



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA DE TELECOMUNICACIÓN

DOBLE GRADO EN INGENIERÍA EN SISTEMAS DE
TELECOMUNICACIÓN Y ADMINISTRACIÓN Y DIRECCIÓN DE
EMPRESAS

TRABAJO FIN DE GRADO

SELECCIÓN DE CARACTERÍSTICAS Y ANÁLISIS DE
CORRESPONDENCIAS PARA EXTRAER CONOCIMIENTO DE
GRUPOS DE PACIENTES CRÓNICOS

Autora: Natalia Alonso Arteaga

Tutora: Inmaculada Mora Jiménez

Cotutora: Cristina Soguero Ruíz

Curso académico 2019/2020

Este trabajo ha sido parcialmente financiado por el Ministerio de Educación, Cultura y Deporte como parte del proyecto TEC2016-75361-R, y por el Instituto de Salud Carlos III como parte del proyecto DTS17/0058.

Tanto si piensas que puedes, como si piensas que no puedes, estás en lo cierto.

Henry Ford

Piensa mal, pero en todos los casos, piensa por ti mismo.

Doris Lessing

Agradecimientos

Ahora sí, por fin puedo decir que lo he conseguido. Después de varios años estudiando y preparándome para el mundo laboral, por fin ha llegado el momento. Con este proyecto culmina este largo proceso de formación, me despido de esta etapa y comienzo con ilusión los nuevos propósitos. Pero antes me gustaría agradecer el apoyo que me han prestado algunas personas durante este ciclo y también a los que han colaborado en el desarrollo de este proyecto.

En primer lugar, quiero agradecer a mi familia, pero muy en especial a mis padres, todo el esfuerzo que han hecho para que yo esté aquí, para que mi formación fuera la mejor y lo más completa posible. En este periodo, ha habido mucho esfuerzo, no siempre recompensado, y en muchas ocasiones esto me ha llevado a la frustración y a pensar que no lo iba a conseguir. Sin embargo, mis padres nunca han tirado la toalla, aún cuando yo he estado a punto de hacerlo, siempre han apostado por mi y me han animado a seguir adelante. Hoy por fin puedo decir, que todo el sacrificio y el esfuerzo ha merecido la pena. También a mi hermano, a mis abuelos y a mis tíos, porque aunque en silencio, siempre han estado ahí para apoyarme.

En segundo lugar, a mis tutoras, Inma y Cristina, por todo lo que me han enseñado y aportado durante este proyecto. Durante mi etapa en la universidad no todas las asignaturas me han gustado por igual, incluso llegándose a plantear si había elegido bien al decidirme por este doble grado. Sin embargo, con este trabajo he disfrutado mucho y he recuperado las ganas de seguir estudiando y formándome. También quiero agradecer la colaboración en este proyecto del profesor José Luis Rojo, y de nuestros colaboradores en el Hospital Universitario de Fuenlabrada y en el Hospital Universitario del Sureste, quienes no han dudado en ayudar en todo lo posible, gracias por aportar vuestro conocimiento y así enriquecer este trabajo.

También quiero dar las gracias a mis compañeros sin los cuales hubiera sido muy difícil seguir adelante. Gracias a ellos el agobio antes de los exámenes y las largas tardes de estudio en la biblioteca, se llevaban un poco mejor. Hemos reído, hemos llorado, hemos compartido palmeras de chocolate y mucho cafés, hemos celebrado los aprobados, hemos sufrido con los suspensos, pero al final lo hemos conseguido.

Por último, a mi pareja, porque estos últimos años, no han sido pocas las veces que ha tenido que aguantar mis frustraciones y mi mal humor cuando las cosas no han salido como yo esperaba a la primera. También ha compartido todo mi agobio y mi esfuerzo, y espero que a partir de ahora podemos compartir los éxitos.

¡Muchas gracias a todos!

Resumen

La sociedad española está sufriendo un crecimiento negativo de la población, con su consecuente envejecimiento. Esto trae consigo la prevalencia de algunas enfermedades crónicas, de progresión lenta y larga duración, que se han convertido en uno de los principales problemas para los sistemas sanitarios. Este Trabajo de Fin de Grado (TFG) se centra en la extracción de conocimiento de algunos grupos de pacientes crónicos, en concreto de pacientes con Hipertensión, Diabetes, Depresión y Enfermedad Pulmonar Obstructiva Crónica.

En concreto, en este TFG se hace uso de un Sistema de Clasificación de Pacientes poblacional para agrupar a la población en diferentes estados de salud. Cada grupo está formado por pacientes con características clínicas, demográficas y de consumo de recursos similares. Además, se hace uso de varias bases de datos, facilitadas por el Hospital Universitario de Fuenlabrada, donde se recogen los contactos de los pacientes del Área de Salud de Fuenlabrada con el sistema sanitario público entre los años 2012 y 2015. Estos datos corresponden a características demográficas (género y edad) y clínicas (diagnósticos y fármacos, ambos codificados).

Tras comprobar la calidad de los datos, para extraer conocimiento de los mismos, primero se realiza un test de hipótesis con remuestreo *bootstrap*. Este test permitirá seleccionar las características (diagnósticos y fármacos) a analizar haciendo uso de un Análisis de Correspondencias Múltiple (ACM). El ACM permitirá visualizar asociaciones tanto entre características como entre pacientes, pudiendo poner de manifiesto relaciones no evidentes en primera instancia.

Del análisis llevado a cabo, por un lado, se concluyen algunas evidencias que permiten validar las herramientas empleadas en este trabajo. Por otro lado, se observa que los pacientes hipertensos también tienen asociados trastornos de ansiedad, y que los diabéticos tienen diagnósticos y fármacos relacionados con hipertensión. Estas conclusiones han sido validadas en la literatura clínica. Además, el análisis realizado ha capturado algunas diferencias entre los grupos en 2012 y 2015, diferencias que podrían estar asociadas a los cambios normativos introducidos con el Real Decreto-ley 16/2012 en el que se recogían algunas medidas urgentes para garantizar la sostenibilidad del sistema sanitario.

Por último, con los datos disponibles hasta el momento, se han extraído algunas conclusiones que podrían ayudar a la toma de decisiones desde el punto de vista clínico, como por ejemplo, ofrecer una atención más especializada a diferentes grupos de pacientes clasificados en un mismo estado de salud. Además, como línea futura se plantea considerar nuevos datos, como el peso, la tensión, o resultados de análisis de sangre, para diseñar modelos predictivos del empeoramiento del estado de salud y poder prevenirlo con las medidas sanitarias adecuadas.

Índice general

Agradecimientos

Resumen

Índice de figuras	V
Lista de acrónimos y abreviaturas	XII
1. Introducción y objetivos	1
1.1. Contexto y motivación	1
1.2. Objetivos	4
1.3. Metodología	5
1.4. Estructura de la memoria	7
2. Cronicidad. Conceptos previos	9
2.1. Enfermedades crónicas	9
2.1.1. Hipertensión Arterial	10
2.1.2. Diabetes Mellitus	11
2.1.3. Depresión	12
2.1.4. Enfermedad Pulmonar Obstructiva Crónica	13
2.2. Sistema de Clasificación de Pacientes poblacional	14
2.3. Codificación de diagnósticos y fármacos	16
2.3.1. Clasificación Internacional de Enfermedades	16

2.3.2. Sistema de Clasificación Anatómica, Terapéutica, Química	17
3. Bases de datos y análisis exploratorio	19
3.1. Bases de datos	19
3.2. Análisis exploratorio	21
3.2.1. Características demográficas y clínicas	26
3.2.2. Perfiles de estados de salud	29
4. Métodos	37
4.1. Test de proporciones y remuestreo <i>bootstrap</i> para seleccionar características . .	37
4.2. Análisis de Correspondencias Múltiple	42
5. Experimentos y resultados	49
5.1. Selección de características mediante test de proporciones y remuestreo <i>bootstrap</i>	49
5.1.1. Caracterización del estado de salud	50
5.1.2. Evolución temporal	54
5.2. Análisis de Correspondencias Múltiple	55
5.2.1. Caracterización del estado de salud	56
5.2.2. Evolución temporal	61
6. Conclusiones y líneas futuras	63
6.1. Conclusiones	63
6.2. Líneas futuras	64
A. Extracción de datos	67
B. Diagnósticos considerados de HTA, DM, depresión y EPOC	71
C. Códigos ATC considerados de HTA, DM, depresión y EPOC	79
D. Distribución del número total de diagnósticos y códigos ATC por paciente	83

E. Análisis comparativo de los perfiles poblaciones desde 2010 hasta 2013	87
F. Códigos seleccionados para analizar la evolución del estado de salud	91
G. Códigos seleccionados para analizar la evolución temporal	99
H. Implementación del ACM	103
I. Resultados del ACM para el análisis de la evolución temporal	105
Bibliografía	109

Índice de figuras

1.1. Aumento de la prevalencia de algunas enfermedades crónicas en España.	3
1.2. Principales cronicidades en España en 2017 entre la población con más de 15 años.	3
1.3. Detalle de las actividades llevadas a cabo.	7
1.4. Diagrama de Gantt.	7
2.1. Principales causas de muerte en España: 2007 frente a 2017.	10
2.2. Estructura de un CRG.	15
2.3. Código ATC A10BA02 desglosado por niveles.	18
3.1. Evolución de la tasa de presencia de los diagnósticos 250, 272, 311, 401, 496, 719, 724, 780 para el CRG-base 3755 (a), 5133 (b), 5192 (c), 5424 (d), 6313 (e) y 7140 (f).	21
3.2. Evolución de la tasa de presencia de los códigos ATC A10AB, A10BA, C09AA, C10AA, N06AB y R03BB para el CRG-base 3755 (a), 5133 (b), 5192 (c), 5424 (d), 6313 (e) y 7140 (f).	22
3.3. Número anual de pacientes entre 2012 y 2015 para los CRGs-base 1000 y 5192 (a), y para el resto de CRGs-base de interés (b).	24
3.4. Número total de pacientes por CRG-base durante el año 2012, omitiendo el CRG-base 1000 para propósitos de visualización.	24
3.5. Distribución de probabilidad del número de contactos por paciente para el CRG-base 3755 (a), 5133 (b), 5192 (c), 5424 (d), 6313 (e), y 7140 (f) en 2012.	25

3.6. Distribución de probabilidad de la edad de los pacientes de los CRG-base: 3755 (a), 5133 (b), 5192 (c), 5424 (d), 6313 (e), y 7140 (f) en 2012.	26
3.7. Distribución de probabilidad del género de los pacientes de los CRG-base: 3755 (a), 5133 (b), 5192 (c), 5424 (d), 6313 (e), y 7140 (f) en 2012.	27
3.8. Perfiles de presencia de diagnósticos (a) y códigos ATC (b), y perfiles de ocu- rrencia de diagnósticos (c) y códigos ATC (d) del CRG-base 3755.	30
3.9. Perfiles de presencia de diagnósticos (a) y códigos ATC (b), y perfiles de ocu- rrencia de diagnósticos (c) y códigos ATC (d) del CRG-base 5133.	31
3.10. Perfiles de presencia de diagnósticos (a) y códigos ATC (b), y perfiles de ocu- rrencia de diagnósticos (c) y códigos ATC (d) del CRG-base 5192.	32
3.11. Perfiles de presencia de diagnósticos (a) y códigos ATC (b), y perfiles de ocu- rrencia de diagnósticos (c) y códigos ATC (d) del CRG-base 5424.	33
3.12. Perfiles de presencia de diagnósticos (a) y códigos ATC (b), y perfiles de ocu- rrencia de diagnósticos (c) y códigos ATC (d) del CRG-base 6313.	34
3.13. Perfiles de presencia de diagnósticos (a) y códigos ATC (b), y perfiles de ocu- rrencia de diagnósticos (c) y códigos ATC (d) del CRG-base 7140.	34
4.1. Estadístico de contraste $T \sim N(0, 1)$ y regiones de rechazo (RR) y aceptación (RA) según la definición de H_0 y H_1 para, (a) $H_0 : \theta \leq \theta_0, H_1 : \theta > \theta_0$, (b) $H_0 : \theta \geq \theta_0, H_1 : \theta < \theta_0$, (c) $H_0 : \theta = \theta_0, H_1 : \theta \neq \theta_0$	38
4.2. Región de rechazo para $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$ y $\alpha = 0.05$	39
4.3. Cantidad pivotal e intervalo de confianza.	40
4.4. Espacio de perfiles asociado al ejemplo considerado.	44
5.1. CRG-base 1000 versus 3755 en 2012. Intervalos de confianza de Δp al 95 % pa- ra diagnósticos (a) y fármacos (b). Detalle para los diagnósticos (c) y fármacos (d) seleccionados.	51
5.2. Códigos seleccionados al comparar el CRG-base 1000 frente al CRG-base 5133 en 2012: diagnósticos (a) y fármacos (b).	52
5.3. Códigos seleccionados al comparar el CRG-base 1000 frente al CRG-base 5192 en 2012: diagnósticos (a) y fármacos (b).	52

5.4. Códigos seleccionados al comparar el CRG-base 1000 frente al CRG-base 5424 en 2012: diagnósticos (a) y fármacos (b)	53
5.5. Códigos seleccionados al comparar el CRG-base 1000 frente al CRG-base 6313 en 2012: diagnósticos (a) y fármacos (b)	53
5.6. Códigos seleccionados al comparar el CRG-base 1000 frente al CRG-base 7140 en 2012: diagnósticos (a) y fármacos (b)	54
5.7. Proporción de inercia explicada por cada autovalor (a) y los tres autovectores con mayor inercia (b)	57
5.8. Asociación entre categorías de diagnósticos, fármacos y CRG-base.	58
5.9. Asociación entre categorías de la variable diagnósticos y CRGs-base (a), y entre fármacos y CRGs-base (b)	59
5.10. Asociación entre categorías de la variable diagnósticos y CRGs-base, representando de manera independiente los que están cerca (a) y lejos (b) del origen.	59
5.11. Asociación entre categorías de la variable fármacos y CRGs-base, representando de manera independiente los que están cerca (a) y lejos (b) del origen.	60
5.12. Representación de individuos de los CRGs-base de interés.	61
A.1. Administrador de origen de datos ODBC de Windows.	68
A.2. Nueva consulta a la base de datos.	69
A.3. Acceso a la base de datos completa.	69
A.4. Acceso a la fecha de nacimiento.	70
D.1. Distribución del número total de diagnósticos por paciente relacionados con depresión (a), EPOC (b), HTA (c) y DM (d); y códigos ATC por paciente relacionados con depresión (e), EPOC (f), HTA (g) y DM (h) para el CRG-base 3755 en 2012.	83
D.2. Distribución del número total de diagnósticos por paciente relacionados con depresión (a), EPOC (b), HTA (c) y DM (d); y códigos ATC por paciente relacionados con depresión (e), EPOC (f), HTA (g) y DM (h) para el CRG-base 5133 en 2012.	84

D.3. Distribución del número total de diagnósticos por paciente relacionados con depresión (a), EPOC (b), HTA (c) y DM (d); y códigos ATC por paciente relacionados con depresión (e), EPOC (f), HTA (g) y DM (h) para el CRG-base 5192 en 2012.	84
D.4. Distribución del número total de diagnósticos por paciente relacionados con depresión (a), EPOC (b), HTA (c) y DM (d); y códigos ATC por paciente relacionados con depresión (e), EPOC (f), HTA (g) y DM (h) para el CRG-base 5424 en 2012.	85
D.5. Distribución del número total de diagnósticos por paciente relacionados con depresión (a), EPOC (b), HTA (c) y DM (d); y códigos ATC por paciente relacionados con depresión (e), EPOC (f), HTA (g) y DM (h) para el CRG-base 6313 en 2012.	85
D.6. Distribución del número total de diagnósticos por paciente relacionados con depresión (a), EPOC (b), HTA (c) y DM (d); y códigos ATC por paciente relacionados con depresión (e), EPOC (f), HTA (g) y DM (h) para el CRG-base 7140 en 2012.	86
E.1. Perfiles de presencia de diagnósticos (a) y códigos ATC (b), y perfiles de ocurrencia de diagnósticos (c) y códigos ATC (d) del CRG-base 3755 en 2010. . . .	88
E.2. Perfiles de presencia de diagnósticos (a) y códigos ATC (b), y perfiles de ocurrencia de diagnósticos (c) y códigos ATC (d) del CRG-base 3755 en 2011. . . .	88
E.3. Perfiles de presencia de diagnósticos (a) y códigos ATC (b), y perfiles de ocurrencia de diagnósticos (c) y códigos ATC (d) del CRG-base 3755 en 2012. . . .	89
E.4. Perfiles de presencia de diagnósticos (a) y códigos ATC (b), y perfiles de ocurrencia de diagnósticos (c) y códigos ATC (d) del CRG-base 3755 en 2013. . . .	89
I.1. Evolución temporal del CRG-base 3755: asociaciones entre categorías (a) y entre pacientes (b).	105
I.2. Evolución temporal del CRG-base 5133: asociaciones entre categorías (a) y entre pacientes (b).	106
I.3. Evolución temporal del CRG-base 5192: asociaciones entre categorías (a) y entre pacientes (b).	106

I.4. Evolución temporal del CRG-base 5424: asociaciones entre categorías (a) y entre pacientes (b).	107
I.5. Evolución temporal del CRG-base 6313: asociaciones entre categorías (a) y entre pacientes (b).	107
I.6. Evolución temporal del CRG-base 7140: asociaciones entre categorías (a) y entre pacientes (b).	108

Lista de acrónimos y abreviaturas

AC Análisis de Correspondencias

ACG *Ambulatory Care Groups*

ACM Análisis de Correspondencias Múltiple

AE Atención Especializada

AP Atención Primaria

ATC *Anatomical, Therapeutic, Chemical Classification System*

CIE – 9 Clasificación Internacional de Enfermedades-9^a Revisión

CRG *Clinical Risk Groups*

CS Centro de Salud

DM Diabetes Mellitus

DVS Descomposición en Valores Singulares

EPOC Enfermedad Pulmonar Obstructiva Crónica

GRD Grupos Relacionados por el Diagnóstico

HTA Hipertensión Arterial

HUF Hospital Universitario de Fuenlabrada

OMS Organización Mundial de la Salud

SCP Sistemas de Clasificación de Pacientes

SNS Sistema Nacional de Salud

TFG Trabajo Fin de Grado

Capítulo 1

Introducción y objetivos

En este capítulo se presenta el contexto en el que se desarrolla este Trabajo de Fin de Grado (TFG) y los objetivos que se van a abordar. Como parte de este contexto se detallará el funcionamiento de los sistemas sanitarios y se justificará la elección de las enfermedades crónicas objeto de estudio. Se detallará también la metodología seguida y la estructura de la presente memoria.

1.1. Contexto y motivación

Según la Organización Mundial de la Salud (OMS) se entiende por salud el “estado completo de bienestar físico, mental y social”. Con el objetivo de conseguir dicho bienestar, surgen en las sociedades los sistemas sanitarios. Estos sistemas, están formados por todas las organizaciones, instituciones y recursos orientados a la consecución de dicho objetivo [1].

En España, el sistema que integra todos los servicios públicos sanitarios, es el Sistema Nacional de Salud (SNS). Entre sus características fundamentales, se encuentran la extensión de sus servicios a toda la población, la prestación de una atención integral, la promoción de la salud, la prevención de enfermedades, y la coordinación e integración de todos los recursos a fin de conseguir altos niveles de calidad en la prestación de sus servicios [2].

Para intentar gestionar los servicios sanitarios de forma eficiente, las competencias sanitarias del SNS se transfieren a las Comunidades Autónomas (CCAA). El SNS se estructura en Servicios de Salud y Áreas de Salud [3]. Un Servicio de Salud agrupa a todos los centros y establecimientos de una Comunidad Autónoma gestionados por ella misma. Cada comunidad debe

elaborar un Plan de Salud con todas las acciones necesarias para cumplir los objetivos aprobados por el Gobierno. Las Comunidades Autónomas delimitan su territorio en Áreas de Salud. Las Áreas de Salud son las estructuras básicas del SNS y deben cubrir las necesidades tanto de atención primaria como especializada, así como procurar la integración de toda la información asociada a cada paciente [4].

La Atención Primaria (AP) es el nivel asistencial que procura la promoción y prevención de la salud, así como el mantenimiento y la recuperación de la misma. En este nivel, se concentra la mayor parte de la actividad sanitaria. Por otra parte, la Atención Especializada (AE) engloba actividades diagnósticas, terapéuticas y de rehabilitación, de mayor complejidad que las consideradas en AP. El acceso a la AE se realiza por requerimiento del profesional de AP, que así lo considere oportuno, o por razones de urgencia [5].

Uno de los grandes problemas al que se enfrentan los sistemas sanitarios hoy en día, es el aumento de la prevalencia de las enfermedades crónicas. Se estima que estas enfermedades consumen aproximadamente el 70 % del total de los recursos sanitarios y que el 80 % de las consultas en AP se deben precisamente a dichas enfermedades [6]. El Informe Anual del Sistema Nacional de Salud 2018 (publicado en diciembre de 2019), recoge un análisis sobre el estado de salud de la población española en 2017. En dicho informe, se resaltan las enfermedades crónicas como uno de los principales problemas de salud [7]. El auge de las enfermedades crónicas, se debe en cierta medida al envejecimiento progresivo de la población.

Este envejecimiento se debe a que la tasa de mortalidad está aumentando en los últimos años (8,17 defunciones por cada 1.000 habitantes en 2010, frente a 9,1 defunciones por cada 1.000 habitantes en 2018 [8]) mientras que el número de nacimientos decrece. Esto da lugar a un crecimiento vegetativo negativo (el número de defunciones es superior al de nacimientos), y al consiguiente envejecimiento de la población. Esto se traduce en un mayor riesgo de padecer una enfermedad crónica. Por ejemplo, en la Figura 1.1, se muestra el aumento en la proporción de individuos que presentan hipertensión, colesterol o diabetes en España entre 1993 y 2017, frente al total de la población española.

El análisis realizado en este TFG, tiene como objetivo extraer conocimiento de datos asociados a grupos de pacientes crónicos y crónicos pluripatológicos: con hipertensión, diabetes, depresión y enfermedad pulmonar obstructiva crónica. La pluripatología o comorbilidad hace referencia al hecho de que un paciente presente dos o más enfermedades crónicas simultáneamente. En la Figura 1.2, se puede ver una lista de las principales cronicidades en España en 2017 y su tasa de presencia en pacientes con más de 15 años.

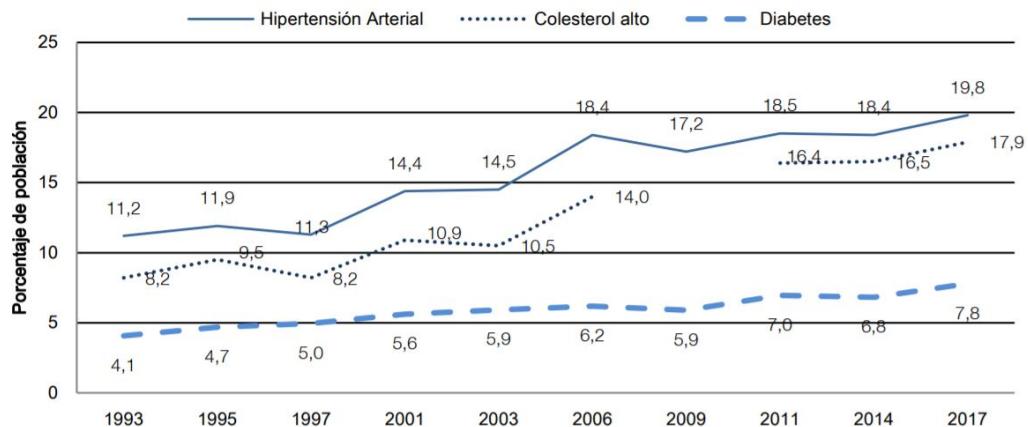


Figura 1.1: Aumento de la prevalencia de algunas enfermedades crónicas en España.
Obtenida de [7].

	Ambos sexos	Hombres	Mujeres	Razón Mujer/Hombre
Hipertensión arterial	19,8	20,0	19,7	1,0
Dolor de espalda crónico (lumbar)	18,5	14,7	22,1	1,5
Colesterol alto	17,9	18,1	17,7	1,0
Artrosis (excluyendo artritis)	17,5	11,3	23,5	2,1
Alergia crónica (rinitis, conjuntivitis...)	15,3	13,2	17,2	1,3
Dolor de espalda crónico (cervical)	14,6	8,9	20,0	2,2
Varices en las piernas	9,1	4,2	13,7	3,3
Migraña o dolor de cabeza frecuente	8,5	4,6	12,3	2,7
Diabetes	7,8	8,5	7,1	0,8
Ansiedad crónica	6,7	4,3	9,1	2,1
Depresión	6,7	4,0	9,2	2,3
Problemas crónicos de piel	5,5	4,9	6,0	1,2
Problemas de tiroides	5,3	1,6	8,9	5,6
Cataratas	5,2	4,0	6,3	1,6

Figura 1.2: Principales cronicidades en España en 2017 entre la población con más de 15 años.
Obtenida de [7].

Este trabajo se ha realizado en colaboración con personal del Hospital Universitario de Fuenlabrada (HUF). Conjuntamente, se plantea la necesidad de realizar un análisis de algunos grupos de pacientes crónicos. Para llevar a cabo este análisis nos han proporcionado varias bases de datos. Estas bases de datos corresponden a los contactos de los pacientes adscritos al HUF con el sistema sanitario: diagnósticos realizados, fármacos dispensados en farmacia, fecha del contacto, género y edad del paciente. Los datos han sido sometidos a un proceso de anonimización, llevado a cabo por el personal del HUF, de forma que a cada paciente le corresponde un identificador único con el que poder realizar un seguimiento de su estado de salud.

Una vez seleccionadas las enfermedades o patologías sobre las que realizar al estudio, se lleva a cabo un análisis exploratorio de los grupos de población considerados, en base a sus diagnósticos y fármacos asociados, al género y a la edad. El objetivo de este primer análisis, es tener una visión general de las características clínicas y demográficas asociadas a grupos pacientes. Posteriormente, se realiza una selección de características basada en un test de proporciones y un remuestreo *bootstrap* [9]. Esta selección de características se realiza para elegir aquéllas más relevantes, que se usarán en un posterior análisis. Como resultado del test, se obtiene una lista de diagnósticos y fármacos cuya tasa de presencia varía significativamente entre los diferentes grupos de pacientes crónicos en estudio. Por último, se realiza un Análisis de Correspondencias Múltiple (ACM) para representar gráficamente, en un espacio de dimensionalidad reducida, tanto asociaciones entre los diferentes grupos de pacientes como entre las características seleccionadas. Los métodos aplicados tienen un doble objetivo: por un lado, el estudio de los estados de salud considerados durante un año concreto y, por otro, el análisis de la evolución temporal de cada estado de salud.

1.2. Objetivos

El objetivo principal en torno al cual gira este proyecto, es caracterizar diferentes grupos de pacientes crónicos. Esta caracterización se realiza en base a datos demográficos y clínicos recogidos por el HUF. El objetivo principal se concreta en otros objetivos más específicos.

- **OE1.** Encontrar diferencias estadísticas significativas en la tasa de presencia de diagnósticos y fármacos asociados a diferentes grupos de pacientes crónicos. Este procedimiento ayudará a seleccionar las características (diagnósticos y fármacos) que mejor diferencian a los grupos de pacientes objeto de estudio.
- **OE2.** Encontrar asociaciones entre las características seleccionadas de los grupos de pacientes crónicos. Se trata de ver gráficamente qué fármacos y diagnósticos están más relacionados con determinadas enfermedades crónicas (de las consideradas a estudio). Estas asociaciones podrían advertir, por ejemplo, que el consumo de un determinado fármaco, o la presencia de cierto diagnóstico, podría estar relacionado con un futuro empeoramiento en el estado de salud de un paciente.
- **OE3.** Visualizar asociaciones entre pacientes clasificados en diferentes estados de salud en un espacio transformado de dimensión reducida, utilizando para ello las características

seleccionadas. Esto haría posible encontrar segmentos de pacientes clasificados en un determinado grupo, pero que poseen algunas características que les diferencian en cierta medida del resto del grupo. Clínicamente, se podrían tomar decisiones para tratar a estos pacientes de forma más personalizada, ya que poseen algunas particularidades que les diferencian del resto.

- **OE4.** Obtener información sobre la evolución temporal de los estados de salud considerados en el estudio. Se trata de encontrar qué características presentan diferencias significativas entre los pacientes clasificados en un mismo estado de salud en un intervalo temporal acotado. Concretamente se buscan diferencias en la tasa de presencia de diagnósticos y fármacos en cada estado de salud entre 2012 y 2015.

1.3. Metodología

Para lograr los objetivos, se han seguido los siguientes pasos:

- Revisar trabajos previos llevados a cabo en la misma temática. También se lleva a cabo un estudio de los métodos y técnicas estadísticas que se van a implementar para realizar el análisis.
- Decidir qué grupos de pacientes se considerarán en este TFG. Estos grupos de pacientes se seleccionan en función de las enfermedades crónicas que presentan dado el interés que suscitan estas enfermedades en el entorno sanitario. Concretamente el estudio se centrará en algunas de las cronicidades más prevalentes, como la hipertensión, diabetes, depresión y enfermedad pulmonar obstructiva crónica. También se considerarán grupos de pacientes que presentan más de una enfermedad crónica (dando lugar a una pluripatología), como por ejemplo, pacientes con hipertensión y diabetes.
- Extraer datos. Antes de proceder con el análisis se examinan las bases de datos para saber con qué datos se cuenta y cómo se estructuran. Las bases de datos recogen datos de todos los pacientes del HUF, sin embargo este TFG se centra sólo en algunos grupos de pacientes. Por esta razón se realiza una extracción de los datos pertenecientes a los grupos de pacientes crónicos objeto de estudio. Posteriormente, se lleva a cabo una limpieza de los mismos. En este proceso, se buscan inconsistencias, pacientes repetidos, o datos incongruentes. El objetivo de esta etapa es que los datos que se usarán posteriormente para el análisis sean de la mayor calidad posible, permitiendo una correcta toma de decisiones.

- Realizar un análisis exploratorio para conocer mejor los grupos de pacientes. En este primer análisis se calculan las distribuciones de edad y género, se estudia la tasa de presencia y la distribución de algunos fármacos y diagnósticos de interés, y se elaboran perfiles poblacionales. En esta etapa, se estudia también la evolución temporal de la tasa de presencia de algunas características relevantes desde el punto de vista clínico (diagnósticos y fármacos).
- Realizar una selección de características basada en test de hipótesis y remuestreo *bootstrap*, para encontrar diferencias entre los grupos de pacientes, en términos de diagnósticos y fármacos. El análisis de pares de grupos de pacientes permite encontrar diferencias significativas en la tasa de presencia de fármacos y diagnósticos de ambos grupos. Si se analizan, por ejemplo, pacientes sanos frente a pacientes diabéticos, se podrán encontrar diferencias entre ambos. Por un lado, se obtendrán diferencias evidentes, como por ejemplo la insulina consumida por diabéticos, que a su vez servirán para verificar que el método usado es correcto. Por otro lado, se obtendrán algunas características o diferencias no tan claras, que podrían aportar conocimiento en el estudio de estas cronicidades.
- Visualizar asociaciones tanto entre grupos de pacientes como entre características haciendo uso de los diagnósticos y fármacos seleccionados previamente.
- En colaboración con el personal del HUF, analizar los resultados obtenidos, extraer conclusiones clínicas y plantear algunas líneas futuras.
- Paralelamente al desarrollo de las etapas anteriormente descritas, se lleva a cabo la redacción de la presente memoria.

A continuación, en las Figuras 1.3 y 1.4, se detallan las actividades y el cronograma de las mismas, llevadas a cabo para la realización del presente TFG.

PROYECTO			
Selección de características y análisis de correspondencias para la extracción de conocimiento de pacientes crónicos			
AUTOR	TUTOR Y COTUTOR	FECHA DE INICIO	FECHA DE FIN
Natalia Alonso Arteaga	Inmaculada Mora Jiménez y Cristina Soguero Ruiz	29-Jun-18	15-Mar-20
TAREAS	INICIO	FIN	DÍAS
Contextualización, documentación y estado del arte	29-Jun-18	28-Sep-18	92
Estudio de las bases de datos y extracción de datos	28-Sep-18	4-Dec-18	68
Análisis exploratorio de los datos	4-Dec-18	11-Apr-19	129
Elaboración de perfiles de población	21-Jan-19	10-Feb-19	21
Investigación sobre Análisis de Correspondencias	11-Feb-19	25-Feb-19	15
Investigación sobre intervalos de confianza y test de proporciones	15-Feb-19	25-Feb-19	11
Obtención de resultados del Análisis de Correspondencias Múltiple (ACM)	18-Feb-19	27-Jun-19	130
Investigación sobre remuestreo bootstrap	20-Feb-19	21-Mar-19	30
Pruebas con medidas de similitud/correlación	21-Mar-19	23-Jul-19	125
Obtención de resultados remuestreo e intervalos de confianza	21-Mar-19	3-Jul-19	105
Diferentes experimentos y pruebas con remuestreo y ACM	18-Jul-19	12-Dec-19	148
Redacción de la memoria	15-Jan-19	15-Mar-20	426
Elaboración de la presentación	14-Feb-20	15-Mar-20	31

Figura 1.3: Detalle de las actividades llevadas a cabo.

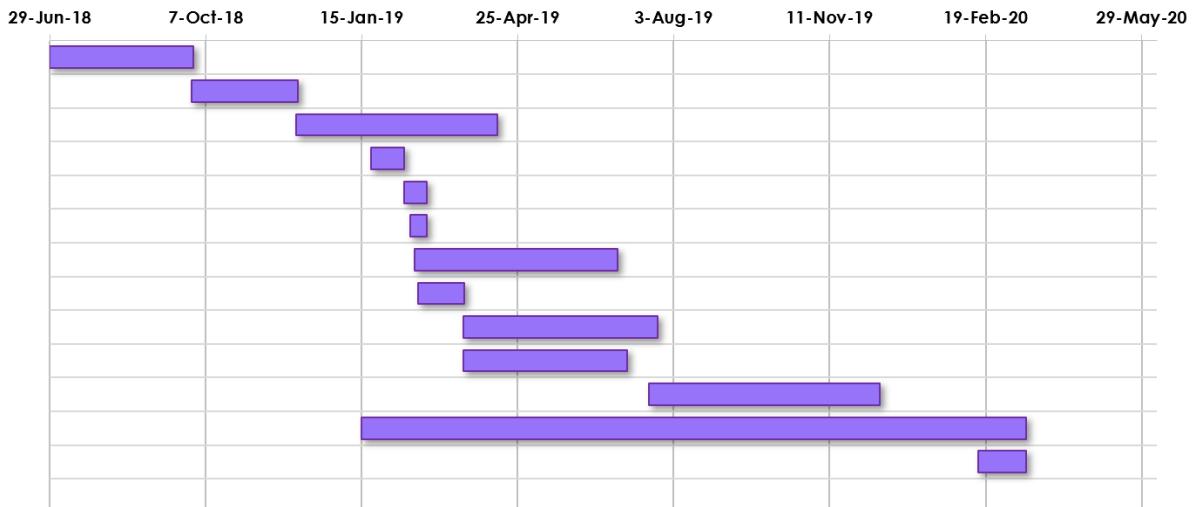


Figura 1.4: Diagrama de Gantt.

1.4. Estructura de la memoria

La memoria de este TFG se estructura en seis capítulos. A continuación, se detalla brevemente el contenido de cada uno de ellos:

- **Capítulo 1. Introducción y objetivos.** En este primer capítulo, se pone en contexto el TFG, detallando el ámbito en el que se desarrolla, la motivación que lo impulsa, y los

objetivos que se persiguen. También se describe la metodología seguida y la estructura de la memoria.

- **Capítulo 2. Cronicidad. Conceptos previos.** El Capítulo 2 introduce algunos conceptos importantes para contextualizar este TFG. En concreto, se describen las enfermedades crónicas elegidas para el estudio y las codificaciones que se usan en el ámbito sanitario para referirse a los diferentes diagnósticos y fármacos. También se describen los sistemas de clasificación de pacientes y cual ha sido usado en este TFG.
- **Capítulo 3. Bases de datos y análisis exploratorio.** A lo largo de este capítulo, se describen las bases de datos proporcionadas y su estructura. También se muestran los resultados del análisis exploratorio y se analizan los perfiles de la población. Por último, se estudia la evolución temporal en la tasa de presencia de algunos fármacos y diagnósticos que se consideran especialmente relevantes.
- **Capítulo 4. Métodos.** Los métodos estadísticos usados para el análisis de datos se describen en este capítulo. En primer lugar, se explica la selección de características basada en un test de hipótesis junto con un remuestreo bootstrap. En segundo lugar, se describe el ACM, empleado para conseguir representaciones gráficas de las asociaciones entre las características seleccionadas en la etapa anterior y entre los diferentes grupos de pacientes crónicos.
- **Capítulo 5. Experimentos y resultados.** Resultados obtenidos con los métodos descritos en el capítulo anterior. En primer lugar, como resultado de la selección de características, se analizan las diferencias estadísticas entre grupos de población en diferentes estados de salud en un mismo año, y la evolución temporal de cada estado de salud de los considerados. En segundo lugar, se analizan los resultados del ACM, es decir, se analizan las asociaciones encontradas entre las características seleccionadas, así como entre los grupos de pacientes considerados a estudio.
- **Capítulo 6. Conclusiones y líneas futuras.** En este capítulo se presentan las principales conclusiones que se deducen de los análisis. También se plantean algunas líneas de trabajo futuro.

Capítulo 2

Cronicidad. Conceptos previos

En este capítulo se explican algunos conceptos previos relacionados con las enfermedades crónicas y el contexto en el que se desenvuelve este TFG. Estos conceptos son importantes para sentar las bases de los siguientes capítulos. Primero se describirán las enfermedades crónicas en estudio (Sección 2.1), luego los Sistemas de Clasificación de Pacientes (Sección 2.2), y por último, los sistemas de codificación de diagnósticos y fármacos (Sección 2.3).

2.1. Enfermedades crónicas

Según la OMS, "las enfermedades crónicas son enfermedades de larga duración y por lo general de progresión lenta". En la actualidad, estas enfermedades son una de las principales causas de mortalidad en el mundo. Las sociedades cada vez más envejecidas, y los estilos de vida poco saludables, están muy relacionados con la prevalencia de estas enfermedades. Se estima que en 2020 serán responsables del 73 % de las muertes a nivel global [10].

Según el informe del Ministerio de Sanidad, Consumo y Bienestar Social, *Hábitos de vida*, sobre los estilos de vida de la sociedad española en 2018, la mala alimentación, la falta de actividad física y el consumo de tabaco o alcohol, son algunos de los hábitos más comunes entre la población. Estos hábitos, son al mismo tiempo factores de riesgo asociados en gran medida a la aparición de enfermedades crónicas [11].

Como ya se ha comentado en el capítulo anterior, el auge de las enfermedades crónicas, junto con el aumento en la utilización de los recursos sanitarios de estos pacientes, supone un reto para los sistemas sanitarios en la actualidad. Las características que presentan los pacientes cró-

nicos, hacen que haya que plantearse una nueva forma de abordar estas patologías, incidiendo en mayor medida en la prevención, la atención integral y la continuidad asistencial. Las enfermedades crónicas, en general, no tienen un tratamiento con resolución inmediata, por tanto, el objetivo no está tan centrado en “curar” a estos pacientes, sino más bien, en aumentar su calidad de vida y evitar futuras complicaciones [12]. Con este TFG se pretende extraer conocimiento ligado a estos pacientes, como por ejemplo, la caracterización de la evolución de su estado de salud. Esto podría tenerse en cuenta para implementar medidas sanitarias con el fin de evitar o ralentizar un futuro empeoramiento en el estado de salud de estos pacientes.

En los últimos años se viene apreciando un aumento de la comorbilidad en los pacientes crónicos. La comorbilidad o pluripatología se refiere a la presencia de una o más enfermedades crónicas, además de la principal o dominante, en un mismo paciente que puede denominarse paciente crónico complejo. Las enfermedades crónicas, y en especial las pluripatológicas, se suelen caracterizar por una complejidad elevada y en algunos casos por el consumo de múltiples fármacos, complicando todavía más la situación de los pacientes crónicos complejos [13].

Como ya se ha mencionado en el Capítulo 1, la hipertensión, la diabetes y la depresión, son algunas de las principales cronicidades en España. Además, como se muestra en la Figura 2.1, según el *Institute for Health Metrics and Evaluation* de EEUU, la Enfermedad Pulmonar Obstructiva Crónica está entre las 10 principales causas de mortalidad en España en 2017.

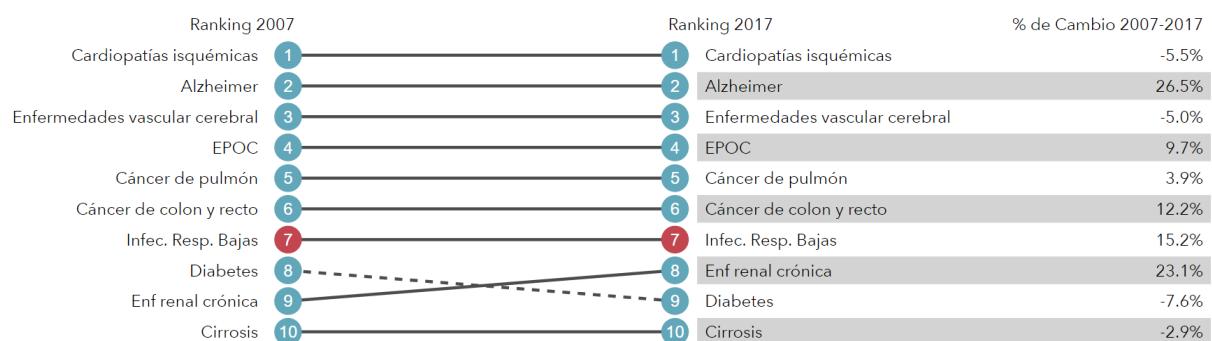


Figura 2.1: Principales causas de muerte en España: 2007 frente a 2017.
Obtenida de [14].

2.1.1. Hipertensión Arterial

La Hipertensión Arterial (HTA) es una enfermedad crónica que se caracteriza por un aumento de la presión arterial. La presión arterial, es la fuerza que la sangre ejerce contra las

paredes de las arterias y se mide en milímetros de mercurio (mm Hg). Esta medida, se puede tomar en dos instantes diferentes, cuando el corazón bombea sangre, entonces se obtiene la presión sistólica, y entre latido y latido, obteniendo el valor de la presión diastólica. El valor de la presión sistólica es más elevado que el de la diastólica. Para medir la tensión arterial, se deben conocer ambos valores [15].

En base a la pareja de valores obtenida de presión arterial, un paciente podría estar en uno de estos escenarios:

- Presión arterial normal: alrededor de 120/80 mm Hg.
- Presión arterial baja: inferior a 120/80 mm Hg.
- Presión arterial elevada o HTA: puede estar en estadio 1: entre 140/90 y 159/99 mm Hg; en estadio 2: entre 160/109 y 179/109 mm Hg; y en estadio 3: superiores a 179/109 mm Hg.

El tratamiento de la HTA suele consistir fundamentalmente en la mejora de los hábitos de vida, mejorando la alimentación y aumentando la actividad física. Adicionalmente, si esto no fuera suficiente el médico podría considerar necesario administrar algún medicamento. Las familias de medicamentos que se suelen usar para tratamientos hipertensivos son: diuréticos, inhibidores del sistema renina angiotensina (IECA), antagonistas de los receptores de angiotensina (ARA-II), calcioantagonistas o betabloqueantes [16].

2.1.2. Diabetes Mellitus

Reciben el nombre de Diabetes Mellitus (DM) una serie de trastornos que se caracterizan por la imposibilidad de regular los niveles de azúcar en sangre. Esto hace que generalmente, aparezcan altos niveles de glucosa en sangre, pudiendo producir graves consecuencias para el organismo.

La glucosa es una fuente de energía para músculos, tejidos y para el cerebro. El páncreas, se encarga de producir insulina y ésta es la encargada de transportar la glucosa, por medio de la sangre, al resto de órganos [17]. Un paciente con DM tiene problemas para hacer llegar la glucosa al resto del organismo. En función de la causa de este problema, se pueden encontrar varios tipos de DM:

- DM Tipo I. Suele diagnosticarse antes de los 35 años. El páncreas deja de generar insulina, por lo que la glucosa no puede transportarse hasta los tejidos y aumenta su concentración en sangre.

- DM Tipo II. En este caso, el páncreas sí que genera insulina, pero las células se vuelven resistentes a ella. La causa del desarrollo de esta resistencia no se sabe con seguridad, pero hay determinados factores de riesgo que están relacionados.
- DM gestacional o inducida. La DM gestacional, aparece durante el embarazo y suele desaparecer tras el parto. La inducida, aparece por el consumo de algún fármaco o por enfermedades genéticas.

Algunas comorbilidades o complicaciones asociadas a la DM, son: problemas cardiovasculares, daños en los nervios, daños renales, oculares, auditivos, en los pies, en la piel, o incluso depresión. La depresión puede incluso afectar al manejo de la enfermedad [18].

En cuanto al tratamiento, en la DM Tipo I consiste en la administración de insulina. En la DM Tipo II, el tratamiento comienza con un programa de dieta y ejercicio cardiosaludable y si no fuera suficiente, se comienza con la administración de antidiabéticos orales y en caso de ser necesario, administración de insulina.

2.1.3. Depresión

La depresión es un trastorno en el estado anímico, prolongado en el tiempo, y que afecta la capacidad de la persona de relacionarse y enfrentarse a las actividades cotidianas. Suele caracterizarse por sentimientos severos de tristeza, melancolía, vacío, pérdida, culpa o irritabilidad [19].

Los síntomas que suele presentar un paciente con depresión, son: irritabilidad, bajo estado de ánimo, disminución del interés en todas las actividades, cansancio o falta de energía, insomnio o exceso de sueño, cambios en el apetito y dificultades para concentrarse [20].

En cuanto al tratamiento, fundamentalmente puede ser de dos tipos: farmacológico o psicoterapéutico. Dependiendo del caso, puede aplicarse uno u otro, o combinar ambos. Los tratamientos farmacológicos, consisten en antidepresivos para regular los niveles en algunas sustancias del cerebro, generalmente en la serotonina. La serotonina es un neurotransmisor relacionado con el control de las emociones y estados de ánimo. Cada tipo de antidepresivo actúa de una forma diferente con el objetivo de incrementar los niveles de serotonina. Los tratamientos psicoterapéuticos, ayudan a los pacientes a cambiar la forma de pensar y sentirse, identificando comportamientos negativos y enseñando al paciente como reemplazarlos por otros más saludables [21].

Algunos de los medicamentos usados en los tratamientos farmacológicos son: inhibidores selectivos de la recaptación de serotonina, inhibidores de la recaptación de serotonina y norepinefrina, antidepresivos atípicos, antidepresivos tricíclicos e inhibidores de la monoaminooxidasa, entre otros [22].

Según algunos estudios [23], la depresión está muy relacionada con otras enfermedades crónicas de manera bidireccional. Por un lado, la ansiedad propia de ser diagnosticado con una enfermedad crónica, puede desencadenar en depresión. La depresión es frecuente en personas que sufren cáncer, epilepsia o Alzheimer, por ejemplo. Sin embargo, también sucede lo contrario, una persona que padece depresión, tiene mayor riesgo de sufrir DM, enfermedades cardiovasculares o derrames cerebrales, entre otras. Aunque las causas del desarrollo de estas comorbilidades todavía no está clara, parece ser que los fármacos usados en el tratamiento de algunas enfermedades crónicas pueden provocar depresión, y los cambios fisiológicos que se producen en el organismo de un paciente depresivo pueden contribuir al desarrollo de otras enfermedades [24].

2.1.4. Enfermedad Pulmonar Obstructiva Crónica

La Enfermedad Pulmonar Obstructiva Crónica (EPOC) se caracteriza por una inflamación de los pulmones que obstruyen el flujo de aire. Los síntomas son: dificultad para respirar, presión en el pecho, tos, mucosidad y silbido al respirar. Estos síntomas empiezan gradualmente asociados al esfuerzo y empeoran hasta aparecer en reposo [25].

La EPOC puede deberse a dos causas, una bronquitis aguda o un enfisema. La bronquitis, es una inflamación de los bronquios que deriva en una producción de mucosidad y tos continuada. Un enfisema destruye las paredes y fibras de los alvéolos, dificultando la respiración.

En cuanto a los factores de riesgo de esta enfermedad, algunos de los más destacados son: el tabaco, la contaminación en los entornos interiores o exteriores o la exposición a polvo, gases y productos químicos, entre otros. Además, pueden aparecer algunas complicaciones como infecciones respiratorias, problemas de corazón, cáncer de pulmón, presión arterial alta en las arterias pulmonares o depresión.

Los tratamientos para la EPOC pueden ser farmacológicos y terapeúticos. Estos tratamientos pueden ralentizar la evolución y reducir los índices de mortalidad. En cuanto a los farmacológicos, los medicamentos que se pueden administrar para el tratamiento de esta enfermedad son: broncodilatadores, esteroides inhalados, inhaladores combinados, esteroides orales, inhibidores de la fosfodiesterasa tipo 4, teofilina o antibióticos. En cuanto a los tratamientos terapeúticos,

se recomienda la oxigenoterapia o rehabilitación pulmonar [26].

Según la OMS, la EPOC causó en 2015 3.17 millones de muertes en todo el mundo, y la prevalencia de la EPOC ascendió en 2016 a 251 millones de casos. Además, se prevé que la presencia de la EPOC continúe creciendo debido al tabaquismo y al envejecimiento de la población. Estos datos, hacen que la EPOC, junto con la HTA, la DM y la depresión, sea otra de las cronicidades de especial interés en el entorno sanitario y se haya seleccionado para el análisis realizado en este TFG.

2.2. Sistema de Clasificación de Pacientes poblacional

Los Sistemas de Clasificación de Pacientes (SCP) son sistemas que tratan de clasificar a los pacientes en un número determinado de grupos homogéneos. Los individuos pertenecientes a un mismo grupo, comparten características clínicas similares y tienen un consumo de recursos sanitarios parecido. Agrupar a los pacientes en estratos similares, puede ayudar al análisis de datos y a comparar grupos de pacientes y resultados.

Existen distintos SCPs según el criterio de clasificación, el ámbito y el objetivo que se persigue. Inicialmente, estos sistemas nacen en el entorno hospitalario para controlar el consumo de recursos y clasificar los episodios de hospitalización. Con el aumento de la actividad ambulatoria, y desde un punto de vista más clínico, surgen nuevos sistemas para clasificar pacientes, con y sin hospitalización, incorporando otros aspectos como la gravedad, la probabilidad de fallecimiento, o el consumo en farmacia, o para integrar de forma global todo el proceso de cuidados de un paciente en los distintos niveles asistenciales [27].

Algunos ejemplos de estos sistemas son: el sistema de clasificación de pacientes hospitalarios Grupos Relacionados por el Diagnóstico (GRDs); el sistema de clasificación de pacientes hospitalarios y ambulatorios *International Refined GDR* (IR-GRDs); el sistema de clasificación de pacientes en atención primaria *Ambulatory Care Groups* (ACGs); o el sistema de clasificación global de pacientes *Clinical Risk Groups* (CRGs) [27].

Este TFG se centra en el sistema CRG. El sistema de clasificación CRG se caracteriza principalmente porque no clasifica episodios clínicos, sino a los propios pacientes en base a sus características clínicas, demográficas y farmacológicas. Por esta razón, es un SCP poblacional. La asignación de un paciente a un determinado grupo, se hace en base a todos los contactos que ese paciente ha tenido con el sistema sanitario, bien sea en atención primaria o especializada, durante un periodo de tiempo, normalmente un año.

El software que se utiliza para clasificar a pacientes mediante el sistema CRG, pertenece a la compañía multinacional *Minnesota Mining and Manufacturing Company* (3M). Este software, usa datos de diagnósticos y procedimientos clínicos, información farmacéutica y del estado de salud, para asignar a cada individuo a un grupo único y mutuamente excluyente (estos grupos también reciben el nombre de CRGs). Cada grupo está asociado a un estado de salud con validez clínica en un periodo de tiempo acotado, típicamente un año.

Los individuos clasificados dentro de un mismo CRG tienen características clínicas y un consumo de recursos similares. El objetivo de los CRGs, es facilitar tanto el análisis del estado de salud de los pacientes como del aspecto financiero, como por ejemplo, el consumo de recursos. A través de este sistema de clasificación se pueden identificar pacientes con un alto nivel de complejidad médica. Este sistema está centrado en el paciente, identificando estados de salud, y midiendo los niveles de comorbilidad.

Cada CRG, es decir, cada grupo mutuamente excluyente, se identifica mediante 5 dígitos. El primero, hace referencia al estado de salud general, los tres siguientes conforman el CRG-base que identifica una patología concreta, y el último, indica el nivel de severidad o gravedad [28].

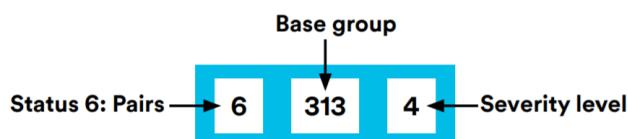


Figura 2.2: Estructura de un CRG.
Obtenida de [28].

Hay 9 grupos principales de estados de salud, desde el estado 1 que indica pacientes sanos, al estado 9 que hace referencia a pacientes en estado catastrófico. A partir del estado 3, se identifican las enfermedades crónicas. En la Figura 2.2, a modo de ejemplo, se desglosa la estructura del CRG 6-313-4. El 6 indica el estado de salud principal “enfermedad crónica significativa en múltiples sistemas orgánicos”. El estado de salud principal, junto con el grupo base, se denomina CRG-base. El CRG-base 6313 indica “HTA y DM”. Finalmente, el último dígito indica el nivel de gravedad: nivel 4 de entre 6 posibles niveles en ese CRG-base.

Originalmente, el software del 3M surge a finales de 1990 para uso del equipo del 3M *Clinical and Economic Research* (CER). Sin embargo, el lanzamiento formal de la primera versión no tiene lugar hasta el año 2000. A partir de entonces, el software ha ido sufriendo modificaciones y se han lanzado varias actualizaciones hasta llegar a la versión 2.0 en Abril del 2016. Es necesario puntualizar que el análisis llevado a cabo en este TFG se realiza sobre los

CRGs definidos en la Versión 2.0. Los códigos CRGs varían entre la Versión 1.0 y la Versión 2.0. Los CRGs-base que incluimos en el estudio son el CRG-base 1000 (pacientes sanos), CRG-base 3755 (depresión), CRG-base 5133 (enfermedad pulmonar obstructiva crónica y bronquiectasia), CRG-base 5192 (HTA), CRG-base 5424 (DM), CRG-base 6313 (HTA y DM), CRG-base 7140 (HTA, DM y otra cronicidad). Se ha decidido no incluir en el estudio el último dígito (nivel de gravedad), decisión que fue tomada junto con los colaboradores del HUF, para el análisis realizado en este TFG, pero que podría extenderse en un futuro.

2.3. Codificación de diagnósticos y fármacos

Como se ha comentado en la sección anterior, en este TFG se usa el sistema de clasificación CRG para determinar el estado de salud de pacientes en base a características clínicas y demográficas. Las características clínicas que consideraremos en el estudio serán diagnósticos y fármacos. Para la identificación y análisis de estas características se usan sistemas de codificación, en concreto, el sistema de Clasificación Internacional de Enfermedades (CIE) en el caso de diagnósticos, y el Sistema de Clasificación Anatómica, Terapéutica, Química (ATC) en el caso de fármacos.

2.3.1. Clasificación Internacional de Enfermedades

El sistema CIE es una herramienta utilizada para la clasificación y codificación de diagnósticos, enfermedades y lesiones, procedimientos clínicos, factores que influyen en el estado de salud, causas externas, o localización de neoplasias [29]. Este instrumento permite compartir información mediante un lenguaje común entre los profesionales del mundo sanitario.

Este sistema de codificación ha ido sufriendo modificaciones a lo largo de los años, utilizándose actualmente el CIE-10 (CIE-10^a revisión). Sin embargo, los datos que se usan en este TFG corresponden a un periodo temporal donde se usaba CIE-9-MC (CIE-9^a revisión-Modificación Clínica). Por este motivo, en los capítulos posteriores se usará la nomenclatura de la codificación CIE-9-MC.

La codificación CIE-9, se estructura en base a 5 bloques [30]:

- **Clasificación de enfermedades y lesiones.** A su vez este bloque, se divide en otros 17. Cada código dentro de este bloque, está formado por 3 dígitos, un punto, y uno o dos dígitos más (que indican la subcategoría). El primer código es el 001.0 (Cólera debida a

Vibrio cholerae) y el último el 999.9 (Otras complicaciones y complicaciones no especificadas de cuidados médicos, no clasificadas bajo otro concepto). En este TFG tendremos en cuenta sólo los 3 primeros dígitos.

- **Códigos V.** Este bloque contiene la clasificación suplementaria de factores que influyen en el estado de salud y contacto con servicios sanitarios. Los códigos dentro de este bloque, también están constituido por 3 caracteres, una letra V y dos números, seguidos de un punto, y uno o dos dígitos más. Por ejemplo, el código V01.0 indica “Cólera”, y el código V01.71 indica “Varicela”. En este TFG tendremos en cuenta sólo los 3 primeros caracteres.
- **Códigos E.** Para la clasificación suplementaria de causas externas de lesiones e intoxicaciones. Cada código dentro de este bloque, esta constituido por la letra E seguida de 3 dígitos. Por ejemplo, el código E000 indica “estado asociado a causa externa”.
- **Códigos M.** Este bloque hace referencia a la morfología de las neoplasias. Los códigos en este bloque se forman con un M seguida de 3 dígitos y van desde el código M800 (Neoplasias NEOM) hasta el código M997 (Enfermedad linfoproliferativa crónica).
- **Clasificación de procedimientos.** Estos códigos se crean con 2 números, y van desde el código 00 que indica “procedimientos e intervenciones, no clasificados bajo otros conceptos” hasta el 99, “otros procedimientos no quirúrgicos”.

Para la realización de este TFG se han considerado los códigos de clasificación de enfermedades y lesiones, los códigos V, los códigos E y los códigos M. Con esta selección, se obtienen un total de 1517 códigos diagnósticos, o dicho de otro modo, 1517 características que estudiar en la población.

2.3.2. Sistema de Clasificación Anatómica, Terapéutica, Química

El sistema de codificación ATC es un sistema que se usa para la clasificación de sustancias farmacológicas y medicamentos [31]. Este sistema asigna a cada fármaco un código ATC en base al sistema orgánico sobre el que actúa, el efecto farmacológico, las indicaciones terapéuticas y la estructura química. Cada código ATC, está formado por cinco niveles y un total de 7 caracteres.

El primer nivel, es una letra y hace referencia al grupo anatómico principal. Puede ser: A (Tracto alimentario y metabolismo), B (Sangre y órganos hematopoyéticos), C (Sistema cardiovascular), D (Dermatológicos), G (Sistema genitourinario y hormonas sexuales), H (Preparados hormonales sistémicos, excl. hormonas sexuales e insulinas), J (Antiinfecciosos para uso sistémico), L (Agentes antineoplásicos e inmunomoduladores), M (Sistema musculoesquelético), N (Sistema nervioso), P (Productos antiparasitarios, insecticidas y repelentes), R (Sistema Respiratorio), S (Órganos de los sentidos), y por último, V (Varios). El nivel 2, indica el subgrupo terapéutico y está formado por 2 números. El nivel 3, hace referencia al subgrupo terapéutico farmacológico identificado por una letra. En el nivel 4, se especifica el subgrupo químico-terapéutico a través de otra letra. Y por último, el nivel 5, indica el principio activo del ATC y se forma con dos números [32].

En la Figura 2.3 se muestra a modo de ejemplo el código ATC A10BA02 (Metformina) desglosado para cada nivel. Como se ha indicado, cada código se forma con 7 caracteres, pero para reducir el número de códigos a analizar, en este TFG se tienen en cuenta los cinco primeros, ignorando los dos últimos dígitos, es decir, los que hacen referencia al principio activo. De esta forma, contamos con un total de 746 posibles códigos ATC.

Código ATC: A10BA02		
Nivel 1	A	Tracto alimentario y metabolismo
Nivel 2	A10	Fármacos usados en diabetes
Nivel 3	A10B	Fármacos hipoglucemiantes orales
Nivel 4	A10BA	Biguanidas
Nivel 5	A10BA02	Metformina

Figura 2.3: Código ATC A10BA02 desglosado por niveles.

Capítulo 3

Bases de datos y análisis exploratorio

En la Sección 3.1 de este capítulo se describen las bases de datos usadas para el desarrollo del TFG. Después se muestra la evolución temporal de la tasa de presencia de algunos diagnósticos y fármacos para cada grupo de pacientes. Por último, en la Sección 3.2, se muestran los resultados del análisis exploratorio para caracterizar a los grupos de pacientes crónicos incluidos en el estudio.

3.1. Bases de datos

Para llevar a cabo este TFG se ha contado con varias fuentes de datos. Es preciso recalcar, que todas ellas han sido previamente anonimizadas y a cada paciente se le ha asignado un identificador único. El proceso de anonimización se ha llevado a cabo por los colaboradores del HUF. Mediante este identificador es posible hacer un seguimiento de cada paciente. Los datos analizados corresponden al Área de Salud que gestiona el propio hospital, es decir, se cuenta con datos de los municipios de Fuenlabrada, Moraleja de Enmedio y Humanes.

En primer lugar, se ha tenido acceso a seis bases de datos, cada una asociada a un año diferente (desde 2010 hasta 2015). Cada base de datos recogen los contactos de un paciente con el sistema sanitario durante un año. Dicho contacto puede tener lugar tanto en atención primaria como especializada, y también en farmacia. Para cada contacto de un paciente con el SNS, se recogen los siguientes datos: identificador anónimo del paciente, la fecha del contacto, la fecha de nacimiento, el género, y por último, una lista de fármacos, diagnósticos y procedimientos derivados del contacto. Los datos de farmacia que se recogen en estas bases de datos corresponden a dispensación en toda la Comunidad de Madrid. Puesto que en las bases de datos de los contac-

tos no se tiene información del CRG en el que ha sido clasificado cada paciente, adicionalmente se cuenta con otro fichero donde se recoge la relación entre paciente (mediante su identificador anónimo) y grupo de riesgo clínico. Esta clasificación ha sido realizada por el personal del HUF usando el software que se mencionó en el Capítulo 2 (3M-CRGs), considerando todos los datos clínicos y demográficos de cada paciente. El resultado de la clasificación permite conocer el grupo de riesgo clínico al que pertenece cada paciente, utilizando para ello 5 dígitos (estado de salud principal + grupo base + nivel de gravedad). En este TFG se ha obviado el nivel de gravedad para que los grupos sean más numerosos y estudiar así el CRG-base.

Además de las fuentes de datos anteriormente descritas, se ha contado con otro fichero donde se recoge, por un lado, la relación entre cada identificador de paciente y un código que hace referencia al Centro de Salud (CS) al que está adscrito, y por otro, una lista con los códigos de los CS de Fuenlabrada, Moraleja de Enmedio y Humanes. De esta forma, se puede conocer si un paciente que ha tenido un contacto con el Área de Salud de Fuenlabrada está adscrito a algún CS de dicho área. En caso contrario, se determinará que es un paciente itinerante. Los pacientes itinerantes son pacientes que, en algún momento concreto han tenido algún contacto con el Área de Salud de Fuenlabrada, pero realmente no pertenecen a dicho área. Será complicado hacer un seguimiento de estos pacientes porque la información disponible es muy limitada, y por tanto no los tendremos en cuenta en este análisis.

Una vez conocidas las bases de datos y cómo se estructuran, se prosigue con la extracción de los datos de los pacientes clasificados en los CRGs de interés. Para ello, primero se acude al fichero con el resultado de la clasificación en CRGs y se realiza un filtrado para obtener la lista de identificadores de pacientes por grupo de interés y año de estudio. Como se mencionó en el Capítulo 2, los grupos de interés corresponden a los CRGs-base: 1000 (pacientes sanos), 3755 (pacientes con depresión), 5133 (pacientes con EPOC), 5192 (pacientes con HTA), 5424 (pacientes con DM), 6313 (pacientes con HTA y DM) y 7140 (pacientes con HTA, DM y otra cronicidad).

Una vez se tiene la lista de los identificadores que pertenecen a cada grupo de población, se acude a las bases de datos con los contactos y se extraen todos los datos asociados a cada identificador de paciente en cada año. Al finalizar este proceso, se obtienen tantas matrices de datos como CRGs-base de interés y años en estudio, con toda la información agrupada para cada paciente. En cada una de estas matrices, se recoge: el identificador, la lista de todos los diagnósticos y códigos ATC asociados a ese paciente durante todo el año, el género y la edad del paciente. En el Apéndice A se muestra una breve explicación de cómo se han extraído los datos.

A continuación, se realiza una limpieza de los datos. Se comprueba que no aparezcan identificadores repetidos, y sí es así, que tengan asociados el mismo CRG-base. En caso de haber identificadores repetidos con distinto CRG-base asociado, se descartarían estos pacientes y todos sus datos asociados. Durante un año, un paciente sólo puede estar agrupado en un CRG, ya que éstos son mutuamente excluyentes. Por último, se acude al fichero con los datos de los códigos de los CS, y para cada identificador se comprueba si su código pertenece a alguno de los municipios de interés. En caso de ser un paciente itinerante, también se eliminaría del estudio. Tras el filtrado, unificación de los datos y la limpieza de los mismos, se cuenta con una matriz con datos asociados a cada grupo de pacientes en estudio y para cada año de los considerados.

3.2. Análisis exploratorio

Antes de proceder con el análisis exploratorio se comprueba si hay grandes cambios en la tasa de presencia de los diagnósticos y fármacos más relevantes en los grupos de pacientes de interés entre 2010 y 2015. Los códigos CIE-9 (diagnósticos) y ATC (fármacos) seleccionados para este análisis temporal, han sido elegidos en colaboración con el personal del HUF, por ser algunos de los principales códigos asociados a las cronicidades en estudio. Los resultados se detallan a continuación.

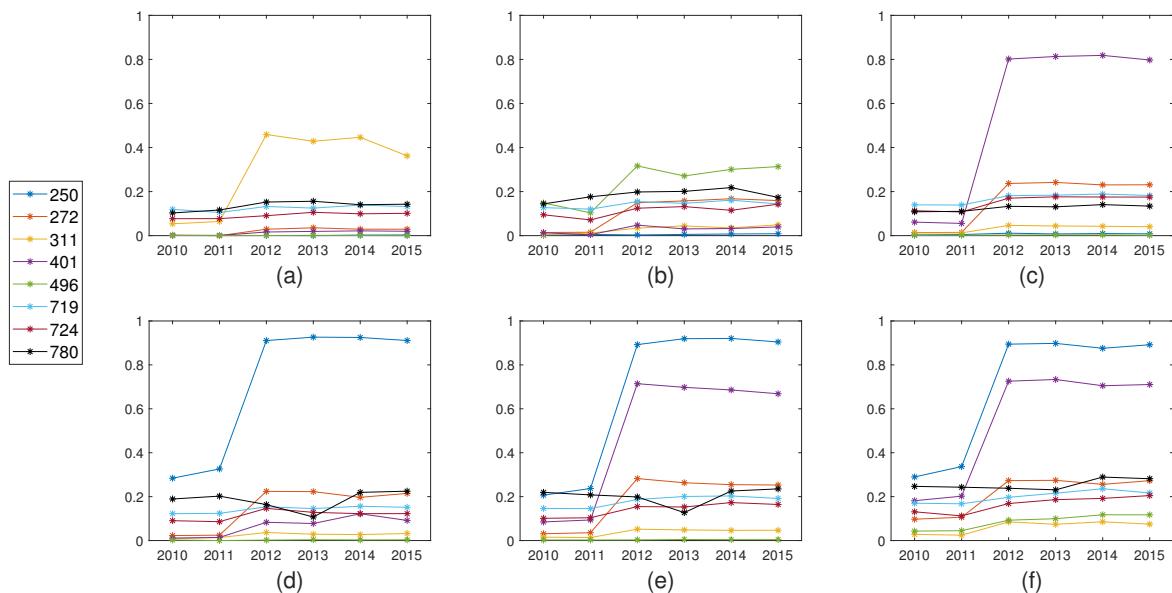


Figura 3.1: Evolución de la tasa de presencia de los diagnósticos 250, 272, 311, 401, 496, 719, 724, 780 para el CRG-base 3755 (a), 5133 (b), 5192 (c), 5424 (d), 6313 (e) y 7140 (f).

En la Figura 3.1, se muestra como varía la tasa de presencia de los códigos diagnósticos 250 (diabetes mellitus), 272 (trastornos del metabolismo de los lípidos), 311 (trastorno depresivo, no clasificado bajo otros conceptos), 401 (hipertensión esencial), 496 (obstrucción crónica de vías respiratorias, no clasificada bajo otros conceptos), 719 (otros trastornos y trastornos no especificados de articulación), 724 (otros trastornos y trastornos no especificados de la espalda) y 780 (síntomas generales), entre 2010 y 2015, para cada uno de los CRGs-base de interés. En el CRG-base 3755, el diagnóstico 311 (desorden depresivo), pasa de un 5% en 2010-2011 a un 45% en 2012, manteniéndose estable hasta 2015, que disminuye casi un 10%. En el CRG-base 5133, el diagnóstico 496 (obstrucción crónica de las vías respiratorias) tiene un aumento en la tasa de presencia entre 2011 y 2012, pasando de un 10% a un 30%. En cuanto al CRG-base 5192, el diagnóstico 401 (hipertensión esencial) sufre en 2012 un aumento, pasando de una presencia en 2010 de un 5% a un 80% en 2012. Lo mismo ocurre con el diagnóstico 250 (diabetes mellitus) en el CRG-base 5424, que pasa de una presencia del 30% en 2011 al 90% en 2012. Este comportamiento se repite en el CRG-base 6313 y en el CRG-base 7140, para los dos diagnósticos en cuestión (250 y 401). Además, a partir de 2012 la tasa de presencia se mantiene estable.

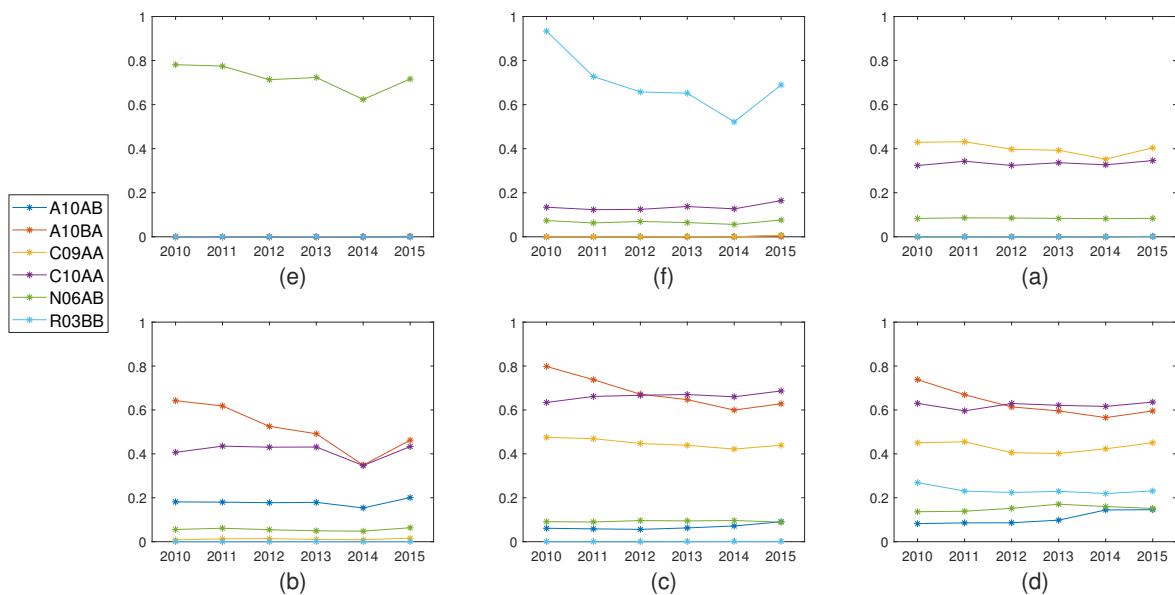


Figura 3.2: Evolución de la tasa de presencia de los códigos ATC A10AB, A10BA, C09AA, C10AA, N06AB y R03BB para el CRG-base 3755 (a), 5133 (b), 5192 (c), 5424 (d), 6313 (e) y 7140 (f).

La evolución en la tasa de presencia de los ATCs A10AB (insulinas y análogos de ac-

ción rápida para inyección), A10BA (biguanidas), C09AA (inhibidores de la enzima convertasa angiotensina), C10AA (modificadores de los lípidos), N06AB (antidepresivos) y R03BB (anti-asmáticos), se muestra en la Figura 3.2. En el CRG-base 3755, predominan los antidepresivos, cuya tasa de presencia se mantiene estable entre 2010 y 2013 en torno al 75 %, cae en 2014 hasta 60 % y remonta en 2015 hasta el 70 %. Dentro del CRG-base 5133, destacan los antiasmáticos con una bajada entre 2010 y 2014, que pasa de un 93 % a un 52 % y remonta en 2015 hasta el 70 %. En el CRG-base 5192, destacan los ATCs C09AA y C10AA que actúan sobre el sistema cardiovascular, y que se mantienen estables en torno al 40 % y 30 %, respectivamente. En el CRG-base 5424 sobresalen los ATCs A10BA y C10AA, para el primero, la tasa disminuye entre 2010 y 2015, pasando de un 64 % a un 34 %, y asciende en 2015 hasta el 46 %, en cuanto al segundo, se mantiene estable entre 2010 y 2013, en un 43 %, baja en 2014 hasta 34 % y remonta en 2015 hasta el 43 %. En los CRGs-base 6313 y 7140, destacan los fármacos A10BA, C09AA y C10AA. En ambos CRGs-base el ATC A10BA sigue una tendencia decreciente pasando de aproximadamente un 75 % de presencia a un 60 %; el ATC C10AA, aumenta ligeramente de un 60 % a un 68 %; y el C09AA se mantiene estable alrededor del 45 %.

Después de analizar la evolución en la tasa de presencia de algunos códigos, se concluye que hay diferencias significativas entre 2010 y 2011, y el resto de años, que afectan sobre todo a la tasa de presencia de diagnósticos. Además, a partir de 2012, las tasas se estabilizan. Después de consultar con el personal del HUF a qué podrían deberse estos cambios, se concluye que no hay una justificación clínica detrás, y qué podrían deberse, en mayor medida, a la forma en la que se recogían los datos. Por este motivo, se elige 2012 para mostrar los resultados del análisis exploratorio y no extender demasiado la memoria. El objetivo de este análisis es conocer más a fondo a los pacientes de los distintos grupos de población.

Para una mejor interpretación de los resultados mostrados a continuación, primero se realiza un recuento del número de pacientes por cada CRG-base de interés. Como ya se ha comentado, los grupos seleccionados para el estudio son el CRG-base 1000, que engloba a los pacientes sanos, el CRG-base 3755 con pacientes con depresión, el CRG-base 5133 que agrupa pacientes con EPOC, el CRG-base 5192 para pacientes con HTA, el CRG-base 5424 con pacientes con DM, el CRG-base 6313 para HTA y DM, y el CRG-base 7140 con las dos cronicidades anteriores y otra más.

En la Figura 3.3, se muestra la evolución del número de pacientes entre 2012 y 2015 para los siete CRGs-base de interés. Para apreciar mejor las diferentes escalas, en la Figura 3.3(a), se representa el número de pacientes para el CRG-base 1000 y para el CRG-base 5192, y en la Figura 3.3(b) para el resto de CRGs-base. El número de pacientes sanos se encuentra en los

4 años en torno a los 77.000 pacientes, y el de HTA en torno a 12.500. En cuanto al resto de CRGs-base, el número anual de pacientes con depresión varía entre 1.500 y 1.700 pacientes (aproximadamente), el número de pacientes con EPOC está alrededor de los 1.000 pacientes y los diabéticos en torno a los 2.500, aproximadamente. Por otro lado, en cuanto a los CRGs-base asociados a crónicos complejos, el número de pacientes con HTA y DM asciende a los 4.000 pacientes los 4 años, y los que además tienen otra cronicidad, oscilan entre los 700 y los 1.000 pacientes.

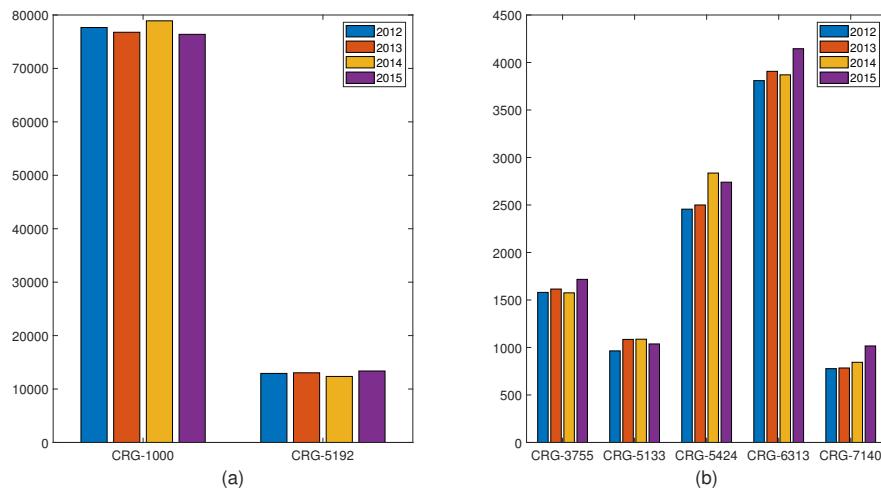


Figura 3.3: Número anual de pacientes entre 2012 y 2015 para los CRGs-base 1000 y 5192 (a), y para el resto de CRGs-base de interés (b).

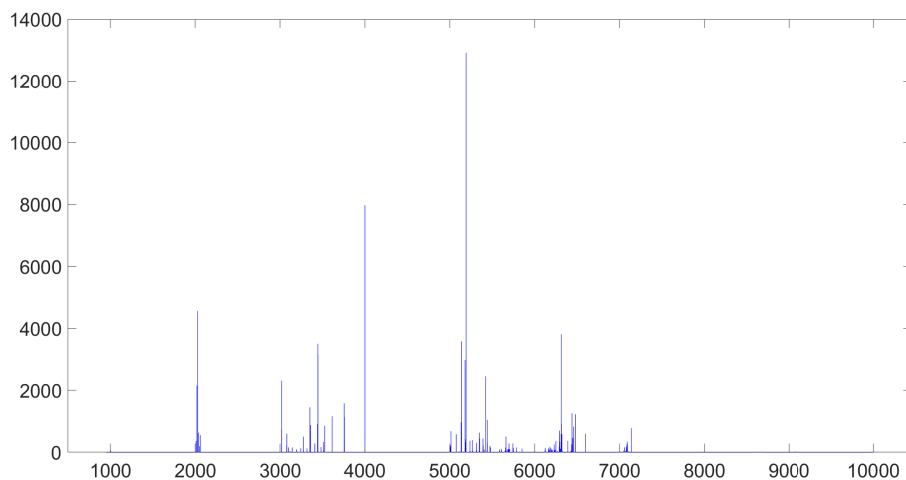


Figura 3.4: Número total de pacientes por CRG-base durante el año 2012, omitiendo el CRG-base 1000 para propósitos de visualización.

Además del número de pacientes, también resulta interesante conocer los CRGs-base que mayor número de pacientes agrupan, y de esta forma conocer el estado de salud general de la población de la que se tiene información y las patologías más relevantes. En la Figura 3.4, se muestra el número total de pacientes por CRG-base durante el año 2012. Nótese que se ha omitido el número de pacientes sanos para apreciar mejor el número de pacientes asignados al resto de CRGs-base. Los CRGs con más de 3.000 pacientes son los siguientes; CRG-base 2030 (una enfermedad aguda significativa, excluyendo enfermedades no transmisibles), 3445 (hiperlipidemia), 3446 (enfermedad tiroidea crónica), 4000 (múltiples cronicidades primarias menores), 5138 (asma), 5192 (HTA), y 6313 (HTA y DM).

Por otro lado, para conocer los patrones de uso de los recursos sanitarios de los pacientes de cada grupo, se analiza el número medio de contactos anuales con el sistema sanitario por paciente (Figura 3.5). En el CRG-base 3755, es de unos 15 contactos al año; en los CRGs-base 5133, 5192, y 5424, es de casi 20 contactos; en el CRG-base 6313 aumenta a unos 23 contactos anuales de media; y en el CRG-base 7140 asciende hasta los 33, concluyendo de este modo, que al aumentar el nivel de cronicidad, o empeorar el estado de salud, el número de contactos con el sistema sanitario aumenta, como cabría esperar.

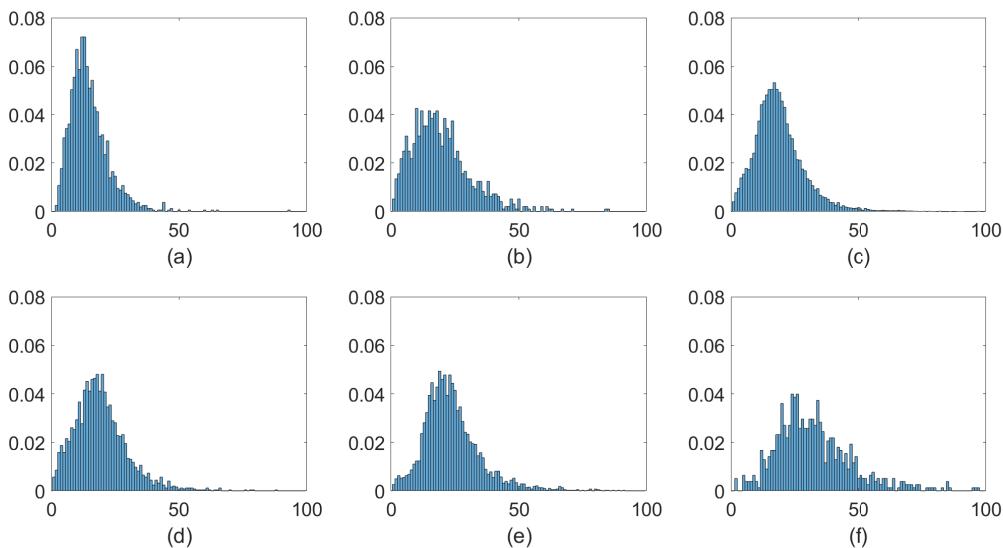


Figura 3.5: Distribución de probabilidad del número de contactos por paciente para el CRG-base 3755 (a), 5133 (b), 5192 (c), 5424 (d), 6313 (e), y 7140 (f) en 2012.

3.2.1. Características demográficas y clínicas

En cuanto a la edad de los pacientes de los CRGs-base de interés (Figura 3.6), la edad media de los pacientes en el grupo 3755 es de unos 42 años, y en el 5133 de 45 años. En la Figura 3.6 (b), que corresponde al CRG-base 5133, llaman la atención dos máximos, uno en torno a los 2 años, y otro en torno a los 55 años, en el CRG-base 5133 se incluyen dos patologías (EPOC y bronquiectasia) y quizás haya dos tipos de pacientes distintos en este CRG. Los pacientes con HTA (CRG-base 5192) tienen una edad media de 57 años, mientras que en los diabéticos (CRG-base 5424) es de 52 años. Para el CRG-base 6313, la edad media asciende a 62 años, y para el CRG-base 7140 es de 67 años. Al igual que sucedía en el caso de los contactos, la edad aumenta al considerar grupos asociados a peores estados de salud, pasando de los 42 años en el estado de salud 3 a los casi 70 años en el estado de salud 7.

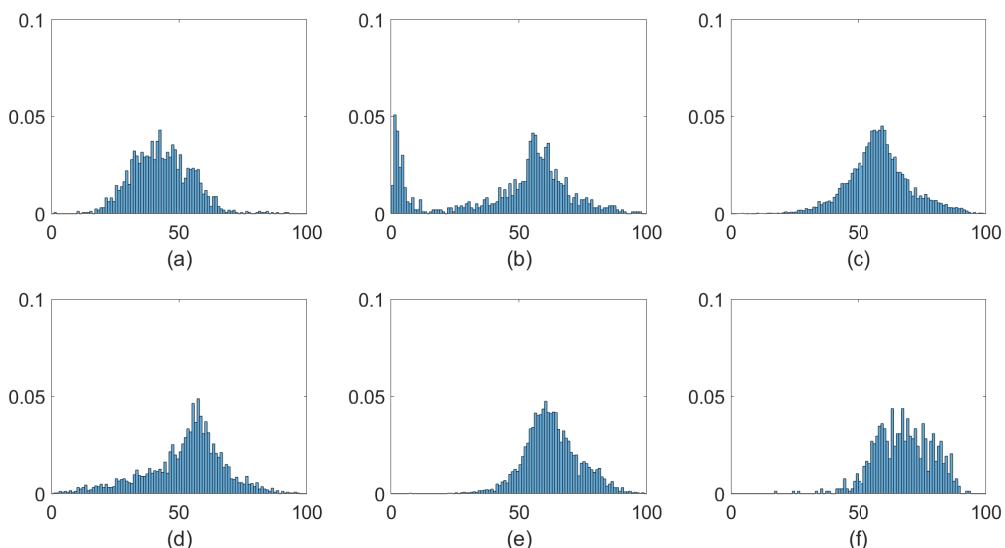


Figura 3.6: Distribución de probabilidad de la edad de los pacientes de los CRG-base: 3755 (a), 5133 (b), 5192 (c), 5424 (d), 6313 (e), y 7140 (f) en 2012.

En cuanto al género de los pacientes (Figura 3.7), en el CRG-base 3755 predomina el género femenino con un 73 % frente al 27 % masculino. En el CRG-base 5133, hay un mayor número de hombres (59 %) que de mujeres (41 %). En el CRG-base 5192, hay un 55 % de mujeres frente a un 45 % de hombres, al contrario que en el CRG-base 5424, que hay un 60 % de hombres frente a un 40 % de mujeres. En el CRG-base 6313, hay un 54 % de hombres y un 46 % de mujeres. Y por último, en el CRG-base 7140, el porcentaje de mujeres asciende al 58 % frente al 42 % de hombres. Se concluye que, en los grupos de depresión y de HTA hay un mayor porcentaje

de mujeres, mientras que en los grupos de EPOC y DM hay un mayor número de hombres. Además, la diferencia de porcentaje en el caso de la depresión, podría estar asociada a que la mujer acude antes a pedir ayuda cuando tiene problemas emocionales.

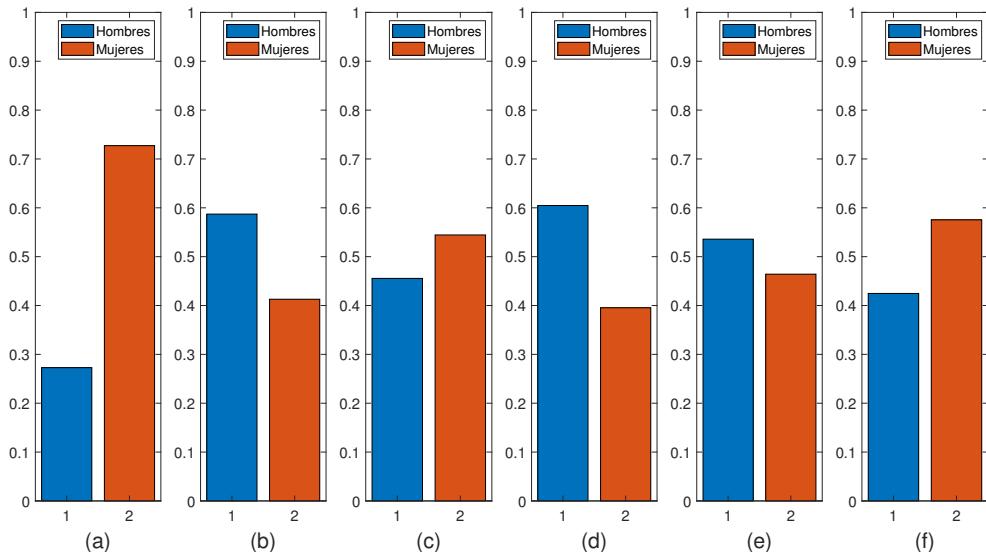


Figura 3.7: Distribución de probabilidad del género de los pacientes de los CRG-base: 3755 (a), 5133 (b), 5192 (c), 5424 (d), 6313 (e), y 7140 (f) en 2012.

Una vez analizadas las características demográficas, se prosigue con el análisis de algunas características clínicas. En concreto, se analiza la presencia de determinados diagnósticos y fármacos (códigos CIE-9, en el caso de diagnósticos, y códigos ATC para fármacos, como se mostró en el capítulo anterior). Para este análisis, se realiza una lista de diagnósticos y otra de ATCs, relacionados con DM, con HTA, con depresión y con EPOC. Con estas listas, se realiza un recuento de pacientes que presentan alguno de los códigos considerados, por un lado, para diagnósticos, y por otro, para fármacos. Las listas de códigos considerados en los recuentos se muestran en los Apéndices B y C. Estas listas han sido realizadas en colaboración con personal del HUF, seleccionando aquellos diagnósticos y fármacos de mayor interés. En el Cuadro 3.1 se pueden observar todos los resultados.

En cuanto al CRG-base 3755, en el año 2012, los pacientes clasificados en este grupo, no poseen prácticamente ningún diagnóstico, ni ATC, relacionado con HTA, DM, ni EPOC. Por otro lado, casi el 90 % de los pacientes con depresión tienen algún código ATC relacionado con depresión, sin embargo, sólo en torno al 50 % tienen algún diagnósticos de depresión (de los considerados a estudio, véase Apéndice B).

Los pacientes con EPOC (CRG-base 5133) no cuentan prácticamente con diagnósticos rela-

cionados con HTA, DM, o depresión, pero un 71 % de los pacientes presenta algún diagnóstico de EPOC. En cambio, en cuanto a ATCs, no cuentan prácticamente con ATCs de HTA, ni DM, pero un 81 % de los pacientes tiene algun ATC relacionado con EPOC, y un 11 % posee códigos asociados a depresión.

En el CRG-base 5192, ningún paciente posee diagnósticos de DM, depresión, ni EPOC, pero sí de HTA (80 % de los pacientes). Tampoco poseen ATCs de DM, ni EPOC, sin embargo sí de depresión (13 % de los pacientes) y un 92 % tiene algún ATC relacionado con HTA.

En relación a los pacientes en el CRG-base 5424, no presentan diagnósticos de HTA, ni depresión, ni EPOC, sólo relacionados con DM (91 %). En cuanto a códigos ATCs, un 82 % de pacientes presenta algún código relacionado con DM, y algunos ATCs relacionados con HTA y con depresión (casi un 9 % de pacientes, respectivamente).

Como cabría esperar, para los CRGs-base 6313 y 7140, se obtienen resultados similares, ya que ambos agrupan pacientes con HTA y DM. En ambos casos, hay un alto porcentaje de pacientes con diagnósticos y códigos ATC de HTA y DM. En cuanto a los diagnósticos de depresión y EPOC, hay un pequeño porcentaje de pacientes que presentan alguno de los diagnósticos considerados. Es destacable que la presencia de diagnósticos de EPOC pasa de un 5 % en el CRG-base 6313, a un 20 % en el CRG-base 7140 y para el caso de la depresión pasa de un 5 % en el CRG-base 6313, a un 8 % en el CRG-base 7140. En cuanto a los ATCs de depresión y EPOC, en el CRG-base 6313 hay un 15 % de pacientes con algún ATC relacionado con depresión, y un 5 % con algún ATC de EPOC. En el CRG-base 7140, el porcentaje de pacientes con algún ATC relacionado con depresión aumenta hasta un 23 %, y en el caso de EPOC, hasta un 28 %. El CRG-base 7140 agrupa a pacientes con HTA, DM y otra cronicidad, por tanto podría concluirse del aumento de diagnósticos y ATCs de depresión y EPOC, que la tercera cronicidad en algunos pacientes podría ser bien depresión o EPOC.

Adicionalmente, se repitió este análisis pero en lugar de estudiar la ausencia o presencia de los códigos incluidos en el análisis, se halló la distribución del número total de códigos por paciente en cada grupo en estudio. Estos resultados se muestran en el Apéndice D.

	% de códigos relacionados con depresión		% de códigos relacionados con EPOC		% de códigos relacionados con HTA		% de códigos relacionados con DM	
	Diagnósticos	ATCs	Diagnósticos	ATCs	Diagnósticos	ATCs	Diagnósticos	ATCs
CRG-base 3755	45,89	89,43	3,35	2,85	1,65	0,89	0,13	0,00
CRG-base 5133	3,42	11,51	71,47	81,54	4,77	2,59	0,31	0,00
CRG-base 5192	4,64	13,11	2,92	2,90	80,43	92,42	1,10	0,00
CRG-base 5424	3,66	9,00	2,81	2,28	8,79	4,52	91,08	82,61
CRG-base 6313	5,17	14,44	4,57	5,49	71,67	97,27	89,24	90,39
CRG-base 7140	8,49	23,17	20,46	28,44	72,97	97,30	89,45	89,45

Cuadro 3.1: Porcentaje de pacientes con diagnósticos y códigos ATCs relacionados con depresión, EPOC, HTA y DM para los CRGs-base de interés.

3.2.2. Perfiles de estados de salud

Una vez analizadas algunas características demográficas y clínicas de los grupos de pacientes de interés, se prosigue con la obtención de perfiles de los grupos de interés en base a sus diagnósticos y ATCs, en términos de presencia, y en términos de ocurrencia.

Un perfil indica la proporción de presencia u ocurrencia de las características estudiadas en una población. En el caso de este TFG, las características estudiadas son los 1517 diagnósticos y los 746 fármacos que se incluyen en el estudio. Cuando se habla de presencia se hace referencia a si un paciente presenta o no en su historial una determinada característica (con independencia del número de veces que se presenta). Cuando se habla de ocurrencia, se incide en el número de veces que aparece el diagnóstico o el fármaco en el historial del paciente. Nótese que, estos perfiles se elaboran anualmente, por tanto cuando se habla de presencia u ocurrencia de una característica en el historial de un paciente, se refiere al año en cuestión.

Para elaborar estos perfiles en términos de presencia, primero se obtiene el vector de características de cada paciente (1517 diagnósticos + 746 códigos ATC). Para crear cada vector de características, se recorre la lista de diagnósticos y códigos ATC de cada paciente, y para cada código de los analizados se le asigna un “0” si no lo presenta, y un “1” en caso contrario. En la Figura 3.2 se muestra un ejemplo de cómo se crean los vectores de características, en concreto para los pacientes del CRG-base 3755 durante el año 2012. Del vector de características asociado al paciente 1, se puede concluir que presenta los códigos ATC A01AA y el V30ZZ, y que no presenta los diagnósticos 001, 002, M997, ni el código ATC A01AB. Por último, para obtener el perfil de presencia, se calcula la proporción de pacientes que presentan cada uno de los códigos.

Para la elaboración de los perfiles en términos de ocurrencia, se repite el procedimiento anteriormente descrito. Primero, se crean los vectores de características, uno para cada paciente, pero en este caso, en lugar de asignar un “0” o un “1” a cada código de los estudiados, se asigna el recuento del número de veces que el paciente tiene ese código durante todo el año. Por último, se crea el perfil de ocurrencia como la proporción de ocurrencia de cada código frente al total de pacientes en cada grupo.

Analizando los perfiles en términos de presencia, se puede observar la tasa de presencia de cada diagnóstico y fármaco entre los pacientes de un grupo. Con esta información, se puede por ejemplo, elaborar un ranking de los diagnósticos con mayor presencia o los fármacos más dispensados en cada grupo de población, para cada año. Cuando analizamos los perfiles en términos de ocurrencia, lo que se deduce es el número medio de veces que se presenta un

CRG-base 3755 en 2012							
	Códigos-diagnósticos				Códigos-ATC		
	Código '001'	Código '002'	(...)	Código 'M997'	Código 'A01AA'	Código 'A01AB'	(...)
Paciente 1	0	0	(...)	0	1	0	(...)
Paciente 2	1	0	(...)	1	1	1	(...)
(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)
Paciente N	1	0	(...)	1	1	1	(...)

(a)

Perfil presencia del CRG-base 3755 en 2012	$\sum \text{Código '001'}$ N	$\sum \text{Código '002'}$ N	(...)	$\sum \text{Código 'M997'}$ N	$\sum \text{Código 'A01AA'}$ N	$\sum \text{Código 'A01AB'}$ N	(...)	$\sum \text{Código 'V30ZZ'}$ N
--	---------------------------------	---------------------------------	-------	----------------------------------	-----------------------------------	-----------------------------------	-------	-----------------------------------

(b)

Cuadro 3.2: Vectores de características (a) y perfil de presencia (b).

diagnóstico o fármaco. Cuando se analizan conjuntamente los perfiles de presencia y ocurrencia se podrían extraer conclusiones del tipo: “el código 001 está presente en el 80% de pacientes del grupo 3755 en el año 2012, y además entre los pacientes que sí lo presentan, en media aparece 4 veces”.

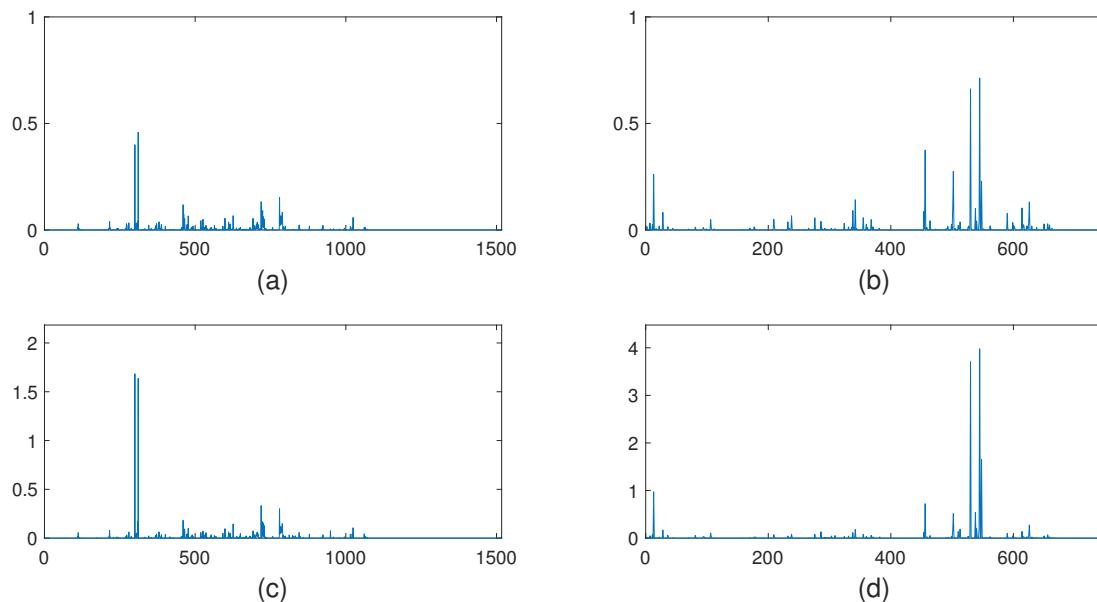


Figura 3.8: Perfiles de presencia de diagnósticos (a) y códigos ATC (b), y perfiles de ocurrencia de diagnósticos (c) y códigos ATC (d) del CRG-base 3755.

En cuanto al perfil del CRG-base 3755 (véase la Figura 3.8), los diagnósticos más presen-

tes en distintos pacientes de este grupo son: el código 311 (desorden depresivo) con una tasa de presencia del 45 %, el diagnóstico 300 (trastornos de ansiedad, dissociativos y somatomorfos) que lo presenta un 40 % de los pacientes y, en menor medida, los códigos 780 (síntomas generales), 719 (trastornos de la articulación), y 460 (nasofaringitis aguda) con una presencia del 15 %, 13 % y 11 %, respectivamente. Por otro lado, los ATCs con mayor presencia son: el N06AB (antidepresivos) con una presencia del 71 %, N05BA (ansiolíticos) con un 66 % de presencia, M01AE (antiinflamatorios) presente en el 37 % de los pacientes, N02BE (analgésicos) con un 27 % y el código A02BC (inhibidores de la bomba de protones) con un 26 %. Al analizar los perfiles de ocurrencia, se observa que los diagnósticos y fármacos con mayor presencia son también los que tienen una mayor ocurrencia. Por ejemplo, el número medio de veces que se presentan los diagnósticos 311 y 300 es de 1.6 veces al año, y el número medio de ansiolíticos y antidepresivos dispensados al año es de 4.

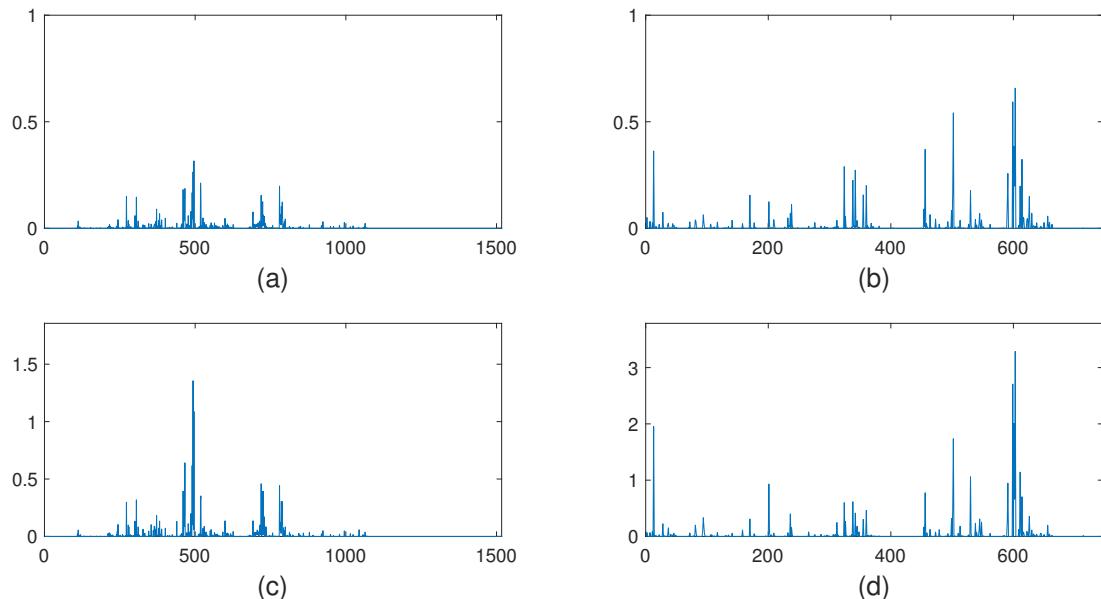


Figura 3.9: Perfiles de presencia de diagnósticos (a) y códigos ATC (b), y perfiles de ocurrencia de diagnósticos (c) y códigos ATC (d) del CRG-base 5133.

Entre los pacientes clasificados en el CRG-base 5133 (véase la Figura 3.9), los diagnósticos con mayor presencia son: el 496 (obstrucción crónica de las vías respiratorias) con una tasa de presencia del 31 %, el 493 (asma) con un 26 %, el 518 (otras enfermedades pulmonares) presente en el 21 % de los pacientes, el 780 (síntomas generales) con un 20 %, y el 466 (bronquitis aguda y bronquiolitis) con una presencia del 19 %. Entre los fármacos más presentes encontramos: el R03BB (antiasmáticos) con una presencia del 65 %, el R03AC (agonistas selectivos de receptores beta-2 adrenérgicos) dispensado al 59 % de pacientes, el N02BE (analgésicos)

entre el 54%, el R03AK (otros antiasmáticos) con un 38%, y el M01AE (antiinflamatorios) entre el 37 % de pacientes. Al analizar la ocurrencia, es decir, el número medio de veces que se presenta cada código, aparecen, además de los anteriores, los diagnósticos 490 y 719 (trastornos de la articulación) con una alta frecuencia. Además, se concluye que los ATCs mencionados se presentan más de una vez a lo largo del año, concretamente entre 3.2 veces (para el caso del R03BB) y 1.7 (en el caso del N02BE).

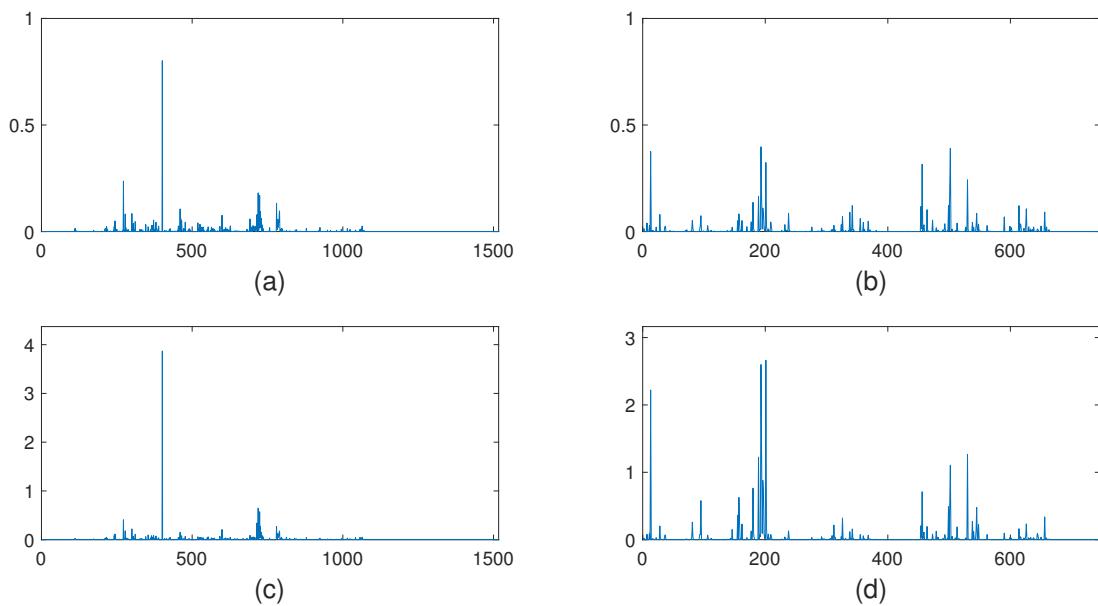


Figura 3.10: Perfiles de presencia de diagnósticos (a) y códigos ATC (b), y perfiles de ocurrencia de diagnósticos (c) y códigos ATC (d) del CRG-base 5192.

En el CRG-base 5192 (véase la Figura 3.10), los diagnósticos con mayor presencia son: el 401 (HTA) 80 %, el 272 (trastornos del metabolismo de los lípidos) 24 %, el 719 (trastornos de la articulación) 18 %, el 724 (trastornos de la espalda) 17%, y el 780 (síntomas generales) 13 %. En cuanto a los ATCs, los que tienen una mayor presencia, entre el 30 y 40 %, son: el C09AA (inhibidores enzima convertasa angiotensina), el N02BE (analgésicos), el A02BC (inhibidores de la bomba de protones), el C10AA (modificadores de los lípidos) y el M01AE (antiinflamatorios). Al analizar la ocurrencia, destaca el código 401, que aparece de media, 4 veces al año por paciente, el resto de códigos tiene una ocurrencia inferior a 1. En relación a la ocurrencia de ATCs, los códigos que mayor número medio de veces se repiten al año son: el C10AA, el C09AA, y A02BC, que también destacaban en presencia, y además los códigos C09BA (diuréticos) y N05BA (ansiolíticos) todos ellos tienen una ocurrencia de más de un código al año en media.

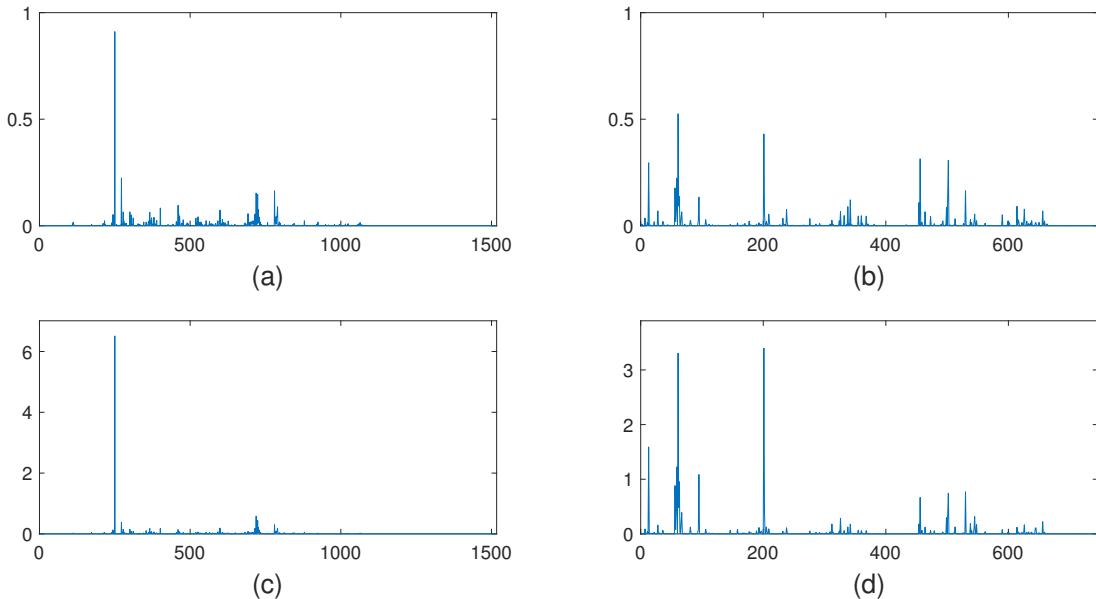


Figura 3.11: Perfiles de presencia de diagnósticos (a) y códigos ATC (b), y perfiles de ocurrencia de diagnósticos (c) y códigos ATC (d) del CRG-base 5424.

En los pacientes pertenecientes al CRG-base 5424 (véase la Figura 3.11), destacan por su presencia los códigos diagnósticos: 250 (diabetes mellitus) con una presencia del 91 %, el 272 (trastornos del metabolismo de los lípidos) con un 22 %, el 780 (síntomas generales) con un 16 %, 719 (trastornos de la articulación) que lo presentan el 15 % de los pacientes y el 724 (trastornos de la espalda) con un 14 %. Por otro lado, los ATCs con mayor presencia son: el A10BA (antidiabéticos orales) con una tasa de presencia del 52 %, el C10AA (modificadores de los lípidos) con un 43 %, y los códigos M01AE (antiinflamatorios), N02BE (analgésicos), y A02BC (inhibidores de la bomba de protones) con un 30 %. En términos de ocurrencia, en diagnósticos destaca el 250 que aparece una media de unas 6 veces al año por paciente, y en ATCs destacan los códigos C10AA y A10BA con una ocurrencia de más de 3 veces al año de media.

Para el CRG-base 6313 (véase la Figura 3.12), los diagnósticos con una presencia más alta son: el 250 (diabetes mellitus) con casi un 90 %, el 401 (hipertensión esencial) con una presencia del 71 %, el 272 (trastornos del metabolismo de los lípidos) con un 28 %, y los códigos 780 (síntomas generales) y 719 (trastornos de la articulación), ambos con una presencia cerca del 20 %. En cuanto a fármacos, con una presencia superior al 40 % destacan: el A10BA (antidiabéticos orales), el C10AA (modificadores de los lípidos), el A02BC (inhibidores de la bomba de protones), el C09AA (inhibidores enzima convertasa angiotensina), y el N02BE (analgésicos). Al analizar la ocurrencia, los códigos diagnósticos que más número de veces aparecen al año

son el 250 (casi 6 veces al año de media por paciente) y el 401 (con casi 3 diagnósticos al año por paciente de media). Para ATCs, destacan los códigos C10AA y A10BA que aparecen una media de entre 5 y 6 veces al año por paciente.

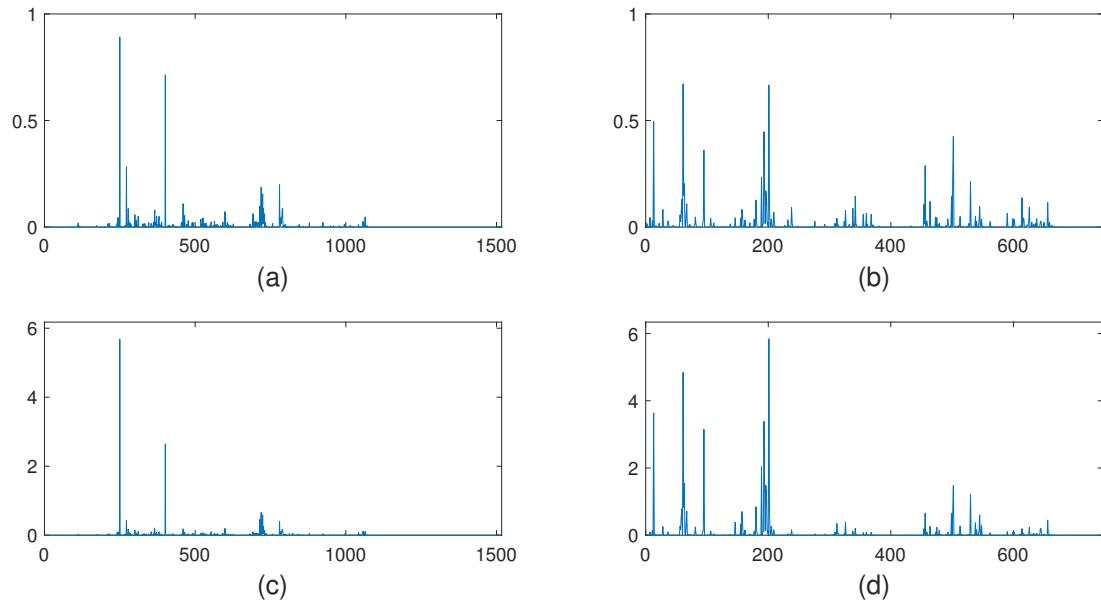


Figura 3.12: Perfiles de presencia de diagnósticos (a) y códigos ATC (b), y perfiles de ocurrencia de diagnósticos (c) y códigos ATC (d) del CRG-base 6313.

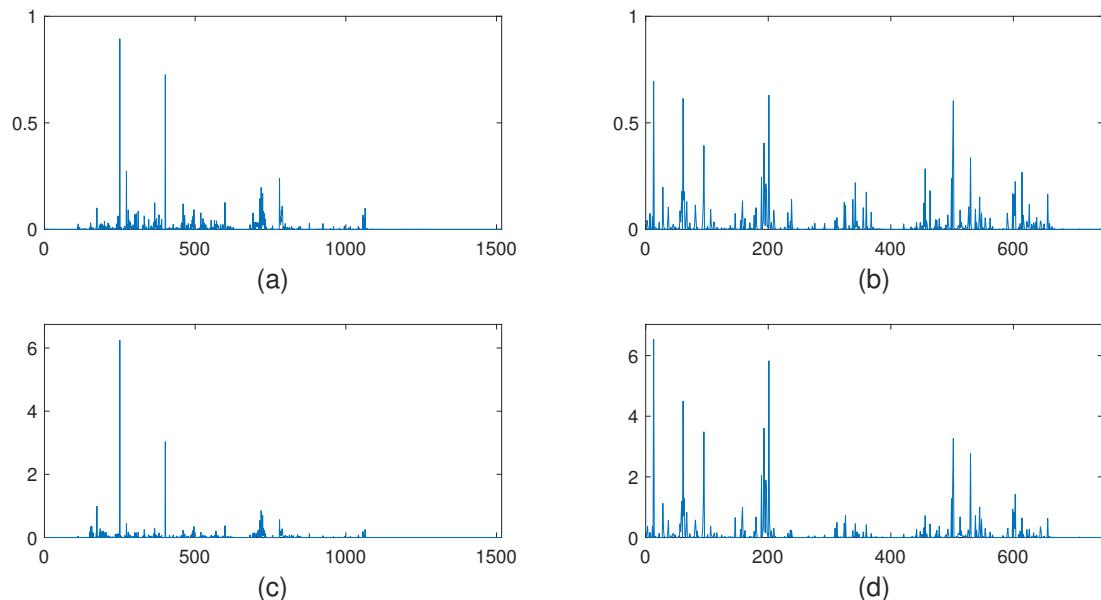


Figura 3.13: Perfiles de presencia de diagnósticos (a) y códigos ATC (b), y perfiles de ocurrencia de diagnósticos (c) y códigos ATC (d) del CRG-base 7140.

En el CRG-base 7140 (véase la Figura 3.13) en términos de presencia, los diagnósticos que más destacan son los mismos que en el CRG-base 6313 y con una presencia muy similar. Lo mismo sucede con los ATCs. En términos de ocurrencia también se obtienen resultados muy parecidos, salvo para el ATC A02BC (inhibidores de la bomba de protones), que en el CRG-base 6313 aparece unas 3 veces de media al año por paciente y en el CRG-base 7140 pasa a ser el doble.

Con estos resultados, se puede concluir que hay varios diagnósticos y ATCs se resaltan en casi todos los pacientes, por ejemplo, el diagnóstico 780 (síntomas generales) o el 719 (trastornos de la articulación). Estos diagnósticos no servirán para caracterizar a los diferentes grupos de pacientes crónicos, puesto que son características que presentan todos ellos, y que puede estar asociados a la edad (la edad media de estos pacientes es superior a la media de la población). Sin embargo, hay otros que caracterizan claramente a cada CRG-base, por ejemplo, el diagnóstico 311 (desorden depresivo), el 496 (obstrucción crónica de las vías respiratorias), el 401 (hipertensión esencial), o el 250 (diabetes mellitus).

Aunque el análisis exploratorio completo se ha mostrado sólo para el año 2012, se repite para el resto de años. Tal y como ya se evidenciaba al comienzo de esta sección, los resultados para el intervalo 2012-2015, eran muy similares. Sin embargo, para 2010 y 2011 había cambios significativos. Se ha incluido en el Apéndice E algunos de los resultados de 2010 y 2011, y su comparación con los años 2012 y 2013. Finalmente, se decide que el análisis que se realizará a continuación para seleccionar características y conseguir representaciones gráficas de sus asociaciones, se centrará en el intervalo temporal 2012-2015 y no se tendrán en cuenta los años 2010 y 2011.

Capítulo 4

Métodos

En la Sección 4.1 de este capítulo, se detallan los fundamentos teóricos asociados al test de proporciones y al remuestreo *bootstrap*, considerados para determinar diferencias significativas en la tasa de presencia de diagnósticos y fármacos de diferentes grupos de pacientes. Las características que presenten diferencias significativas serán seleccionadas para su posterior análisis. En la Sección 4.2 se presentan los conceptos teóricos relacionados con el Análisis de Correspondencias, que permitirá determinar la asociación entre diagnósticos/fármacos/grupos de pacientes.

4.1. Test de proporciones y remuestreo *bootstrap* para seleccionar características

Una etapa muy importante en el análisis de datos es la selección de variables o características, para determinar cuáles son importantes para el estudio y cuáles no aportan información. Para llevar a cabo esta selección, se emplea un contraste de hipótesis basado en un test de proporciones y un remuestreo *bootstrap*. Ambos métodos se describen a continuación.

Un contraste o test de hipótesis es un procedimiento para contrastar la validez de una hipótesis nula (H_0) frente a una alternativa (H_1), a partir de un conjunto de n muestras de una población. En Estadística, una población X es el conjunto total de elementos sobre los que se desea estudiar una característica o parámetro θ , y una muestra \underline{X}_n es un subconjunto finito de dicha población [33]. Como generalmente no se puede conocer el parámetro θ de la población X , se hace uso de un estadístico de contraste T , observado en las muestras \underline{X} , para estimarlo o

contrastarlo. Para llevar a cabo un test de hipótesis, hay que seguir una serie de pasos:

- **Establecer la hipótesis nula (H_0) y la hipótesis alternativa (H_1).** H_0 , es la hipótesis o idea inicialmente aceptada, que se desea contrastar. En caso de que el test aporte suficiente información como para rechazar H_0 , entonces se aceptará H_1 , que es la hipótesis frente a la cual se contrasta H_0 .
- **Decidir el parámetro θ que se desea estudiar en la población y establecer un estadístico de contraste T .** La distribución del estadístico debe ser conocida y estar relacionada con el parámetro de estudio θ y de las muestras de la población ($\underline{X}_n = X_1, X_2, \dots, X_n$). El estadístico de contraste se denota como $T(\theta, \underline{X}_n)$. Este estadístico, es una medida de la discrepancia entre lo que dicta H_0 y los resultados observados en la muestra.
- **Establecer un nivel de significación α y seleccionar una muestra X de tamaño n .** El nivel de significación α es la probabilidad de rechazar H_0 cuando realmente es cierta. Al establecer el nivel de significación y el estadístico de contraste, se crea una región de rechazo y una región de aceptación. Si el estadístico de contraste observado en la muestra (t) está dentro de la región de rechazo, se rechazará H_0 , en caso contrario se aceptará. Hay que tener en cuenta que, en función de cómo se establezca H_0 y H_1 , la región de rechazo puede ser unilateral o bilateral (véase la Figura 4.1).

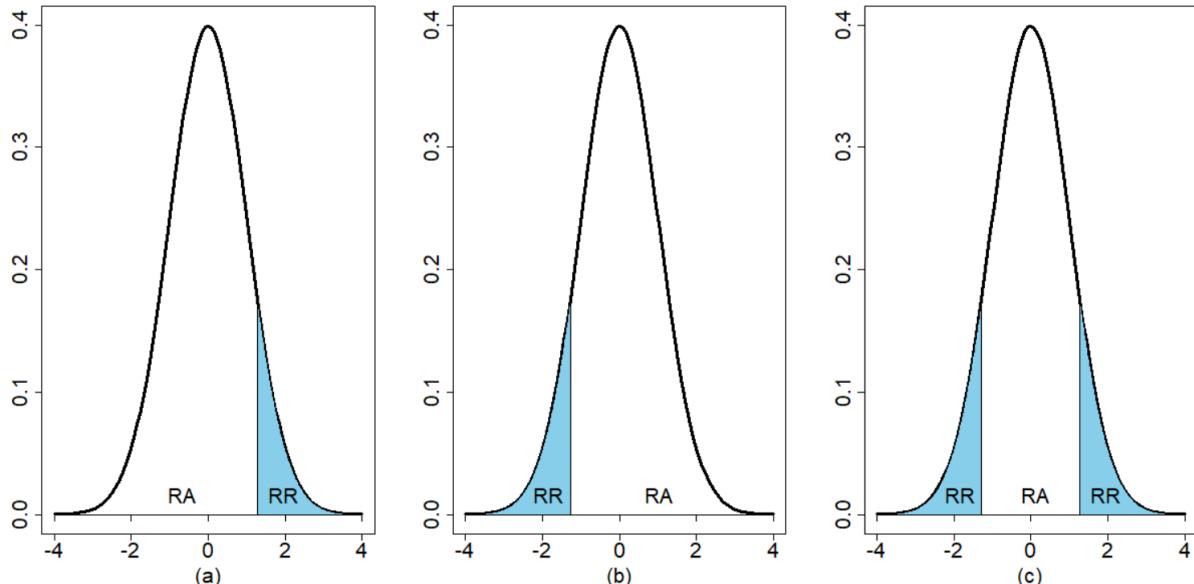


Figura 4.1: Estadístico de contraste $T \sim N(0, 1)$ y regiones de rechazo (RR) y aceptación (RA) según la definición de H_0 y H_1 para, (a) $H_0 : \theta \leq \theta_0, H_1 : \theta > \theta_0$, (b) $H_0 : \theta \geq \theta_0, H_1 : \theta < \theta_0$, (c) $H_0 : \theta = \theta_0, H_1 : \theta \neq \theta_0$.

A continuación, se detalla un ejemplo de cómo realizar un contraste de hipótesis. Sea X una población que sigue una distribución de probabilidad normal, de media μ y varianza $\sigma^2 = 5^2$, se quiere contrastar si $\mu \leq 50$ (H_0), o por el contrario, $\mu > 50$ (H_1). Se decide que el parámetro a estudiar es la media ($\theta \sim \mu$) y se define el estadístico de contraste como $Z(\mu, \underline{X}_n) = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$, donde \bar{X} es la media muestral, μ_0 es la media bajo H_0 , es decir, $\mu_0 = 50$, y n es el tamaño de la muestra. Nótese que, cuando se habla de distribuciones normales en lugar de usar T para denotar el estadístico de contraste, se suele usar Z . De igual modo, el estadístico de contraste observado en distribuciones normales se denominaría como z . Se establece que el nivel de significación deseado α es del 5 %. A partir de las definiciones anteriores, se selecciona una muestra de la población de tamaño $n=15$, se calcula la media observada de la muestra, $\bar{x} = 51,3$, y se calcula el estadístico de contraste observado en la muestra: $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{51,3 - 50}{5/\sqrt{15}} = 1,007$. Con estos valores, ya se puede definir la región de rechazo al nivel de significación del 5 %, $RR_{0,05} = \{z : z > z_{0,05}\} = \{z : z > 1,645\}$. En la Figura 4.2, se muestra gráficamente la región de rechazo (sombreado azul) para un nivel de significación $\alpha = 5\%$. Dado que el estadístico observado en la muestra ($z=1.007$) no está dentro de la RR ($z: z > 1.645$), no se rechaza H_0 , y se acepta que $\mu \leq 50$. Se concluye que la muestra seleccionada no aporta suficiente evidencia estadística como para poder rechazar H_0 .

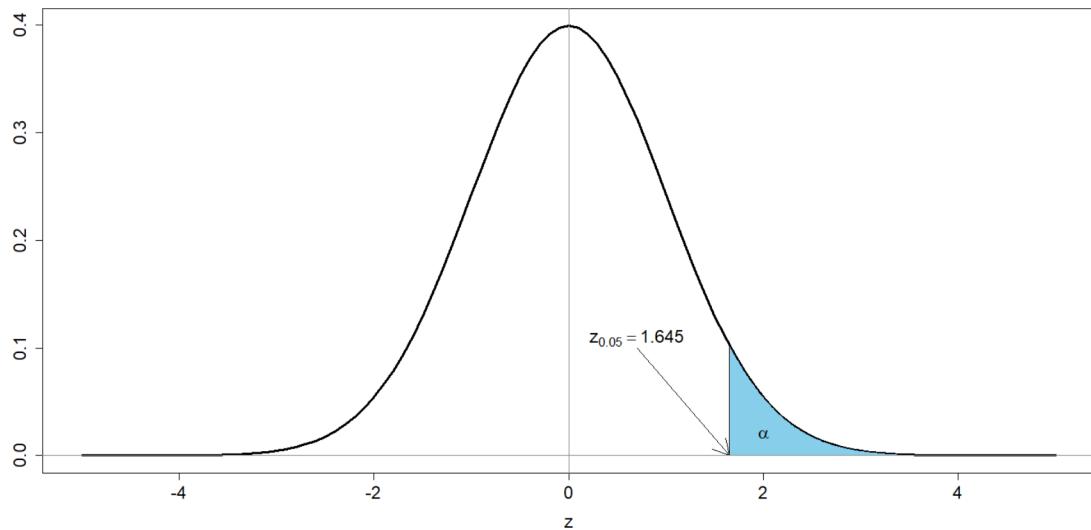


Figura 4.2: Región de rechazo para $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$ y $\alpha = 0.05$.

El objetivo de un contraste de hipótesis es diferenciar cuándo una propiedad estudiada en una muestra de una población es fruto del azar (en este caso se rechazaría H_0) y cuándo hay evidencia estadística de que esa propiedad es significativa, es decir, no es sólo obra del azar (en

cuyo caso se aceptaría H_0).

Un contraste de hipótesis a un nivel de significación α , además de con el procedimiento anteriormente descrito, también se puede realizar a través del cálculo de intervalos de confianza. A continuación, se detalla la estimación por intervalos de confianza y su dualidad con los contrastes de hipótesis [34].

Sea X una población en estudio, θ el parámetro que se desea estimar, y dado un nivel de significación α , un estimador de θ por intervalos de confianza marcados por los límites T_1 y T_2 , es un intervalo (T_1, T_2) que cumple que la probabilidad P de que θ esté en dicho intervalo es de $1 - \alpha$ (i.e. $P[\theta \in (T_1, T_2)] = 1 - \alpha$). Para obtener el estimador de θ , primero se define un pivote o cantidad pivotal $C(\theta, \underline{X}_n)$, que es una función dependiente del parámetro θ y de la muestra \underline{X} de tamaño n y cuya distribución sea conocida. Se define el pivote como $C(\mu, \underline{X}_n) = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$, donde el parámetro θ a estimar es μ , y \bar{X} es la media muestral. Para encontrar los extremos del intervalo de confianza al $(1-\alpha)\%$, se debe resolver la doble desigualdad para μ : $z_{1-\alpha/2} < C(\mu, \underline{X}_n) < z_{\alpha/2}$, o lo que es lo mismo, $-1,96 < C(\mu, \underline{X}_n) < 1,96$. Finalmente se obtiene que μ está comprendido entre $T_1 = \bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ y $T_2 = \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$. Esto quiere decir que el 95 % de la cantidad pivotal está comprendida entre $z_{1-\alpha/2}$ y $z_{\alpha/2}$, y que con un 95 % de confianza μ estará entre T_1 y T_2 . En la Figura 4.3 se muestra gráficamente la distribución de la cantidad pivotal.

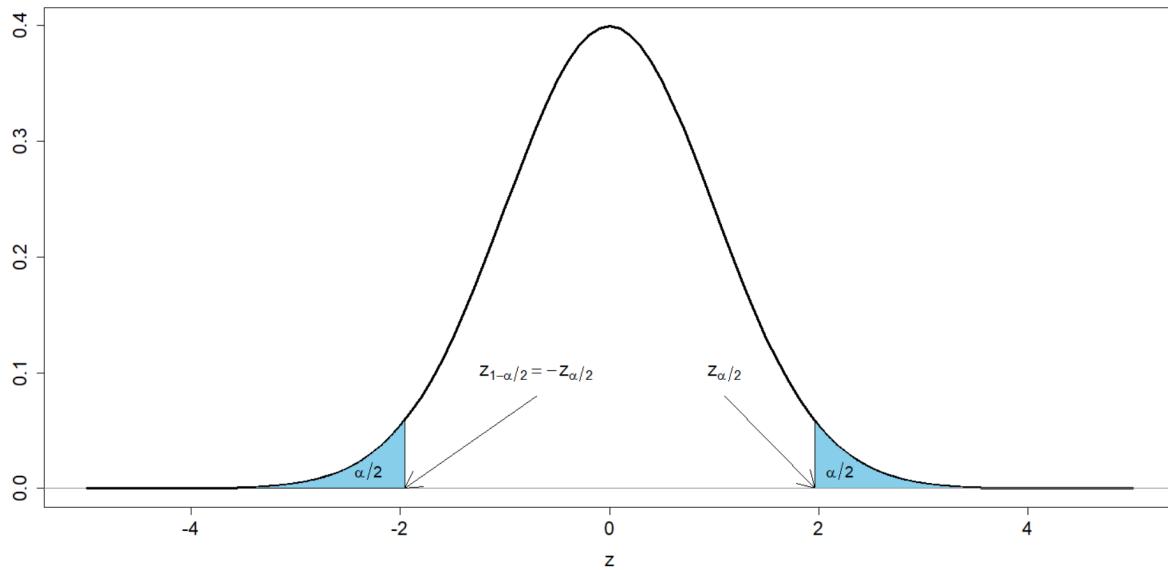


Figura 4.3: Cantidad pivotal e intervalo de confianza.

Se concluye que existe una dualidad entre el contraste de hipótesis de un parámetro y su

estimación por intervalos de confianza: en el contraste se define un estadístico de contraste, que en intervalos de confianza se denomina cantidad pivotal. Lo que se definía como región de aceptación en los contrastes, es lo que en el segundo caso se denomina intervalo de confianza. El intervalo de confianza de un parámetro construido al 95 %, quiere decir que el conjunto de valores contenidos dentro del intervalo son plausibles (con una confianza del 95 %) y no lo son los valores que quedan fuera. Por tanto, si el valor del parámetro especificado por H_0 está contenido en el intervalo, la hipótesis será aceptada; en caso contrario, se rechazará.

En este TFG, se quiere estudiar si la diferencia de proporción en la tasa de presencia de un determinado código (diagnóstico o ATC) es estadísticamente significativa en dos poblaciones. Los códigos o características con diferencias significativas serán seleccionadas para el posterior análisis de correspondencias. El contraste de hipótesis que se realiza en este TFG se basa en intervalos de confianza. Para analizar los diagnósticos y fármacos que mejor caracterizan a los diferentes grupos de pacientes crónicos, se considera como estadístico de contraste o pivote la diferencia de proporciones (que se denominará Δp_i) en la presencia de un determinado código i (diagnóstico o ATC) entre dos grupos de población. Para ello, se contrastarán las diferencias Δp_i entre la proporción del código i en el grupo de pacientes sanos ($p_{X1,i}$) frente a la proporción de ese mismo código en cada uno de los grupos de pacientes crónicos ($p_{X2,i}$). Por ejemplo, si se realiza el test de proporciones comparando el CRG-base 1000 y el CRG-base 5192, se verán las diferencias significativas en la tasa de presencia de los diagnósticos y ATCs de esos dos grupos. Las diferencias significativas que se obtengan de este test caracterizan a los grupos de pacientes crónicos (en este caso, de los hipertensos).

Para llevar a cabo el contraste, inicialmente, se tiene solo una muestra de cada grupo de población estudiado (CRGs-base: 1000, 3755, 5133, 5192, 5424, 6313 y 7140). Por tanto, al realizar el test para cada pareja de CRGs-base en busca de aumentos o disminuciones en la presencia de un determinado código, sólo se dispone de un valor del estadístico de contraste o pivote ($\Delta p_i = p_{X1,i} - p_{X2,i}$). Para elaborar un intervalo de confianza, se necesita conocer la distribución de Δp_i , y no sólo un valor aislado. Para obtener la distribución de Δp_i de cada código, se hará uso del método de **remuestreo bootstrap**. El remuestreo *bootstrap* consiste en considerar la muestra inicial (por ejemplo, la muestra del CRG-base 5192) como el total de la población (X_{5192}), y a partir de ella, obtener N submuestras o subconjuntos, $X_{1,5192}, X_{2,5192} \dots X_{N,5192}$. Ahora ya no se tiene un valor único de Δp_i , sino N valores de Δp_i , que siguen una distribución, y con los que se puede construir un intervalo de confianza de Δp_i para cada código i .

A continuación, se detalla un ejemplo del contraste que se realiza en este TFG. Sean las

poblaciones:

$X1 \sim$ pacientes clasificados en 2012 en el CRG-base 1000

$X2 \sim$ pacientes clasificados en 2012 en el CRG-base 5192

Se quiere seleccionar los códigos cuya tasa de presencia varíe significativamente en los conjuntos $X1$ y $X2$. Para ello, se establece el nivel de significación al 7% y se quiere, además, que la diferencia de proporciones sea al menos del 5%. Se definen las hipótesis:

$$H_0 : p_{X1,i} - p_{X2,i} \leq 0,05, \text{ donde } i \text{ es cada uno de los códigos considerados.}$$

$$H_1 : p_{X1,i} - p_{X2,i} > 0,05, \text{ donde } i \text{ es cada uno de los códigos considerados.}$$

El estadístico de contraste o pivote usado es la diferencia de proporción en la tasa de presencia de cada código entre $X1$ y $X2$, es decir, $\Delta p_i = p_{X1,i} - p_{X2,i}$. Por ejemplo, si el código diagnóstico $i = 1$ (es el código diagnóstico “001”) en la población $X1$ está presente en 500 pacientes de los 1000 que hay, y en $X2$ en 50 pacientes de los 200 que hay, entonces:

$$\Delta p_1 = p_{X1,1} - p_{X2,1} = \frac{500}{1000} - \frac{50}{200} = 0,25$$

Para conseguir una distribución de Δp_i y construir intervalos de confianza al 7%, se seleccionan mediante remuestreo *bootstrap*, N remuestras de $X1$ y N remuestras de $X2$. Para cada una de las remuestras se calcula Δp_i de cada código i , obteniendo al final N valores de Δp_i . Finalmente, se construyen los intervalos de confianza y se comprueba si el intervalo está por encima de 0.05, en cuyo caso se rechaza H_0 y se acepta que la diferencia de proporciones es significativa (mayor que 0.05~5%) y en caso contrario se acepta H_0 y ese código no sería seleccionado para el análisis de correspondencia dado que la diferencia de proporciones es muy pequeña (menor del 5%).

4.2. Análisis de Correspondencias Múltiple

En el contexto de este TFG, el **Análisis de Correspondencias** (AC) se puede entender como una manera de representar gráficamente asociaciones tanto entre variables y categorías, como entre individuos. El AC hace uso de la **tabla de contingencia**, que permite recoger información sobre la frecuencia de co-ocurrencia de variables, generalmente categóricas.

Aprovechando la temática de este TFG, se propone analizar la relación que existe entre 3 grupos de pacientes y 3 diagnósticos. Se consideran datos de 110 pacientes, 35 pacientes diabéticos, de los cuales 20 presentan el diagnóstico 1 (D1), 10 el diagnóstico 2 (D2) y 5 el diagnóstico 3 (D3); 40 pacientes hipertensos, 10 de ellos presentan el D1, 20 el D2 y 10 el D3; y por último, 35 pacientes con depresión, donde 2 de ellos tienen el D1, 18 el D2 y 15 el D3. El Cuadro 4.1 recoge los datos descritos en forma de tabla de contingencia con dos variables. En este ejemplo en particular, entre grupos de pacientes, y diagnósticos que pueden presentar. En dicha tabla, para cada pareja grupo/diagnóstico se muestra el número de pacientes en ese grupo y con ese diagnóstico, o lo que es lo mismo la frecuencia absoluta de esa pareja.

	D1	D2	D3	Total
Grupo-DM	20	10	5	35
Grupo-HTA	10	20	10	40
Grupo-Depresión	2	18	15	35
Total	32	48	30	110

Cuadro 4.1: Tabla de contingencia del ejemplo propuesto: estados de salud versus diagnósticos.

Un caso particular del AC, cuando se analizan dos variables, es el diagrama de dispersión. Los **diagramas de dispersión** permiten representar de forma visual dos variables en relación a dos ejes. Además, el AC permite analizar la asociación entre ambas variables de un modo visual y sencillo. Cuanto más grandes sean las tablas y más variables contengan, más difícil se hace su representación e interpretación gráfica. En estos casos, dejará de ser útil el diagrama de dispersión, y será más conveniente usar el AC [35].

Para entender el AC, es conveniente presentar los conceptos sobre los que se fundamenta, para lo cual haremos uso del ejemplo anterior. Así, si de la tabla de contingencia se calculan las frecuencias relativas, se obtienen los **profiles-fila** o los **profiles-columna**, según se normalice respecto al total de la fila (total de pacientes en cada grupo) o de la columna (total de pacientes con cada diagnóstico). El AC se puede centrar en el análisis de las filas (que llamaremos **análisis-fila**) tomando como referencia las columnas o en las columnas (que llamaremos **análisis-columna**) tomando como referencia las filas. Si el análisis se centra en las filas, entonces será preciso calcular los profiles-fila y en caso contrario, los profiles-columna. En el ejemplo propuesto se decide llevar a cabo el análisis-fila, por tanto, se calculan los profiles-fila que serán los profiles de cada grupo de pacientes. Adicionalmente, hay que calcular el perfil-fila medio, que se obtiene normalizando el total de cada columna, respecto al total de pacientes. Estos profiles se muestran en el Cuadro 4.2.

	D1	D2	D3	Masas-fila
Perfil del grupo-DM	20/35=0,57	10/35=0,29	5/35=0,14	35/110=0,32
Perfil del grupo-HTA	10/40=0,25	20/40=0,5	10/40=0,25	40/110=0,36
Perfil del grupo-depresión	2/35=0,06	18/35=0,51	15/35=0,43	35/110=0,32
Perfil-fila medio	32/110=0,29	48/110=0,44	30/110=0,27	

Cuadro 4.2: Perfiles-fila de la tabla de contingencia (Cuadro 4.1), en función de los diagnósticos.

Un perfil formado por N elementos se puede representar en un espacio de N-1 dimensiones, ya que la suma de todos los elementos de un perfil es 1. En este ejemplo, los perfiles-fila tienen 3 elementos, por tanto, para representarlos se necesita un espacio de dos dimensiones. El AC asigna un punto en el espacio de 2D a cada elemento del perfil, de tal forma que la relación entre puntos (en términos de proximidad) refleje asociaciones entre las variables [36]. Para reflejar en el subespacio creado las asociaciones reales entre variables, se usa la llamada **distancia Chi-Cuadrado**. Esta distancia es un métrica para medir la proximidad, que además de reflejar la asociación, pondera cada uno de los perfiles por un peso o masa. Este peso da una idea de la importancia del perfil respecto al total de observaciones, pues no todos los perfiles tienen la misma relevancia, dado que no en todos los grupos hay el mismo número de pacientes. En el caso de llevar a cabo el análisis-fila, el peso de cada fila se calcula como la suma de los pacientes en cada fila, frente al total de pacientes.

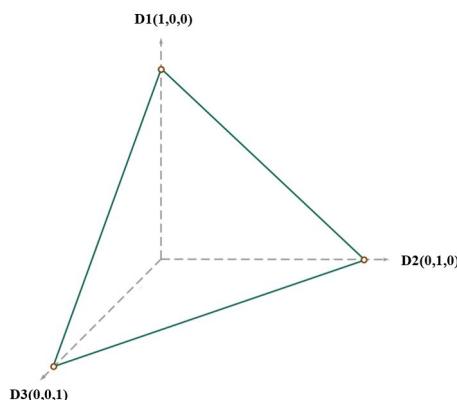


Figura 4.4: Espacio de perfiles asociado al ejemplo considerado.

Otro concepto importante es el de **vértice**. Los vértices definen los extremos del espacio en el que podrían representarse los perfiles (fila o columna, según el análisis). Si el análisis se centra en los perfiles-fila, los vértices (o vértices-fila) serían los hipotéticos perfiles-fila extremos. En el ejemplo considerado en esta sección, estos vértices corresponderían al perfil de un

grupo de pacientes donde todos los integrantes tuvieran únicamente el diagnóstico D1 (1, 0, 0); el D2 (0, 1, 0); o el D3 (0, 0, 1). Volviendo al ejemplo, se concluye que se podrían representar los tres perfiles-fila en un espacio de dos dimensiones delimitado por tres vértices (véase Figura 4.4). Dicha representación contendría la misma información que la tabla de datos dada y permitiría encontrar asociaciones entre las variables grupo de pacientes y diagnósticos. Sin embargo, normalmente los perfiles están formados por más elementos y consecuentemente para su representación se necesitarían espacios de más dimensiones.

Esta reducción se puede llevar a cabo a costa de cierta **pérdida de información**. El objetivo del AC es identificar espacios de pocas dimensiones que contengan todos los perfiles, pero de forma aproximada y con la mínima pérdida de información. Para medir esta pérdida de información se usa el concepto de **inercia**. La inercia es una medida de la variabilidad de los perfiles respecto al perfil medio. Un baja inercia implica poca variabilidad, por tanto los perfiles estarán cerca del perfil medio y se dice que existe poca relación entre filas y columnas. Para expresar la calidad de la aproximación entre el espacio real en el que podrían representarse los perfiles fielmente y el subespacio creado (de menor dimensionalidad), se habla de porcentaje de inercia. Si el porcentaje de inercia representado es del 97 %, esto quiere decir que, perdiendo un 3 % de la variabilidad, se pueden representar los perfiles en un espacio dimensional de fácil interpretación. La inercia perdida es una medida de la calidad de la aproximación, es decir, la diferencia entre las posiciones que un elemento tendría en su espacio multidimensional ideal y el que realmente tiene en el espacio obtenido aproximado.

Otro aspecto clave del AC es la **simetría** entre el análisis de los perfiles-fila (análisis-fila) y el análisis de los perfiles-columna (análisis-columna). Analizando los perfiles-fila se toma como referencia las columnas de la tabla de contingencia, y viceversa. Estos análisis tienen una serie de peculiaridades. Aunque los espacios donde se encuentran los perfiles-fila o perfiles-columna sean distintos, su dimensionalidad es la misma. Por otro lado, la calidad que se podría obtener de uno u otro análisis también coincide, en términos de inercia. Por último, las coordenadas de los perfiles en el espacio (también llamadas **coordenadas principales**) en un análisis (el análisis-fila, por ejemplo) corresponden a las coordenadas que tendrían los vértices (también llamadas **coordenadas estándares**) en el otro análisis (por ejemplo, el análisis-columna) multiplicadas por un factor de escala [35].

En los párrafos anteriores se ha considerado el análisis de dos variables, estado de salud y diagnósticos. En el caso de analizar más de dos variables, por ejemplo, estado de salud, diagnósticos y fármacos, se habla de **Análisis de Correspondencias Múltiple** (ACM). Este análisis puede realizarse a través de la **matriz binaria** (con los datos codificados de forma binaria), o a

través de la **matriz de Burt** (con todos los cruces entre variables) [37].

Para construir cualquiera de las matrices necesaria para el ACM, primero se seleccionan las variables a analizar, en este caso: diagnósticos, fármacos y CRGs-base. En segundo lugar, por cada variable se consideran las categorías en las que se divide. Así, para la variable *diagnósticos* se tendrán tantas categorías como diagnósticos seleccionados previamente; para la variable *fármacos*, tantas categorías como códigos ATC previamente seleccionados; y la variable *CRG-base* tendrá 6 categorías, que corresponden a los 6 estados de salud objeto de estudio. Para construir la matriz binaria (Z) se crea una tabla donde las categorías de cada variable se sitúan en las columnas, y los datos de cada paciente en las filas. Para la variable CRG-base, sus 6 categorías son mutuamente excluyentes, es decir, si un paciente ha sido clasificado en el CRG-base 5192, tendrá un 1 en esa columna y un 0 en el resto de columnas asociadas a la variable estado de salud. Nótese que la codificación es ligeramente diferente para el caso de diagnósticos o fármacos, ya que un mismo paciente podría tener más de un diagnóstico y más de un código ATC. Una vez construida esta matriz binaria Z , la matriz de Burt (B) se obtiene como resultado de multiplicar la matriz binaria (Z) por su traspuesta, y muestra el número de pacientes por cada par de categorías ($B = ZZ^T$). A continuación, los Cuadros 4.3 y 4.4 muestran un ejemplo de cómo se generan Z y B .

		Variable 1: Diagnósticos (CIE-9)			Variable 2: Fármacos (ATCs)			Variable 3: CRGs-base		
		Categoría 1: Diagnóstico 001	Categoría 2: Diagnóstico 002	(...)	Categoría 1: Código A01AA	Categoría 2: Código A01AB	(...)	Categoría 1: CRG-base 3755	Categoría 2: CRG-base 5133	(...)
Pacientes	Id 1	0	1	(...)	0	1	(...)	1	0	(...)
	Id 2	0	0	(...)	1	0	(...)	0	1	(...)
	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)

Cuadro 4.3: Matriz binaria Z .

		Variable 1: Diagnósticos (CIE-9)			Variable 2: Fármacos (ATCs)			Variable 3: CRGs-base		
		Categoría 1: Diagnóstico 001	Categoría 2: Diagnóstico 002	(...)	Categoría 1: Código A01AA	Categoría 2: Código A01AB	(...)	Categoría 1: CRG-base 3755	Categoría 2: CRG-base 5133	(...)
Variable 1: Diagnósticos (CIE-9)	Categoría 1: Diagnóstico 001	69	23	(...)	36	47	(...)	41	63	(...)
	Categoría 2: Diagnóstico 002	23	15	(...)	53	29	(...)	5	34	(...)
	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)
Variable 2: Fármacos (ATCs)	Categoría 1: Código A01AA	36	53	(...)	28	13	(...)	12	85	(...)
	Categoría 2: Código A01AB	47	29	(...)	13	21	(...)	7	71	(...)
	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)
Variable 3: CRGs-base	Categoría 1: CRG-base 3755	41	63	(...)	12	85	(...)	700	0	(...)
	Categoría 2: CRG-base 5133	5	34	(...)	7	71	(...)	0	1200	(...)
	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)

Cuadro 4.4: Matriz de Burt B .

Una vez construida la matriz de Burt, se busca un subespacio S de baja dimensionalidad que minimice el siguiente criterio:

$$\text{Proximidad a } \mathbf{S} = \sum_i m_i [d_i(\mathbf{S})]^2 \quad (4.1)$$

Donde i es cada uno de los perfiles, m_i es el peso del perfil i -ésimo y $d_i(\mathbf{S})$ es la distancia entre el perfil i -ésimo y el subespacio \mathbf{S} . Para encontrar este subespacio, se usa la **Descomposición en Valores Singulares** (DVS), que permite factorizar cualquier matriz rectangular en tres matrices de la forma: $\mathbf{S} = \mathbf{U}\mathbf{D}_\alpha\mathbf{V}^T$, donde \mathbf{U} y \mathbf{V} son los vectores singulares y los elementos de la diagonal de \mathbf{D}_α son los valores singulares.

Cada dimensión del subespacio \mathbf{S} contendrá un porcentaje de inercia. Así, cuantas más dimensiones se consideren en \mathbf{S} , más fiel será la representación en \mathbf{S} a la del espacio original. En nuestro caso, se tomarán las tres dimensiones con mayor inercia, obteniendo un espacio tridimensional donde analizar gráficamente la asociación entre las categorías de cada variable. Como el análisis de este TFG se centra en las columnas, se representarán por un lado los perfiles-columna en el subespacio hallado, lo que permitirá visualizar asociaciones entre las categorías. Por otro lado, se proyectarán los individuos (en este TFG, pacientes) en el subespacio creado, haciendo uso de la expresión $Z_{\text{proyectada}} = \mathbf{V}^T Z^T$. Esto permitirá visualizar asociaciones entre pacientes o grupos de pacientes. Los detalles de la implementación del ACM, se adjuntan en el Apéndice H.

Capítulo 5

Experimentos y resultados

En este capítulo se muestran los resultados de aplicar los métodos presentados en el Capítulo 4 sobre los datos detallados en el Capítulo 3. En la Sección 5.1, se mostrarán los resultados del test de hipótesis y del remuestreo *bootstrap* para realizar selección de características. En la Sección 5.2, se presentarán los resultados del ACM, tanto para visualizar asociaciones entre las características seleccionadas como de los grupos de pacientes. Ambos análisis se realizarán, por un lado, para caracterizar los diferentes estados de salud considerados, y por otro, la evolución temporal de cada uno de ellos.

5.1. Selección de características mediante test de proporciones y remuestreo *bootstrap*

Como se indicó en el capítulo anterior, la selección de características realizada en este TFG se basa en un test de proporciones y un remuestreo *bootstrap*. El test de proporciones es un contraste de hipótesis sobre la diferencia (Δp_i) en la tasa de presencia p de un código i (diagnóstico o fármaco) entre dos grupos de pacientes. Por otra parte, el remuestreo *bootstrap* se usa para estimar intervalos de confianza de cada Δp_i , permitiéndonos seleccionar aquellos códigos cuya diferencia de proporciones sea significativa con un nivel de confianza del 95 % (es decir, se establece el nivel de significación al 5 %). Además, se decide que el incremento o disminución de la presencia de cada código entre los dos grupos de pacientes sea, al menos, de un 5 %.

También se ha comentado en los capítulos anteriores que el análisis se puede hacer desde dos puntos de vista. Si comparamos los CRGs-base de pacientes crónicos (CRGs-base: 3755,

5133, 5192, 5424, 6313 y 7140) con el CRG-base de pacientes sanos (CRG-base 1000) en un mismo año (en este caso 2012) encontraremos las diferencias significativas en la tasa de presencia de diagnósticos y fármacos que mejor caractericen esos estados de salud crónicos. Otra forma de interpretar esta comparación entre pacientes sanos y crónicos, es pensar que se están eliminando a las características de los pacientes hipertensos aquéllas que también están en los pacientes sanos. A modo de ejemplo, por un lado, los pacientes sanos generalmente consumen *Ibuprofeno*, por otro lado, los pacientes hipertensos también consumen *Ibuprofeno*, y además *Simvastatina* (un fármaco muy usado en el tratamiento de la HTA). En este ejemplo, se podría concluir que la diferencia en la tasa de proporción de la *Simvastatina* es superior a la de *Ibuprofeno*, pues un medicamento está sólo presente en pacientes hipertensos y el otro en ambos grupos. De esta forma se seleccionaría la *Simvastatina* como característica de los pacientes hipertensos, y no el *Ibuprofeno*.

El otro enfoque que se puede considerar en este análisis es comparar las características de los pacientes asociados al mismo estado de salud en distintos años. Por ejemplo, buscar diferencias entre el grupo de pacientes hipertensos en 2012 (CRG-base 5192 en 2012) y 2015 (CRG-base 5192 en 2015). De esta forma, no se caracterizaría el estado de salud, si no la evolución temporal del grupo de pacientes hipertensos en términos de presencia de diagnósticos y fármacos. Se podría ver qué medicamentos o diagnósticos han aumentado o disminuido su presencia en ese intervalo temporal.

5.1.1. Caracterización del estado de salud

Antes de comentar los resultados, se van a explicar las figuras que se muestran a continuación. La Figura 5.1 muestra, por un lado, los intervalos de confianza de Δp para los 1517 diagnósticos (Figura 5.1 (a)) y los 746 fármacos (Figura 5.1 (b)) analizados, al comparar el CRG-base 1000 y el CRG-base 3755. Por otro lado, como son muchos códigos y no se aprecian los índices de aquéllos que cumplen la condición establecida (diferencias superiores al 5 %), en los paneles inferiores se muestran únicamente los códigos seleccionados junto con el límite inferior o superior del intervalo de confianza: Figura 5.1 (c) para diagnósticos y Figura 5.1 (d) para fármacos. Se mostrará el límite inferior para los códigos que disminuyen su presencia en el CRG-base 3755 frente al CRG-base 1000; en caso contrario, se mostrará el límite superior. Para el resto de estados de salud (Figuras 5.2- 5.6) se mostrara sólo la representación equivalente del panel inferior.

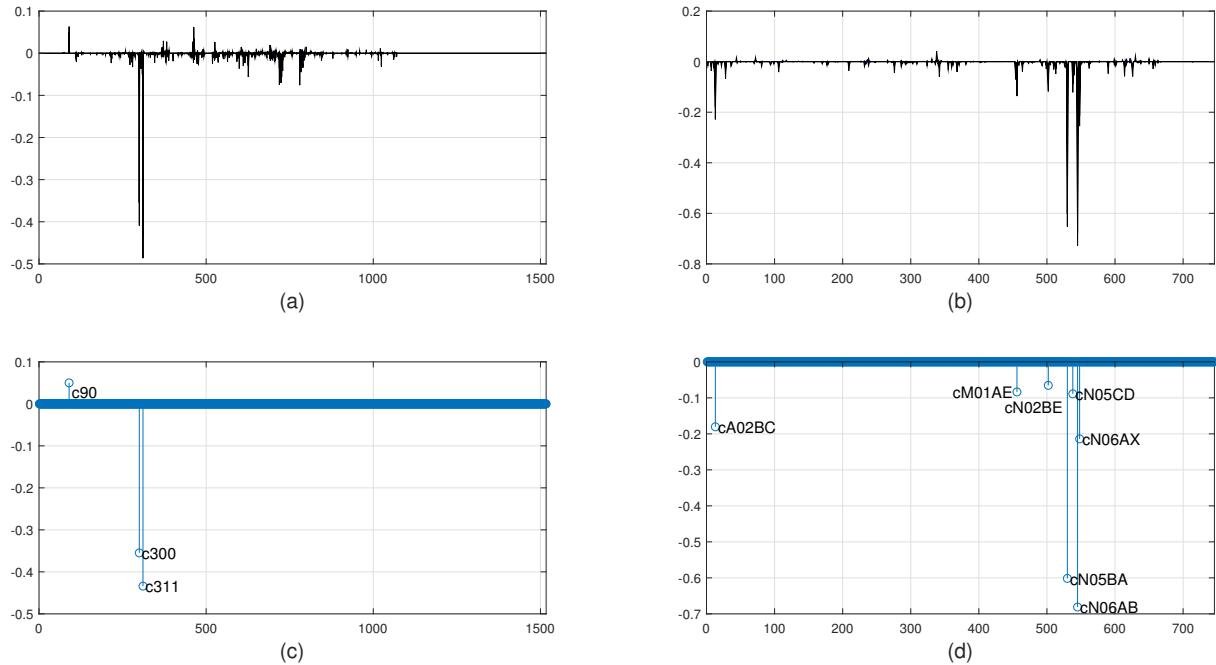


Figura 5.1: CRG-base 1000 versus 3755 en 2012. Intervalos de confianza de Δp al 95 % para diagnósticos (a) y fármacos (b). Detalle para los diagnósticos (c) y fármacos (d) seleccionados.

Para no extender la memoria, en el Apéndice F se adjuntan las listas de códigos seleccionados junto con sus descripciones al comparar cada CRG-base de grupos de pacientes crónicos con el CRG-base de pacientes sanos. A continuación se describen brevemente los resultados.

En cuanto al CRG-base 3755 (Figura 5.1), los códigos diagnósticos 311 (trastorno depresivo, no clasificado bajo otros conceptos) y 300 (trastornos de ansiedad, dissociativos y somatomorfos) son los que experimentan mayor diferencia (por encima del 35 %). Por otro lado, los códigos ATC N06AB (inhibidores selectivos recaptación serotonina-antidepresivos) y N05BA (benzodiazepinas-ansiolíticos) experimentan un incremento por encima del 60 %; y los códigos N06AX (otros antidepresivos) y A02BC (inhibidores de la bomba de protones-antiúlcera péptica) por encima del 20 %.

En el CRG-base 5133 (véase la Figura 5.2), destacan los códigos diagnósticos 496 (obstrucción crónica de vías respiratorias, no clasificada bajo otros conceptos), 493 (asma) y 518 (otras enfermedades pulmonares), con incrementos superiores al 20 %. Los códigos ATC R03BB (anticolinérgicos-antiasmáticos) y R03AC (agonistas beta-2-adrenergicos selectivos) presentan incrementos superiores al 50 %, mientras que R03AK (adrenérgicos asociados a otros antiasmáticos) y N02BE (anilinas-analgésicos) aumentan su presencia en más del 30 %.

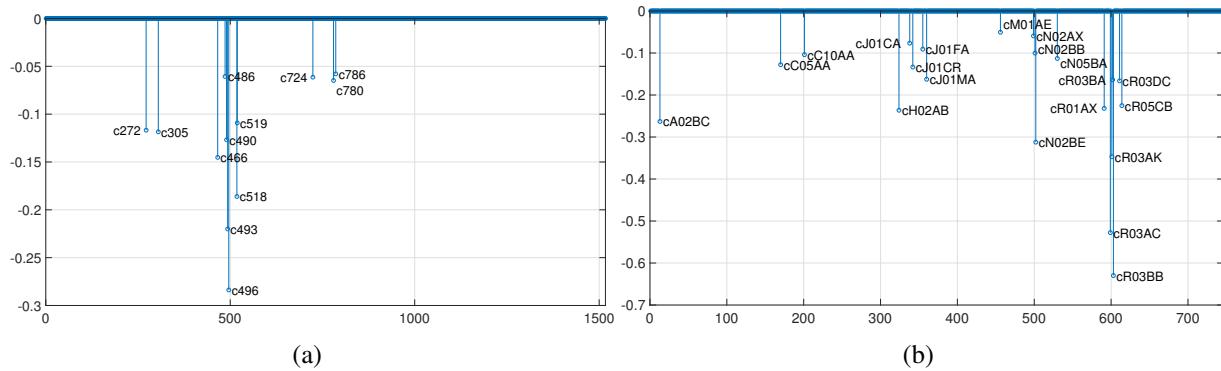


Figura 5.2: Códigos seleccionados al comparar el CRG-base 1000 frente al CRG-base 5133 en 2012: diagnósticos (a) y fármacos (b).

En cuanto al CRG-base 5192 (véase la Figura 5.3), el diagnóstico 401 (hipertensión esencial) aumenta su presencia respecto a los pacientes sanos en un 80 %, como cabría esperar. También destacan con incrementos superiores al 10 % los diagnósticos 272 (trastornos del metabolismo de los lípidos), 724 (otros trastornos y trastornos no especificados de la espalda) y 719 (otros trastornos y trastornos no especificados de articulación). En cuanto a fármacos, los códigos C09AA (inhibidores enzima convertasa angiotensina solos), C10AA (inhibidores de la hmg-coa reductasa-modificadores lípidos), A02BC (inhibidores de la bomba de protones-antiúlcera péptica), N05BA (benzodiazepinas-ansiolíticos) y N02BE (anilinas-analgésicos) incrementan su presencia en más de un 10 %.

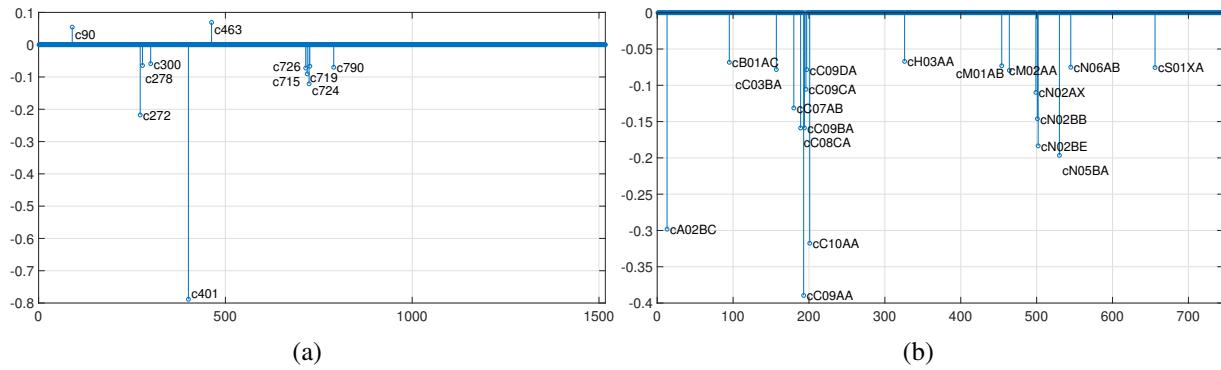


Figura 5.3: Códigos seleccionados al comparar el CRG-base 1000 frente al CRG-base 5192 en 2012: diagnósticos (a) y fármacos (b).

En cuanto al CRG-base 5424 (véase la Figura 5.4), el diagnóstico predominante es el 250 (diabetes mellitus) que aumenta un 90 %. Le siguen los diagnósticos 272 (trastornos del me-

tabolismo de los lípidos) y 724 (otros trastornos y trastornos no especificados de la espalda), que aumentan entre un 10% y un 20%. Por otro lado, los códigos ATC A10BA (biguanidas-antidiabéticos orales), C10AA (inhibidores de la hmg-coa reductasa-modificadores lípidos), A02BC (inhibidores de la bomba de protones-antiúlcera péptica) y A10AE (insulinas y análogos acción lenta para inyección) se incrementan en más de un 20%.

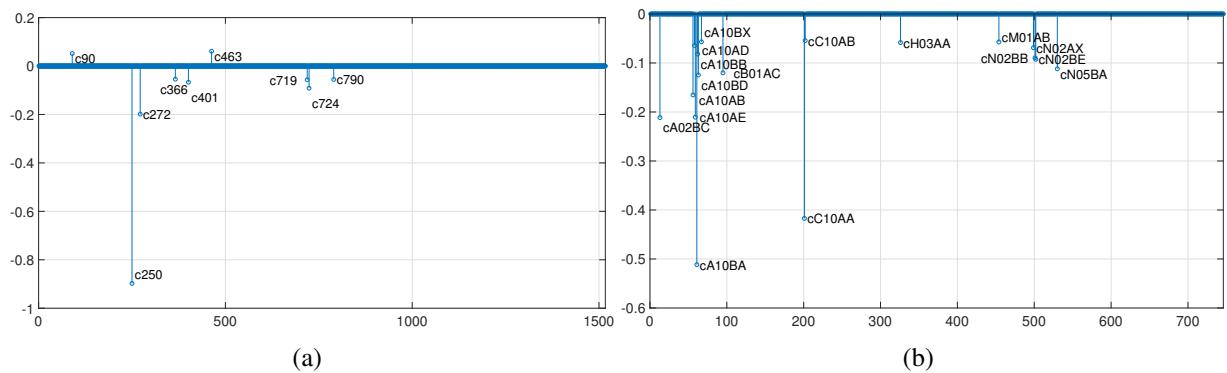


Figura 5.4: Códigos seleccionados al comparar el CRG-base 1000 frente al CRG-base 5424 en 2012: diagnósticos (a) y fármacos (b).

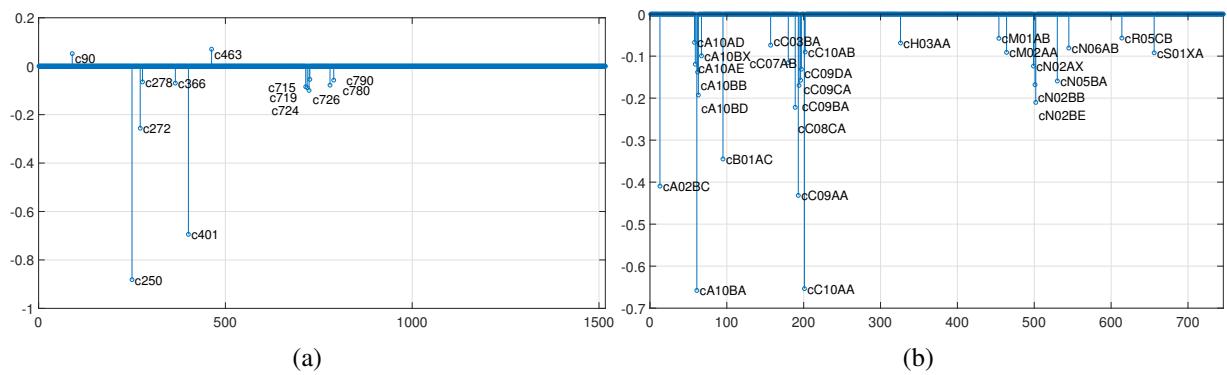


Figura 5.5: Códigos seleccionados al comparar el CRG-base 1000 frente al CRG-base 6313 en 2012: diagnósticos (a) y fármacos (b).

Por último, los resultados para los CRGs-base 6313 (Figura 5.5) y 7140 (Figura 5.6), y se detallan conjuntamente. Destacan los diagnósticos 250 (diabetes mellitus) y 401 (hipertensión esencial) con incrementos del 90% y 80%, respectivamente; y 272 (trastornos del metabolismo de los lípidos) con un incremento por encima del 20%. Por otro lado, los fármacos C10AA (inhibidores de la hmg-coa reductasa-modificadores lípidos), A02BC (inhibidores de la bomba de protones-antiúlcera péptica), A10BA (biguanidas-antidiabéticos orales), N02BE (anilinas-

analgésicos), C09AA (inhibidores enzima convertasa angiotensina solos) y B01AC (inhibidores de la agregación plaquetaria) presentan incrementos superiores al 35 % en ambos CRGs-base.

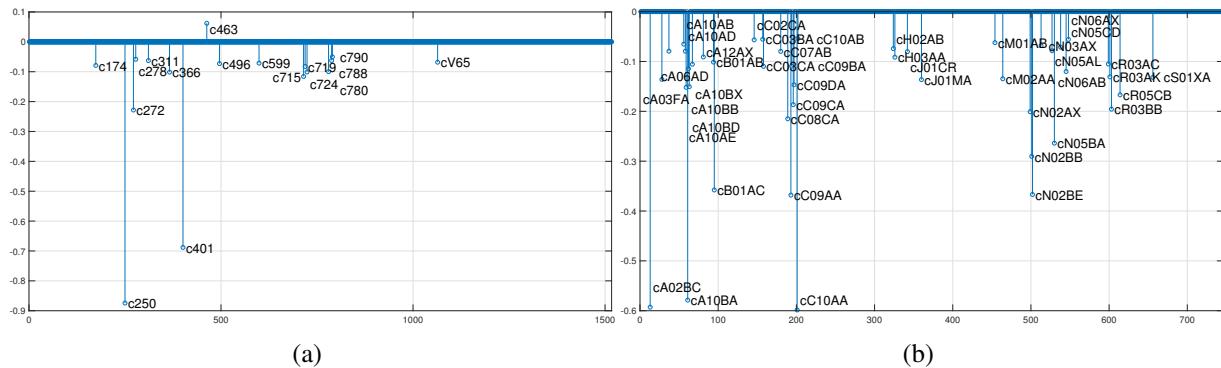


Figura 5.6: Códigos seleccionados al comparar el CRG-base 1000 frente al CRG-base 7140 en 2012: diagnósticos (a) y fármacos (b).

De las figuras anteriores se concluyen algunas evidencias, como que el diagnóstico que más aumenta su presencia en el grupo de pacientes con depresión es el 300 (trastorno depresivo), en pacientes con EPOC el 496 (obstrucción crónicas de las vías respiratorias), en pacientes hipertensos es el 401 (hipertensión esencial) y en diabéticos el 250 (diabetes mellitus).

Sin embargo, también se pudo concluir que tanto los pacientes hipertensos como diabéticos, tienen asociados problemas de espalda y articulaciones. Así mismo, se observa que los pacientes hipertensos tienen asociados diagnósticos y fármacos relacionados con problemas de peso y ansiedad. Por otro lado, los pacientes diabéticos también presentan un incremento del 8 % en el diagnóstico de hipertensión esencial (401) respecto a los pacientes sanos, y también en el 366 (cataratas). Por último, en el CRG-base 7140, además de diagnósticos y fármacos relacionados con HTA y DM, se incrementa la presencia de códigos relacionados con problemas respiratorios (496), trastornos depresivos (311) y neoplasias malignas (174).

5.1.2. Evolución temporal

Tras haber caracterizado a los grupos de pacientes crónicos, se realiza el test de proporciones y el remuestreo *bootstrap* comparando cada estado de salud en dos años diferentes, 2012 y 2015. Ello pondrá de relieve diferencias significativas en un intervalo temporal. En el Apéndice G se muestra una tabla con los códigos cuya diferencia de proporción entre 2012 y 2015, para cada CRG-base asociado a los grupos de pacientes crónicos, sea superior al 5 %.

Se observa, que el código R05CB (mucolíticos) desciende su presencia en todos los grupos, al igual que el código M02AA (antiinflamatorios no esteroideos para uso tópico). También descienden de manera generalizada en prácticamente todos los grupos los fármacos: S01XA (otros oftalmológicos) y M01AB (derivados del ácido acético y sustancias relacionadas). Tras consultar estos resultados con el personal del HUF, se ha concluido que estos descensos podrían estar relacionados con el Real Decreto-ley 16/2012, de 20 de abril, donde se recogieron algunas medidas para la sostenibilidad del SNS. Con este Real Decreto-ley algunos medicamentos dejaron de financiarse y quizás sea la razón de esta disminución.

El resto de diagnósticos y fármacos no sufren cambios notables. De hecho, aunque han sido seleccionados para el posterior análisis, la tasa de aumento o disminución está muy próxima al umbral establecido (5 %). También se comentaron estas observaciones con los colaboradores clínicos, y se planteó la posibilidad de que algunos de estos ligeros incrementos podrían estar relacionados con una buena adherencia al tratamiento. Este podría ser el caso, por ejemplo, de los fármacos A10BD (hipoglucemiantes orales en asociación), A10AE (insulinas y análogos acción lenta para inyección), J01CR (asociaciones de penicilinas, incluidas inhibidores betalactamasa) y C09AA (inhibidores enzima convertasa angiotensina solos).

5.2. Análisis de Correspondencias Múltiple

Tras la selección de características basada en las diferencias significativas entre diferentes grupos de pacientes, se realiza el ACM. Como ya se ha comentado, el ACM permite visualizar gráficamente las asociaciones recogidas en una tabla. En este TFG primero se visualizarán las asociaciones entre las características seleccionadas y, en segundo lugar, entre los individuos (pacientes de los diferentes estados de salud).

Al igual que se realizó en la selección de características, primero se aplica el ACM sobre los estados de salud de un año concreto (2012), y después para cada estado de salud en un intervalo temporal (2012-2015). Para el primer caso, se usarán las características seleccionadas en la Sección 5.1.1. Para el segundo caso, se usarán las características seleccionadas al caracterizar la evolución temporal (Sección 5.1.2, donde se comparaba cada CRG-base en 2012 y en 2015).

Algunas consideraciones a tener en cuenta antes de presentar los resultados:

- Tras la selección de características había códigos que se seleccionaban para varios estados de salud crónicos. Antes de realizar el ACM, se eliminan códigos repetidos, solo

manteniendo códigos únicos. Por ejemplo, el diagnóstico 250 mostraba una diferencia significativa en el CRG-base 5424, en el 6313 y en el 7140. Este código se considerará sólo una vez.

- El ACM que se lleva a cabo en este TFG se realiza submuestreando cada grupo de pacientes crónicos para que todos ellos tengan el mismo número de individuos, consiguiendo así grupos balanceados. Por tanto puede ocurrir que algún diagnóstico o ATC de los seleccionados no esté presente en ninguno de los grupos submuestreados, lo que daría lugar a una columna de ceros. Esto produciría una indeterminación al calcular las matrices inversas necesarias para obtener \mathbf{S} (véase el Apéndice H, Ecuación H.2). Para solucionarlo, se añade un paciente adicional que se denominará “paciente desconocido” y que recogerá las categorías de las variables no presentes en los pacientes del submuestreo.
- En el ACM, la suma de los valores asociados a las categorías de una misma variable debe ser igual a 1 para cada paciente (véase el Apéndice H, Ecuación H.4). Sin embargo, puesto que en nuestro caso un paciente puede tener simultáneamente varios diagnósticos o ATCs, se ha normalizado la presencia de cada categoría respecto al total de la variable a la que pertenece. Además, también puede ocurrir que alguno de los pacientes no posean ninguno de los diagnósticos o fármacos considerados. Para solucionar esto, se ha añadido la categoría “Ningún diagnóstico” y “Ningún fármaco” a cada variable respectivamente, para recoger estos casos.

5.2.1. Caracterización del estado de salud

A continuación se muestra la caracterización del estado de salud de los grupos crónicos haciendo uso del ACM. Primero se muestra la proporción de inercia explicada por cada autovalor, junto con los tres autovectores asociados a los tres autovalores que mayor inercia explican. Después se presenta la asociación entre las categorías seleccionadas de las tres variables (estado de salud, diagnósticos y fármacos). Para obtener una mejor visualización, se muestra también la asociación haciendo uso de diferentes filtros (aplicados sólo a la visualización).

Tras la selección de características de la Sección 5.1.1 se obtiene 84 características a analizar: 28 diagnósticos + el campo “Ningún diagnóstico” + 48 ATCs + el campo “Ningún ATC” + 6 CRGs-base. Las columnas de la matriz binaria Z están organizadas siguiendo ese orden, así como las representaciones mostradas a continuación. Concretamente, las componentes 79, 80, 81, 82, 83 y 84, corresponden a los CRGs-base 3755, 5133, 5192, 5424, 6313 y 7140, respec-

tivamente. Por tanto, el espacio original en el que podrían representarse los perfiles sin pérdida de información, estaría formado por 84 dimensiones. Para construir el subespacio transformado \mathbf{S} , de dimensionalidad reducida, se considerarán los tres autovectores de mayor inercia.

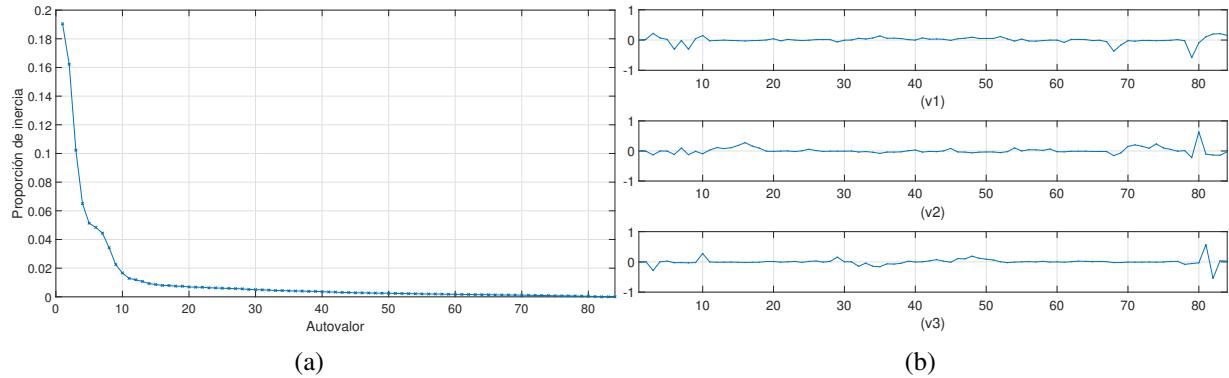


Figura 5.7: Proporción de inercia explicada por cada autovalor (a) y los tres autovectores con mayor inercia (b).

La Figura 5.7 (a) representa, en el panel izquierdo, la proporción de inercia que posee cada autovalor al aplicar el ACM sobre grupos balanceados que consideran los seis CRGs-base de pacientes crónicos. Obsérvese que las tres dimensiones consideradas (con mayor inercia) tienen una proporción de inercia explicada de 0.19, 0.16 y 0.10, respectivamente. Estas tres dimensiones serán las que se usen para visualizar las asociaciones entre categorías e individuos.

La Figura 5.7 (b) muestra los valores de las componentes asociadas para cada uno los tres autovectores con mayor inercia (v_1 , v_2 y v_3). Se puede comprobar, por ejemplo, que la componente 79 de v_1 (CRG-base 3755) tiene un valor de -0.58 y las componentes 80, 81, 82, 83 y 84 tienen valores muy cercanos a 0. Se concluye que el autovector v_1 está discriminando al CRG-base 3755 (depresión) frente al resto de CRGs-base. En cuanto al autovector v_2 , la componente 80 (corresponde al CRG-base 5133) tiene un valor de 0.63 y las componentes asociadas al resto de CRGs-base tienen valores cercanos a 0. Por tanto, el autovector v_2 discrimina a los pacientes con EPOC frente al resto de grupos. Por último, en el autovector v_3 , las componentes 81 (CRG-base 5192) y 82 (CRG-base 5424) tienen valores de 0.57 y -0.54, respectivamente, y para el resto de CRGs-base valores próximos a cero. Este autovector discrimina a los pacientes hipertensos y diabéticos frente al resto.

La Figura 5.8 representa la proyección de las categorías (perfils-columna) en el subespacio tridimensional creado con los autovectores v_1 , v_2 y v_3 . Obsérvese cómo los 6 CRGs-base (asteriscos verdes) rodean una nube de puntos que corresponde a las categorías de diagnósticos

(cruces azules) y fármacos (círculos rojos), y cómo algunas de esas categorías se acercan más a cierto CRG-base (por ejemplo, el diagnóstico 250 al CRG-base 5424, o los códigos 401 y C09AA al CRG-base 5192). Por otro lado, al crear un espacio tridimensional y no poder rotar la imagen sobre el papel, las asociaciones no se visualizan claramente. Es por ello que, en las siguientes representaciones, se ha decidido emplear sólo dos dimensiones, aunque las interpretaciones se realizan considerando las tres.

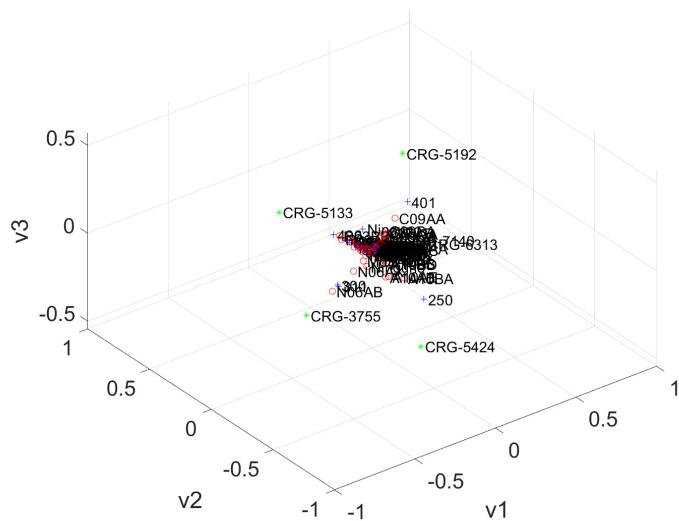


Figura 5.8: Asociación entre categorías de diagnósticos, fármacos y CRG-base.

En la Figura 5.9 se muestran las asociaciones filtrando, por un lado, diagnósticos y CRGs-base (Figura 5.9 (a)) y por otro lado fármacos y CRGs-base (Figura 5.9 (b)). Se visualiza ahora una especie de triángulo externo en uno de cuyos vértices queda el CRG-base 3755, en otro el CRG-base 5133, y en el otro los CRGs-base 5192, 5424, 6313, 7140. Por otro lado, en el interior del triángulo están las categorías de diagnósticos y fármacos, respectivamente. También se empieza a observar cómo algunos códigos se asocian más a algunos CRGs-base. Por ejemplo, en la Figura 5.9 (a), el código diagnóstico 496 (obstrucción crónica de vías respiratorias, no clasificada bajo otros conceptos) se asocia al CRG-base 5133 (EPOC), mientras que los códigos 300 (trastornos de ansiedad, disociativos y somatomorfos) y 311 (trastorno depresivo, no clasificado bajo otros conceptos) se asocian al CRG-base 3755. Por otro lado, en la Figura 5.9 (b), se observa que los códigos ATC N06AB (inhibidores selectivos recaptación serotonina-antidepresivos) y N06AX (otros antidepresivos) están más asociados al CRG-base 3755, mientras que los ATCs R03BB (anticolinérgicos-antiasmáticos) y R03AC (agonistas beta-2-adrenérgicos selectivos) lo están al CRG-base 5133, como cabría esperar.

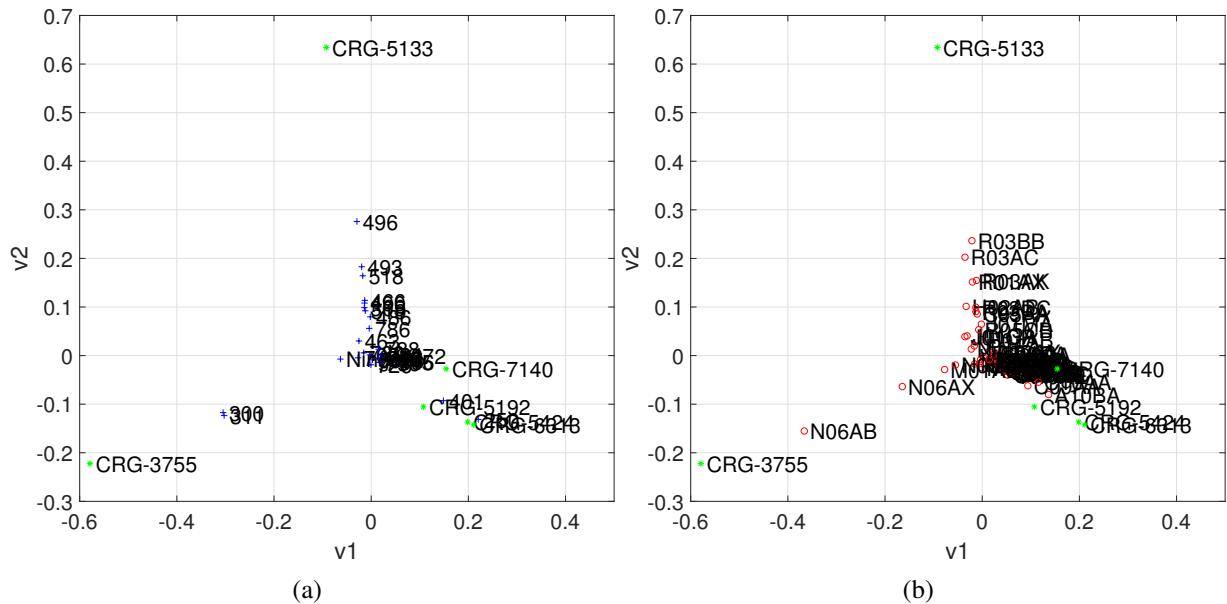


Figura 5.9: Asociación entre categorías de la variable diagnósticos y CRGs-base (a), y entre fármacos y CRGs-base (b).

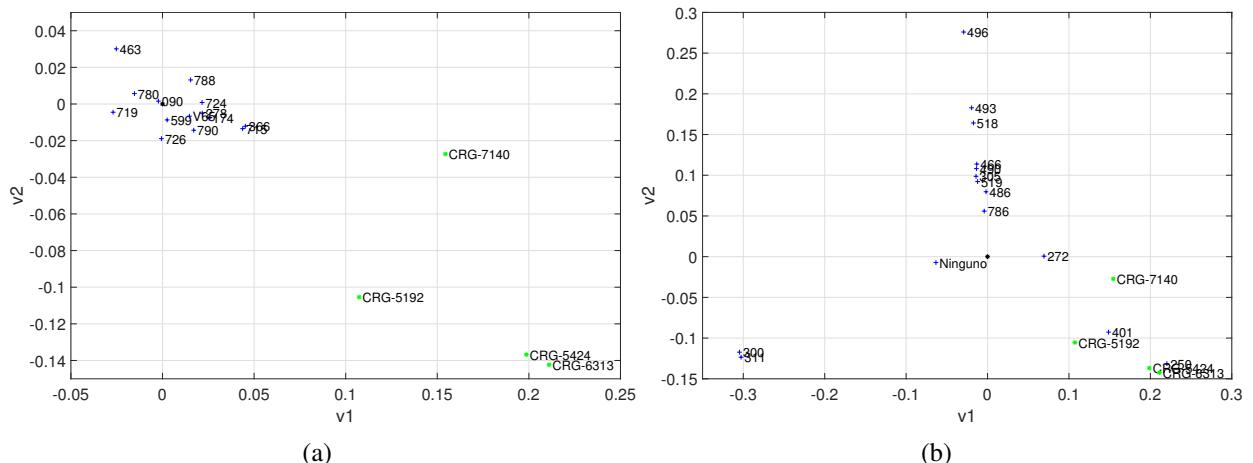


Figura 5.10: Asociación entre categorías de la variable diagnósticos y CRGs-base, representando de manera independiente los que están cerca (a) y lejos (b) del origen.

En las Figuras 5.10 (a) y (b) se muestran las asociaciones entre diagnósticos y CRGs-base. Para mejorar la visualización, se han filtrado, por un lado, los diagnósticos que aparecían cerca del origen de coordenadas, y por otro lado, los que están más lejos del origen. Nótese que al ampliar la imagen se han perdido del campo visual algunos CRGs-base. Además, como se comentó en el capítulo anterior, al reducir la dimensionalidad se pierde información. Así, en

la Figura 5.8 se observa que los CRGs-base 6313 y 7140 quedan situados entre los CRGs-base 5192 y 5424. En cambio, en la proyección 2D mostrada en la Figura 5.10 (a), los CRGs-base 5424 y 6313 están más próximos entre sí que a los CRGs-base 5192 y 7140. Por otro lado, también en la Figura 5.10 (a) se observan diagnósticos (366, 715, 726, 174, entre otros) asociados con los CRGs de HTA y DM, y en menor medida con depresión y EPOC. Por otro lado, la Figura 5.10 (b) muestra cómo los diagnósticos 496, 493 y 518 están muy asociados a EPOC; 300 y 311 a depresión; 401 muy asociado al CRG-base 5192, 250 al CRG-base 5424 y los diagnósticos 272, 786 y 486 a los CRGs de HTA y DM.

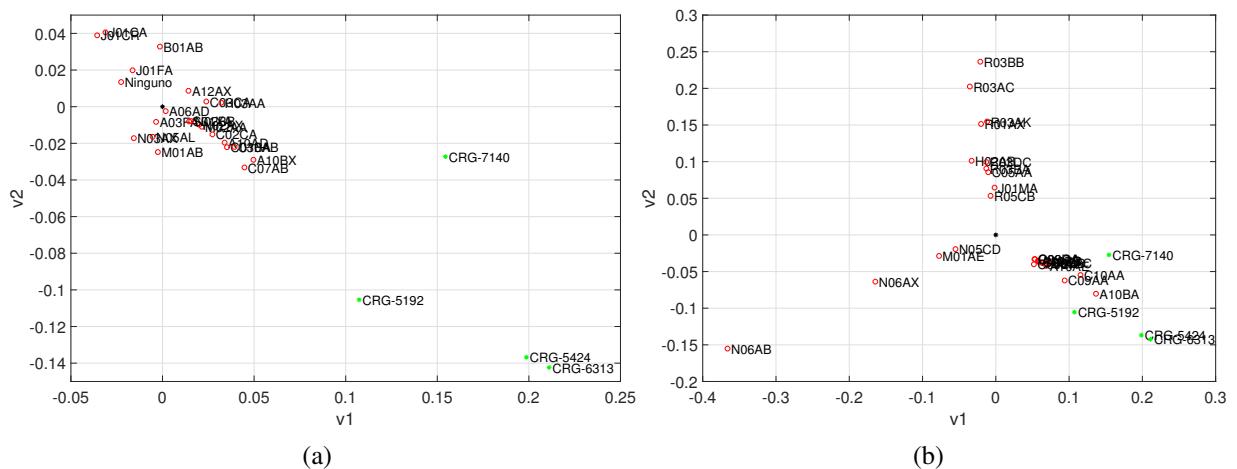


Figura 5.11: Asociación entre categorías de la variable fármacos y CRGs-base, representando de manera independiente los que están cerca (a) y lejos (b) del origen.

En cuanto a las asociaciones entre fármacos y CRGs-base, en la Figura 5.11 se aprecia una nube de puntos cercanos al origen de coordenadas (véase la Figura 5.11 (a)), y algunos códigos que se asocian en mayor medida a determinados CRGs-base (véase la Figura 5.11 (b)). Al ampliar la gráfica para visualizar mejor los códigos, se dejan de representar algunos CRGs-base, pero estarían en la misma posición que en la Figura 5.9 (b). Por ejemplo, volviendo a la Figura 5.11 (b), los códigos R03BB, R03AC, R03AK y R01AX se asocian al CRG-base 5133, los códigos N06AB, N06AX, M01AE, N05CD al CRG-base 3755, y los códigos A10BA, C10AA, C09AA, A10AE, B01AC, A10BD, C08CA, A10BB, A10AB, C09BA, C09DA, C09CA, a los CRGs-base 5192, 5424, 6313 y 7140.

Por último, la Figura 5.12 muestra las proyecciones de los individuos (profiles-fila) en el subespacio creado por v1, v2 y v3, identificando con un color distinto a los pacientes de cada grupo. Se observa de nuevo cómo los CRGs-base 3755, 5133, 5192 y 5424 quedan situados en los extremos, y los CRGs-base 6313 y 7140 quedan entre los CRGs-base 5192 y 5424.

Precisamente los pacientes de los CRGs-base 6313 y 7140 comparten HTA y DM, siendo lógico que se asocien en mayor medida con los grupos de CRGs-base 5192 y 5424 y en menor medida con los CRGs-base 3755 y 5133. También se puede apreciar cómo los pacientes de un mismo CRG-base no se organizan en agrupaciones compactas y en algunos casos (aunque no se aprecie en la figura por la imposibilidad de rotar la imagen sobre el papel), se generan *subgrupos*. En el caso del CRG-base 7140, estos subgrupos podrían corresponder a pacientes con “otra cronicidad”. En el caso de los CRGs-base 5424 y 6313, los subgrupos podrían estar asociados a diferentes tipos de diabetes.

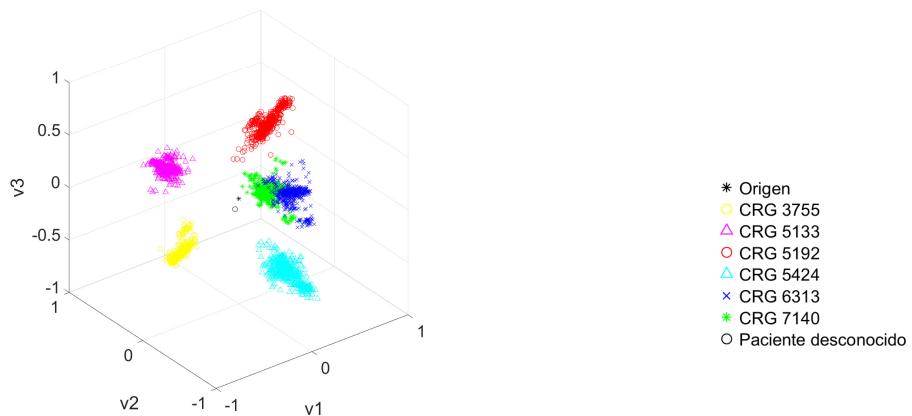


Figura 5.12: Representación de individuos de los CRGs-base de interés.

5.2.2. Evolución temporal

En esta sección se presenta la aplicación del ACM (con las características seleccionadas en la Sección 5.1.2) para caracterizar la evolución temporal de cada grupo (o estado de salud) entre 2012 y 2015. De esta forma, se podrá caracterizar la evolución temporal en la dispensación de fármacos seleccionados o en la presencia de los diagnósticos considerados, entre 2012 y 2015. Las representaciones gráficas para caracterizar la asociación tanto entre categorías como entre pacientes se muestran en el Apéndice I. Para no extender la memoria, las descripciones de los códigos se pueden consultar en el Apéndice G. A continuación se detallan algunas conclusiones.

En cuanto al CRG-base 3755, de la Figura I.1 (a), se concluye que en general los códigos seleccionados, como ya se evidenciaba en el análisis *bootstrap*, están asociados en mayor medida a 2012 (es el caso de los códigos: N05BA, A02BC, M01AB, M02AA, R05CB y 311). Para el CRG-base 5133 (véase la Figura I.2 (a)), casi todos los códigos están igualmente representados

en ambos años (2012 y 2015), salvo por el diagnóstico 599 (otros trastornos de uretra y tracto urinario), mayormente asociado a 2015, y el fármaco R05CB (mucolíticos) que está presente en mayor medida en 2012. En el CRG-base 5192 (véase la Figura I.3 (a)), se observa que los fármacos M01AE y M01AB están presentes en una proporción muy similar en 2012 y 2015. Sin embargo, los códigos M02AA, S01XA y R05CB están más presente al inicio del periodo, concluyendo que esos fármacos se dispensaron en mayor medida en 2012 frente a 2015. Para el CRG-base 5424 (véase la Figura I.4 (a)), ocurre lo mismo salvo por el diagnóstico 526 (enfermedades de los maxilares) más presente en el año 2015. En el caso del CRG-base 6313 (véase la Figura I.5 (a)), hay dos fármacos (A10AE y A10BD) igualmente representados en 2012 y 2015, y otros (M01AE, M02AA, R05CB y S01XA) que se aproximan en mayor medida a 2012. Por último, el CRG-base 7140 es en el que más diferencias significativas se han encontrado entre 2012 y 2015 (véase la Figura I.6 (a)), sin embargo casi todas con diferencias de proporciones muy cercanas al umbral (5 %), salvo para los fármacos S01XA, M02AA, R05CB que se presentan con mayor frecuencia en 2012.

En cuanto a la proyección de los pacientes, en el panel derecho de las Figuras I.1 - I.6, se ilustra cómo en el subespacio creado se distinguen perfectamente los pacientes de cada grupo en 2012 y 2015. Dentro del mismo grupo también se observan *subgrupos* para los años considerados. Esto podría deberse a que un grupo recoge diferentes patologías (como es el caso de los CRGs-base 6313 y 7140) o diferentes perfiles de pacientes (como en el CRG-base 5133 que como ya se evidenciaba en el Capítulo 3, aparecen pacientes claramente diferenciados por edad y posiblemente también por tratamiento).

Se podría concluir, que en cada CRG-base, no se observan grandes cambios entre 2012 y 2015. Aunque se han encontrado algunas diferencias significativas, se encuentran próximas al umbral establecido y no está claro si estos cambios en los fármacos/diagnósticos considerados son relevantes. Indicar que estos pequeños cambios se producen en mayor medida en los ATCs que en los diagnósticos. Destacan los siguientes ATCs: R05CB (mucolíticos), M02AA (antiinflamatorios no esteroideos para uso tópico), S01XA (otros oftalmológicos) y M01AB (derivados del ácido acético y sustancias relacionadas).

Capítulo 6

Conclusiones y líneas futuras

En este capítulo se exponen algunas conclusiones extraídas del análisis realizado y se plantean algunas líneas futuras que se podrían seguir para extender este trabajo.

6.1. Conclusiones

Las enfermedades crónicas son uno de los principales problemas de salud hoy en día. Concretamente, se ha podido corroborar con los datos recogidos por el HUF de los pacientes que pertenecen al Área de Salud de Fuenlabrada, que algunas de las enfermedades crónicas que afectan a un mayor número de pacientes son, entre otras, el asma, la hipertensión o la diabetes.

A través del análisis exploratorio, se ha comprobado cómo la edad media de los grupos de pacientes crónicos aumentaba al empeorar el estado de salud. En el grupo de pacientes con depresión (estado de salud de nivel 3) la edad media de los pacientes era de 42 años, mientras que en el grupo de pacientes con hipertensión, diabetes y otra cronicidad (estado de salud de nivel 7), asciende a una edad media de 67 años. Además, en el grupo con EPOC parece que hay dos patrones de pacientes, diferenciados por las dos modas en la distribución de edad de este grupo. También se ha podido comprobar cómo, en los grupos de EPOC y DM, hay mayor número de hombres que de mujeres, al contrario de lo que sucede en los grupos de depresión e hipertensión, donde predominan las mujeres. En el CRG-base 7140, generalmente con pacientes en edad avanzada, destaca el mayor número de mujeres, lo que podría justificarse por una esperanza de vida más alta para las mujeres.

Por otro lado, el test de proporciones y el remuestreo *bootstrap* realizado en este TFG, ha

permitido encontrar diferencias significativas en la tasa de presencia de diagnósticos y fármacos entre diferentes grupos. Al comparar los grupos de crónicos frente al grupo de pacientes sanos, se ha concluido que los pacientes diabéticos e hipertensos, además del trastorno principal, también tienen asociados problemas de espalda, de articulaciones y trastornos alimentarios. Además, los pacientes diabéticos poseen también diagnósticos relacionados con hipertensión, lo que podría indicar que ambas enfermedades están relacionadas, y cataratas. Por otro lado, los hipertensos tienen asociados trastornos de ansiedad, típicos de los pacientes con depresión.

El test de proporciones también ha permitido encontrar diferencias relevantes entre grupos de pacientes en 2012 y 2015, caracterizando así la evolución temporal de cada estado de salud. En este aspecto, llama la atención la disminución generalizada de los ATCs R05CB (mucolíticos), M02AA (antiinflamatorios no esteroideos para uso tópico) y S01XA (otros oftalmológicos), sobre todo por el hecho de que estas disminuciones se observan en todos los estados de salud. Este descenso podría estar justificado por el hecho de que, en 2012, estos medicamentos dejaron de ser financiados por la Seguridad Social. Estos cambios se introdujeron justo en ese año con el Real Decreto-ley 16/2012 del 20 de abril de 2012.

Por último, las características (diagnósticos y fármacos) que presentan diferencias significativas, por un lado, al comparar los grupos de crónicos con los pacientes sanos, y por otro, al comparar los estados de salud en años diferentes, han sido seleccionadas para el posterior ACM. Este análisis ha permitido representar gráficamente asociaciones entre las características seleccionadas. Esta representación se consigue mediante la transformación de un espacio original de 84 dimensiones a un subespacio de dimensionalidad reducida que, aún con cierta pérdida de información, captura las asociaciones entre diagnósticos, fármacos y CRGs-base. También se han encontrado asociaciones entre los pacientes de diferentes grupos. Esta visualización ha permitido encontrar subgrupos de pacientes con características ligeramente diferentes al resto de individuos del grupo al que pertenecen, que podrían ser tratados de una forma más especializada.

6.2. Líneas futuras

Para extender el análisis realizado en este TFG, se plantean varias líneas futuras:

- Este TFG ha considerado CRGs-base, determinados con cuatro dígitos. No se ha tenido en cuenta el último dígito del CRG, que hace referencia al nivel de gravedad. Se propone considerar este último dígito, algo que quizás permitiría caracterizar los diferentes tipos

de diabetes. Se podría considerar este aspecto en futuros análisis. Algo similar ocurre con los diagnósticos y ATCs. En ambos casos se ha reducido la precisión a la hora de seleccionar los códigos diagnósticos (CIE-9) y fármacos (ATCs) considerados a estudio, sin tener en cuenta los últimos dígitos de su codificación. Una posible línea futura podría desarrollarse este sentido.

- También se podría considerar dividir cada grupo de pacientes por rangos de edad. Por ejemplo, en el caso de los pacientes con EPOC, se podrían capturar los dos perfiles de pacientes que se observaron al realizar el análisis exploratorio en el Capítulo 3.
- Otro aspecto que se podría contemplar a la hora de encontrar diferencias significativas, es el análisis en términos de ocurrencia, ya que este TFG se ha centrado sólo en la presencia o ausencia de las características analizadas, pero no ha tenido en cuenta el número de veces que aparece cada una de ellas anualmente en la historia clínica de cada paciente.
- Por último, si en un futuro se contara con datos relacionados con hábitos de vida de los pacientes (consumo de alcohol, tabaco, dieta o ejercicio) o se pudiera ampliar las variables disponibles en este momento (peso, análisis de sangre o niveles colesterol, entre otros), se podría plantear el diseño de un modelo predictivo. El objetivo de este modelo sería predecir un futuro empeoramiento del estado de salud de un paciente para tomar medidas clínicas preventivas a fin de evitar o ralentizar esta evolución.

Apéndice A

Extracción de datos

Para empezar a analizar las bases de datos, lo primero que se ha decidido hacer es unificar en una sola matriz todos los datos de interés relativos a cada paciente, de manera ordenada, para cada año analizado. A continuación, se describe brevemente el proceso seguido.

Para llevar a cabo dicha unificación debemos extraer los datos necesarios de las bases de datos proporcionadas. Estas bases de datos son archivos .accdb, es decir, bases de datos de Microsoft Access. Matlab (el software usado para llevar a cabo este TFG) no posee ninguna función para abrir o leer este tipo de archivos, pero posee una herramienta, llamada *Database Explorer*. Esta herramienta permite establecer una conexión con la base de datos, realizar consultas, extraer las partes de la base de datos que sean de interés y cerrar la conexión. Una vez se ha descargado esta aplicación, tenemos que configurar el origen de los datos, lo que se hace a través de una aplicación de Windows llamada *Administrador de origen de datos ODBC*. Seleccionamos la pestaña DNS de usuario, agregamos el controlador que necesitamos para establecer la conexión con nuestra base de datos, en este caso *Microsoft Access Driver (*.mdb, *.accdb)*, asignamos un nombre con el que acceder a nuestros datos y el origen de dichos datos, en este caso, la base de datos de los pacientes del Hospital de Fuenlabrada.

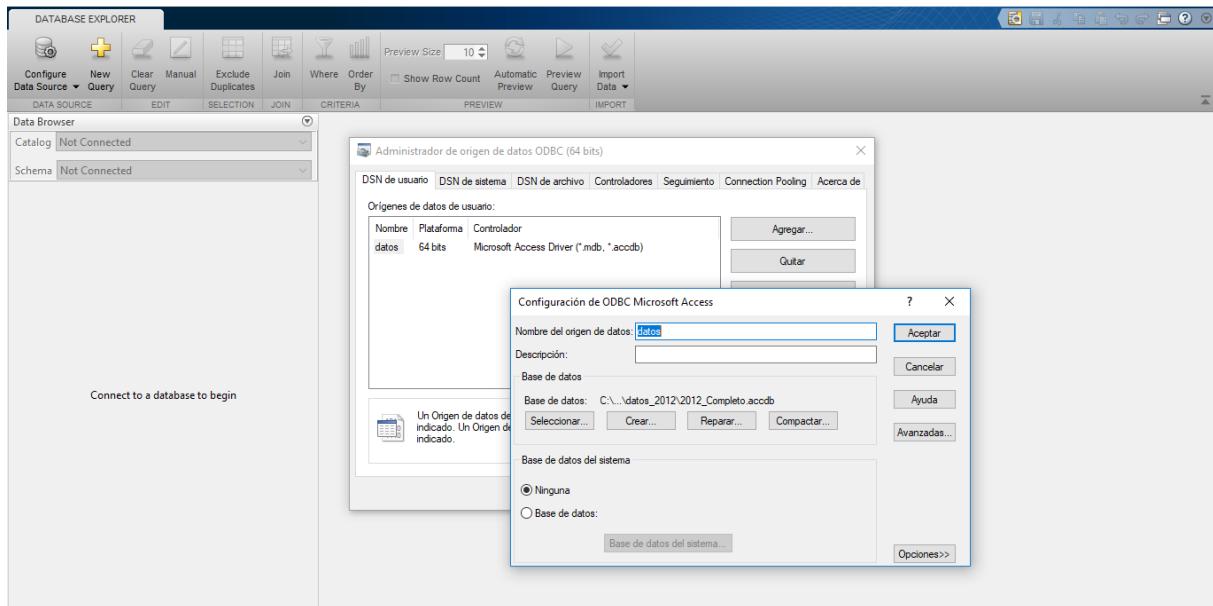


Figura A.1: Administrador de origen de datos ODBC de Windows.

En la Figura A.1 se muestra cómo se ha realizado el procedimiento anteriormente descrito. Una vez hecho esto, podemos realizar una consulta a la base de datos pinchando en la pestaña *New Query*. Se puede ver en la Figura A.2 y A.3 cómo se establece la conexión y cómo se accede a los datos. Es importante realizar este paso, pues al acceder a la base de datos desde Matlab no es necesario extraer toda la base de datos, siendo posible extraer solo las columnas de interés, ahorrando así tiempo de procesado y ejecución.

Para extraer los datos necesarios desde el programa y crear una matriz con la que realizar el análisis, se usa la función `conn = database(datasource,username,password)` que permite establecer conexión a nuestra base de datos indicándole el origen de los datos, previamente configurado. Una vez establecida la conexión podemos acceder a los datos mediante la función `data = select(conn,selectquery)`, donde `selectquery` es la consulta *SQL Query* que se muestra en la aplicación *Database Explorer*, pudiendo extraer así las columnas que resulten más interesantes.

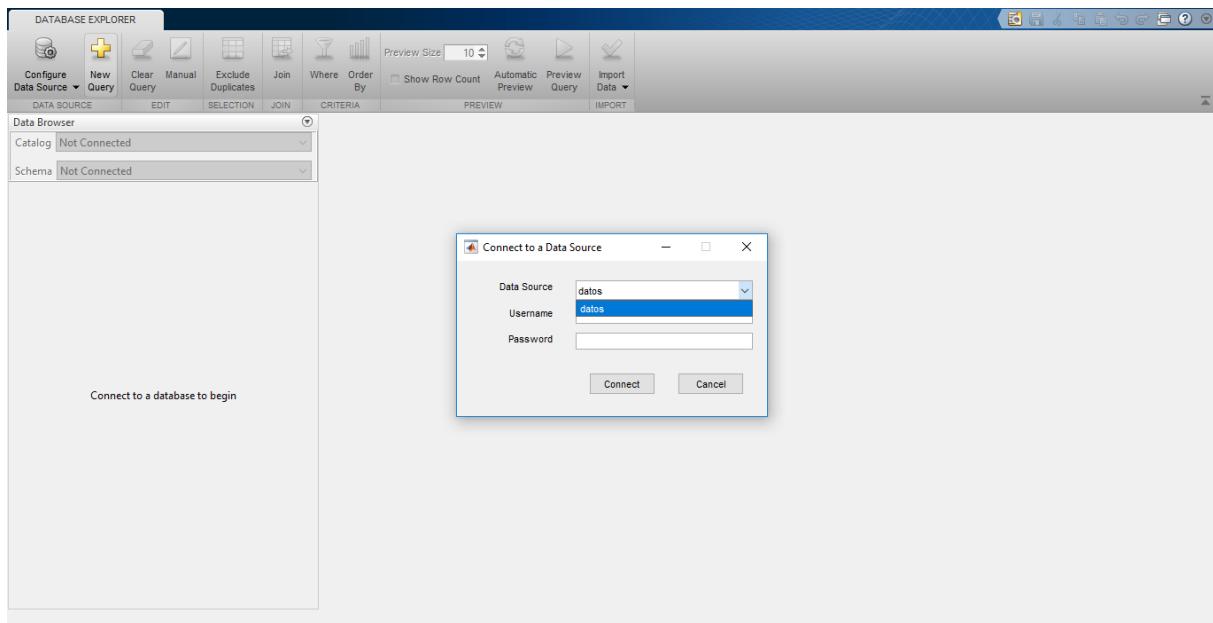


Figura A.2: Nueva consulta a la base de datos.

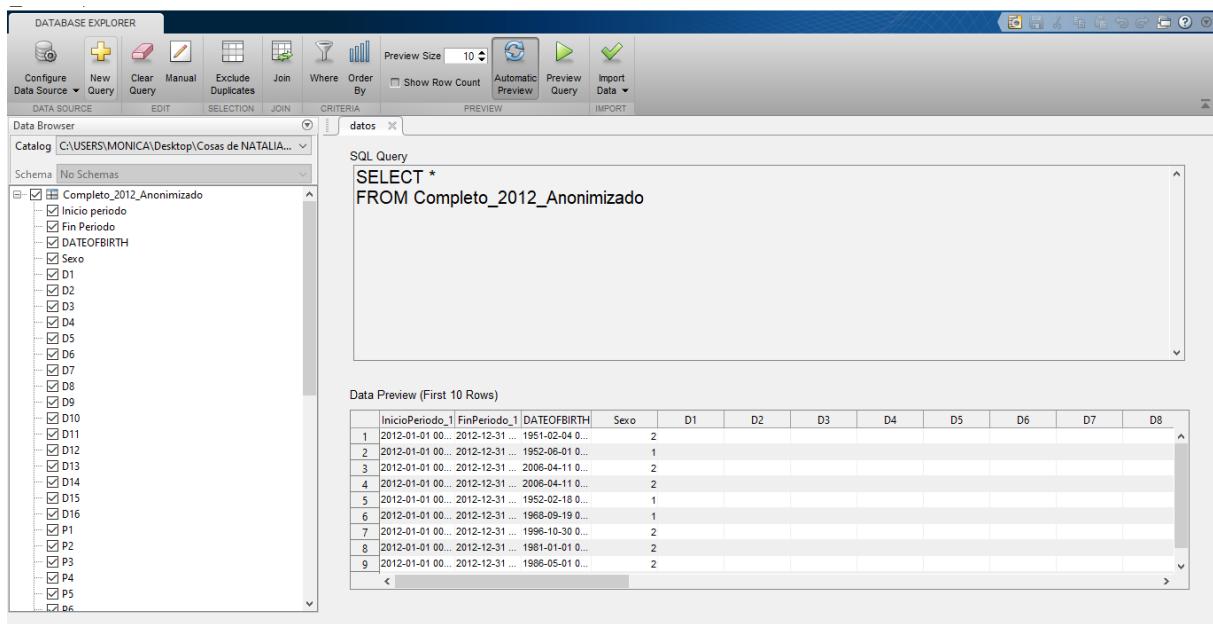


Figura A.3: Acceso a la base de datos completa.

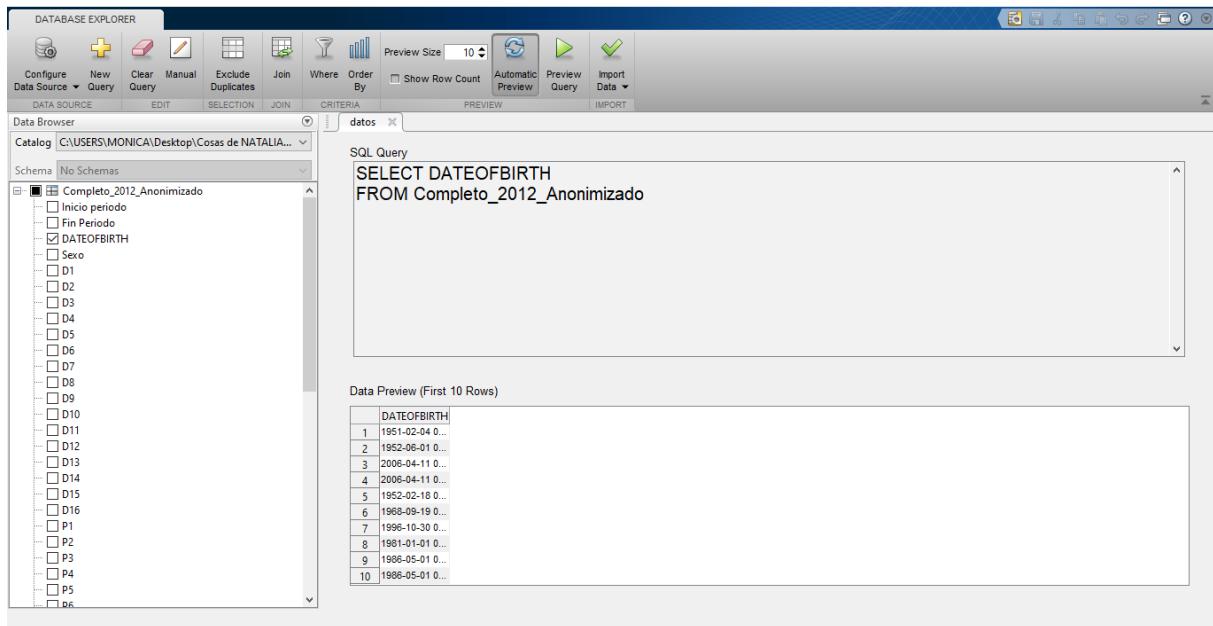


Figura A.4: Acceso a la fecha de nacimiento.

En la Figura A.4 se observa cómo, mediante la consulta *SQL Query*: *SELECT DATEOFBIRTH FROM completo_2012_Anonimizado* podemos acceder únicamente a la columna fecha de nacimiento.

Apéndice B

Diagnósticos considerados de HTA, DM, depresión y EPOC

A continuación, se muestran los códigos utilizados en el análisis exploratorio relacionados con HTA (Cuadro B.1), DM (Cuadro B.2), depresión (Cuadro B.3) y EPOC (Cuadro B.4). Estas listas han sido realizadas en colaboración con el personal sanitario, siendo algunos de los diagnósticos de mayor interés en el entorno clínico.

Código	Descripción
362.11	Retinopatía hipertensiva
401.0	Hipertensión esencial. Maligna
401.1	Hipertensión esencial. Benigna
401.9	Hipertensión esencial. No especificada
402.0	Cardiopatía hipertensiva. Maligna
402.00	Cardiopatía hipertensiva. Sin insuficiencia cardiaca
402.01	Cardiopatía hipertensiva. Con insuficiencia cardiaca
402.1	Cardiopatía hipertensiva. Benigna
402.10	Cardiopatía hipertensiva. Sin insuficiencia cardiaca
402.11	Cardiopatía hipertensiva. Con insuficiencia cardiaca
402.9	Cardiopatía hipertensiva. No especificada
402.90	Cardiopatía hipertensiva. Sin insuficiencia cardiaca
402.91	Cardiopatía hipertensiva. Con insuficiencia cardiaca
403.0	Enfermedad del riñón hipertensiva crónica. Maligna

403.00	Con enfermedad crónica del riñón estadio I a estadio IV, o sin especificar
403.01	Con enfermedad crónica del riñón estadio V o enfermedad renal terminal
403.1	Enfermedad del riñón hipertensiva crónica. Benigna
403.10	Con enfermedad crónica del riñón estadio I a estadio IV, o sin especificar
403.11	Con enfermedad crónica del riñón estadio V o enfermedad renal terminal
403.9	Enfermedad del riñón hipertensiva crónica. No especificada
403.90	Con enfermedad crónica del riñón estadio I a estadio IV, o sin especificar
403.91	Con enfermedad crónica del riñón estadio V o enfermedad renal terminal
404.0	Enfermedad cardiaca y renal crónica hipertensiva. Maligna
404.00	Enfermedad cardiaca y renal crónica hipertensiva. Sin insuficiencia cardiaca y con enfermedad crónica del riñón estadio I a estadio IV, o sin especificar
404.01	Enfermedad cardiaca y renal crónica hipertensiva. Con insuficiencia cardiaca y con enfermedad crónica del riñón estadio I a estadio IV, o sin especificar
404.02	Enfermedad cardiaca y renal crónica hipertensiva. Sin insuficiencia cardiaca y con enfermedad crónica del riñón estadio V o estadio final de enfermedad renal
404.03	Enfermedad cardiaca y renal crónica hipertensiva. Con insuficiencia cardiaca y enfermedad crónica del riñón estadio V o estadio final de enfermedad renal
404.1	Enfermedad cardiaca y renal crónica hipertensiva. Benigna
404.10	Enfermedad cardiaca y renal crónica hipertensiva. Sin insuficiencia cardiaca y con enfermedad crónica del riñón estadio I a estadio IV, o sin especificar
404.11	Enfermedad cardiaca y renal crónica hipertensiva. Con insuficiencia cardiaca y con enfermedad crónica del riñón estadio I a estadio IV, o sin especificar
404.12	Enfermedad cardiaca y renal crónica hipertensiva. Sin insuficiencia cardiaca y con enfermedad crónica del riñón estadio V o estadio final de enfermedad renal
404.13	Enfermedad cardiaca y renal crónica hipertensiva. Con insuficiencia cardiaca y enfermedad crónica del riñón estadio V o estadio final de enfermedad renal
404.9	Enfermedad cardiaca y renal crónica hipertensiva. No especificada
404.90	Enfermedad cardiaca y renal crónica hipertensiva. Sin insuficiencia cardiaca y con enfermedad crónica del riñón estadio I a estadio IV, o sin especificar
404.91	Enfermedad cardiaca y renal crónica hipertensiva. Con insuficiencia cardiaca y con enfermedad crónica del riñón estadio I a estadio IV, o sin especificar
404.92	Enfermedad cardiaca y renal crónica hipertensiva. Sin insuficiencia cardiaca y con enfermedad crónica del riñón estadio V o estadio final de enfermedad renal

404.93	Enfermedad cardiaca y renal crónica hipertensiva. Con insuficiencia cardiaca y enfermedad crónica del riñón estadio V o estadio final de enfermedad renal
405.0	Hipertensión secundaria. Maligna
405.01	Hipertensión secundaria. Renovascular
405.09	Hipertensión secundaria. Otra
405.1	Hipertensión secundaria. Benigna
405.11	Hipertensión secundaria. Renovascular
405.19	Hipertensión secundaria. Otra
405.9	Hipertensión secundaria. No especificada
405.91	Hipertensión secundaria. Renovascular
405.99	Hipertensión secundaria. Otra
760.0	Feto o recién nacido afectado por situaciones maternas no necesariamente relacionadas con el embarazo actual. Trastornos hipertensivos maternos

Cuadro B.1: Diagnósticos relacionados con HTA.

Código	Descripción
250.13	Tipo I [tipo juvenil], incontrolada
250.2	Diabetes con hiperosmolaridad
250.20	Tipo II o de tipo no especificado, no establecida como incontrolada
250.21	Tipo I [tipo juvenil], no indicada como incontrolada
250.22	Tipo II o de tipo no especificado, incontrolada
250.23	Tipo I [tipo juvenil], incontrolada
250.3	Diabetes con otro tipo de coma
250.30	Tipo II o de tipo no especificado, no establecida como incontrolada
250.31	Tipo I [tipo juvenil], no indicada como incontrolada
250.32	Tipo II o de tipo no especificado, incontrolada
250.33	Tipo I [tipo juvenil], incontrolada
250.4	Diabetes con manifestaciones renales
250.40	Tipo II o de tipo no especificado, no establecida como incontrolada
250.41	Tipo I [tipo juvenil], no indicada como incontrolada
250.42	Tipo II o de tipo no especificado, incontrolada
250.43	Tipo I [tipo juvenil], incontrolada

250.5	Diabetes con manifestaciones oftálmicas
250.50	Tipo II o de tipo no especificado, no establecida como incontrolada
250.51	Tipo I [tipo juvenil], no indicada como incontrolada
250.52	Tipo II o de tipo no especificado, incontrolada
250.53	Tipo I [tipo juvenil], incontrolada
250.6	Diabetes con manifestaciones neurológicas
250.60	Tipo II o de tipo no especificado, no establecida como incontrolada
250.61	Tipo I [tipo juvenil], no indicada como incontrolada
250.62	Tipo II o de tipo no especificado, incontrolada
250.63	Tipo I [tipo juvenil], incontrolada
250.7	Diabetes con trastornos circulatorios periféricos
250.70	Tipo II o de tipo no especificado, no establecida como incontrolada
250.71	Tipo I [tipo juvenil], no indicada como incontrolada
250.72	Tipo II o de tipo no especificado, incontrolada
250.73	Tipo I [tipo juvenil], incontrolada
250.8	Diabetes con otras manifestaciones especificadas
250.80	Tipo II o de tipo no especificado, no establecida como incontrolada
250.81	Tipo I [tipo juvenil], no indicada como incontrolada
250.82	Tipo II o de tipo no especificado, incontrolada
250.83	Tipo I [tipo juvenil], incontrolada
250.9	Diabetes con complicación no especificada
250.90	Tipo II o de tipo no especificado, no establecida como incontrolada
250.91	Tipo I [tipo juvenil], no indicada como incontrolada
250.92	Tipo II o de tipo no especificado, incontrolada
250.93	Tipo I [tipo juvenil], incontrolada
588.1	Diabetes insípida nefrogénica
648.0	Diabetes mellitus
648.00	Sin especificación con respecto a episodio de cuidados
648.01	Parto, con o sin mención de estado anteparto
648.02	Parida, con mención de complicación postparto
648.03	Estado o complicación anteparto
648.04	Estado o complicación postparto

775.1	Diabetes mellitus neonatal
-------	----------------------------

Cuadro B.2: Diagnósticos relacionados con DM.

Código	Descripción
296.2	Trastorno depresivo mayor, episodio crónico
296.20	No especificada
296.21	Leve
296.22	Moderada
296.23	Grave, sin mención de comportamiento psicótico
296.24	Grave, especificada acompañada de comportamiento psicótico
296.25	En vías de remisión parcial o no especificada
296.26	En vías de remisión total
296.3	Trastorno depresivo mayor, episodio recurrente
296.30	No especificada
296.31	Leve
296.32	Moderada
296.33	Grave, sin mención de comportamiento psicótico
296.34	Grave, especificada acompañada de comportamiento psicótico
296.35	En vías de remisión parcial o no especificada
296.36	En vías de remisión total
296.5	Trastorno bipolar I, episodio mas reciente (o actual) depresivo
296.50	No especificada
296.51	Leve
296.52	Moderada
296.53	Grave, sin mención de comportamiento psicótico
296.54	Grave, especificada acompañada de comportamiento psicótico
296.55	En vías de remisión parcial o no especificada
296.56	En vías de remisión total
296.82	Trastorno depresivo atípico
298.0	Psicosis de tipo depresivo
301.12	Trastorno depresivo crónico de la personalidad
309.1	Reacción depresiva prolongada

311	Trastorno depresivo, no clasificado bajo otros conceptos
779.2	Depresión cerebral, coma y otros signos cerebrales anormales
968	Envenenamiento por otros depresores y anestésicos del sistema nervioso central
968.0	Depresores del tono muscular a través del sistema nervioso central
969.0	Antidepresivos
969.00	Antidepresivo, sin especificar
969.01	Inhibidores de la monoamino-oxidasa
969.02	Inhibidores selectivos de la recaptación de serotonina y norepinefrina
969.03	Inhibidores selectivos de la recaptación de serotonina
969.04	Antidepresivos tetracíclicos
969.05	Antidepresivos tricíclicos
969.09	Otros antidepresivos

Cuadro B.3: Diagnósticos relacionados con depresión.

Código	Descripción
490	Bronquitis no especificada como aguda ni como crónica
491	Bronquitis crónica
491.0	Bronquitis crónica simple
491.1	Bronquitis crónica mucopurulenta
491.2	Bronquitis crónica obstructiva
491.20	Sin exacerbación
491.21	Con exacerbación
491.22	Con bronquitis aguda
491.8	Otra bronquitis crónica
491.9	Bronquitis crónica no especificada
492	Enfisema
492.0	Vesícula enfisematosas
492.8	Otros enfisemas
493	Asma
493.0	Asma extrínseca
493.00	No especificada
493.01	Con estado asmático

493.02	Con exacerbación (aguda)
493.1	Asma intrínseca
493.10	No especificada
493.11	Con estado asmático
493.12	Con exacerbación (aguda)
493.2	Asma obstructiva crónica
493.20	No especificada
493.21	Con estado asmático
493.22	Con exacerbación (aguda)
493.8	Otras formas de asma
493.81	Broncoespasmo por ejercicio
493.82	Tos como variante asmática
493.9	Asma no especificada
493.90	No especificada
493.91	Con estado asmático
493.92	Con exacerbación (aguda)
494	Bronquiectasia
494.0	Bronquiectasia sin exacerbación aguda
494.1	Bronquiectasia con exacerbación aguda
495	Alveolitis alérgica extrínseca
495.0	Pulmón de los granjeros
495.1	Bagazosis
495.2	Pulmón de los criadores de pájaros
495.3	Suberosis
495.4	Pulmón de los manipuladores de malta
495.5	Enfermedad de los manipuladores de champiñones
495.6	Pulmón de los descortezadores de arce
495.7	Neumonitis de "ventilación"
495.8	Otras alveolitis y neumonitis alérgicas especificadas
495.9	Alveolitis y neumonitis alérgicas no especificadas
496	Obstrucción crónica de vías respiratorias, no clasificada bajo otros conceptos
500	Neumoconiosis de los trabajadores del carbón

501	Asbestosis
502	Neumoconiosis por otro sílice o silicatos
503	Neumoconiosis por otro polvo inorgánico
504	Neumopatía por inhalación de otro tipo de polvo
505	Neumoconiosis no especificada
506	Enfermedades respiratorias por humos y vapores químicos
506.0	Bronquitis y neumonitis por humos y vapores
506.1	Edema pulmonar agudo por humos y vapores
506.2	Inflamación de las vías respiratorias superiores por humos y vapores
506.3	Otras enfermedades respiratorias agudas o subagudas por humos y vapores
506.4	Afecciones respiratorias crónicas por humos y vapores
506.9	Enfermedades respiratorias no especificadas por humos y vapores
508	Enfermedades respiratorias por otros agentes externos y agentes externos no especificados
508.0	Manifestaciones pulmonares agudas por radiaciones
508.1	Manifestaciones pulmonares crónicas y otras manifestaciones pulmonares por radiación
508.2	Enfermedades respiratorias por inhalación de humo
508.8	Enfermedades respiratorias por otros agentes externos especificados
508.9	Enfermedades respiratorias por agente externo no especificado

Cuadro B.4: Diagnósticos relacionados con EPOC.

Apéndice C

Códigos ATC considerados de HTA, DM, depresión y EPOC

En este apéndice, se muestran los códigos ATC utilizados en este TFG para el análisis exploratorio en relación con HTA (Cuadro C.1), DM (Cuadro C.2), depresión (Cuadro C.3) y EPOC (Cuadro C.4).

Código	Descripción
C02AB	Metildopa (antihipertensivos)
C02AC	Agonistas receptores imidazolina (antihipertensivos)
C02CA	Bloqueantes alfa-adrenérgicos (antihipertensivos)
C02DC	Derivados de la pirimidina (antihipertensivos)
C02DD	Derivados de la nitroferricianida (antihipertensivos)
C02K	Otros antihipertensivos
C02LC	Agonistas receptor imidazólico en asociación con diurético
C03AA	Tiazidas solas (diuréticos)
C03AX	Tiazidas asociados a otras sustancias (diuréticos)
C03BA	Sulfonamidas solas (diuréticos bajo techo)
C03CA	Sulfonamidas solas (diuréticos alto techo)
C03DA	Antagonistas de la aldosterona (diuréticos)
C03EA	Asociaciones diuréticos bajo techo+ahorradores de potasio
C03EB	Asociaciones diuréticos alto techo+ahorradores de potasio
C04AD	Derivados de las purinas (vasodilatadores periféricos)

C04AE	Alcaloides ergotaminicos (vasodilatadores periféricos)
C04AX	Otros vasodilatadores periféricos
C07AA	Betabloqueantes no selectivos solos
C07AB	Betabloqueantes selectivos solos
C07AG	Alfa y betabloqueantes
C07B	Betabloqueantes y tiazidas
C07C	Betabloqueantes y otros diuréticos
C07D	Betabloqueantes, tiazidas y otros diuréticos
C07E	Betabloqueantes y vasodilatadores
C08CA	Derivados de dihidropiridina (bloqueantes calcio efect.vasc)
C08DA	Derivados de fenilalquilamina (bloqueantes calcio selectiv.)
C08DB	Derivados de benzotiapina (bloqueantes calcio cardio-select)
C08G	Bloqueantes del canal de calcio y diuréticos
C09AA	Inhibidores enzima convertasa angiotensina solos
C09BA	Inhib.enzima convertasa angiotensina asociados a diuréticos
C09BB	Inhib.enzim.convert.angiot.asoc. bloqueantes canales calcio
C09CA	Antagonistas angiotensina II solos
C09DA	Antagonistas angiotensina II asociados a diuréticos
C09DB	Antagonistas angiotensina II asociados a bloq.canales calcio
C09XA	Inhibidores de la renina (acc.sist.renina-angiotensina.)

Cuadro C.1: Códigos ATC relacionados con HTA.

Código	Descripción
A10AB	Insulinas y análogos de acción rápida para inyección
A10AC	Insulinas y análogos acción intermedia para inyección
A10AD	Insulinas y análogos acción intermedia+rápida para inyección
A10AE	Insulinas y análogos acción lenta para inyección
A10AF	Insulinas y análogos para inhalación
A10BA	Biguanidas (antidiabéticos orales)
A10BB	Sulfonilureas (antidiabéticos orales)
A10BD	Hipoglucemiantes orales en asociación
A10BF	Inhibidores alfa-glucosidasa (antidiabéticos orales)

A10BG	Tiazolidinodionas (antidiabéticos orales)
A10BH	Inhibidores de la dipeptidil peptidasa 4 (DPP-4)
A10AB	Insulinas y análogos de acción rápida para inyección
A10BX	Otros antidiabéticos orales, excluidas insulinas

Cuadro C.2: Códigos ATC relacionados con DM.

Código	Descripción
N06AA	Inhib. no selectivos recaptación monoamina (antidepresivos)
N06AB	Inhib. selectivos recaptación serotonina (antidepresivos)
N06AF	Inhib. no selectivos monoamino-oxidasa (antidepresivos)
N06AG	Inhib. monoamino-oxidasa tipo A (antidepresivos)
N06AX	Otros antidepresivos

Cuadro C.3: Códigos ATC relacionados con depresión.

Código	Descripción
R03AA	Agonistas alfa y beta adrenérgicos (antiasmáticos)
R03AH	Adrenérgicos en asociación (antiasmáticos)
R03AK	Adrenérgicos asociados a otros antiasmáticos
R03BA	Glucocorticoides (antiasmáticos)
R03BB	Anticolinérgicos (antiasmáticos)
R03BX	Otros para obstrucción de vías respiratorias, inhalados
R03CA	Agonistas alfa y beta adrenérgicos (antiasmáticos)
R03CC	Agonistas beta-2-adrenérgicos selectivos (antiasmáticos)
R03DA	Xantinas (antiasmáticos sistémicos)
R03DB	Xantinas asociadas a adrenérgicos (antiasmáticos sistémicos)
R03DC	Antagonistas receptores del leukotrieno (antiasmáticos)
R03DX	Otros para obstrucción de vías respiratorias, uso sistemico

Cuadro C.4: Códigos ATC relacionados con EPOC.

Apéndice D

Distribución del número total de diagnósticos y códigos ATC por paciente

A continuación, se muestran los resultados de la distribución normalizada del número total de diagnósticos y códigos ATC por paciente y CRG-base de interés para el año 2012, como parte del análisis exploratorio. Nótese que cuando se habla del número total de diagnósticos y códigos ATC, se refiere al total de los códigos considerados en el estudio (Apéndice B y C).

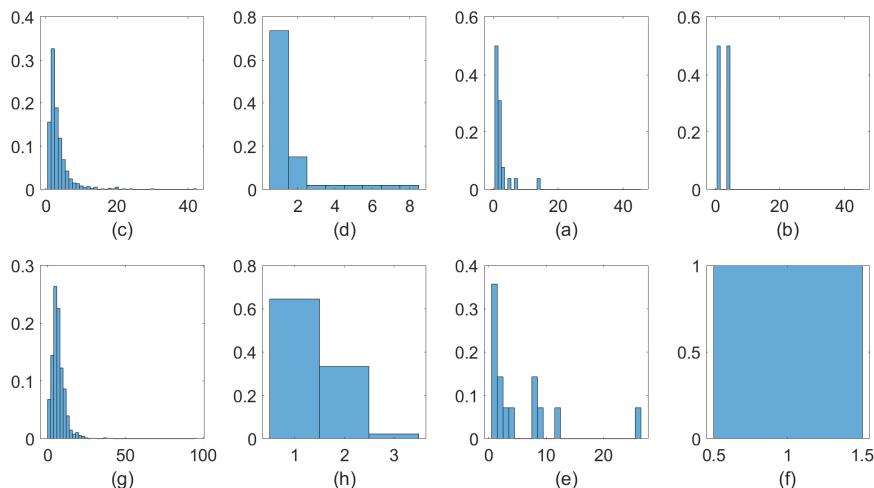


Figura D.1: Distribución del número total de diagnósticos por paciente relacionados con depresión (a), EPOC (b), HTA (c) y DM (d); y códigos ATC por paciente relacionados con depresión (e), EPOC (f), HTA (g) y DM (h) para el CRG-base 3755 en 2012.

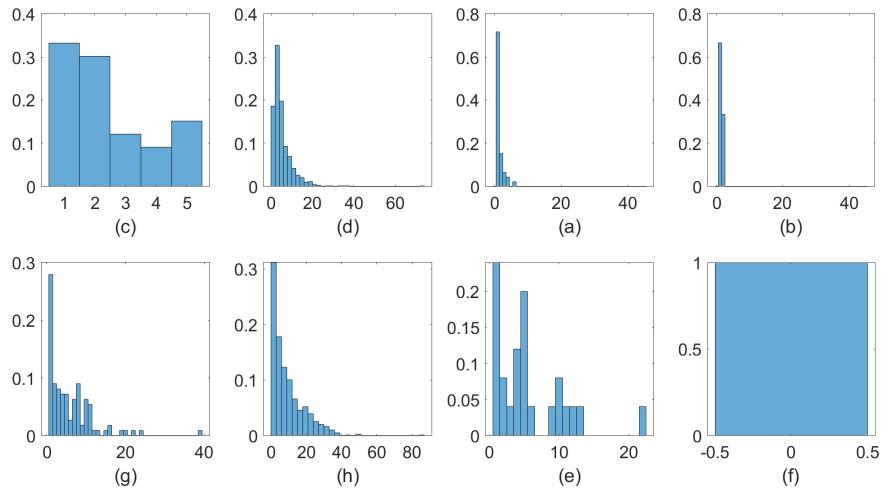


Figura D.2: Distribución del número total de diagnósticos por paciente relacionados con depresión (a), EPOC (b), HTA (c) y DM (d); y códigos ATC por paciente relacionados con depresión (e), EPOC (f), HTA (g) y DM (h) para el CRG-base 5133 en 2012.

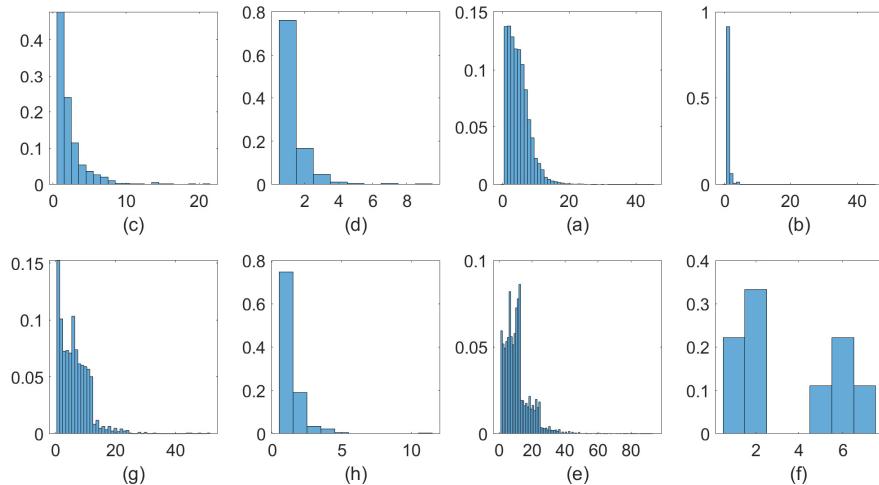


Figura D.3: Distribución del número total de diagnósticos por paciente relacionados con depresión (a), EPOC (b), HTA (c) y DM (d); y códigos ATC por paciente relacionados con depresión (e), EPOC (f), HTA (g) y DM (h) para el CRG-base 5192 en 2012.

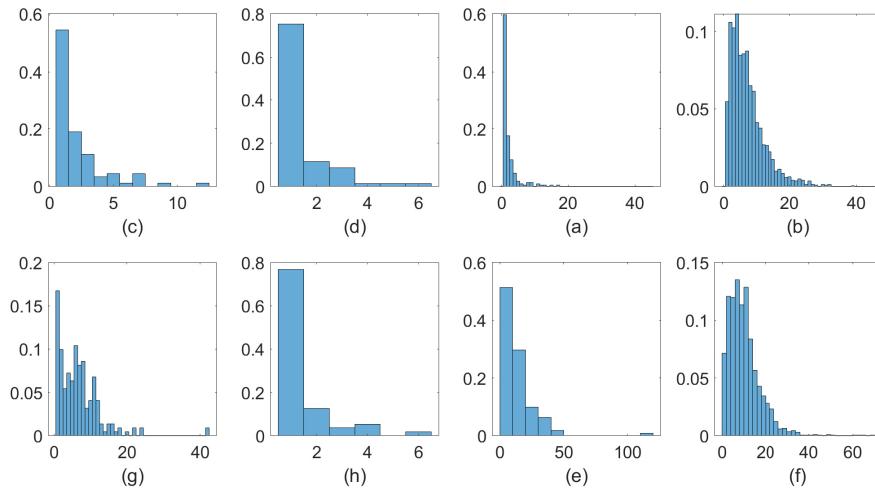


Figura D.4: Distribución del número total de diagnósticos por paciente relacionados con depresión (a), EPOC (b), HTA (c) y DM (d); y códigos ATC por paciente relacionados con depresión (e), EPOC (f), HTA (g) y DM (h) para el CRG-base 5424 en 2012.

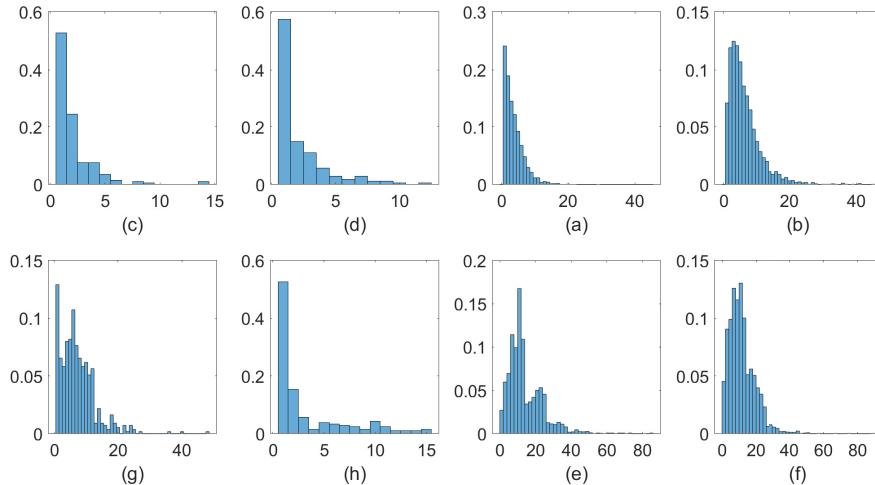


Figura D.5: Distribución del número total de diagnósticos por paciente relacionados con depresión (a), EPOC (b), HTA (c) y DM (d); y códigos ATC por paciente relacionados con depresión (e), EPOC (f), HTA (g) y DM (h) para el CRG-base 6313 en 2012.

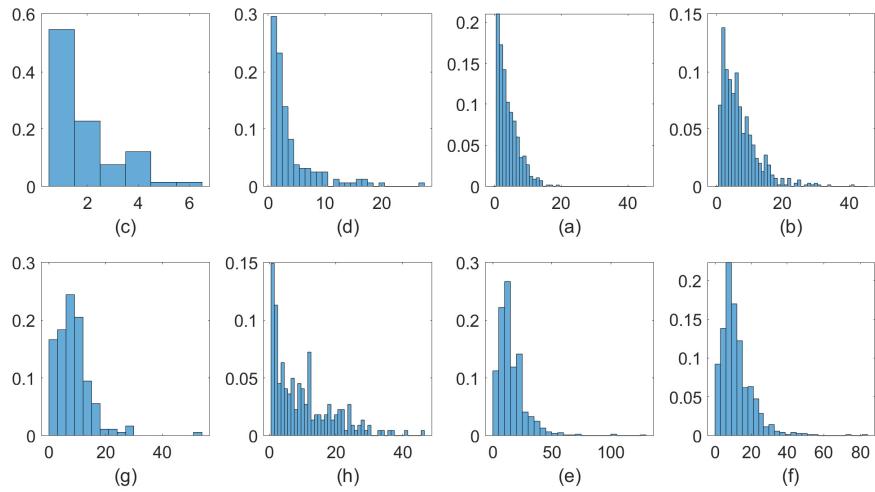


Figura D.6: Distribución del número total de diagnósticos por paciente relacionados con depresión (a), EPOC (b), HTA (c) y DM (d); y códigos ATC por paciente relacionados con depresión (e), EPOC (f), HTA (g) y DM (h) para el CRG-base 7140 en 2012.

Apéndice E

Análisis comparativo de los perfiles poblaciones desde 2010 hasta 2013

A continuación, se muestran los perfiles de presencia y ocurrencia del CRG-base 3755 para los años 2010, 2011, 2012 y 2013. Se observa que, entre 2012 y 2013, los perfiles (tanto de presencia como de ocurrencia) apenas varían; sin embargo, en 2010 y 2011, hay diferencias notables con respecto a 2012 y 2013. Se ha mostrado este CRG-base a modo de ejemplo, pero sucede lo mismo con el resto de CRGs-base considerados. De igual modo, se han elegido los perfiles (como parte del análisis exploratorio) para mostrar estos cambios, pero también se observaba este comportamiento en los resultados del porcentaje con/sin diagnósticos y ATCs relacionados con HTA, DM, depresión y EPOC, y en los resultados de la distribución del número total de diagnósticos y códigos ATC por paciente relacionados con HTA, DM, depresión y EPOC. También se observa sistemáticamente, que los cambios se producen en mayor medida en el análisis de diagnósticos y en menor medida en el análisis de códigos ATC.

Tras comentar estas observaciones con el personal clínico, se concluye que estos cambios podrían estar justificados por modificaciones en la política de recogida de datos, y no tanto por cambios significativos en el estado de salud de los pacientes de cada grupo.

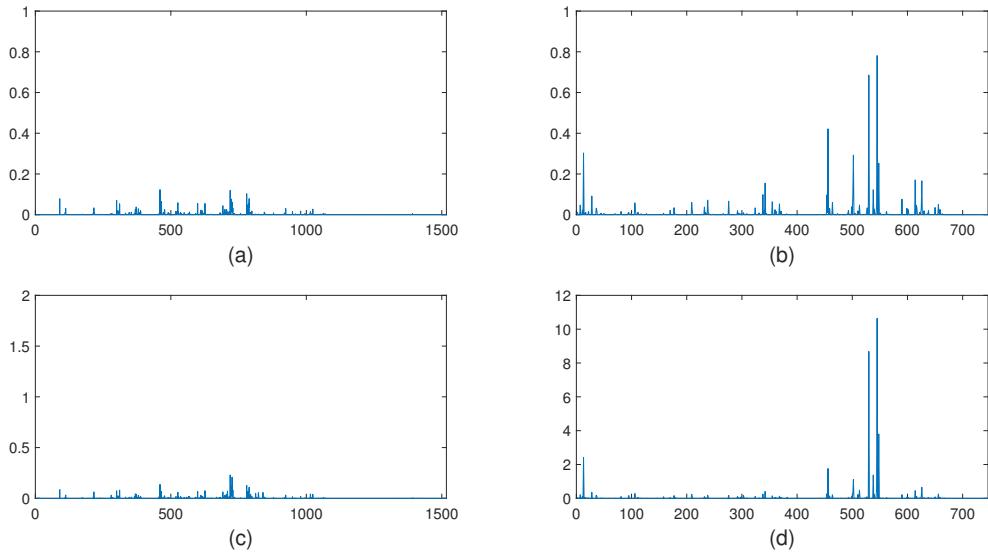


Figura E.1: Perfiles de presencia de diagnósticos (a) y códigos ATC (b), y perfiles de ocurrencia de diagnósticos (c) y códigos ATC (d) del CRG-base 3755 en 2010.

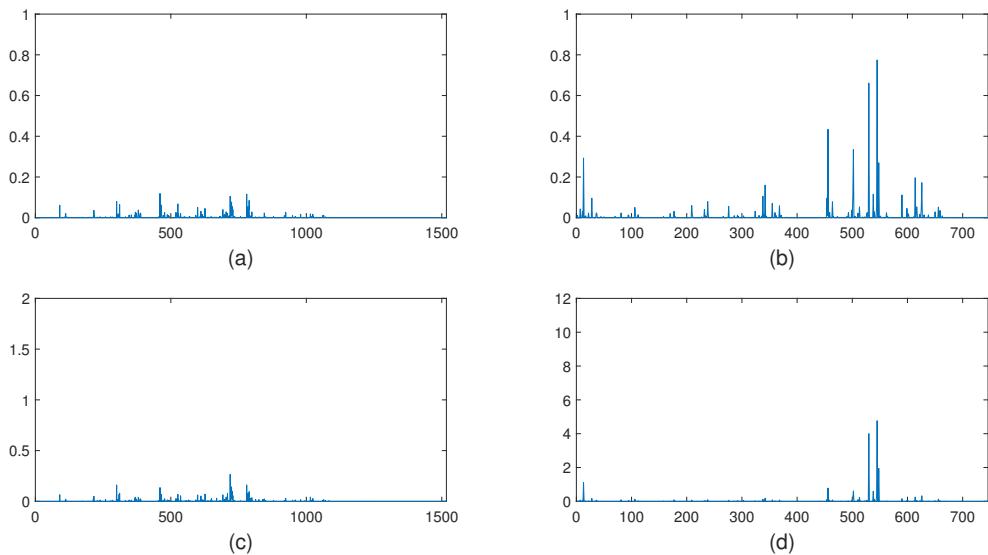


Figura E.2: Perfiles de presencia de diagnósticos (a) y códigos ATC (b), y perfiles de ocurrencia de diagnósticos (c) y códigos ATC (d) del CRG-base 3755 en 2011.

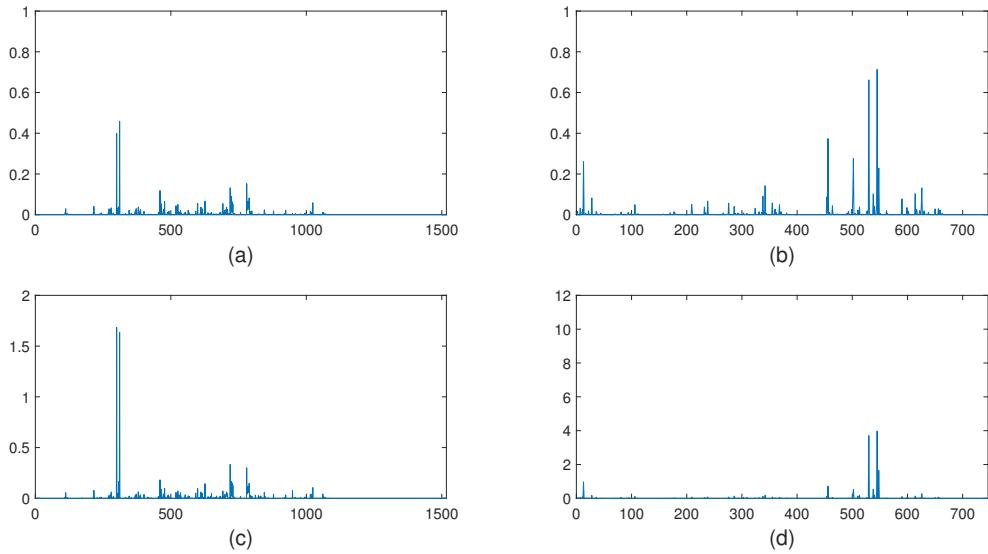


Figura E.3: Perfiles de presencia de diagnósticos (a) y códigos ATC (b), y perfiles de ocurrencia de diagnósticos (c) y códigos ATC (d) del CRG-base 3755 en 2012.

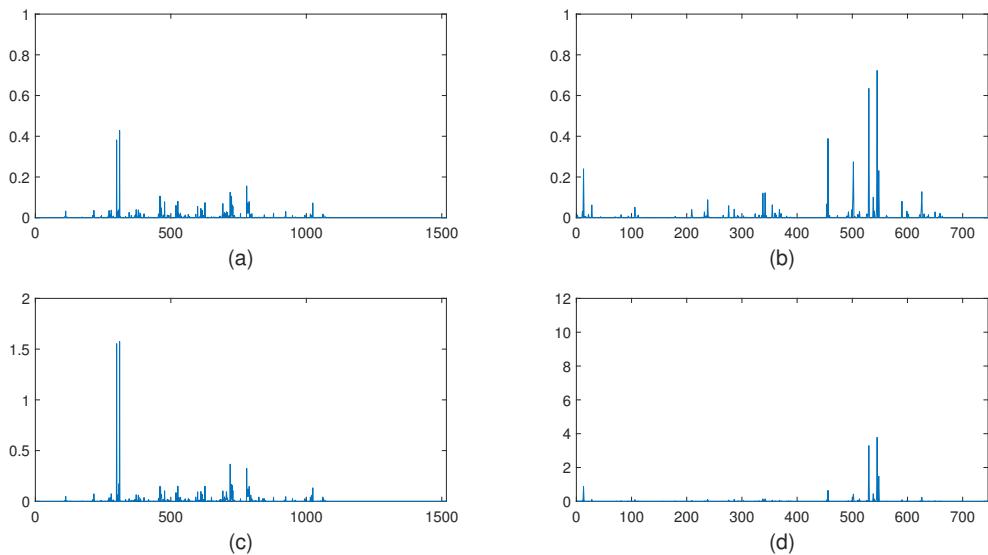


Figura E.4: Perfiles de presencia de diagnósticos (a) y códigos ATC (b), y perfiles de ocurrencia de diagnósticos (c) y códigos ATC (d) del CRG-base 3755 en 2013.

Apéndice F

Códigos seleccionados para analizar la evolución del estado de salud

En este apéndice se adjuntan las listas de códigos diagnósticos y ATCs con diferencias significativas entre el grupo de pacientes sanos (CRG-base 1000) y cada uno de los grupos de pacientes crónicos bajo estudio (CRGs-base: 3755, 5133, 5192, 5424, 6313, 7140). También se muestra la diferencia de proporciones (Δp) media en tanto por uno. Nótese que Δp negativos corresponden a incrementos en la tasa de presencia de un código en el grupo de crónicos frente al de sanos, y viceversa.

Δp	Código	Descripción
-0,46	311	Trastorno depresivo, no clasificado bajo otros conceptos
-0,38	300	Trastornos de ansiedad, disociativos y somatomorfos
0,06	090	Sífilis congénita
-0,70	N06AB	Inhib, selectivos recaptación serotonina (antidepresivos)
-0,63	N05BA	Benzodiazepinas (ansiolíticos)
-0,23	N06AX	Otros antidepresivos
-0,20	A02BC	Inhibidores de la bomba de protones (antiúlcera péptica)
-0,11	M01AE	Derivados del ácido propiónico (AINEs)
-0,10	N05CD	Benzodiazepinas (hipnóticos y sedantes)
-0,09	N02BE	Anilinas (analgésicos)

Cuadro F.1: Códigos seleccionados al comparar el CRG-base 1000 y el CRG-base 3755 en 2012.

Δp	Código	Descripción
-0,31	496	Obstrucción crónica de vías resp. no clasificada bajo otros conceptos
-0,25	493	Asma
-0,21	518	Otras enfermedades pulmonares
-0,17	466	Bronquitis y bronquiolitis agudas
-0,15	490	Bronquitis no especificada como aguda ni como crónica
-0,14	305	Abuso de drogas, sin dependencia
-0,14	272	Trastornos del metabolismo de los lípidos
-0,13	519	Otras enfermedades del aparato respiratorio
-0,09	780	Síntomas generales
-0,08	724	Otros trastornos y trastornos no especificados de la espalda
-0,08	786	Síntomas que implican al ap. respiratorio y otros síntomas torácicos
-0,08	486	Neumonía, organismo no especificado
-0,66	R03BB	Anticolinérgicos (antiasmáticos)
-0,56	R03AC	Agonistas beta-2-adrenergicos selectivos
-0,38	R03AK	Adrenérgicos asociados a otros antiasmáticos
-0,35	N02BE	Anilinas (analgésicos)
-0,29	A02BC	Inhibidores de la bomba de protones (antiúlcera péptica)
-0,26	H02AB	Glucocorticoides
-0,26	R01AX	Otros preparados nasales
-0,26	R05CB	Mucolíticos
-0,19	R03DC	Antagonistas receptores del leukotrieno (antiasmáticos)
-0,19	R03BA	Glucocorticoides (antiasmáticos)
-0,19	J01MA	Fluorquinolonas (antiinfecciosos sistémicos)
-0,16	J01CR	Asociaciones de penicilinas, incluidas inhibid. betalactamasa
-0,15	C05AA	Antihemorroidales con corticoides (vasoprotectores)
-0,14	N05BA	Benzodiazepinas (ansiolíticos)
-0,13	N02BB	Pirazolonas (analgésicos)
-0,12	C10AA	Inhibidores de la hmg-coa reductasa (modificadores lípidos)
-0,11	J01FA	Macrolidos (antiinfecciosos sistémicos)
-0,11	J01CA	Penicilinas de amplio espectro
-0,09	M01AE	Derivados del ácido propionico (AINEs)

-0,08	N02AX	Otros analgésicos opiaceos
-------	-------	----------------------------

Cuadro F.2: Códigos seleccionados al comparar el CRG-base 1000 y el CRG-base 5133 en 2012.

Δp	Código	Descripción
-0,80	401	Hipertensión esencial
-0,23	272	Trastornos del metabolismo de los lípidos
-0,13	724	Otros trastornos y trastornos no especificados de la espalda
-0,10	719	Otros trastornos y trastornos no especificados de articulación
-0,08	715	Osteoartrosis y trastornos conexos
-0,08	790	Hallazgos no específicos en el análisis de sangre
-0,07	726	Tendinitis de las inserciones periféricas y síndromes conexos
-0,07	278	Sobrepeso, obesidad y otros tipos de hiperalimentación
-0,06	300	Trastornos de ansiedad, disociativos y somatomorfos
0,07	463	Amigdalitis aguda
0,06	090	Sífilis congénita
-0,40	C09AA	Inhibidores enzima convertasa angiotensina solos
-0,33	C10AA	Inhibidores de la hmg-coa reductasa (modificadores lípidos)
-0,31	A02BC	Inhibidores de la bomba de protones (antiúlcera péptica)
-0,20	N05BA	Benzodiazepinas (ansiolíticos)
-0,19	N02BE	Anilinas (analgésicos)
-0,17	C08CA	Derivados de dihidropiridina (bloqueantes calcio efect. vasc.)
-0,17	C09BA	Inhib. enzima convertasa angiotensina asociados a diuréticos
-0,15	N02BB	Pirazolonas (analgésicos)
-0,14	C07AB	Betabloqueantes selectivos solos
-0,12	N02AX	Otros analgésicos opiáceos
-0,11	C09CA	Antagonistas angiotensina II solos
-0,08	M02AA	Antiinflamatorios no esteroideos para uso tópico
-0,08	C09DA	Antagonistas angiotensina II asociados a diuréticos
-0,08	C03BA	Sulfonamidas solas (diuréticos bajo techo)
-0,08	S01XA	Otros oftalmológicos
-0,08	N06AB	Inhib. selectivos recaptación serotonina (antidepresivos)

-0,08	M01AB	Derivados del ácido acético y sustanc. relacionadas (AINEs)
-0,07	B01AC	Inhibidores de la agregación plaquetaria
-0,07	H03AA	Hormonas tiroideas

Cuadro F.3: Códigos seleccionados al comparar el CRG-base 1000 y el CRG-base 5192 en 2012.

Δp	Código	Descripción
-0,91	250	Diabetes mellitus
-0,21	272	Trastornos del metabolismo de los lípidos
-0,11	724	Otros trastornos y trastornos no especificados de la espalda
-0,08	401	Hipertensión esencial
-0,07	719	Otros trastornos y trastornos no especificados de articulación
-0,07	790	Hallazgos no específicos en el análisis de sangre
-0,06	366	Catarata
0,07	463	Amigdalitis aguda
0,06	090	Sífilis congénita
-0,53	A10BA	Biguanidas (antidiabéticos orales)
-0,44	C10AA	Inhibidores de la hmg-coa reductasa (modificadores lípidos)
-0,23	A02BC	Inhibidores de la bomba de protones (antiúlcera péptica)
-0,23	A10AE	Insulinas y análogos acción lenta para inyección
-0,18	A10AB	Insulinas y análogos de acción rápida para inyección
-0,14	A10BD	Hipoglucemiantes orales en asociación
-0,13	B01AC	Inhibidores de la agregación plaquetaria
-0,13	N05BA	Benzodiazepinas (ansiolíticos)
-0,11	N02BE	Anilinas (analgésicos)
-0,10	N02BB	Pirazolonas (analgésicos)
-0,09	A10BB	Sulfonilureas (antidiabéticos orales)
-0,08	N02AX	Otros analgésicos opiáceos
-0,07	A10AD	Insulinas y análogos acción intermedia + rápida para inyección
-0,07	M01AB	Derivados del ácido acético y sustanc. relacionadas (AINEs)
-0,07	H03AA	Hormonas tiroideas
-0,07	A10BX	Otros antidiabéticos orales, excluidas insulinas

-0,06	C10AB	Fibratos (modificadores de lípidos)
-------	-------	-------------------------------------

Cuadro F.4: Códigos seleccionados al comparar el CRG-base 1000 y el CRG-base 5424 en 2012.

Δp	Código	Descripción
-0,89	250	Diabetes mellitus
-0,71	401	Hipertensión esencial
-0,27	272	Trastornos del metabolismo de los lípidos
-0,11	724	Otros trastornos y trastornos no especificados de la espalda
-0,10	719	Otros trastornos y trastornos no especificados de articulación
-0,09	715	Osteoartrosis y trastornos conexos
-0,09	780	Síntomas generales
-0,08	366	Catarata
-0,07	278	Sobrepeso, obesidad y otros tipos de hiperalimentación
-0,07	790	Hallazgos no específicos en el análisis de sangre
-0,07	726	Tendinitis de las inserciones periféricas y síndromes conexos
0,08	463	Amigdalitis aguda
0,06	090	Sífilis congénita
-0,67	A10BA	Biguanidas (antidiabéticos orales)
-0,67	C10AA	Inhibidores de la hmg-coa reductasa (modificadores lípidos)
-0,45	C09AA	Inhibidores enzima convertasa angiotensina solos
-0,43	A02BC	Inhibidores de la bomba de protones (antiúlcera péptica)
-0,36	B01AC	Inhibidores de la agregación plaquetaria
-0,23	C08CA	Derivados de dihidropiridina (bloqueantes calcio efecto. vasc)
-0,23	N02BE	Anilinas (analgésicos)
-0,20	A10BD	Hipoglucemiantes orales en asociación
-0,18	N02BB	Pirazolonas (analgésicos)
-0,18	C09BA	Inhib. enzima convertasa angiotensina asociados a diuréticos
-0,17	N05BA	Benzodiazepinas (ansiolíticos)
-0,17	C09CA	Antagonistas angiotensina II solos
-0,15	A10BB	Sulfonilureas (antidiabéticos orales)
-0,14	C09DA	Antagonistas angiotensina II asociados a diuréticos

-0,14	N02AX	Otros analgésicos opiáceos
-0,13	A10AE	Insulinas y análogos acción lenta para inyección
-0,13	C07AB	Betablockantes selectivos solos
-0,11	A10BX	Otros antidiabéticos orales, excluidas insulinas
-0,10	S01XA	Otros oftalmológicos
-0,10	M02AA	Antiinflamatorios no esteroideos para uso tópico
-0,10	C10AB	Fibratos (modificadores de lípidos)
-0,09	N06AB	Inhib. selectivos recaptación serotonina (antidepresivos)
-0,08	C03BA	Sulfonamidas solas (diuréticos bajo techo)
-0,08	H03AA	Hormonas tiroideas
-0,08	A10AD	Insulinas y análogos acción intermedia + rápida para inyección
-0,07	R05CB	Mucolíticos
-0,07	M01AB	Derivados del ácido acético y sustanc. relacionadas (AINEs)

Cuadro F.5: Códigos seleccionados al comparar el CRG-base 1000 y el CRG-base 6313 en 2012.

Δp	Código	Descripción
-0,90	250	Diabetes mellitus
-0,72	401	Hipertensión esencial
-0,26	272	Trastornos del metabolismo de los lípidos
-0,14	715	Osteoartrosis y trastornos conexos
-0,13	780	Síntomas generales
-0,13	724	Otros trastornos y trastornos no especificados de la espalda
-0,12	366	Catarata
-0,11	719	Otros trastornos y trastornos no especificados de articulación
-0,10	174	Neoplasia maligna de la mama femenina
-0,09	599	Otros trastornos de uretra y tracto urinario
-0,09	496	Obstrucción crónica de vías respiratorias, no clasificada bajo otros conceptos
-0,09	V65	Otras personas que solicitan consulta
-0,08	788	Síntomas que afectan al aparato urinario
-0,08	311	Trastorno depresivo, no clasificado bajo otros conceptos

-0,08	278	Sobrepeso, obesidad y otros tipos de hiperalimentación
-0,07	790	Hallazgos no específicos en el análisis de sangre
0,08	463	Amigdalitis aguda
-0,63	C10AA	Inhibidores de la hmg-coa reductasa (modificadores lípidos)
-0,63	A02BC	Inhibidores de la bomba de protones (antiúlcera péptica)
-0,61	A10BA	Biguanidas (antidiabéticos orales)
-0,41	N02BE	Anilinas (analgésicos)
-0,40	C09AA	Inhibidores enzima convertasa angiotensina solos
-0,39	B01AC	Inhibidores de la agregación plaquetaria
-0,33	N02BB	Pirazolonas (analgésicos)
-0,30	N05BA	Benzodiazepinas (ansiolíticos)
-0,24	C08CA	Derivados de dihidropiridina (bloqueantes calcio efecto. vasc)
-0,23	N02AX	Otros analgésicos opiáceos
-0,22	R03BB	Anticolinérgicos (antiasmáticos)
-0,21	C09CA	Antagonistas angiotensina II solos
-0,20	R05CB	Mucolíticos
-0,18	A10AE	Insulinas y análogos acción lenta para inyección
-0,18	A10BD	Hipoglucemiantes orales en asociación
-0,17	C09DA	Antagonistas angiotensina II asociados a diuréticos
-0,16	A03FA	Propulsivos
-0,16	M02AA	Antiinflamatorios no esteroideos para uso tópico
-0,16	J01MA	Fluorquinolonas (antiinfecciosos sistémicos)
-0,16	R03AK	Adrenérgicos asociados a otros antiasmáticos
-0,16	S01XA	Otros oftalmológicos
-0,15	N06AB	Inhib. selectivos recaptación serotonina (antidepresivos)
-0,14	A10BB	Sulfonilureas (antidiabéticos orales)
-0,14	C09BA	Inhib. enzima convertasa angiotensina asociados a diuréticos
-0,13	C03CA	Sulfonamidas solas (diuréticos alto techo)
-0,13	R03AC	Agonistas beta-2-adrenergicos selectivos
-0,13	A10BX	Otros antidiabéticos orales, excluidas insulinas
-0,12	B01AB	Heparinas
-0,11	H03AA	Hormonas tiroideas

-0,11	A12AX	Calcio asociado a otras sustancias
-0,11	J01CR	Asociaciones de penicilinas, incluidas inhibid. betalactamasa
-0,10	C07AB	Betabloqueantes selectivos solos
-0,10	N05AL	Benzamidas (antipsicóticos)
-0,10	A10AD	Insulinas y análogos acción intermedia + rápida para inyección
-0,10	A06AD	Laxantes osmóticos
-0,10	H02AB	Glucocorticoides
-0,09	N05CD	Benzodiazepinas (hipnóticos y sedantes)
-0,09	N03AX	Otros antiepilépticos
-0,09	A10AB	Insulinas y análogos de acción rápida para inyección
-0,09	M01AB	Derivados del ácido acético y sustanc. relacionadas (AINEs)
-0,07	N06AX	Otros antidepresivos
-0,07	C02CA	Bloqueantes alfa-adrenérgicos (antihipertensivos)
-0,07	C10AB	Fibratos (modificadores de lípidos)
-0,07	C03BA	Sulfonamidas solas (diuréticos bajo techo)

Cuadro F.6: Códigos seleccionados al comparar el CRG-base 1000 y el CRG-base 7140 en 2012.

Apéndice G

Códigos seleccionados para analizar la evolución temporal

En este apéndice se muestran los códigos con diferencias significativas en su tasa de presencia entre dos grupos de pacientes clasificados en el mismo CRG-base en 2012 y 2015, caracterizando así los estados de salud en un intervalo temporal.

Δp	Código	Descripción
CRG-base 3755		
0,11	R05CB	Mucolíticos
0,07	N05BA	Benzodiazepinas (ansiolíticos)
0,05	A02BC	Inhibidores de la bomba de protones (antiúlcera péptica)
0,05	M02AA	Antiinflamatorios no esteroideos para uso tópico
0,05	M01AB	Derivados del ácido acético y sustanc. relacionadas (AINEs)
0,08	311	Trastorno depresivo, no clasificado bajo otros conceptos
CRG-base 5133		
-0,06	N05BA	Benzodiazepinas (ansiolíticos)
-0,06	J01MA	Fluorquinolonas (antiinfecciosos sistémicos)
-0,05	N02BB	Pirazolonas (analgésicos)
-0,05	R06AX	Otros antihistaminicos de uso sistémico
-0,05	599	Otros trastornos de uretra y tracto urinario
0,30	R05CB	Mucolíticos
0,08	C05AA	Antihemorroidales con corticoides (vasoprotectores)

0,07	R01AX	Otros preparados nasales
0,06	M02AA	Antiinflamatorios no esteroideos para uso tópico
0,06	R03DC	Antagonistas receptores del leukotrieno (antiasmáticos)
0,06	S01XA	Otros oftalmológicos
0,05	H02AB	Glucocorticoides
0,05	R03BA	Glucocorticoides (antiasmáticos)
0,05	R05DA	Alcaloides del opio y derivados (antitusivos)
0,08	518	Other diseases of lung
0,07	490	Bronquitis no especificada como aguda ni como crónica
CRG-base 5192		
0,12	R05CB	Mucolíticos
0,10	M02AA	Antiinflamatorios no esteroideos para uso tópico
0,09	S01XA	Otros oftalmológicos
0,06	M01AE	Derivados del ácido propiónico (AINEs)
0,06	M01AB	Derivados del ácido acético y sustanc. relacionadas (AINEs)
CRG-base 6313		
-0,06	A10BD	Hipoglucemiantes orales en asociación
-0,05	A10AE	Insulinas y análogos acción lenta para inyección
-0,08	780	Síntomas generales
-0,05	526	Enfermedades de los maxilares
0,10	R05CB	Mucolíticos
0,07	M01AE	Derivados del ácido propiónico (AINEs)
0,07	M02AA	Antiinflamatorios no esteroideos para uso tópico
0,06	S01XA	Otros oftalmológicos
0,06	A10BA	Biguanidas (antidiabéticos orales)
0,05	M01AB	Derivados del ácido acético y sustanc. relacionadas (AINEs)
CRG-base 6313		
-0,06	A10BD	Hipoglucemiantes orales en asociación
-0,05	A10AE	Insulinas y análogos acción lenta para inyección
0,13	R05CB	Mucolíticos
0,12	M02AA	Antiinflamatorios no esteroideos para uso tópico
0,11	S01XA	Otros oftalmológicos

0,08	M01AE	Derivados del ácido propiónico (AINEs)
0,05	401	Hipertensión esencial
CRG-base 7140		
-0,09	J01CR	Asociaciones de penicilinas, incluidas inhibid. betalactamasa
-0,07	A10AB	Insulinas y análogos de acción rápida para inyección
-0,06	C09AA	Inhibidores enzima convertasa angiotensina solos
-0,06	J01XX	Otros antibacterianos (antiinfecciosos sistémicos)
-0,05	P02DX	Otros cestocidas
-0,08	780	Síntomas generales
-0,08	V58	Admisión para otros cuidados posteriores y los no especificados
-0,07	519	Otras enfermedades del aparato respiratorio
-0,05	790	Hallazgos no específicos en el análisis de sangre
0,26	R05CB	Mucolíticos
0,18	M02AA	Antiinflamatorios no esteroideos para uso tópico
0,15	S01XA	Otros oftalmológicos
0,11	M01AE	Derivados del ácido propiónico (AINEs)
0,08	C09CA	Antagonistas angiotensina II solos
0,08	A02AD	Asociaciones y compl. de comp. de aluminio, calcio y magnesio
0,07	A03FA	Propulsivos
0,07	M01AB	Derivados del ácido acético y sustanc. relacionadas (AINEs)
0,06	C05CA	Bioflavonoides (vasoprotectores)
0,06	B01AC	Inhibidores de la agregación plaquetaria
0,06	R05DA	Alcaloides del opio y derivados (antitusivos)
0,06	A06AD	Laxantes osmóticos
0,05	C09DA	Antagonistas angiotensina II asociados a diuréticos

Cuadro G.1: Códigos seleccionados al comparar cada CRG-base en 2012 y el mismo CRG-base en 2015.

Apéndice H

Implementación del ACM

A continuación, se describe el procedimiento seguido para la implementación del ACM:

- Cálculo de la matriz de correspondencias \mathbf{P} , construida a través de la matriz de datos \mathbf{N} y la suma total de elementos n de \mathbf{N} . Nótese que, en el caso del ACM, \mathbf{N} es la matriz de Burt, I es el número de filas y J el número de columnas:

$$\mathbf{P} = \frac{\mathbf{N}}{n} ; n = \sum_{i=1}^I \sum_{j=1}^J n_{ij} \quad (\text{H.1})$$

- Cálculo de las masas de filas r_i y columnas c_j :

$$r_i = \sum_{j=1}^J p_{ij} ; c_j = \sum_{i=1}^I p_{ij} \quad (\text{H.2})$$

Donde $\mathbf{r} = \mathbf{P}\mathbf{1}$, $\mathbf{c} = \mathbf{P}^T\mathbf{1}$, y $\mathbf{1}$ es un vector de unos con la longitud adecuada.

- Cálculo de las matrices diagonales de masas de filas \mathbf{D}_r y columnas \mathbf{D}_c , donde $\mathbf{r}=\{r_i\}$ y $\mathbf{c}=\{c_j\}$:

$$\mathbf{D}_r = \text{diag}(\mathbf{r}) ; \mathbf{D}_c = \text{diag}(\mathbf{c}) \quad (\text{H.3})$$

- Subespacio \mathbf{S} :

$$\mathbf{S} = \mathbf{D}_r^{-\frac{1}{2}} (\mathbf{P} - \mathbf{rc}^T) \mathbf{D}_c^{-\frac{1}{2}} \quad (\text{H.4})$$

- DVS, siendo \mathbf{D}_α es la matriz diagonal de valores singulares:

$$\mathbf{S} = \mathbf{U}\mathbf{D}_\alpha\mathbf{V}^T \text{ donde: } \mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = 1 \quad (\text{H.5})$$

- Cálculo de las coordenadas estándares de las filas ϕ y columnas Γ :

$$\phi = \mathbf{D}_r^{-\frac{1}{2}}\mathbf{U}; \Gamma = \mathbf{D}_c^{-\frac{1}{2}}\mathbf{V} \quad (\text{H.6})$$

- Cálculo de las coordenadas principales de las filas \mathbf{F} y columnas \mathbf{G} :

$$\mathbf{F} = \mathbf{D}_r^{-\frac{1}{2}}\mathbf{U}\mathbf{D}_\alpha = \phi\mathbf{D}_\alpha; \mathbf{G} = \mathbf{D}_c^{-\frac{1}{2}}\mathbf{V}\mathbf{D}_\alpha = \Gamma\mathbf{D}_\alpha \quad (\text{H.7})$$

- Cálculo de la inercia λ_k de cada autovalor k -ésimo:

$$\lambda_k = \alpha_k^2, k = 1, 2, \dots, K, \text{ con } K = \min\{I-1, J-1\} \quad (\text{H.8})$$

- Inercia total:

$$\text{Inercia total} = \sum_{k=1}^K \alpha_k^2 = \sum_{k=1}^K \lambda_k \quad (\text{H.9})$$

- Proporción de contribución de cada autovalor k a la inercia total:

$$\text{Prop. contrib. } \lambda_k = \frac{\lambda_k}{\sum_{k=1}^K \lambda_k} \quad (\text{H.10})$$

Apéndice I

Resultados del ACM para el análisis de la evolución temporal

A continuación se adjuntan los resultados del ACM al considerar la evolución temporal de cada estado de salud. En las siguientes figuras se muestra en el panel izquierdo las asociaciones entre características seleccionadas y en el panel derecho, la proyección de los individuos de cada grupo en 2012 y en 2015.

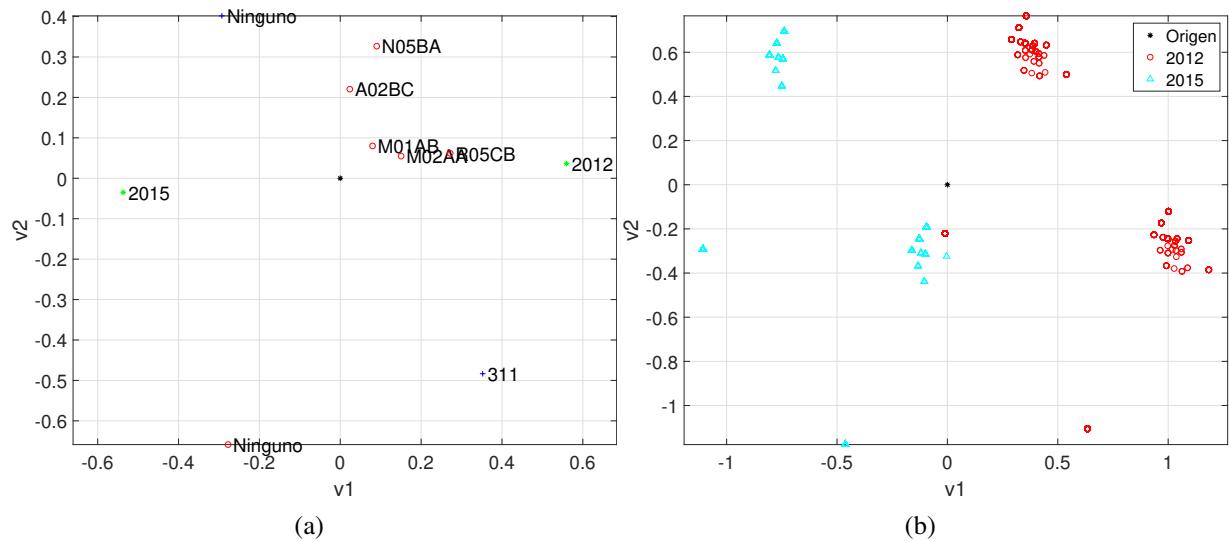


Figura I.1: Evolución temporal del CRG-base 3755: asociaciones entre categorías (a) y entre pacientes (b).

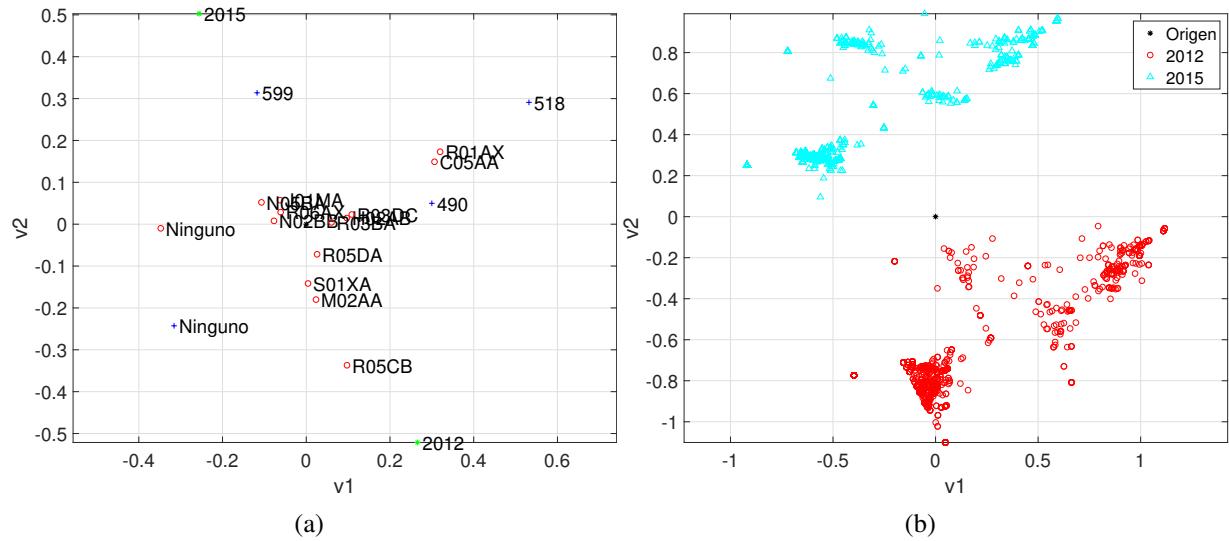


Figura I.2: Evolución temporal del CRG-base 5133: asociaciones entre categorías (a) y entre pacientes (b).

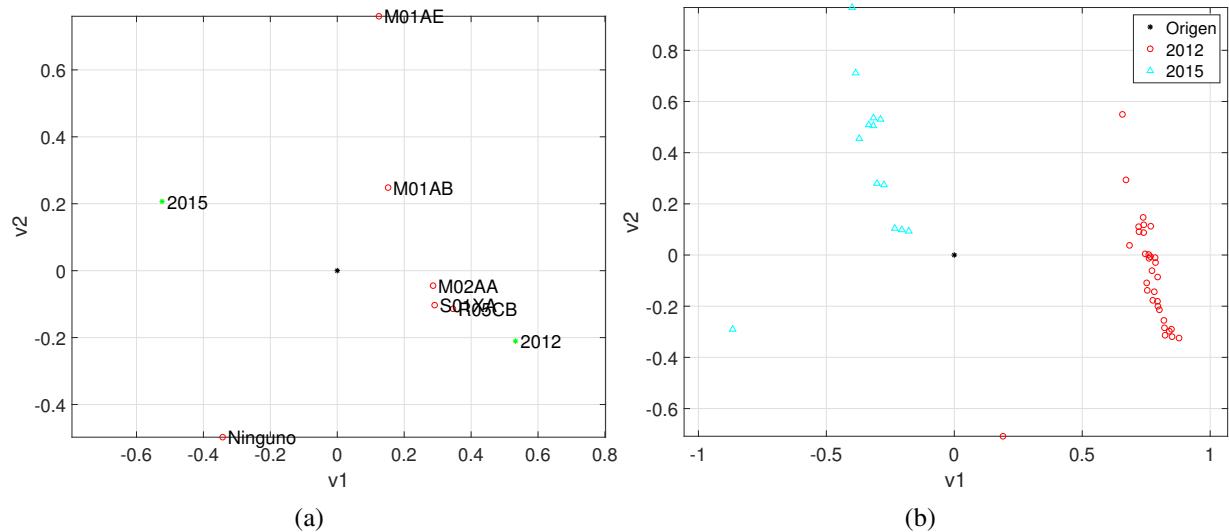


Figura I.3: Evolución temporal del CRG-base 5192: asociaciones entre categorías (a) y entre pacientes (b).

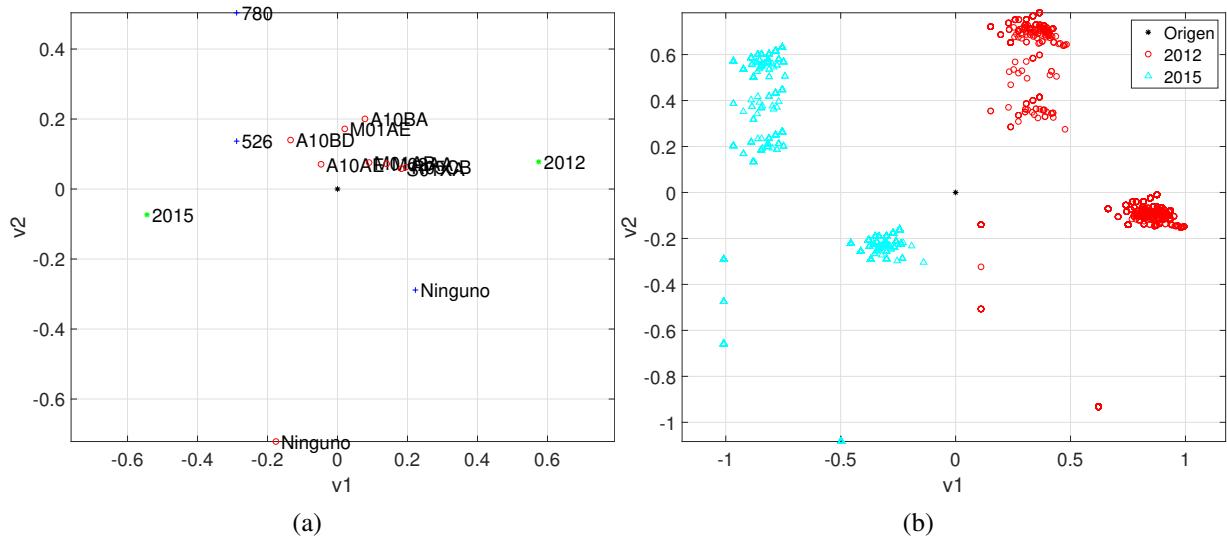


Figura I.4: Evolución temporal del CRG-base 5424: asociaciones entre categorías (a) y entre pacientes (b).

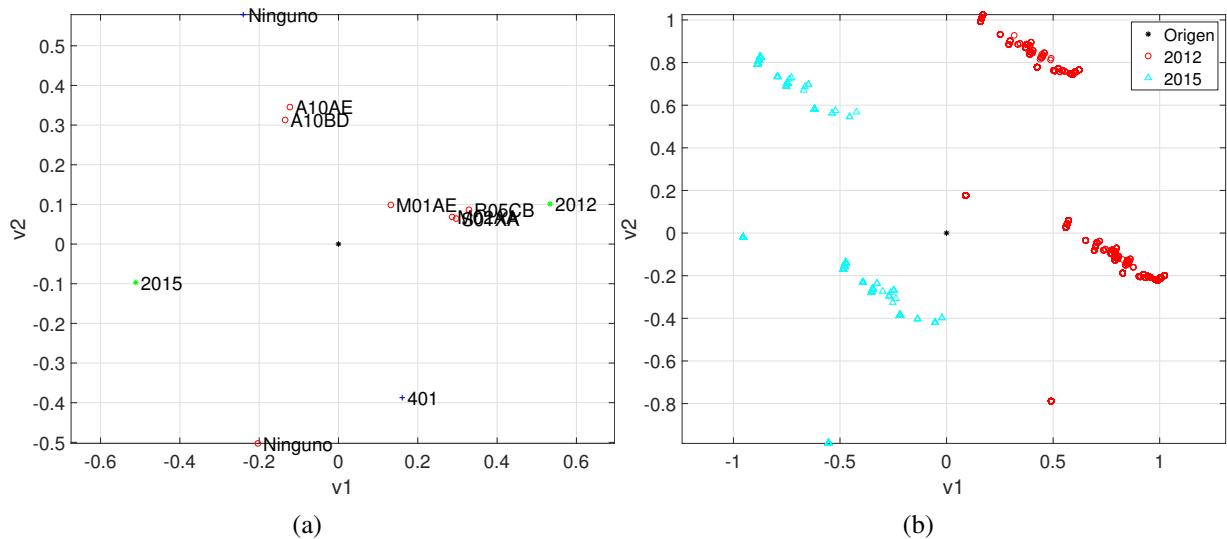


Figura I.5: Evolución temporal del CRG-base 6313: asociaciones entre categorías (a) y entre pacientes (b).

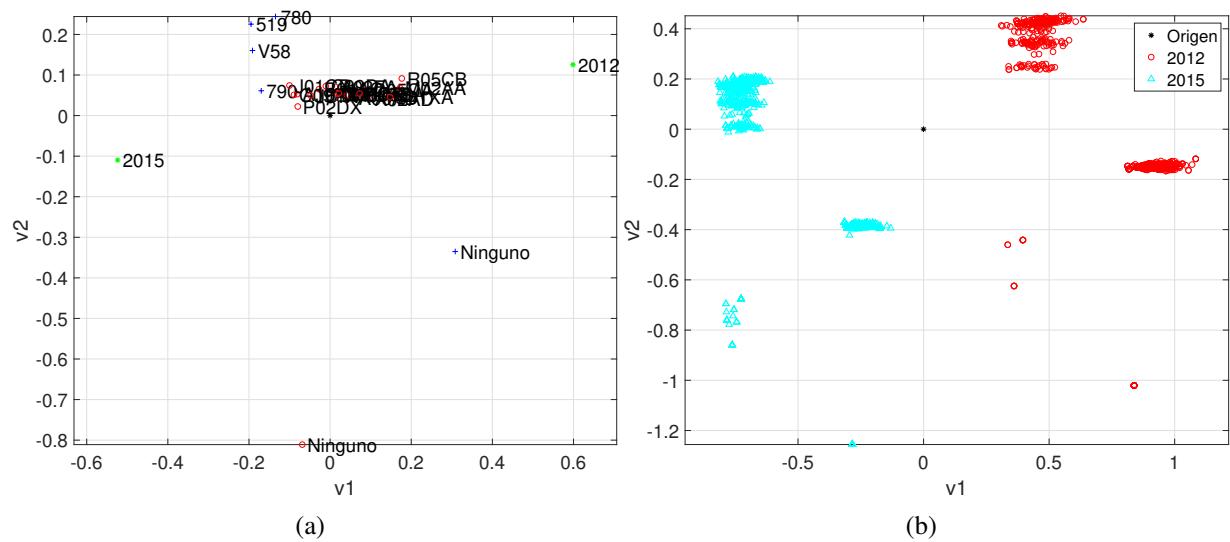


Figura I.6: Evolución temporal del CRG-base 7140: asociaciones entre categorías (a) y entre pacientes (b).

Bibliografía

- [1] Organización Mundial de la Salud. Constitución, 2019. Disponible en: <https://www.who.int/es/about/who-we-are/constitution> [Último acceso: 15 de octubre de 2019].
- [2] Ley 14/1986, de 25 de abril, General de Sanidad, 2019. Disponible en: <https://www.boe.es/buscar/pdf/1986/BOE-A-1986-10499-consolidado.pdf> [Último acceso: 15 de octubre de 2019].
- [3] El sistema nacional de salud, 2019. Disponible en: https://rodas5.us.es/file/f8facbca-7d3f-978d-9c76-ec85884f8abf/1/tema11_SCORM.zip/page_10.htm [Último acceso: 21 de enero de 2020].
- [4] Sistema nacional de salud, 2019. Disponible en: https://guiasjuridicas.wolterskluwer.es/Content/Documento.aspx?params=H4sIAAAAAAAEAMtMSbF1jTAAAUMTQ0sLtbLUouLM_DxbIwMDCwNzA7BAZlqlS35ySGVBqm1aYk5xKgC_ymVZNQAAAA==WKE [Último acceso: 15 de octubre de 2019].
- [5] Consumo y Bienestar Social Ministerio de Sanidad. Prestaciones del Sistema Nacional de Salud. Disponible en: <http://www.mscbs.gob.es/organizacion/sns/docs/prestaciones08.pdf> [Último acceso: 30 de octubre de 2019].
- [6] SEMERGEN. La cronicidad, mucho más que un problema de salud: un nuevo modelo de atención sanitaria, 2018. Disponible en: <https://www.semegen.es/?seccion=noticias&subSeccion=detalleNoticia&idN=188> [Último acceso: 21 de enero de 2020].
- [7] Consumo y Bienestar Social Ministerio de Sanidad. Situación de salud. Informe Anual del Sistema Nacional de Salud 2018, 2019. Disponible en:

- https://www.mscbs.gob.es/estadEstudios/estadisticas/sisInfSanSNS/tablasEstadisticas/InfAnualSNS2018/Cap.2_SituacionSalud.pdf [Último acceso: 21 de enero de 2020].
- [8] Instituto Nacional de Estadística. Tasa Bruta de Mortalidad, 2019. Disponible en: <https://www.ine.es/jaxiT3/Datos.htm?t=1411> [Último acceso: 10 de febrero de 2020].
- [9] T. Hesterberg. Bootstrap. *WIREs Computational Statistics*, 3(6):497–526, 2011.
- [10] Observatorio de Prácticas Innovadoras en el Manejo de Enfermedades Crónicas Complejas. Disponible en: <https://www.opimec.org/glosario/chronic-diseases/> [Último acceso: 8 de noviembre de 2019].
- [11] Consumo y Bienestar Social Ministerio de Sanidad. Hábitos de vida. Informe Anual del Sistema Nacional de Salud 2018, 2019. Disponible en: https://www.mscbs.gob.es/estadEstudios/estadisticas/sisInfSanSNS/tablasEstadisticas/InfAnualSNS2018/Cap.3_HabitosVida.pdf [Último acceso: 24 de enero de 2020].
- [12] Msdsalud.es, 2019. Disponible en: https://www.msdsalud.es/Assets/docs/dossier_msd/el-paciente-cronico.pdf [Último acceso: 8 de noviembre de 2019].
- [13] L. Palomo, C. Rubio, and J. Gérvás. La comorbilidad en atención primaria. *Gaceta Sanitaria*, 20:182–191, 2006.
- [14] Institute for Health Metrics and Evaluation. Disponible en: <http://www.healthdata.org/spain?language=149> [Último acceso: 13 de noviembre de 2019].
- [15] Cuidateplus.com. Hipertensión Arterial. Disponible en: <https://cuidateplus.marca.com/enfermedades/enfermedades-vasculares-y-del-corazon/hipertension-arterial.html> [Último acceso: 9 de noviembre de 2019].
- [16] Fundación Española del Corazón. Hipertensión. Disponible en: <https://fundaciondelcorazon.com/prevencion/riesgo-cardiovascular/hipertension-tension-alta.html> [Último acceso: 9 de noviembre de 2019].
- [17] Fundación Española del Corazón. Diabetes. Disponible en: <https://fundaciondelcorazon.com/prevencion/riesgo-cardiovascular/diabetes.html> [Último acceso: 13 de noviembre de 2019].

- [18] Mayo Clinic.org. Diabetes. Disponible en: <https://www.mayoclinic.org/es-es/diseases-conditions/diabetes/symptoms-causes/syc-20371444> [Último acceso: 13 de noviembre de 2019].
- [19] Medlineplus.gov. Depresión. Disponible en: <https://medlineplus.gov/spanish/ency/article/003213.htm> [Último acceso: 13 de noviembre de 2019].
- [20] National Institute of Mental Health (NIMH). Depresión. Disponible en: <https://www.nimh.nih.gov/health/publications/espanol/depresion-sp/index.shtml> [Último acceso: 13 de noviembre de 2019].
- [21] Cuidateplus.com. Depresión. Disponible en: <https://cuidateplus.marca.com/enfermedades/psiquiatricas/depresion.html> [Último acceso: 13 de noviembre de 2019].
- [22] Mayo Clinic.org. Depresión (trastorno depresivo mayor). Disponible en: <https://www.mayoclinic.org/es-es/diseases-conditions/depression/diagnosis-treatment/drc-20356013> [Último acceso: 13 de noviembre de 2019].
- [23] J.I. Castillo-Quan, D.J. Barrera-Buenfil, J.M. Pérez-Osorio, and F.J. Álvarez Cervera. Depresión y diabetes: de la epidemiología a la neurobiología. *Rev Neurol*, 51:347–359, 2010.
- [24] National Institute of Mental Health (NIMH). Las enfermedades crónicas y la salud mental: Cómo reconocer y tratar la depresión. Disponible en: <https://www.nimh.nih.gov/health/publications/espanol/las-enfermedades-cronicas-y-la-salud-mental-como-reconocer-y-tratar-la-depresion/index.shtml> [Último acceso: 13 de noviembre de 2019].
- [25] Organización Mundial de la Salud. Enfermedad pulmonar obstructiva crónica. Disponible en: [https://www.who.int/es/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-\(copd\)](https://www.who.int/es/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd)) [Último acceso: 13 de noviembre de 2019].
- [26] Mayo Clinic.org. Enfermedad pulmonar obstructiva crónica. Disponible en: <https://www.mayoclinic.org/es-es/diseases-conditions/copd/diagnosis-treatment/drc-20353685> [Último acceso: 13 de noviembre de 2019].
- [27] A. Albarracín Serra. La producción sanitaria: codificación de enfermedades y procedimientos, Madrid: Escuela Nacional de Sanidad, 2012. Dispo-

- nible en: http://e-spacio.uned.es/fez/eserv/bibliuned%3A500676/n8.1_La_producci_n_sanitaria.pdf [Último acceso: 27 de enero de 2019].
- [28] Multimedia.3m.com, 2019. Disponible en: <https://multimedia.3m.com/mws/media/7658330/3m-crgs-measuring-risk-managing-care-white-paper.pdf> [Último acceso: 15 de noviembre de 2019].
- [29] Consumo y Bienestar Social Ministerio de Sanidad. Clasificación Internacional de Enfermedades 9^a Revisión Modificación Clínica. Disponible en: https://www.mscbs.gob.es/estadEstudios/estadisticas/docs/CIE9MC_2014_def_accesible.pdf [Último acceso: 16 de noviembre de 2019].
- [30] Consumo y Bienestar Social Ministerio de Sanidad. eCIE9MC. Disponible en: https://eciemaps.mscbs.gob.es/ecieMaps/browser/index_9_mc.html [Último acceso: 17 de noviembre de 2019].
- [31] M.V. Saladrigas. El sistema de clasificación atc de sustancias farmacéuticas para uso humano. *Panace@*, V:58–60, marzo 2004. Disponible en: https://www.tremedica.org/wp-content/uploads/n15_tribuna-Saladrigas.pdf [Último acceso: 28 de enero de 2019].
- [32] Sanidad de Castilla y León. Clasificación de medicamentos. Disponible en: <http://antia.usal.es/TOLes/MM/MOD2%20Final%20v2/Presen2/b31bltt.htm> [Último acceso: 17 de noviembre de 2019].
- [33] G.C. Canavos. *Probabilidad y estadística: aplicaciones y métodos*. McGraw-Hill Interamericana de España, 1992.
- [34] M.H. DeGroot and M.J. Schervish. *Probability and Statistics*. Pearson Education, 2012.
- [35] M. Greenacre. *La práctica del análisis de correspondencias*. Bilbao: Fundación BBVA, 2008.
- [36] S. de la Fuente Fernández. Análisis correspondencias simples y múltiples, 2019. Disponible en: <http://www.estadistica.net/ECONOMETRIA/REDUCIR-DIMENSION/CORRESPONDENCIAS/correspondencias.pdf> [Último acceso: 20 de junio de 2019].
- [37] F. Husson and J. Josse. *The Visualization and Verbalization of Data. Chapter: Multiple Correspondence Analysis*. CRC/PRESS, 2014.