaskaran et al. also had the limitation of not including data regarding ART, viral load, or CD4 + T lymphocytes status, but they did not consider that to have distorted their findings [10].

Conclusions

PLWH have a greater chance of being hospitalized due to COVID-19 than non-PLWH. PLWH are 25% more likely to die if coinfected with SARS-CoV-2 than non-PLWH. Our results indicate that PLWH should be considered at risk for both hospitalization and adverse outcomes, including mortality. The effects of age, sex, and other comorbidities should also be considered as adjusting estimators because they can modify the clinical course of COVID-19 in PLWH.

Author Contributions Dr. Garcia-Carretero conceived and designed the study, wrote the first draft of the manuscript, and preprocessed and analyzed the data. Drs. Vazquez-Gomez and Rodriguez-Maya made substantial contributions to the interpretation of the results, critically reviewed the first draft of the manuscript, and made valuable suggestions. Drs. Gil-Prieto and Gil-de-Miguel supervised the project and critically reviewed and edited the final draft of the manuscript. All authors read and approved the final manuscript.

Funding This work was supported by AstraZeneca.

Data Availability A contract signed with the Spanish Health Ministry, which provided the data set, prohibits the authors from providing their data to any other researcher. Furthermore, the authors must destroy the data upon the conclusion of their investigation. The data cannot be uploaded to any public repository.

Declarations

Ethics Approval and Consent to Participate This study was approved by the Ethical Board of Universidad Rey Juan Carlos (ID number 2610202334423). No identifying information was included in the manuscript. Because the authors used historical data, informed consent was not necessary. All procedures involving human participants were conducted in accordance with the ethical standards of the responsible institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Consent for Publication Not applicable.

Competing Interests The authors have no conflicts of interest to declare.

References

1. Informes. COVID-19 [Internet]. [cited 2024 Jan 13]. https://www.isciii.es/QueHacemos/Servicios/VigilanciaSaludPublicaRENAVE/EnfermedadesTransmisibles/Paginas/InformesCOVID-19.aspx.
2. Russell CD, Lone NI, Baillie JK. Comorbidities, multimorbidity and COVID-19. Nat Med. 2023;29(2):334–43.
3. Danwang C, Noubiap JJ, Robert A, Yombi JC. Outcomes of patients with HIV and COVID-19 co-infection: a systematic review and meta-analysis. AIDS Res Ther. 2022;19(1):3.
4. Informe no 128 Situación de COVID. -19 en España a 10 de mayo de 2022.pdf [Internet]. [cited 2022 Jun 9]. https://www.isciii.es/QueHacemos/Servicios/VigilanciaSaludPublicaRENAVE/EnfermedadesTransmisibles/Documents/INFORMES/Informes%20COVID-19/INFORMES%20COVID-19%202022/

Informe%20n%C2%BA%20128%20Situaci%C3%B3n%20de%20COVID-19%20en%20Espa%C3%B1a%20a%2010%20de%20mayo%20de%202022.pdf.
5. Tibshirani R. Regression shrinkage and selection via the Lasso. J Royal Stat Soc Ser B (Methodological). 1996;58(1):267–88.
6. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized Linear models via Coordinate Descent. J Stat Softw Articles. 2010;33(1):1–22.
7. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. BMC Med. 2015;13(1):1.
8. Breiman L. Random forests. Mach Learn. 2001;45(1):5–32.
9. Cutler DR, Edwards TC Jr, Beard KH, Cutler A, Hess KT, Gibson J, et al. Random forests for classification in ecology. Ecology. 2007;88(11):2783–92.
10. Bhaskaran K, Rentsch CT, MacKenna B, Schultze A, Mehrkar A, Bates CJ, et al. HIV infection and COVID-19 death: a population-based cohort analysis of UK primary care data and linked national death registrations within the OpenSAFELY platform. Lancet HIV. 2021;8(1):e24–32.
11. Biswas M, Rahaman S, Biswas TK, Haque Z, Ibrahim B. Association of Sex, Age, and comorbidities with mortality in COVID-19 patients: a systematic review and Meta-analysis. Intervirology. 2020;1–12.
12. Costenaro P, Minotti C, Barbieri E, Giaquinto C, Donà D. SARS-CoV-2 infection in people living with HIV: a systematic review. Rev Med Virol. 2021;31(1):1–12.
13. Kopel J, Perisetti A, Roghani A, Aziz M, Gajendran M, Goyal H. Racial and gender-based differences in COVID-19. Front Public Health. 2020;8:418.
14. Danwang C, Endomba FT, Nkeck JR, Wouna DLA, Robert A, Noubiap JJ. A meta-analysis of potential biomarkers associated with severity of coronavirus disease 2019 (COVID-19). Biomark Res. 2020;8:37.
15. Jin JM, Bai P, He W, Wu F, Liu XF, Han DM, et al. Gender differences in patients with COVID-19: focus on severity and mortality. Front Public Health. 2020;8:152.
16. Falutz J, Brañas F, Erlandson KM. Frailty: the current challenge for aging people with HIV. Curr Opin HIV AIDS. 2021;16(3):133–40.
17. Ponatshego PL, Youssouf NF, Mosepele M. Current co-morbidities burden in patients living with HIV in low- and middle-income countries. Curr Opin HIV AIDS. 2021;16(3):163–7.
18. Wang H, Peng J, Wang B, Lu X, Zheng Jz, Wang K et al. Inconsistency between Univariate and multiple logistic regressions. Shanghai Arch Psychiatry 29(2):124–8.
19. Karmen-Tuohy S, Carlucci PM, Zervou FN, Zacharioudakis IM, Rebick G, Klein E, et al. Outcomes among HIV-Positive patients hospitalized with COVID-19. J Acquir Immune Defic Syndr. 2020;85(1):6–10.
20. Freiberg MS. HIV and Cardiovascular Disease — an ounce of Prevention. N Engl J Med. 2023;389(8):760–1.
21. Feinstein MJ. HIV and Cardiovascular Disease: from insights to interventions. Top Antivir Med. 2021;29(4):407–11.
22. Toribio M, Neilan TG, Zanni MV. Heart failure among people with HIV: evolving risks, mechanisms, and preventive considerations. Curr HIV/AIDS Rep. 2019;16(5):371–80.
23. Freiberg MS, Chang CCH, Skanderson M, Patterson OV, DuVall SL, Brandt CA, et al. Association between HIV infection and the risk of heart failure with reduced ejection fraction and preserved ejection fraction in the antiretroviral therapy era: results from the veterans Aging Cohort Study. JAMA Cardiol. 2017;2(5):536–46.
24. Suárez-García I, Gutierrez F, Pérez-Molina JA, Moreno S, Aldamiz T, Valencia Ortega E, et al. Mortality due to non-AIDS-defining cancers among people living with HIV in Spain over 18 years of follow-up. J Cancer Res Clin Oncol. 2023;149(20):18161–71.
25. 90-90-90. - An ambitious treatment target to help end the AIDS epidemic [Internet]. [cited 2024 Jan 13]. https://www.unaids.org/en/resources/documents/2017/90-90-90.
26. Sistemas. de información sobre prevalencia de VIH [Internet]. [cited 2024 Jan 13]. https://www.isciii.es/QueHacemos/Servicios/VigilanciaSaludPublicaRENAVE/EnfermedadesTransmisibles/Paginas/Resultados_Vigilancia_VIH.aspx.

9 COVID-19 and Hematological Malignancies

Journal of
Clinical Medicine

MDPI

Article

Outcomes and Patterns of Evolution of Patients with Hematological Malignancies during the COVID-19 Pandemic: A Nationwide Study (2020–2022)

Rafael Garcia-Carretero [1,*,†], Maria Ordoñez-Garcia [2,*,†], Ruth Gil-Prieto [3] and Angel Gil-de-Miguel [3]

1 Internal Medicine Department, Mostoles University Hospital, Rey Juan Carlos University, 29835 Mostoles, Spain
2 Hematology Department, Mostoles University Hospital, Rey Juan Carlos University, 29835 Mostoles, Spain
3 Department of Preventive Medicine and Public Health, Rey Juan Carlos University, 28922 Alcorcón, Spain; ruth.gil@urjc.es (R.G.-P.); angel.gil@urjc.es (A.G.-d.-M.)
* Correspondence: rgcarretero@salud.madrid.org (R.G.-C.); maria.ordonez@salud.madrid.org (M.O.-G.)
† These authors contributed equally to this work.

Abstract: Background: Early reports suggest that hematological malignancy (HM) is a relevant risk factor for morbidity and mortality in COVID-19. We investigated the characteristics, outcomes, and risk factors for mortality in patients hospitalized with HM and COVID-19. **Methods**: We conducted a retrospective, nationwide study using data from hospitalized patients that were provided by the Spanish Ministry of Health including all patients admitted to a Spanish hospital from 2020 to 2022 with a COVID-19 diagnosis. A descriptive analysis and correlational analyses were conducted. Logistic regression was used to assess mortality in these patients and to calculate odds ratios (ORs). **Results**: We collected data on 1.2 million patients with COVID-19, including 34,962 patients with HMs. The incidence of hospitalization for patients with HMs was 5.8%, and the overall mortality rate was higher than for patients without HMs (19.8% versus 12.7%, $p = 0.001$). Mortality rates were higher for patients with lymphomas, multiple myelomas, and leukemias than for those with myeloproliferative disorders. Having HMs was a risk factor for mortality, with OR = 1.7 (95% CI 1.66–1.75, $p = 0.001$). By subtype, non-Hodgkin lymphomas were the highest risk factor for mortality (OR = 1.7), followed by leukemias (OR = 1.6), Hodgkin lymphomas (OR = 1.58), and plasma cell dyscrasias (OR = 1.24). **Conclusions**: This study identifies the different clinical profiles of patients with HMs who are at a high risk for mortality when hospitalized with COVID-19. As members of a vulnerable population, these patients deserve special prophylactic and therapeutic measures to minimize the effects of SARS-CoV-2 infection.

Keywords: SARS-CoV-2; COVID-19; hematologic malignancy

Citation: Garcia-Carretero, R.; Ordoñez-Garcia, M.; Gil-Prieto, R.; Gil-de-Miguel, A. Outcomes and Patterns of Evolution of Patients with Hematological Malignancies during the COVID-19 Pandemic: A Nationwide Study (2020–2022). *J. Clin. Med.* **2024**, *13*, 5400. https://doi.org/10.3390/jcm13185400

Academic Editor: Swaminathan P. Iyer

Received: 22 August 2024
Revised: 9 September 2024
Accepted: 10 September 2024
Published: 12 September 2024

1. Introduction

Coronavirus disease 2019 (COVID-19) is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), identified in late 2019 [1,2]. Over the course of the pandemic, several variants of the virus presented modified transmissibility or immune escape from previous immunization [3,4]. The COVID-19 pandemic had a significant impact on the world, with 6.7 million confirmed deaths by December 2022 [5]. Currently, the disease is now considered an endemic, involving decreasing severity and mortality in the general population.

Patients with hematological malignancies (HMs) are at high risk for mortality from COVID-19 and are more susceptible to poor prognoses [6]. The clinical course of COVID-19 in patients with HM can be severe, as has been reported in several meta-analyses, with higher rates of hospitalization and higher intensive care unit (ICU) admissions [7–9]. In addition, the case fatality rate is higher in patients who have HMs [6,10]. It has been noted that the mortality rate associated with COVID-19 can reach 34%, although some reviews have reported mortalities up to 51% [7], or 3.5 times the rate in the general population [11].

Patients with HMs have a significantly reduced immune response to COVID-19, either due to the underlying disease itself or to the use of treatments that induce immunosuppression. Using vaccines to induce immunization in such patients could prevent a significant impact of COVID-19, but only half of patients develop a measurable amount of antibodies against SARS-CoV-2 [6].

Apart from the profound immune dysfunction in these patients, several risk factors for mortality have been identified, such as age, cardiovascular comorbidities, and HM subtype [12,13]. However, mortality in HM patients has reportedly been declining since the beginning of the pandemic. Moreover, the demographic and clinical profiles of the patients have changed over time, likely due to vaccination, the breakthrough of new variants of SARS-CoV-2, preventive measures, and effective treatments [9,11,14]. These events may have played a role in the changes in mortality.

HMs, either in the form of solid tumors or hematological neoplasms, are often underreported, as series for hematologic patients are scarce or heterogeneous. Patients with HMs are more vulnerable and deserve special attention, particularly in cases of COVID-19. An improved understanding of the demographic features, clinical profiles, and outcomes of specific populations can improve the management of this population. It should be noted that the future of the evolving endemic remains uncertain, and studies analyzing the epidemiology and risks factors of patients with HMs can offer valuable insight into the interplay between HMs and COVID-19.

This study describes the clinical impact of hospitalization for COVID-19 in patients with HMs. We designed a retrospective, population-based, nationwide study among hospitalized COVID-19 patients with HMs to assess their clinical characteristics and mortality relative to COVID-19 patients without HMs. We also evaluated the predictors associated with mortality in COVID-19 patients with HMs.

2. Materials and Methods

2.1. Study Design and Data Collection

We conducted a population-based, retrospective study using microdata extracted from the National Hospital Data Information System at Hospitalization (MBDS-H) from the Spanish Ministry of Health between the years 2020 and 2022. This registry includes patient data and their diseases, encoded using the 10th Clinical Revision of the International Classification of Diseases (ICD-10-CM). We used the coding for COVID-19 (U07.1) in any diagnostic position (either main or secondary diagnosis) when collecting data.

MBDS-H is an administrative registry that is constructed from discharge reports covering nearly 95% of hospitals in Spain, both public and private. It is estimated that 97% of all discharge reports are collected in this database. The registry includes age, sex, date of admission/discharge, type of hospital, place of residence, and diagnoses at discharge. A new dataset is generated in January of each year. Due to the large amount of data, the data availability has a delay of 1 year. At present, the Spanish Ministry of Health has provided us with data up to 31 December 2022.

In addition to the confirmed diagnosis of COVID-19, patients were categorized using codes to include different types of lymphoma (Hodgkin, follicular, B-cell, T-cell, and NK-cell lymphoma), multiple myeloma and plasma cell leukemia, acute and chronic leukemias (lymphoid, myeloid, and monocytic), and myeloproliferative disorders (polycythemia vera, essential thrombocytosis, myelofibrosis, and chronic myelomonocytic leukemia), which are listed in Appendix A Table A1. We collected data on age, sex, date of admission and discharge, intensive care unit (ICU) admission, and death (if it occurred) for each hospitalized patient. Some chronic conditions of interest were also collected, such as diabetes, hypertension, and other comorbidities. Patients with incomplete data regarding ICU admission, mortality, length of hospitalization, and diagnosed conditions of interest were excluded. No names or personal information were recorded, and data were de-identified to ensure patients' privacy. No data on treatment or immunological status were provided. This study was approved by the Research Committee of our institution.

2.2. Definition of Waves

The observation period covered 3 years (January 2020 to December 2022) and was split into several periods according to local epidemic outbreaks. We based the categorization of the pandemic on the classification of the epidemiological research by Epidemiological National Surveillance Net, which exclusively used data from Spain. Epidemic waves were based on 14-day cumulative incidence. Each turning point indicated the end of one wave and the beginning of the next [15]. Appendix A Table A2 presents the dates that we used to define the epidemic waves.

2.3. Statistical Analyses

We performed descriptive and correlational analyses. We calculated medians and interquartile ranges (IQRs) for continuous variables and absolute numbers and percentages for categorical variables. Hospital length of stay was defined as the total number of days of stay divided by the total number of hospitalizations and expressed as medians and IQRs. Deaths and ICU admissions were divided by the total number of hospitalizations to calculate the mortality rate and ICU admission rate, respectively. Both of these parameters were expressed as absolute numbers and percentages. Chi-square and Mann–Whitney U tests were performed as tests of independence wherever appropriate. In multivariate analysis, we performed binary logistic regressions to assess the effects of explanatory variables on mortality (considered the response variable). We calculated the regression coefficients and odds ratios (ORs) of the variables of interest.

For all tests, the level of statistical significance was set at $p < 0.05$. We used Python language version 3.11.2 and R language version 4.3.2 (Vienna, Austria) on a Debian 12 GNU/Linux workstation.

3. Results

3.1. Descriptive Analyses

The first known COVID-19 case in Spain was detected in late January 2020, but the first significant hospitalization of a patient due to COVID-19 occurred in early February 2020, coinciding with the spread of the virus in European countries like Italy. Spain's first confirmed death from COVID-19, which was retrospectively identified, happened on 13 February 2020. Between 1 February 2020 and 31 December 2022, almost 1.2 million patients were hospitalized due to COVID-19 (Table 1). The cohort of patients with HMs included 34,962 individuals (2.94% of patients hospitalized with COVID-19). It was estimated by the Spanish Oncology Society [16] that there were 198,507 patients with HMs in Spain in 2022. That is, the prevalence of HMs in Spain was estimated to be 0.41%. The rate of hospitalization of patients in the general population was 0.83% per year between 2020 and 2022. For individuals with HM, we calculated an estimated ratio for hospitalized patients of 5.87% per year, that is, greater than the general population.

Men were more likely to be admitted to the hospital (54.9% globally). The median age of patients with HM was 75, and HM patients had a longer hospital stay than patients without HM (13.8 vs. 10.8 days). The ICU admission rate and mortality rate were also higher in patients with HM (9.8% and 19.8%, respectively). Additionally, we analyzed some comorbidities, including diabetes, hypertension, coronary disease, and so on. Table 1 presents the demographic and clinical features of our population. It should be noted that the prevalence of solid malignancies was lower in patients with HM.

Table 1. Demographic and clinical characteristics of patients in our cohort.

	Patients with HM	Patients without HM	All Patients	p-Value
Patients	34,962	1,153,198	1,188,160	NA
Sex (men, %)	59.4%	54.8%	54.9%	0.001
Age (IQR)	75 (19)	73 (27)	73 (27)	0.001
Hospital LOS in days (IQR)	13.8 (11)	10.8 (8)	10.9 (8)	0.001
ICU admissions	3420	97,158	100,578	NA
ICU (%)	9.8%	8.4%	8.5%	0.001
ICU LOS in days (IQR)	8.5 (17.2)	8 (20.3)	8 (20.2)	0.598
Deaths	6925	146,219	153,144	NA
Mortality rate (%)	19.8%	12.7%	12.9%	0.001
Comorbidities				
Diabetes	22.2%	24.5%	24.4%	0.001
Hypertension	32%	31.9%	31.9%	0.805
Coronary disease	8.6%	9.8%	9.7%	0.001
Heart failure	15.9%	17%	17%	0.001
Dementia	3.9%	6.6%	6.5%	0.001
Chronic kidney disease	17%	15.5%	15.6%	0.001
Chronic liver disease	0.7%	0.6%	0.6%	0.017
Solid tumor	5.8%	7.4%	7.4%	0.001
Obesity	7.1%	12%	11.8%	0.001
Chronic pulmonary disease	12.7%	14.8%	14.8%	0.001
Cerebrovascular disease	0.8%	1.4%	1.4%	0.001

Categorical variables are expressed in absolute numbers and percentages. Continuous variables are expressed as medians (interquartile ranges). Chi-square test for continuous variables and Mann–Whitney U test for categorical variables were performed as tests of independence. NA: not applicable. ICU: intensive care unit. IQR: interquartile range. HM: hematological malignancy. LOS: length of stay.

We split the whole observation period into seven epidemic waves, as shown in Appendix A Table A2. The daily admissions are shown in Figure A1, while the monthly evolution of the pandemic during the 3 years of observation is shown in Figure 1. The evolution of mortality is shown in Figure 2. The admission and mortality trends match. As shown in Appendix A Table A3, in-hospital mortality reached its highest values in January 2021 and January 2022 (up to 26.6% in the third period) and its lowest value in December 2022 (14.7%). In spite of the overall decreasing mortality rate, cumulative deaths were 6925, with almost 1000 of them having occurred by May 2020 (Figure A2).

We also explored the relationship between age and sex in a population pyramid (Figure 3). The main characteristics related to sex among individuals with HM are given in Table 2. Admission was more likely among men (59.4%). The mortality rate was similar between men and women (20.1% vs. 19.3%). In addition, the prevalences of diabetes, coronary disease, chronic kidney disease, solid malignancies, and chronic pulmonary diseases were higher in men, while hypertension, heart failure, and obesity were more prevalent in women. The rate of ICU admission was higher in men than in women (10.6% vs. 8.6%).

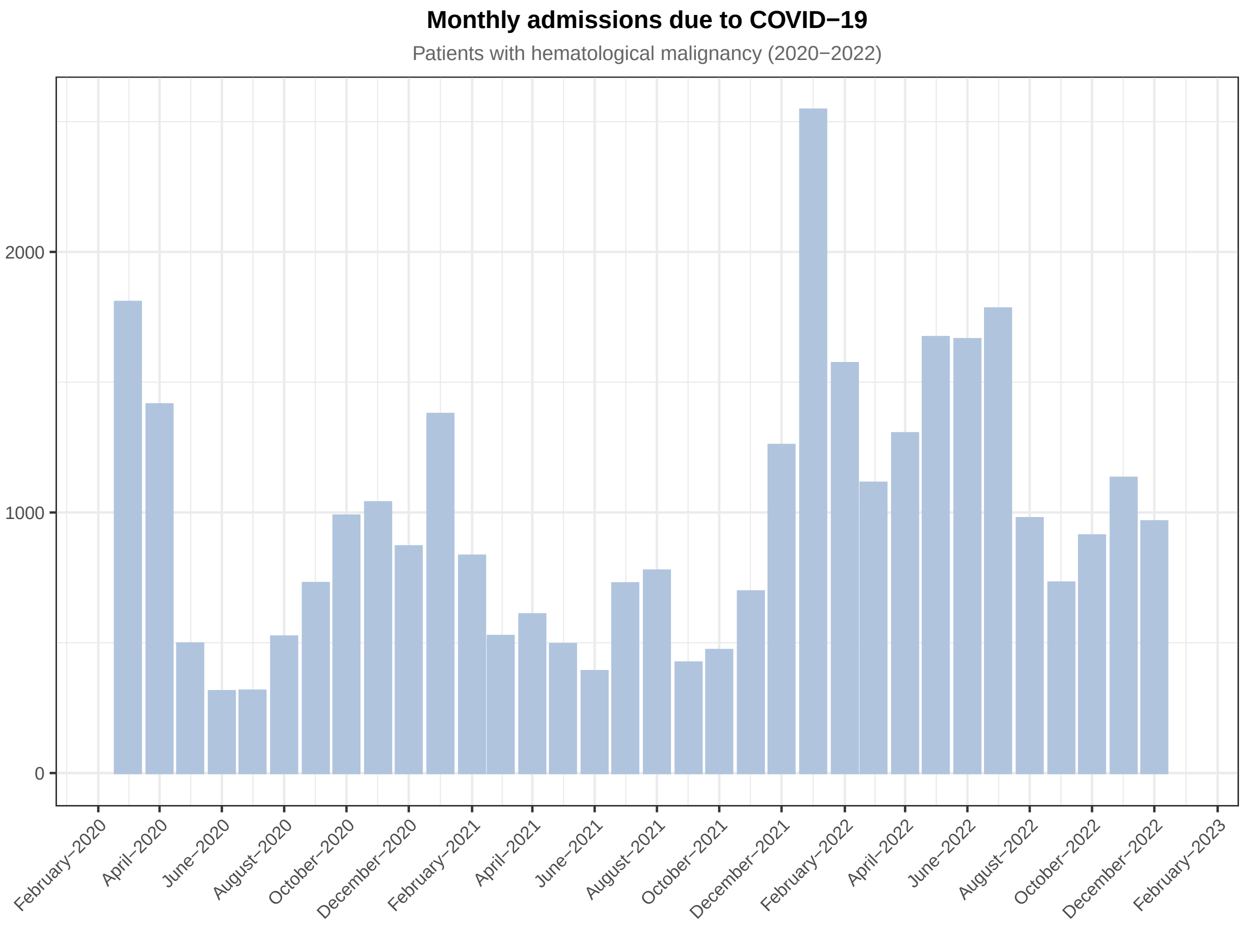

Figure 1. Evolution and trend of the COVID-19 pandemic with respect to hospitalized patients with hematological malignancies over the studied period, aggregated by month.

Table 2. Demographic and clinical characteristics of the studied cohort of patients with hematological malignancy.

	Men	**Women**	**Total**	***p* Value**
No. patients	(*n* = 20,764)	(*n* = 14,198)	(*n* = 34,962)	NA
Age, years (median, IQR)	74 (19)	76 (19)	75 (19)	0.001
Hospital LOS in days (median, IQR)	14 (12)	13.5 (11)	13.8 (11)	0.005
ICU admissions	2194	1226	3420	NA
ICU (%)	10.6	8.6	9.8	0.001
ICU LOS in days (median, IQR)	9 (17.6)	8 (16.2)	8.5 (17.2)	0.028
Deaths	4182	2743	6925	NA
Mortality rate (%)	20.1%	19.3%	19.8%	0.008
Comorbidities				
Diabetes	23.7%	20%	22.2%	0.001
Hypertension	31.4%	32.8%	32%	0.003

Table 2. *Cont.*

	Men	Women	Total	*p* Value
Coronary disease	11%	5.1%	8.6%	0.001
Heart failure	14.5%	17.8%	15.9%	0.001
Dementia	3.2%	5%	3.9%	0.001
Chronic kidney disease	18.1%	15.4%	17%	0.001
Chronic liver disease	0.8%	0.6%	0.7%	0.037
Solid tumor	7%	4%	5.8%	0.001
Obesity	6.2%	8.4%	7.1%	0.001
Chronic pulmonary disease	17.6%	5.5%	12.7%	0.001
Cerebrovascular disease	0.8%	0.9%	0.8%	0.457

Categorical variables are expressed in absolute numbers and percentages. Continuous variables are expressed as medians (interquartile ranges). Chi-square test for continuous variables and Mann–Whitney U test for categorical variables were performed as tests of independence. NA: not applicable. ICU: intensive care unit. IQR: interquartile range. HM: hematological malignancy. LOS: length of stay.

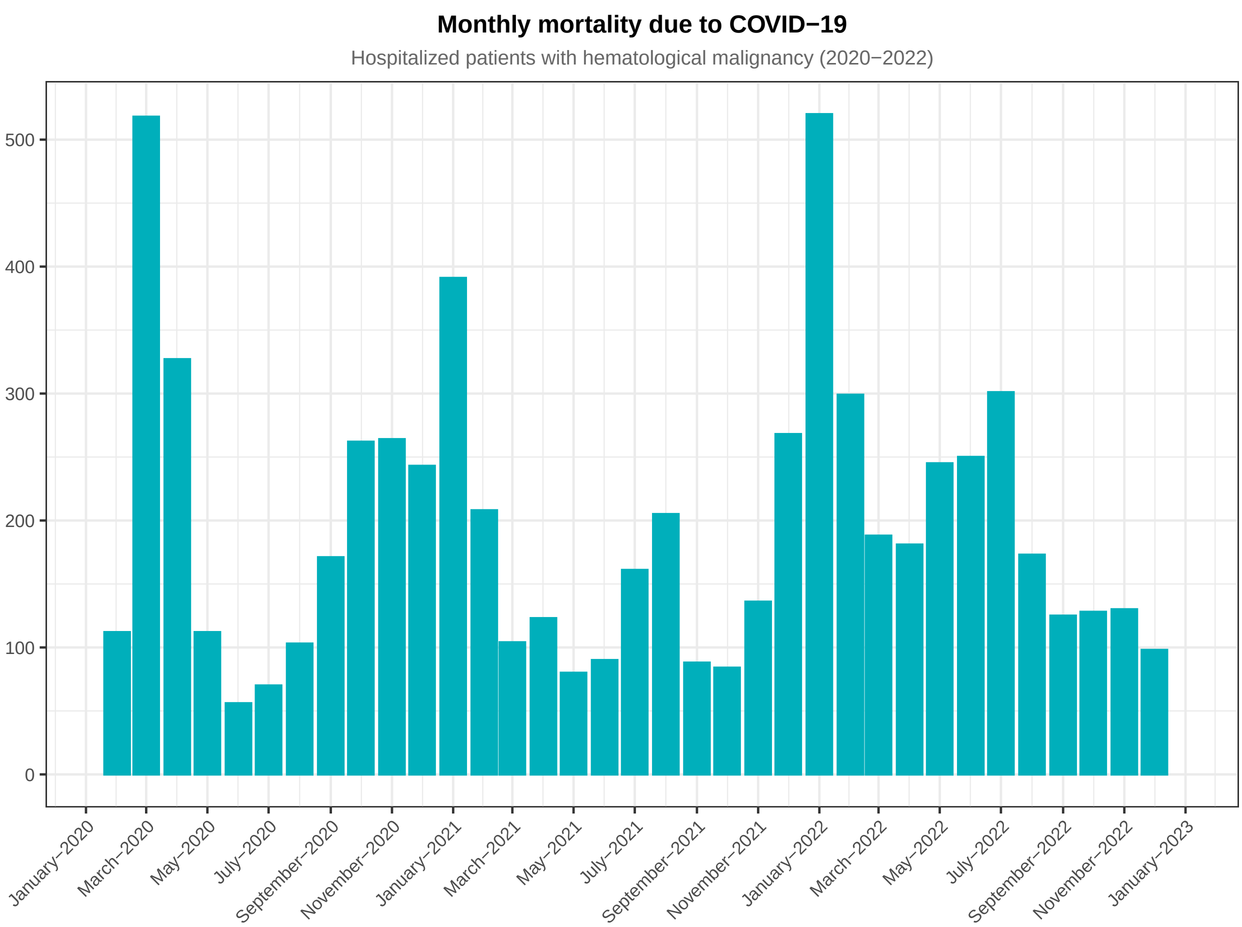

Figure 2. In-hospital mortality of patients with COVID-19 and hematological malignancy.

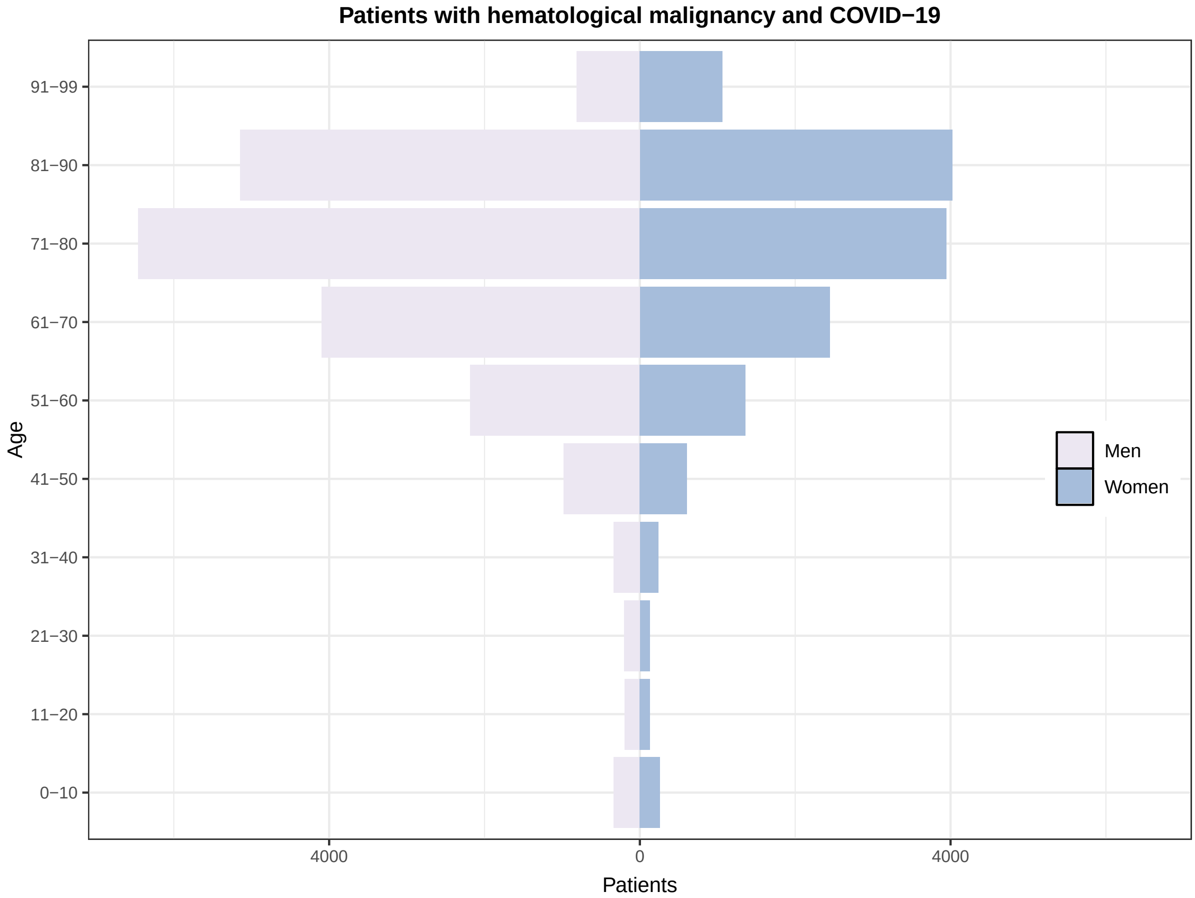

Figure 3. Population pyramid showing the distribution according to sex and age between 2020 and 2022.

3.2. Comorbidities

The subgroup of patients with myeloproliferative disorders had the highest prevalence of diabetes, coronary disease, heart failure, dementia, chronic kidney disease, chronic pulmonary disease, and cerebrovascular disease (Table 3).

Table 3. Incident cases, deaths, and demographic characteristics of hospitalized patients with hematological malignancy.

	Hodgkin Lymphoma	Follicular and B-Cell Lymphoma	T/NK-Cell Lymphomas	Multiple Myeloma and Malignant Plasma Cell Neoplasms	Leukemias	Chronic Myeloproliferative Disorders
No. patients	860	10,482	766	6171	12,628	5308
Adjusted hospitalization rate per year	1%	3.7% *		12.6%	7.9%	ND
Sex (men, %)	64.4	59.1	64.8	57.3	61.4	58.6
Age, years (median, IQR)	61 (32.2)	73 (18)	69 (21)	75 (16)	76 (19)	79 (18)
Hospital length of stay in days (median, IQR)	14.3 (12.2)	14.7 (13)	15.4 (13)	12.7 (10)	14 (12)	12.3 (9)
ICU admissions	100	1111	90	499	1242	483

Table 3. *Cont.*

	Hodgkin Lymphoma	Follicular and B-Cell Lymphoma	T/NK-Cell Lymphomas	Multiple Myeloma and Malignant Plasma Cell Neoplasms	Leukemias	Chronic Myeloproliferative Disorders
ICU (%)	11.6	10.6	11.7	8.1	9.8	9.1
ICU length of stay in days (median, IQR)	6 (12)	9 (18)	9.5 (19)	7 (15)	9 (16)	9 (15)
Deaths	143	2104	155	1123	2757	976
Mortality rate (%)	16.6%	20.1%	20.2%	18.2%	21.8%	18.4%
Comorbidities						
Diabetes	17.4%	21.5%	20.9%	20.7%	23.7%	24.2%
Hypertension	22.1%	32.2%	31.3%	33.8%	31.7%	31.3%
Coronary disease	5.5%	7.3%	6.5%	8.2%	9.2%	12.1%
Heart failure	8.8%	12.2%	12.5%	17.6%	16.4%	23.8%
Dementia	1.5%	3%	2.5%	3.8%	3.9%	6.7%
Chronic kidney disease	8.4%	13.1%	10.7%	23.4%	16.2%	23.4%
Chronic liver disease	1.2%	1%	0.4%	0.6%	0.5	0.7%
Solid tumor	4.5%	6.6%	6.7%	5.2%	5.1%	6.7%
Obesity	6.4%	7.1%	6.3	7.0	7.0	7.9
Chronic pulmonary disease	11.6%	11.9%	11.7%	12.7%	12.0%	17.6%
Cerebrovascular disease	0.5%	0.5%	0.5%	0.8%	0.9%	1.3%

Categorical variables are expressed as absolute numbers and percentages. Continuous variables are expressed as medians (interquartile ranges). NA: not applicable. ICU: intensive care unit. IQR: interquartile range. ND: non-available data. LOS: length of stay. * Data for non-Hodgkin lymphoma are combined data from follicular, B-cell, and T/NK-cell lymphomas.

3.3. Overall and Subtype-Related Mortality

According to the Spanish Oncology Society [16], the hospitalization rates per year were 1% (Hodgkin lymphoma), 3.7% (non-Hodgkin lymphoma), 12.6% (multiple myeloma), and 7.9% (leukemias). No data were available on chronic myeloproliferative disorders. Table 3 presents the characteristics of our population by subtype of HM.

Mortality rates were higher in patients with lymphomas, multiple myelomas, and leukemias than in those with myeloproliferative disorders. In-hospital mortality rates are plotted in Figures 4 and A3. Overall, we observed a decline in mortality for all types of HM over time. Hodgkin lymphoma maintained a certain stability over the 3 years of observation, while the remaining malignancies settled down below 15% of mortality by the end of 2022 (see Appendix A Table A4).

3.4. Age-Adjusted Mortality

We also calculated the crude and age-standardized (adjusted) mortality rates using the direct method, as shown in Figure 5 and Appendix A Table A5. The age groups remained stable across the 3 years of the observation period. Figure 5 shows that, in 2022, patients over 60 experienced an increase in mortality rate, but no significant changes were observed in the rest of the groups.

3.5. Multivariate Analyses

We performed logistic regression to assess the effect of HM on mortality due to COVID-19 in the general population. The unadjusted OR was 1.7 (95%CI 1.66–1.75, p = 0.001), that is, hospitalized patients with HM had a 70% greater chance of dying due to COVID-19. Adjusted by other estimators (i.e., sex, age, and comorbidities), the OR did not change, as shown in Table 4.

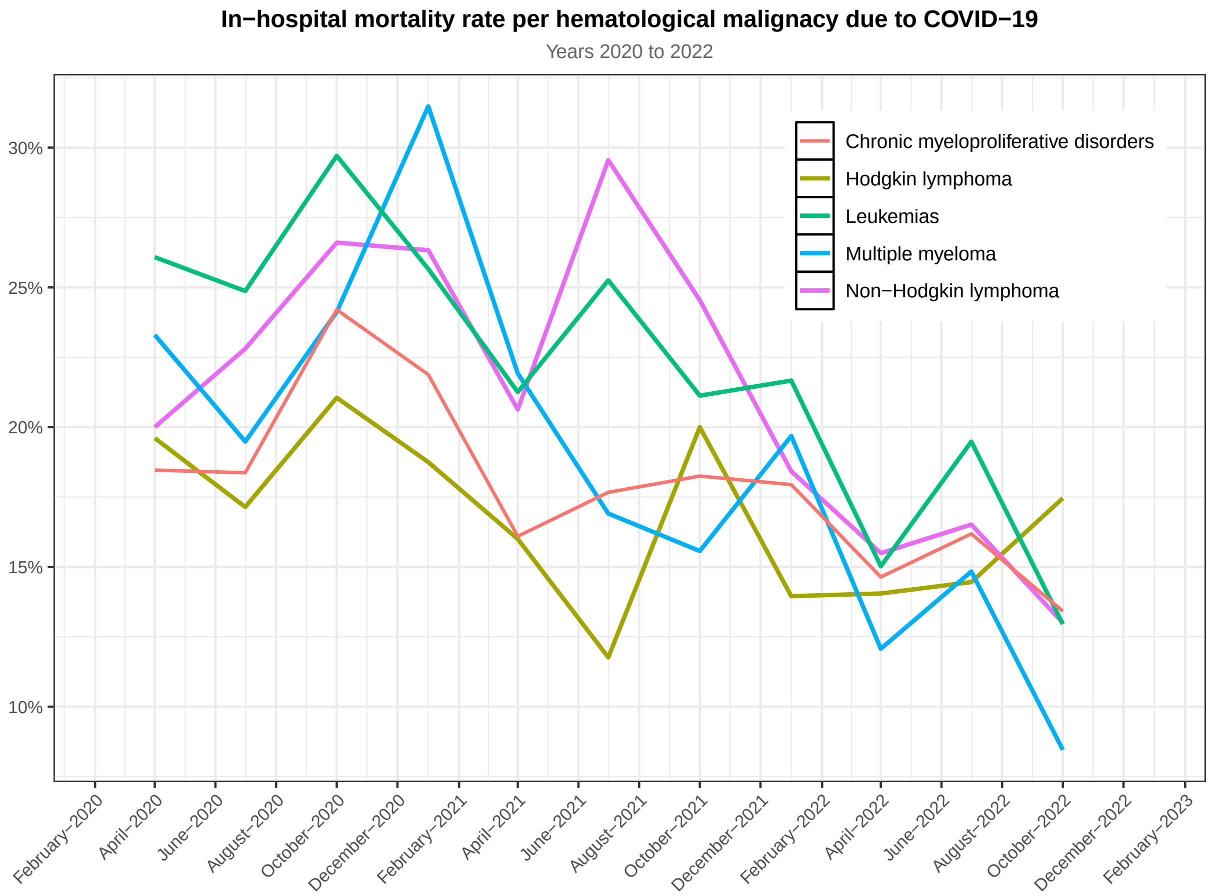

Figure 4. Mortality rates per type of malignancy. Non-Hodgkin lymphomas are a combination of follicular, B-cell, and T/NK-cell lymphomas. Multiple myeloma also includes plasma cell neoplasms.

Table 4. Multivariate analyses (logistic regression) for the entire cohort using mortality as the dependent variable.

	OR (Adjusted)	95%CI	*p* Value
Sex (men)	1.35	1.34–1.37	0.001
Age	1.05	1.05–1.05	0.001
Hematological malignancy	1.71	1.66–1.76	0.001
Hodgkin lymphoma	2.22	1.83–2.67	0.001
Follicular and B-cell lymphoma	1.86	1.77–1.95	0.001
T/NK-cell lymphomas	2.14	1.77–2.57	0.001
Multiple myeloma	1.49	1.39–1.59	0.001
Leukemias	1.82	1.74–1.91	0.001
Chronic myeloproliferative disorders	1.11	1.03–1.2	0.001
Comorbidities			
Solid tumor	2.33	2.29–2.38	0.001
Diabetes	1.01	1–1.02	0.15
Coronary disease	1.09	1.07–1.11	0.001
Heart failure	1.07	1.05–1.08	0.001
Hypertension	0.93	0.92–0.95	0.001
Obesity	0.96	0.94–0.98	0.001
Dementia	1.24	1.22–1.27	0.001
Cerebrovascular disease	2.06	1.99–2.14	0.001
Chronic liver disease	1.6	1.5–1.7	0.001
Chronic kidney disease	1.14	1.12–1.16	0.001
Chronic pulmonary disease	0.71	0.7–0.73	0.001

Next, we performed a multivariate analysis of the cohort of patients with HM, that is, we explored the effects of several predictors in this specific population (Table 5). The model applying binary logistic regression identified non-Hodgkin lymphomas as major risk factors for mortality, followed by leukemias, Hodgkin lymphomas, and plasma cell dyscrasias. The model also evaluated the effects of age, sex, and included comorbidities. ICU admission and length of ICU stay were considered predictors of mortality, as they can increase the risk of in-hospital death.

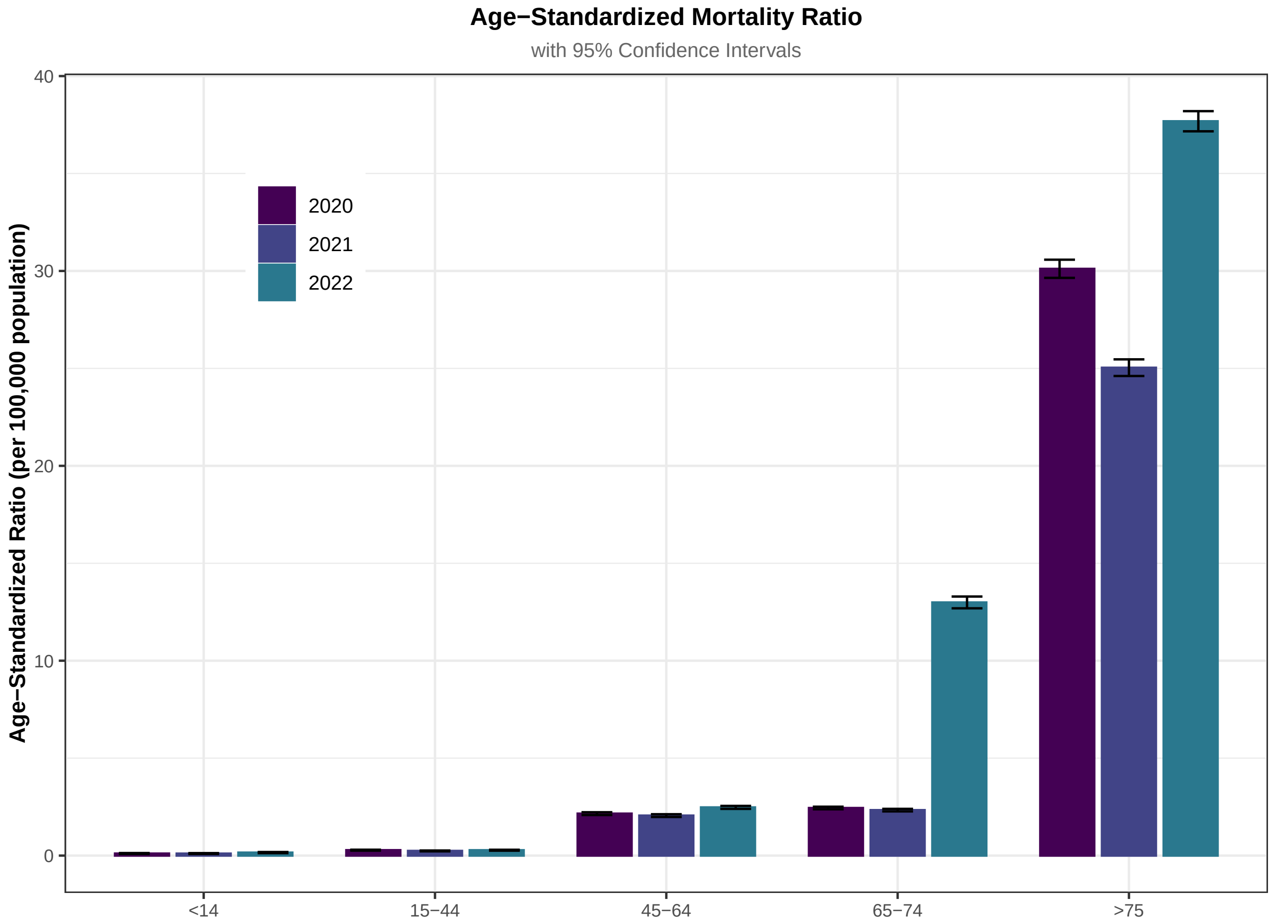

Figure 5. Age-standardized (adjusted) mortality rate using the direct method.

Table 5. Multivariate analyses of the cohort with hematological patients only (binary logistic regression).

	OR (Adjusted)	95%CI	*p* Value
Sex (men)	1.12	1.06–1.19	0.001
Age	1.03	1.02–1.03	0.001
Length of hospital stay	0.97	0.97–0.98	0.001
ICU admission	6.68	6.15–7.26	0.001
Length of ICU stay	1.03	1.03–1.04	0.001
Type of malignancy			
Hodgkin lymphoma	1.58	1.25–1.98	0.001
Follicular and B-cell lymphoma	1.52	1.31–1.77	0.001
T/NK-cell lymphomas	1.68	1.33–2.1	0.001
Multiple myeloma	1.24	1.06–1.45	0.01
Leukemias	1.6	1.39–1.84	0.001
Chronic myeloproliferative disorders	1.05	0.92–1.2	0.48

Table 5. *Cont.*

	OR (Adjusted)	95%CI	*p* Value
Comorbidities			
Solid tumor	1.55	1.4–1.72	0.001
Diabetes	0.92	0.87–0.99	0.02
Coronary disease	1.05	0.96–1.15	0.3
Heart failure	1.25	1.17–1.35	0.001
Hypertension	1.03	0.96–1.09	0.4
Obesity	0.97	0.87–1.08	0.55
Dementia	1.12	0.99–1.27	0.07
Cerebrovascular disease	2.04	1.59–2.62	0.001
Chronic liver disease	1.28	0.94–1.73	0.11
Chronic kidney disease	1.15	1.07–1.24	0.001
Chronic pulmonary disease	0.73	0.67–0.79	0.001

4. Discussion

We report the first population-based, nationwide study of the epidemiology and the risk of in-hospital death in patients with prevalent HMs and COVID-19 in Spain. We included almost all patients hospitalized with COVID-19 from 2020 to 2022. Our aim was to gain insight into the characteristics of this specific population. The first concern that may arise regards the high rate of hospitalization (5.8%) in patients with HM relative to the general population (0.84%), which indicates the impact of infection with the SARS-CoV-2 virus in this specific population. The overall mortality rate of patients with HM in our study was 19.8%. We also found that those with male sex and elderly age had a significant risk for hospitalization, in line with other studies [11,14]. Demographic profiles in terms of sex, age, and comorbidities in patients with HM were similar to those found in the general population.

4.1. ICU Admissions

We identified a decreased ICU admission rate (9.8%) relative to early meta-analyses [7,17], which reported a pooled ICU admission of 21%. A meta-analysis by Langerbeins et al. [7] included studies of patients with COVID-19 and HM during the early stages of the pandemic. The researchers collected data from retrospective cohorts, prospective registries, and population surveys, identifying an ICU admission rate between 10% and 24%. We cannot explain these discrepancies among studies. One hypothesis could be that those meta-analyses reported data from the earliest stages of the pandemic, when ICU settings were overwhelmed and the disease was at its most severe. However, we observed increasing rates of ICU admissions over the analyzed period, with the lowest rate during the earliest stages. Moreover, in our cohort, we did not find rates above 14% in our analysis of the observation period by waves. Another concern that may arise regards the convenience of transferring the more severe patients with HM to these clinical areas. The percentage found in our study (9.8%) could not represent the severity of patients with HM. In fact, in an overwhelmed ICU setting, patients with any malignancy may have been denied admittance to these units. Therefore, our figures for ICU admissions could under-represent the severity of COVID-19 in this population. Unfortunately, we do not have data on the criteria for ICU admissions for our patients.

4.2. Overall Mortality

Early reports have reported different outcomes for patients with HM. A meta-analysis reported a mortality rate between 14% and 51% [7]. A Turkish study [18] estimated the risk of death in patients with HM at 14% compared to the general population (7%), which is similar to our results. However, overall, the outcomes and conclusions differ from ours. The significant disparity in the range of mortality across studies cannot be easily explained. The cited studies were based on large cohorts of patients with HM, but some were heterogeneous [7]. Some studies have reported rates of mortality in both the ambulatory and hospitalized population, while others have reported data from patients only with HM [12]. Another plausible explanation is that, similar to the results for ICU admission, these other studies and meta-analyses were published during the early stages of the pandemic, when mortality was the highest and ICU admissions were reserved for selected patients only.

A Spanish study of 1166 hospitalized patients with HM reported an overall mortality rate of 32% [19]. This rate is quite different from the rate we found (19.8% in patients with HM). This discrepancy may be explained by the difference in observation periods and the constrained area where the other study was conducted. That study only considered patients in Madrid (with a population of 6.6 million, 14% of the total Spanish population) over the first 12 months of the pandemic, before vaccination was widely available. It should also be noted that the region of Madrid was the most affected by COVID-19 during the first year of the pandemic [11], which might explain the high mortality rate.

We reported a noticeable increase in COVID-19-related mortality in Spain during late 2022, although it was milder compared to earlier waves. Several factors may have contributed to this increase. We can speculate about some factors. Omicron subvariants like BA.4, BA.5, and later BA.2.75 were circulating in Europe, including Spain. There were more immune-evasive. Although they caused mild disease in healthy individuals, these strains could still lead to severe outcomes in vulnerable populations like the elderly or immunocompromised. Immunity waning can be also an interesting factor. By late 2022 most of the population in Spain had been vaccinated, but their immunity might have waned. The reduced immunity against SARS-CoV-2 might have contributed to increase mortality during this period. Another factor could be the impact on vulnerable populations. The elderly, the immunocompromised, and those patients with comorbidities might have been at a high risk even during mild peaks. Also, the co-circulation of influenza and Respiratory Syncytial Virus in late 2022, alongside SARS-CoV-2, could have contributed to the increase in mortality. So, although merely speculation, we cannot attribute the increase in mortality solely to the variants but to a more complex interplay of the mentioned factors.

4.3. Mortality in Subtypes of Malignancy

According to Vijenthira et al. [17], who examined 2192 patients in a meta-analysis, the pooled risk for mortality was 41% for leukemias, 33% for plasma cell dyscrasias, 32% for lymphoma, and 34% for myeloproliferative disorders. The aforementioned Spanish study reported that the most common subtypes were non-Hodgkin lymphoma and multiple myeloma [19]. Another Spanish study that included immunosuppressed patients found the highest risk of mortality in patients with leukemia and lymphoma [13], although no subtype of malignancy was reported. Overall, the results for mortality in these studies are higher than ours. Here as well, we consider the heterogeneity of the given studies as the cause of the difference between their results and ours.

The most relevant study that examined patients with HM infected by SARS-CoV-2 is EPICOVIDEHA [6,12], a survey supported by the European Hematology Association (EHA). EPICOVIDEHA found that non-Hodgkin lymphoma (31%) and plasma cell disorders (17%) were the most prevalent. That study also found that leukemias and myelodysplastic syndrome had the highest mortality rate (in a range from 18% to 21%), in line with our study. EPICOVIDEHA reported an overall decrease in mortality over time and across all subtypes and a steady decline in mortality per type of malignancy, from a global 38.2% to 5.3% in late December 2022. While these results are consistent with the global trends found in our study, our results are less optimistic. We identified a decline in mortality across all subtypes of HM, beginning in the final quarter of 2021, producing an overall rate of 15% in late December 2022. However, during that period, only multiple myeloma and plasma cell dyscrasias showed a rate as low as 8.5%, with the remaining subtypes being above 13%, and Hodgkin lymphoma having a 17.5% mortality rate. That is, our reported mortality rate is higher than the results from EPICOVIDEHA. In spite of these discrepancies, the results are consistent with an overall decline in mortality rate in Spain [11], which was probably concurrent with the beginning of vaccination rollout in patients with malignancies in Spain in March 2021. Unfortunately, we have no data regarding the role of vaccination in patients with HM, and research on this topic is purely speculative.

4.4. Comorbidities and Predictors of Mortality

We observed that the prevalence of comorbidities in COVID-19 patients with HM did not differ from that of the general population. It should be noted that solid malignancies were more prevalent in patients without HM. We also found that the subgroup with myeloproliferative disorders had the highest prevalence of cardiovascular comorbidities such as cerebrovascular disease, diabetes, coronary disease, and heart failure, as well as dementia, chronic kidney disease, and chronic pulmonary disease, probably because this subgroup includes an older population that is vulnerable to chronic diseases. From 60 years old and older, we observed an increase in deaths, indicating that mortality is strongly associated with age.

The interest in comorbidities is not only descriptive but also predictive. Some studies have associated certain conditions with poor prognosis. The meta-analysis by Langerbeins et al. [7] summarized the main risk factors for severity. Among comorbidities, the authors identified demographic features, such as older age and male sex, as well as several comorbidities, such as cardiovascular features and chronic diseases. The meta-analysis reported factors such as race, cytotoxicity, anti-cancer therapy, and type of malignancy. For HM, lymphomas were the most important predictive marker related to mortality. Regalado et al. [20] found that older age, heart disease, and chronic kidney disease were the most important risk factors for mortality in a multivariate analysis of patients with lymphoma.

The EPICOVIDEHA registry [6] found that cardiomyopathy (35%), diabetes (14%), and chronic pulmonary disease (13%) were the most common conditions in patients with HM. It also reported several predictors associated with mortality using Cox regression and hazard ratios (HRs) as follows: age (HR 1.03), two or more comorbidities (HR 1.24), and non-HM (HR 1.83). Chronic cardiomyopathy, chronic liver disease, and chronic kidney disease were the most relevant predictors in a multivariate analysis assessing mortality [12].

Chronic pulmonary obstructive disease (COPD) is commonly associated with smoking. Interestingly, the EPICOVIDEHA registry found that chronic pulmonary disease was frequent among patients with HM, but it was not related to higher mortality. This is surprising, as COPD is generally considered a risk factor for severe COVID-19, where an increased odds ratio (OR) would be expected. Several factors may explain this lower-than-expected mortality, but they are purely speculative. First, COPD patients are likely to receive closer monitoring and earlier therapeutic interventions, potentially reducing the severity of COVID-19. Additionally, respiratory therapies, such as bronchodilators or corticosteroids, could help manage symptoms early on. The anti-inflammatory effects of corticosteroids may also lower the risk of the hyperinflammatory *cytokine storm* associated with severe COVID-19. Survivor bias might play a role, as HM patients with COPD may be younger or better managed, making them more resilient. Furthermore, surviving the early pandemic waves could have provided these patients with better care protocols or timely vaccinations, reducing their mortality risk.

4.5. Limitations

Our study had several limitations. Its main limitation was the use of an administrative database. Although electronic health records help researchers collect data and make them available for research, some clinical data were not available. Information on vaccination and medications was not available, so we could not assess the effect of immunization or certain treatments, such as specific anti-cancer treatments or immunotherapies. As a result, the clinical stage of the patients could not be evaluated, so we could not assess risk stratification in patients with HM. This issue is inherent to the characteristics of this dataset and the coding process for MBDS-H.

In addition, some patients' data may have been duplicated. The database provided by the Ministry of Health is anonymized by default, so we could not identify duplicated patient data. It is important to remember that patients with HM tend to have frequent medical visits; thus, each entry of data in our study would represent one given hospitalization rather than one given patient. While this issue may result in bias, every hospitalization is unique. We consider that overestimation is unlikely, due to the large number of patients included.

We included hospitalized patients, so another limitation is that we did not include ambulatory patients. This may have led our outcomes to under-represent reality.

Another limitation is that we assumed that mortality was directly related to COVID-19. However, some HMs are life-threatening on their own. We are aware that COVID-19 can be a healthcare-associated infection, and was particularly so in later 2022. In fact, other authors have reported that in-hospital mortality in patients with HM may increase by 50% if it overlaps with COVID-19 infection [21,22].

4.6. Strengths

Regardless of these limitations, we consider that the main strength of our study is its inclusion of almost all hospitalized patients with COVID-19 and HM, such that our results (ICU admission of 9.8% and mortality rate of 19.8%) are reliable and capture the impact of COVID-19 in this specific population. Our study represents the largest set of data on hospitalized patients with HM and COVID-19. We are aware of the heterogeneity of results across different studies, but this can be explained by the bias introduced to a given study according to the specific population considered in that study. By conducting a Spain-wide study, we reliably assessed the risk of in-hospital mortality in this patient group based on data from an unrestricted population in terms of place or region.

Our findings may be useful for understanding the impact of COVID-19 in patients with HM. Meta-analyses and nationwide studies such as ours may be very relevant because they analyze large cohorts and can identify trends across specific populations. Collecting knowledge and analyzing data on the impact of COVID-19 in specific at-risk groups is critical for the endemic phase of COVID-19. Our reported outcomes highlight the burden of healthcare in patients with HM. Future prevention strategies should emphasize this population. Moreover, patients with HM should be considered for ICU admission when appropriate, given that they can recover from COVID-19 [6].

Early prophylactic or therapeutic measures should be promptly initiated in such patients to prevent severe forms of COVID-19. Our study demonstrates the vulnerability of patients with HM. We suggest that treatment for COVID should not be delayed. Physicians should also be aware of the need for close monitoring and tailored care for these patients.

5. Conclusions

This study presents the first comprehensive, nationwide analysis of the epidemiology and risk factors for in-hospital mortality among patients with hematological malignancies (HMs) and COVID-19 in Spain. It highlights the unique vulnerabilities of this population. Our findings indicate a significantly higher hospitalization rate (5.8%) for patients with HM compared to the general population (0.84%), and an overall mortality rate of 19.8%.

We found discrepancies in mortality rates between our research and other studies, potentially due to differences in study populations, observation periods, and pandemic phases. Notably, the mortality rate has declined over time, possibly reflecting improved management and vaccination, although we are aware that specific data on vaccination were not available in this study.

Our research highlights the need for targeted prophylactic and therapeutic interventions. Close monitoring and tailored care are essential to mitigate the severe outcomes of COVID-19 in this high-risk population. Our study provides valuable insights into the impact of COVID-19 on patients with HM in a population-based setting.

COVID-19 has turned from a pandemic to an endemic, and it will pose future healthcare challenges in both general and specific populations. This is why a better understanding will benefit the creation of optimized strategies for patients with HM. Although the severity of the disease and mortality have decreased over time, patients who have HM are at high risk, as the impact of COVID-19 remains high. Our study highlights the importance of preventive measures, targeted interventions, vaccination, and prompt therapy to improve survival and prevent severe forms of COVID-19.

Author Contributions: Conceptualization, R.G.-C. and M.O.-G.; methodology, R.G.-C.; software, R.G.-C.; validation, R.G.-C. and M.O.-G.; formal analysis, R.G.-C.; investigation, R.G.-C. and M.O.-G.; data curation, R.G.-C.; writing—original draft preparation, R.G.-C.; writing—review and editing, R.G.-P. and A.G.-d.-M.; visualization, R.G.-C.; supervision, R.G.-P. and A.G.-d.-M.; project administration, R.G.-P. and A.G.-d.-M.; funding acquisition, R.G.-C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by AstraZeneca.

Institutional Review Board Statement: This study was approved by the Ethical Board of Universidad Rey Juan Carlos (ID number 2610202334423). No identifying information was included in the manuscript. Because the authors used historical data, informed consent was not necessary. All procedures involving human participants were conducted in accordance with the ethical standards of the responsible institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed Consent Statement: Not applicable.

Data Availability Statement: A contract signed with the Spanish Health Ministry, which provided the dataset, prohibits the authors from providing their data to any other researcher. Furthermore, the authors must destroy the database upon the conclusion of their investigation. The database cannot be uploaded to any public repository. However, we uploaded Python code and some pieces of the whole dataset to a public repository at https://github.com/rafalinux/covid-hematological.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A

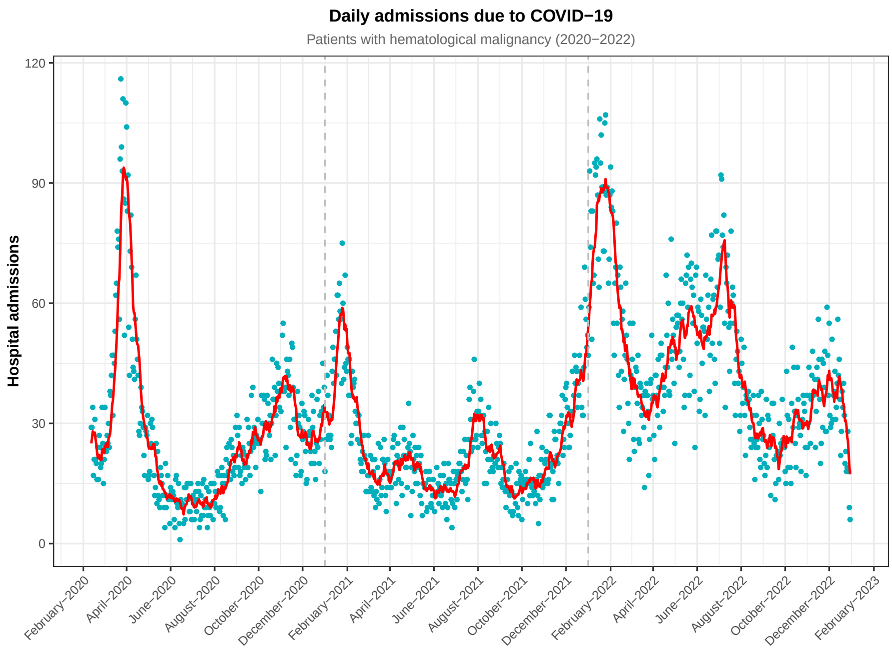

Figure A1. Evolution and trend of the COVID-19 pandemic regarding hospitalized patients with hematological malignancies. Figure shows daily admission, both in raw data (scatter points) and in 7-day moving average (red line).

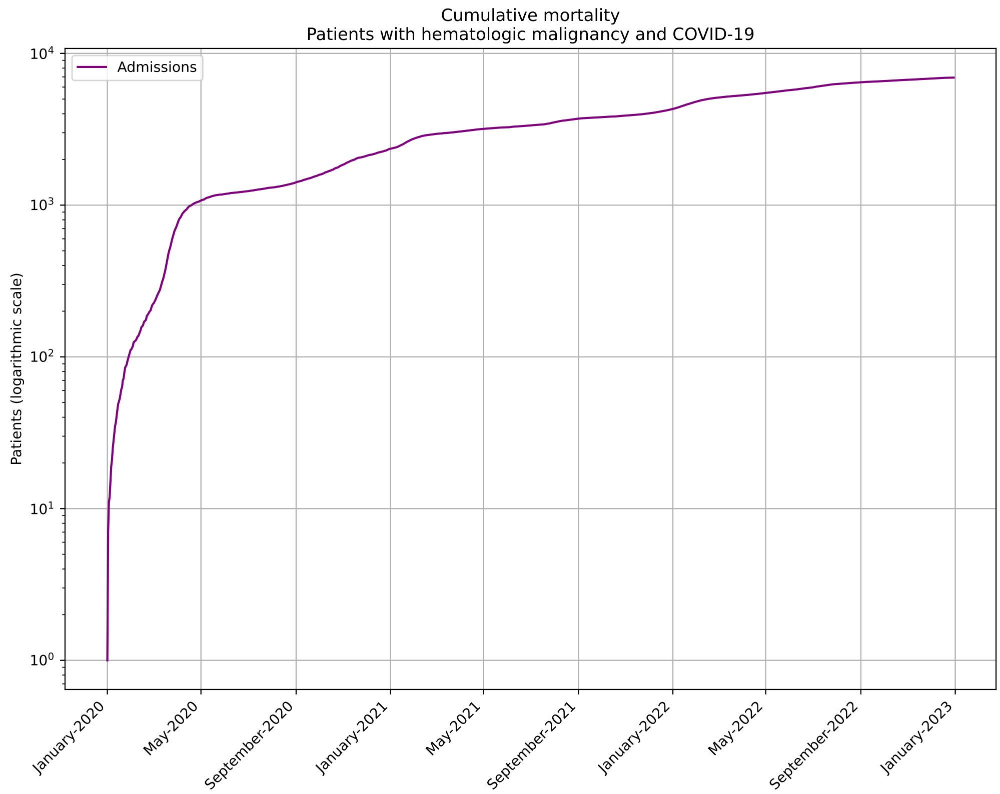

Figure A2. Cumulative sum of daily mortality in a semi-logarithmic-scale plot, which highlights the impact on the healthcare: the almost vertical initial trend showed that almost a thousand of all deaths occurred in the first four months.

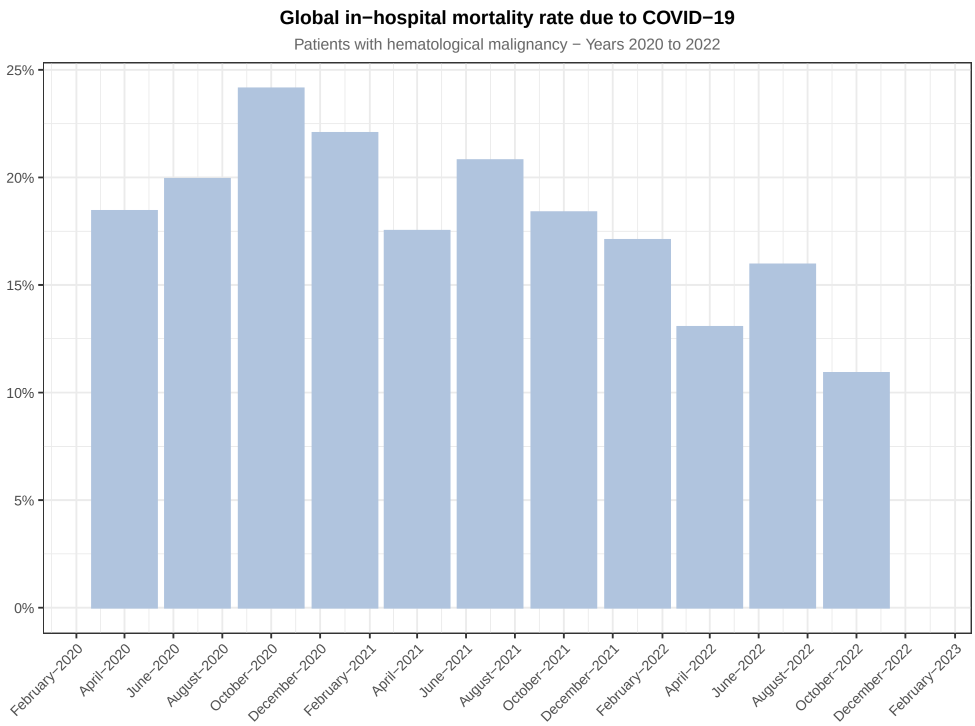

Figure A3. Global in-hospital mortality for patients with hematological malignancies over time.

Table A1. Included ICD-10 codes to categorize hematological malignancies and solid tumors.

Diagnosis	ICD-10 Code
Hodgkin lymphoma	C81 (Hodgkin lymphoma)
Follicular and B-cell lymphoma	C82 (follicular lymphoma) C83 (small cell B-cell lymphoma) C85 (unspecified types of non-Hodgkin lymphoma) C88 (other B-cell lymphomas)
T/NK-cell lymphomas	C84 (Mature T/NK-cell lymphomas) C86 (other specified types of T/NK-cell lymphoma)
Multiple myeloma and plasma cell leukemia	C90 (m. myeloma and malignant plasma cell neoplasms) C90.1 (plasma cell leukemia)
Acute/chronic leukemias	C91 (lymphoid leukemia) C92 (myeloid leukemia) C93 (monocytic leukemia) C94 (other leukemias) C95 (leukemia of unspecified cell type)
Chronic myeloproliferative disorders	D47.1 (chronic myeloproliferative disease) D47.4 (osteomyelofibrosis) D45 (polycythemia vera) D47.3 (essential thrombocythemia) C93.1 (chronic myelomonocytic leukemia)
Solid tumors	C00 – C26, C30–C34, C37–C41, C43, C45–C80

Table A2. Epidemic waves in Spain.

Periods	Dates
First period	From 1 January 2020 to 21 June 2020
Second period	From 22 June 2020 to 6 December 2020
Third period	From 7 December 2020 to 14 March 2021
Fourth period	From 15 March 2021 to 19 June 2021
Fifth period	From 20 June 2021 to 13 October 2021
Sixth period	From 14 October 2021 to 27 March 2022
Seventh period	From 28 March 2022 to 31 December 2022

Table A3. Demographic and clinical characteristics in the studied cohort of patients with hematological malignancy, regarding the epidemic wave.

	Total	First	Second	Third	Fourth	Fifth	Sixth	Seventh
Patients	34,962	5425	3856	3153	1650	2267	7297	11,314
Sex (men)	59.4%	60.4%	59.4%	60.2%	59.1%	59.8%	58.6%	59.1%
Age, years (median, IQR)	75 (19)	74 (20)	74 (20	75 (19)	73 (21)	74 (20)	74 (20)	75 (18)
Hospital stay in days (median, IQR)	9 (11)	13 (10)	10 (14)	15.1 (13)	10 (13)	15 (13)	9 (12)	8 (11)
ICU admissions	3420	437	462	378	216	318	776	833
ICU (%)	9.8	8.1	12	12	13.1	14	10.6	7.4
ICU stay in days (median, IQR)	8.5 (17.2)	8 (14.9)	8.0 (20.3)	12 (18.5)	12 (17.5)	10 (19)	10 (16.9)	6 (14.3)
Deaths	6925	1217	928	839	332	517	1428	1664
Mortality rate (%)	19.8	22.4	24.1	26.6	20.1	22.8	19.6	14.7

Table A4. Mortality rates (in percentages) per type of malignancy and quarters. Non-Hodgkin lymphoma is a combination of follicular, B-cell, and T/NK-cell lymphomas. Multiple myeloma also includes plasma cells neoplasms.

	Hodgkin lymphoma	Non-Hodgkin Lymphoma	Multiple Myeloma and Malignant Plasma Cell Neoplasms	Leukemias	Chronic Myeloproliferative Disorders
January–March 2020	19.5	22	20.7	25.3	20.8
April–June 2020	19.6	20	23.3	26.1	18.5
July–September 2020	17.1	22.8	19.5	24.9	18.4
October–December 2020	21.1	26.6	24.1	29.7	24.2
January–March 2021	18.8	26.3	31.5	25.7	21.9
April–June 2021	16	20.6	21.9	21.3	16.1
July–September 2021	11.8	29.6	16.9	25.3	17.7
October–December 2021	20	24.5	15.6	21.1	18.2
January–March 2022	14	18.4	19.7	21.7	17.9
April–June 2022	14	15.5	12.1	15	14.6
July–September 2022	14.5	16.5	14.8	19.5	16.2
October–December 2021	17.5	13	8.5	13	13.4

Table A5. Age-adjusted (standardized) mortality rates, per 100,000 population.

	2020		2021			2022		
Age Group	Deaths	Age-Adjusted Rate (95% CI)	Deaths	Age-Adjusted Rate (95% CI)		Deaths	Age-Adjusted Rate (95% CI)	
<14	7	0.10 (0.08–0.13)	7	0.10	(0.08–0.12)	10	0.15	(0.12–0.18)
15–44	48	0.28 (0.25–0.30)	42	0.24	(0.21–0.26)	47	0.27	(0.25–0.29)
45–64	303	2.15 (2.08–2.23)	284	2.05	(1.98–2.12)	354	2.47	(2.39–2.54)
65–74	458	2.44 (2.38–2.51)	431	2.33	(2.26–2.39)	620	12.98	(12.68–13.29)
>74	1394	30.11 (29.65–30.58)	1147	25.03	(24.61–25.46)	1773	37.68	(37.16–38.2)

References

1. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species Severe acute respiratory syndrome-related coronavirus: Classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.* **2020**, *5*, 536–544. [CrossRef] [PubMed]
2. Lovato, A.; de Filippis, C. Clinical Presentation of COVID-19: A Systematic Review Focusing on Upper Airway Symptoms. *Ear Nose Throat J.* **2020**, *99*, 569–576. [CrossRef] [PubMed]
3. Volz, E.; Mishra, S.; Ch, M.; Barrett, J.C.; Johnson, R.; Geidelberg, L.; Hinsley, W.R.; Laydon, D.J.; Dabrera, G.; O'Toole, Á.; et al. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature* **2021**, *593*, 266–269. [CrossRef] [PubMed]
4. Zhou, D.; Dejnirattisai, W.; Supasa, P.; Liu, C.; Mentzer, A.J.; Ginn, H.M.; Zhao, Y.; Duyvesteyn, H.M.E.; Tuekprakhon, A.; Nutalai, R.; et al. Evidence of escape of SARS-CoV-2 variant B.1.351 from natural and vaccine-induced sera. *Cell* **2021**, *184*, 2348–2361.e6. [CrossRef] [PubMed]
5. World Health Organization. COVID-19 Deaths | WHO COVID-19 Dashboard. 2024. Available online: http://data.who.int/dashboards/covid19/cases (accessed on 8 April 2024).
6. Salmanton-García, J.; Marchesi, F.; Farina, F.; Weinbergerová, B.; Itri, F.; Dávila-Valls, J.; Martín-Pérez, S.; Glenthøj, A.; Hersby, D.S.; da Silva, M.G.; et al. Decoding the historical tale: COVID-19 impact on haematological malignancy patients-EPICOVIDEHA insights from 2020 to 2022. *EClinicalMedicine* **2024**, *71*, 102553. [CrossRef]
7. Langerbeins, P.; Hallek, M. COVID-19 in patients with hematologic malignancy. *Blood* **2022**, *140*, 236–252. [CrossRef]
8. Passamonti, F.; Nicastri, E.; Di Rocco, A.; Guarini, A.; Ibatici, A.; Luminari, S.; Mikulska, M.; Visco, C. Management of patients with lymphoma and COVID-19: Narrative review and evidence-based practical recommendations. *Hematol. Oncol.* **2023**, *41*, 3–15. [CrossRef] [PubMed]
9. Chien, K.S.; Peterson, C.B.; Young, E.; Chihara, D.; Manasanch, E.E.; Ramdial, J.L.; Thompson, P.A. Outcomes of breakthrough COVID-19 infections in patients with hematologic malignancies. *Blood Adv.* **2023**, *7*, 5691–5697. [CrossRef] [PubMed]
10. He, W.; Chen, L.; Chen, L.; Yuan, G.; Fang, Y.; Chen, W.; Wu, D.; Liang, B.; Lu, X.; Ma, Y.; et al. COVID-19 in persons with haematological cancers. *Leukemia* **2020**, *34*, 1637–1645. [CrossRef] [PubMed]
11. Garcia-Carretero, R.; Vazquez-Gomez, O.; Gil-Prieto, R.; Gil-de Miguel, A. Hospitalization burden and epidemiology of the COVID-19 pandemic in Spain (2020–2021). *BMC Infect. Dis.* **2023**, *23*, 476. [CrossRef] [PubMed]

12. Pagano, L.; Salmanton-García, J.; Marchesi, F.; Busca, A.; Corradini, P.; Hoenigl, M.; Klimko, N.; Koehler, P.; Pagliuca, A.; Passamonti, F.; et al. COVID-19 infection in adult patients with hematological malignancies: A European Hematology Association Survey (EPICOVIDEHA). *J. Hematol. Oncol.* **2021**, *14*, 168. [CrossRef] [PubMed]
13. Suárez-García, I.; Perales-Fraile, I.; González-García, A.; Muñoz-Blanco, A.; Manzano, L.; Fabregate, M.; Díez-Manglano, J.; Aizpuru, E.F.; Fernández, F.A.; García, A.G.; et al. In-hospital mortality among immunosuppressed patients with COVID-19: Analysis from a national cohort in Spain. *PLoS ONE* **2021**, *16*, e0255524. [CrossRef] [PubMed]
14. Garcia-Carretero, R.; Vazquez-Gomez, O.; Ordoñez-Garcia, M.; Garrido-Peño, N.; Gil-Prieto, R.; Gil-de Miguel, A. Differences in Trends in Admissions and Outcomes among Patients from a Secondary Hospital in Madrid during the COVID-19 Pandemic: A Hospital-Based Epidemiological Analysis (2020–2022). *Viruses* **2023**, *15*, 1616. [CrossRef] [PubMed]
15. Instituto de Salud Carlos III. Centro Nacional de Epidemiología. Informe no 128 Situación de COVID-19 en España a 10 de Mayo de 2022.pdf. 2022. Available online: https://repisalud.isciii.es/handle/20.500.12105/19497 (accessed on 9 June 2022).
16. Sociedad Española de Oncología Médica. Las Cifras del Cáncer en España. 2022. Available online: https://seom.org/prensa/el-cancer-en-cifras (accessed on 29 March 2022).
17. Vijenthira, A.; Gong, I.Y.; Fox, T.A.; Booth, S.; Cook, G.; Fattizzo, B.; Martín-Moro, F.; Razanamahery, J.; Riches, J.C.; Zwicker, J.; et al. Outcomes of patients with hematologic malignancies and COVID-19: A systematic review and meta-analysis of 3377 patients. *Blood* **2020**, *136*, 2881–2892. [CrossRef] [PubMed]
18. Yigenoglu, T.N.; Bascı, S.; Dal, M.S.; Korkmaz, S.; Turgut, B.; Altuntas, F. The outcome of COVID-19 in patients with hematological malignancy. *J. Med Virol.* **2021**, *93*, 1099–1104. [CrossRef] [PubMed]
19. Martínez-López, J.; De la Cruz, J.; Gil-Manso, R.; Alegre, A.; Ortiz, J.; Llamas, P.; Martínez, Y.; Hernández-Rivas, J.Á.; González-Gascón, I.; Benavente, C.; et al. COVID-19 severity and survival over time in patients with hematologic malignancies: A population-based registry study. *Cancers* **2023**, *15*, 1497. [CrossRef] [PubMed]
20. Regalado-Artamendi, I.; Jimenez-Ubieto, A.; Hernandez-Rivas, J.A.; Navarro, B.; Nuñez, L.; Alaez, C.; Córdoba, R.; Peñalver, F.J.; Cannata, J.; Estival, P.; et al. Risk factors and mortality of COVID-19 in patients with lymphoma: A multicenter study. *Hemasphere* **2021**, *5*, e538. [CrossRef] [PubMed]
21. Biernat, M.M.; Zińczuk, A.; Biernat, P.; Bogucka-Fedorczuk, A.; Kwiatkowski, J.; Kalicińska, E.; Marciniak, D.; Simon, K.; Wróbel, T. Nosocomial outbreak of SARS-CoV-2 infection in a haematological unit - High mortality rate in infected patients with haematologic malignancies. *J. Clin. Virol.* **2020**, *130*, 104574. [CrossRef] [PubMed]
22. Ramos-Rincon, J.M.; Lopez-Sampalo, A.; Cobos-Palacios, L.; Ricci, M.; Rubio-Rivas, M.; Díaz-Simón, R.; Martín-Escalante, M.D.; Castañeda-Pérez, S.; Fernández-Madera-Martínez, R.; Beato-Perez, J.L.; et al. Nosocomial COVID-19: A nationwide Spanish study. *Gerontology* **2023**, *69*, 671–683. [CrossRef] [PubMed]

Discussion

Discussion 10

In the research focused on analyzing comorbidities in patients with COVID-19 in a pilot study [42], we explored the association between insulin resistance, metabolic syndrome, and hospitalization risk in COVID-19 patients using both traditional statistical methods and advanced machine learning techniques. Our findings demonstrate that metabolic syndrome and its components, such as BMI, insulin resistance (determined as HOMA-IR), and systolic blood pressure, were significant predictors of severe COVID-19, and so requiring hospitalization. These results align with previous studies that highlight the role of metabolic syndrome in worsening COVID-19 outcomes, providing additional evidence that these comorbidities exacerbate disease severity [43, 44].

COVID-19 and metabolic syndrome.

Our data reveal that men with metabolic syndrome have a 58% increased risk of hospitalization, while individuals with obesity face a 23% higher risk. These findings are consistent with prior research, which suggests that metabolic syndrome's combination of obesity, insulin resistance, dyslipidemia, and hypertension creates a pro-inflammatory state that exacerbates COVID-19 outcomes [45, 46]. Specifically, the activation of cytokine storms and the dysregulated immune response linked to metabolic syndrome can contribute to the progression of acute respiratory distress syndrome (ARDS) in COVID-19 patients [47].

Machine learning techniques, including recursive feature elimination (RFE) and LASSO, played a key role in identifying the most relevant predictors of hospitalization among our cohort. By refining the model to focus on some critical variables, i.e., sex, age, BMI, metabolic syndrome, systolic blood pressure, and HOMA-IR, we could develop a more parsimonious and interpretable model. These results higlight the importance of incorporating machine learning methods into clinical research, as they can uncover hidden relationships between predictors and clinical outcomes that traditional methods may overlook [48].

We consider that the clinical implications of our findings are significant. Given the high prevalence of metabolic syndrome in the population, clinicians should be aware of the increased risk in these patients when infected with SARS-CoV-2. Early identification of individuals with metabolic syndrome, combined with aggressive management of their comorbidities, could mitigate the risk of severe COVID-19 outcomes. Specifically, we propose that managing insulin resistance and controlling blood pressure may reduce hospitalization rates in this high-risk group.

Metabolic syndrome and its clinical implications in patients with COVID-19.

Furthermore, the ability to use machine learning models to predict hospitalization risk at an individual level, as demonstrated by our use of the LIME technique, suggests that these tools could be integrated into clinical practice to support decision-making. Personalized risk assessments based on metabolic factors could allow for more targeted interventions, potentially improving patient outcomes.

Limitations and future research.

Our study offers valuable insights, but we are aware of limitations which should be acknowledged. First, this is a retrospective cross-sectional study, and the data were collected from a single secondary care hospital during a specific period of the pandemic. This may limit the generalizability of our findings to broader populations. Additionally, our study did not account for other significant COVID-19-related laboratory parameters, such as inflammatory markers, which could provide further context regarding the disease's progression. Finally, while machine learning techniques allowed for a more refined model, there is still a need for external validation of our predictive model in larger and more diverse cohorts. Therefore, we consider that future lines should focus on expanding this pilot study to include a larger, more heterogeneous population. Further exploration of the additive and synergistic effects of metabolic syndrome components could provide a more comprehensive understanding of their role in COVID-19 severity. Additionally, integrating other relevant biomarkers, such as C-reactive protein and interleukin-6, into future models may offer deeper insights into the pathophysiological mechanisms that contribute to poor outcomes in these patients. Nonetheless, our findings confirm the critical role of metabolic syndrome and its components in predicting severe outcomes in COVID-19 patients. Machine learning models offer a promising approach for enhancing clinical decision-making and identifying high-risk individuals. Continued research in this area is essential to further refine predictive models and improve patient care.

COVID-19 in a secondary hospital.

Once we finished the pilot study, we expanded the period of observation, so we could provide a detailed and retrospective analysis of COVID-19's impact on a single Spanish hospital over the course of 27 months. The main objective was to assess the impact on the healthcare system, particularly during the first wave of the pandemic, when admissions and mortality rates peaked. The findings highlight how public health measures, hospital preparedness, and vaccination strategies can alter the course of the pandemic and mitigate its impacts on both healthcare infrastructure and patient outcomes.

The demographic characteristics of the patients admitted with COVID-19 revealed significant patterns. Over the course of the pandemic, the median age of hospitalized patients fluctuated, with older populations more severely impacted during the early waves. Our data showed that patients admitted during the first wave had a median age of 70 years, whereas by the fifth wave, the median age had dropped to 47 years. This shift likely reflects the early prioritization of vaccination among older and high-risk populations, resulting in fewer hospitalizations and deaths among these groups during later waves.

Epidemiological patterns and healthcare burden.

The first wave of the pandemic turned out to be an overwhelming burden on the hospital, with nearly one-third of all admissions and deaths occurring during this period. The mortality rate of 16.6% during the first wave contrasts with the lower mortality rates observed in subsequent waves, likely due to the combination of improved clinical management, widespread testing, and the eventual introduction of vaccines. These findings align with national trends in Spain, where healthcare systems experienced significant strain during the initial phase of the pandemic [49].

Our study's time series analysis revealed distinct waves of COVID-19 admissions and ICU occupancy. Notably, ICU admissions followed a similar trend to general hospital admissions, with the highest peaks in the early months of the pandemic. The need to expand ICU capacity by repurposing other hospital spaces, such as surgical rooms, demonstrates the immense pressure placed on hospital resources. The median ICU stay of 19 days, significantly longer than the median general hospital stay of 7 days, further emphasizes the intensive care requirements of critically ill COVID-19 patients.

The role of vaccination in the daily burden of a secondary hospital.

The implementation of vaccination programs in early 2021 played a crucial role in altering the course of the pandemic. Our results show a marked decrease in both hospital admissions and mortality rates from the fourth wave onwards. The significant reduction in mortality during the fourth and fifth waves coincides with increased vaccination coverage among the elderly and high-risk populations, further emphasizing the effectiveness of vaccines in preventing severe outcomes.

Despite these advances, a slight increase in both hospitalizations and mortality was observed during the sixth wave, which likely reflects the emergence of more transmissible variants and waning immunity among certain population groups. This finding aligns with global trends, where the rise of new variants such as Omicron led to renewed waves of infection, even in highly vaccinated populations.

Demographic differences in outcomes: sex and age

Our study also highlights significant sex-based and age-related differences in COVID-19 outcomes. Men accounted for a higher proportion of admissions (55%) and experienced higher mortality rates (10.8%) compared to women (7.6%). This trend was consistent across all waves, particularly during the first wave, when the mortality rate among men reached 19.5%. These findings are consistent with existing literature, which suggests that biological and behavioral factors may contribute to sex differences in COVID-19 susceptibility and outcomes. Older patients remained disproportionately affected by severe outcomes throughout the pandemic, with mortality rates highest among those aged 70 and above. However, the fifth wave saw a shift in the demographic profile of hospitalized patients, with younger individuals becoming more frequently hospitalized. This likely reflects the increased exposure among younger, less-vaccinated populations during this period.

Limitations of a secondary hospital

We were aware that the main limitation, as the previous study, was carried out in a single-center study, so our findings may not be fully generalizable to other regions or hospitals. The specific strategies implemented by this hospital, including ICU capacity expansion and the local testing policies, may have influenced the outcomes observed. Additionally, the study did not include detailed information on individual patient comorbidities or vaccination status, which could provide further insights into the factors influencing hospitalization and mortality.

Furthermore, while we used time series analysis and data visualization to uncover trends in hospital admissions and mortality, our analysis did not account for potential seasonal effects or external factors such as public health interventions that may have influenced these trends. Future research could benefit from incorporating more detailed patient-level data and exploring the long-term effects of COVID-19, including post-hospitalization outcomes.

Epidemiology of COVID-19 in Spain.

We analized Spanish data on COVID-19 to try to describe the epidemiology and burden of COVID-19 hospitalizations across our country during the first two years of the pandemic (2020–2021). Our results demonstrate that Spain experienced six distinct waves of COVID-19 hospitalizations, with significant geographic and temporal variability. The overall mortality rate among hospitalized patients was 14.3%, and while mortality decreased from 16.1% in 2020 to 12.5% in 2021, the burden on the healthcare system remained substantial, particularly during the early waves of the pandemic. The data confirm that age is the most significant predictor of hospitalization and mortality, with 77.4% of hospitalizations occurring in individuals over the age of 50. This finding is consistent with global studies indicating that advanced age is a major risk factor for severe outcomes in COVID-19 [49, 50]. Mortality was highest in patients aged 70–90 years, which aligns with other studies that highlight the vulnerability of elderly populations to COVID-19 [50].

Regional variability of COVID-19 in Spain.

The regional disparities observed in hospitalization and mortality rates suggest that socioeconomic factors and healthcare capacity played an important role in shaping the impact of the pandemic in Spain. Regions such as Madrid and Cataluña experienced the highest rates of hospitalizations, while regions like Ceuta and Castilla-La Mancha recorded the highest mortality rates per hospitalized patient. These findings align with previous research indicating that regions with higher population density and socioeconomic challenges faced more significant healthcare burdens during the pandemic [51].

Public health interventions, vaccines, treatment.

The evolution of the pandemic in Spain reflects the impact of public health measures, vaccination efforts, and changes in SARS-CoV-2 variants. The first wave in early 2020 was driven by the initial outbreak, and strict lockdown measures helped mitigate the spread of the virus. However, the easing of restrictions during the summer of 2020 led to the second wave, followed by subsequent peaks around holidays and social events [52, 53]. These trends suggest a clear link between social behavior and COVID-19 waves, emphasizing the importance of sustained public health interventions.

The introduction of COVID-19 vaccines in late 2020 had a profound effect on reducing severe outcomes, as evidenced by the decline in hospitalizations and mortality in 2021 [50]. Vaccination coverage likely contributed to the milder clinical presentations and fewer ICU admissions observed during the later waves, demonstrating the critical role of vaccination in managing the pandemic.

Our findings underscore the need for targeted public health interventions to mitigate the impact of COVID-19, particularly in high-risk populations such as the elderly and those with preexisting conditions. Early vaccination campaigns and the prioritization of vulnerable groups were crucial in reducing hospital admissions and mortality during the later waves. Future pandemic preparedness plans should emphasize the importance of rapid vaccine distribution and public health measures to prevent the healthcare system from becoming overwhelmed.

Additionally, the regional variability in outcomes suggests that healthcare resources and access to care should be tailored to the needs of each region. Ensuring equitable access to healthcare and addressing underlying socioeconomic disparities will be essential to improving outcomes in future health crises.

The main limitation of this study was the lack of data for 2022, which prevented us from analyzing the full impact of vaccination and newer SARS-CoV-2 variants on hospitalization trends. Future studies incorporating more granular data on individual patient characteristics and vaccination status will be important to fully understand the drivers of severe COVID-19 outcomes.

Assessment of vaccines against COVID-19 using machine learning.

We analized the differences between epidemic waves and so assessed the impact of COVID-19 vaccination on hospitalizations and mortality in Spain during the pandemic. Our results show that the vaccination campaign significantly reduced both hospitalizations and deaths across all age groups. Using machine learning models, we estimated that vaccines prevented between 115,172 and 170,959 hospitalizations and between 24,546 and 25,078 deaths in Spain between March and December 2021. These findings provide robust evidence for the effectiveness of vaccines as a critical public health intervention against COVID-19.

The sharp decline in hospitalizations and deaths observed after the rollout of vaccines highlights their impact on mitigating the severity of COVID-19. This effect was particularly noticeable in older populations, which were prioritized in the vaccination campaign. By April 2021, over 75% of individuals aged 80 and above were fully vaccinated, which corresponds with the substantial drop in hospitalizations and deaths in this age group during the fourth and fifth waves of the pandemic [54]. Previous studies, such as those by Barandalla et al. (2021) and Sentís et al. (2022), also reported significant decreases in severe outcomes among vaccinated individuals, reinforcing the effectiveness of these vaccines in reducing the burden on healthcare systems [50, 54].

The use of machine learning and predictive models.

The use of machine learning models, specifically ElasticNet (EN) and RandomForest (RF), allowed for accurate estimation of the potential outcomes in a scenario where vaccination had not been implemented. Both models produced comparable results, with RF estimating 115,172 averted hospitalizations and 25,078 averted deaths, and EN estimating 170,959 averted hospitalizations and 24,546 averted deaths. These findings underscore the importance of advanced statistical and machine learning techniques in public health research, particularly in simulating and forecasting the effects of large-scale interventions [55, 56].

Our analysis also demonstrates that machine learning models are well-suited for handling the complexity of pandemic data, where time-dependent factors, changing public health measures, and vaccine rollout need to be integrated into predictive models. Previous studies, such as those by Breiman (2001) and Makridakis et al. (2023), have highlighted the robustness of RandomForest in dealing with high-dimensional datasets, making it a valuable tool for forecasting future public health scenarios [57, 58].

Our results show that the demographic and clinical characteristics of hospitalized patients changed over the course of the pandemic. During the first three waves, older adults were disproportionately affected, with a median age of 66 years in our cohort. However, in the fourth and fifth waves, the median age of hospitalized patients decreased to 57 years, likely reflecting the protective effects of vaccination in older populations and the increased transmission of the virus among younger, unvaccinated individuals. This shift in the age distribution is consistent with findings from other studies that report a decline in severe outcomes among vaccinated older adults [50].

Age and comorbidities as markers of the change of the course of the pandemic.

Comorbidities such as type 2 diabetes, hypertension, and kidney disease were strongly associated with increased risk of hospitalization and death, as seen throughout the pandemic. However, our study shows that the prevalence of these comorbidities among hospitalized patients decreased in the fourth and fifth waves, further indicating the protective effects of vaccines in high-risk populations [59].

However, we were unable to account for other important factors such as variants of concern, which may have influenced hospitalization and mortality rates. Although machine learning models provided robust estimates, the projections were based on assumptions and may not fully capture the complexity of the evolving pandemic. These limitations deserve future research that should aim to include more granular data on individual vaccination status, comorbidities, and the effects of different COVID-19 variants. Expanding the use of machine learning in epidemiological research could also offer more accurate forecasting models, allowing for better preparedness and response to future public health crises. Furthermore, studies that assess the long-term impact of vaccination on both morbidity and mortality, especially as new variants emerge, will be crucial to fully understanding the benefits of vaccination.

COVID-19 and HIV.

With the nationwide study including hospitalized patients with both SARS-CoV-2 and HIV, we aimed to evaluate the impact of HIV infection on the risk of hospitalization and mortality in patients with COVID-19. Our findings demonstrate that people living with HIV (PLWH) were at higher risk of hospitalization and mortality due to COVID-19 compared to non-PLWH, with an estimated 25% increased likelihood of death. These results demonstrate the importance of HIV as a significant risk factor when managing COVID-19 in hospitalized patients.

Our descriptive analyses showed that PLWH were younger than non-PLWH at the time of hospitalization, with a median age of 54 years, compared to 73 years in the general population. Moreover, PLWH had a lower prevalence of traditional cardiovascular comorbidities such as diabetes, hypertension, and coronary disease but a higher prevalence of chronic liver disease, malignancy, and chronic pulmonary disease. These findings suggest that the more common comorbidities in PLWH were not cardiovascular risk factors, but malignancy and liver disease. These risk factor may contribute significantly to their poor outcomes when coinfected with SARS-CoV-2.

Our analysis further revealed that malignancy and acute heart failure were the most significant comorbidities contributing to mortality among PLWH. These results align with existing literature, which identifies both conditions as critical predictors of adverse outcomes in COVID-19 patients. The higher prevalence of malignancy in PLWH may be explained by the increased incidence of non-AIDS-defining cancers in this population, which has been linked to long-term antiretroviral therapy and chronic inflammation. In addition, PLWH are known to have a higher risk of developing cardiovascular conditions, such as heart failure, which may be exacerbated by COVID-19 infection [60, 61].

It is also worth noting that while the overall mortality rate for PLWH was 7.7%, this varied significantly depending on comorbid conditions and age. As in previous studies in the current dissertation, the use of machine learning techniques, such as LASSO regression, allowed us to refine our model and identify key predictors of mortality. Once more, by selecting the most relevant variables, i.e., age, malignancy, and heart failure, we developed a more parsimonious and interpretable model that accurately predicts mortality in this population.

Clinical importance of identifying HIV as a risk factor.

Given the increased risk of hospitalization and mortality in PLWH with COVID-19, healthcare systems should prioritize this population for preventive measures, including vaccination and early therapeutic interventions. Furthermore, recognizing malignancy and heart failure as key risk factors allows clinicians to identify high-risk individuals and tailor their treatment accordingly. The findings also highlight the importance of addressing HIV-related comorbidities, particularly chronic liver disease and malignancy, in the management of COVID-19. As PLWH continue to age and experience long-term effects of antiretroviral therapy, these comorbidities may become more prevalent and complicate their clinical outcomes.

Future research in HIV.

Despite the strengths of our study, several limitations should be acknowledged. First, the absence of immunovirological data, such as CD4 counts and viral loads, limits our ability to assess the impact of HIV disease progression on COVID-19 outcomes. Previous studies have suggested that well-controlled HIV infection may mitigate the severity of COVID-19, but our study was unable to explore this hypothesis due to the lack of such data [CITA]. Future research should incorporate immunovirological markers to better stratify the risk for PLWH. Additionally, while our study was based on a large, nationwide dataset, the reliance on administrative data may have led to the underreporting of certain comorbidities or misclassification of outcomes. Finally, although we adjusted for a broad range of confounders, residual confounding cannot be entirely ruled out. Future research should aim to clarify the role of immunovirological status in determining COVID-19 outcomes for PLWH. Studies incorporating CD4 counts, viral loads, and antiretroviral therapy status could provide valuable insights into the mechanisms by which HIV influences COVID-19 severity. Moreover, expanding the scope of research to include post-acute sequelae of COVID-19 (i.e., long COVID) in PLWH would offer a more comprehensive understanding of the long-term effects of the disease in this vulnerable population.

Our study on COVID-19 and HIV confirms that PLWH are at increased risk for both hospitalization and mortality due to COVID-19. Malignancy and acute heart failure are the most critical predictors of mortality in this population, underscoring the need for targeted interventions and tailored clinical management strategies for PLWH during the ongoing pandemic.

COVID-19 and hematological malignancies.

When analyzing hematological malignancies in a COVID-19 pandemic, we provided comprehensive insights into the outcomes and evolution of this vulnerable population when inifected by SARS-CoV-2 between 2020 and 2022. Our results demonstrated that these patients were at a significantly higher risk for both hospitalization and mortality when infected with SARS-CoV-2 compared to the general population. The overall mortality rate in patients with HM was 19.8%, considerably higher than the 12.7% observed in patients without hematological malignancies. This highlights the vulnerability of this population during the pandemic.

Our findings align with previous studies showing that patients with hematological malignancies are more susceptible to severe outcomes when infected with SARS-CoV-2 [62, 63]. This increased risk can be attributed to the profound immune dysfunction inherent to hematological malignancies and the immunosuppressive therapies often used to treat them. Additionally, the mortality rates in our cohort were notably higher in subgroups such as patients with non-Hodgkin lymphomas, leukemias, and plasma cell disorders, supporting earlier evidence that specific subtypes of HM confer varying degrees of risk [64].

Being aware of COVID-19 severity in patients with hematological malignancies.

The elevated risk of ICU admission and prolonged hospital stays among patients with HM further underscores the need for early intervention and close monitoring of these patients. While ICU admissions were lower than those reported in earlier meta-analyses, likely due to improvements in disease management and healthcare capacity throughout the pandemic, the 9.8% ICU admission rate in our cohort still reflects the severe clinical burden COVID-19 imposes on patients with HM [64, 65]. Additionally, the longer hospital stays observed in patients with HM (median of 13.8 days) compared to those without HM (10.8 days) suggest a more complicated clinical course, potentially due to the higher prevalence of comorbidities such as diabetes, heart failure, and chronic kidney disease.

Mortality analysis in patients with hematological malignancies.

The multivariate analysis identified specific subgroups of HM patients who are at the highest risk of mortality. Non-Hodgkin lymphomas, leukemias, and plasma cell disorders were found to be strong predictors of mortality, with odds ratios of 1.7, 1.6, and 1.24, respectively. This aligns with findings from other studies, such as the EPICOVIDEHA registry, which also highlighted non-Hodgkin lymphomas and leukemias as particularly high-risk groups [62, 63]. The identification of these high-risk subtypes emphasizes the need for tailored interventions and treatment strategies aimed at mitigating the impact of COVID-19 in these patients.

Trends and the role of vaccination.

We observed a significant decline in mortality rates over the study period, particularly in 2022, coinciding with the widespread availability of COVID-19 vaccines. This findings are in line with other global observations of reduced COVID-19 severity following vaccination, even in immunocompromised populations [59]. However, despite the decline in overall mortality, patients with HM continued to experience high mortality rates compared to the general population, likely due to the suboptimal immune response to vaccines observed in these patients [62, 63]. Future studies should focus on assessing the long-term efficacy of vaccination in patients with HM and exploring the potential role of booster doses or alternative vaccine strategies.

The need of further research in hematological malignancies.

There are several limitations to this study that should be acknowledged. First, the use of an administrative database limited our access to detailed clinical information such as the specific treatments patients received or their immunological status, including CD4 counts and viral loads. This restricts our ability to assess the direct impact of immunosuppression or specific therapeutic regimens on COVID-19 outcomes. Additionally, the absence of vaccination data prevented us from fully evaluating the protective effect of vaccines in this population, though the temporal decline in mortality rates suggests a beneficial impact. Another limitation is the lack of data on ambulatory patients, which may have led to an underestimation of the true burden of COVID-19 in this population. Future research should aim to include more detailed clinical data, particularly on the immunological status and treatment regimens of patients with hematological malignancies. Understanding the impact of different therapies and the role of vaccination in reducing the severity of COVID-19 in these patients will be critical for guiding future preventive and therapeutic strategies. Additionally, expanding studies to include ambulatory patients and those in outpatient settings would provide a more complete picture of the impact of COVID-19 on patients with hematological malignancies.

When analyzing hematological malignancies, we highlighted the significant burden of COVID-19 on these patients. Despite advances in treatment and the availability of vaccines, this population remains at high risk for severe outcomes, including increased mortality and prolonged hospital stays. Tailored preventive and therapeutic interventions, including early vaccination and close monitoring, are essential to improving outcomes for these vulnerable patients as the COVID-19 pandemic transitions into its endemic phase.

11 Key Conclusions of the Dissertation

The topic of this thesis was focused on the profound impact of COVID-19 on healthcare systems, patient outcomes, and the role of public health interventions, particularly vaccination, in mitigating the pandemic's burden. Through a combination of epidemiological analyses, machine learning techniques, and retrospective hospital data, this dissertation offers valuable insights into how different aspects of the pandemic unfolded in Spain.

The first major conclusion from the collected studies is that the early phases of the pandemic were marked by extreme strain on hospital systems, especially during the initial waves. A substantial proportion of admissions and deaths occurred during the first wave, with older populations being disproportionately affected. This finding highlights the importance of early pandemic preparedness and the need for robust health system responses to manage such crises effectively. The pressure on intensive care units was particularly pronounced, underscoring the need of flexible hospital infrastructure capable of adapting to rapid surges in critically ill patients.

A critical finding across the chapters is the impact of vaccination on reducing the severity of COVID-19 outcomes. The studies demonstrate that vaccination not only reduced hospital admissions and mortality but also shifted the demographic profile of those hospitalized. By the fourth and fifth waves, the median age of hospitalized patients had dropped significantly, reflecting the protection conferred to older, more vulnerable populations through vaccination campaigns. The dissertation highlights vaccination as the most effective intervention to date in facing the pandemic, with estimates suggesting that thousands of hospitalizations and deaths were averted as a direct result of widespread immunization.

Machine learning techniques applied in several manuscripts allowed for precise modeling of the pandemic's effects, particularly in predicting outcomes based on various demographic and clinical factors. These models further reinforced the critical role of age, sex, and comorbidities in determining patient outcomes. For instance, men were consistently shown to have higher mortality rates than women, and patients with pre-existing conditions like hypertension, diabetes, and heart disease were more likely to experience severe COVID-19 outcomes. These findings emphasize the need for personalized risk assessments and targeted interventions for high-risk populations during pandemics.

The geographic variability in the pandemic's impact, explored in the manuscripts, also suggests that regional differences in healthcare capacity, public health measures, and socioeconomic factors played a crucial role in shaping outcomes. Hospitals in highly populated or socioeconomically disadvantaged regions faced greater challenges, with higher rates of admissions and mortality. These disparities underscore the importance of equitable resource allocation and healthcare access to ensure that all populations are adequately protected during public health emergencies.

Finally, the dissertation underscores the evolving nature of the COVID-19 pandemic, with emerging variants and shifting public health policies influencing the course of the disease. The slight resurgence in hospitalizations and mortality during later waves, despite vaccination, highlights the ongoing threat posed by new viral variants. This finding suggests that continuous surveillance, booster vaccination strategies, and adaptive healthcare responses are essential in managing the long-term effects of COVID-19.

Regarding vulnerable population, we focused on patients with HIV and patients with hematological malignancies. This dissertation on COVID-19 and population at risk confirms that both people living with HIV and patients with previous hematological malignancy were at increased risk for both hospitalization and mortality due to COVID-19. Despite advances in treatment and the availability of vaccines, this population remains at high risk for severe outcomes, including increased mortality and prolonged hospital stays. Adapted preventive and therapeutic interventions, including early vaccination and close monitoring, are essential to improving outcomes for these vulnerable patients as the COVID-19 pandemic transitions into its endemic phase.

In conclusion, this body of work provides a comprehensive analysis of the COVID-19 pandemic's impact on healthcare systems, specifically in Spain. It underscores the importance of early preparedness, the critical role of vaccination, and the need for adaptive strategies to respond to evolving challenges. The insights gained from these studies can guide future public health efforts, ensuring that healthcare systems are better equipped to handle pandemics and other large-scale health crises.

Appendix

A Contributions to Conferences and Workshops

Table A.1: Posters, communications and proceedings publications related to the current dissertation.

Workshop	Title	Date
V Congreso De La Escuela Internacional De Doctorado Universidad Rey Juan Carlos	• Evolution of COVID-19 in the absence of vaccination: Practical use of artificial intelligence	May 2024
44º Congreso de la Sociedad Española de Medicina Interna (SEMI) – Valencia	• Importancia De Predictores En El Riesgo De Hospitalización Por Covid-19: Uso Práctico De Machine Learning • Impacto De Las Neoplasias Hematológicas En La Gravedad De Las Hospitalizaciones Por Covid-19 A Nivel Nacional (2020-2021) • Impacto De La Vacunación Contra Sars-Cov-2 En Las Hospitalizaciones: Estimación Mediante Modelos De Machine Learning	November 2023
XXVII Congreso Nacional de la Sociedad Española de Enfermedades Infecciosas y Microbiología Clínica	• Machine learning como herramienta para predecir la evolución de la pandemia COVID-19: simulación de la efectividad vacunal • Características demográficas y clínicas de los pacientes con VIH que precisan hospitalización por COVID-19: estudio nacional (2020-2021)	June 2024
LXVI Congreso Nacional de la Sociedad Española de Hematología y Hemoterapia (SEHH) y XL Congreso Nacional de la Sociedad Española de Trombosis y Hemostasia (SETH)	• Impacto del COVID-19 en Pacientes Hospitalizados con Linfoma: Análisis de Factores de Riesgo y Mortalidad en un Estudio Poblacional en España • Mortalidad de los Pacientes Hematológicos con COVID-19 en España: Un Estudio Retrospectivo • Evaluación del Efecto de las Vacunas Contra SARS-CoV-2 en Pacientes con Neoplasias Hematológicas Mediante Simulación y Algoritmos de Machine Learning	October 2024

Bibliography

[1] Rodriguez-Arrastia M, García-Martín M, Romero-López A, Ropero-Padilla C, Ruiz-Gonzalez C, Roman P, et al. Evolution of the Public-Health Response to COVID-19 Pandemic in Spain: A Descriptive Qualitative Study. International Journal of Environmental Research and Public Health. 2022;19(7):3824.

[2] Jiang F, Deng L, Zhang L, Cai Y, Cheung CWM, Xia Z. Review of the Clinical Characteristics of Coronavirus Disease 2019 (COVID-19). Journal of General Internal Medicine. 2020;35(5):1545–1549.

[3] Phelan AL, Katz R, Gostin LO. The COVID-19 Pandemic and the World Health Organization: Global Health in Crisis. JAMA. 2020;323(23):2337–2338.

[4] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet. 2020;395(10223):497–506.

[5] Johns Hopkins University. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE); 2023. Available at: https://coronavirus.jhu.edu/map.html.

[6] Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. The Lancet Infectious Diseases. 2020;20(5):533–534.

[7] Centro Nacional de Epidemiología. Informe sobre la situación de COVID-19 en España; 2023. Available at: https://cnecovid.isciii.es/covid19/.

[8] Instituto de Salud Carlos III. Informe no 128 Situación de COVID-19 en España a 10 de mayo de 2022.pdf; 2022. Accessed: 2022-06-09. https://www.isciii.es/QueHacemos/Servicios/VigilanciaSaludPublicaRENAVE/EnfermedadesTransmisibles/Documents/INFORMES/Informes%20COVID-19/INFORMES%20COVID-19%202022/Informe%20n%C2%BA%20128%20Situaci%C3%B3n%20de%20COVID-19%20en%20Espa%C3%B1a%20a%2010%20de%20mayo%20de%202022.pdf.

[9] Nicola M, Alsafi Z, Sohrabi C, Kerwan A, Al-Jabir A, Iosifidis C, et al. The socio-economic implications of the coronavirus pandemic (COVID-19): A review. International Journal of Surgery. 2020;78:185–193.

[10] United Nations. Policy Brief: Education during COVID-19 and beyond; 2020. Accessed: 2024-09-24.

[11] Guan Wj, Ni Zy, Hu Y, Liang Wh, Ou Cq, He Jx, et al. Clinical characteristics of coronavirus disease 2019 in China. New England Journal of Medicine. 2020;382(18):1708–1720.

[12] Wu Z, McGoogan JM. Characteristics of and Important Lessons from the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases from the Chinese Center for Disease Control and Prevention. JAMA. 2020;323(13):1239–1242.

[13] Jordan RE, Adab P, Cheng KK. COVID-19: risk factors for severe disease and death. BMJ. 2020;368:m1198.

[14] Stefan N, Birkenfeld AL, Schulze MB. Obesity and impaired metabolic health in patients with COVID-19. Nature Reviews Endocrinology. 2020;16(7):341–342.

[15] Sanchis-Gomar F, Lavie CJ, Mehra MR, Henry BM, Lippi G. Obesity and outcomes in COVID-19: when an epidemic and pandemic collide. Mayo Clinic Proceedings. 2020;95(7):1445–1453.

[16] Kupferschmidt K, Cohen J. Race to find COVID-19 treatments accelerates. Science. 2020;367(6485):1412–1413.

[17] Mullard A. Flooded by the torrent: the COVID-19 drug pipeline. The Lancet. 2020;395(10232):1245–1246.

[18] Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. The Lancet. 2020;395(10229):1054–1062.

[19] Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. JAMA. 2020;323(20):2052–2059.

[20] World Health Organization. Impact of COVID-19 on people's livelihoods, their health, and our food systems; 2020. Accessed: 2024-09-24. https://www.who.int/news/item/13-10-2020-impact-of-covid-19-on-people's-livelihoods-their-health-and-our-food-systems.

[21] Hiscott J, Alexandridi M, Muscolini M, Tassone E, Palermo E, Soultsioti M, et al. The global impact of the coronavirus pandemic. Cytokine & growth factor reviews. 2020;53:1–9.

[22] Sawicka B, Aslan I, Della Corte V, Periasamy A, Krishnamurthy SK, Mohammed A, et al. The coronavirus global pandemic and its impacts on society. 2022;p. 267–311.

[23] Naseer S, Khalid S, Parveen S, Abbass K, Song H, Achim MV. COVID-19 outbreak: Impact on global economy. Frontiers in public health. 2023;10:1009393.

[24] Mishra NP, Das SS, Yadav S, Khan W, Afzal M, Alarifi A, et al. Global impacts of pre-and post-COVID-19 pandemic: Focus on socio-economic consequences. Sensors International. 2020;1:100042.

[25] He J, Kraemer MUG, Wilson K, Moffatt P, Dunn AG, Roy A, et al. Impact of COVID-19 on research productivity across all areas of knowledge. BMJ Global Health. 2023;8(4):e011514.

[26] Dissanayake H. COVID-19 and metabolic syndrome. Best Practice & Research Clinical Endocrinology & Metabolism. 2023;37(4):101753.

[27] Makhoul E, Aklinski JL, Miller J, Leonard C, Backer S, Kahar P, et al. A review of COVID-19 in relation to metabolic syndrome: obesity, hypertension, diabetes, and dyslipidemia. Cureus. 2022;14(7).

[28] Elemam NM, Hannawi H, Al Salmi I, Naeem KB, Alokaily F, Hannawi S. Diabetes mellitus as a comorbidity in COVID-19 infection in the United Arab Emirates. Saudi medical journal. 2021;42(2):170.

[29] Mueller-Wieland D, Marx N, Dreher M, Fritzen K, Schnell O. COVID-19 and cardiovascular comorbidities. Experimental and Clinical Endocrinology & Diabetes. 2022;130(03):178–189.

[30] He ZF, Zhong NS, Guan WJ. Impact of chronic respiratory diseases on the outcomes of COVID-19. Archivos de Bronconeumología. 2022;58(1):5.

[31] Calver JF, Fabbri L, May J, Jenkins RG. COVID-19 in patients with chronic lung disease. Clinics in Chest Medicine. 2023;44(2):385–393.

[32] Antinori A, Bausch-Jurken M. The burden of COVID-19 in the immunocompromised patient: implications for vaccination and needs for the future. The Journal of Infectious Diseases. 2023;228(Supplement_1):S4–S12.

[33] Martins MP, de Oliveira RB. COVID-19 and chronic kidney disease: A narrative review. COVID. 2023;3(8):1092–1105.

[34] Brogan M, Ross MJ. COVID-19 and kidney disease. Annual review of medicine. 2023;74(1):1–13.

[35] Vogt F, Rebuli N, Cretikos M, Liu B, Macartney K, Kaldor J, et al. Assessing the effects of SARS-CoV-2 vaccination on the risk of household transmission during delta variant circulation: a population-based data linkage cohort study. The Lancet Regional Health–Western Pacific. 2024;42.

[36] Martínez-Baz I, Trobajo-Sanmartín C, Miqueleiz A, Casado I, Navascués A, Burgui C, et al. Risk reduction of hospitalisation and severe disease in vaccinated COVID-19 cases during the SARS-CoV-2 variant Omicron BA. 1-predominant period, Navarre, Spain, January to March 2022. Eurosurveillance. 2023;28(5):2200337.

[37] Marchand G, Masoud AT, Medi S. Risk of all-cause and cardiac-related mortality after vaccination against COVID-19: A meta-analysis of self-controlled case series studies. Human vaccines & immunotherapeutics. 2023;19(2):2230828.

[38] Waseel F, Streftaris G, Rudrusamy B, Dass SC. Assessing the dynamics and impact of COVID-19 vaccination on disease spread: A data-driven approach. Infectious Disease Modelling. 2024;9(2):527–556.

[39] Lam ICH, Zhang R, Man KKC, Wong CKH, Chui CSL, Lai FTT, et al. Persistence in risk and effect of COVID-19 vaccination on long-term health consequences after SARS-CoV-2 infection. Nature Communications. 2024;15(1):1716.

[40] Watson OJ, Barnsley G, Toor J, Hogan AB, Winskill P, Ghani AC. Global impact of the first year of COVID-19 vaccination: a mathematical modelling study. The Lancet infectious diseases. 2022;22(9):1293–1302.

[41] Varrelman TJ, Rader B, Remmel C, Tuli G, Han AR, Astley CM, et al. Vaccine effectiveness against emerging COVID-19 variants using digital health data. Communications Medicine. 2024;4(1):81.

[42] Garcia-Carretero R, Vazquez-Gomez O, Lopez-Lomba M, Gil-Prieto R, Gil-de Miguel A. Insulin Resistance and Metabolic Syndrome as Risk Factors for Hospitalization in Patients with COVID-19: Pilot Study on the Use of Machine Learning. Metab Syndr Relat Disord. 2023 Oct;21(8):443–452.

[43] Xie J, Zu Y, Alkhatib A, Pham TT, Gill F, Jang A, et al. Metabolic Syndrome and COVID-19 Mortality Among Adult Black Patients in New Orleans. Diabetes Care. 2020 Aug 25;44(1):188–193.

[44] Lohia P, Kapur S, Benjaram S, Pandey A, Mir T, Seyoum B. Metabolic syndrome and clinical outcomes in patients infected with COVID-19: Does age, sex, and race of the patient with metabolic syndrome matter? J Diabetes. 2021 Jan 16;13(5):420–429.

[45] Denson JL, Gillet AS, Zu Y, Brown M, Pham T, Yoshida Y, et al. Metabolic Syndrome and Acute Respiratory Distress Syndrome in Hospitalized Patients With COVID-19. JAMA Network Open. 2021 Dec 22;4(12):e2140568.

[46] Stefan N, Birkenfeld AL, Schulze MB. Global pandemics interconnected - obesity, impaired metabolic health and COVID-19. Nat Rev Endocrinol. 2021 Mar;17(3):135–149.

[47] Chen G, Wu D, Guo W, Cao Y, Huang D, Wang H, et al. Clinical and immunological features of severe and moderate coronavirus disease 2019. J Clin Invest. 2020 May 1;130(5):2620–2629.

[48] Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. The New England journal of medicine. 2019 Apr;380(14):1347–1358.

[49] Casas-Rojo J, Antón-Santos J, Millán-Núñez-Cortés J, Lumbreras-Bermejo C, Ramos-Rincón J, Roy-Vallejo E, et al. Clinical characteristics of patients hospitalized with COVID-19 in Spain: Results from the SEMI-COVID-19 Registry. Revista Clínica Española (English Edition). 2020;220(8):480–494.

[50] Barandalla I, Alvarez C, Barreiro P, de Mendoza C, González-Crespo R, Soriano V. Impact of scaling up SARS-CoV-2 vaccination on COVID-19 hospitalizations in Spain. International Journal of Infectious Diseases. 2021;112:81–88.

[51] Palomino J, Rodríguez J, Sebastian R. The COVID-19 shock on the labour market: poverty and inequality effects across Spanish regions. Regional Studies. 2022 Sep 23;0(0):1–15.

[52] Taboada M, González M, Alvarez A, Eiras M, Costa J, Álvarez J, et al. First, second and third wave of COVID-19. What have we changed in the ICU management of these patients? Journal of Infection. 2021;82(6):e14–e15.

[53] Soriano V, de Mendoza C, Gómez-Gallego F, Corral O, Barreiro P. Third wave of COVID-19 in Madrid, Spain. International Journal of Infectious Diseases. 2021 June;107:212–214.

[54] Sentís A, Kislaya I, Nicolay N, Meijerink H, Starrfelt J, Martínez-Baz I, et al. Estimation of COVID-19 vaccine effectiveness against hospitalisation in individuals aged $\geq$ 65 years using electronic health registries; a pilot study in four EU/EEA countries, October 2021 to March 2022. Eurosurveillance. 2022;27(30):2200551.

[55] Kırbaş İ, Sözen A, Tuncer AD, Kazancıoğlu FŞ. Comparative analysis and forecasting of COVID-19 cases in various European countries with ARIMA, NARNN and LSTM approaches. Chaos Solitons Fractals. 2020 Sep;138:110015.

[56] Nabi KN, Tahmid MT, Rafi A, Kader ME, Haider MA. Forecasting COVID-19 cases: A comparative analysis between recurrent and convolutional neural networks. Results Phys. 2021 May;24:104137.

[57] Breiman L. Random Forests. Machine Learning. 2001 Oct;45(1):5–32.

[58] Makridakis S, Spiliotis E, Assimakopoulos V, Semenoglou AA, Mulder G, Nikolopoulos K. Statistical, machine learning and deep learning forecasting methods: Comparisons and ways forward. Journal of the Operational Research Society. 2023 Mar 4;74(3):840–859.

[59] Garcia-Carretero R, Vazquez-Gomez O, Gil-Prieto R, Gil-de Miguel A. Hospitalization burden and epidemiology of the COVID-19 pandemic in Spain (2020-2021). BMC Infect Dis. 2023 Jul 18;23(1):476.

[60] Bhaskaran K, Rentsch CT, MacKenna B, Schultze A, Mehrkar A, Bates CJ, et al. HIV infection and COVID-19 death: a population-based cohort analysis of UK primary care data and linked national death registrations within the OpenSAFELY platform. Lancet HIV. 2021 Jan;8(1):e24–e32.

[61] Suárez-García I, Gutierrez F, Pérez-Molina JA, Moreno S, Aldamiz T, Valencia Ortega E, et al. Mortality due to non-AIDS-defining cancers among people living with HIV in Spain over 18 years of follow-up. Journal of Cancer Research and Clinical Oncology. 2023;149(20):18161–18171.

[62] Salmanton-García J, Marchesi F, Farina F, Weinbergerová B, Itri F, Dávila-Valls J, et al. Decoding the historical tale: COVID-19 impact on haematological malignancy patients-EPICOVIDEHA insights from 2020 to 2022. EClinicalMedicine. 2024 May;71:102553.

[63] Pagano L, Salmanton-García J, Marchesi F, Busca A, Corradini P, Hoenigl M, et al. COVID-19 infection in adult patients with hematological malignancies: a European Hematology Association Survey (EPICOVIDEHA). Journal of Hematology & Oncology. 2021 Oct;14(168).

[64] Vijenthira A, Gong IY, Fox TA, Booth S, Cook G, Fattizzo B, et al. Outcomes of patients with hematologic malignancies and COVID-19: a systematic review and meta-analysis of 3377 patients. Blood. 2020 Dec;136(25):2881–2892.

[65] Langerbeins P, Hallek M. COVID-19 in patients with hematologic malignancy. Blood. 2022 Jul;140(3):236–252.

Et tandem, totus audientia tacite ad te spectat.
Aut mentes eorum commovisti
aut perterritos reliquisti verbis carentes.
Ite, missa est.