



**Universidad
Europea**

Máster Universitario en Bioinformática

**“BÚSQUEDA E IDENTIFICACIÓN DE
PACIENTES CON INFECCION
ACTIVA POR HEPATITIS C Y CON
NECESIDAD DE VINCULACIÓN AL
SEGUIMIENTO Y TRATAMIENTO EN LOS
REGISTROS SANITARIOS”**

Autor: Miguel Sánchez Rodríguez

Tutor: Pablo Ryan Murua

Curso 2023-24

AGRADECIMIENTOS

Estimado Pablo Ryan:

Me siento en deuda contigo por todo lo que has hecho por mí; no solo a estos últimos meses con todo lo relacionado a la realización de este Trabajo de Fin de Máster, si no también por la labor y el especial cuidado que has tenido conmigo durante todo el curso. Este trabajo para mí ha sido muy inspirador, ya que es la primera vez que he usado mi conocimiento y he puesto en práctica mis habilidades para poder ayudar a bastante gente en un futuro inmediato; lo que me enorgullece enormemente. Muchas gracias por contar conmigo, espero que más pronto que tarde, podamos llegar a trabajar juntos como lo hemos hecho en esta ocasión.

Miguel Sánchez Rodríguez.

RESUMEN

A día de hoy, la hepatitis C continúa siendo uno de los retos más desafiantes que enfrenta la salud pública a nivel global, considerándose una de las enfermedades infecciosas más prevalentes en todo el mundo. En este trabajo, hemos llevado a cabo una búsqueda activa de pacientes con infección no resuelta de hepatitis C haciendo uso de registros electrónicos médicos provenientes del Hospital Universitario Infanta Leonor, con el objetivo de poder identificarlos y ponerlos en manos de los profesionales de la salud. Asimismo, se han analizado las características y tendencias de los pacientes a lo largo de los 14 años de seguimiento continuo, identificando en el proceso posibles factores de riesgo asociados a la enfermedad y comprobando la eficacia de las políticas médicas aplicadas. Todo el análisis se realizó en la plataforma *RStudio*, lo cual nos permitió una evaluación exhaustiva de los datos. Los resultados obtenidos destacan la necesidad inherente de reforzar las prácticas clínicas de cribado y tratamiento para la mejora de la salud pública general.

Palabras clave: Hepatitis C, HCV, Registros Electrónicos Médicos, infección activa.

ABSTRACT

Currently, hepatitis C remains one of the most challenging issues facing global public health, being considered a prevalent infectious diseases worldwide. In this work, we have conducted an active search for patients with unresolved hepatitis C infection using electronic medical records from the Hospital Universitario Infanta Leonor, with the aim of identifying them and referring them to health authorities. Additionally, we have analyzed the characteristics and trends of the patients over 14 years of continuous follow-up, identifying possible risk factors associated with the disease and assessing the effectiveness of the implemented medical policies. The entire analysis was carried out on the *RStudio* platform, which allowed us to conduct a comprehensive evaluation of the data. The results obtained highlight the inherent need to strengthen screening and treatment practices to improve overall public health.

Keywords: Hepatitis C, HCV, Electronic Health Records (EHRs), active infection.

ÍNDICE

REVISAR ÍNDICE DE NUEVO QUE LO HE CAMBIADO

1. INTRODUCCIÓN.....	pág 5
2. HIPÓTESIS DEL TRABAJO Y OBJETIVOS.....	pág 13
3. MATERIALES Y MÉTODOS.....	pág 15
4. RESULTADOS OBTENIDOS.....	pág 23
5. DISCUSIÓN.....	pág 42
6. CONCLUSIONES.....	pág 46
BIBLIOGRAFÍA.....	pág 48
ANEXOS.....	pág 51
DECLARACIÓN DE DEL USO DE HERRAMIENTAS DE IA.....	pág 51

1. INTRODUCCIÓN

REGISTROS ELECTRÓNICOS SANITARIOS – EL BIG DATA AL SERVICIO DE LA SALUD

En las últimas décadas, el avance de la tecnología ha transformado la gestión de la información en el ámbito sanitario. Los Registros Electrónicos de Salud (“*Electronic Health Records*”: EHRs, por sus siglas en inglés) han surgido como una herramienta indispensable para almacenar, gestionar y compartir datos médicos de los pacientes. Desde su aparición, los EHR han revolucionado la práctica clínica, haciendo que la información médica sea más accesible para los profesionales de la salud y mejorando como consecuencia la calidad de los cuidados médicos (Evans, 2016).

Una de las principales ventajas de los EHR es su capacidad para facilitar el intercambio de información entre diferentes entidades y niveles de atención sanitaria. Hoy en día, cada vez más organizaciones de salud han adoptado sistemas informáticos que permiten compartir estos registros de forma segura y eficiente, agilizando procesos administrativos complejos y facilitando la coordinación de cuidados entre especialistas, hospitales y centros de atención primaria. Esta capacidad de integrar información de diversas fuentes es especialmente relevante en el contexto de las enfermedades infecciosas o bio-vigilancia de la salud pública, donde la rápida transmisión de datos es crucial para la detección y control de posibles brotes. De esta forma, podemos lograr identificar cambios en la incidencia de enfermedades infecciosas, evaluar la efectividad de las intervenciones sanitarias realizadas y ajustar las políticas de salud en consecuencia como última instancia (Evans, 2016). Un ejemplo destacado y bastante desarrollado del uso de EHR en salud pública es el sistema de vigilancia de salud de Massachusetts (EEUU), *ESO*, el cual permite calcular la incidencia y prevalencia de enfermedades en tiempo real, proporcionando a los profesionales de la salud una visión detallada y actualizada de la situación sanitaria de la población (Willis et al., 2019).

Sin embargo, a pesar de los avances logrados hasta el día de hoy, el uso de EHR en investigación clínica aún enfrenta a diversos desafíos. Algunas limitaciones surgen del

hecho de que alguna información relevante como los síntomas o detalles específicos de la evolución de una enfermedad a menudo no se registran adecuadamente en los EHR o se recogen únicamente en notas clínicas físicas; lo que dificulta su acceso y análisis. Para solucionar este problema se están desarrollando nuevas estrategias basadas en el Procesamiento del Lenguaje Natural (NLP; de sus siglas en inglés “*Natural Processing Language*”); técnica que permite extraer de manera más eficiente y estandarizada esta información (Willis et al., 2019). Por otro lado, tradicionalmente el área de las enfermedades infecciosas no ha contado con suficientes recursos tecnológicos y ha sido considerada una especialidad médica muy técnica, hecho que ha dificultado la adopción de sistemas electrónicos de manera generalizada. Esta condición ha sido muy relevante para que exista un vacío en la literatura científica sobre el uso de EHR en enfermedades infecciosas. Sin embargo, esta situación está comenzando a cambiar con la incorporación de expertos en bioinformática, quienes están desempeñando un papel crucial en la modernización de la vigilancia epidemiológica (Jhaveri et al., 2021).

EL VIRUS DE LA HEPATIS C: BIOLOGÍA GENERAL Y SITUACIÓN ACTUAL.

La hepatitis C es una infección viral que afecta aproximadamente a 180 millones de personas en todo el mundo (Preciado et al., 2014). Con una prevalencia estimada del 1.5% en la población general y hasta un 56% en grupos de alto riesgo, representa un problema muy importante para la salud pública (Ridruejo et al., 2016). Como podemos observar en la **FIGURA 1**, la historia del virus de la hepatitis C está marcada por descubrimientos científicos e hitos claves:

Todo comienza en la década de 1980, momento en el que se detectó un aumento considerable del número de pacientes que presentaban problemas hepáticos. Sin embargo, estos no estaban infectados por los agentes virales de la hepatitis A ni B, los cuales eran los únicos conocidos hasta ese momento en este contexto. Por este motivo, esta entidad viral fue denominada en primera instancia como hepatitis *non-A* y *non-B*, respectivamente. No fue hasta 1989 que el investigador Michael Houghton y su equipo lograron clonar y caracterizar el virus, el cual fue conocido como el virus de la hepatitis C (HCV por sus siglas en inglés “*Hepatitis C Virus*”). Tras este descubrimiento, se identificaron al poco tiempo siete genotipos diferentes de este nuevo virus, lo que

marcó un hito en la comprensión de su diversidad genética y su transmisión. Posteriormente, se desarrollaron pruebas de detección especializadas contra el patógeno, como los test ELISA o PCRs específicas, mejorando la precisión de los estudios epidemiológicos. Años después del descubrimiento del HCV, se llevó a cabo una búsqueda intensa de vacunas y tratamientos antivirales contra el patógeno. Sin embargo, no se llegó a desarrollar una vacuna totalmente viable debido a la alta variabilidad genética que presentaba el virus, hecho que redirigió los esfuerzos hacia el diseño de tratamientos antivirales (Ward et al., 2014). A principios de los años 90, surgió el primer tratamiento basado en interferones, que, aunque en un primer momento se consideró muy prometedor, únicamente curaba al 50% de los pacientes y presentaba demasiados efectos secundarios. Sin embargo, en la última década, el desarrollo de antivirales directos de nueva generación ha permitido curar completamente la infección en un alto porcentaje de pacientes, con efectos secundarios mínimos; disminuyendo de esta forma la severidad y la prevalencia total de esta enfermedad por todo el mundo. A fecha de hoy, la hepatitis C es una de las pocas enfermedades que tienen cura (Lagging et al., 2017).

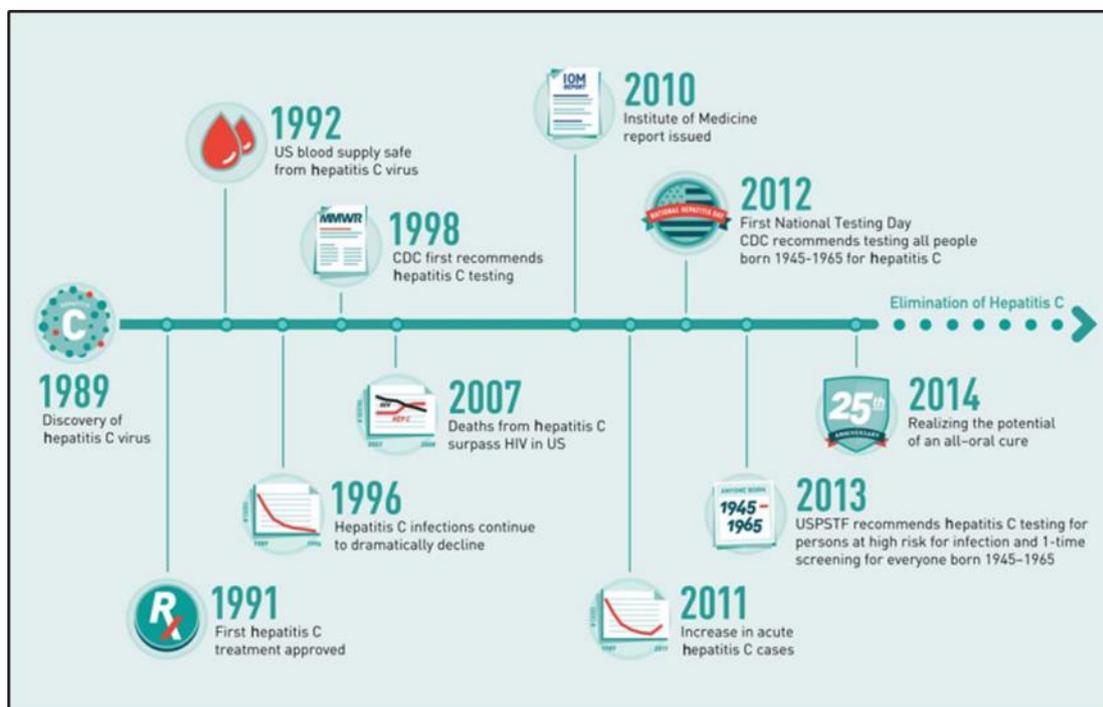


FIGURA 1. Historia e hitos claves de la Hepatitis C. Extraída de Ward et al., 2014

Desde el punto de vista biológico, el virus de la hepatitis C (o en inglés “*Hepatitis C Virus*”: HCV) es un pequeño virus de ARN de cadena simple positiva perteneciente al género *Hepacivirus* de la familia *Flaviviridae*. Posee envuelta y su genoma se comprende aproximadamente de 9,6 kb en el que se incluyen secuencias altamente conservadas y regiones hipervariables que desempeñan un papel crucial en la evasión del sistema inmune (Zeisel et al., 2013). Una de sus características principales es su elevada tasa de mutación, lo que le permite adaptarse rápidamente y evolucionar dentro del organismo del huésped. Esta notoria capacidad está retroalimentada a su vez por la “compartimentalización” que presenta el virus en diferentes tejidos del cuerpo humano, concepto representado en la **FIGURA 2**. De esta forma, mientras que el hígado es la ubicación para su supervivencia y replicación (su reservorio), el sistema digestivo y la sangre son las vías de transmisión principales. Asimismo, su capacidad para generar variantes genéticas dentro de un mismo huésped le permite evadir exitosamente el sistema inmunológico y adaptarse a los diferentes entornos corporales. La compartimentalización no solo afecta a la progresión de la enfermedad, sino que también influye negativamente en la eficacia del tratamiento antiviral. Además, la dinámica del parásito también se ve afectada por la co-infección con otros virus como el VIH o estados fisiológicos como el embarazo (Preciado et al., 2014).

Como ya hemos mencionado, el reservorio principal del HCV es el hígado, premisa que explica su fuerte asociación con enfermedades hepáticas crónicas como la cirrosis y el carcinoma hepatocelular: aproximadamente el 30% de los pacientes infectados desarrollan cirrosis durante el transcurso de su infección, y el 10% padecen un cáncer hepático fatal. Sumado a esto, la infección por HCV se asocia con un mayor riesgo de desarrollar otro tipo de enfermedades metabólicas, como diabetes mellitus y esteatosis hepática (acumulación de grasa en el hígado), complicaciones que agravan aún más el estado de salud de los pacientes (Preciado et al., 2014).

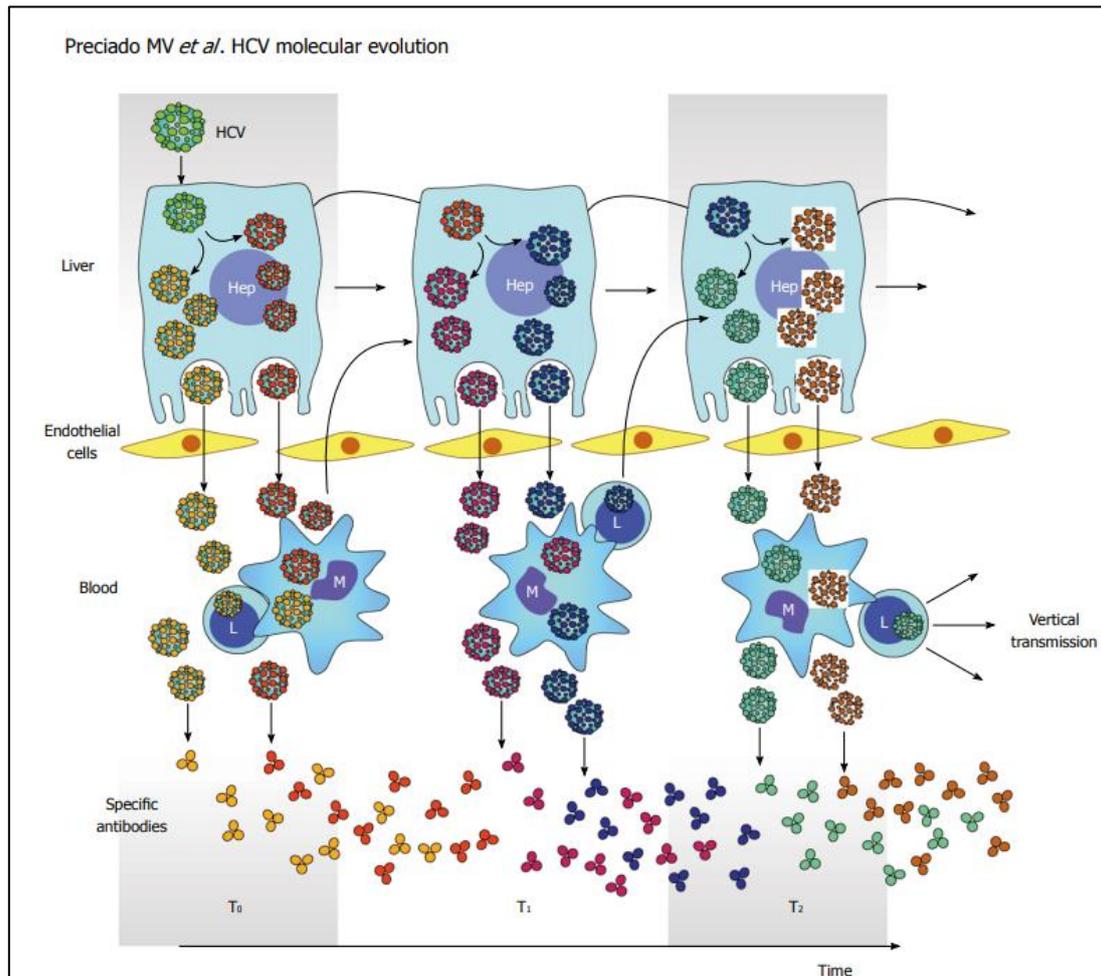


FIGURA 2. La compartimentalización del virus de la Hepatitis C en diferentes órganos del cuerpo humano. Se observan las múltiples subpoblaciones del HCV en distintos colores. En la parte superior, se encuentran las células del hígado, las cuales son perforadas tras la salida de la partícula viral. Posteriormente estas pasan a la sangre para su transmisión vertical. Se aprecian también los anticuerpos específicos. Imagen extraída de Preciado MV *et al.*

En cuanto a su transmisión, esta ocurre principalmente por vía parenteral, es decir, a través de la sangre. Entre los factores de riesgo más importantes están el uso compartido de jeringuillas (IDUs, conocido por sus siglas “*Injection Drug Users*”; Usuarios de drogas inyectables) y las relaciones sexuales sin protección entre hombres (MSM: “*Men Sex with Men*”). Aunque la transmisión heterosexual del virus es rara (<5% de los casos), el sexo anal sin protección es uno de los factores de riesgo más comunes en la transmisión del HCV. Cabe señalar que una de las vías principales de transmisión del virus en el transcurso del siglo pasado fue por medio de transfusiones de sangre contaminada, hasta que se tomaron las medidas oportunas para solventarlo. Por otro lado, los brotes emergentes son muy difíciles de identificar; los largos periodos de

incubación de este hacen muy difícil vincular casos a una fuente común de infección. Aparte, es preciso destacar que la mayoría de las infecciones se producen de forma asintomática. Debido a esto, el screening y la identificación de pacientes positivos es una de las formas más eficaces de combatir esta epidemia. Se puede detectar un individuo portador con la realización de una prueba serológica o una PCR dirigida; aunque pueden existir otros biomarcadores útiles para sospechar de la posible infección: como niveles anormales de transaminasas en sangre y de alanina aminotransferasas. Los anticuerpos contra la enfermedad tardan mucho en aparecer en el paciente infectado, sobre 1 año (Chan et al., 2016).

A lo largo de las últimas décadas los avances en el conocimiento de la biología molecular del HCV han sido esenciales para el desarrollo de terapias antivirales. Actualmente, los Antivirales de Acción Directa (DAA por su nombre en inglés: “*Direct-Acting Antivirals*”) han supuesto un cambio de paradigma en el tratamiento del HCV, logrando tasas de curación superiores al 95% en la mayoría de los pacientes (independientemente de su edad o comorbilidades). El mecanismo de acción de estos tratamientos se basa en interferir directamente en el proceso de replicación del virus, eliminando completamente la carga viral en los pacientes y reduciendo significativamente la mortalidad asociada. También, tienen la característica de poder ser administrados por vía oral. Sin embargo, el genotipo del virus sigue siendo un factor clave en la selección del tratamiento adecuado, lo que resalta la necesidad de una atención personalizada en el manejo de esta enfermedad (Lagging et al., 2017).

IDENTIFICACIÓN ACTIVA DE PACIENTES CON HEPATITIS C Y ANÁLISIS DE UN EHR

Como ya hemos comentado anteriormente, la hepatitis C continúa siendo uno de los retos más desafiantes que enfrenta la salud pública a nivel global, considerándose como una de las infecciones transmitidas por la sangre más prevalentes en todo el mundo. La gran prevalencia de esta enfermedad, acompañada por las limitaciones presupuestarias de los sistemas de salud, ha impulsado la necesidad de implementar estrategias racionales de cribado y tratamiento haciendo uso de herramientas informáticas como los EHR. De esta forma, la identificación activa de los pacientes infectados y el análisis

de los datos clínicos recopilados en los EHR se han convertido en las armas más pertinentes en la lucha contra esta afección (Li et al., 2017).

En este estudio, nos enfocaremos en la identificación activa de pacientes con hepatitis C mediante el análisis de un EHR. La integración de los EHR con el cribado de la hepatitis C permite una identificación rigurosa de los pacientes, lo que a su vez mejora la eficiencia del sistema de salud al priorizar a aquellos con mayor riesgo de complicaciones. Al mismo tiempo, realizaremos un estudio longitudinal retrospectivo de los casos de hepatitis C reportados en prácticas de atención primaria. Este planteamiento nos permitirá analizar la evolución de la enfermedad a lo largo del tiempo, poniendo especial atención en las cohortes más vulnerables. Para la estructuración de nuestro análisis, hemos tomado como referencia diferentes estudios previos, como el de Guiriguet et al. 2024 (Guiriguet et al., 2024); el cual dispone los pasos a seguir para el tratamiento de datos epidemiológicos, aunque esté centrado en infecciones de transmisión sexual en general. Este enfoque metodológico nos permitirá replicar estrategias que han resultado eficaces en el manejo de grandes volúmenes de datos clínicos, adaptándonos a las características específicas de la hepatitis C.

Para maximizar la eficiencia de estos programas y sacarles el mayor provecho, se recomienda filtrar a los pacientes según su edad y sexo, priorizando aquellos más vulnerables (Ridruejo et al., 2016). Varios estudios han demostrado que esta estrategia permite no solo mejorar la calidad de vida de los pacientes diagnosticados (ya que la mayoría de estos se logran tratar adecuadamente), sino también ayuda a prevenir nuevas infecciones. Esto es debido al hecho de que una vez que los pacientes conocen su condición, tienden a ser más cuidadosos en sus comportamientos de riesgo (Li et al., 2017). La hepatitis C afecta en su mayoría a personas nacidas entre 1945 y 1975, y se estima que más del 50% de los pacientes de esta cohorte no son conscientes de estar infectados, ya que la enfermedad se presenta de forma asintomática en la mayoría de los casos (Ward et al., 2014). Sin un diagnóstico y tratamiento conveniente, estas personas corren un riesgo considerable de desarrollar complicaciones graves como fibrosis hepática, cirrosis o carcinoma hepatocelular (Li et al., 2017).

Asimismo, en el contexto de este estudio se considera fundamental la colaboración entre médicos y bioinformáticos, como se muestra en la **FIGURA 3**. Los médicos juegan un papel clave en la confirmación del diagnóstico y la derivación de los pacientes para su tratamiento. Es aquí donde los bioinformáticos, como parte de este proceso, tienen un rol fundamental. Estos poseen las competencias necesarias para analizar y gestionar grandes cantidades de datos clínicos, optimizando así el diagnóstico y el seguimiento de los pacientes con hepatitis C. De igual forma, son capaces de comprender el contexto de una enfermedad a través de los datos que disponen (Jhaveri et al., 2021; Ridruejo et al., 2016).

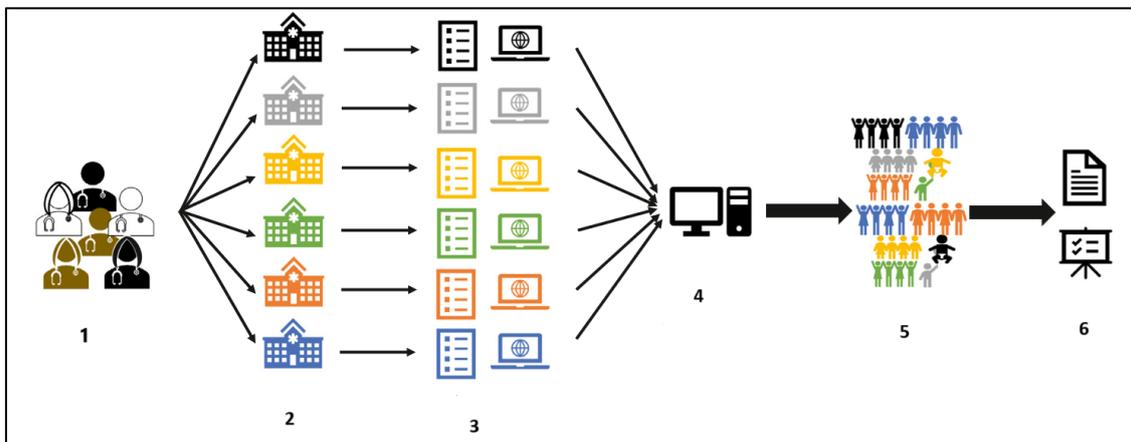


FIGURA 3. Flujo de Identificación y Segregación de Pacientes con Hepatitis C mediante Análisis de EHR. Podemos observar un flujo de etapas bien definidas: 1 Identificación mediante diagnóstico de la población de interés por parte del personal médico. 2 Recopilación de datos clínicos. 3 Centralización. 4 Análisis. 5 Estratificación de pacientes. 6 Acciones y seguimiento. Imagen extraída de Jhaveri et al., 2021

2. HIPÓTESIS DE TRABAJO Y OBJETIVOS

PREGUNTA DE INVESTIGACIÓN E HIPÓTESIS DEL TRABAJO

PREGUNTA DE INVESTIGACIÓN: pregunta PICO (Santos et al., 2007):

¿La identificación de pacientes mediante el uso de registros electrónicos digitales permite una detección temprana efectiva de individuos que presentan infección activa por hepatitis C que facilite intervenciones posteriores y reduzca la prevalencia y complicaciones de la enfermedad?

HIPÓTESIS:

Pensamos que analizar una base de datos de un Hospital que utiliza historia clínica digital nos va a permitir detectar aquellos casos de hepatitis C que no han recibido hasta la fecha tratamiento antiviral.

Nuestra idea se basa en buscar en listados de pacientes que se han realizado una prueba de Hepatitis C en los últimos 10 años en un centro sanitario de la Comunidad de Madrid casos de individuos que padecen una infección activa por Hepatitis C y no han sido tratados aún. De tal forma que, mediante la realización de este estudio, podremos analizar las características de estos pacientes y vincularlas a las consultas de Digestivo o Enfermedades Infecciosas para que puedan recibir una atención de calidad.

Por tanto, en lo que se refiere a la pregunta PICO:

- **Población:** Pacientes que se han realizado una prueba de Hepatitis C en los últimos 10 años en un centro sanitario de la CM.
- **Intervención:** Buscar en listados de pacientes electrónicos y localizar sujetos que siguen infectados actualmente y que no han recibido tratamiento.
- **Comparación:** Aquellos pacientes que no presentan la enfermedad o anteriormente se han curado.
- **“Outcome” o desenlace:** Detección temprana de pacientes con hepatitis C, reducción de prevalencia, incremento en el número de pacientes tratados y reducción de complicaciones relacionadas.

OBJETIVO PRINCIPAL Y OBJETIVOS SECUNDARIOS DEL TRABAJO

El **objetivo principal** de este trabajo es **identificar a los pacientes diagnosticados con hepatitis C activa** (denominados como pacientes positivos persistentes, los cuales no han recibido tratamiento antiviral) mediante el análisis de una base de datos procedente de las RCH y vincularlos nuevamente a las consultas médicas oportunas para que puedan recibir los cuidados apropiados para su enfermedad.

Por otro lado, algunos de los **objetivos secundarios** de este trabajo son los siguientes:

Realizaremos un estudio descriptivo: de toda la población general y otro por grupos de pacientes (sexo, edades, año especialidades...).

Clasificaremos los pacientes en distintas categorías: Segregaremos los pacientes según el resultado de su prueba y la dinámica de su infección.

Evaluaremos las tendencias generales de la enfermedad: Estudiaremos la evolución de la enfermedad en los pacientes a lo largo de los 15 años de registro para esclarecer su evolución y hacia dónde se dirige en el futuro. Para lograrlo, realizaremos un análisis estratificado por sexo, edad y otros factores relevantes, lo que nos permitirá identificar patrones específicos en subgrupos de pacientes. Este enfoque nos ayudará a comprender mejor la progresión de la hepatitis C en diferentes cohortes.

3. MATERIALES Y MÉTODOS

ORIGEN DE LOS REGISTROS ELECTRÓNICOS MÉDICOS UTILIZADOS

Los datos de los que disponemos en este trabajo provienen de registros clínicos electrónicos obtenidos del Sistema Hospitalario de la Comunidad de Madrid: el registro fue solicitado al laboratorio encargado de realizar las pruebas diagnósticas en esta materia del Hospital Universitario Infanta Leonor de Madrid. Es importante destacar que los pacientes se encuentran encriptados mediante un código, es decir, están pseudonimizados para proteger la privacidad de estos; cumpliendo en todo momento con el Reglamento General de Protección de Datos (Samper, 2020). Estos se corresponden a pacientes a los que se les ha realizado una prueba ELISA para detectar antígenos anti VHC y/o una qPCR para el cribado del virus de la Hepatitis C. La qPCR es una técnica altamente sensible y específica que permite detectar la presencia del ARN del virus en el organismo y cuantificar el nivel de unidades virales en el suero de los pacientes. Esta información es crucial para la identificación de pacientes portadores del virus, monitorear la progresión de la infección y determinar la necesidad de tratamiento (Garinis et al., 1999).

FORMATO ORIGINAL DE LA BASE DE DATOS UTILIZADA PARA EL ESTUDIO

En cuanto a las características del archivo original, este se corresponde a un archivo tabulado en formato “.xls” compuesto por 13 hojas diferentes, cada una correspondiente a un año de registro, desde 2010 a 2023. El propio registro contiene información muy relevante para el análisis, con un total de 28.414 entradas que reflejan las diversas pruebas o solicitudes realizadas a cada paciente individual. Por otro lado, las diferentes columnas del archivo están vinculadas a las características de los consultantes y los resultados de cada prueba realizada (9 columnas en total). De izquierda a derecha, se procederá a describir cada columna (o característica) de este:

- **Petición: Identificador único para cada solicitud** o prueba realizada. Necesario para la gestión de las peticiones dentro del sistema hospitalario. Resaltar que un paciente puede tener más de una petición, será lo común.
- **Fecha:** Esta columna contiene la fecha y hora en la que se registró la solicitud. Está en un formato estándar: "AAAA-MM-DD HH:MM".
- **Servicio:** Indica el servicio dentro del hospital donde se realizó la prueba.
- **Ámbito:** Este campo especifica el ámbito sanitario donde se realizó la prueba.
- **Urgente:** Se trata de un indicador binario que señala si la prueba fue solicitada con carácter de urgencia ("S") o no ("N").
- **Código Local o NHC:** El Número de Historia Clínica del paciente. Cada paciente individual posee uno.
- **Edad:** La edad del paciente en el momento en el que se realiza la prueba.
- **Sexo:** Género del paciente al nacimiento: "H" hombre o "M" mujer.
- **Resultado:** Este campo muestra el resultado de la prueba realizada en cada paciente.

	Peticion	Fecha	Servicio	Ambito	Urgente	NHC	Edad	Sexo	Resultado
1	17547933	2023-01-01 11:57:00	Z6_URGENCIAS GENERALES HOSP	HOSP	N	6/928588	58	H	No procede. Ac anti VHC negativos.
2	17547937	2023-01-02 07:37:00	Z6_MEDICINA INTERNA HOSP	HOSP	N	6/1192365	59	H	514
3	19996470	2023-01-02 08:37:00	Z6_REUMATOLOGÍA CCEE	CCEE	N	6/296208	57	H	No procede. Ac anti VHC negativos.
4	19204207	2023-01-02 09:47:00	z6_VS DIGESTIVO	CCEE	N	6/853465	57	H	6979
5	19987340	2023-01-02 09:47:00	Z6_REUMATOLOGÍA CCEE	CCEE	N	6/884999	45	M	No procede. Ac anti VHC negativos.
6	19207581	2023-01-02 10:47:00	Z6_DIGESTIVO CCEE	CCEE	N	6/149447	88	H	2475490
7	19991560	2023-01-02 11:17:00	Z6_MEDICINA INTERNA CCEE	CCEE	N	6/1092929	55	H	No detectable
8	19988669	2023-01-02 11:47:00	Z6_DERMATOLOGÍA CCEE	CCEE	N	6/998177	49	H	No procede. Ac anti VHC negativos.
9	17548073	2023-01-03 07:37:00	Z6_MEDICINA INTERNA HOSP	HOSP	N	6/772549	57	M	No procede. Ac anti VHC negativos.

FIGURA 4. Ejemplo del formato original de los datos que contenía el archivo.

HERRAMIENTAS Y RECURSOS UTILIZADAS PARA EL MANEJO DE LOS DATOS: RStudio.

Para llevar a cabo la parte práctica del trabajo, hemos hecho uso de la plataforma RStudio, la cual lleva implementada el lenguaje de programación R. Adjunto a esta memoria se encuentra el ANEXO, el cual es un archivo extraído directamente de RStudio en el que se describen todos los recursos informáticos utilizados y los distintos pasos llevados a cabo en el análisis de los datos pertinentes.

AUTORIZACIÓN DEL COMITÉ DE ÉTICA

Debido a la propia naturaleza de los datos (proviene de registros clínicos y contienen información de estudios realizados en humanos), se ha solicitado la autorización del Comité de Ética correspondiente. Se adjunta la copia de esta autorización al presente informe para verificar que se cumple con los requisitos éticos y legales necesarios para el manejo de datos sensibles de pacientes. Asimismo, hemos seguido en todo momento las directrices del “Manual de Buenas Prácticas de Investigación” extraído del propio “MÓDULO 10 – TRABAJO DE FIN DE MÁSTER” y los principios establecidos en la Declaración de Helsinki (Asamblea Médica Mundial, 2000), garantizando de este modo la estricta confidencialidad de los datos y el respeto de los derechos de los pacientes involucrados en el estudio.

PLAN DE MANEJO DEL DATO Y METODOLOGÍA SEGUIDA

La parte práctica del trabajo constó de varias fases que permitieron abordar de manera estructurada el análisis de los datos. Adjunto a esta memoria se encuentra el Anexo extraído directamente de RStudio el cual cuenta con el código implementado para llevar a cabo el análisis.

El fundamento principal del trabajo se basa en clasificar a los pacientes en distintas categorías dependiendo del estado de su enfermedad (si son positivos o negativos respecto a carga viral) y de la fecha asociada a este resultado y analizar las características de los grupos segregados: Inicialmente los clasificamos según los resultados de la primera prueba que se realizaron, obteniendo la prevalencia de la enfermedad. Después, se llevó a cabo una clasificación más general diferenciando entre pacientes negativos y pacientes que han sido positivos por lo menos una vez en el registro. Por último, establecimos una clasificación de pacientes más compleja basada en el concepto de “líneas temporales continuas” en la que diferenciamos cuatro categorías posibles: pacientes curados, pacientes positivos persistentes (los más importantes), pacientes negativos persistentes y pacientes que han adquirido la infección durante el seguimiento. A continuación, se procede a describir paso por paso el análisis llevado a cabo y el procedimiento de clasificación en cada caso:

0. Carga, Unificación y preparación de los datos clínicos. Primeros pasos.

Una vez que los datos fueron suministrados directamente por el tutor, se procedió con su manipulación. En primer lugar, se unificaron. Los registros clínicos originalmente estaban contenidos en múltiples hojas de *Excel*, separadas por cada año diferente de registro. De este modo, se combinaron todas ellas para obtener un único *DataFrame* sobre el cual poder trabajar.

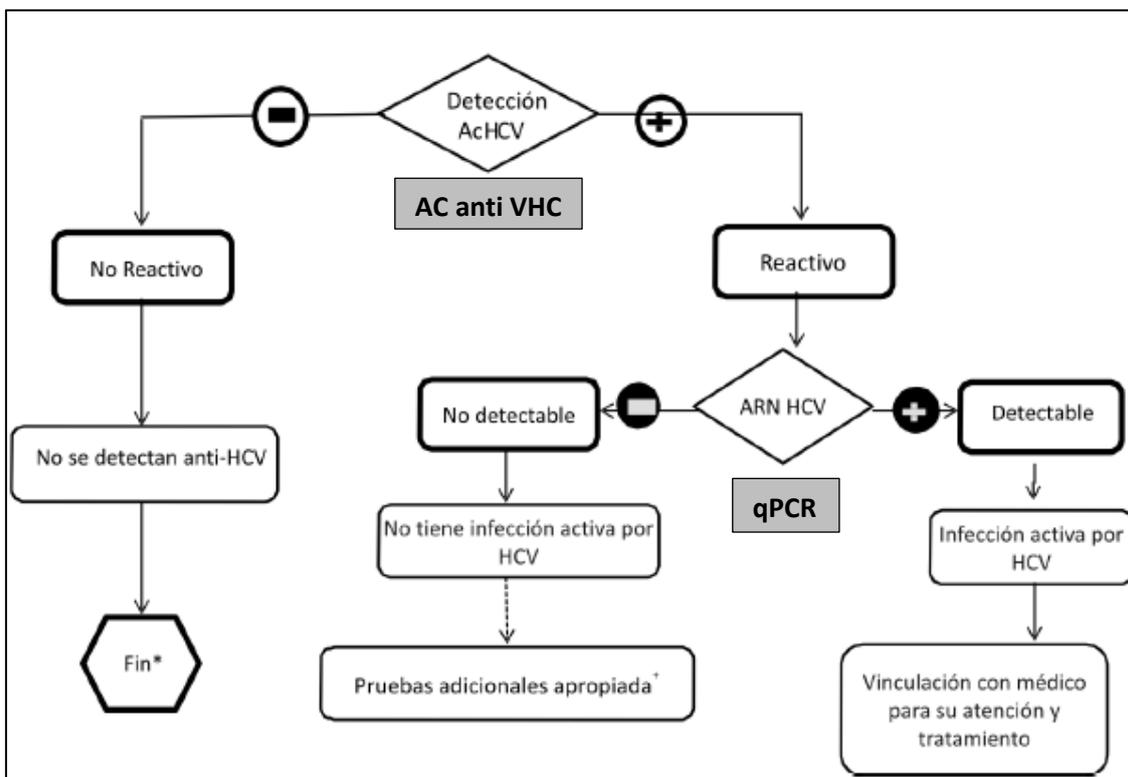
1. Identificación de Pacientes Únicos y realización de un estudio descriptivo global: EDA general.

En este apartado, identificamos a todos los pacientes únicos mediante el NHC (Número de Historia Clínica), rastreando todas las interacciones que estos han tenido con el Sistema de Hospitales de la Comunidad de Madrid. Esta etapa nos permitió no solo conocer cuántos pacientes únicos componen el estudio (la población total), sino también cuántas pruebas se han realizado a cada uno de ellos, obteniendo estadísticas descriptivas como la mediana de pruebas por paciente o la variabilidad en las frecuencias de las mismas. Además, para realizar un análisis más preciso, en este primer EDA general trabajamos con la primera prueba realizada a cada paciente en relación con el virus de la Hepatitis C. Este enfoque nos permitió estudiar la primera interacción de los pacientes con el sistema de salud. A partir de esta subpoblación, pudimos responder preguntas clave como la edad y el sexo de los pacientes que acuden a la consulta, el servicio y el ámbito hospitalario donde acude más gente en esta materia, etc.

2. Clasificación de pacientes 1: Identificación de pacientes que son clasificados como positivos en la primera prueba y evaluación de la tendencia temporal.

El análisis del resultado de la primera prueba en cada paciente es esencial para identificar quiénes resultaron positivo en su primer contacto con el sistema de salud, concepto conocido como Prevalencia. Este indicador proporciona una visión clara de la proporción de la población afectada en un momento específico.

Para ello, en primer lugar, se clasificaron los sujetos en pacientes positivos o pacientes negativos según la presencia de anticuerpos hacia el patógeno y su carga viral; características que encontramos en la variable “Resultado”. Un resultado positivo se asigna a aquellos con carga viral detectable (ARN viral), mientras que un resultado negativo corresponde a aquellos individuos que resultan negativos en el test de antígenos anti HCV o no se les detecta carga viral en la qPCR (o cuyo nivel de partículas virales está por debajo del umbral de detección de la máquina). Para facilitar este proceso, se creó una nueva columna que resume de manera clara si el resultado de cada paciente fue positivo o negativo. Este filtro se empleó durante todo el trabajo para la categorización del estado de los pacientes. En el **ESQUEMA 1** se ilustra el proceso de detección y clasificación de los casos clínicos en función de la presencia de anticuerpos contra el HCV y/o su carga viral que se realiza en el ámbito clínico.



ESQUEMA 1. Diagrama de flujo del proceso de detección y clasificación de pacientes con Hepatitis C. Los pacientes son clasificados como reactivos o no reactivos según la presencia de anticuerpos anti-HCV. En caso de resultado reactivo, se realiza una prueba adicional de qPCR para determinar la presencia de infección activa. Imagen extraída de Ridruejo et al., 2016.

De igual forma, se creó una nueva columna llamada “Year” cuyo objetivo es simplificar la fecha de realización de la prueba a formato de año, facilitando el posterior análisis temporal.

Clasificados los pacientes y asignándole el año en el que se realizaron la prueba, obtuvimos métricas muy interesantes como la proporción de positivos y negativos (Prevalencia) y la proporción de pacientes con resultado positivo por el número de pruebas realizadas por año. Además, se analizaron las tendencias temporales de la enfermedad, estratificando los datos por edad y sexo, lo que permitió identificar diferentes patrones en la prevalencia de la hepatitis C a lo largo del tiempo.

3. Clasificación de pacientes 2: Comparación global entre pacientes positivos vs pacientes negativos.

En este apartado identificamos a los pacientes que han sido positivos al menos una vez durante su seguimiento clínico. Como en la anterior clasificación, esta analogía nos brindó la oportunidad de analizar si existían diferencias significativas entre los distintos grupos (positivos vs negativos) en variables como la edad, sexo o número de pruebas realizadas. Asimismo, además de estudiar las tendencias de los pacientes positivos a lo largo del tiempo (segregando de nuevo por sexo y edad), también evaluamos otros aspectos relevantes como el número de pruebas totales realizadas por año, la comparación del número de pruebas que se realiza cada grupo (pacientes positivos y negativos), etc.

4. Clasificación de pacientes 3: Clasificación de los pacientes según los diferentes eventos de la enfermedad.

Este es el apartado más relevante del trabajo ya que se llevó a cabo el objetivo principal de este estudio. En este caso el análisis se centró en la evolución de los pacientes a lo largo del tiempo, segregándolos en distintos grupos según la dinámica de su enfermedad. De esta forma, podemos obtener 4 supuestos diferentes:

- **Paciente curado:** Paciente positivo que negativiza con el tiempo. Esta clasificación incluye tres fechas críticas: la primera prueba realizada, la primera prueba positiva, y la última prueba en la que se detecta la negativización del virus (es decir, la prueba que confirma que el paciente se ha curado). Esta estructura permite identificar aquellos pacientes que en algún momento fueron positivos pero que al final del seguimiento lograron negativizar. Se garantiza que las fechas estén ordenadas cronológicamente, asegurando que la primera prueba sea la más antigua y la última la más reciente, con una validación adicional de que la última prueba es negativa, indicando la curación completa. Asimismo, se analizó la tendencia del número de pacientes curados por año del seguimiento.
- **Paciente positivo persistente:** En esta clasificación el paciente continúa siendo positivo en la última prueba que se realizó, lo que indica una infección no resuelta. Identificar a estos pacientes es el objetivo principal de este trabajo, para posteriormente poder remitirlos a las autoridades sanitarias y que estas puedan proceder con su seguimiento.
- **Paciente negativo persistente:** Aquellos cuya infección por hepatitis C ha sido descartada o resuelta, ya que han mantenido un resultado negativo hasta su última prueba registrada. Estos pacientes también presentan dos fechas clave: la primera prueba y la última prueba.
- **Paciente negativo que positiviza.** Se trata de pacientes que inicialmente fueron negativos y posteriormente positivizaron, es decir, aquellos que se infectaron durante el seguimiento. Para identificarlos, se necesitan dos fechas clave: la primera fecha que corresponde a la primera prueba negativa y la última fecha que corresponde a la primera prueba positiva (después de la negativa). La clasificación de estos pacientes resultó muy útil a la hora de explicar la diferencia del número de individuos entre las diferentes subpoblaciones del estudio.

Igualmente, en este apartado se llevó a cabo una validación de los resultados con el objetivo de comprobar que segregamos a los pacientes exitosamente.

5. Estudio descriptivo de los pacientes positivos persistentes.

Se realizó un análisis detallado de las características de los pacientes que siguen siendo positivos hasta el día de hoy. Para ello se hizo uso de la matriz generada en el apartado anterior que contenía la subpoblación de positivos persistentes. Por último, guardamos los registros y los pusimos en manos de los profesionales sanitarios para que pudiesen comenzar a realizar una búsqueda activa sobre ellos.

4. RESULTADOS OBTENIDOS

Una vez descritos los diferentes pasos metodológicos a seguir en el análisis de los datos, en esta sección se presentan los resultados obtenidos a lo largo del estudio:

1. IDENTIFICACIÓN DE PACIENTES ÚNICOS Y REALIZACIÓN DE ESTUDIO DESCRIPTIVO

GLOBAL: EDA GENERAL.

Identificación de pacientes únicos mediante el NHC:

El primer paso consistió en identificar a todos los pacientes únicos por su ID NHC. Para ello, inicialmente se llevó a cabo un pequeño análisis para evaluar el número de pruebas al que cada paciente se ha sometido, obteniendo los siguientes resultados:

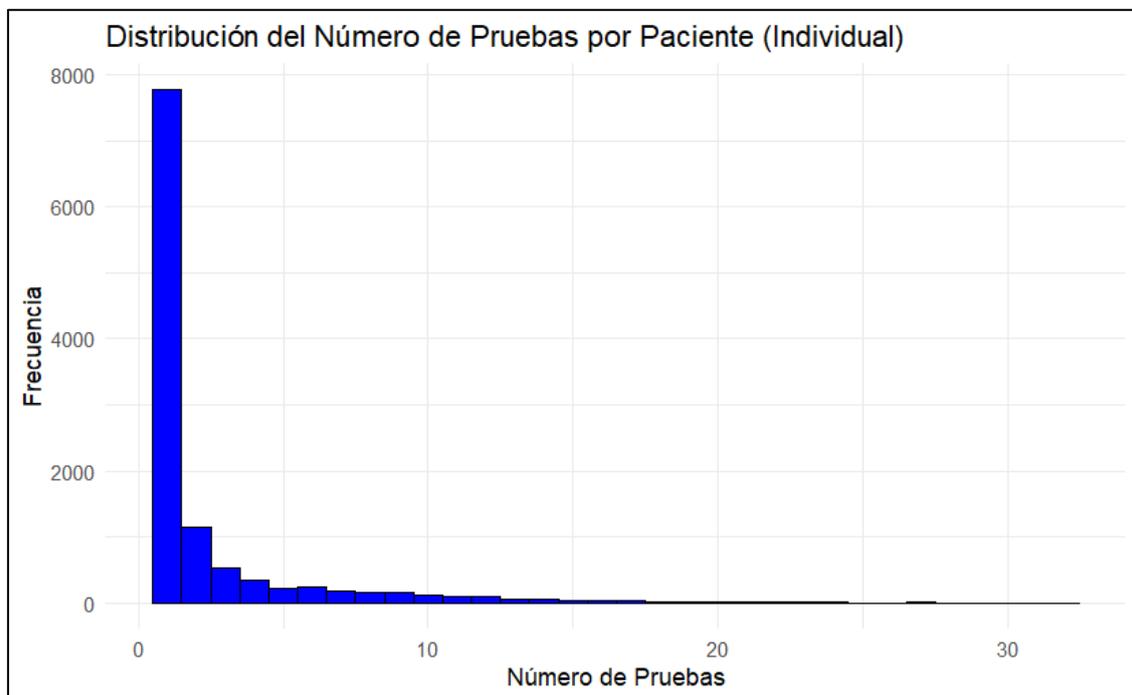


FIGURA 5. Gráfico de barras de la distribución del Número de Pruebas realizadas por paciente individual.

TABLA 1. Resumen estadístico del conteo medio de pruebas que se realizan los pacientes individualmente.

Mín.	1st Qu.	Mediana	Media	3rd Qu.	Max.
1,00	1,00	1,00	2,482	2,00	32,00

Como podemos observar tanto en la **TABLA 1** como en la **FIGURA 5**, la mayoría de los pacientes se realizaron una sola prueba. Sin embargo, el promedio de pruebas realizadas por pacientes es 2.49.

Por otra parte, una vez que filtramos la matriz de datos original para que únicamente estuvieran presentes los resultados de la primera prueba realizada en cada paciente, obtuvimos **una población total de 11404 pacientes.**

1.1 ANÁLISIS DESCRIPTIVO GLOBAL: EDA GENERAL. Respectivo a la 1ª prueba.

Obtenida la población total, realizamos el primer Análisis Exploratorio de los Datos, con el objetivo de tener una mejor comprensión de las variables claves de los pacientes en general, como la edad, el sexo y el ámbito de la consulta:

Distribución de la Edad en la población:

TABLA 2. Resumen estadístico de la distribución de la EDAD

Mín.	1st Qu.	Mediana	Media	3rd Qu.	Max.
0,00	37	49	49,5	60	100

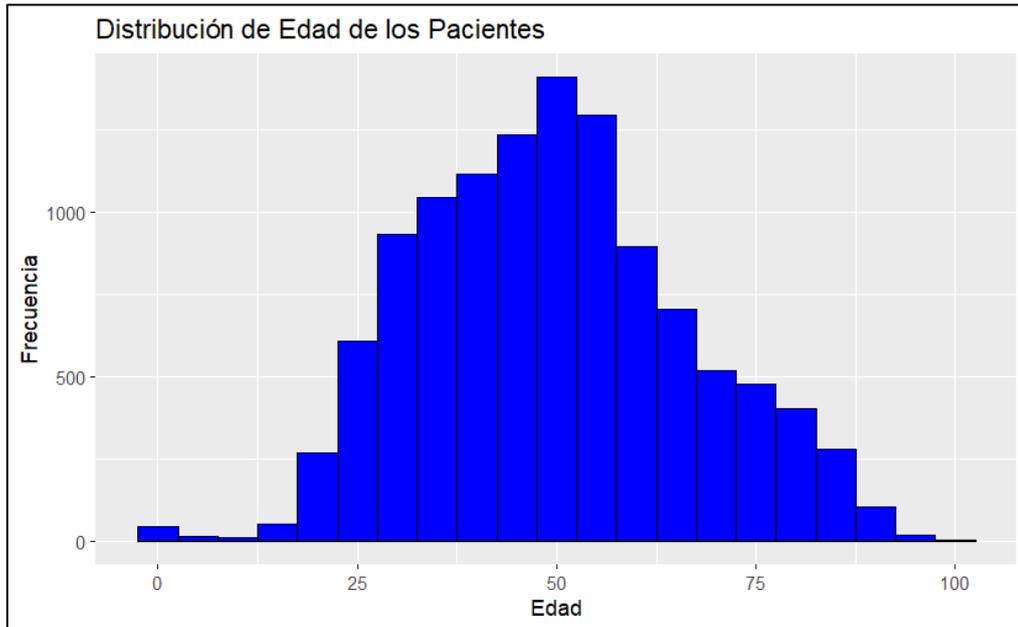


FIGURA 6. Histograma de la Distribución de la edad de los Pacientes que se someten a las

Las dos representaciones, la **TABLA 2** y la **FIGURA 6**, nos muestran la distribución de la edad de los pacientes que se realizaron pruebas de Hepatitis C. El histograma muestra que la mayoría de los pacientes tienen edades comprendidas entre 40 y 60 años, con un pico en la mediana señalando que la mayoría de los pacientes tienen 50 años; el resumen estadístico corrobora esta información.

Distribución del sexo en la población:

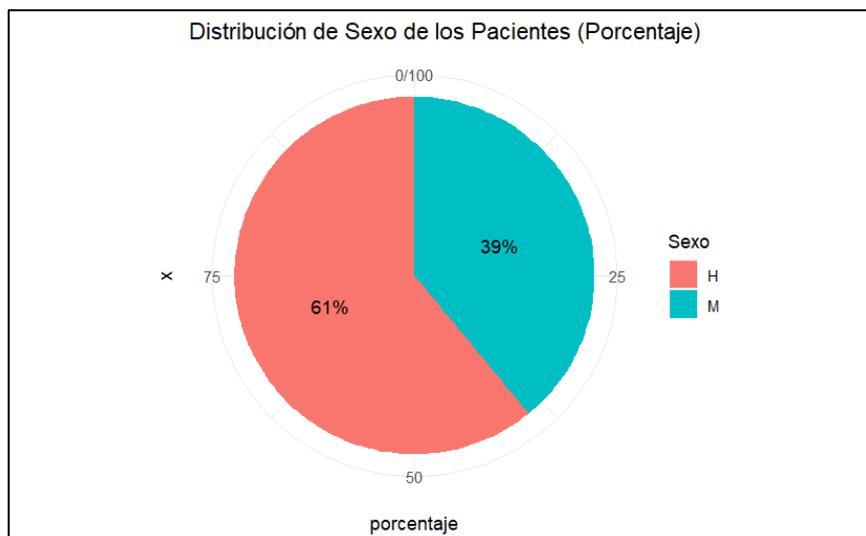


FIGURA 7. Gráfico de sectores de la distribución del sexo de los pacientes.

Como podemos observar en la **FIGURA 7**, hay un mayor número de hombres que de mujeres sometiéndose a este tipo de pruebas: aproximadamente 6000 hombres y menos de 5000 mujeres, el 60% y el 40% respectivamente. Asimismo, el valor estadístico p-valor (Alpha = 0.05) es extremadamente pequeño, indicando que hay una diferencia altamente significativa entre las dos distribuciones.

Distribución del Servicio de realización de la prueba en la población:

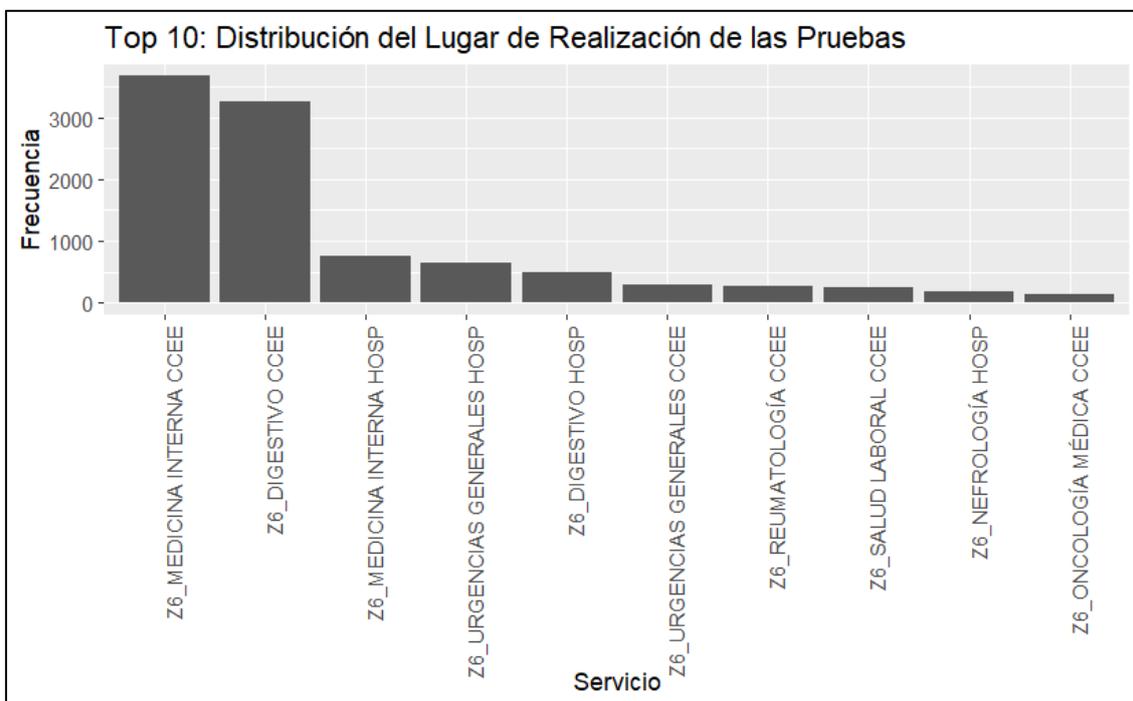


FIGURA 8. Gráfico de barras de la distribución del lugar de Realización de las pruebas

El gráfico de barras de la **FIGURA 8** muestra la distribución del lugar de realización de pruebas. Observamos que la mayoría de las pruebas se concentran en unos pocos servicios, siendo Medicina Interna el servicio con el mayor número de pruebas realizadas, seguido por Digestivo. Esta gran concentración de pacientes en las consultas de Digestivo se relaciona principalmente con el hecho de que la hepatitis C está asociada fuertemente a problemas relacionados con el hígado.

2. CLASIFICACIÓN DE PACIENTES 1: Identificación de pacientes que son clasificados como positivos en la 1ª prueba.

2.1 CLASIFICACIÓN DE PACIENTES POR EL RESULTADO DE LA 1ª PRUEBA.

El análisis del resultado de la primera prueba de cada paciente es crucial para conocer la prevalencia que presenta la enfermedad en la población. Este paso permite identificar a los pacientes que presentaron un resultado positivo desde su primera interacción con el sistema de salud, lo cual indica la necesidad inmediata de seguimiento y tratamiento, frente a aquellos que dieron negativo y pueden requerir solo observación. Al centrarnos en la primera prueba realizada, se puede analizar el estado inicial de los pacientes y las características que puedan estar asociadas con un mayor riesgo de infección, como el sexo y la edad. Este análisis nos ayuda a entender el comportamiento inicial de la población en términos de infección:

Número de Casos Positivos vs Número de casos Negativos:

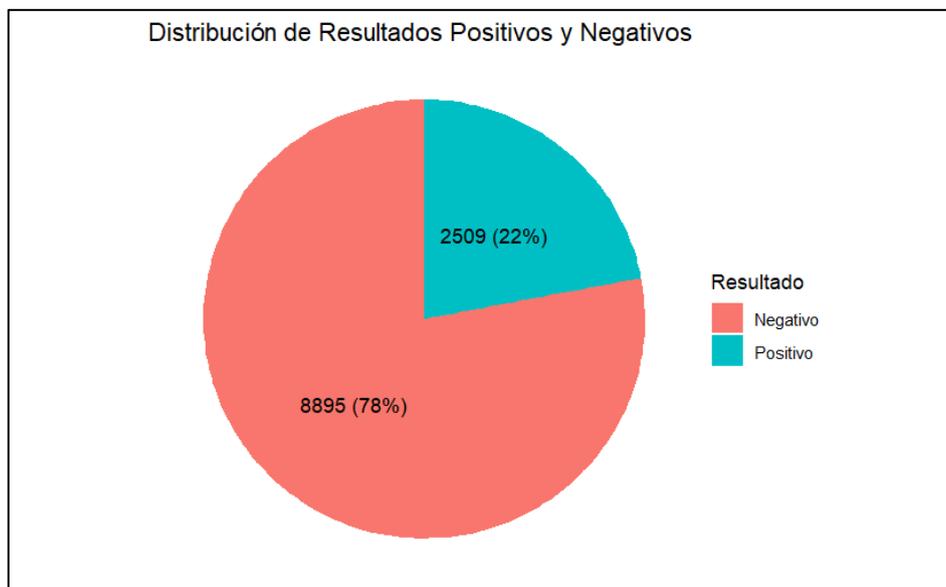


FIGURA 9. Gráfico de sectores del número de casos positivos frente a los casos negativos.

En la **FIGURA 9** se muestra la proporción de pacientes con resultados de hepatitis C positivos y negativos en la primera prueba que se les realizó. De los pacientes analizados, el 22% resultaron positivos (2509 pacientes); en contraste, un 78% (8895 pacientes) resultaron negativo. Esto muestra una relación global aproximada de 1 paciente diagnosticado como positivo por cada 5 pacientes nuevos.

Número de pacientes nuevos por año:

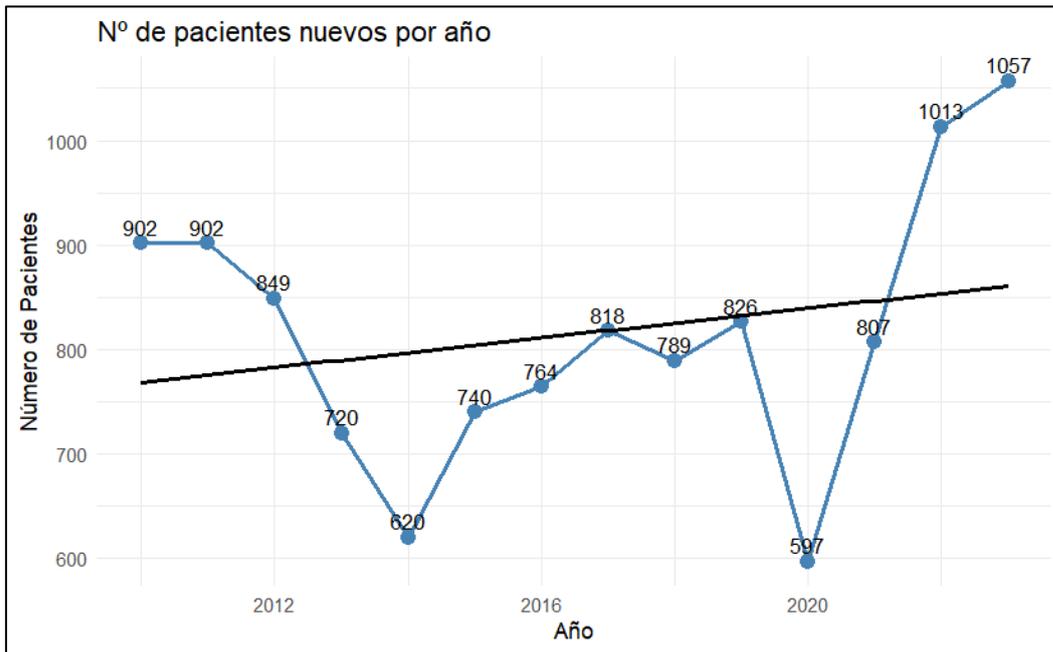


FIGURA 10. Gráfica del número de pacientes nuevos que ingresan al sistema por cada año de registro.

La **FIGURA 10** muestra el número de pacientes nuevos que se han añadido cada año al registro clínico. Observamos un incremento constante en los últimos años, con varias fluctuaciones apreciables, sobre todo en 2014 y 2020.

Tendencia general del número de pruebas realizadas por año.

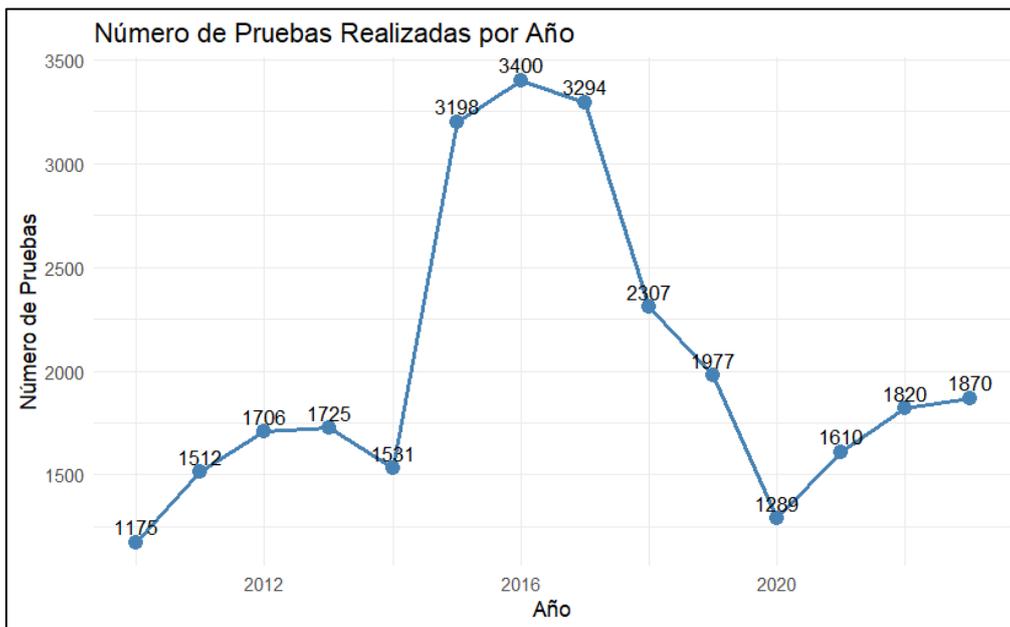


FIGURA 11. Gráfica de la tendencia del número de pruebas que se realizaron cada año del registro.

En la **FIGURA 11** podemos observar varios periodos con características diferentes: un aumento ligero en el número de pruebas entre 2010 y 2014, seguido de una fuerte subida a partir de 2015. Por otro lado, después de este fuerte repunte, sigue una abrupta bajada y luego una posterior estabilización a los mismos números que hace 10 años.

2.2 TENDENCIA TEMPORAL GENERAL DE LA ENFERMEDAD: 1ª VISITA.

En este caso, analizaremos la tendencia temporal general de la enfermedad en el transcurso del tiempo. De tal modo que utilizaremos la clasificación de pacientes llevada a cabo en el apartado anterior 2.1 para realizar un conteo del número de pacientes con resultado positivo por cada año del registro disponible. Este tipo de análisis temporal es fundamental para observar si la prevalencia de la enfermedad ha ido aumentando o disminuyendo con el tiempo:

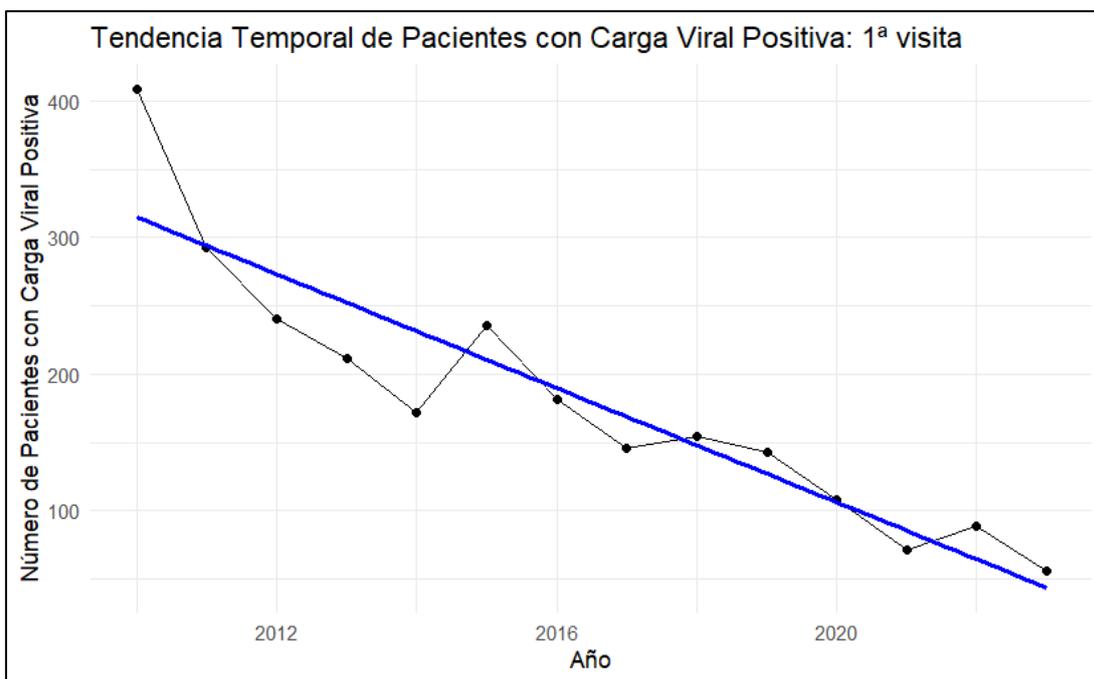


FIGURA 12. Gráfica de la tendencia del número de pacientes positivos por cada año de registro.

En la **FIGURA 12** se representa la tendencia temporal del número de pacientes detectados con carga viral positiva a lo largo de los años. Se observa un descenso constante en el número de casos desde el momento que se creó el registro, donde inicialmente había más de 400 pacientes con carga viral positiva, hasta el momento más actual (2023), donde la cifra ha descendido por debajo de los 100 pacientes diagnosticados.

2.3 TENDENCIAS TEMPORALES ESTRATIFICADAS:

En este apartado profundizaremos en el análisis estratificando estos datos según la edad y el sexo de los pacientes:

Tendencia temporal según el Sexo:

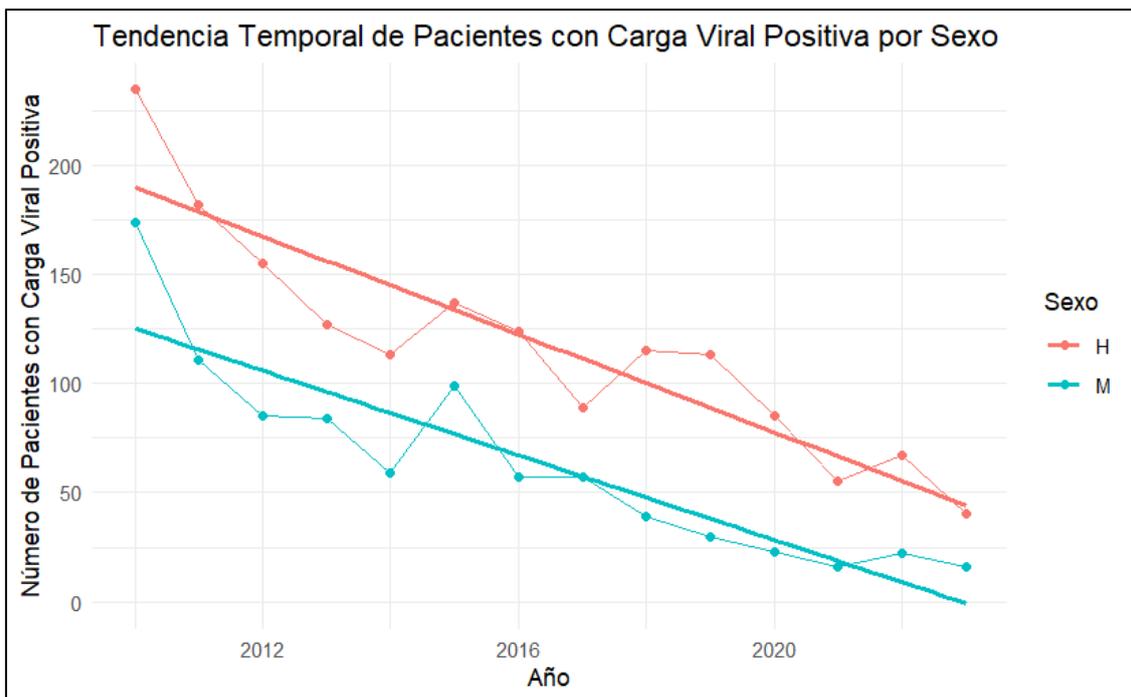


FIGURA 13. Gráfica de la tendencia temporal del número de pacientes positivos estratificada por sexo.

En la **FIGURA 13** se analiza la tendencia temporal del número de pacientes con carga viral positiva desglosado por sexo. A lo largo del periodo de estudio, se observa claramente que el número de hombres con resultado positivo ha sido mucho mayor que el de las mujeres. Asimismo, detectamos que ambos grupos presentan una tendencia similar: descendente.

Tendencia temporal según el Rango de Edad:

La **FIGURA 14** muestra la tendencia temporal de pacientes con carga viral positiva, segmentada por rangos de edad. Esta visualización nos permite identificar qué grupos etarios han sido más afectados por la enfermedad a lo largo del tiempo. Al desglosar los resultados por rangos de edad; como individuos entre 20 y 39, 40 y 59, y mayores de 60, se pueden observar posibles variaciones en la prevalencia del virus dentro de diferentes grupos poblacionales:

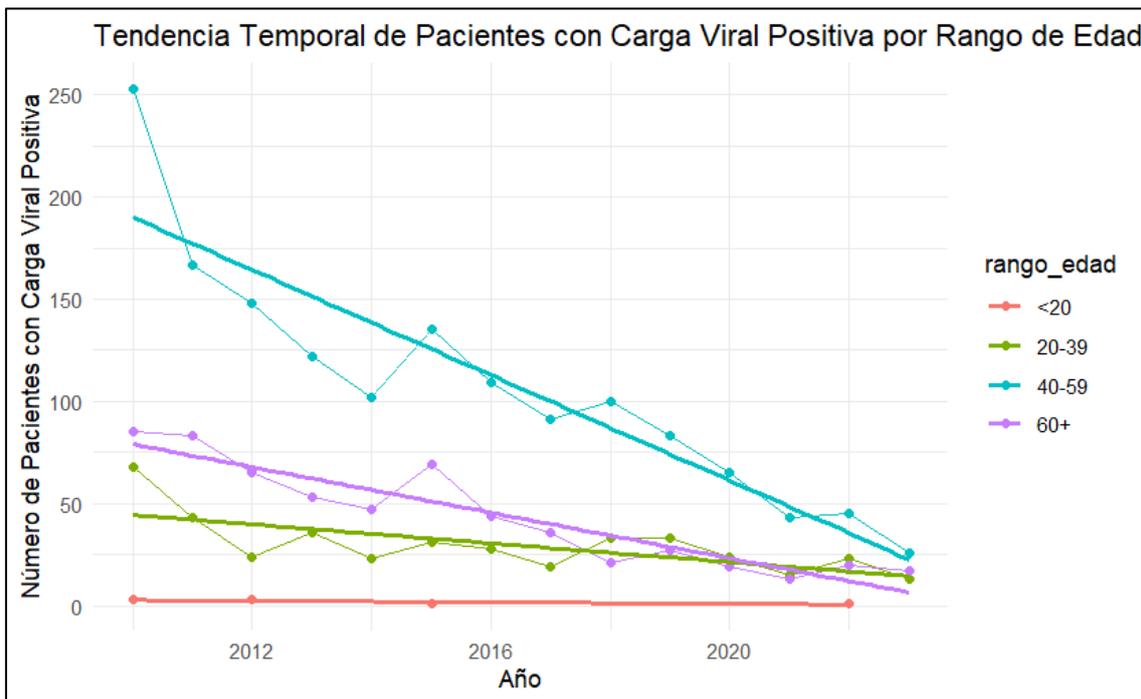


FIGURA 14. Gráfica de la tendencia temporal de los pacientes positivos estratificada por grupos etarios.

En la gráfica contemplamos que el mayor número de casos de hepatitis C con resultado positivo se concentra en el grupo de 40-59 años, grupo etario que ha mantenido un número elevado de casos a lo largo del tiempo, aunque en los últimos años parece que se han reducido drásticamente. Por otro lado, la tendencia del número de casos nuevos en el grupo de 20-39 parece no disminuir al mismo ritmo que los demás, si no que se mantiene constante en el tiempo. En contraste, los grupos de 60+ años presentan una tendencia descendente.

Tendencia temporal de Pacientes con Carga Viral Positiva por Sexo y Rango de Edad.

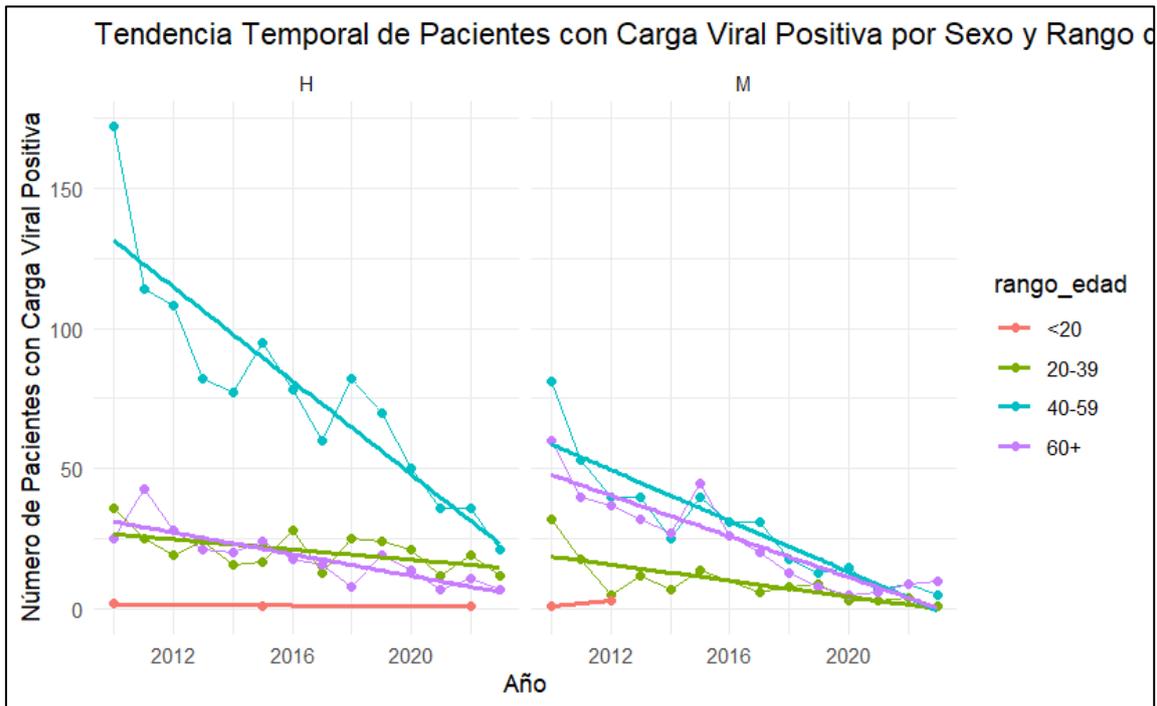


FIGURA 15. Gráfica de la tendencia temporal del número de pacientes con resultado positivo estratificado por sexo y grupos etarios

En la **FIGURA 15** se representan a la izquierda los hombres y, a la derecha las mujeres. Se observa que el grupo etario comprendido entre 40-59 años (línea azul) sigue siendo el más afectado en ambos sexos. Sin embargo, tanto en hombres como en mujeres se ha reducido el número de casos. En cuanto al grupo más joven (20-39 años) (línea verde), destaca el mantenimiento en el número de casos en los hombres, presentando una tendencia estable. En contraste, la tendencia de las mujeres jóvenes parece disminuir con el paso de los años de manera similar al resto de grupos. Por otro lado, los grupos que comprenden los pacientes más envejecidos (60+ años) (línea morada) han mostrado una disminución progresiva del número de casos a lo largo del tiempo. Llama la atención la diferencia en la cantidad de casos de mujeres de +60 que presentaban la enfermedad alrededor del 2010 comparada con los hombres de la misma franja de edad.

3. CLASIFICACIÓN 2: COMPARACIÓN GLOBAL ENTRE PACIENTES POSITIVOS VS PACIENTES NEGATIVOS.

En esta sección, en lugar de centrarnos en los pacientes que ingresan a la consulta por primera vez, nos enfocamos en la proporción de pacientes que han sido positivos al menos una vez a lo largo del tiempo. Igualmente, esta analogía nos permitirá analizar si existen diferencias significativas entre los distintos grupos en variables como la edad, el sexo, o el número de pruebas realizadas.

Esta clasificación resultó en un total de 8790 individuos clasificados como negativos y 2626 pacientes que en algún momento del seguimiento resultaron positivos de una población total de 11404 pacientes. Comparados con los datos de la primera visita de los enfermos (**APARTADO 2, FIGURA 10**) los cuales resultaron en 8895 pacientes negativos y 2509 pacientes positivos, podemos observar que los pacientes con resultado positivo en la primera visita (2509) son ligeramente menores que los pacientes positivos totales que se han detectado a lo largo del seguimiento (2626). Esta diferencia se debe a que probablemente estos hayan contraído la infección durante el seguimiento: en un primer momento eran negativos y posteriormente han positivizado. Este fenómeno lo abordaremos en el siguiente apartado.

Comparación de la distribución de Edad de pacientes positivos vs negativos:

TABLA 3. Análisis descriptivo de la EDAD de pacientes Positivos vs Negativos.						
PACIENTES	Min.	1st Qu	Median	Mean	3rd Qu.	Max.
POSITIVO	0	43	50	51.92	58	97
NEGATIVO	0	35	48	48.9	61	100

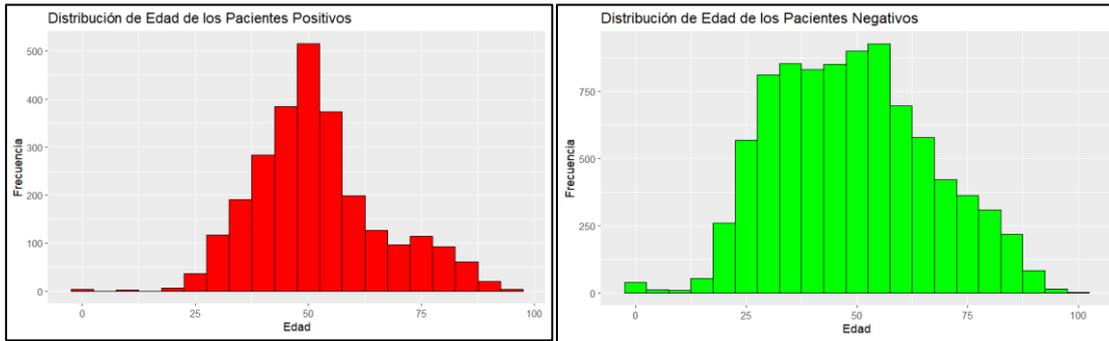


FIGURA 16 & FIGURA 17. Comparación de la distribución de la edad de los pacientes positivos (izquierda) frente a los pacientes negativos (derecha).

Al comparar las distribuciones de edad entre pacientes positivos (**FIGURA 16**) y negativos (**FIGURA 17**) se observa que los pacientes positivos tienden a concentrarse en un rango de edad más específico, con un pico significativo alrededor de los 50 años. Por otro lado, la distribución de los pacientes negativos es más uniforme, con un pico más amplio entre los 40 y 60 años, pero también incluye una mayor proporción de individuos más jóvenes y ancianos. Los resultados observados en las diferentes distribuciones se corroboran con la información de la **TABLA 3**. Se comprobó la veracidad de estas premisas con la realización de un test estadístico, más concretamente la prueba de Mann-Whitney U, cuyo el objetivo es comprobar si existe una diferencia considerable entre las distribuciones de los dos grupos: resultando en un p-valor de 2.2×10^{-16} , lo que significa que si existe una disparidad estadísticamente significativa entre las distribuciones de edades de estos grupos ($\text{Alpha} = 0,05$).

Comparación de la distribución del Sexo de los pacientes Positivos vs Negativo:

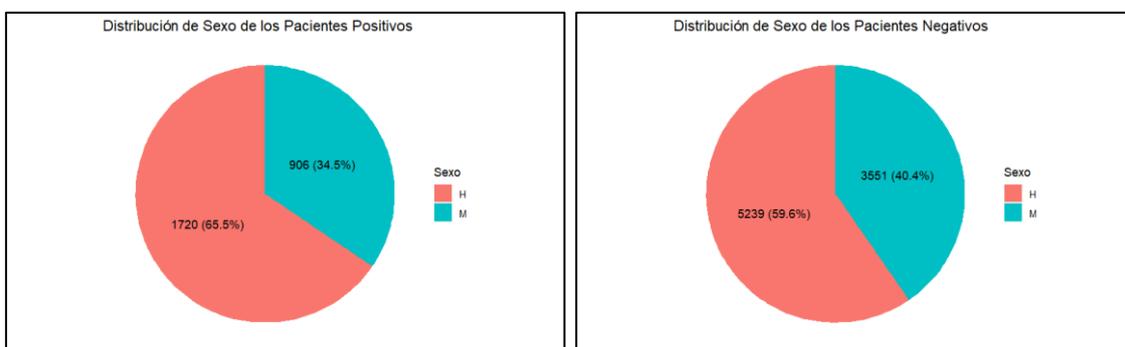


FIGURA 18 & FIGURA 19. Gráficos de sectores de la distribución del sexo en pacientes positivos (izquierda) y pacientes negativos (derecha).

Los gráficos muestran la distribución de sexo entre los pacientes positivos (**FIGURA 18**) y negativos (**FIGURA 19**). En ambos casos, los hombres representan la mayoría, pero hay una diferencia significativa en las proporciones: en el caso de los pacientes positivos, el 65% son hombres y el 35% son mujeres. Estas observaciones las corroboramos mediante la realización del test estadístico Chi-cuadrado, el cual resultó un p-valor de 6.209e-8, lo que indica que existe una diferencia estadísticamente significativa entre ambas distribuciones (Alpha = 0.05).

Comparación del número de pruebas que se realizan los pacientes Positivos frente a los Negativos:

TABLA 4. Comparación del nº de pruebas que se realizan los Positivos vs Negativos						
PACIENTES	Min.	1st Qu	Median	Mean	3rd Qu.	Max.
POSITIVO	1	2	5	6.313	9	32
NEGATIVO	1	1	1	1.352	1	26

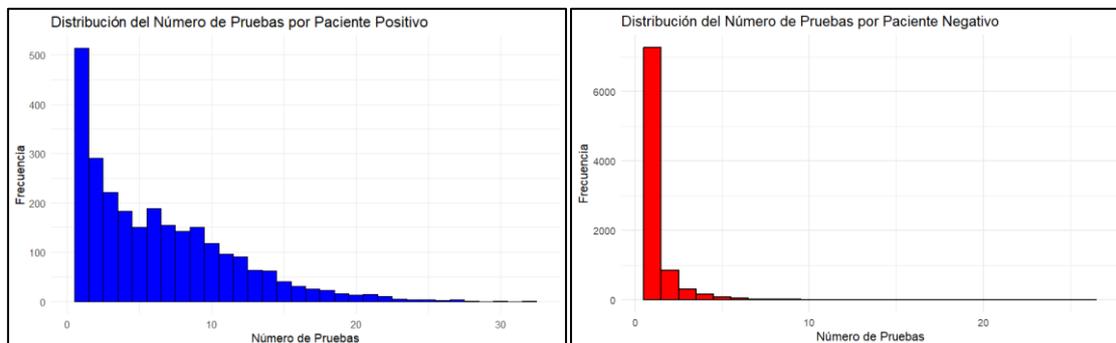


FIGURA 20 & FIGURA 21. Gráfico de barras de la distribución del número de pruebas en pacientes positivos (izquierda) y pacientes negativos (derecha)

Al comparar los dos gráficos (**FIGURA 20 & FIGURA 21**) y analizar la **TABLA 4**, podemos concluir que los pacientes con resultado positivo tienden a realizarse un número mucho mayor de pruebas a lo largo del tiempo proporcionalmente. En cambio, los pacientes con resultado negativo suelen realizarse pocas pruebas, y la mayoría se detienen después de una o dos pruebas.

3.2 TENDENCIAS GENERALES DE LOS PACIENTES POSITIVOS GLOBALES.

En este caso no se van a mostrar las gráficas, ya que son exactamente iguales a las generadas en el apartado 2.2 y 2.3 (Tendencia general del número de positivos a lo largo de los años, y sus estratificaciones por sexo y edad). Esta coincidencia se debe a que gracias a la primera visita se logra diagnosticar a casi la totalidad de los pacientes que resultan positivos, a excepción a los que posteriormente se infectan durante el seguimiento.

4. CLASIFICACIÓN DE PACIENTES 3: Clasificación de los pacientes según los diferentes eventos de la enfermedad.

Esta segregación se centra en la dinámica de los pacientes a lo largo del tiempo en relación al estado de su enfermedad. A continuación, se comentará los resultados obtenidos en cada clasificación:

PACIENTES CURADOS.

	NHC	PRIMERA_PRUEBA	PRIMERA_PRUEBA_POSITIVA	ULTIMA_PRUEBA	resumen_prueba_PRIMERA	resumen_prueba_POSITIVA	resumen_prueba_ULTIMA
1	6/1000156	2014-11-13 14:04:00	2015-04-30 12:54:00	2022-10-04 16:37:00	Negativo	Positivo	Negativo
2	6/1000274	2018-02-09 16:01:00	2018-02-09 16:01:00	2018-07-17 14:37:00	Positivo	Positivo	Negativo
3	6/1000313	2016-09-12 07:35:00	2016-09-12 07:35:00	2018-04-05 08:27:00	Positivo	Positivo	Negativo
4	6/1000533	2022-04-06 12:07:00	2022-04-06 12:07:00	2022-09-15 16:57:00	Positivo	Positivo	Negativo
5	6/1000643	2014-12-04 11:55:00	2014-12-04 11:55:00	2018-01-26 09:14:00	Positivo	Positivo	Negativo
6	6/1001086	2016-05-18 14:04:00	2017-02-16 11:05:00	2020-01-15 14:27:00	Negativo	Positivo	Negativo
7	6/1001462	2015-07-31 10:54:00	2015-07-31 10:54:00	2022-01-21 08:17:00	Positivo	Positivo	Negativo

FIGURA 22. Muestra del dataset que comprende a los pacientes curados. Se muestra de 7 sujetos diferentes, las fechas y resultados claves del proceso de su infección.

En esta clasificación se agruparon 1764 pacientes; de tal forma que se incluyen los pacientes que resultando positivos, al paso del tiempo negativizan; desapareciendo la carga viral en ellos y curándose. Es considerado el grupo más complejo de categorizar ya que se necesitan 3 fechas diferentes, como se observa en la FIGURA 22: la primera prueba que se le realiza, la prueba en el que resultan positivos y finalmente, la prueba en la que pierden la carga viral.

En el ejemplo, podemos identificar los dos diferentes casos que abarca este grupo: Por un lado, existen pacientes que un primer momento son negativos y posteriormente adquieren la infección, aunque terminan curándose, como es el caso del paciente de la fila 1. Sin embargo, en la mayoría de los casos en la primera prueba que se realizan resultan positivos.

Tendencia del número de pacientes curados por año:

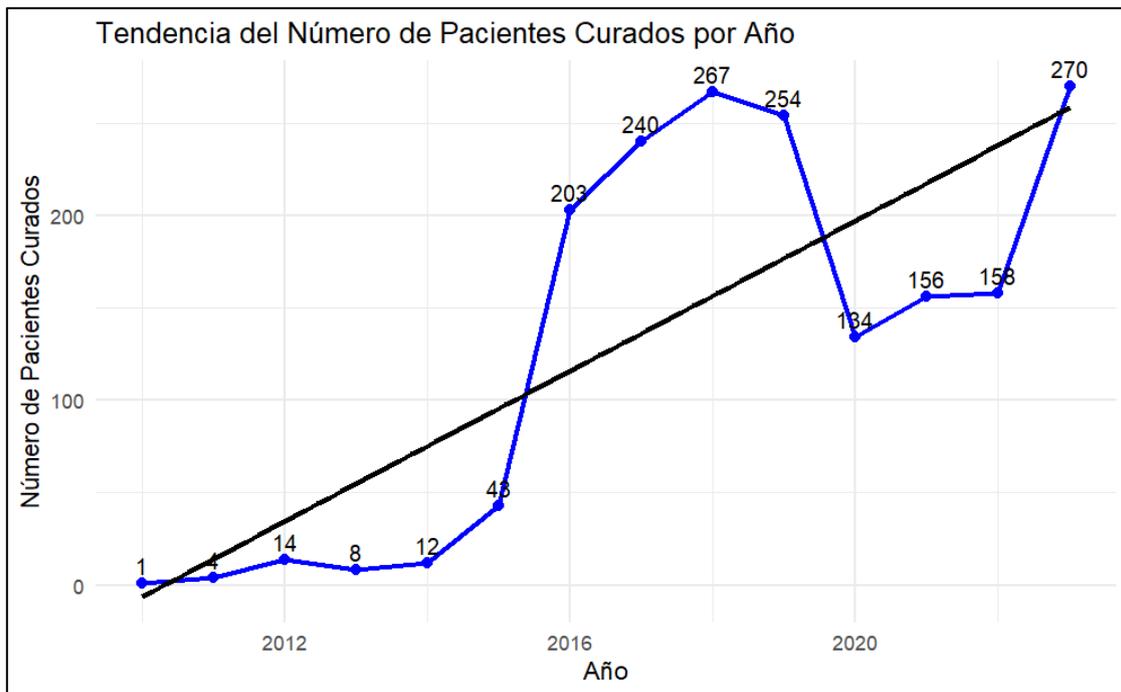


FIGURA 23. Gráfica del número de pacientes curados por año. Se trata de pacientes que pasan de un resultado positivo durante el seguimiento a un resultado negativo definitivo. Se muestran los pacientes según la fecha de su curación: la última prueba y con resultado negativo.

En la **FIGURA 23** contemplamos la tendencia de Negativización de los pacientes Infectados, es decir, cuando se curan o desaparece la carga viral de su infección. En este caso, se muestran los pacientes desde el momento en el que negativizan (última prueba del registro y negativa). Como observamos, los primeros años, del 2010 al 2015, muy pocos individuos se curaban prácticamente. Sin embargo, no es hasta el año 2015 que el número de pacientes curados se disparó positivamente, incluso llegando a cifras de 250 pacientes curados al año. Esta tendencia continuó hasta un pequeño bajón en 2020. Sin embargo, la tendencia continuó al alza de nuevo.

PACIENTES POSITIVOS PERSISTENTES.

	NHC	PRIMERA_PRUEBA	ULTIMA_PRUEBA	resumen_prueba_PRIMERA	resumen_prueba_ULTIMA
1	6/1001514	2015-01-16 14:14:00	2015-01-16 14:14:00	Positivo	Positivo
2	6/1003788	2015-01-20 11:44:00	2015-01-20 11:44:00	Positivo	Positivo
3	6/1004495	2023-06-28 10:13:00	2023-06-28 10:13:00	Positivo	Positivo
4	6/1007312	2015-03-18 13:44:00	2018-02-09 11:04:00	Negativo	Positivo
5	6/1007428	2021-07-02 15:18:00	2023-12-19 18:37:00	Negativo	Positivo
6	6/1007795	2015-05-05 07:34:00	2015-05-05 07:34:00	Positivo	Positivo
7	6/1010174	2015-05-26 09:45:00	2019-10-11 11:27:00	Positivo	Positivo

FIGURA 24. Muestra del dataset que comprende a los Pacientes positivos persistentes. Se muestra de 7 sujetos diferentes, las fechas y resultados claves del proceso de su infección.

En la **FIGURA 24** visualizamos una muestra de la población de los pacientes que continúan siendo positivos hasta el día de hoy. **Identificar a los sujetos que pertenecen a este grupo constituye el objetivo principal de este trabajo.** Tras el proceso de clasificación, se reveló una población de 862 pacientes que aún presentan la infección activa. Tal como se observa en la figura, distinguimos dos situaciones: aquellos que han sido positivos a lo largo de todo el registro y aquellos pacientes que inicialmente se identificaron como negativos pero posteriormente positivizaron. Por consiguiente, esta información será presentada a los profesionales sanitarios para facilitar el inicio del proceso de localización y tratamiento de estos pacientes, contribuyendo a mejorar el control de la enfermedad.

PACIENTES NEGATIVOS PERSISTENTES.

	NHC	PRIMERA_PRUEBA	ULTIMA_PRUEBA	resumen_prueba_PRIMERA	resumen_prueba_ULTIMA
1	6/1000087	2014-11-12 09:55:00	2014-11-12 09:55:00	Negativo	Negativo
2	6/1000156	2014-11-13 14:04:00	2022-10-04 16:37:00	Negativo	Negativo
3	6/1000274	2018-02-09 16:01:00	2018-07-17 14:37:00	Positivo	Negativo
4	6/1000313	2016-09-12 07:35:00	2018-04-05 08:27:00	Positivo	Negativo
5	6/1000330	2023-07-21 07:47:00	2023-07-21 07:47:00	Negativo	Negativo
6	6/1000340	2018-04-05 08:27:00	2018-04-05 08:27:00	Negativo	Negativo
7	6/1000412	2021-01-05 12:47:00	2021-12-02 11:57:00	Negativo	Negativo

FIGURA 25. Muestra del dataset que comprende a los Pacientes negativos persistentes. Se muestra de 7 sujetos, las fechas y resultados claves del proceso de su infección.

En este grupo, la población total identificada asciende a 10542 pacientes que, hasta la fecha, se mantienen con un resultado negativo. Dentro de este grupo se distinguen dos situaciones principales mostradas en la **FIGURA 25**: los pacientes que han sido negativos

desde la primera prueba, y aquellos que inicialmente resultaron positivos pero más tarde lograron curarse, pasando de un estado positivo a negativo (ejem: los pacientes de la fila 3 y 4).

PACIENTES NEGATIVOS QUE POSITIVIZAN: INFECTADOS DURANTE EL SEGUIMIENTO.

	NHC	PRIMERA_FECHA	ULTIMA_FECHA	resumen_prueba_PRIMERA	resumen_prueba_ULTIMA
1	6/1000156	2014-11-13 14:04:00	2017-11-02 17:40:00	Negativo	Positivo
2	6/1001086	2016-05-18 14:04:00	2017-05-25 10:20:00	Negativo	Positivo
3	6/1001462	2016-04-29 10:44:00	2019-02-21 10:07:00	Negativo	Positivo
4	6/1002724	2018-05-30 11:37:00	2021-01-27 10:57:00	Negativo	Positivo
5	6/1006990	2015-03-13 13:24:00	2020-11-06 12:27:00	Negativo	Positivo
6	6/1007312	2015-03-18 13:44:00	2018-02-09 11:04:00	Negativo	Positivo
7	6/1007428	2021-07-02 15:18:00	2023-12-19 18:37:00	Negativo	Positivo

FIGURA 26. Muestra del dataset que comprende a los Pacientes infectados durante el seguimiento. Se muestra de 7 sujetos diferentes, las fechas y resultados claves del proceso de su infección.

Este grupo incluye a los pacientes que se infectaron durante el periodo de seguimiento, es decir, aquellos que pasaron de un estado negativo a positivo. La población total identificada en este grupo es de 317 pacientes. Tal como se observa en la imagen, este cambio en el estado de los pacientes indica la adquisición de la infección en algún momento posterior a su primera prueba.

VALIDACIÓN E INTEGRACIÓN DE LOS RESULTADOS OBTENIDOS.

TABLA 5. Nº DE PACIENTES QUE COMPRENEN LAS DIFERENTES CLASIFICACIONES.	
POBLACIÓN DETERMINADA	Nº DE INDIVIDUOS
TOTAL	11404
CLASIFICACIÓN 1: PACIENTES POSITIVOS 1ª VISITA	2509
CLASIFICACIÓN 1: PACIENTES NEGATIVOS 1ª VISITA	8895
CLASIFICACIÓN 2: PACIENTES TOTALES POSITIVOS	2626*
CLASIFICACIÓN 2: PACIENTES TOTALES NEGATIVOS	8790
CLASIFICACIÓN 3: PACIENTES POSITIVOS PERSISTENTES	862*
CLASIFICACIÓN 3: PACIENTES NEGATIVOS PERSISTENTES	10542
CLASIFICACIÓN 3: PACIENTES CURADOS	1764*
CLASIFICACIÓN 3: PACIENTES INFECTADOS SEGUIMIENTO	317

La **TABLA 5** representa el número de pacientes clasificados en función de su estado respecto a la infección por hepatitis C. Estas relaciones reflejan cómo se distribuye la población en función de su evolución frente a la infección, permitiendo identificar tanto a aquellos que requieren seguimiento y tratamiento continuo, como a quienes que ya no presentan riesgos adicionales de infección en este momento:

Por un lado, la **CLASIFICACIÓN 2** de los pacientes que han sido positivos alguna vez resultó en 2626 pacientes. De estos, de acuerdo con la **CLASIFICACIÓN 3**, 1764 pacientes se curaron mientras que los 862 pacientes restantes forman parte del grupo clasificados como positivos persistentes.

Por otro lado, la diferencia existente entre la población que fue detectada como positivo en la primera visita (2509) (**CLASIFICACIÓN 1**) y los que han sido positivos alguna vez durante todo el seguimiento (2626) (**CLASIFICACIÓN 2**) se debe a los pacientes que se infectaron durante el seguimiento (**CLASIFICACIÓN 3**). Sin embargo, aunque la población que adquirió la infección asciende a 317 pacientes, la diferencia entre estos dos grupos es únicamente 117 (2626-2509). Esto indica que los 200 pacientes restantes (317-117) son aquellos que han estado alternando entre resultados positivos y negativos a lo largo de todo el seguimiento, explicando porque algunos pacientes se les ha realizado más de 30 pruebas; reflejando una evolución fluctuante en su estado viral.

5. ESTUDIO DESCRIPTIVO DE LOS PACIENTES POSITIVOS PERSISTENTES.

En este apartado se realizará un análisis de las características de los pacientes que siguen siendo positivos hasta el día de hoy. Este análisis es crucial, ya que permitirá identificar aquellos pacientes que requieren una revisión individual en su historial clínico para determinar las causas de la persistencia del virus y la necesidad de un seguimiento más cercano:

Distribución de la edad en pacientes positivos persistentes:

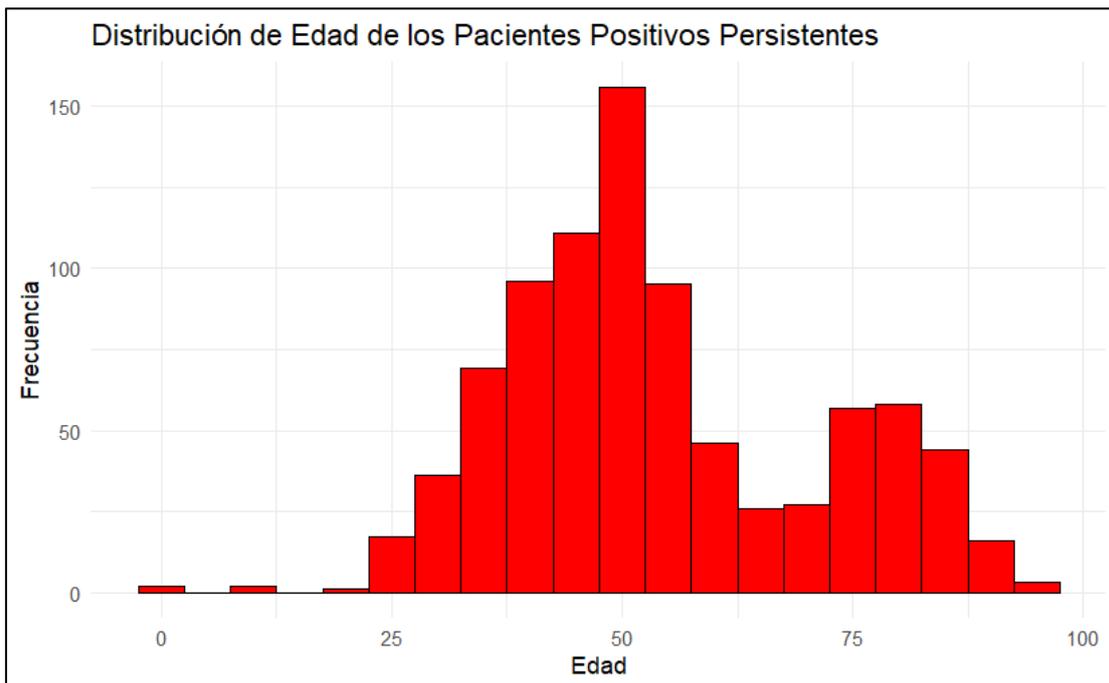


FIGURA 27. Gráfica que muestra el Histograma de la distribución de la edad en la subpoblación de pacientes positivos persistentes.

La **FIGURA 27** muestra la distribución de la edad de los pacientes positivos persistentes para hepatitis C. Podemos observar un pico bien definido alrededor de los 50 años. Por otro lado, también se observan grupos significativos en edades más avanzadas hacia los 70 años.

Distribución del sexo de los pacientes positivos persistentes:

Existe un mayor predominio de hombres en comparación con las clasificaciones anteriores (un 2% más), lo que sugiere que los hombres no solo representan una mayor proporción de los casos totales, sino que también tienen más probabilidades de ser positivos persistentes en el tiempo.

5. DISCUSIÓN

En este trabajo, además de cumplir con el objetivo principal de identificar a los pacientes que presentan una infección no resuelta de hepatitis C y poner dichos documentos en manos de las autoridades sanitarias (un extracto de este archivo lo podemos observar en la **FIGURA 24**), hemos realizado un estudio retrospectivo descriptivo en la población a lo largo de los últimos 13 años. De tal forma que, en el proceso hemos descubierto hallazgos muy reveladores.

Uno de ellos es el hecho de la reducción del número de casos nuevos de hepatitis C con el paso de los años, como podemos observar en la **FIGURA 12**. Este descenso demuestra la gran efectividad de las políticas de intervención médica, tales como la implementación de tratamientos antivirales y el screening masivo de pacientes. Esta tendencia se ha visto en todos los estratos sociales, tanto por edad y sexo como por la combinación de ambos **FIGURA 13 & FIGURA 14 & FIGURA 15**; observándose como disminuye la prevalencia de esta enfermedad año a año. Sin embargo, a pesar de estos avances, se estima que solo el 4% de los pacientes con hepatitis C están diagnosticados, cifra alarmantemente baja (Ridruejo et al., 2016). Esto subraya que, a pesar de los avances significativos en el tratamiento, el mayor desafío sigue siendo aumentar la tasa de diagnóstico. No obstante, se observa como cada año se aumenta el número de pacientes nuevos que se añaden al registro clínico **FIGURA 10**. Asimismo, si nos fijamos bien, la **FIGURA 10** y la **FIGURA 12** muestran una relación inversa entre el número de pacientes nuevos introducidos y la cantidad de pacientes positivos a lo largo de los años. Esto refleja que, aunque en los últimos años más personas se someten a pruebas de diagnóstico o cribado masivo, el porcentaje que resultan positivos ha disminuido considerablemente, claro signo de la disminución de la prevalencia de la enfermedad. Por este motivo, solo a través de un cribado masivo, una detección temprana y un tratamiento oportuno se podrá reducir de manera sustancial el impacto de esta enfermedad en la población, como se ha hecho hasta el día de hoy.

Otra premisa que cabe destacar es la eficacia del cribado inicial (la primera prueba realizada a los sujetos), ya que permite detectar a la gran mayoría de los pacientes positivos en esta primera instancia, salvo aquellos que adquirieron la infección durante

el seguimiento (**TABLA 5**). De ahí la relevancia de la primera prueba diagnóstica y la de la introducción continua de nuevos pacientes en el sistema. Los servicios médicos públicos poseen un sistema de cribado muy eficiente. Sin embargo, es igualmente relevante no solo detectar a los pacientes inicialmente, sino también llevar a cabo un seguimiento adecuado para identificar posibles infecciones posteriores. Como podemos observar en la **FIGURA 5**, la mayoría de los pacientes se realizaron una sola prueba. No obstante, el promedio de pruebas realizadas por pacientes es 2.49. Esto explica porque al comparar los gráficos de la **FIGURA 20** y la **FIGURA 21**, observamos que los pacientes con resultados positivos tienden a realizarse un número mucho mayor de pruebas a lo largo del tiempo proporcionalmente. En cambio, aquellos pacientes con resultados negativos suelen realizarse pocas pruebas, y la mayoría se detienen después de una o dos pruebas. Esto sugiere que, tras un resultado negativo, generalmente no se realiza seguimiento, reflejando una práctica médica donde una vez que se confirma un resultado negativo los pacientes son dados de alta del sistema de vigilancia, mientras que aquellos con resultado positivo se les realiza un monitoreo continuo posiblemente debido a complicaciones en su enfermedad.

Como hemos comprobado varias veces a lo largo del trabajo, esta enfermedad la sufren mayoritariamente personas de mediana edad, de unos 40 a 65 años, siendo 50 años la edad más recurrente **FIGURA 16 & FIGURA 17 & FIGURA 14**. Asimismo, este grupo etario es más susceptible a desarrollar complicaciones graves, como cirrosis o carcinoma hepatocelular; de este modo, la detección temprana permite no solo la prevención de la progresión de la enfermedad a través de un tratamiento adecuado, sino también promueve comportamientos sexuales más responsables, lo que contribuye a evitar nuevas infecciones. Sin embargo, a su vez, este grupo etario es el grupo al que más se ha disminuido el número de casos de HCV con el tiempo **FIGURA 14**. Por otro lado, nuestros resultados revelan que la mayoría de los pacientes con HCV son varones **FIGURA 18 & FIGURA 19**, representando aproximadamente el 64% de los casos totales. Tras los resultados de este análisis, podríamos sugerir en el contexto de la realización una presunta campaña de cribado universal, que sería crucial implementar un cribado centrado principalmente en varones alrededor de los 50 años; dado que esta subpoblación presenta la mayor concentración de casos positivos. De esta forma, se

maximizaría la eficiencia del screening, permitiendo detectar y tratar de manera temprana a aquellos con mayor riesgo de infección activa, es decir, las personas más vulnerables. De igual modo, destaca el mantenimiento en el número de casos positivos en HCV en hombres jóvenes de una franja de edad entre 20 a 39 años a lo largo de los años **FIGURA 15**. Este fenómeno sugiere que los jóvenes adultos varones podrían estar exponiéndose a nuevos factores de riesgo o que se les está detectando la enfermedad en fases más tempranas.

Como destacan muchos estudios, la pandemia de COVID-19 ha tenido un impacto significativo en la detección y reporte de infecciones de transmisión sexual (Guiriguet et al., 2024), incluyendo la hepatitis C. Durante este periodo, muchos países de la UE experimentaron una disminución notable en el número de casos de ETS, en gran parte debido a que la atención médica y los recursos se enfocaron en mayor medida a la crisis pandémica. Este fenómeno también se observa en gran medida en nuestro análisis, donde se registran varios acontecimientos negativos asociados a esta situación extraordinaria: Una reducción sustancial en el número de pacientes nuevos que se añadieron al registro clínico alrededor del 2020 **FIGURA 10**, disminución del número de pruebas para el diagnóstico de la hepatitis C en el 2020 **FIGURA 11** y un descenso muy pronunciado en el número de pacientes que superan la enfermedad (pacientes curados) en los años más críticos de la pandemia **FIGURA 23**. No obstante, observamos una estabilización positiva en todos estos casos una vez se apaciguó la crisis sanitaria; incluso mejorando estos valores respecto a años anteriores a esta.

Este trabajo ha permitido evaluar el impacto de las intervenciones sanitarias aplicadas durante la última década, mostrando sin lugar a dudas la efectividad de los tratamientos antivirales, los cuales han logrado curar a un número considerable de pacientes. En la **FIGURA 23** contemplamos la tendencia de negativización de los pacientes infectados: como observamos, en los primeros años del registro muy pocos individuos se curaban prácticamente ya que las terapias que se aplicaban eran mediante tratamiento con base de Interferón; una intervención muy poco efectiva, general y con muchos efectos secundarios. Sin embargo, no es hasta el año 2015, que se introdujeron en el sistema hospitalario de la CM los antivirales de nueva generación frente a la hepatitis C, los

DAA, que el número de pacientes curados se disparó positivamente al alza, demostrando la gran eficacia de estos. De esta forma el número de pacientes que superan la enfermedad aumenta cada año desde su introducción. Además, se puede observar el gran esfuerzo que llevaron a cabo las instituciones médicas para comprobar la eficacia de los DAAs en primera instancia: se llevaron a cabo grandes labores de monitoreo de pacientes, aumentando considerablemente el número de pruebas realizadas en los primeros años de implementación del tratamiento **FIGURA 11**. Por otro lado, cabe mencionar que hay cierto tipo de pacientes a los que no se les administra este tratamiento: la aplicación en los pacientes más envejecidos (+70 años) no es siempre la mejor opción; ya que, en algunos casos los tratamientos pueden tener ciertos efectos secundarios que podrían perjudicar su salud en lugar de mejorarla **FIGURA 27**. Es por eso que en el archivo que contiene a los pacientes positivos persistentes (el documento que entregamos a los profesionales sanitarios) hay una subpoblación con pacientes en edades bastante avanzadas, a los que posiblemente se reevaluará su estado para la decisión de tratarlos o no.

Uno de los puntos fuertes de nuestro estudio ha sido la duración y la vasta cantidad de pacientes incluidos. Contar con una base de datos longitudinal nos ha permitido observar tendencias a largo plazo y analizar la efectividad de los tratamientos implementados. Sin embargo, nuestra investigación también presenta ciertas limitaciones. El uso de EHRs primarios puede desembocar en la omisión de ciertos diagnósticos relevantes, especialmente aquellos documentados solo en notas clínicas rutinarias. Además, la falta de información sobre ciertos aspectos relevantes como la nacionalidad de los pacientes, sus comportamientos sexuales, el uso de drogas recreativas inyectables, el estadio que presenta la enfermedad en ellos... puede haber limitado nuestra capacidad de interpretar completamente los resultados.

6. CONCLUSIONES

Las conclusiones principales que se pueden extraer de este estudio son las siguientes:

En primer lugar, se ha demostrado que la búsqueda activa de pacientes es una estrategia muy útil para la detección de casos positivos de HCV. Esta metodología ha permitido identificar eficientemente a pacientes infectados, facilitando su posterior tratamiento. En segundo lugar, el perfil de personas con infección activa se ha caracterizado claramente: se trata principalmente de hombres de entre 40 y 60 años. Este hallazgo es relevante para dirigir mejor los esfuerzos de cribado y prevención en este grupo de población. Otro aspecto importante es la disminución de la prevalencia del HCV, atribuida principalmente al uso de los Antivirales de Acción Directa (DAAs). Estos tratamientos han demostrado ser altamente efectivos en la lucha contra la enfermedad. No obstante, a pesar de la disponibilidad de pruebas diagnósticas y tratamientos eficaces, se siguen detectando numerosos nuevos casos de hepatitis C, sobre todo en hombres jóvenes que tienen sexo con hombres (MSM), lo que indica que la transmisión sexual sigue siendo una vía relevante de contagio. Asimismo, la pandemia de COVID-19 ha tenido una influencia considerable en este estudio, lo que indica la importancia de las políticas aplicadas en materia de salud pública en cada momento.

Por último y como aspecto más relevante, se han identificado 862 casos de pacientes con infección activa de HCV que serán revisados por el personal clínico, los cuales se les someterá a un re-test para valorar la necesidad de iniciar un tratamiento con ADDs y que no queden postergados en un archivo clínico. De esta manera, se contribuye a la mejora continua de la salud pública, aportando un pequeño pero importante avance en la lucha contra la hepatitis C **FIGURA 28**.

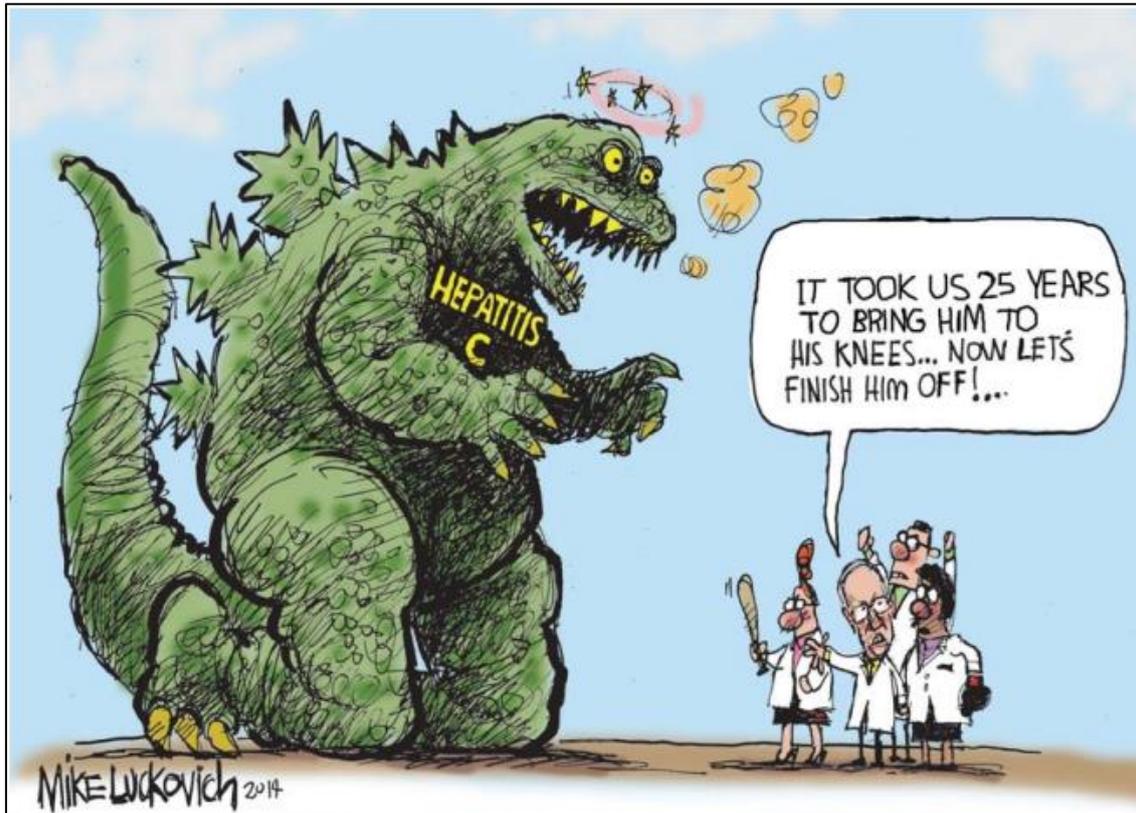


FIGURA 28. Situación contemporánea del virus de la hepatitis C. Extraída de Ward, J.W. et al., 2014. Autor Mike Luckovich. Realizado en el 25º aniversario del descubrimiento del HCV.

BIBLIOGRAFÍA

- Allaire, J., Xie, Y., Dervieux, C., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W., & Iannone, R. (2014). rmarkdown: Dynamic Documents for R (p. 2.28) [Dataset]. <https://doi.org/10.32614/CRAN.package.rmarkdown>
- Asamblea Médica Mundial, 2000; Declaración de Helsinki.
- Chan, D. P. C., Sun, H.-Y., Wong, H. T. H., Lee, S.-S., & Hung, C.-C. (2016). Sexually acquired hepatitis C virus infection: A review. *International Journal of Infectious Diseases*, 49, 47-58. <https://doi.org/10.1016/j.ijid.2016.05.030>
- Evans, R. S. (2016). Electronic Health Records: Then, Now, and in the Future. *Yearbook of Medical Informatics*, 25(S 01), S48-S61. <https://doi.org/10.15265/iys-2016-s006>
- Garinis, G., Spanakis, N., Theodorou, V., Gorgoulis, V., Manolis, E., Karameris, A., & Valis, D. (1999). Comparison of the enzyme-linked immunosorbant assay III, recombinant immunoblot third generation assay, and polymerase chain reaction method in the detection of hepatitis C virus infection in Haemodialysis patients. *Journal of Clinical Laboratory Analysis*, 13(3), 122-125. [https://doi.org/10.1002/\(SICI\)1098-2825\(1999\)13:3<122::AID-JCLA6>3.0.CO;2-A](https://doi.org/10.1002/(SICI)1098-2825(1999)13:3<122::AID-JCLA6>3.0.CO;2-A)
- Guiriguet, C., Alberny, M., Mora, N., Rebagliato, O., Roca, C., Fina, F., Fàbregas, M., Benítez, M., De La Poza, M., Medina, M., Flayeh, S., Pedrazas, D., Sabatés, M., & Coma, E. (2024). Sexually transmitted infections, the epidemic that persists after the COVID-19 pandemic: An analysis of the primary care electronic health records covering about 5 million people in Catalonia. *BMC Primary Care*, 25(1). <https://doi.org/10.1186/s12875-024-02395-4>
- Iryopogu, J. (2007). *Análisis de datos con Power BI, R-RStudio y Knime: Curso práctico*. RA-MA Editorial; eBook. <https://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=edselb&AN=edselb.ELB222665&authtype=sso&custid=s1136447&lang=es&site=eds-live&scope=site&authtype=sso&custid=s1136447>

- Jhaveri, R., John, J., & Rosenman, M. (2021). Electronic Health Record Network Research in Infectious Diseases. *Clinical Therapeutics*, 43(10), 1668-1681. <https://doi.org/10.1016/j.clinthera.2021.09.002>
- Lagging, M., Wejstål, R., Norkrans, G., Karlström, O., Aleman, S., Weiland, O., Castedal, M., Westin, J., & for the Swedish Consensus Group. (2017). Treatment of hepatitis C virus infection: Updated Swedish Guidelines 2016. *Infectious Diseases*, 49(8), 561-575. <https://doi.org/10.1080/23744235.2017.1300682>
- Li, Y., Huang, H., Zabinsky, Z. B., & Liu, S. (2017). Optimizing Implementation of Hepatitis C Birth-Cohort Screening and Treatment Strategies: Model-Based Projections. *MDM Policy & Practice*, 2(1), 238146831668679. <https://doi.org/10.1177/2381468316686795>
- OpenAI 2024; GPT-4 Technical Report: <https://arxiv.org/abs/2303.08774>
- Preciado, M. V., Valva, P., Escobar-Gutierrez, A., Rahal, P., Ruiz-Tovar, K., Yamasaki, L., Vazquez-Chacon, C., Martinez-Guarneros, A., Carpio-Pedroza, J. C., Fonseca-Coronado, S., & Cruz-Rivera, M. (2014). Hepatitis C virus molecular evolution: Transmission, disease progression and antiviral therapy. *World Journal of Gastroenterology*, 20(43), 15992. <https://doi.org/10.3748/wjg.v20.i43.15992>
- Ridruejo, E., Fainboim, H., & Villamil, A. (2016). CRIBADO, DIAGNÓSTICO Y TRATAMIENTO DE LA HEPATITIS POR VIRUS C EN LA PRÁCTICA CLÍNICA.
- Sabanés Bové, D., Mouksassi, S., & Sachs, M. (2023). ggplot2.utils: Selected Utilities Extending «ggplot2» (p. 0.3.2) [Dataset]. <https://doi.org/10.32614/CRAN.package.ggplot2.utils>
- Samper, M. B. (2020). Protección de datos personales: Esquemas (1.a ed.). Dykinson. <https://doi.org/10.2307/j.ctv17hm980>
- Santos, C. M. D. C., Pimenta, C. A. D. M., & Nobre, M. R. C. (2007). The PICO strategy for the research question construction and evidence search. *Revista Latino-Americana de Enfermagem*, 15(3), 508-511. <https://doi.org/10.1590/S0104-11692007000300023>
- Spinu, V., Grolemond, G., & Wickham, H. (2010). lubridate: Make Dealing with Dates a Little Easier (p. 1.9.3) [Dataset]. <https://doi.org/10.32614/CRAN.package.lubridate>

- Vihta, K.-D., Stoesser, N., Llewelyn, M. J., Quan, T. P., Davies, T., Fawcett, N. J., Dunn, L., Jeffery, K., Butler, C. C., Hayward, G., Andersson, M., Morgan, M., Oakley, S., Mason, A., Hopkins, S., Wyllie, D. H., Crook, D. W., Wilcox, M. H., Johnson, A. P., ... Walker, A. S. (2018). Trends over time in Escherichia coli bloodstream infections, urinary tract infections, and antibiotic susceptibilities in Oxfordshire, UK, 1998–2016: A study of electronic health records. *The Lancet Infectious Diseases*, 18(10), 1138-1149. [https://doi.org/10.1016/S1473-3099\(18\)30353-0](https://doi.org/10.1016/S1473-3099(18)30353-0)
- Ward, Hepatitis C virus: The 25-year journey from Discovery to cure.
- Wickham, H., & Bryan, J. (2015a). readxl: Read Excel Files (p. 1.4.3) [Dataset]. <https://doi.org/10.32614/CRAN.package.readxl>
- Wickham, H., & Bryan, J. (2015b). readxl: Read Excel Files (p. 1.4.3) [Dataset]. <https://doi.org/10.32614/CRAN.package.readxl>
- Wickham, H., & Henry, L. (2015). purrr: Functional Programming Tools (p. 1.0.2) [Dataset]. <https://doi.org/10.32614/CRAN.package.purrr>
- Wickham, H., Hester, J., & Bryan, J. (2015). readr: Read Rectangular Text Data (p. 2.1.5) [Dataset]. <https://doi.org/10.32614/CRAN.package.readr>
- Willis, S. J., Cocoros, N. M., Randall, L. M., Ochoa, A. M., Haney, G., Hsu, K. K., DeMaria, A., & Klompas, M. (2019). Electronic Health Record Use in Public Health Infectious Disease Surveillance, USA, 2018–2019. *Current Infectious Disease Reports*, 21(10). <https://doi.org/10.1007/s11908-019-0694-5>
- Zeileis, A., Grothendieck, G., & Ryan, J. A. (2004). zoo: S3 Infrastructure for Regular and Irregular Time Series (Z's Ordered Observations) (p. 1.8-12) [Dataset]. <https://doi.org/10.32614/CRAN.package.zoo>
- Zeisel, M. B., Felmler, D. J., & Baumert, T. F. (s. f.). Hepatitis C virus entry.

ANEXOS

Link del anexo complementario:

Enlace al archivo .PDF:

https://drive.google.com/file/d/1uH_aJYQhocAEILS2_kuAGjImGCVpkURs/view?usp=sharing

Enlace al archivo .HTML:

<https://drive.google.com/file/d/1bNjCz08RpFdBsQSGmA23H0C0I9QMB2w5/view?usp=sharing>

Para quienes deseen revisar el proceso detallado de cómo se generaron las distintas clasificaciones y gráficas, pueden consultar el archivo Anexo a esta memoria. Se trata de un archivo *rmarkdown* donde se documenta el código utilizado y se explica de manera reproducible cada uno de los pasos ejecutados durante el análisis.

DECLARACIÓN OBLIGATORIA DEL USO DE HERRAMIENTAS DE IA

La rúbrica de la DECLARACIÓN DEL USO DEL USO DE HERRAMIENTAS IA se adjunta al trabajo.