



UNIVERSIDAD DE SEVILLA

Facultad de Matemáticas

TRABAJO FIN DE GRADO

ANÁLISIS DE MULTIMORBILIDAD EN PACIENTES PLURIPATOLÓGICOS: BASE POBLACIONAL DE SALUD ANDALUCÍA

Grado en Estadística

Departamento de Análisis Económico y Economía Política

Sevilla, Junio de 2020

Autor: Pablo de la Hera Martínez

Tutora: Myriam Luisa González Limón

Resumen

En este trabajo, realizamos un análisis de multimorbilidad sobre pacientes pluripatológicos para identificar sus características, así como las relaciones entre las patologías que presentan. Nos planteamos dos objetivos principalmente. Por un lado, un análisis descriptivo de las enfermedades crónicas, por grupos de edad y sexo, además de las combinaciones de dichas patologías; por otro lado, un análisis de secuencias a lo largo de la vida de los pacientes, basado en dos estudios. Uno sobre los pacientes que presentan cáncer y otro sobre cinco de las patologías más frecuentes. A fin de alcanzar el objeto de este trabajo, que no es otro que “Obtener un mayor conocimiento de este tipo de pacientes”, el Servicio Andaluz de Salud nos ha proporcionado una muestra de la base de datos poblacional de Andalucía del año 2017, formada mediante la unión de distintos registros médicos electrónicos.

Palabras claves:

Multimorbilidad, registro médicos electrónicos, atención primaria de salud, enfermedades crónicas, prevalencia.

Abstract

In this paper, we performed a Multimorbidity analyses from multipathological patients in order to identify their features, as well as the relationship between the pathologies they have. There are two main objectives. On the one hand, a descriptive analyses of chronic diseases, sorted by age and sex groups, in addition to the combination of the pathologies; on the other hand, a sequencing analyses of the patients throughout life, based on two different studies. The first concerning the patients which cancer has been diagnosed and the second upon the five most frequent pathologies. To achieve our goal, which is “getting the most knowledge as possible from this kind of patients”, the “Servicio Andaluz de Salud” has provided us a sample of the population database from the year 2017, created by joining different set of electronic medical records.

Key words: Multimorbidity, primary health care, electronic medical records, chronic disease, prevalence.

Índice:

1	INTRODUCCIÓN.....	10
2	SERVICIO ANDALUZ DE SALUD.....	13
2.1	Estructura y funciones.....	14
2.2	Organización y servicios.....	14
2.3	Atención primaria.....	15
2.4	Atención hospitalaria.....	15
2.5	Organizaciones específicas.....	15
2.6	Recursos.....	16
3	SUBDIRECCIÓN TÉCNICA ASESORA DE GESTIÓN DE LA INFORMACIÓN.....	16
3.1	Organigrama del Servicio Andaluz de Salud.....	18
4	PREGUNTA DE INVESTIGACIÓN Y OBJETIVOS.....	18
4.1	Objetivo general:.....	18
4.2	Objetivos específicos:.....	18
4.3	Hipótesis del estudio.....	19
4.4	Protección de datos.....	19
5	MATERIAL Y MÉTODO.....	19
5.1	Variables del estudio.....	22
5.2	Aspectos éticos de la investigación.....	23
6	RESULTADOS.....	23
6.1	Análisis de patologías crónicas en pacientes con multimorbilidad.....	23
6.1.1	Características demográficas del estudio.....	26
6.1.2	Combinaciones en pacientes con multimorbilidad.....	30
6.1.3	Características de la actividad sanitaria realizada en atención primaria.....	32
6.1.4	Costes por actividad sanitaria en atención primaria.....	34
6.2	Análisis de secuencias.....	36
6.2.1	Para el Estudio 1:.....	41
6.2.2	Para el Estudio 2:.....	44
7	CONCLUSIONES.....	50
8	BIBLIOGRAFÍA.....	52
9	ANEXO: Dictamen favorable y Estudio en RStudio.....	56

Índice de Tablas

Tabla 1. Costes por Grupo.....	12
Tabla 2. Índice Barthel	24
Tabla 3. Índice Pfeiffer	25
Tabla 4. Patologías frecuentes	25
Tabla 5. Grupos de Edad.....	26
Tabla 6. Análisis descriptivo Edad.....	26
Tabla 7. Media Enfermedades según Grupo de Edad.....	27
Tabla 8. Media Enfermedades según Sexo.....	27
Tabla 9. Número de Enfermedades, por Sexo y Grupo de Edad	28
Tabla 10. Usuarios según Grupo de Edad y Sexo.....	29
Tabla 11. Combinaciones más frecuentes en pacientes con multimorbilidad	31
Tabla 12. Consultas Atención Primaria por Grupo de Edad y Sexo	32
Tabla 13. Tipo de Actividad en Atención Primaria según Proveedor Sanitario	33
Tabla 14. Costes por Sexo según tipo de Actividad Sanitaria en Atención Primaria.....	35
Tabla 15. Costes por Grupo de Edad según tipo de Actividad Sanitaria en Atención Primaria	35
Tabla 16. Patologías de interés	37
Tabla 17. Patologías de interés y Abreviaturas	38
Tabla 18. Patologías_m3 con abreviaturas	38
Tabla 19. Tipos de Cáncer.....	39
Tabla 20. Nº de Cánceres por persona	39
Tabla 21. Patologías seleccionadas	40
Tabla 22. Nº Patologías por persona.....	40
Tabla 23. Alfabeto.....	42
Tabla 24. Tiempo (años) en cada estado	42
Tabla 25. Media del tiempo (años) dedicada a cada tipo de estado	42
Tabla 26. Alfabeto (no completo).....	45
Tabla 27. Tiempo (años) medio en cada estado	45
Tabla 28. Media del tiempo (años) dedicada a cada tipo de evento	46

Índice de Ilustraciones

Ilustración 1. Pirámide Población de Andalucía (2017)	11
Ilustración 2. Agrupación según nº de patologías crónicas	12
Ilustración 3. Distribución según Sexo del paciente	28
Ilustración 4. Distribución de pacientes según número de Enfermedades	29
Ilustración 5. Distribución por Grupo de Edad y Sexo	30
Ilustración 6. Consultas Atención Primaria por Grupo de Edad y Sexo.....	33
Ilustración 7. Tipo de Actividad en AP según Proveedor Sanitario.....	34
Ilustración 8. Costes según Actividad Sanitaria en Atención Primaria	36
Ilustración 9. Análisis de secuencias. Estudio 1	43
Ilustración 10. Análisis de secuencias. Estudio 2	48

1 INTRODUCCIÓN

La coexistencia de múltiples enfermedades crónicas dentro de un individuo es lo que conocemos como multimorbilidad, la aparición de dos o más enfermedades crónicas en un mismo paciente cada vez es más frecuente en nuestra sociedad (Denton & Spencer, 2010; Nicholson, 2017).

La Organización Mundial de la Salud (OMS) define la multimorbilidad como la presencia de dos o más enfermedades crónicas en una misma persona. En una población que envejece acumulando enfermedades crónicas la multimorbilidad está pasando actualmente a ser la norma más que la excepción (World Health Organization. 2008).

La presencia de múltiples enfermedades es muy frecuente en edades avanzadas, y estudios poblacionales han comprobado cómo el número de patologías que presentan los sujetos aumenta con la edad. Se ha descrito que el 24% de los mayores de 65 años y el 31,4% de los mayores de 85 años presentan cuatro o más enfermedades crónicas (Bartke *et al.*, 2001; O Halloran *et al.*, 2004).

El envejecimiento progresivo de la población ocasiona un aumento de la morbilidad y como consecuencia un incremento en el uso de los recursos sanitarios (World Health Organization. 2002; Starfield *et al.*, 2003). Los cambios demográficos y epidemiológicos están provocando un aumento del número de pacientes que presentan múltiples enfermedades crónicas. Conocer la prevalencia y el consumo de recursos sanitarios, ayuda a una mejor gestión y planificación sanitaria (Sánchez Mollá *et al.*, 2017).

Según nos indican Denton & Spencer (2010) el aumento de la esperanza de vida, la mejora en los tratamientos para los procesos agudos, los avances en el diagnóstico y tratamiento de numerosas enfermedades están produciendo una mayor supervivencia de la población con el aumento progresivo de la prevalencia de las enfermedades crónicas.

El concepto de enfermo crónico se ha modificado en las últimas décadas. Actualmente, una representación más real es la del paciente con varias patologías crónicas, que presenta además incapacidad, deterioro en su autonomía y fragilidad clínica. Siguiendo a Alguacil Herrero *et al.* (2012) la enfermedad de base ha dejado de ser lo relevante para incidir más en la importancia de la multimorbilidad y la limitación en las funciones en la vida diaria del paciente.

En este sentido, desde la Consejería de Salud de Andalucía, la respuesta a esta nueva situación surgió en 2004, con la publicación del Proceso Asistencial Integrado Atención al Paciente

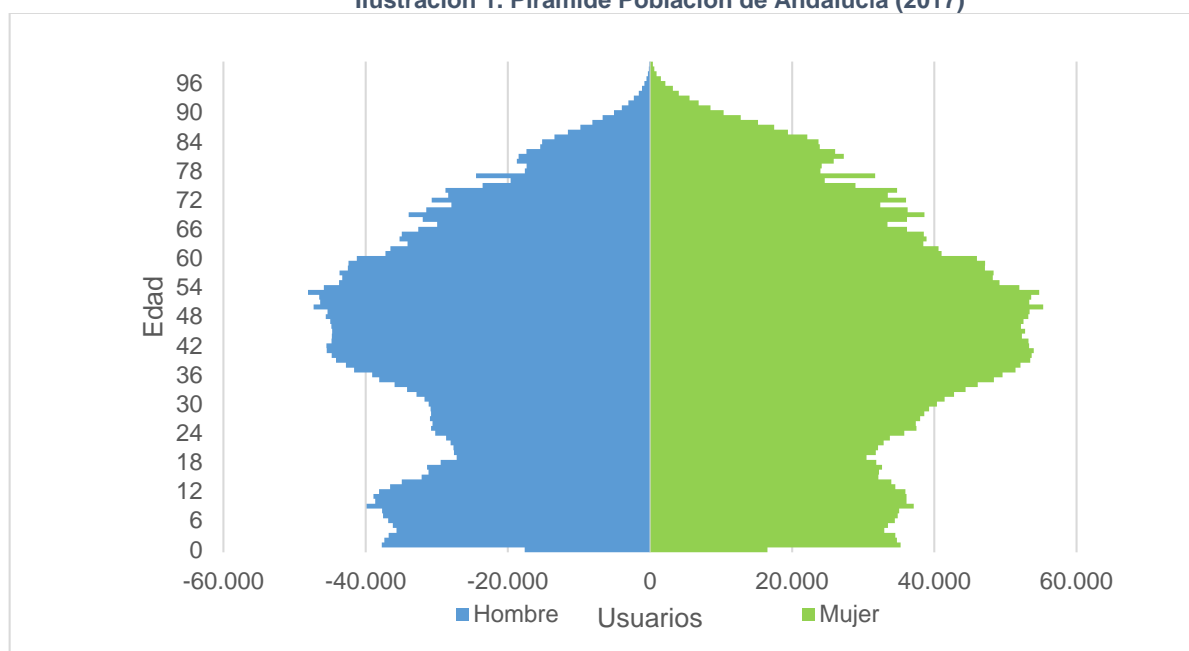
Pluripatológico (PP) y posteriormente con la publicación del Plan Andaluz de Atención Integrada a Pacientes con Enfermedades Crónicas (Alguacil Herrero *et al.*, 2012).

La definición de PP identifica a aquellos individuos con dos o más enfermedades crónicas no curables que generan deterioro progresivo y pérdida gradual de la autonomía funcional, con riesgo de sufrir nuevas comorbilidades y complicaciones, ocasionando todo ello importantes repercusiones sociales y económicas (Bohórquez Colombo *et al.*, 2014).

Actualmente la digitalización de los servicios sanitarios está generando un volumen de datos sin precedentes (Kostkova, 2015; Schlegel & Ficheur, 2017), tanto para la atención sanitaria como para su uso en planificación y gestión de los sistemas de información sanitarios. Estos datos están dispersos, en múltiples bases de datos y sistemas de información diferentes, por lo que su disponibilidad es limitada (Muñoyerro-Muñiz *et al.*, 2019).

Un estudio interno reciente titulado “CuidandodeTi”, realizado por la dirección de enfermedades crónicas del Servicio Andaluz de Salud (SAS), detecta que la población de personas activas en la Base de Datos de Usuarios (BDU, 2017) en Andalucía es de 8.402.590 usuarios (Ilustración 1). Como expresa Lafuente Robles & Rodríguez Gómez (2019), el 84,17% de los usuarios han utilizado los servicios sanitarios durante ese mismo año.

Ilustración 1. Pirámide Población de Andalucía (2017)

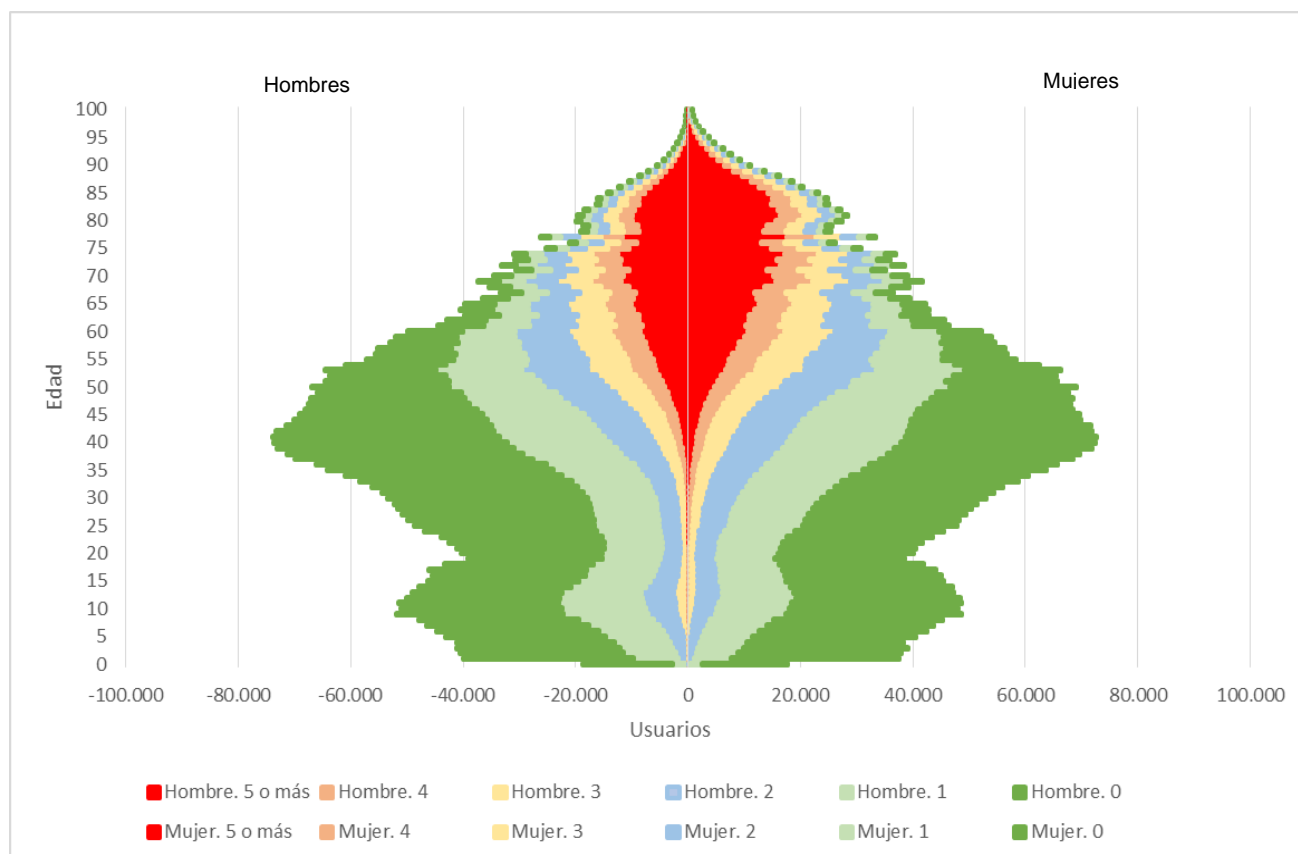


Fuente: Subdirección técnica asesora de Gestión de la Información

En este mismo estudio, encontramos que el 91,84% de las personas que han tenido contacto con los servicios sanitarios tienen codificación diagnóstica registrada, pudiendo obtenerse datos explotables

que nos permitan mostrar una imagen muy fidedigna de las morbilidades que nuestra ciudadanía padece (Ilustración 2) y del uso de recursos que están realizando.

Ilustración 2. Agrupación según nº de patologías crónicas



Fuente: Subdirección técnica asesora de Gestión de la Información

Existen 918.626 personas en Andalucía que tienen 5 o más enfermedades, esto equivale al 11% de la población según la BDU (Ilustración 2). El 5% de la población (374.083) que ha tenido contacto con el Sistema Sanitario Público de Andalucía (SSPA) utiliza el 48% de los recursos anuales del SSPA (3.815.820.874,29 €). Con 74.817 personas (el 1%) utiliza el 21% de estos recursos (1.677.283.312€) (Tabla 1).

Tabla 1. Costes por Grupo

G_GASTO	N	Coste	% Pob	% Coste
No contacto con el Sistema	1.433.992	-		
[0.19€, 306.40€]	3.740.838	480.922.367,68	50	6,04
(306.40€, 656.67€]	1.496.327	678.227.793,07	20	8,52
(656.67€, 1522.83€]	1.122.230	1.109.641.915,55	15	13,94
(1522.83€, 4380.88€]	748.164	1.874.664.665,79	10	23,55
(4380.88€, 12699.20€]	299.266	2.138.537.561,92	4	26,87
(12699.20€, 731125.90€]	74.817	1.677.283.312,37	1	21,07

Fuente: Subdirección técnica asesora de Gestión de la Información

Por todo ello, se necesita una mejor comprensión de la multimorbilidad, puesto que es uno de los problemas más comunes y complejos de la medicina moderna. Dicho problema requiere un enfoque clínico más eficaz.

Hay tres áreas de conocimiento que vamos a tratar:

- 1) La prevalencia y características comunes de multimorbilidad.
- 2) Los patrones que siguen los pacientes con múltiples enfermedades crónicas.
- 3) La progresión natural de los pacientes, así como las variables en el paciente y su análisis de secuencias.

La Base Poblacional de Salud (BPS) es un sistema de información sanitario, que ha permitido la conexión de los diferentes sistemas de información que había en el Servicio Andaluz de Salud (*Servicio Andaluz de Salud. 2018. Resolución 0068/18 Dirección Gerencia del SAS; Davidson, 2015*). Este trabajo se basa en la explotación de una muestra de la BPS, para elaborar información útil que permita una mejora en la gestión en pacientes pluripatológicos y en el uso de los recursos que utilizan (*Starfield et al., 2003*).

A continuación, vamos a describir las características del Servicio Andaluz de Salud, órgano que nos ha facilitado el acceso a la información y datos para realizar el estudio.

2 SERVICIO ANDALUZ DE SALUD¹

El Servicio Andaluz de Salud (SAS), creado en 1986, de acuerdo con lo previsto en el artículo 64 de la ley 2/1998, de 15 de junio, de Salud de Andalucía está adscrito a la Consejería de Salud y Familias y desarrolla las funciones que le están atribuidas bajo la supervisión y control de la misma.

Su misión es prestar atención sanitaria a las ciudadanas y ciudadanos andaluces, ofreciendo servicios sanitarios públicos de calidad, asegurando la accesibilidad, equidad y satisfacción de los usuarios, buscando la eficiencia y el aprovechamiento óptimo de los recursos.

¹ La información se ha obtenido a través de la página web del SAS. <https://www.sspa.juntadeandalucia.es/servicioandaluzdesalud/>

2.1 Estructura y funciones

El Servicio Andaluz de Salud es una agencia administrativa de las previstas en el artículo 65 de la Ley 9/2007, de 22 de octubre, se adscribe a la Consejería de Salud y Familias dependiendo específicamente de la Viceconsejería.

Corresponde al Servicio Andaluz de Salud el ejercicio de las funciones que se especifican en el Decreto 105/2019, de 12 de febrero, por el que se establece la estructura orgánica de la Consejería de Salud y Familias y del Servicio Andaluz de Salud, y, en particular, son las siguientes:

- a. La gestión del conjunto de prestaciones sanitarias en el terreno de la promoción y protección de la salud, prevención de la enfermedad, asistencia sanitaria y rehabilitación que le corresponda en el territorio de la Comunidad Autónoma de Andalucía.
- b. La administración y gestión de las instituciones, centros y servicios sanitarios que actúan bajo su dependencia orgánica y funcional.
- c. La gestión de los recursos humanos, materiales y financieros que se le asignen para el desarrollo de sus funciones.

Al Servicio Andaluz de Salud se adscriben la Empresa Pública de Emergencias Sanitarias, la Agencia Pública Empresarial Sanitaria Costa del Sol, la Agencia Pública Empresarial Sanitaria Hospital de Poniente de Almería, la Agencia Pública Empresarial Sanitaria Hospital Alto Guadalquivir y la Agencia Pública Empresarial Sanitaria Bajo Guadalquivir, sin perjuicio de su dependencia de la Consejería de Salud y Familias. Asimismo, se le adscribe el Consorcio Sanitario Público del Aljarafe (San Juan de Dios). El Servicio Andaluz de Salud cuenta con los siguientes órganos o centros directivos:

- Dirección Gerencia, con rango de Viceconsejería.
- Dirección General de Asistencia Sanitaria y Resultados en Salud.
- Dirección General de Personal.
- Dirección General de Gestión Económica y Servicios.

2.2 Organización y servicios

El Servicio Andaluz de Salud dispone de una red de servicios asistenciales integrada y organizada para garantizar la accesibilidad de la población y la equidad en la distribución de recursos. En su Carta de Servicios se describen con carácter genérico, aquellos servicios asistenciales que recibe directamente la ciudadanía.

2.3 Atención primaria

El primer nivel de atención es la atención primaria de salud, que integra la asistencia preventiva, curativa, rehabilitadora y la promoción de la salud de los ciudadanos. Los servicios de atención primaria de salud están organizados en Andalucía en distritos de atención primaria, estructuras organizativas para la planificación operativa, dirección, gestión y administración en ese ámbito. Existen actualmente 1.518 centros de atención primaria de salud, de manera que todos los andaluces cuentan con uno de ellos a pocos minutos de su domicilio en transporte estándar.

2.4 Atención hospitalaria

El segundo nivel, de atención hospitalaria, atiende a los pacientes que precisan de hospitalización. Dispone de consultas externas ambulatorias en hospitales y en centros periféricos.

El Servicio Andaluz de Salud cuenta con 28 hospitales, distribuidos por toda la geografía andaluza. Asimismo se hace cargo funcionalmente de los centros pertenecientes a Agencias Públicas Empresariales Sanitarias y del Consorcio Sanitario Público del Aljarafe.

2.5 Organizaciones específicas

Además, existen 15 Áreas de Gestión Sanitaria. Se basa en un modelo de organización de gestión unitaria de los niveles de Atención Primaria y de Hospitalaria, en una demarcación territorial específica.

Los Centros de Transfusión Sanguínea configuran una red específica con el fin de asegurar la disponibilidad de componentes sanguíneos y la correcta preservación de tejidos.

Otras áreas de atención tienen una organización especial, dadas sus peculiaridades:

- Programa de Salud Mental.
- Coordinación Autonómica de Trasplantes.
- Atención a Urgencias y Emergencias.

2.6 Recursos

El Servicio Andaluz de Salud cuenta con 100.304 profesionales (media anual) para atender su red asistencial: 25.831 en atención primaria y 74.473 atención hospitalaria. (Año 2018).

El presupuesto definitivo para el ejercicio 2018 es de 9.013 millones de euros.

3 SUBDIRECCIÓN TÉCNICA ASESORA DE GESTIÓN DE LA INFORMACIÓN

La Subdirección Técnica de Gestión de la Información es el centro de gestión de datos del Servicio Andaluz de Salud. Podríamos decir que dicha Subdirección es el *Data Management Center* del SAS.

Durante los últimos años hemos visto como los volúmenes de datos han sido incrementados en las organizaciones. Hoy en día, todos los movimientos que hacemos son registrados por nuestros teléfonos inteligentes, sensores, logs de páginas web, coordenadas de posicionamiento y hábitos de compra. La competencia entre empresas está empezando a enfocarse más en la experiencia de los usuarios que en los productos en sí. Ante este panorama abrumador de volúmenes de datos, distintas fuentes, velocidad y formas surge el *Data Management*. Esta nueva disciplina es también conocida como un proceso de negocios de alto nivel. Consiste en diferentes sub-procesos: planificación y ejecución de políticas, prácticas y proyectos que adquieren, controlan, protegen, entregan y refinan el valor de los activos de datos e información.

Como bien nos indica Lewer *et al.* (2019), mejorando la disponibilidad, seguridad y calidad de información es como debemos dar respuesta a las necesidades de información de todos los sectores de la empresa.

Los objetivos estratégicos de la función de *Data Management* son: entender las necesidades de información de la empresa, capturar, almacenar, proteger y asegurar la integridad de los activos de información (Lewer *et al.*, 2019; Rahman Safwan *et al.*, 2016). Además, se debe mejorar la calidad de los datos e información, incluyendo: exactitud, integridad, integración de la información, tiempo de captura, presentación y definición de los datos, asegurar la confidencialidad y maximizar el valor de los activos de datos.

A cargo del mismo hay un Chief Data Officer (CDO) y un Chief Analytical Officer (CAO). En la disciplina de datos ven al CDO como una persona que se centra en la gestión de datos tácticos, mientras que

el CAO se concentra en el despliegue estratégico de los análisis. El movimiento de un CAO se basa en una función de análisis, no en una función de datos. El papel del CDO está evolucionando progresivamente a CAO. El papel del CAO como estrategia de funciones cruzadas, combinado con el aumento de la dependencia de las organizaciones en los datos para impulsar en las empresas la toma de decisiones, apoya la sugerencia de que el título no es “flor de un día”. El CAO es una extensión natural del análisis; es algo que está incrustado en toda la organización; más se necesita una posición de nivel ejecutivo para transmitir el énfasis estratégico.

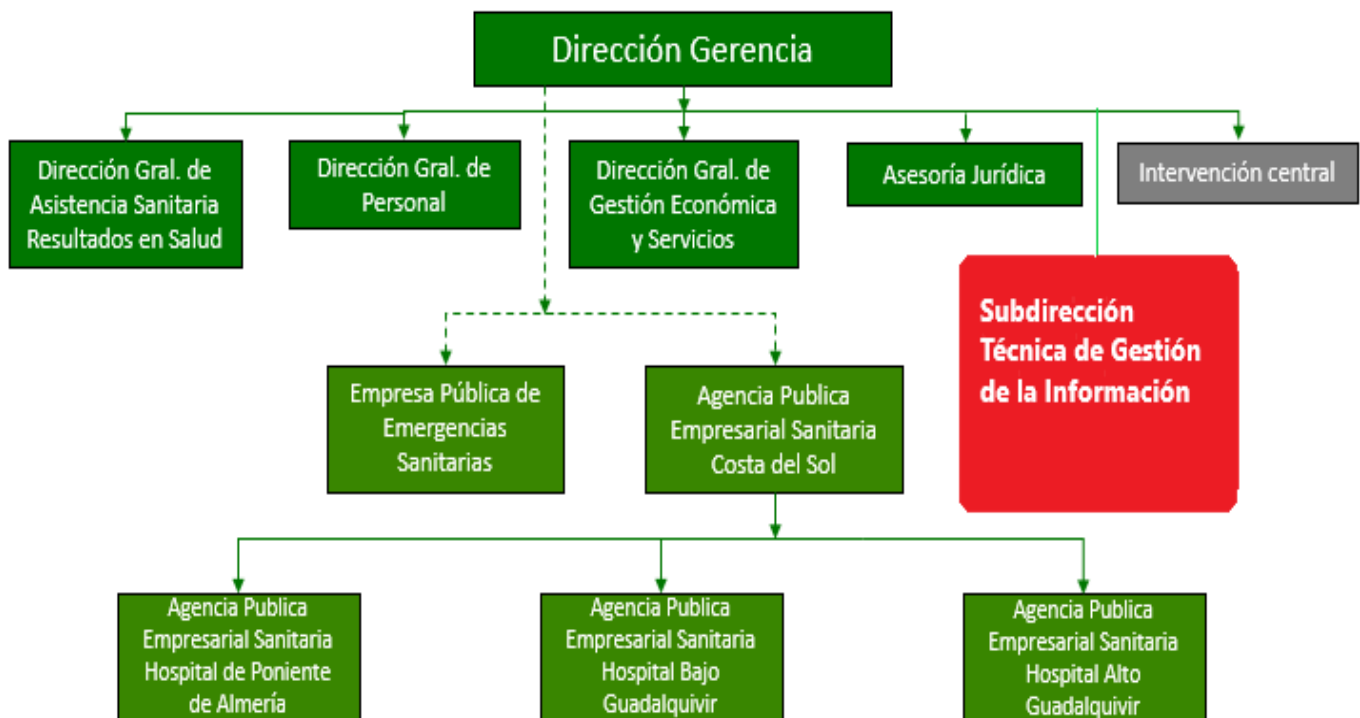
La Subdirección Técnica Asesora de Gestión de la Información posee amplios recursos dentro de su personal de trabajo. Podemos encontrar en el mismo servicio tanto a médicos, como economistas, informáticos, estadísticos y una amplia lista de personal totalmente cualificado para el análisis y la información a nivel sanitario.

En este servicio se manejan distintos sistemas de información sanitarias. Estos sistemas son, entre otros, el Conjunto Mínimo Básico de Datos (CMBD), Aplicación para la Gestión de la Demanda (AGD), Aplicación INFHOS que facilita información on-line de Consultas Externas y Pruebas Diagnósticas, Contrato Programa SAS-Centros, Base de datos Poblacional de Salud, etc.

En especial, nos vamos a centrar en la Base Poblacional de Salud (BPS) de 2017, dónde extraeremos a los pacientes pluripatológicos que vamos a analizar y de los cuales sacaremos las conclusiones sobre nuestro estudio de multimorbilidad.

El SSPA, la historia clínica electrónica y otros sistemas de información, generan datos cuya explotación se realiza en módulos específicos de tratamiento. Esta fragmentación de datos crea duplicidades, propicia su infrautilización y afecta a su calidad y costes de gestión. Esta situación está evolucionando hacia modelos de análisis más globales, gracias a la conexión de registros de los sistemas de información. La integración de datos genera un conocimiento superior al obtenido a partir del análisis aislado e independiente de cada sistema de información. La conexión de registros no es tan habitual como actividad continuada, sistemática e institucional, para el uso secundario de sistemas de datos administrativos. Estas iniciativas institucionales suelen tener como finalidad la evolución de políticas públicas o el desarrollo de infraestructuras para investigación social. El SSPA puso en marcha un proyecto de “Gobernanza de datos”, para superar las dificultades que nos encontramos en la gestión de datos dispersos en los múltiples sistemas de información y contribuir a una mejora de la atención sanitaria, al ayudar a la toma de decisiones en un entorno cada vez más complejo.

3.1 Organigrama del Servicio Andaluz de Salud.



4 PREGUNTA DE INVESTIGACIÓN Y OBJETIVOS

La pregunta de investigación que nos planteamos en este trabajo fin de grado es la siguiente:

PI: ¿Cuáles son los patrones que sigue un paciente que presenta multimorbilidad?

4.1 Objetivo general:

El objetivo general de este estudio es identificar los patrones de multimorbilidad que poseen los pacientes en Andalucía.

4.2 Objetivos específicos:

- Describir el número de enfermedades crónicas por grupos de edad y sexo en los pacientes con multimorbilidad.
- Analizar las combinaciones de enfermedades más frecuentes de los pacientes pluripatológicos.
- Representar los patrones de secuencias en pacientes con múltiples patologías en Andalucía.

4.3 Hipótesis del estudio

Las hipótesis planteadas en este trabajo, son las siguientes:

HI1: ¿Son comunes las características de los pacientes con multimorbiolidad?

HI2: ¿Qué combinaciones de enfermedades son más frecuentes en los pacientes pluripatológicos?

HI3: ¿Qué patrones siguen los pacientes con multimorbiolidad?

4.4 Protección de datos

El derecho a la protección de datos personales persigue garantizar a la persona el control de sus datos personales, su uso y su destino con el propósito de impedir su tratamiento ilícito y lesivo para sus derechos y libertades personales.

En virtud del artículo 12.1, artículo 13 y artículo 14 del Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales, regulada en el ámbito nacional por la Ley Orgánica 3/2018 de Protección de Datos Personales, y la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento General de Protección de Datos), se establece la obligación del responsable (Servicio Andaluz de Salud) de informar al interesado sobre una serie de aspectos en relación al tratamiento de sus datos personales.

Además de la información expresada a continuación, con este mismo fin pero con distinto alcance, están disponibles las siguientes fuentes de consulta:

- Política de privacidad unificada de los sistemas de información online del SAS
- Página web dedicada a la protección de datos personales de la Administración de la Junta de Andalucía.

5 MATERIAL Y MÉTODO

Tal como hemos explicado anteriormente la multimorbiolidad es la coexistencia de dos o más enfermedades dentro de un individuo. En este sentido, para nuestro estudio vamos a usar una base de datos poblacional y longitudinal. Consiste en estudiar y evaluar a las mismas personas por un período prolongado de tiempo (año 2017).

Esta investigación consta de tres pilares relacionados: conexión de registros sanitarios (Base Poblacional de Salud en Andalucía), análisis de datos sobre multimorbilidad y su posterior análisis de secuencias.

El primer pilar se basa en la descripción y el desarrollo de un sistema de información que conecta datos procedentes de múltiples registros sanitarios, para su uso con fines asistenciales de administración, gestión, evaluación, investigación y salud pública (Servicio Andaluz de Salud. 2018. Resolución 0068/18 de la Dirección Gerencia del SAS).

El identificador de cada persona es el Número Único de Historia de Salud de Andalucía (NUHSA) que figura en la Base de Datos de Usuarios (BDU) del SSPA. Los datos fueron extraídos de sistemas operacionales o de ficheros externos, depositados en un servidor y cargados en la Base de Datos Oracle, tras controles de recepción y detección de inconsistencias. En el proceso de transformación, los datos fueron enriquecidos para facilitar su explotación. Estos fueron cargados en un modelo de datos en estrella, modelo dimensional, y tablas agregadas para favorecer su explotación y rendimiento. Los diagnósticos clínicos fueron codificados por documentalistas clínicos para el CMBD.

Se ha utilizado la lista de enfermedades crónicas para su seguimiento, definidas en la Base de Poblacional de Salud. Además, se han establecido los indicadores de prevalencia, expresados como proporciones, utilizando como numerador el número de personas que presentan una condición durante el periodo de estudio y como denominador el total de personas en seguimiento durante el periodo de estudio.

Se han definido 82 patologías crónicas para su seguimiento, se ha utilizado un codificador automático del Servicio Producto Sanitario para los diagnósticos clínicos (Goicoechea-Salazar *et al.*, 2013).

Conjuntamente, a cada paciente se le ha aplicado los costes unitarios de cada servicio recibido, obtenidos del sistema de contabilidad analítica Contabilidad Analítica de Gestión del Servicio Sanitario Público de Andalucía (COAN) del SSPA.

Para recopilar la base de datos proporcionada por el Servicio Andaluz de Salud, tuvimos que realizar en primer lugar, mediante lenguaje SQL, la unión de distintas tablas anidadas entre sí por variables comunes. Las distintas tablas hacían referencia por ejemplo al Número de Historia de cada Paciente (NUHSA), número de patologías de cada paciente, índice barthel, índice pfeiffer, número de consultas de atención primaria de salud, número de consultas de enfermería (en el centro o en el domicilio), número de consultas médicas (en el centro o en el domicilio), número de consultas de pediatría, número de urgencias, número de consultas atención hospitalaria, número de altas

hospitalarias, además del coste de cada servicio utilizado. La fusión y depuración de dichas tablas nos proporciona la base de datos con la que hemos trabajado.

La población consta de 8,5 millones de habitantes, procedentes de la BDU, historia clínica electrónica y sistemas de información de salud mental, pruebas analíticas, vacunas, pacientes renales y farmacia, conectadas de forma determinística. En total hemos extraído una muestra de 30.000 usuarios de dicha base de datos.

El segundo pilar, se realiza un análisis descriptivo de la población de estudio, para conocer las características que presentan según números de enfermedades, edad y sexo, así como las combinaciones más frecuentes de los pacientes con multimorbilidad.

Por último, realizaremos un análisis de los patrones de secuencias en pacientes con múltiples patologías, con dos tipos de estudio, un estudio para los pacientes con cáncer y un segundo estudio para las cinco patologías más frecuentes.

Para la obtención de los resultados hemos utilizado el lenguaje de programación “R” y hemos detallado los pasos a seguir a medida que íbamos analizando los datos en el script (Ver Anexo). Una vez generadas las tablas en “R” se ha mejorado su formato con el programa “Microsoft Office (Excell)”. Las Ilustraciones del análisis descriptivo se han obtenido para facilitar la lectura de los resultados, dichas Ilustraciones se han realizado en “Excell”. En el análisis de secuencias los clúster han sido creados mediante “R”. Las características de atención primaria y los costes por actividad sanitaria, tanto tablas como Ilustraciones, se han trabajado en “Excell”.

En definitiva, las características del estudio se especifican en la siguiente ficha técnica.

FICHA TÉCNICA DEL ESTUDIO

▪ **Ámbito:**

- Andalucía (España).

▪ **Población objetivo:**

- Personas en activo durante el año 2017, con al menos un registro en la Base Poblacional de Salud de Andalucía.

▪ **Tamaño de la muestra:**

- n = 30.000 para el estudio en general.
- n = 3.000 para el análisis de secuencias.

▪ **Enfermedades crónicas:**

- En total, 56 grupos de patologías que hemos seleccionado de las 82 patologías iniciales.

▪ **Grupos de pacientes:**

- Pacientes sin ninguna enfermedad crónica, con una, con dos, con tres, con cuatro y con cinco o más enfermedades crónicas.

▪ **Pacientes elegibles:**

- Todos los pacientes, con sexo y edad conocida.

▪ **Inicio de observación del periodo:**

- Fecha de nacimiento de cada individuo de la muestra.

▪ **Fin de observación del periodo:**

- 31 de diciembre de 2017.

▪ **Diseño muestral:**

- Muestreo aleatorio simple sin reemplazamiento.
-

5.1 Variables del estudio

Las variables que vamos a analizar en nuestro estudio sobre multimorbilidad son las siguientes:

- Conjunto de patologías a estudiar
- Sexo del paciente
- Edad del paciente
- Codificación del paciente según NUHSA
- Número de enfermedades crónicas del paciente
- Tiempo transcurrido en el desarrollo de nuevas enfermedades
- Número de consultas en atención primaria (Médico General, Enfermería y Pediatra)
- Fechas de inicio y duración de las enfermedades, según el paciente
- Costes de los servicios por actividad

5.2 Aspectos éticos de la investigación

Los datos que se van a analizar son de rigurosa confidencialidad y, por lo tanto, es fundamental la no promulgación de dichos datos fuera del análisis. Para ello, se ha tenido en cuenta el compromiso de confidencialidad y se ha tenido constancia de los artículos en referencia a la confidencialidad en la Ley Orgánica 3/2018 de Protección de Datos Personales y garantía de los derechos digitales (LOPD-GDD) y el Reglamento General de Protección de Datos (RGPD).

El protocolo de estudio fue aprobado por el Comité de Ética de la Junta de Andalucía (Ver Anexo).

6 RESULTADOS

6.1 Análisis de patologías crónicas en pacientes con multimorbilidad

Vamos a realizar un estudio estadístico sobre multimorbilidad, es decir, pacientes que sufren más de una enfermedad crónica. Así mismo queremos observar la evolución de cada paciente y sus enfermedades a lo largo de su vida, mediante un análisis de secuencias. Para ello, se nos proporciona una base de datos de pacientes de Andalucía que han sido atendidos y registrados por el Servicio Sanitario Andaluz, a través de sus múltiples registros médicos electrónicos puestos en común.

En primer lugar, cargamos la base de datos que se nos ha proporcionado.

```
load("C:/Users/Administrador/Desktop/TFG/datosBPS.RData")
```

Nuestra base de datos contiene 4 tablas diferenciadas: barthel_m3, pfeiffer_m3, patologías_m3 y usuarios_m3.

Acto seguido cargamos las librerías necesarias para utilizar los paquetes correspondientes en este estudio.

```
library(dplyr)
```

```
library(TraMineR)
```

```
library(ggplot2)
```

```
library(base)
```

Índice de BARTHEL

Vamos a trabajar con la tabla **barthel_m3**, la cual nos indica una muestra de 1.620 pacientes que sufren algún déficit funcional, que les impida ser independientes para realizar las actividades cotidianas, dentro de una escala de rangos, en concreto 5:

##	n_orden	FECHA_BARTHEL	PUNTUACION_BARTHEL	RESULTADO_BARTHEL
## 1:	7819158	20150428	100	Independencia
## 2:	1156919	20140805	90	Dependencia moderada
## 3:	3491536	20170426	100	Independencia
## 4:	4475760	20170504	100	Independencia
## 5:	1282445	20170919	10	Dependencia total
## 6:	5204789	20170119	100	Independencia

La columna que nos indica el grado de índice BARTHEL de cada paciente, la convertimos en factor. Creamos una tabla de frecuencias con los cinco rangos en orden, de mayor a menor número de pacientes en cada grado, con su respectivo porcentaje respecto al total.

Tabla 2. Índice Barthel

	RESULTADO_BARTHEL	N	Porcentaje
1	Independencia	493	30,43
2	Dependencia moderada	460	28,40
3	Dependencia severa	345	21,30
4	Dependencia total	198	12,22
5	Dependencia escasa	124	7,65
	Total	1.620	100

Fuente: Elaboración propia

Se puede observar en esta tabla 2 de resultados Barthel, que el 30% de la muestra (1.620), presenta independencia, frente al 70%, que presenta alguna dependencia y sólo el 12,22%, dependencia total.

Índice PFEIFFER

Por otro lado, hemos trabajado con la tabla **pfeiffer_m3**, que nos desprende una muestra de 1.139 pacientes que sufren algún tipo de demencia (deterioro cognitivo), y a su vez, ver el grado de deterioro que presenta cada uno.

Pasamos la columna de los resultados del test de pfeiffer a factor para poder trabajar con ella. Creamos una tabla de frecuencias con los cuatro grados del test en orden, de mayor a menor número de pacientes en cada grado, con su respectivo porcentaje respecto al total:

Tabla 3. Índice Pfeiffer

	RESULTADO_PFEIFFER	N	Porcentaje
1	Compatible con la normalidad	687	60,32
2	Deterioro intelectual moderado	161	14,14
3	Deterioro intelectual leve	152	13,35
4	Deterioro intelectual severo	139	12,20
	Total	1.139	100

Fuente: Elaboración propia

En la tabla 3 **pfeiffer_m3** se observa que el deterioro intelectual severo lo presentan el 12,20% de los pacientes.

Prevalencia de patologías

A continuación trabajamos con la tabla **patologias_m3**, creamos una tabla de frecuencias, ordenada de mayor a menor en función del número de pacientes que sufren cada patología, con su respectivo porcentaje y la prevalencia de cada una de ellas.

Tabla 4. Patologías frecuentes

	PATOLOGIA_BPS	N	Porcentaje	Prevalencia
1	Trastorno de ansiedad	6.469	10,69	21,56
2	Hipertensión	6.278	10,37	20,93
3	Dislipemia	6.154	10,17	20,51
4	Artrosis - espondilosis	4.444	7,34	14,81
5	Diabetes	2.843	4,70	9,48
6	Asma	2.719	4,49	9,06

Fuente: Elaboración propia

Dónde podemos observar que las patologías más frecuentes en nuestra base de datos son Trastorno de ansiedad (10,69%), Hipertensión (10,37%), Dislipemia (10,17%) y Artrosis (7,34%).

Del total de las muestra de nuestro estudio, la prevalencia de las patologías más frecuentes se presentan en Trastorno de ansiedad con 21,56%, seguida de la Hipertensión con una prevalencia de 20,93% y de la Dislipemia con una prevalencia del 20,51%. La Diabetes y el Asma presentan una prevalencia de 9,48% y de 9,06%, respectivamente. Este dato nos indica que tanto el Trastorno de ansiedad, como la Hipertensión y la Dislipemia se refleja en más de un 20% de los pacientes de nuestra muestra.

Eliminamos Sexo que no sean Hombre o Mujer, de la tabla **usuarios_m3**.

```
usuarios_m4 <- usuarios_m3 %>%
  dplyr::filter(!(SEXO%in%
```

```
c("Desconocido", "Indeterminado",  
"No especificado"))
```

Opcional; si queremos ocupar menos memoria, borramos la tabla **usuarios_m3** inicial y nos quedamos con **usuarios_m4**.

Al objeto de analizar los grupos de edad que presenta la tabla **usuarios_m4** de nuestro estudio, se han creado 7 grupos de edad.

6.1.1 Características demográficas del estudio

Tabla 5. Grupos de Edad

Grupos de Edad	N	Porcentaje
[0,15)	4.630	15,44
[15,45)	11.842	39,49
[45,65)	8.283	27,62
[65,75)	2.829	9,43
[75,85)	1.702	5,68
[85,95)	660	2,20
>95 años	45	0,15
Total	29.991	100

Fuente: Elaboración propia

El análisis descriptivo respecto a la edad de estos 29.991 usuarios se puede observar en la siguiente tabla:

Tabla 6. Análisis descriptivo Edad

Estadísticos	Edades
Mín.	0
1st Qu.	24
Median	42
Mean	41
3rd Qu.	58
Máx.	114
Var	520
Sd	23
Cv	0,55

Fuente: Elaboración propia

La edad media de los pacientes es de 41,47 años, el primer cuartil se posiciona para los individuos de 24 años, y la edad máxima es de 114 años.

En la tabla 7 se recoge el número medio de enfermedades crónicas que presenta cada grupo de edad.

Tabla 7. Media Enfermedades según Grupo de Edad

Grupos de Edad	N	Porcentaje	Media de Enfermedades
[0,15)	4.630	15,44	0,6507559
[15,45)	11.842	39,49	0,9754264
[45,65)	8.283	27,62	2,3602559
[65,75)	2.829	9,43	4,4393779
[75,85)	1.702	5,68	5,7497062
[85,95)	660	2,20	5,8833333
>95 años	45	0,15	3,6222222
Total	29.991		

Fuente: Elaboración propia

Podemos observar que a medida que van cumpliendo años nuestros pacientes, presentan más enfermedades crónicas, salvo los de +95 años por la escasez de casos.

Los pacientes en los grupos de edad de 75 a 95 años presentan una media de más de 5 enfermedades crónicas.

Otro de los análisis realizados ha sido la distribución según el sexo y la media de enfermedades presentadas según sexo. Se han excluido 9 pacientes que presentaban la variable sexo indeterminado.

Tabla 8. Media Enfermedades según Sexo

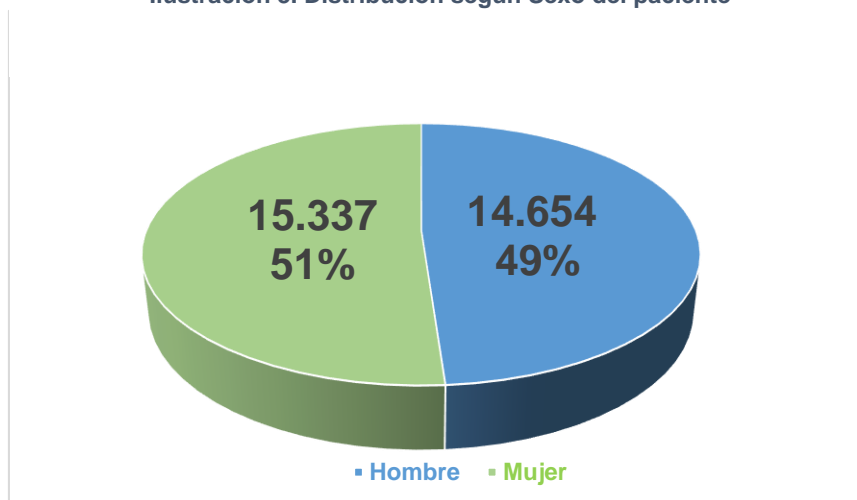
Sexo	N	Porcentaje	Media de Enfermedades
Hombre	14.654	48,86	1,818343
Mujer	15.337	51,14	2,207668
Total	29.991		

Fuente: Elaboración propia

En la tabla 8 se aprecia que las mujeres poseen un promedio de más de dos enfermedades crónicas por persona, en cambio los hombres no llegan a dos enfermedades (1,82).

El 51,14% son mujeres frente al 48,86% son hombres (Ilustración 3).

Ilustración 3. Distribución según Sexo del paciente



Fuente: Elaboración propia

Tabla 9. Número de Enfermedades, por Sexo y Grupo de Edad

Variables	Ninguna_Enf	Una_Enf	Dos_Enf	Tres_Enf	Cuatro_Enf	Cincomás_Enf
Número de pacientes	10.585 (35,29%)	6.576 (21,93%)	4.011 (13,37%)	2.645 (8,82%)	1.885 (6,29%)	4.292 (14,31%)
Sexo	Ninguna_Enf	Una_Enf	Dos_Enf	Tres_Enf	Cuatro_Enf	Cincomás_Enf
Hombres	5.653 (18,85%)	3.324 (11,08%)	1.869 (6,23%)	1.189 (3,96%)	815 (2,72%)	1.804 (6,02%)
Mujeres	4.932 (16,44%)	3.251 (10,84%)	2.142 (7,14%)	1.454 (4,85%)	1.070 (3,57%)	2.488 (8,3%)
Grupos de Edad	Ninguna_Enf	Una_Enf	Dos_Enf	Tres_Enf	Cuatro_Enf	Cinco o más_Enf
[0,15)	2.656 (8,86%)	1.251 (4,17%)	499 (1,66%)	158 (0,53%)	49 (0,16%)	17 (0,06%)
[15,45)	5.630 (18,77%)	3.259 (10,87%)	1.661 (5,54%)	691 (2,3%)	332 (1,11%)	269 (0,9%)
[45,65)	1.979 (3,8%)	1.693 (5,65%)	1.368 (4,56%)	1.140 (3,8%)	785 (2,62%)	1.318 (4,39%)
[65,75)	211 (0,7%)	262 (0,87%)	325 (1,08%)	418 (1,39%)	393 (1,31%)	1.220 (4,07%)
[75,85)	68 (0,23%)	80 (0,27%)	110 (0,37%)	160 (0,53%)	242 (0,81%)	1.042 (3,47%)
[85,95)	28 (0,09%)	28 (0,09%)	44 (0,15%)	73 (0,24%)	76 (0,25%)	411 (1,37%)
+95 años	13 (0,04%)	2 (0,01%)	4 (0,01%)	3 (0,01%)	8 (0,03%)	15 (0,05%)

Fuente: Elaboración propia

En esta tabla 9 se aprecia cómo la mayoría de nuestros pacientes tienen 0 o 1 enfermedad crónica, no obstante, más de un 14% presenta 5 enfermedades o más.

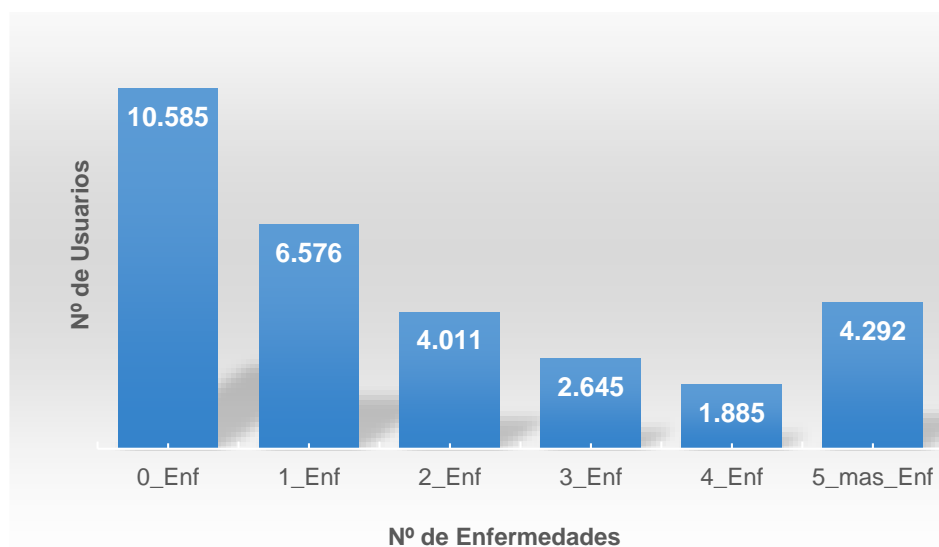
En cuanto al sexo, hay más hombres que presentan 0 o 1 enfermedad (18,85% y 11%), a partir de 2 enfermedades crónicas los porcentajes son mayores en las mujeres donde predomina tener 2, 3, 4 y 5 o más enfermedades crónicas, este último grupo (5 o más enfer.) representa el 8,3% en las mujeres.

En los grupos de edad podemos observar que lo normal en niños de 0 a 15 años es no presentar ninguna patología crónica. Entre los pacientes de 15 a 45 años engloba un alto porcentaje con 1 o ninguna enfermedad crónica. De los 45 a los 65 años está muy repartido el número de enfermedades

crónicas de cada paciente. A partir, de los 65 años, la mayoría de los pacientes registrados en la BPS presentan 5 enfermedades o más.

A continuación, vamos a realizar la representación gráfica (Ilustración 4) de las variables estudiadas en la tabla anterior.

Ilustración 4. Distribución de pacientes según número de Enfermedades



Fuente: Elaboración propia

Hemos creado una tabla de frecuencias respecto al sexo de los pacientes y su grupo de edad correspondiente, incluidos en la base de datos poblacional, diferenciando entre hombres y mujeres y los 7 grupos de edad de cada uno, cuyo porcentaje sería respecto al total de la población a estudiar (n= 29.991).

Tabla 10. Usuarios según Grupo de Edad y Sexo

Grupos de Edad	HOMBRES		MUJERES	
	N	Porcentaje	N	Porcentaje
[0,15)	2.365	16,14	2.265	14,77
[15,45)	5.934	40,49	5.908	38,52
[45,65)	4.073	27,79	4.210	27,45
[65,75)	1.316	8,98	1.513	9,87
[75,85)	703	4,80	999	6,51
[85,95)	249	1,70	411	2,68
>95 años	14	0,10	31	0,20
TOTAL	14.654	100	15.337	100

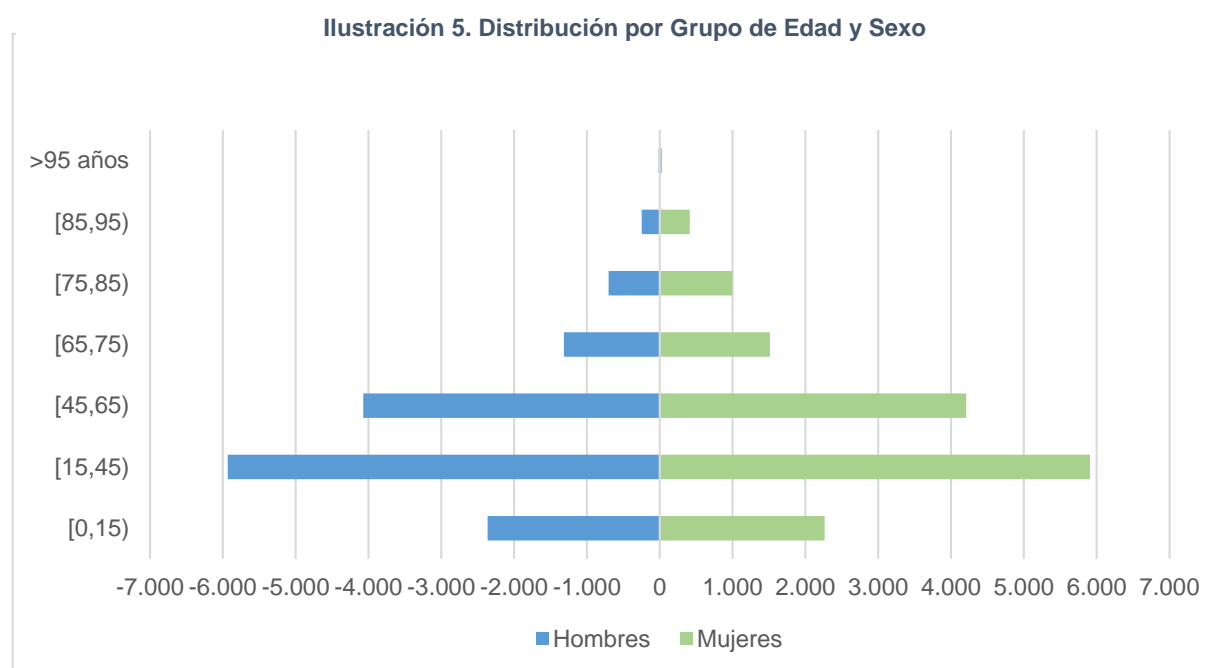
Fuente: Elaboración propia

Se puede observar como la mayoría de nuestros usuarios, tanto hombres como mujeres, se encuentran entre los 15 y 65 años de edad, representando un total de 67,11% entre hombres y mujeres.

En la población masculina el grupo de edad con mayor frecuencia es el comprendido en el intervalo de 15 a 45 años, con más del 40%.

En la población femenina se observa que el grupo de 15 a 45 años es el más representativo con 38,52%.

Se puede decir que el comportamiento de nuestra muestra, tanto en hombres como en mujeres, es similar (Ilustración 5).



Fuente: Elaboración propia

6.1.2 Combinaciones en pacientes con multimorbilidad

Entendiendo por multimorbilidad la aparición de 2 enfermedades o más en el paciente, existen 12.833 usuarios de los 29.991 de nuestra base de datos que presentan multimorbilidad. Un 42,79% de los pacientes la presentan. A continuación, estudiamos las combinaciones de patologías más frecuentes en nuestro análisis.

Vamos a analizar las combinaciones que presentan los pacientes con multimorbilidad para 2, 3, 4 y hasta 5 combinaciones.

Tabla 11. Combinaciones más frecuentes en pacientes con multimorbilidad

N° de Enfermedades	Patologías Unidas	N	%
2 (N = 4.011)	Dislipemia e Hipertensión	191	4,76
	Dislipemia y Trastorno de ansiedad	154	3,84
	Asma y Dermatitis atópica	141	3,52
	Litiasis urinaria y Trastorno de ansiedad	137	3,42
	Asma y Trastorno de ansiedad	129	3,22
3 (N = 2.645)	Artrosis-espondilosis , Dislipemia e Hipertensión	89	3,37
	Diabetes , Dislipemia e Hipertensión	61	2,31
	Dislipemia , Hipertensión y Trastorno de ansiedad	61	2,31
	Artrosis-espondilosis , Dislipemia y Trastorno de ansiedad	34	1,29
	Dislipemia , Hipotiroidismo y Trastorno de ansiedad	31	1,17
4 (N = 1.885)	Artrosis-espondilosis , Dislipemia , Hipertensión y Trastorno de ansiedad	33	1,75
	Artrosis-espondilosis , Diabetes , Dislipemia e Hipertensión	26	1,38
	Diabetes , Dislipemia , Hipertensión y Obesidad	17	0,90
	Artrosis_ espondilosis , Dislipemia , Hipertensión y Obesidad	14	0,74
	Artrosis_ espondilosis , Dislipemia , Hipertensión y Osteoporosis	14	0,74
5 (N = 1.294)	Artrosis-espondilosis , Dislipemia , Hipertensión , Litiasis urinaria y Trastorno de ansiedad	10	0,77
	Artrosis-espondilosis , Diabetes , Dislipemia , Hipertensión y Trastorno de ansiedad	8	0,62
	Artrosis-espondilosis , Dislipemia , Hipertensión , Trastorno de ansiedad y Trastorno estado animo	8	0,62
	Artrosis-espondilosis , Diabetes , Dislipemia , Glaucoma e Hipertensión	6	0,46
	Artrosis-espondilosis , Dislipemia , Hipertensión , Osteoporosis y Trastorno de ansiedad	6	0,46

Fuente: Elaboración propia

En la tabla 11 se aprecia que, los usuarios con dos enfermedades representan el 31,25% de las personas con multimorbilidad. Dentro de los pacientes con dos enfermedades, encontramos que Dislipemia e Hipertensión es la pareja más frecuente. La presentan 191 pacientes de 4.011, un 4,76%.

Los pacientes con tres enfermedades representan algo más del 20% de las personas con multimorbilidad. El trío de enfermedades más frecuente es Artrosis, Dislipemia e Hipertensión, que representa un 3,37% del total de combinaciones de tres.

Por otro lado, los pacientes con 4 y 5 enfermedades equivalen casi al 25% de los usuarios con multimorbilidad. La combinación de cuatro enfermedades más frecuentes es Artrosis, Dislipemia, Hipertensión y Trastorno de ansiedad, la cual la presentan 33 personas (1,75%). Del mismo modo, las combinaciones de 5 enfermedades la presentan un total de 1.294 pacientes, de ellas las más

frecuentes son para el grupo de Artrosis, Dislipemia, Hipertensión, Litiasis urinaria y Trastorno de ansiedad.

Por último, se puede concluir que estos grupos son combinaciones de las cuatro enfermedades más frecuentes.

6.1.3 Características de la actividad sanitaria realizada en atención primaria

En **usuarios_m4** podemos observar distintas variables de atención primaria, las cuales vamos a analizar a continuación.

En la tabla 12 se representa el número de consulta de atención primaria que han realizado por edad y sexo:

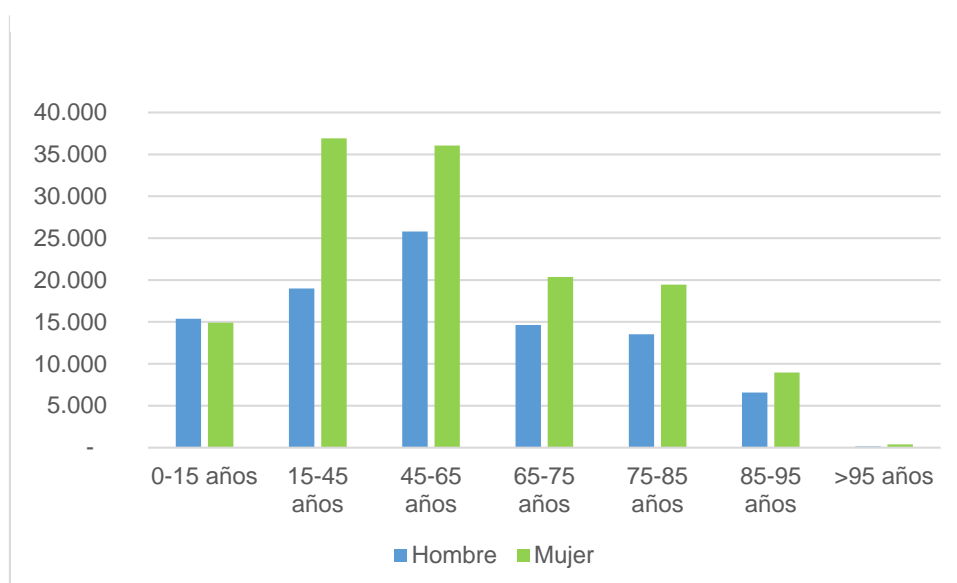
Tabla 12. Consultas Atención Primaria por Grupo de Edad y Sexo

Grupos de Edad	Hombre	Mujer	Total general	%
0-15 años	15.374	14.894	30.268	13,04
15-45 años	18.998	36.919	55.917	24,09
45-65 años	25.789	36.044	61.833	26,64
65-75 años	14.642	20.370	35.012	15,09
75-85 años	13.537	19.457	32.994	14,22
85-95 años	6.574	8.962	15.536	6,69
>95 años	135	391	526	0,23
Total general	95.049	137.037	232.086	100
%	40,95	59,05	100	

Fuente: Elaboración propia

El 59% de las consultas de atención primaria están realizadas por mujeres y el grupo de edad que más demanda es el de 45-65 años, con 26,64%.

Ilustración 6. Consultas Atención Primaria por Grupo de Edad y Sexo



Fuente: Elaboración propia

En todos los grupos de edad se manifiesta que es la mujer la que acude con más frecuencias a las consultas de atención primaria, excepto en la edad pediátrica, que presentan valores muy similares entorno a las 15.000 consultas en ambos sexos.

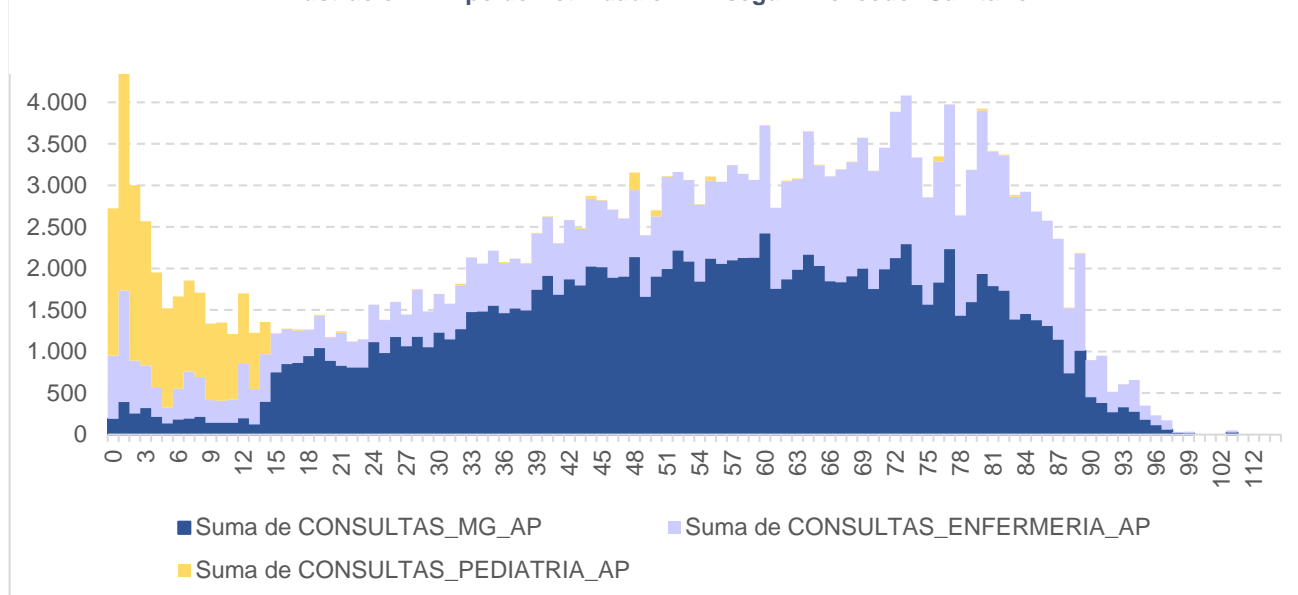
Por otro lado, sí diferenciamos el tipo de consulta que utilizan los usuarios según proveedor sanitario, ya sea consulta de medicina general, consulta de enfermería o consulta de pediatría.

Tabla 13. Tipo de Actividad en Atención Primaria según Proveedor Sanitario

Grupos de Edad	Suma de CONSULTAS_MG_A P	Suma de CONSULTAS_ENFERMERIA_AP	Suma de CONSULTAS_PEDIATRIA_AP	% MG	% Enferm	%Pediatría
0-15 años	3.219	7.700	18.752	2,56	9,52	95,92
15-45 años	37.970	15.263	213	30,20	18,86	1,09
45-65 años	40.341	19.616	402	32,09	24,24	2,06
65-75 años	19.566	14.750	31	15,56	18,23	0,16
75-85 años	16.939	15.452	135	13,47	19,10	0,69
85-95 años	7.452	7.842	17	5,93	9,69	0,09
>95 años	229	293	0	0,18	0,36	0,00
Total general	125.716	80.916	19.550	100	100	100

Fuente: Elaboración propia

Ilustración 7. Tipo de Actividad en AP según Proveedor Sanitario



Fuente: Elaboración propia

En el Ilustración 7 se puede observar, como los usuarios con edades comprendidas entre los 15-65 años tienen un mayor consumo de consultas de medicina general (62,29%), entre los 45-65 años las consultas de enfermería presentan una demanda mayor (24,24%). Cabe destacar que las personas entre 75-85 años ocupan casi un 20% de las consultas de enfermería, ya que son más longevos. En contraposición se observa que de 0 a 15 años predominan las consultas de pediatría (95,92%), dado que es la edad pediátrica de la población.

6.1.4 Costes por actividad sanitaria en atención primaria

Los altos costes que presenta el sistema sanitario han hecho que diferentes disciplinas formen parte de la planificación sanitaria. En este sentido, la OMS, en su 50 aniversario, recogió como una preocupación creciente la complejidad y altos costes de los servicios de salud, reconociendo el reto de perseguir la equidad en el acceso a éstos en un mundo cada vez más rico, pero con un mayor número de pobres, y en el que se cuestiona la financiación pública, a pesar de que es la única que asegura la sostenibilidad social y la disminución de las desigualdades (Martínez Castells, 2014). Con este fin, la economía de la salud es una ciencia que permite el análisis y la toma de decisiones a los gestores sanitarios.

Tal como expresa Ortún-Rubio *et al.* (2001), la economía de la salud en estos últimos 20 años, ha tenido una gran aplicación en la evaluación de las políticas sanitarias, siendo esta una gran herramienta en la planificación y gestión sanitaria.

A continuación vamos a analizar los costes de la actividad sanitaria en atención primaria, como uno de los pilares del gasto de los servicios sanitarios.

Se nos ha facilitado los costes finalistas de cada servicio recibido, obtenido del sistema de contabilidad analítica (COAN SSPA. 2017). En nuestra muestra sólo hemos tenido acceso a los costes de atención primaria.

Tabla 14. Costes por Sexo según tipo de Actividad Sanitaria en Atención Primaria

Sexo	N	COSTE_CONSULTAS_AP	COSTE_URGENCIAS_AP	COSTE_FARMACIAS_AP	COSTE_VACUNAS_AP	COSTE_TOTAL_AP
Hombre	14.654	1.902.880,98	209.609,40	2.480.701,24	45.812,71	4.639.004,33
Mujer	15.337	2.743.480,74	254.794,54	3.024.105,61	67.813,60	6.090.194,49
Total	29.991	4.646.361,72	464.403,94	5.504.806,85	113.626,31	10.729.198,82
% Coste AP		43,31	4,33	51,31	1,06	100

Fuente: Elaboración propia

En atención primaria, de los 10,7 millones de gastos imputados en nuestra muestra, el 51,31% de los costes son debidos a la farmacia; es decir a los fármacos que se prescriben a los usuarios, seguidos del 43,31% que corresponden a las consultas de atención primaria que engloban las consultas de medicina general, de enfermería y consultas de pediatría.

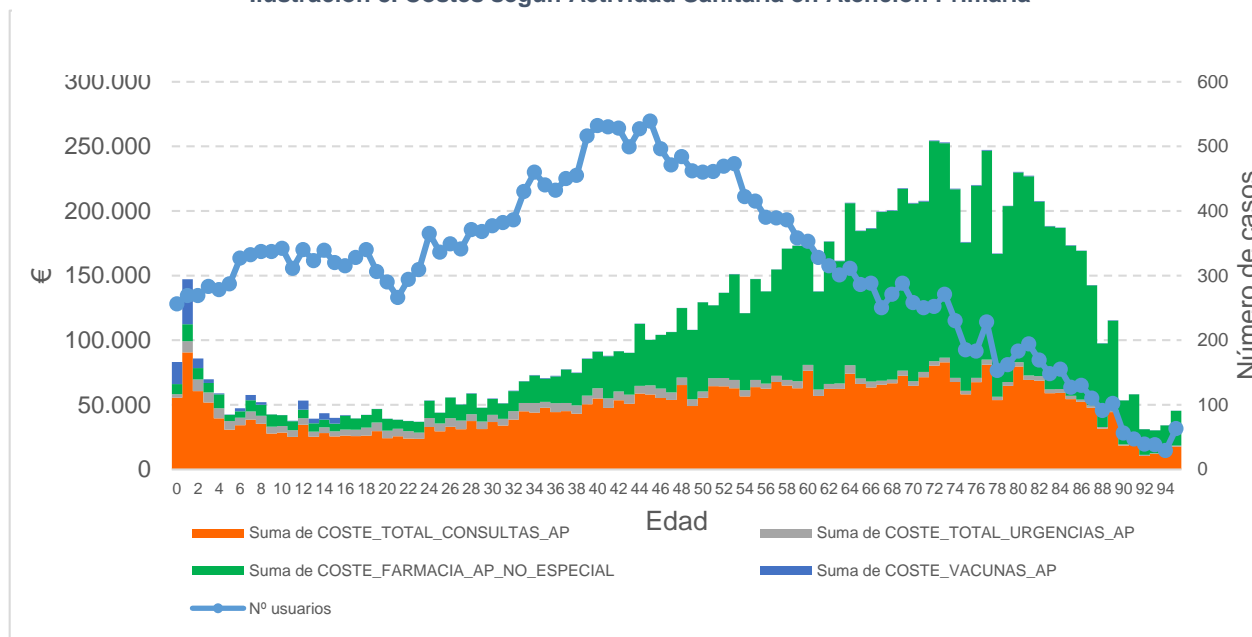
Tabla 15. Costes por Grupo de Edad según tipo de Actividad Sanitaria en Atención Primaria

Grupos de Edad	N	Suma de COSTE_TOTAL_CONSULTAS_AP	Suma de COSTE_TOTAL_URGENCIAS_AP	Suma de COSTE_TOTAL_FARMACIAS_AP	Suma de COSTE_TOTAL_VACUNAS_AP	Total AP	Coste por Persona	% Consultas AP	% Farmacia	% Total AP
0-15 años	4.630	605.965,36	90.049,96	118.901,49	87.659,97	902.576,78	194,94	13,04	2,16	8,41
15-45 años	11.842	1.119.458,34	183.703,52	522.076,84	11.270,56	1.836.509,26	155,08	24,09	9,48	17,12
45-65 años	8.283	1.237.896,66	108.408,30	1.505.095,22	3.659,73	2.855.059,91	344,69	26,64	27,34	26,61
65-75 años	2.644	700.940,24	36.916,88	1.385.433,00	5.054,41	2.128.344,53	804,97	15,09	25,17	19,84
75-85 años	1.760	660.539,88	32.052,02	1.358.464,17	4.174,67	2.055.230,74	1.167,74	14,22	24,68	19,16
85-95 y más años	832	321.561,24	13.273,26	614.836,13	1.806,97	951.477,60	1.143,60	6,92	11,17	8,87
Total	29.991	4.646.361,72	464.403,94	5.504.806,85	113.626,31	10.729.199	357,75	100	100	100
%		43,31	4,33	51,31	1,06	100				

Fuente: Elaboración propia

La tabla 15 nos proporciona los datos necesarios para calcular el gasto por persona en atención primaria según grupos de edad. Dividiendo el coste total por el número de personas de cada grupo de edad. El coste por persona de 0-15 años y de 15-45 años es de 195€ y 155€ respectivamente. Lo cual es lógico, ya que presentan menos patologías en estas edades. A partir de los 75 años el coste por persona asciende a más de 1.000€, ya que en estas edades las personas presentan más patologías.

Ilustración 8. Costes según Actividad Sanitaria en Atención Primaria



Fuente: Elaboración propia

Como se puede apreciar en el Ilustración 8, el consumo de costes según la actividad realizada en atención primaria es mayor a medida que aumenta la edad del paciente.

En este sentido, el coste en Farmacia, es decir los medicamentos dispensados a los pacientes es el más elevado, representando el 51,31% del total del coste de atención primaria. El coste de fármacos de los pacientes menores de 45 años representa sólo el 11,64% de total de gasto en farmacia, frente al 88,36% que consumen los pacientes de más de 45 años.

6.2 Análisis de secuencias

En este estudio queremos evaluar las enfermedades que han ido teniendo los distintos usuarios a lo largo de su vida mediante un análisis de secuencias. Para ello tendremos que preparar la base de datos y modificarla de tal manera que las variables sean las edades de cada individuo y podamos comprobar las patologías que iban apareciendo, año a año, en cada uno de ellos. Vamos a realizar dos estudios, separados en 2 fases, el primero (Estudio 1) sobre cánceres y el segundo (Estudio 2)

sobre 5 patologías de interés. Para realizar el Estudio 2, debemos ejecutar el script desde el principio (ignorando la parte del Estudio 1), para obtener los resultados de manera óptima. Si lo ejecutamos todo de corrido solo obtendríamos los resultados del primer estudio (Estudio 1).

Comenzamos creando la variable “Estudio” y añadiéndola a la tabla 4, para posteriormente, poder eliminar las patologías que no van a ser de interés, de la tabla patologias_m3, en nuestra base de datos. Para ello, llamamos a dicha variable “dic_sel” (diccionario de selección) y eliminamos todas las patologías que vengan acompañadas por FALSE de la variable “Estudio”. Véase el ejemplo:

```
Estudio <- c(T,T,T,T,T,T,T,F,T,T,T,T,F,T,T,T,T,T,T,F,F,T,T,F,T,F,F,F,T,T,F,T,F,
F,T,T,F,T,F,T,F,F,F,F,T,T,F,F,T,F,T,F,F,F,T,T,F,T,T,T,T,T,F,T,T,F,T,T,T,T,T,T,
T,T,T,T,T,T,T,T,T)
tabla_patologias$Estudio <- Estudio

dic_sel <- tabla_patologias$Estudio
names(dic_sel) <- tabla_patologias$PATOLOGIA_BPS
patologias_m3 <- patologias_m3[dic_sel[patologias_m3$PATOLOGIA_BPS],]
```

En la tabla 16 nos deshacemos de las patologías que no son de nuestro interés, eliminando las patologías que posean FALSE en la variable Estudio.

Tabla 16. Patologías de interés

PATOLOGIA_BPS	N	Porcentaje	Prevalencia	Estudio
Trastorno de ansiedad	6.469	10,69	21,56	TRUE
Hipertensión	6.278	10,37	20,93	TRUE
Dislipemia	6.154	10,17	20,51	TRUE
Artrosis, espondilosis	4.444	7,34	14,81	TRUE
Diabetes	2.843	4,70	9,48	TRUE
Asma	2.719	4,49	9,06	TRUE

Fuente: Elaboración propia

Añadimos una nueva variable a nuestra tabla 16, llamada “Abreviaturas”, para agrupar las patologías en grupos, que contengan 1 o más de nuestras patologías a estudiar y reducir el nombre con dichas abreviaturas. Cada abreviatura debe corresponder a su patología.

Tabla 17. Patologías de interés y Abreviaturas

PATOLOGIA_BPS	N	Porcentaje	Prevalencia	Estudio	Abreviaturas
Trastorno de ansiedad	6.469	10,69	21,56	TRUE	T
Hipertensión	6.278	10,37	20,93	TRUE	H
Dislipemia	6.154	10,17	20,51	TRUE	Dp
Artrosis, espondilosis	4.444	7,34	14,81	TRUE	Art
Diabetes	2.843	4,70	9,48	TRUE	D
Asma	2.719	4,49	9,06	TRUE	As

Fuente: Elaboración propia

Vamos a incorporar la variable “Abreviaturas” a la tabla patologias_m3, de la misma forma que lo hemos hecho anteriormente, llamando a dicha variable “dic_sel” y que a cada patología se le asocie su abreviatura correspondiente. A su vez, ordenamos la tabla para que la columna de las abreviaturas aparezca a continuación de la de patologías.

Tabla 18. Patologías_m3 con abreviaturas

n_orden	CODIGO_PATOLOGIA_BPS	PATOLOGIA_BPS	Abreviaturas	FECHA_INICIO	FECHA_FIN
587	1304	Otra artropatía	Art	14/10/2013	31/12/9999
1806600	1002	Asma	As	19/03/2012	31/12/9999
441220	1002	Asma	As	18/02/2010	31/12/9999
4709259	1002	Asma	As	26/02/2013	31/12/9999
3644186	1001	EPOC	As	19/11/2008	31/12/9999
3644186	909	Fibrilación auricular	Cai	19/11/2008	31/12/9999

Fuente: Elaboración propia

A continuación, se pegan las tablas usuarios_m4 y patologias_m3, puesto que ambas poseen en común la variable “n_orden”, creando la tabla Result para el posterior análisis de secuencias. Así mismo vamos a cambiar el formato de la variable “fecha inicio de patología” y “fecha fin de patología” a un número entero, quedándonos solo con el año final para poder hacer la operación correspondiente para obtener la edad inicio de patología y la edad fin de patología.

```
## [1] "2011-10-10" "2006-08-29" "2004-11-15" "2010-06-14" "2009-10-07"
## [6] "2014-04-07"
## [1] "2011" "2006" "2004" "2010" "2009" "2014"
## [1] 2011 2006 2004 2010 2009 2014
## [1] "9999-12-31" "9999-12-31" "9999-12-31" "9999-12-31" "9999-12-31"
## [6] "9999-12-31"
## [1] "9999" "9999" "9999" "9999" "9999" "9999"
## [1] 9999 9999 9999 9999 9999 9999
```

Para la edad del paciente a la que finaliza la patología, vamos a usar un condicionante, para que si el resultado de la edad fin es mayor que la edad actual del paciente, nos proporcione, en ese caso, la edad actual del paciente. Ejemplo:

```
Result$EDAD_FIN <- ifelse(Result$EDAD_FIN>Result$EDAD,Result$EDAD,Result$EDAD_FI
N)
```

De esta forma, nos quedamos solo con los tipos de cáncer para reducir el número de patologías lo máximo posible y así poder realizar un estudio sobre cánceres. (Estudio 1).

En nuestra muestra nos encontramos con 22 tipos de cánceres diferentes.

Tabla 19. Tipos de Cáncer

PATOLOGIA_BPS	N	Porcentaje	Prevalencia	Estudio	Abreviaturas	Estudio1
Cáncer de mama	275	0,45	0,92	TRUE	C	TRUE
Cáncer de próstata	189	0,31	0,63	TRUE	C	TRUE
Cáncer colorrectal	152	0,25	0,51	TRUE	C	TRUE
Cáncer de vejiga	117	0,19	0,39	TRUE	C	TRUE
Cáncer de bronquio y pulmón	59	0,10	0,20	TRUE	C	TRUE
Cáncer de cabeza y cuello	56	0,09	0,19	TRUE	C	TRUE

Fuente: Elaboración propia

En la tabla 19 podemos observar los seis tipos de cánceres más frecuentes de nuestro estudio. El 3,8% del total de usuarios presenta algún cáncer, es decir 1.139 personas tienen cáncer. Entre las personas con cáncer en nuestro estudio, casi un 25% de pacientes tiene diagnosticado cáncer de mama. Del mismo modo, hemos analizado los individuos que presentan 1 o más cánceres a la vez.

Tabla 20. Nº de Cánceres por persona

Numero de Cánceres	Personas
1	1.029
2	107
3	2
4	1
Total	1.139

Fuente: Elaboración propia

En la tabla 20, existe una persona que presenta 4 cánceres diferentes, por el contrario 1.029 personas poseen 1 cáncer.

Por otro lado, para comenzar el Estudio 2 se debe ejecutar desde el inicio (ignorando el Estudio 1), como hemos explicado al comienzo del análisis de secuencias.

En la primera fase del Estudio 2, se han seleccionado un total de 10 patologías. En una segunda fase, se han agrupado según afinidad de las patologías en 5 grupos. De este modo ha resultado la

siguiente agrupación: Asma y EPOC = Asma; Artrosis y Artritis reumatoide = Artrosis; etc. Al objeto de representar los patrones de multimorbilidad de forma más eficiente.

Tabla 21. Patologías seleccionadas

PATOLOGIA_BPS	N	Porcentaje	Prevalencia	Estudio	Abreviaturas	Estudio2
Trastorno de ansiedad	6.469	10,69	21,56	TRUE	T	TRUE
Hipertensión	6.278	10,37	20,93	TRUE	H	TRUE
Artrosis-espondilosis	4.444	7,34	14,81	TRUE	Art	TRUE
Diabetes	2.843	4,70	9,48	TRUE	D	TRUE
Asma	2.719	4,49	9,06	TRUE	As	TRUE

Fuente: Elaboración propia

En esta tabla 21 distinguimos las cinco patologías seleccionadas en nuestro Estudio 2. De este modo, se observa que el 21,56% del total de usuarios presentan Trastornos de ansiedad, al igual que más de un 20% de pacientes muestran Hipertensión.

En la tabla 22 hemos analizado el número de patologías por persona.

Tabla 22. Nº Patologías por persona

Nº Patologías	Personas
1	7.296
2	3.717
3	2.026
4	895
5	362
6	100
7	27
8	5
9	1
Total	14.429

Fuente: Elaboración propia

Observamos que 7.296 personas muestran al menos una de las patologías objeto de estudio y existe una persona con 9 patologías de nuestro estudio. Se reduce nuestro análisis a un total de 14.429 pacientes.

De la siguiente forma vamos a ordenar las abreviaturas en el análisis de secuencias final para cada paciente (orden alfabético):


```
cadena <- c("Art", "As", "Cai")
cadena2 <- c("C")
cad <- sort(c(cadena, cadena2))
```

Creamos un nuevo data.frame, llamado Fin, con las variables "n_orden", "EDAD" y "Fallecido" de la tabla usuario_m4.

A su vez, añadimos las variables "Edad" desde la EDAD000 a la Edad114, que son las edades que pueden tener nuestros pacientes, para el posterior análisis de secuencias.

Vamos a crear una nueva tabla, llamada temp_pac (paciente temporal), donde el paciente en la posición 1 de la tabla Fin sea nuestro único dato, pero lo extraemos de la tabla Result con todas las variables que posea en dicha tabla. Nos dedicaremos a analizar un paciente exclusivamente, por ahora.

Vamos a crear un bucle "for" para poder realizar el análisis de secuencias con el paquete TRAMINER, en nuestra tabla Fin podemos ver todas las variables EDAD de los pacientes, donde se le va a asignar una "S", a cada edad del paciente, hasta que aparezca su primera patología y seguirá teniendo una "S" desde que se le cierre dicha patología hasta llegar a la EDAD actual del paciente, que se le asignara una "N" hasta la EDAD máxima. En el caso de que tenga múltiples patologías a la vez, hasta que no finalicen todas las patologías, no le volverá a aparecer una "S", o bien una "N" si llega a su EDAD actual. Si el paciente ha fallecido, se le asignará una "M".

Dentro del bucle, se genera el método necesario para que, en caso de contar con múltiples patologías, nos indique cada una de ellas, así como la secuencia de patologías que presentará cada paciente pluripatológico.

Generamos una muestra aleatoria simple sin reemplazamiento de 3.000 pacientes a partir de la tabla Fin, para poder ejecutar el paquete TRAMINER.

```
muestra.Fin <- Fin[sample(nrow(Fin), size=3000, replace=F),]
```

6.2.1 Para el Estudio 1:

Preparamos la matriz para poder ejecutar el paquete TRAMINER:

```
Fin.labels <- c("Presenta Cáncer", "Fallecido", "No vivido", "Sin patología")
Fin.alphabet <- c("C", "M", "N", "S")
```

```
Fin.cpal <- c("pink", "black", "gray", "lightblue")
```

Preparamos la secuencia definiendo el alfabeto a utilizar, los colores y las etiquetas, así como la matriz de donde vamos a realizar el análisis de secuencia. Lo nombramos: Ansec_OSEQ.

Tabla 23. Alfabeto

Etiquetas	Niveles
C	Presenta Cáncer
M	Fallecido
N	No vivido
S	Sin patología

Fuente: Elaboración propia

Tabla 24. Tiempo (años) en cada estado

n_orden	C	M	N	S
6517	0	0	96	19
10078	0	0	84	31
14068	0	0	89	26
25769	0	0	49	66

Fuente: Elaboración propia

La tabla 24 nos indica los años que ha vivido cada paciente en cada estado o etiqueta. Por ejemplo, el paciente 6517 (n_orden) ha vivido 19 años sano y le quedan 96 por vivir. La suma de los valores de las etiquetas de cada paciente es 115 años, puesto que contamos las edades desde 0 a 114 años.

Tabla 25. Media del tiempo (años) dedicada a cada tipo de estado

Estado	Media
C	0,22
M	0,28
N	72,01
S	42,49

Fuente: Elaboración propia

La tabla 25 representa el tiempo medio que cada individuo dedica a cada estado. Podemos observar que, en el estado “S” los individuos están 42 años y medio. Por el contrario, solo presentan cáncer durante aproximadamente 3 meses de media.

Se generan 4 clúster para poder observar el comportamiento del análisis de secuencias en cada uno de ellos. Generamos los clúster del análisis de secuencias:

Ilustración 9. Análisis de secuencias. Estudio 1

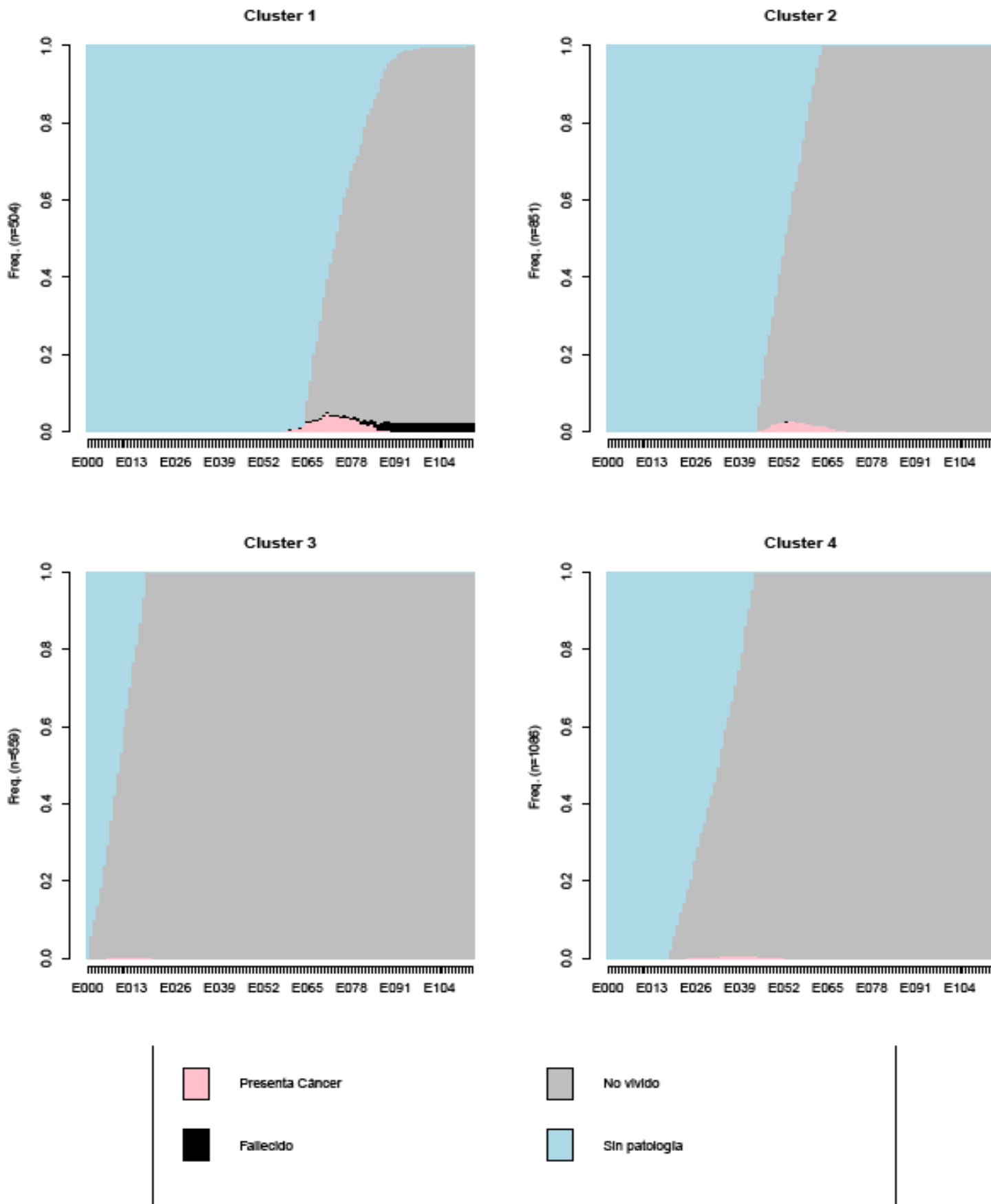


Ilustración Estudio 1:

En el primer clúster, n=504 pacientes, se observan pacientes que han llegado a la edad de (55-60) años, completamente sanos. A partir de dicha edad hasta los (85-90) años, se le ha diagnosticado algún tipo de cáncer, que en la mayoría de casos les ha causado la muerte. También se observa una minoría de pacientes sanos hasta edades muy longevas.

En el segundo clúster, n=851 pacientes, se observan pacientes sanos hasta los 45 años aproximadamente. A partir de dicha edad, presentan algún tipo de cáncer, con el que conviven. Se observan pacientes con cáncer hasta los 70 años. No se aprecian pacientes fallecidos, salvo una minoría en torno a los 55 años. La mayoría de los pacientes con cáncer tienen una edad media entre 50 y 70 años.

En el tercer clúster, n=559 pacientes, se observan pacientes de edad temprana, aproximadamente hasta los 20 años. Una minoría de ellos presenta algún tipo de cáncer infantil (leucemia, cáncer de encéfalo,...). La mayoría de pacientes se encuentran sanos.

En el cuarto clúster, n=1086 pacientes, se observan pacientes de edad media-joven, aproximadamente hasta los (45-50) años. Una minoría de ellos presenta algún tipo de cáncer, entre los 20 y los 50 años. Los demás pacientes de este clúster se encuentran sanos hasta la actualidad.

6.2.2 Para el Estudio 2:

Preparamos la matriz para poder ejecutar el paquete TRAMINER:

```
Fin.labels <- c("Presenta Artrosis", "Presenta Artrosis y Asma", "Presenta Artrosis, Asma y Diabetes", "Presenta Artrosis, Asma, Diabetes e Hipertensión", "Presenta Artrosis, Asma, Diabetes, Hipertensión y Trastorno", "Presenta Artrosis, Asma, Diabetes y Trastorno", "Presenta Artrosis, Asma e Hipertensión", "Presenta Artrosis, Asma, Hipertensión y Trastorno", "Presenta Artrosis, Asma y Trastorno", "Presenta Artrosis y Diabetes", "Presenta Artrosis, Diabetes e Hipertensión", "Presenta Artrosis, Diabetes, Hipertensión y Trastorno", "Presenta Artrosis, Diabetes y Trastorno", "Presenta Artrosis e Hipertensión", "Presenta Artrosis, Hipertensión y Trastorno", "Presenta Artrosis y Trastorno", "Presenta Asma", "Presenta Asma y Diabetes", "Presenta Asma, Diabetes e Hipertensión", "Presenta Asma, Diabetes, Hipertensión y Trastorno", "Presenta Asma, Diabetes y Trastorno", "Presenta Asma e Hipertensión", "Presenta Asma, Hipertensión y Trastorno", "Presenta Asma y Trastorno", "Presenta Diabetes", "Presenta Diabetes e Hipertensión", "Presenta Diabetes, Hipertensión y Trastorno", "Presenta Diabetes y Trastorno", "Presenta Hipertensión", "Presenta Hipertensión y Trastorno", "Fallecido", "No vivido", "Sin patología", "Presenta Trastorno")
```

```
Fin.alphabet <- c("Art", "ArtAs", "ArtAsD", "ArtAsDH", "ArtAsDHT", "ArtAsDT", "ArtAsH", "ArtAsHT", "ArtAsT", "ArtD", "ArtDH", "ArtDHT", "ArtDT", "ArtH", "ArtHT")
```

```
", "ArtT", "As", "AsD", "AsDH", "AsDHT", "AsDT", "AsH", "AsHT", "AsT", "D", "DH", "DHT", "DT", "H", "HT", "M", "N", "S", "T")
```

```
Fin.cpal <- c("white", "aliceblue", "aquamarine", "antiquewhite1", "antiquewhite4", "aquamarine4", "azure3", "beige", "blue", "blueviolet", "brown", "burlywood", "chocolate", "chocolate4", "coral", "cornflowerblue", "cornsilk", "pink", "cyan", "darkblue", "darkgoldenrod", "darkgoldenrod4", "darkgreen", "darkkhaki", "darkmagenta", "darkolivegreen", "darkolivegreen1", "darkorange", "darkred", "black", "gray", "darksalmon", "yellow")
```

Preparamos la secuencia definiendo el alfabeto a utilizar, los colores y las etiquetas, así como la matriz de donde vamos a realizar el análisis de secuencia. Lo llamamos Ansec_OSEQ.

```
## [>] 34 distinct states appear in the data:
```

Tabla 26. Alfabeto (no completo)

Etiquetas	Niveles
Art	Presenta Artrosis
ArtAs	Presenta Artrosis y Asma
ArtAsD	Presenta Artrosis, Asma y Diabetes
ArtAsDH	Presenta Artrosis, Asma, Diabetes e Hipertensión
ArtAsDHT	Presenta Artrosis, Asma, Diabetes, Hipertensión y Trastorno
ArtAsDT	Presenta Artrosis, Asma, Diabetes y Trastorno
ArtAsH	Presenta Artrosis, Asma e Hipertensión
ArtAsHT	Presenta Artrosis, Asma, Hipertensión y Trastorno
ArtAsT	Presenta Artrosis, Asma y Trastorno
ArtD	Presenta Artrosis y Diabetes
ArtDH	Presenta Artrosis, Diabetes e Hipertensión
ArtDHT	Presenta Artrosis, Diabetes, Hipertensión y Trastorno

Fuente: Elaboración propia

Tabla 27. Tiempo (años) medio en cada estado

Usuarios	Art	ArtAs	ArtAsD	ArtAsDH	ArtAsDHT	ArtAsDT	ArtAsH	ArtAsHT	ArtAsT	ArtD			
20381	0	4	0	0	0	0	0	0	0	0			
18480	0	0	0	0	0	0	0	0	0	0			
17631	0	0	0	0	0	0	0	0	0	0			
17591	0	0	0	0	0	0	0	0	0	0			
13295	0	0	0	0	0	0	0	0	0	0			
4491	0	0	0	0	0	0	0	0	0	0			
	ArtDH	ArtDHT	ArtDT	ArtH	ArtHT	ArtT	As	AsD	AsDH	AsDHT	AsDT	AsH	AsHT
20381	0	0	0	0	0	0	7	0	0	0	0	0	0
18480	0	0	0	0	0	0	0	0	0	0	0	0	0
17631	0	0	0	0	0	0	0	0	0	0	0	0	0
17591	0	0	0	0	0	0	0	0	0	0	0	0	0
13295	0	0	0	0	0	0	0	0	0	0	0	0	0
4491	0	0	0	0	0	0	0	0	0	0	0	0	0

	AsT	D	DH	DHT	DT	H	HT	M	N	S	T
20381	0	0	0	0	0	0	0	0	61	43	0
18480	0	14	0	0	0	0	0	0	53	48	0
17631	0	0	0	0	0	0	0	0	68	47	0
17591	0	0	0	0	0	17	0	0	66	32	0
13295	0	0	0	0	1	0	0	0	73	39	2
4491	0	0	0	0	0	0	0	0	100	14	1

Fuente: Elaboración propia

La tabla 27 nos indica los años que ha vivido cada paciente en cada estado o etiqueta. Por ejemplo, el paciente 20381 (n_orden) ha vivido 43 años sano, lleva 7 años con Asma, 4 años con Asma y Artrosis y le quedan 61 años por vivir. Tal como hemos comentado anteriormente, la suma de los valores de las etiquetas de cada paciente es 115 años, puesto que contamos las edades desde 0 a 114 años.

Tabla 28. Media del tiempo (años) dedicada a cada tipo de evento

Estado	Media
Art	0,364
ArtAs	0,043
ArtAsD	0,006
ArtAsDH	0,0357
ArtAsDHT	0,0103
ArtAsDT	0,0017
ArtAsH	0,0423
ArtAsHT	0,025
ArtAsT	0,022
ArtD	0,0593
ArtDH	0,11
ArtDHT	0,072
ArtDT	0,033
ArtH	0,2783
ArtHT	0,152
ArtT	0,1293
As	0,6067
AsD	0,0153
AsDH	0,0247
AsDHT	0,0123
AsDT	0,0093
AsH	0,0707
AsHT	0,0227
AsT	0,0913
D	0,2263
DH	0,201
DHT	0,0443
DT	0,0597
H	0,74
HT	0,1333

M	0,242
N	72,740
S	37,328
T	1,0480

Fuente: Elaboración propia

La tabla 28 representa el tiempo medio que se dedica a cada estado. Podemos observar que, el estado “N” es el que más tiempo ocupa en nuestra muestra. Lo cual tiene sentido, porque a un gran número de usuarios les falta muchos años para llegar a 114, a no ser que hayan fallecido, que como podemos ver en la etiqueta “M”, no ocupa mucho tiempo en el análisis de secuencia de nuestros usuarios. Por el contrario, vemos que de media los pacientes presentan algún Trastorno durante aproximadamente 1 año e Hipertensión durante 9 meses.

Se generan 4 clúster para poder observar el comportamiento del análisis de secuencias en cada uno de ellos. Generamos los clúster del análisis de secuencias.

Ilustración 10. Análisis de secuencias. Estudio 2

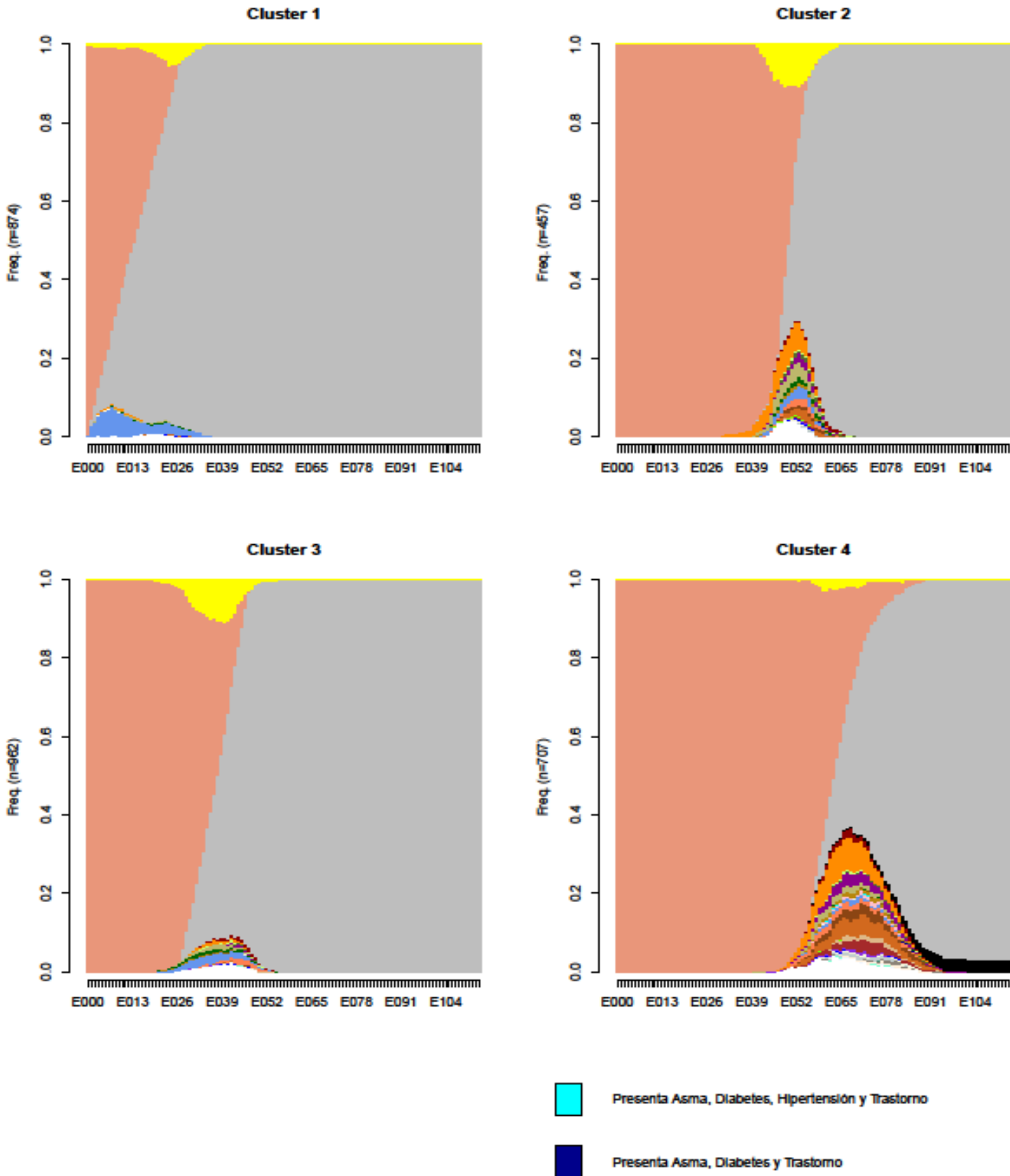


Ilustración Estudio 2:

En el primer clúster, n=874 pacientes, se observan pacientes muy jóvenes, en concreto de 0 a 35 años, en los cuales predomina el Asma desde su nacimiento hasta la actualidad. Cabe señalar también una gran parte de pacientes que sufren algún Trastorno (de ansiedad o de estado de ánimo), cuyo apogeo se alcanza de los 15 a los 30 años.

En el segundo clúster, n=457 pacientes, se observan pacientes sanos hasta los 30 años aproximadamente. A partir de dicha edad, comienzan a presentar Hipertensión (naranja), aunque es a los 40 años cuando alcanza su mayor frecuencia. Cabe señalar también a partir de dicha edad un gran número de patologías, entre las que destacan Artrosis (blanco), Asma (celeste), Diabetes (kaki), y la anteriormente citada Hipertensión. Aunque la que vemos en un mayor número de pacientes son los Trastornos de ansiedad y de estado de ánimo, que oscilan entre los 40 y 60 años. A partir de los 50 años se encuentra la combinación que más predomina en este clúster, Artrosis e Hipertensión (chocolate), que se le puede unir algún Trastorno. Se observan pacientes hasta los 70 años (año 2017).

En el tercer clúster, n=962 pacientes, se observan pacientes de edad media, aproximadamente hasta los 52 años. Los cuales presentan, a partir de los 20 años, el conjunto de patologías y combinaciones de enfermedades con una frecuencia similar, aunque predominan los pacientes con Asma e Hipertensión. Cabe señalar un gran número de pacientes con algún Trastorno entre los 20 y los 50 años.

En el cuarto clúster, n=707 pacientes, se observan pacientes que han llegado a la edad de (45-50) años, completamente sanos. A partir de dicha edad hasta los 90 años, se le han diagnosticado patologías, las cuales han sido combinaciones de las enfermedades de nuestro estudio, entre las que predomina la Hipertensión, Hipertensión y Trastorno (rojo), Artrosis, Hipertensión y Trastorno (chocolate⁴), entre otras. La mayoría de estos pacientes han fallecido a consecuencia de la presencia de varias patologías en cada uno de ellos. También se observa una minoría de pacientes, que presentan algún tipo de Trastorno entre los 60 y 80 años.

7 CONCLUSIONES

- Se puede observar que los resultados del índice Barthel, el índice Pfeiffer presentan muy baja cumplimentación en los registros clínicos, sólo se notifican en el 5% y 4% cada uno de ellos.
- Las patologías más frecuentes en nuestros pacientes son Trastorno de ansiedad (10,69%), Hipertensión (10,37%), Dislipemia (10,17%) y Artrosis (7,34%).
- La prevalencia es superior al 20% en Trastornos de ansiedad, Hipertensión y Dislipemia.
- Los paciente que se encuentran en el grupos de edad de (65,75) años presentan una media de enfermedades de 4,44, los de (75,85) años presentan una media de 5,75 enfermedades crónicas y por último los del grupo de (85,95) años presentan la media de enfermedades más alta de 5,88 enfermedades crónicas.
- Destacamos como las mujeres poseen una media de enfermedades del 2,20, frente a las 1,80 enfermedades en los hombres.
- Las combinaciones más frecuentes para los pacientes que presentan dos enfermedades son Dislipemia e Hipertensión.
- Cuando padecen tres enfermedades se añade Dislipemia, Hipertensión y Artrosis.
- Si los pacientes tienen cuatro y cinco enfermedades se les agrega Trastorno de ansiedad y Litiasis urinaria respectivamente.
- En todos los grupos de edad, se manifiesta que la mujer es la que acude con más frecuencia a las consultas de atención primaria de los servicios sanitarios. Excepto en la edad pediátrica (0-15 años), que presentan valores muy similares entorno a las 15.000 consultas en ambos sexos
- En atención primaria, de los 10,7 millones de gastos imputados en nuestra muestra, el 51,31% de los costes son debidos a la farmacia; es decir a los fármacos que se prescriben a los usuarios.

- El consumo de costes según la actividad realizada en atención primaria es mayor a medida que aumenta la edad del paciente.
- El análisis de secuencias del Estudio 1 se refiere a los diferentes tipos de cáncer que hay en nuestra muestra. Solo un 3,8% (1.139) de los pacientes presentan algún tipo de cáncer.
- Se observa que la mayoría de los cánceres se diagnostican a partir de los 45-60 años, excepto para una minoría de cánceres juveniles que suelen aparecer en edades tempranas hasta los 20 años.
- En el análisis realizado en el Estudio 2, se aprecia que en las edades jóvenes (0-35 años) predominan las patologías de Asma y Trastorno de ansiedad o de estado de ánimo.
- A partir de los 50 años la secuencia de enfermedades más frecuente es Artrosis, Hipertensión y algún Trastorno mental.
- Por último, se necesita una mejor comprensión de la multimorbilidad, puesto que es uno de los problemas más comunes y complejos de la medicina moderna y requiere un enfoque clínico más eficaz.

8 BIBLIOGRAFÍA

- Alguacil Herrero, M. D., Álvarez Tello, M., Barón Franco, B., Cabrera León, A., et al. (2012). Plan de Atención a Pacientes Crónicos con necesidades de salud complejas. Estrategia de implantación. Servicio Andaluz de Salud. Consejería de Salud. Junta de Andalucía: D.L.: SE 1291-2012. Disponible en <https://www.juntadeandalucia.es/servicioandaluzdesalud/.../Plan-Atención-Pacientes-con-Enfermedades-Crónicas>.
- Bartke, A., Coschigano, K., Kopchick, J., Chandrashekar, V., Mattison, J., Kinney, B. & Hauck, S. (2001). Genes that prolong life: Relationships of growth hormone and growth to aging and life span. *J Gerontol Biol Sci*. 56A:B340-9. doi: 10.1093/gerona/56.8.B340.
- Bohórquez Colombo, P., Nieto Martín, M. D., Pascual de la Pisa, B., García Lozano, M. J., Ortiz Camúñez, M. A. & Wittel, M. B. (2014). Validación de un modelo pronóstico para pacientes pluripatológicos en atención primaria: Estudio PROFUND en atención primaria. *Aten Primaria*. 46(Supl 3):41-48. doi: 10.1016/S0212-6567(14)70064-2.
- Davidson, A. J. (2015). Creating value: unifying silos into public health business intelligence. *EGEMS (Wash DC)*. 2:1172. doi: 10.13063 / 2327-9214.1172.
- Denton, F. T. & Spencer, B. G. (2010). Chronic health conditions: changing prevalence in an aging population and some implications for the delivery of health care services. *Can J Aging*. Mar;29(1):11-21. doi: 10.1017/S0714980809990390.
- Goicoechea-Salazar, J. A., Nieto-García, M. A., Laguna-Téllez, A., et al. (2013) Desarrollo de un sistema de codificación automática para recuperar y analizar textos diagnósticos de los registros de servicios de urgencias hospitalarios. *Emergencias*. 25: 430–436.
- Jefatura del Estado. Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales. BOE, núm. 294, de 6 de diciembre de 2018.
- Junta de Andalucía. Decreto 105/2019, de 12 de febrero, por el que se establece la estructura orgánica de la Consejería de Salud y Familias y del Servicio Andaluz de Salud. Boletín Oficial de la Junta de Andalucía (BOJA), núm. 31, de 14 de febrero de 2019.

- Kostkova, P. (2015). Grand challenges in digital health. *Front Public Health*. doi: 10.3389/fpubh.2015.00134.
- Lafuente Robles, N., Rodríguez Gómez, S., Casado Mora, M. I., Ayuso Fernández, M., et al. (2015). Estrategia de Cuidados de Andalucía: nuevos retos en el cuidado de la ciudadanía. Consejería de Salud: Servicio Andaluz de Salud. Disponible en: <https://www.picuida.es/estrategia-de-cuidados-de-andalucia/>.
- Lafuente Robles, N. & Rodríguez Gómez, S. (2019). CuidandodeTi. Proyecto para el abordaje de la cronicidad. (Informe interno). Servicio Andaluz de Salud. Consejería de Salud y Familias.
- Lewer, D., Bourne, T., George, A., Abi-Aad, G., Taylor, C. & George, J. (2019). Data resource: the Kent Integrated Dataset (KID). *Int J Popul Data Sci*. 2018:3. (Consultado el 6/3/2019). doi: 10.23889/ijpds.v3i1.427.
- Martínez Castells, A. (2014). La economía de la salud y la salud de la economía. *El Farmacéutico*. num. 500, 15 de febrero de 2014. Disponible en: <https://elfarmacéutico.es/index.php/la-revista/item/4361-la-economia-de-la-salud-y-la-salud-de-la-economia>.
- Muñozerro-Muñiz, D., Goicoechea-Salazar, J. A., García-León, F. J., Laguna-Téllez, A., Larrocha-Mata, D. & Cardero-Rivas, M. (2019). Conexión de registros sanitarios: base poblacional de salud de Andalucía. *Gaceta Sanitaria*. doi: 10.1016/j.gaceta.2019.03.003.
- Nicholson, K. (2017). Multimorbidity Among Adult Primary Health Care Patients In Canada: Examining Multiple Chronic Diseases Using An Electronic Medical Record Database. *Electronic Thesis and Dissertation Repository*. 4483. Disponible en <https://ir.lib.uwo.ca/etd/4483>.
- O'Halloran, J., Miller, G. C. & Britt, H. (2004). Defining chronic conditions for primary care with ICPC-2. *Fam Pract*. 21:381---6. doi: 10.1093/fampra/cmh407.
- Ortún-Rubio, V., Pinto-Prades, J. L. & Puig-Junoy, J. (2001). La economía de la salud y su aplicación a la evaluación. *Atención Primaria*. Vol. 27. Núm. 1. Enero 2001. Disponible en: <https://jaumepuigjunoy.cat/wp-content/uploads/2017/11/abc1.pdf>.

- Rahman Safwan, E., Meredith, R. & Burstein, F. (2016). Evolución del sistema de Business Intelligence (BI): un caso en una institución de salud. *Journal of Decision Systems*. 25:sup1, 463-475. doi: 10.1080/12460125.2016.1187384.
- Sánchez Mollá, M., Candela García, I., Gómez-Romero, F. J., Orozco Beltrán, D. & Ollero Baturone, M. (2017). Concordancia entre sistemas de estratificación e identificación de pacientes crónicos complejos en Atención Primaria. *Rev Calid Asist*. 32(1):10---16. doi: 10.1016/j.cali.2016.07.006.
- Schlegel, D. R. & Ficheur, G. (2017). Secondary use of patient data: review of the literature published in 2016. *Yearb Med Inform*. 26:68–70. doi: 10.15265/IY-2017-032.
- Servicio Andaluz de Salud. (2018). Contabilidad Analítica de Gestión del Servicio Sanitario Público de Andalucía (COAN SSPA). Recuperado 13 de diciembre 2018. https://ws027.juntadeandalucia.es/principal/documentosacc.asp?pagina=pr_coan.
- Servicio Andaluz de Salud. (2018). Resolución 0068/18 de la Dirección Gerencia del Servicio Andaluz de Salud por la que se crea la Base Poblacional de Salud. Disponible en http://www.sas.junta-andalucia.es/principal/documentosacc.asp?pagina=pr_base_poblac.
- Starfield, B., Lemke, K. W., Bernhardt, T., Foldes, S. S., Forrest, C. B. & Weiner, J. P. (2003). Comorbidity: implications for the importance of primary care in 'case' management. *Ann Fam Med*. 1:8---14. doi: 10.1370.
- World Health Organization. (2002). The World Health Report: reducing risk, promoting healthy life. Ginebra: World Health Organization; 2002. ISBN 92 4 156207 2.
- World Health Organization. (2008). The World Health Report: The Challenges of changing world. Primary Health Care (Now More Than Ever). 2008, p. 21.

Páginas Web:

- <http://mephisto.unige.ch/pub/TraMineR/doc/TraMineR-Users-Guide.pdf>
- https://cran.r-project.org/doc/contrib/rdebuts_es.pdf
- <https://www.juntadeandalucia.es/servicioandaluzdesalud/>

9 ANEXO: Dictamen favorable y Estudio en RStudio

JUNTA DE ANDALUCÍA

CONSEJERÍA DE SALUD Y FAMILIAS

DICTAMEN ÚNICO EN LA COMUNIDAD AUTÓNOMA DE ANDALUCÍA

D/Dª: Carlos García Pérez como secretario/a del CEI de los hospitales universitarios Virgen Macarena-Virgen del Rocío

CERTIFICA

Que este Comité ha evaluado la propuesta del promotor/investigador (No hay promotor/a asociado/a) para realizar el estudio de Investigación titulado:

TÍTULO DEL ESTUDIO: ANÁLISIS DE MULTIMORBILIDAD EN PACIENTES PLURIPATOLÓGICOS: BASE POBLACIONAL DE SALUD- ANDALUCÍA 2019
Protocolo, Versión: versión 3
HIP, Versión:
CI, Versión:

Y que considera que:

Se cumplen los requisitos necesarios de idoneidad del protocolo en relación con los objetivos del estudio y se ajusta a los principios éticos aplicables a este tipo de estudios.

La capacidad de la investigadora y los medios disponibles son apropiados para llevar a cabo el estudio.

Están justificados los riesgos y molestias previsibles para los participantes.

Que los aspectos económicos involucrados en el proyecto, no interfieren con respecto a los postulados éticos.

Y que este Comité considera, que dicho estudio puede ser realizado en los Centros de la Comunidad Autónoma de Andalucía que se relacionan, para lo cual corresponde a la Dirección del Centro correspondiente determinar si la capacidad y los medios disponibles son apropiados para llevar a cabo el estudio.

Lo que firmo en Sevilla a 02/09/2019

D/Dª. Carlos García Pérez, como Secretario/a del CEI de los hospitales universitarios Virgen Macarena-Virgen del Rocío



Código Seguro De Verificación:	4ad4c08de0f8ae3be93ccceb845391847da4f0d	Fecha	02/09/2019	
Normativa	Este documento incorpora firma electrónica reconocida de acuerdo a la Ley 59/2003, de 19 de diciembre, de firma electrónica.			
Firmado Por	Carlos García Pérez			
Uri De Verificación	https://www.juntadeandalucia.es/salud/portaldesitica/xhtml/ayuda/verificarfirmaDocumento.iface/code/4ad4c08de0f8ae3be93ccceb845391847da4f0d	Página	1/2	

CERTIFICA

Que este Comité ha ponderado y evaluado en sesión celebrada el 29/07/2019 y recogida en acta 07/2019 la propuesta del/de la Promotor/a (No hay promotor/a asociado/a), para realizar el estudio de investigación titulado:

TÍTULO DEL ESTUDIO: ANÁLISIS DE MULTIMORBILIDAD EN PACIENTES PLURIPATOLÓGICOS: BASE POBLACIONAL DE SALUD- ANDALUCÍA 2019
Protocolo, Versión: versión 3
HIP, Versión:
CI, Versión:

Que a dicha sesión asistieron los siguientes integrantes del Comité:

President/a

D/Dª. Víctor Sánchez Margalet

Vicepresident/a

D/Dª.

Secretari/a

D/Dª. Carlos García Pérez

Vocales

D/Dª. Enrique Calderón Sandubete
D/Dª. José Gamacho Montero
D/Dª. Gabriel Ramírez Soto
D/Dª. Cristina Pichardo Guerrero
D/Dª. Javier Vitorica Fernandez
D/Dª. MARIA EUGENIA ACOSTA MOSQUERA
D/Dª. Luis Lopez Rodriguez
D/Dª. Enrique de Álava Casado
D/Dª. ANGELA CEJUDO LOPEZ
D/Dª. Amancio Camero Moya
D/Dª. Jose Salas Turrents
D/Dª. María Pilar Guadix Martín
D/Dª. ESPERANZA GALLEGU CALVENTE
D/Dª. Ana Melcon -

Que dicho Comité, está constituido y actúa de acuerdo con la normativa vigente y las directrices de la Conferencia Internacional de Buena Práctica Clínica.



Lo que firmo en Sevilla a 02/09/2019

Código Seguro De Verificación:	4ed4c08de0f8ae3be93ccceb845391847da4f0d	Fecha:	02/09/2019
Normativa	Este documento incorpora firma electrónica reconocida de acuerdo a la Ley 59/2003, de 19 de diciembre, de firma electrónica.		
Firmado Por	Carlos García Pérez		
Url De Verificación	https://www.juntadeandalucia.es/salud/portaldedeitica/xhtml/ayuda/verificarfirmaDocumento.iface/code/4ed4c08de0f8ae3be93ccceb845391847da4f0d	Página	1/2



Estudio en RStudio:

#Análisis de patologías crónicas sobre pacientes con multimorbilidad.

Vamos a realizar un estudio estadístico sobre multimorbilidad, es decir, pacientes que sufren más de una enfermedad crónica. Así mismo queremos observar la evolución de cada paciente y sus enfermedades a lo largo de su vida, mediante un análisis de secuencias. Para ello se nos proporciona una base de datos de pacientes de Andalucía que han sido atendidos y registrados por el servicio sanitario andaluz, a través de sus múltiples registros médicos electrónicos puestos en común.

En primer lugar, vamos a cargar la base de datos que se nos ha proporcionado:

```
````{r}
load("C:/Users/Administrador/Desktop/TFG/datosBPS.RData")
````
```

Acto seguido las librerías necesarias para utilizar los paquetes correspondientes en este estudio:

```
````{r}
library(dplyr)
library(TraMineR)
library(ggplot2)
library(base)
````
```

Vamos a trabajar con la tabla* `barthel_m3`, la cual nos indica una muestra de 1620 pacientes que sufren algún déficit funcional, que les impida ser independientes para realizar las actividades cotidianas, dentro de una escala de rangos, en concreto 5:

```
````{r echo=F}
head(barthel_m3)
````
```

La columna que nos indica el grado de índice `barthel` de cada paciente, la convertimos en factor.

```
````{r echo=F}
barthel_m3RESULTADO_BARTHEL <- factor(barthel_m3$RESULTADO_BARTHEL,
 levels= c("Independencia",
 "Dependencia escasa",
 "Dependencia moderada"),
```

```

"Dependencia severa",
"Dependencia total"))
....

Creamos una tabla de frecuencias con los 5 rangos en orden, de mayor a menor
número de pacientes
en cada grado, con su respectivo porcentaje respecto al total.
```{r echo=F}
tabla_barthel <- barthel_m3 %>%
  dplyr::group_by(RESPUESTA_BARTHEL) %>%
  dplyr::count() %>%
  arrange(desc(n))
tabla_barthel <- data.frame(tabla_barthel)

tabla_barthel <- tabla_barthel %>%
  dplyr::mutate(
    Porcentaje=round(n/sum(n)*100,2)
  )
tabla_barthel
....

Para acabar con esta tabla, vamos a cambiar el formato de la columna
Fecha_barthel y la pondremos en
formato fecha (as.Date).
```{r echo=F}
barthel_m3$FECHA_NUEVA <- paste(substr(barthel_m3$FECHA_BARTHEL,1,4),
 substr(barthel_m3$FECHA_BARTHEL,5,6),
 substr(barthel_m3$FECHA_BARTHEL,7,8),
 sep="-")
barthel_m3$FECHA_NUEVA <- as.Date(barthel_m3$FECHA_NUEVA,"%Y-%m-%d")

barthel_m3$FECHA_BARTHEL <- NULL
....

Hacemos lo mismo con las columnas Fecha_inicio y Fecha_fin de la tabla*
patologias_m3:
```{r echo=F}
patologias_m3$FECHA_INICIO <-
paste(substr(patologias_m3$FECHA_INICIO_PATOLOGIA_BPS,1,4),

substr(patologias_m3$FECHA_INICIO_PATOLOGIA_BPS,5,6),

```

```

substr(patologias_m3$FECHA_INICIO_PATOLOGIA_BPS,7,8),
      sep="-")
patologias_m3$FECHA_INICIO_PATOLOGIA_BPS      <-
as.Date(patologias_m3$FECHA_INICIO,"%Y-%m-%d")
patologias_m3$FECHA_INICIO <- NULL

```

```

patologias_m3$FECHA_FIN      <-
paste(substr(patologias_m3$FECHA_FIN_PATOLOGIA_BPS,1,4),

substr(patologias_m3$FECHA_FIN_PATOLOGIA_BPS,5,6),

substr(patologias_m3$FECHA_FIN_PATOLOGIA_BPS,7,8),
      sep="-")
patologias_m3$FECHA_FIN_PATOLOGIA_BPS      <-
as.Date(patologias_m3$FECHA_FIN,"%Y-%m-%d")

```

```

patologias_m3$FECHA_FIN <- NULL
****

```

Vamos a trabajar con la tabla* pfeiffer_m3, que nos indica una muestra de 1139 pacientes que sufren algún tipo de demencia (deterioro cognitivo), y a su vez, ver el grado de deterioro que presenta cada uno.

Lo primero que vamos a hacer es pasar la columna fecha_pfeiffer a formato fecha (as.Date).

```

``````{r echo=F}
pfeiffer_m3$FECHA_PFEIFFER_NUEVA <-
paste(substr(pfeiffer_m3$FECHA_PFEIFFER,1,4),

substr(pfeiffer_m3$FECHA_PFEIFFER,5,6),

substr(pfeiffer_m3$FECHA_PFEIFFER,7,8),
 sep="-")
pfeiffer_m3$FECHA_PFEIFFER <-
as.Date(pfeiffer_m3$FECHA_PFEIFFER_NUEVA,"%Y-%m-%d")
pfeiffer_m3$FECHA_PFEIFFER_NUEVA <- NULL

```

Pasamos la columna de los resultados del test de pfeiffer a factor para poder trabajar con ella.

```
````{r echo=F}
pfeiffer_m3RESULTADO_PFEIFFER <- factor(pfeiffer_m3$RESULTADO_PFEIFFER,
                                       levels= c("Compatible con la
normalidad",
                                               "Deterioro intelectual
leve",
                                               "Deterioro intelectual
moderado",
                                               "Deterioro intelectual
severo"))
````
```

Nos creamos una tabla de frecuencias con los 4 grados del test en orden, de mayor a menor número de pacientes en cada grado, con su respectivo porcentaje respecto al total:

```
````{r echo=F}
tabla_pfeiffer <- pfeiffer_m3 %>%
  dplyr::group_by(RERESULTADO_PFEIFFER) %>%
  dplyr::count() %>%
  arrange(desc(n))
tabla_pfeiffer <- data.frame(tabla_pfeiffer)

tabla_pfeiffer <- tabla_pfeiffer %>%
  dplyr::mutate(
    Porcentaje=round(n/sum(n)*100,2)
  )
tabla_pfeiffer
````
```

Con la tabla patologias\_m3, creamos una tabla de frecuencias, ordenada de mayor a menor número de pacientes que sufren cada patología, con su respectivo porcentaje y la prevalencia de cada una de ellas.

```
````{r echo=F}
tabla_patologias <- patologias_m3 %>%
  dplyr::group_by(PATOLOGIA_BPS) %>%
  dplyr::count() %>%
  arrange(desc(n))
tabla_patologias <- data.frame(tabla_patologias)
```

```

tabla_patologias <- tabla_patologias %>%
  dplyr::mutate(
    Porcentaje=round(n/sum(n)*100,2)
  )

```

```

prevalencia <- round(tabla_patologias$n/30000*100,2)
tabla_patologias$prevalencia <- prevalencia
head(tabla_patologias)
****

```

Donde podemos observar que las patologías más frecuentes en nuestros pacientes son Trastorno de ansiedad, Hipertensión, Dislipemia y Artrosis.

Eliminamos Sexo que no sean Hombre o Mujer, de la tabla usuarios_m3.

```

****{r}
usuarios_m4 <- usuarios_m3 %>%
  dplyr::filter(!(SEXO%in%
    c("Desconocido","Indeterminado",
      "No especificado")))
****

```

Opcional; si queremos ocupar menos memoria, borramos la tabla usuarios_m3 inicial y nos quedamos con usuarios_m4.

```

****{r echo=F}
rm(usuarios_m3)

```

Hemos de crear una variable con los pacientes agrupados por edad, la denominaremos GrupoEdad y la colocaremos en la cuarta posición de nuestra tabla, justo antes de la variable edad. Agrupamos los pacientes por GrupoEdad.

```

****{r echo=F}
usuarios_m4$GrupoEdad <-
ifelse(usuarios_m4$EDAD>=0&usuarios_m4$EDAD<15,"[0,15)",
ifelse(usuarios_m4$EDAD>=15&usuarios_m4$EDAD<45,"[15,45)",
ifelse(usuarios_m4$EDAD>=45&usuarios_m4$EDAD<65,"[45,65)",
ifelse(usuarios_m4$EDAD>=65&usuarios_m4$EDAD<=75,"[65,75)",

```

```

ifelse(usuarios_m4$EDAD>=75&usuarios_m4$EDAD<=85,"[75,85)",
ifelse(usuarios_m4$EDAD>=85&usuarios_m4$EDAD<=95,"[85,95)",""+95"))))
usuarios_m4 <- usuarios_m4[with(usuarios_m4, order(usuarios_m4$GrupoEdad)),
c(1:3,27,4:26)]

```

Nos creamos un nuevo data.frame llamado tabla_edad con 9 variables con respecto a la edad de los usuarios registrados en la tabla usuarios_m4, para realizar un pequeño análisis descriptivo:

```

```{r echo=F}
EDADES <- c(summary(usuarios_m4$EDAD), var=var(usuarios_m4$EDAD),
sd=sd(usuarios_m4$EDAD), Cv=sd(usuarios_m4$EDAD)/mean(usuarios_m4$EDAD))
tabla_edad <- data.frame(EDADES)
tabla_edad

```

Vamos a pegar dos tablas (patologias\_m3 y usuarios\_m4) mediante una variable en común, que es la variable n\_orden, con la función "merge", a esta nueva tabla la denominaremos Result2. Así mismo vamos a cambiar el formato de la variable fecha inicio de patología y fecha fin de patología a un número entero, quedándonos solo con el año final y poder hacer la operación correspondiente para obtener la edad inicio de patología y la edad fin de patología.

```

```{r echo=F}
Result2 <- merge(usuarios_m4, patologias_m3, by="n_orden")

```

También es necesario crear una tabla de frecuencias únicamente respecto al grupo de edad, diferenciando entre los 7 grupos de edad de cada uno y añadiendo el número medio de enfermedades crónicas por grupo de edad.

```

```{r echo=F}
tabla_GrupoEdad <- usuarios_m4 %>%
 dplyr::group_by(GrupoEdad) %>%
 dplyr::count()
tabla_GrupoEdad <- data.frame(tabla_GrupoEdad)

tabla_GrupoEdad <- tabla_GrupoEdad %>%
 dplyr::mutate(

```

```

 Porcentaje=round(n/sum(n)*100,2)
)

```

```

tabla_grupoEdad$MediaEnfermedades <-
table(Result2$GrupoEdad)[tabla_grupoEdad$GrupoEdad]/table(usuarios_m4$GrupoEdad)[tabla_grupoEdad$GrupoEdad]
tabla_grupoEdad
````

```

Donde podemos observar que a medida que van cumpliendo años nuestros pacientes, presentan más enfermedades crónicas, salvo los de +95 años por la escasez de casos.

También es necesario crear una tabla de frecuencias únicamente respecto al sexo, diferenciando entre masculino y femenino y añadiendo el número medio de enfermedades crónicas por sexo.

```

````{r echo=F}
tabla_sexo <- usuarios_m4 %>%
 dplyr::group_by(SEXO) %>%
 dplyr::count()
tabla_sexo <- data.frame(tabla_sexo)

tabla_sexo <- tabla_sexo %>%
 dplyr::mutate(
 Porcentaje=round(n/sum(n)*100,2)
)

```

```

tabla_sexo$MediaEnfermedades <-
table(Result2$SEXO)[tabla_sexo$SEXO]/table(usuarios_m4$SEXO)[tabla_sexo$SEXO]
tabla_sexo
````

```

En la cual podemos apreciar que las mujeres poseen un promedio de más de dos enfermedades crónicas por persona, en cambio los hombres no llegan a dos enfermedades.

Vamos a representar una tabla, donde las variables sean el número de Enfermedades que presenta cada paciente según el sexo (Hombre, Mujer), y los

grupos de edad. Para ello, creamos una tabla `N°_Enf`, a partir de nuestra tabla `patologías_m3`, agrupando los `n_orden`, para poder visualizar el número de enfermedades que presenta cada paciente.

```
````{r echo=F}
N_Enf <- patologías_m3 %>%
 dplyr::group_by(n_orden) %>%
 dplyr::count()
head(N_Enf)
````
```

Para obtener los pacientes, que no están registrados en la tabla `patologías_m3`, por lo tanto no presentan ninguna enfermedad, los extraemos de la tabla `usuarios_m4` con el siguiente comando:

```
````{r}
Enf0 <- anti_join(usuarios_m4, N_Enf, by="n_orden")

````
```

Realizamos un filtro de la tabla `N°_Enf`, según tengan 1, 2, 3, 4 o 5 o más enfermedades.

```
````{r echo=F}
Enf1 <- N_Enf %>%
 dplyr::filter(n==1)

Enf2 <- N_Enf %>%
 dplyr::filter(n==2)

Enf3 <- N_Enf %>%
 dplyr::filter(n==3)

Enf4 <- N_Enf %>%
 dplyr::filter(n==4)

Enf5omas <- N_Enf %>%
 dplyr::filter(n>=5)
````
```

Realizamos una primera prueba para ver cómo quedaría nuestro `data.frame` en cuestión, con las variables que hemos obtenido por ahora.

```
````{r echo=F}
```

```

Ntotal_Enf <- data.frame(Variables=c("Número de
pacientes", "Sexo", "Hombres", "Mujeres"),
 Nin_Enf=c(paste0(nrow(Enf0), "
(", round(nrow(Enf0)/nrow(usuarios_m4)*100, 2), "%)"),
 NA, NA, NA),
 Una_Enf=c(paste0(nrow(Enf1), "
(", round(nrow(Enf1)/nrow(usuarios_m4)*100, 2), "%)"),
 NA, NA, NA),
 Dos_Enf=c(paste0(nrow(Enf2), "
(", round(nrow(Enf2)/nrow(usuarios_m4)*100, 2), "%)"),
 NA, NA, NA),
 Tres_Enf=c(paste0(nrow(Enf3), "
(", round(nrow(Enf3)/nrow(usuarios_m4)*100, 2), "%)"),
 NA, NA, NA),
 Cuatro_Enf=c(paste0(nrow(Enf4), "
(", round(nrow(Enf4)/nrow(usuarios_m4)*100, 2), "%)"),
 NA, NA, NA),
 Cincomas_Enf=c(paste0(nrow(Enf5omas), "
(", round(nrow(Enf5omas)/nrow(usuarios_m4)*100, 2), "%)"),
 NA, NA, NA))

```

```
Ntotal_Enf
```

```

```

Es necesario pasar a factor la variable GrupoEdad de la tabla usuarios\_m4, para poder crear la tabla con el número total de enfermedades.

```
****{r echo=F}
```

```
usuarios_m4$GrupoEdad <- factor(usuarios_m4$GrupoEdad, levels =
c("[0,15)", "[15,45)", "[45,65)", "[65,75)", "[75,85)", "[85,95)", "+95"))
```

```

```

Realizamos un filtro de la tabla usuarios\_m4, donde solo aparezcan los pacientes que presentan una enfermedad. Así mismo, volvemos a filtrar la tabla usuarios\_m4, para los pacientes que presentan dos enfermedades, para los pacientes que presentan tres enfermedades, cuatro y 5 o más. Exceptuando la tabla Enf0, ya que los pacientes que no presentan ninguna enfermedad los hemos obtenido de la tabla usuarios\_m4.

```
****{r echo=F}
```

```
Enf1_completo<-usuarios_m4 %>%
```

```
 dplyr::filter(n_orden %in% c(Enf1$n_orden))
```

```
Enf2_completo<-usuarios_m4 %>%
```

```

dplyr::filter(n_orden %in% c(Enf2$n_orden))
Enf3_completo<-usuarios_m4 %>%
 dplyr::filter(n_orden %in% c(Enf3$n_orden))
Enf4_completo<-usuarios_m4 %>%
 dplyr::filter(n_orden %in% c(Enf4$n_orden))
Enf5_completo<-usuarios_m4 %>%
 dplyr::filter(n_orden %in% c(Enf5omas$n_orden))
````

```

También es necesario filtrar según su sexo las tablas que acabamos de obtener.

```

````{r echo=F}
Hombre_G0<-Enf0 %>%
 dplyr::filter(SEXO=="Hombre")
Hombre_G1<-Enf1_completo %>%
 dplyr::filter(SEXO=="Hombre")
Hombre_G2<-Enf2_completo %>%
 dplyr::filter(SEXO=="Hombre")
Hombre_G3<-Enf3_completo %>%
 dplyr::filter(SEXO=="Hombre")
Hombre_G4<-Enf4_completo %>%
 dplyr::filter(SEXO=="Hombre")
Hombre_G5<-Enf5_completo %>%
 dplyr::filter(SEXO=="Hombre")
Mujer_G0<-Enf0 %>%
 dplyr::filter(SEXO=="Mujer")
Mujer_G1<-Enf1_completo %>%
 dplyr::filter(SEXO=="Mujer")
Mujer_G2<-Enf2_completo %>%
 dplyr::filter(SEXO=="Mujer")
Mujer_G3<-Enf3_completo %>%
 dplyr::filter(SEXO=="Mujer")
Mujer_G4<-Enf4_completo %>%
 dplyr::filter(SEXO=="Mujer")
Mujer_G5<-Enf5_completo %>%
 dplyr::filter(SEXO=="Mujer")
````

```

Podemos observar cómo quedaría el data.frame de número total de pacientes que presentan 1, 2, 3, 4 y 5 o más enfermedades según su sexo.

```

````{r echo=F}
Ntotal_Enf <- data.frame(Variables=c("Número
pacientes", "Sexo", "Hombres", "Mujeres", "Grupos Edad"),
 Nin_Enf=c(paste0(nrow(Enf0), "
(", round(nrow(Enf0)/nrow(usuarios_m4)*100,2), "%)"),
 NA,
 paste0(nrow(Hombre_G0), "
(", round(nrow(Hombre_G0)/nrow(usuarios_m4)*100,2), "%)"),
 paste0(nrow(Mujer_G0), "
(", round(nrow(Mujer_G0)/nrow(usuarios_m4)*100,2), "%)"),
 NA),
 Una_Enf=c(paste0(nrow(Enf1), "
(", round(nrow(Enf1)/nrow(usuarios_m4)*100,2), "%)"),
 NA,
 paste0(nrow(Hombre_G1), "
(", round(nrow(Hombre_G1)/nrow(usuarios_m4)*100,2), "%)"),
 paste0(nrow(Mujer_G1), "
(", round(nrow(Mujer_G1)/nrow(usuarios_m4)*100,2), "%)"),
 NA),
 Dos_Enf=c(paste0(nrow(Enf2), "
(", round(nrow(Enf2)/nrow(usuarios_m4)*100,2), "%)"),
 NA,
 paste0(nrow(Hombre_G2), "
(", round(nrow(Hombre_G2)/nrow(usuarios_m4)*100,2), "%)"),
 paste0(nrow(Mujer_G2), "
(", round(nrow(Mujer_G2)/nrow(usuarios_m4)*100,2), "%)"),
 NA),
 Tres_Enf=c(paste0(nrow(Enf3), "
(", round(nrow(Enf3)/nrow(usuarios_m4)*100,2), "%)"),
 NA,
 paste0(nrow(Hombre_G3), "
(", round(nrow(Hombre_G3)/nrow(usuarios_m4)*100,2), "%)"),
 paste0(nrow(Mujer_G3), "
(", round(nrow(Mujer_G3)/nrow(usuarios_m4)*100,2), "%)"),
 NA),
 Cuatro_Enf=c(paste0(nrow(Enf4), "
(", round(nrow(Enf4)/nrow(usuarios_m4)*100,2), "%)"),
 NA,

```

```

paste0(nrow(Hombre_G4), "
(", round(nrow(Hombre_G4)/nrow(usuarios_m4)*100, 2), "%)"),
paste0(nrow(Mujer_G4), "
(", round(nrow(Mujer_G4)/nrow(usuarios_m4)*100, 2), "%)"),
NA),
Cincomas_Enf=c(paste0(nrow(Enf5omas), "
(", round(nrow(Enf5omas)/nrow(usuarios_m4)*100, 2), "%)"),
NA,
paste0(nrow(Hombre_G5), "
(", round(nrow(Hombre_G5)/nrow(usuarios_m4)*100, 2), "%)"),
paste0(nrow(Mujer_G5), "
(", round(nrow(Mujer_G5)/nrow(usuarios_m4)*100, 2), "%)"),
NA))

```

Ntotal\_Enf

```

```

Agrupamos las tablas obtenidas de los pacientes con 0, 1, 2, 3, 4 y 5 o más enfermedades según su grupo de edad.

```

****{r echo=F}

```

```

Grupo0Enf_Edades<-Enf0 %>%
 dplyr::group_by(GrupoEdad) %>%
 dplyr::count()
Grupo1Enf_Edades<-Enf1_completo %>%
 dplyr::group_by(GrupoEdad) %>%
 dplyr::count()
Grupo2Enf_Edades<-Enf2_completo %>%
 dplyr::group_by(GrupoEdad) %>%
 dplyr::count()
Grupo3Enf_Edades<-Enf3_completo %>%
 dplyr::group_by(GrupoEdad) %>%
 dplyr::count()
Grupo4Enf_Edades<-Enf4_completo %>%
 dplyr::group_by(GrupoEdad) %>%
 dplyr::count()
Grupo5Enf_Edades<-Enf5_completo %>%
 dplyr::group_by(GrupoEdad) %>%
 dplyr::count()

```

Por último, realizamos el data.frame correspondiente para generar la tabla de número total de pacientes que presentan 0, 1, 2, 3, 4 y 5 o más enfermedades según su sexo y su grupo de edad.

```

````{r echo=F}
Ntotal_Enf          <-          data.frame(Variables=c("Número          de
pacientes","Sexo","Hombres","Mujeres","Grupos  Edad","[0-15)  años","[15-45)
años","[45-65) años","[65-75) años","[75-85) años","[85-95) años","+95 años"),
          Nin_Enf=c(paste0(nrow(Enf0),"
(",round(nrow(Enf0)/nrow(usuarios_m4)*100,2),"%")),
          NA,
          paste0(nrow(Hombre_G0),"
(",round(nrow(Hombre_G0)/nrow(usuarios_m4)*100,2),"%")),
          paste0(nrow(Mujer_G0),"
(",round(nrow(Mujer_G0)/nrow(usuarios_m4)*100,2),"%")),
          NA,
          paste0(Grupo0Enf_Edades[1,2],"
(",round(Grupo0Enf_Edades[1,2]/nrow(usuarios_m4)*100,2),"%")),
          paste0(Grupo0Enf_Edades[2,2],"
(",round(Grupo0Enf_Edades[2,2]/nrow(usuarios_m4)*100,2),"%")),
          paste0(Grupo0Enf_Edades[3,2],"
(",round(Grupo3Enf_Edades[3,2]/nrow(usuarios_m4)*100,2),"%")),
          paste0(Grupo0Enf_Edades[4,2],"
(",round(Grupo0Enf_Edades[4,2]/nrow(usuarios_m4)*100,2),"%")),
          paste0(Grupo0Enf_Edades[5,2],"
(",round(Grupo0Enf_Edades[5,2]/nrow(usuarios_m4)*100,2),"%")),
          paste0(Grupo0Enf_Edades[6,2],"
(",round(Grupo0Enf_Edades[6,2]/nrow(usuarios_m4)*100,2),"%")),
          paste0(Grupo0Enf_Edades[7,2],"
(",round(Grupo0Enf_Edades[7,2]/nrow(usuarios_m4)*100,2),"%"))),
          Una_Enf=c(paste0(nrow(Enf1),"
(",round(nrow(Enf1)/nrow(usuarios_m4)*100,2),"%")),
          NA,
          paste0(nrow(Hombre_G1),"
(",round(nrow(Hombre_G1)/nrow(usuarios_m4)*100,2),"%")),
          paste0(nrow(Mujer_G1),"
(",round(nrow(Mujer_G1)/nrow(usuarios_m4)*100,2),"%")),
          NA,
          paste0(Grupo1Enf_Edades[1,2],"
(",round(Grupo1Enf_Edades[1,2]/nrow(usuarios_m4)*100,2),"%")),

```

```

paste0(Grupo1Enf_Edades[2,2], "
(", round(Grupo1Enf_Edades[2,2]/nrow(usuarios_m4)*100,2), "%) " ,
paste0(Grupo1Enf_Edades[3,2], "
(", round(Grupo1Enf_Edades[3,2]/nrow(usuarios_m4)*100,2), "%) " ,
paste0(Grupo1Enf_Edades[4,2], "
(", round(Grupo1Enf_Edades[4,2]/nrow(usuarios_m4)*100,2), "%) " ,
paste0(Grupo1Enf_Edades[5,2], "
(", round(Grupo1Enf_Edades[5,2]/nrow(usuarios_m4)*100,2), "%) " ,
paste0(Grupo1Enf_Edades[6,2], "
(", round(Grupo1Enf_Edades[6,2]/nrow(usuarios_m4)*100,2), "%) " ,
paste0(Grupo1Enf_Edades[7,2], "
(", round(Grupo1Enf_Edades[7,2]/nrow(usuarios_m4)*100,2), "%) " ) ,
Dos_Enf=c(paste0(nrow(Enf2), "
(", round(nrow(Enf2)/nrow(usuarios_m4)*100,2), "%) " ) ,
NA,
paste0(nrow(Hombre_G2), "
(", round(nrow(Hombre_G2)/nrow(usuarios_m4)*100,2), "%) " ) ,
paste0(nrow(Mujer_G2), "
(", round(nrow(Mujer_G2)/nrow(usuarios_m4)*100,2), "%) " ) ,
NA,
paste0(Grupo2Enf_Edades[1,2], "
(", round(Grupo2Enf_Edades[1,2]/nrow(usuarios_m4)*100,2), "%) " ,
paste0(Grupo2Enf_Edades[2,2], "
(", round(Grupo2Enf_Edades[2,2]/nrow(usuarios_m4)*100,2), "%) " ,
paste0(Grupo2Enf_Edades[3,2], "
(", round(Grupo2Enf_Edades[3,2]/nrow(usuarios_m4)*100,2), "%) " ,
paste0(Grupo2Enf_Edades[4,2], "
(", round(Grupo2Enf_Edades[4,2]/nrow(usuarios_m4)*100,2), "%) " ,
paste0(Grupo2Enf_Edades[5,2], "
(", round(Grupo2Enf_Edades[5,2]/nrow(usuarios_m4)*100,2), "%) " ,
paste0(Grupo2Enf_Edades[6,2], "
(", round(Grupo2Enf_Edades[6,2]/nrow(usuarios_m4)*100,2), "%) " ,
paste0(Grupo2Enf_Edades[7,2], "
(", round(Grupo2Enf_Edades[7,2]/nrow(usuarios_m4)*100,2), "%) " ) ,
Tres_Enf=c(paste0(nrow(Enf3), "
(", round(nrow(Enf3)/nrow(usuarios_m4)*100,2), "%) " ) ,
NA,
paste0(nrow(Hombre_G3), "
(", round(nrow(Hombre_G3)/nrow(usuarios_m4)*100,2), "%) " ) ,

```

```

paste0(nrow(Mujer_G3), "
(", round(nrow(Mujer_G3)/nrow(usuarios_m4)*100,2), "%)"),
NA,
paste0(Grupo3Enf_Edades[1,2], "
(", round(Grupo3Enf_Edades[1,2]/nrow(usuarios_m4)*100,2), "%)"),
paste0(Grupo3Enf_Edades[2,2], "
(", round(Grupo3Enf_Edades[2,2]/nrow(usuarios_m4)*100,2), "%)"),
paste0(Grupo3Enf_Edades[3,2], "
(", round(Grupo3Enf_Edades[3,2]/nrow(usuarios_m4)*100,2), "%)"),
paste0(Grupo3Enf_Edades[4,2], "
(", round(Grupo3Enf_Edades[4,2]/nrow(usuarios_m4)*100,2), "%)"),
paste0(Grupo3Enf_Edades[5,2], "
(", round(Grupo3Enf_Edades[5,2]/nrow(usuarios_m4)*100,2), "%)"),
paste0(Grupo3Enf_Edades[6,2], "
(", round(Grupo3Enf_Edades[6,2]/nrow(usuarios_m4)*100,2), "%)"),
paste0(Grupo3Enf_Edades[7,2], "
(", round(Grupo3Enf_Edades[7,2]/nrow(usuarios_m4)*100,2), "%)"),
Cuatro_Enf=c(paste0(nrow(Enf4), "
(", round(nrow(Enf4)/nrow(usuarios_m4)*100,2), "%)"),
NA,
paste0(nrow(Hombre_G4), "
(", round(nrow(Hombre_G4)/nrow(usuarios_m4)*100,2), "%)"),
paste0(nrow(Mujer_G4), "
(", round(nrow(Mujer_G4)/nrow(usuarios_m4)*100,2), "%)"),
NA,
paste0(Grupo4Enf_Edades[1,2], "
(", round(Grupo4Enf_Edades[1,2]/nrow(usuarios_m4)*100,2), "%)"),
paste0(Grupo4Enf_Edades[2,2], "
(", round(Grupo4Enf_Edades[2,2]/nrow(usuarios_m4)*100,2), "%)"),
paste0(Grupo4Enf_Edades[3,2], "
(", round(Grupo4Enf_Edades[3,2]/nrow(usuarios_m4)*100,2), "%)"),
paste0(Grupo4Enf_Edades[4,2], "
(", round(Grupo4Enf_Edades[4,2]/nrow(usuarios_m4)*100,2), "%)"),
paste0(Grupo4Enf_Edades[5,2], "
(", round(Grupo4Enf_Edades[5,2]/nrow(usuarios_m4)*100,2), "%)"),
paste0(Grupo4Enf_Edades[6,2], "
(", round(Grupo4Enf_Edades[6,2]/nrow(usuarios_m4)*100,2), "%)"),
paste0(Grupo4Enf_Edades[7,2], "
(", round(Grupo4Enf_Edades[7,2]/nrow(usuarios_m4)*100,2), "%)"),

```



```

Cincomas_Enf=c(paste0(nrow(Enf5omas),"
(",round(nrow(Enf5omas)/nrow(usuarios_m4)*100,2),"%")),
              NA,
              paste0(nrow(Hombre_G5),"
(",round(nrow(Hombre_G5)/nrow(usuarios_m4)*100,2),"%")),
              paste0(nrow(Mujer_G5),"
(",round(nrow(Mujer_G5)/nrow(usuarios_m4)*100,2),"%")),
              NA,
              paste0(Grupo5Enf_Edades[1,2],"
(",round(Grupo5Enf_Edades[1,2]/nrow(usuarios_m4)*100,2),"%")),
              paste0(Grupo5Enf_Edades[2,2],"
(",round(Grupo5Enf_Edades[2,2]/nrow(usuarios_m4)*100,2),"%")),
              paste0(Grupo5Enf_Edades[3,2],"
(",round(Grupo5Enf_Edades[3,2]/nrow(usuarios_m4)*100,2),"%")),
              paste0(Grupo5Enf_Edades[4,2],"
(",round(Grupo5Enf_Edades[4,2]/nrow(usuarios_m4)*100,2),"%")),
              paste0(Grupo5Enf_Edades[5,2],"
(",round(Grupo5Enf_Edades[5,2]/nrow(usuarios_m4)*100,2),"%")),
              paste0(Grupo5Enf_Edades[6,2],"
(",round(Grupo5Enf_Edades[6,2]/nrow(usuarios_m4)*100,2),"%")),
              paste0(Grupo5Enf_Edades[7,2],"
(",round(Grupo5Enf_Edades[7,2]/nrow(usuarios_m4)*100,2),"%"))))
Ntotal_Enf
.....

```

En esta tabla se aprecia como la mayoría de nuestros pacientes tienen 0 o 1 enfermedad crónica, no obstante más de un 14% presenta 5 enfermedades o más. En cuanto al sexo, hay más hombres que presentan 0 o 1 enfermedad, pero a partir de ahí es en las mujeres donde predomina tener 2, 3, 4 y 5 o más enfermedades crónicas. En los grupos de edad podemos observar que lo normal en niños de 0 a 15 años es no presentar ninguna patología crónica. Entre los pacientes de 15 a 45 años cabe destacar que un alto porcentaje se encuentra con 1 o ninguna enfermedad crónica. De los 45 a los 65 años está muy repartido el número de enfermedades crónicas de cada paciente. A partir de los 65 años, la mayoría de los pacientes registrados en la BPS presentan 5 enfermedades o más.

Ilustración de barras según el número de pacientes con 0, 1, 2, 3, 4 y 5 o más enfermedades.

```

````{r echo=F}
tabla_gra <- data.frame(grupos=c("0 Enf","1 Enf","2 Enf","3 Enf","4 Enf","5
Enf"),

n=c(nrow(Enf0),nrow(Enf1_completo),nrow(Enf2_completo),nrow(Enf3_completo),nr
ow(Enf4_completo),nrow(Enf5_completo)))

tabla_gra$grupos <- factor(tabla_gra$grupos)

ggplot(tabla_gra,aes(x=grupos,y=n))+
 geom_histogram(stat = "identity")+
 labs(y="Usuarios",x="N° de Enfermedades")+
 theme(text=element_text(family="Arial", size=15),
 panel.background = element_rect(fill = "white"),
 panel.grid.minor = element_line(color = "#c6c6c6", size = 0.1,linetype
= 'dashed'),
 axis.line.x = element_line(colour = '#c6c6c6', size = 0.1))
````

```

También es necesario crear una tabla de frecuencias respecto al sexo de los pacientes y su grupo de edad correspondiente, incluidos en la base de datos poblacional, diferenciando entre hombres y mujeres y los 7 grupos de edad de cada uno, cuyo porcentaje sería respecto al total de la población a estudiar.

```

````{r echo=F}
tabla_sexoEdad <- usuarios_m4 %>%
 dplyr::group_by(SEXO,GrupoEdad) %>%
 dplyr::count()
tabla_sexoEdad <- data.frame(tabla_sexoEdad)

tabla_sexoEdad <- tabla_sexoEdad %>%
 dplyr::mutate(
 Porcentaje=round(n/sum(n)*100,2)
)
tabla_sexoEdad
````

```

Se puede observar como la mayoría de nuestros usuarios, tanto hombres como mujeres, se encuentran entre los 15 y 65 años de edad.

Hacemos un filtro por sexo de la tabla usuarios_m4.

```
````{r echo=F}
hombres_m4 <- usuarios_m4 %>%
 dplyr::filter(SEXO=="Hombre")

mujeres_m4 <- usuarios_m4 %>%
 dplyr::filter(SEXO=="Mujer")
```

````

Vamos a crear una tabla de frecuencias respecto al sexo masculino, según su grupo de edad.

```
````{r echo=F}
tabla_hombresEdad <- hombres_m4 %>%
 dplyr::group_by(GrupoEdad) %>%
 dplyr::count()
tabla_hombresEdad <- data.frame(tabla_hombresEdad)
```

```
tabla_hombresEdad <- tabla_hombresEdad %>%
 dplyr::mutate(
 Porcentaje=round(n/sum(n)*100,2)
)
```

```
tabla_hombresEdad
```

````

Histograma de la población masculina por grupo de edad.

```
````{r echo=F}
tabla_hombresEdad$GrupoEdad <- factor(tabla_hombresEdad$GrupoEdad)

ggplot(tabla_hombresEdad, aes(x=GrupoEdad, y=n)) +
 geom_histogram(stat = "identity") +
 labs(y="Personas", x="Grupo de edades") +
 theme(text=element_text(family="Arial", size=15),
 panel.background = element_rect(fill = "white"),
 panel.grid.minor = element_line(color = "#c6c6c6", size = 0.1, linetype
= 'dashed'),
 axis.line.x = element_line(colour = '#c6c6c6', size = 0.1))
```

````

Vamos a crear una tabla de frecuencias respecto al sexo femenino, según su grupo de edad.

```
````{r echo=F}
tabla_mujerEdad <- mujeres_m4 %>%
 dplyr::group_by(GrupoEdad) %>%
 dplyr::count()
tabla_mujerEdad <- data.frame(tabla_mujerEdad)
```

```
tabla_mujerEdad <- tabla_mujerEdad %>%
 dplyr::mutate(
 Porcentaje=round(n/sum(n)*100,2)
)
```

```
tabla_mujerEdad
````
```

Histograma de la población femenina por grupo de edad.

```
````{r echo=F}
tabla_mujerEdad$GrupoEdad <- factor(tabla_mujerEdad$GrupoEdad)

ggplot(tabla_mujerEdad,aes(x=GrupoEdad,y=n))+
 geom_histogram(stat = "identity")+
 labs(y="Personas",x="Grupo de edades")+
 theme(text=element_text(family="Arial", size=15),
 panel.background = element_rect(fill = "white"),
 panel.grid.minor = element_line(color = "#c6c6c6", size = 0.1,linetype
= 'dashed'),
 axis.line.x = element_line(colour = '#c6c6c6', size = 0.1))
````
```

```
````{r echo=F}
hist(usuarios_m4$CONSULTAS_AP_TOTALES, main = "Población según consultas de
atención primaria",
 xlab = "Consultas AP", ylab = "usuarios", breaks =
c(0,20,40,60,80,100,120,140,160,180,200,220,240,
260,280,300,320,340,360,380,400,420,447),
 col = "lightgreen")
````
```

En esta ilustración podemos observar, que la mayoría de usuarios están en el intervalo (0,20), en cuanto a consultas de atención primaria totales.

```

````{r echo=F}
tabla_Aitor<-Result2 %>%
 dplyr::group_by(n_orden) %>%
 dplyr::count()
tabla_Aitor2<-tabla_Aitor %>%
 dplyr::group_by(n) %>%
 dplyr::count()
tabla_Aitor2$Porcentaje<-tabla_Aitor2$nn/nrow(tabla_Aitor)*100

Personas_1<-tabla_Aitor %>%
 dplyr::filter(n==1)
Personas_1_pato<-Result2 %>%
 dplyr::filter(n_orden%in%c(Personas_1$n_orden)) %>%
 dplyr::group_by(PATOLOGIA_BPS) %>%
 dplyr::count() %>%
 dplyr::arrange(desc(n)) %>%
 dplyr::mutate(Porcentaje=n/nrow(Personas_1)*100)
Personas_1_pato_final<-head(Personas_1_pato,5)
Personas_1_pato_final

Personas_2<-tabla_Aitor %>%
 dplyr::filter(n==2)
Personas_2_pato<-Result2 %>%
 dplyr::filter(n_orden%in%c(Personas_2$n_orden)) %>%
 dplyr::select(n_orden, PATOLOGIA_BPS)

Personas_2_pato_1<-Personas_2_pato %>%
 dplyr::filter(n_orden%in%Personas_2$n_orden[1])
Personas_2_pato_1$Patologias_Unidas<-NA
Personas_2_pato_1$Patologias_Unidas<-
paste(sort(c(Personas_2_pato_1$PATOLOGIA_BPS[1],Personas_2_pato_1$PATOLOGIA_B
PS[2])) [1],sort(c(Personas_2_pato_1$PATOLOGIA_BPS[1],Personas_2_pato_1$PATOLO
GIA_BPS[2])) [2],sep=" y ")
Personas_2_pato_1<-Personas_2_pato_1 %>%
 dplyr::select(n_orden,Patologias_Unidas) %>%
 unique()

for(i in 2:nrow(Personas_2)){
 Personas_2_pato_prueba2<-Personas_2_pato %>%

```

```

 dplyr::filter(n_orden%in%Personas_2$n_orden[i])
 Personas_2_pato_prueba2$Patologias_Unidas<-NA
 Personas_2_pato_prueba2$Patologias_Unidas<-
paste(sort(c(Personas_2_pato_prueba2$PATOLOGIA_BPS[1],Personas_2_pato_prueba2
$PATOLOGIA_BPS[2])) [1],sort(c(Personas_2_pato_prueba2$PATOLOGIA_BPS[1],Person
as_2_pato_prueba2$PATOLOGIA_BPS[2])) [2],sep=" y ")
 Personas_2_pato_prueba2<-Personas_2_pato_prueba2 %>%
 dplyr::select(n_orden,Patologias_Unidas) %>%
 unique()
 Personas_2_pato_1<-union_all(Personas_2_pato_1,Personas_2_pato_prueba2)
}
Personas_2_pato<-Personas_2_pato_1 %>%
 dplyr::group_by(Patologias_Unidas) %>%
 dplyr::count() %>%
 dplyr::arrange(desc(n)) %>%
 dplyr::mutate(Porcentaje=n/nrow(Personas_2)*100)
Personas_2_pato_final<-head(Personas_2_pato,5)

```

```

Personas_3<-tabla_Aitor %>%
 dplyr::filter(n==3)
Personas_3_pato<-Result2 %>%
 dplyr::filter(n_orden%in%c(Personas_3$n_orden)) %>%
 dplyr::select(n_orden,PATOLOGIA_BPS)

```

```

Personas_3_pato_1<-Personas_3_pato %>%
 dplyr::filter(n_orden%in%Personas_3$n_orden[1])
Personas_3_pato_1$Patologias_Unidas<-NA
Personas_3_pato_1$Patologias_Unidas<-
paste(paste(sort(c(Personas_3_pato_1$PATOLOGIA_BPS[1],Personas_3_pato_1$PATOL
OGIA_BPS[2],Personas_3_pato_1$PATOLOGIA_BPS[3])) [1],sort(c(Personas_3_pato_1$
PATOLOGIA_BPS[1],Personas_3_pato_1$PATOLOGIA_BPS[2],Personas_3_pato_1$PATOLOG
IA_BPS[3])) [2],sep="
"),sort(c(Personas_3_pato_1$PATOLOGIA_BPS[1],Personas_3_pato_1$PATOLOGIA_BPS[
2],Personas_3_pato_1$PATOLOGIA_BPS[3])) [3],sep=" y ")
Personas_3_pato_1<-Personas_3_pato_1 %>%
 dplyr::select(n_orden,Patologias_Unidas) %>%
 unique()

```

```

for(i in 2:nrow(Personas_3)){
 Personas_3_pato_prueba2<-Personas_3_pato %>%
 dplyr::filter(n_orden%in%Personas_3$n_orden[i])
 Personas_3_pato_prueba2$Patologias_Unidas<-NA
 Personas_3_pato_prueba2$Patologias_Unidas<-
paste(paste(sort(c(Personas_3_pato_prueba2$PATOLOGIA_BPS[1],Personas_3_pato_p
rueba2$PATOLOGIA_BPS[2],Personas_3_pato_prueba2$PATOLOGIA_BPS[3]))[1],sort(c(
Personas_3_pato_prueba2$PATOLOGIA_BPS[1],Personas_3_pato_prueba2$PATOLOGIA_BP
S[2],Personas_3_pato_prueba2$PATOLOGIA_BPS[3]))[2],sep="
"),sort(c(Personas_3_pato_prueba2$PATOLOGIA_BPS[1],Personas_3_pato_prueba2$PA
TOLOGIA_BPS[2],Personas_3_pato_prueba2$PATOLOGIA_BPS[3]))[3],sep=" y ")
 Personas_3_pato_prueba2<-Personas_3_pato_prueba2 %>%
 dplyr::select(n_orden,Patologias_Unidas) %>%
 unique()
 Personas_3_pato_1<-union_all(Personas_3_pato_1,Personas_3_pato_prueba2)
}
Personas_3_pato<-Personas_3_pato_1 %>%
 dplyr::group_by(Patologias_Unidas) %>%
 dplyr::count() %>%
 dplyr::arrange(desc(n)) %>%
 dplyr::mutate(Porcentaje=n/nrow(Personas_3)*100)
Personas_3_pato_final<-head(Personas_3_pato,5)

Personas_4<-tabla_Aitor %>%
 dplyr::filter(n==4)
Personas_4_pato<-Result2 %>%
 dplyr::filter(n_orden%in%c(Personas_4$n_orden)) %>%
 dplyr::select(n_orden,PATOLOGIA_BPS)

Personas_4_pato_1<-Personas_4_pato %>%
 dplyr::filter(n_orden%in%Personas_4$n_orden[1])
Personas_4_pato_1$Patologias_Unidas<-NA
Personas_4_pato_1$Patologias_Unidas<-
paste(paste(sort(c(Personas_4_pato_1$PATOLOGIA_BPS[1],Personas_4_pato_1$PATOL
OGIA_BPS[2],Personas_4_pato_1$PATOLOGIA_BPS[3],Personas_4_pato_1$PATOLOGIA_BP
S[4]))[1],

```

```

sort(c(Personas_4_pato_1$PATOLOGIA_BPS[1],Personas_4_pato_1$PATOLOGIA_BPS[2],
Personas_4_pato_1$PATOLOGIA_BPS[3],Personas_4_pato_1$PATOLOGIA_BPS[4])) [2],

sort(c(Personas_4_pato_1$PATOLOGIA_BPS[1],Personas_4_pato_1$PATOLOGIA_BPS[2],
Personas_4_pato_1$PATOLOGIA_BPS[3],Personas_4_pato_1$PATOLOGIA_BPS[4])) [3],se
p=" , "),

sort(c(Personas_4_pato_1$PATOLOGIA_BPS[1],Personas_4_pato_1$PATOLOGIA_BPS[2],
Personas_4_pato_1$PATOLOGIA_BPS[3],Personas_4_pato_1$PATOLOGIA_BPS[4])) [4],se
p=" y ")
Personas_4_pato_1<-Personas_4_pato_1 %>%
 dplyr::select(n_orden,Patologias_Unidas) %>%
 unique()

for(i in 2:nrow(Personas_4)){
 Personas_4_pato_prueba2<-Personas_4_pato %>%
 dplyr::filter(n_orden%in%Personas_4$n_orden[i])
 Personas_4_pato_prueba2$Patologias_Unidas<-NA
 Personas_4_pato_prueba2$Patologias_Unidas<-
paste(paste(sort(c(Personas_4_pato_prueba2$PATOLOGIA_BPS[1],Personas_4_pato_p
rueba2$PATOLOGIA_BPS[2],Personas_4_pato_prueba2$PATOLOGIA_BPS[3],Personas_4_p
ato_prueba2$PATOLOGIA_BPS[4])) [1],

sort(c(Personas_4_pato_prueba2$PATOLOGIA_BPS[1],Personas_4_pato_prueba2$PATOL
OGIA_BPS[2],Personas_4_pato_prueba2$PATOLOGIA_BPS[3],Personas_4_pato_prueba2$
PATOLOGIA_BPS[4])) [2],

sort(c(Personas_4_pato_prueba2$PATOLOGIA_BPS[1],Personas_4_pato_prueba2$PATOL
OGIA_BPS[2],Personas_4_pato_prueba2$PATOLOGIA_BPS[3],Personas_4_pato_prueba2$
PATOLOGIA_BPS[4])) [3],sep=" , "),

sort(c(Personas_4_pato_prueba2$PATOLOGIA_BPS[1],Personas_4_pato_prueba2$PATOL
OGIA_BPS[2],Personas_4_pato_prueba2$PATOLOGIA_BPS[3],Personas_4_pato_prueba2$
PATOLOGIA_BPS[4])) [4],sep=" y ")
 Personas_4_pato_prueba2<-Personas_4_pato_prueba2 %>%
 dplyr::select(n_orden,Patologias_Unidas) %>%
 unique()
 Personas_4_pato_1<-union_all(Personas_4_pato_1,Personas_4_pato_prueba2)

```



```

}
Personas_4_pato<-Personas_4_pato_1 %>%
 dplyr::group_by(Patologias_Unidas) %>%
 dplyr::count() %>%
 dplyr::arrange(desc(n)) %>%
 dplyr::mutate(Porcentaje=n/nrow(Personas_4)*100)
Personas_4_pato_final<-head(Personas_4_pato,5)

Personas_5<-tabla_Aitor %>%
 dplyr::filter(n==5)
Personas_5_pato<-Result2 %>%
 dplyr::filter(n_orden%in%c(Personas_5$n_orden)) %>%
 dplyr::select(n_orden,PATOLOGIA_BPS)

Personas_5_pato_1<-Personas_5_pato %>%
 dplyr::filter(n_orden%in%Personas_5$n_orden[1])
Personas_5_pato_1$Patologias_Unidas<-NA
Personas_5_pato_1$Patologias_Unidas<-
paste(paste(sort(c(Personas_5_pato_1$PATOLOGIA_BPS[1],Personas_5_pato_1$PATOL
OGIA_BPS[2],Personas_5_pato_1$PATOLOGIA_BPS[3],Personas_5_pato_1$PATOLOGIA_BP
S[4],Personas_5_pato_1$PATOLOGIA_BPS[5])) [1],

sort(c(Personas_5_pato_1$PATOLOGIA_BPS[1],Personas_5_pato_1$PATOLOGIA_BPS[2],
Personas_5_pato_1$PATOLOGIA_BPS[3],Personas_5_pato_1$PATOLOGIA_BPS[4],Persona
s_5_pato_1$PATOLOGIA_BPS[5])) [2],

sort(c(Personas_5_pato_1$PATOLOGIA_BPS[1],Personas_5_pato_1$PATOLOGIA_BPS[2],
Personas_5_pato_1$PATOLOGIA_BPS[3],Personas_5_pato_1$PATOLOGIA_BPS[4],Persona
s_5_pato_1$PATOLOGIA_BPS[5])) [3],

sort(c(Personas_5_pato_1$PATOLOGIA_BPS[1],Personas_5_pato_1$PATOLOGIA_BPS[2],
Personas_5_pato_1$PATOLOGIA_BPS[3],Personas_5_pato_1$PATOLOGIA_BPS[4],Persona
s_5_pato_1$PATOLOGIA_BPS[5])) [4],sep=" , "),

sort(c(Personas_5_pato_1$PATOLOGIA_BPS[1],Personas_5_pato_1$PATOLOGIA_BPS[2],
Personas_5_pato_1$PATOLOGIA_BPS[3],Personas_5_pato_1$PATOLOGIA_BPS[4],Persona
s_5_pato_1$PATOLOGIA_BPS[5])) [5],sep=" y ")
Personas_5_pato_1<-Personas_5_pato_1 %>%

```

```

dplyr::select(n_orden,Patologias_Unidas) %>%
unique()

for(i in 2:nrow(Personas_5)){
 Personas_5_pato_prueba2<-Personas_5_pato %>%
 dplyr::filter(n_orden%in%Personas_5$n_orden[i])
 Personas_5_pato_prueba2$Patologias_Unidas<-NA
 Personas_5_pato_prueba2$Patologias_Unidas<-
paste(paste(sort(c(Personas_5_pato_prueba2$PATOLOGIA_BPS[1],Personas_5_pato_p
rueba2$PATOLOGIA_BPS[2],Personas_5_pato_prueba2$PATOLOGIA_BPS[3],Personas_5_p
ato_prueba2$PATOLOGIA_BPS[4],Personas_5_pato_prueba2$PATOLOGIA_BPS[5])) [1],

sort(c(Personas_5_pato_prueba2$PATOLOGIA_BPS[1],Personas_5_pato_prueba2$PATOL
OGIA_BPS[2],Personas_5_pato_prueba2$PATOLOGIA_BPS[3],Personas_5_pato_prueba2$
PATOLOGIA_BPS[4],Personas_5_pato_prueba2$PATOLOGIA_BPS[5])) [2],

sort(c(Personas_5_pato_prueba2$PATOLOGIA_BPS[1],Personas_5_pato_prueba2$PATOL
OGIA_BPS[2],Personas_5_pato_prueba2$PATOLOGIA_BPS[3],Personas_5_pato_prueba2$
PATOLOGIA_BPS[4],Personas_5_pato_prueba2$PATOLOGIA_BPS[5])) [3],

sort(c(Personas_5_pato_prueba2$PATOLOGIA_BPS[1],Personas_5_pato_prueba2$PATOL
OGIA_BPS[2],Personas_5_pato_prueba2$PATOLOGIA_BPS[3],Personas_5_pato_prueba2$
PATOLOGIA_BPS[4],Personas_5_pato_prueba2$PATOLOGIA_BPS[5])) [4],sep=" , "),

sort(c(Personas_5_pato_prueba2$PATOLOGIA_BPS[1],Personas_5_pato_prueba2$PATOL
OGIA_BPS[2],Personas_5_pato_prueba2$PATOLOGIA_BPS[3],Personas_5_pato_prueba2$
PATOLOGIA_BPS[4],Personas_5_pato_prueba2$PATOLOGIA_BPS[5])) [5],sep=" y ")
 Personas_5_pato_prueba2<-Personas_5_pato_prueba2 %>%
 dplyr::select(n_orden,Patologias_Unidas) %>%
 unique()
 Personas_5_pato_1<-union_all(Personas_5_pato_1,Personas_5_pato_prueba2)
}
Personas_5_pato<-Personas_5_pato_1 %>%
 dplyr::group_by(Patologias_Unidas) %>%
 dplyr::count() %>%
 dplyr::arrange(desc(n)) %>%
 dplyr::mutate(Porcentaje=n/nrow(Personas_5)*100)
Personas_5_pato_final<-head(Personas_5_pato,5)

```

```

colnames(Personas_1_pato_final)<-colnames(Personas_5_pato_final)
Personas_1_pato_final<-data.frame(Personas_1_pato_final)
Personas_2_pato_final<-data.frame(Personas_2_pato_final)
Personas_3_pato_final<-data.frame(Personas_3_pato_final)
Personas_4_pato_final<-data.frame(Personas_4_pato_final)
Personas_5_pato_final<-data.frame(Personas_5_pato_final)

Tabla_final<-union_all(Personas_1_pato_final,Personas_2_pato_final)
Tabla_final<-union_all(Tabla_final,Personas_3_pato_final)
Tabla_final<-union_all(Tabla_final,Personas_4_pato_final)
Tabla_final<-union_all(Tabla_final,Personas_5_pato_final)
Tabla_final$Patologias<-c(rep(1,5),rep(2,5),rep(3,5),rep(4,5),rep(5,5))
Tabla_final<-Tabla_final %>%
 dplyr::select(Patologias,Patologias_Unidas,n,Porcentaje)
Tabla_final
````

```

Vamos a analizar las patologías o la combinación de patologías más frecuentes según el nº de patologías que tenga una persona. En nuestra base de datos tenemos 19.406 personas con al menos una patología. A continuación se muestra una tabla, donde se observa

#####ANÁLISIS DE SECUENCIAS#####

Primero realizamos un análisis de secuencias únicamente de los pacientes que presentan algún tipo de cáncer (Estudio2), una vez generados los Ilustraciones del Estudio2, tendremos que volver a cargar la base de datos para empezar a trabajar con la tabla patologías_m3 inicial, para poder realizar el Estudio3 (las 5 patologías más frecuentes de nuestra base de datos).

Comenzamos creando la variable Estudio y añadiéndola a la tabla_patologias, para posteriormente, poder eliminar las patologías que no van a ser de interés, de la tabla patologías_m3 en nuestro estudio. Para ello, llamamos a dicha variable dic_sel y eliminamos todas las patologías que vengan acompañadas por FALSE de la variable Estudio. Veasé el ejemplo:

```

````{r}
Estudio <-
c(T,T,T,T,T,T,T,F,T,T,T,T,F,T,T,T,T,T,T,F,F,T,T,F,T,F,F,F,T,T,F,T,F,F,T,T,F,T

```

```
,F,T,F,F,F,F,T,T,F,F,T,F,T,F,F,F,T,T,F,T,T,T,T,T,F,T,T,F,T,T,T,T,T,T,T,T,T,T,
T,T,T,T,T,T)
```

```
tabla_patologias$Estudio <- Estudio
```

```
dic_sel <- tabla_patologias$Estudio
```

```
names(dic_sel) <- tabla_patologias$PATOLOGIA_BPS
```

```
patologias_m3 <- patologias_m3[dic_sel[patologias_m3$PATOLOGIA_BPS],]
```

```
````
```

En la tabla_patologias nos deshacemos de las patologias que no son de nuestro interés, eliminando las patologias que posean FALSE en la variable Estudio.

```
````{r echo=F}
```

```
tabla_patologias <- tabla_patologias[Estudio,]
```

```
head(tabla_patologias)
```

```
tabla_patologias
```

```
````
```

Añadimos una nueva variable a nuestra tabla_patologia, llamada Abreviaturas, para agrupar las patologias en grupos, que contengan 1 o más de nuestras patologías a estudiar y reducir el nombre con dichas Abreviaturas. Cada abreviatura debe corresponder a su patología.

```
````{r echo=F}
```

```
Abreviaturas <-
```

```
c("T", "H", "Dp", "Art", "D", "As", "Ur", "Ht", "O", "T", "As", "Est", "Cai", "Art", "Ica",
"Cai", "Ost", "Cai", "Cev", "Dp", "Cai", "T", "Mue",
```

```
"Ren", "Dm", "Cev", "C", "C", "Art", "C", "Col", "C", "Cev", "Cev", "Cev", "Hig", "Cai", "C",
"C", "C", "C",
```

```
"C", "C", "C", "C", "C", "C", "C", "C", "C", "C", "Cai", "C", "C", "C", "C", "C")
```

```
tabla_patologias$Abreviaturas <- Abreviaturas
```

```
head(tabla_patologias)
```

```
````
```

Vamos a incorporar la variable Abreviaturas a la tabla patologias_m3 y a la tabla Result de la misma forma que lo hemos hecho anteriormente llamando a dicha variable dic_sel, con una columna en comun entre los dos data.frame, y que cada patologia bps se le asocie su abreviatura correspondiente. A su vez, ordenamos la tabla para que la columna de las abreviaturas aparezca a continuacion de la de patologias.

```
````{r echo=F}
```

```

dic_sel <- tabla_patologias$Abreviaturas
names(dic_sel) <- tabla_patologias$PATOLOGIA_BPS
patologias_m3$Abreviaturas <- dic_sel[patologias_m3$PATOLOGIA_BPS]
Abreviaturas <- patologias_m3$Abreviaturas

patologias_m3 <-
data.frame(patologias_m3[,1:3],Abreviaturas,patologias_m3[,4:7])
head(patologias_m3)
````

De nuevo, se pegan las tablas usuarios_m4 y patologias_m3, creando la tabla
Result, para el posterior análisis de secuencias. Así mismo vamos a cambiar el
formato de la variable fecha inicio de patología y fecha fin de patología a un
número entero, quedándonos solo con el año final para poder hacer la operación
correspondiente para obtener la edad inicio de patología y la edad fin de
patología.
````{r echo=F}
Result <- merge(usuarios_m4, patologias_m3, by="n_orden")

head(as.character(Result$FECHA_INICIO_PATOLOGIA_BPS))
head(substr(as.character(Result$FECHA_INICIO_PATOLOGIA_BPS),1,4))
head(as.integer(substr(as.character(Result$FECHA_INICIO_PATOLOGIA_BPS),1,4)))

Result$EDAD_INICIO <- Result$EDAD - 2017 +
as.integer(substr(as.character(Result$FECHA_INICIO_PATOLOGIA_BPS),1,4))

head(as.character(Result$FECHA_FIN_PATOLOGIA_BPS))
head(substr(as.character(Result$FECHA_FIN_PATOLOGIA_BPS),1,4))
head(as.integer(substr(as.character(Result$FECHA_FIN_PATOLOGIA_BPS),1,4)))

Result$EDAD_FIN <- Result$EDAD - 2017 +
as.integer(substr(as.character(Result$FECHA_FIN_PATOLOGIA_BPS),1,4))
````

Para la edad del paciente a la que finaliza la patología, vamos a usar un
condicionante, para que si el resultado de la edad fin es mayor que la edad
actual del paciente, nos proporcione, en ese caso, la edad actual del paciente.
Ejemplo:
````{r}
Result$EDAD_FIN <-
ifelse(Result$EDAD_FIN>Result$EDAD,Result$EDAD,Result$EDAD_FIN)

```



```

Estudio3 <- c(T,T,F,T,T,T,F,F,F,T,T,F,F,T,F,F,F,F,F,F,T,F,
 F,F,F,F,F,T,F,F,F,F,F,F,F,F,F,F,F,F,F,F,F,F,
 F,F,F,F,F,F,F,F,F,F)
tabla_patologias$Estudio3 <- Estudio3

dic_sel <- tabla_patologias$Estudio3
names(dic_sel) <- tabla_patologias$PATOLOGIA_BPS
patologias_m3 <- patologias_m3[dic_sel[patologias_m3$PATOLOGIA_BPS],]

Result <- Result[dic_sel[Result$PATOLOGIA_BPS],]

```

```

tabla_patologias <- tabla_patologias[Estudio3,]
head(tabla_patologias)
````

```

Creamos una tabla con el número de patologías de nuestro Estudio 2 que presenta cada paciente y seguidamente una tabla con el número de personas que presentan 1 patología, 2, 3, etc.

```

````{r}
prueba3<-Result %>%
 dplyr::group_by(n_orden) %>%
 dplyr::count()
prueba4<-prueba3 %>%
 dplyr::group_by(n) %>%
 dplyr::count()
colnames(prueba4)<-c("Numeropatol","Personas")
prueba4
````

```

Colocamos la columna abreviatura a continuación de las patologías bps.

```

````{r echo=F}
tabla_patologias <- tabla_patologias[with(tabla_patologias,
order(tabla_patologias$Abreviaturas)), c(1,7,2:6)]
head(tabla_patologias)
````

```

De esta forma vamos a ordenar las abreviaturas en el analisis de secuencias final para cada paciente:

```
````{r echo=F}
cadena <- c("Art","As","Cai")
cadena2 <- c("C")
cad <- sort(c(cadena,cadena2))
````
```

Creamos un nuevo data.frame, llamado Fin, con las variables n_orden,EDAD y Fallecido de la tabla usuario_m3.

```
````{r echo=F}
Fin <-
data.frame(n_orden=usuarios_m4$n_orden,EDAD=usuarios_m4$EDAD,Fallecido=usuarios_m4$FALLECIDO_EN_EL_PERIODO)
````
```

A su vez, añadimos las variables Edad desde la EDAD000 a la Edad114, que son las edades que pueden tener nuestros pacientes, para el posterior análisis de secuencias.

```
````{r echo=F}
Fin$E000 <- "-"
Fin$E001 <- "-"
Fin$E002 <- "-"
Fin$E003 <- "-"
Fin$E004 <- "-"
Fin$E005 <- "-"
Fin$E006 <- "-"
Fin$E007 <- "-"
Fin$E008 <- "-"
Fin$E009 <- "-"
Fin$E010 <- "-"
Fin$E011 <- "-"
Fin$E012 <- "-"
Fin$E013 <- "-"
Fin$E014 <- "-"
Fin$E015 <- "-"
Fin$E016 <- "-"
Fin$E017 <- "-"
Fin$E018 <- "-"
Fin$E019 <- "-"
```



Fin\$E020 <- "-"  
Fin\$E021 <- "-"  
Fin\$E022 <- "-"  
Fin\$E023 <- "-"  
Fin\$E024 <- "-"  
Fin\$E025 <- "-"  
Fin\$E026 <- "-"  
Fin\$E027 <- "-"  
Fin\$E028 <- "-"  
Fin\$E029 <- "-"  
Fin\$E030 <- "-"  
Fin\$E031 <- "-"  
Fin\$E032 <- "-"  
Fin\$E033 <- "-"  
Fin\$E034 <- "-"  
Fin\$E035 <- "-"  
Fin\$E036 <- "-"  
Fin\$E037 <- "-"  
Fin\$E038 <- "-"  
Fin\$E039 <- "-"  
Fin\$E040 <- "-"  
Fin\$E041 <- "-"  
Fin\$E042 <- "-"  
Fin\$E043 <- "-"  
Fin\$E044 <- "-"  
Fin\$E045 <- "-"  
Fin\$E046 <- "-"  
Fin\$E047 <- "-"  
Fin\$E048 <- "-"  
Fin\$E049 <- "-"  
Fin\$E050 <- "-"  
Fin\$E051 <- "-"  
Fin\$E052 <- "-"  
Fin\$E053 <- "-"  
Fin\$E054 <- "-"  
Fin\$E055 <- "-"  
Fin\$E056 <- "-"  
Fin\$E057 <- "-"  
Fin\$E058 <- "-"

Fin\$E059 <- "-"  
Fin\$E060 <- "-"  
Fin\$E061 <- "-"  
Fin\$E062 <- "-"  
Fin\$E063 <- "-"  
Fin\$E064 <- "-"  
Fin\$E065 <- "-"  
Fin\$E066 <- "-"  
Fin\$E067 <- "-"  
Fin\$E068 <- "-"  
Fin\$E069 <- "-"  
Fin\$E070 <- "-"  
Fin\$E071 <- "-"  
Fin\$E072 <- "-"  
Fin\$E073 <- "-"  
Fin\$E074 <- "-"  
Fin\$E075 <- "-"  
Fin\$E076 <- "-"  
Fin\$E077 <- "-"  
Fin\$E078 <- "-"  
Fin\$E079 <- "-"  
Fin\$E080 <- "-"  
Fin\$E081 <- "-"  
Fin\$E082 <- "-"  
Fin\$E083 <- "-"  
Fin\$E084 <- "-"  
Fin\$E085 <- "-"  
Fin\$E086 <- "-"  
Fin\$E087 <- "-"  
Fin\$E088 <- "-"  
Fin\$E089 <- "-"  
Fin\$E090 <- "-"  
Fin\$E091 <- "-"  
Fin\$E092 <- "-"  
Fin\$E093 <- "-"  
Fin\$E094 <- "-"  
Fin\$E095 <- "-"  
Fin\$E096 <- "-"  
Fin\$E097 <- "-"

```
Fin$E098 <- "-"
Fin$E099 <- "-"
Fin$E100 <- "-"
Fin$E101 <- "-"
Fin$E102 <- "-"
Fin$E103 <- "-"
Fin$E104 <- "-"
Fin$E105 <- "-"
Fin$E106 <- "-"
Fin$E107 <- "-"
Fin$E108 <- "-"
Fin$E109 <- "-"
Fin$E110 <- "-"
Fin$E111 <- "-"
Fin$E112 <- "-"
Fin$E113 <- "-"
Fin$E114 <- "-"
````
```

Vamos a crear una nueva tabla, llamada `temp_pac`, donde el paciente en la posición 1 de la tabla `Fin` sea nuestro único dato, pero lo extraemos de la tabla `Result` con todas las variables que posea en dicha tabla. Nos dedicaremos a analizar un paciente exclusivamente, por ahora.

Vamos a crear un bucle `for` para poder realizar el análisis de secuencias con el paquete `TRAMINER`, en nuestra tabla `Fin` podemos ver todas las variables `EDAD` de los pacientes, donde se le va a asignar una "S", a cada edad del paciente, hasta que aparezca su primera patología y seguirá teniendo una "S" desde que se le cierre dicha patología hasta llegar a la `EDAD` actual del paciente, que se le asignara una "N" hasta la `EDAD` máxima. En el caso de que tenga múltiples patologías a la vez, hasta que no finalicen todas las patologías, no le volverá a aparecer una "S", o bien una "N" si llega a su `EDAD` actual. Si el paciente ha fallecido, se le asignará una "M".

Dentro del bucle, se genera el método necesario para que, en caso de contar con múltiples patologías, nos indique cada una de ellas, así como la secuencia de patologías que presentara cada paciente pluripatológico.

```
````{r echo=F}
```



```

````{r echo=F}
Para el Estudio2:

  Fin.labels <- c("Presenta Cáncer","Fallecido","No vivido","Sin patología")

Fin.alphabet <- c("C","M","N","S")

Fin.cpal <- c("pink",
             "black",
             "gray",
             "lightblue")
````

##Para el Estudio3:

````{r}

Fin.labels <- c("Presenta Artrosis", "Presenta Artrosis y Asma", "Presenta
Artrosis, Asma y Diabetes", "Presenta Artrosis, Asma, Diabetes e Hipertensión",
"Presenta Artrosis, Asma, Diabetes, Hipertensión y Trastorno", "Presenta
Artrosis, Asma, Diabetes y Trastorno", "Presenta Artrosis, Asma e Hipertensión",
"Presenta Artrosis, Asma, Hipertensión y Trastorno", "Presenta Artrosis, Asma
y Trastorno", "Presenta Artrosis y Diabetes", "Presenta Artrosis, Diabetes e
Hipertensión", "Presenta Artrosis, Diabetes, Hipertensión y Trastorno",
"Presenta Artrosis, Diabetes y Trastorno", "Presenta Artrosis e Hipertensión",
"Presenta Artrosis, Hipertensión y Trastorno", "Presenta Artrosis y Trastorno",
"Presenta Asma", "Presenta Asma y Diabetes", "Presenta Asma, Diabetes e
Hipertensión", "Presenta Asma, Diabetes, Hipertensión y Trastorno", "Presenta
Asma, Diabetes y Trastorno", "Presenta Asma e Hipertensión", "Presenta Asma,
Hipertensión y Trastorno", "Presenta Asma y Trastorno", "Presenta Diabetes",
"Presenta Diabetes e Hipertensión", "Presenta Diabetes, Hipertensión y
Trastorno", "Presenta Diabetes y Trastorno", "Presenta Hipertensión", "Presenta
Hipertensión y Trastorno", "Fallecido", "No vivido", "Sin patología", "Presenta
Trastorno")

Fin.alphabet <- c("Art","ArtAs","ArtAsD","ArtAsDH","ArtAsDHT","ArtAsDT",
                "ArtAsH","ArtAsHT","ArtAsT","ArtD","ArtDH","ArtDHT",
                "ArtDT","ArH","ArtHT","ArtT","As","AsD","AsDH","AsDHT",
                "AsDT","AsH","AsHT","AsT","D","DH","DHT","DT","H","HT","M",
                "N","S","T")

```

```
Fin.cpal <- c("white",  
             "aliceblue",  
             "aquamarine",  
             "antiquewhite1",  
             "antiquewhite4",  
             "aquamarine4",  
             "azure3",  
             "beige",  
             "blue",  
             "blueviolet",  
             "brown",  
             "burlywood",  
             "chartreuse",  
             "chocolate",  
             "chocolate4",  
             "coral",  
             "cornflowerblue",  
             "cornsilk",  
             "pink",  
             "cyan",  
             "darkblue",  
             "darkgoldenrod",  
             "darkgoldenrod4",  
             "darkgreen",  
             "darkkhaki",  
             "darkmagenta",  
             "darkolivegreen",  
             "darkolivegreen1",  
             "darkorange",  
             "darkred",  
             "black",  
             "gray",  
             "darksalmon",  
             "yellow")
```

Preparamos la secuencia definiendo el alfabeto a utilizar, los colores y las etiquetas, así como la matriz de donde vamos a realizar el análisis de secuencia. Lo llamamos Ansec_OSEQ.

```
````{r echo=F}
Ansec_OSEQ <- seqdef(muestra.Fin, var = 4:118,
 cpal = Fin.cpal, alphabet = Fin.alphabet,
 labels = Fin.labels)
````
```

Tiempo medio en cada estado.

```
````{r echo=F}
head(seqistatd (Ansec_OSEQ))
````
```

Media del tiempo dedicada a cada tipo de evento.

```
````{r echo=F}
seqmeant (Ansec_OSEQ)
````
```

```
````{r echo=F}
Fin.om <- seqdist(Ansec_OSEQ, method = "OM", indel = 1, sm = "TRATE")
````
```

Se generan 4 clúster para poder observar el comportamiento del análisis de secuencias en cada uno de ellos.

```
````{r echo=F}
library("clúster")
clústerward <- agnes(Fin.om, diss = TRUE, method = "ward")
muestra.Fin.cl4 <- cutree(clústerward, k = 4)
cl4.lab <- factor(muestra.Fin.cl4, labels = paste("Clúster", 1:4))
````
```

Generamos los Ilustracións de análisis de secuencias.

```
````{r echo=F}
seqdplot(Ansec_OSEQ, group = cl4.lab, border = NA)
````
```