



Este obra está bajo una licencia de Creative Commons

Reconocimiento-CompartirIgual 3.0 España.

http://creativecommons.org/licenses/by-sa/3.0/es/deed.es_ES

Antonio Eleazar Serrano-López : *Análisis de producción y consumo de información en Enfermedades Raras con causa genética identificada (2000-2009)*, Tesis doctoral , Septiembre, 2012

ISBN: 978-84-695-5437-1



Universidad Carlos III de Madrid

Tesis doctoral

*Análisis de producción y consumo de información en
Enfermedades Raras con causa genética identificada
(2000-2009)*

Autor:

Antonio Eleazar Serrano-López

Directora:

Carmen Martín-Moreno

Departamento de Biblioteconomía y Documentación
Facultad de Humanidades, Comunicación y Documentación

Getafe, Septiembre, 2012

Tesis doctoral

Análisis de producción y consumo de información en
Enfermedades Raras con causa genética identificada
(2000-2009)

Autor: Antonio Eleazar Serrano-López

Directora: Carmen Martín-Moreno

Firma del tribunal calificador:

Firma:

Presidente:

Vocal:

Secretario/a:

Calificación:

Getafe, de de

*A los cuatro pilares de mi vida,
sin vosotros no hubiera sido posible.*

Agradecimientos

Nature is nowhere accustomed more openly to display her secret mysteries than in cases where she shows traces of her workings apart from the beaten path; nor is there any better way to advance the proper practice of medicine than to give our minds to the discovery of the usual law of nature, by the careful investigation of cases of rarer forms of disease. — William Harvey.

Carta a John Vlackveld (24 de abril de 1657).

Durante el trabajo de elaborar una tesis doctoral uno espera encontrar escollos, problemas y ser capaz de dar con las soluciones a los mismos. Esas soluciones vienen en muchos casos de manos amigas, que se vuelven cada vez más y más importantes y acaban conformando una segunda casa y una segunda familia.

En mi caso, en esta segunda familia, ha sido Carmen Martín, mi directora, la que se ha comportado como una segunda madre para mí. Gracias a ella he aprendido la importancia de la corrección metodológica, la rigurosidad y el trabajo bien hecho, además del cariño y de las palabras de ánimo que siempre ha tenido y tiene para mí y por supuesto, por servir de guía en este camino, que emprendimos juntos hace tres años, y también a Juanma, por sus consejos, que han sido siempre bienvenidos.

Durante este camino, el Laboratorio de Estudios Métricos de Información ha sido para mí esa segunda familia, que me acogió desde el primer día como uno más,

prestándome su apoyo, su experiencia y los medios técnicos, materiales y sobe todo humanos, que han hecho posible esta investigación a todos, pero en especial a Elias Sanz debo agradecerle la confianza depositada en mi durante estos años, el apoyo prestado y las oportunidades, tanto personales como profesionales, que me ha proporcionado y por despertar en mi el interés por los estudios métricos, que me ha llevado hasta aquí.

También debo agradecerle a María Luisa Lascurain, por su apoyo, su amistad y por su ayuda a la hora de elegir una batería de indicadores adecuada para el objeto de estudio, a Carlos Zorita, por su constante preocupación por el desarrollo de *la bicha* y por su interés y sus ideas respecto al sistema de normalización y los indicadores multidimensionales con los que tanto tiempo le he robado. Además, en esta, mi segunda familia, merece una especial atención mi amigo y compañero Preiddy Efraín, que ha sido para mi un verdadero mentor en el desarrollo técnico de este proyecto, pues suyos son los cimientos de la metodología que yo solo he complementado.

En cuanto a la familia institucional, a la Universidad Carlos III y especialmente al Departamento de Biblioteconomía y Documentación le debo la oportunidad que me ha dado de participar de él, de formarme dentro y fuera de él, y de poder dedicarle mucho más tiempo a la investigación y a estas páginas, así como al colectivo de becarios, ayudantes y asociados, por darme la oportunidad de actuar como su representante en el Consejo de Departamento. Entre los compañeros de la Universidad Carlos III y a riesgo de dejarme algún nombre, me gustaría reconocerle su apoyo, consejos, palabras de ánimo y constante interés por *la tesis* a Rosa San Segundo, Jose Antonio Moreira, Daniela De Filippo, David Rodriguez, Pilar Azcárate, Luis Moreno, Mónica Villegas, Susana Sawas y Pilar Alonso.

Para terminar con mi segunda familia, debo mostrarle mi más sincero agradecimiento a Dani, porque desde que comenzamos juntos esta andadura han pasado diez años y siempre me ha brindado su amabilidad, su apoyo y su amistad. He-

mos llegado juntos hasta aquí y espero que nuestros caminos nunca discurren demasiado separados.

Pero además, en este camino me he encontrado con muchos amigos que han conformado *mi pequeña ONU* y que han sido un gran apoyo, especialmente cuando me he encontrado fuera de casa. Sobre todo tengo que dar las gracias por su hospitalidad a Peter, Irene, Birger, Christina, Toine y Maria Manuel, amigos todos y compañeros magníficos.

Sin embargo, no todo han sido buenos momentos en este camino y es que el tiempo que le he dedicado a este proyecto se lo he robado a mi verdadera familia y amigos, a los que debo agradecerles la comprensión por ese tiempo que he dejado de dedicarles. Prometo recompensároslo, sobre todo a mi madre, a mi padre y a mi Tata, que tanto me han enseñado y aguantado. No sabéis cuanto he echado de menos ese *ten cuidado*, cuando salía de casa y en realidad estaba tan lejos de ella. Muchas gracias por ser como sois y por aguantar a este *rancio* que tantos disgustos os ha dado. Por último, aunque nunca será la última: Muchas gracias Aroa, por sufrir mis ausencias, por haber encontrado el valor para recorrer media Europa y por haber aguantado a esta amante insaciable, que tanto tiempo te ha robado, y a la que pongo fin con estas letras. Hasta aquí es mío, el resto de estas páginas os las debo a vosotros cuatro.

A riesgo de haberme dejado a alguien, mi agradecimiento, de todo corazón, es para todos vosotros.

Abbreviated abstract

Rare diseases are those that have a very low prevalence, lower than 5 cases per 10.000 inhabitants. Our goal is to know the patterns of the research on those diseases, how the flow of the scientific activity is, the use of the different types of information by the researches and the impact that this research generates.

To accomplish this objective, we analyze, for the period 2000-2009, the scientific production of papers by the research community, over 252 rare diseases with an identified genetic cause. The analysis of this production is doing through bibliometric indicators both unidimensional and multidimensional. These indicators are divided and analyzed by several levels of aggregation, from the most general to the most particular, starting by the general characterization of the research and descending to the countries, subjects, sources, institutions and authors analysis.

In order to obtain the bibliometric indicators, we have develop a methodology for the automatic normalization of some bibliographic fields, such as institutions and authors, and to assign MeSH terms to the records obtained from the Web of Science.

These results reveal some interesting patterns on the publishing on rare diseases, with a very high influence of the Genetics area, and showing an inequality between the most producer countries and the countries with a lower capacity of research.

But sometimes, the research of these, less productive countries, can obtain a high impact over the scientific community.

In addition, we have discovered a strong relationship between some social variables like health expenditure, and patient organizations, over the scientific production about this disorders. And we have also found some diseases, like Cystic fibrosis, being the object of more papers than we expected considering their low prevalence.

About the centers conducting this research, we have identified the hospital as the main institutional type, although the national research centers are which, individually, produce a greater number of documents, regardless of they do not receive an equivalent volume of citations.

Finally, we observe some patterns on the collaboration between countries, and the preferences of their researchers over the journals to publish, with some countries publishing in all kind of journals and some other with a clear preference to publish in national journals. In addition, the research centers have a clear preference to collaborate with geographically closer centers.

Sumario

Agradecimientos	v
Abbreviated abstract	ix
Índice de Tablas	xvii
Índice de figuras	xxi
I Introducción, hipótesis y objetivos	1
1 Introducción	3
1.1 Contexto histórico-social	6
1.1.1 Medicamentos huérfanos	6
1.1.2 La paradoja de la rareza	7
1.1.3 Problemática de la investigación en Enfermedades Raras (ER)	8
1.2 Situación actual	12
1.2.1 Programas nacionales y supranacionales	13
1.2.1.1 Europa	13
1.2.1.2 España	18
1.2.2 Genética, células madre y Enfermedades Raras	23
1.3 La evaluación de la Ciencia y la Tecnología	25
1.3.1 Los indicadores bibliométricos	27
1.3.1.1 Características de los indicadores	27
1.3.1.2 Limitaciones de los indicadores	31
1.3.1.3 Tipos de indicadores	35
1.3.2 Bibliometría en Ciencias de la Vida y de la Salud	38
1.3.2.1 Investigaciones bibliométricas en ER	40
1.3.2.2 Uso de indicadores de salud en estudios bibliométricos cos	41
1.4 Justificación científica	44

2	Hipótesis	47
3	Objetivos	49
II	Metodología	51
4	Fuentes, recogida y tratamiento de los datos	53
4.1	Clasificación de las Enfermedades Raras	53
4.1.1	La Clasificación Internacional de Enfermedades (ICD) . . .	54
4.1.2	Sistema de clasificación MIM (Mendelian Inheritance in Man)	56
4.1.3	Medical Subject Headings (MeSH)	57
4.1.4	La utilidad de las clasificaciones de enfermedades	58
4.1.4.1	Acotar por grupos de enfermedades	59
4.1.4.2	Mejorar la recuperación a través de la clasificación	59
4.1.4.3	Comparar con la investigación global en Enfermedades Raras	60
4.2	Selección de la muestra de enfermedades	60
4.2.1	Instituto de Investigación de Enfermedades Raras de Base Genética (INERGEN)	62
4.2.1.1	Enfermedades con causa genética identificada . .	62
4.3	Fuentes de información	62
4.3.1	Bases de datos bibliográficas	63
4.3.2	Bases de datos estadísticas sobre salud	63
4.3.3	Registro de asociaciones de pacientes	63
4.4	Método de recogida y tratamiento de los datos	64
4.4.1	Obtención de los datos	66
4.4.1.1	Estrategias de interrogación a las bases de datos bibliográficas	67
4.4.2	Gestión de las bases de datos: MySQL	68
4.4.2.1	Volcado a la base de datos en MySQL	70
4.4.2.2	Extracción de datos mediante consultas SQL . .	72
4.4.3	Tratamiento de los registros con scripts en Perl	73
4.4.3.1	Descripción del lenguaje Perl y su uso para el análisis de textos	73
4.4.3.2	Scripts para SCI y Pubmed	74
4.4.3.3	Scripts para la normalización	76
5	Análisis estadístico	103
5.1	Definición de indicadores métricos	103
5.1.1	Indicadores de Actividad Científica.	103
5.1.1.1	Indicadores de Producción.	106
5.1.1.2	Indicadores de Consumo.	114
5.1.2	Indicadores de Impacto / Visibilidad.	116
5.1.2.1	Factor de Impacto.	118

5.1.2.2	Análisis de citas	122
5.1.2.3	Ranking de revistas	124
5.1.3	Indicadores conexionistas	125
5.1.3.1	Colaboración	126
5.1.3.2	Cocitación	127
5.1.3.3	Temáticas	129
5.1.4	Indicadores multidimensionales	129
5.1.4.1	Escalados multidimensionales (MDS)	130
5.1.4.2	Análisis de clusters	132
5.1.4.3	Análisis de correspondencias	134
5.2	Análisis estadístico y uso del paquete estadístico R	135
5.2.1	Estadística descriptiva	136
5.2.2	Estadística inferencial	136
5.2.3	Regresión lineal	138
5.2.4	El paquete estadístico R	140
6	Estudio de solapamiento entre Pubmed y SCI	145
6.1	Solapamiento de las publicaciones	145
6.2	Solapamiento de los términos MeSH	147
III	Resultados	149
7	Indicadores unidimensionales	151
7.1	Indicadores de actividad científica (Producción)	151
7.1.1	Producción anual	152
7.1.1.1	Publicaciones por año	152
7.1.1.2	Producción y tasa de crecimiento	152
7.1.2	Países	154
7.1.2.1	Producción anual por país	154
7.1.3	Tipología documental	156
7.1.3.1	SCI	156
7.1.3.2	Pubmed	157
7.1.4	Temática	161
7.1.5	Instituciones	165
7.1.6	Idiomas	171
7.1.7	Colaboración	173
7.1.7.1	Índice de colaboración	173
7.1.7.2	Coautoría	174
7.1.7.3	Colaboración nacional	180
7.1.7.4	Colaboración internacional	180
7.1.8	Dispersión	188
7.1.8.1	Concentración temática de las revistas	192
7.1.9	Productividad de los autores	194
7.1.9.1	Distribución de Lotka	194
7.1.9.2	Índice de élite de Price	198

7.1.9.3	Índice de Yablonsky	198
7.2	Indicadores de actividad científica (Consumo)	199
7.2.1	Distribución de las referencias	199
7.2.2	Obsolescencia	201
7.2.2.1	Semiperiodo de Burton y Kebler	203
7.2.2.2	Índice de Price	203
7.2.3	Temática	205
7.2.4	Aislamiento	205
7.2.5	Dispersión	205
7.3	Indicadores de Impacto / Visibilidad	208
7.3.1	Impacto (Documentos en el primer cuartil)	208
7.3.2	Análisis de citas	210
7.3.2.1	Total de citas	214
7.3.2.2	Citas por documento	216
7.3.2.3	Documentos no citados	223
7.3.3	Ranking de revistas	231
8	Análisis de regresión	237
8.1	Producción por país y porcentaje del gasto sanitario respecto al PIB	237
8.2	Producción frente a número de asociaciones de enfermos por país	250
9	Indicadores conexionistas	253
9.1	Autores	253
9.2	Instituciones	258
9.3	Países	260
9.4	Temáticas	262
10	Indicadores multidimensionales	265
10.1	Autores	265
10.2	Instituciones	272
10.3	Temáticas	275
10.4	Países	280
10.5	Publicaciones	284
IV	Discusión y conclusiones	295
11	Discusión	297
11.1	Países	306
11.2	Temática	314
11.3	Fuentes	318
11.4	Instituciones	323
11.5	Autores	332
12	Conclusions	341
12.1	Methodology	342

12.2 Countries	343
12.3 Subjects	344
12.4 Sources	345
12.5 Institutions	346
12.6 Authors	348
13 Conclusiones	351
13.1 Metodología	352
13.2 Países	353
13.3 Temática	355
13.4 Fuentes	355
13.5 Instituciones	357
13.6 Autores	358
14 Recomendaciones	361
V Bibliografía y anexos	363
Bibliografía	365
Apéndices	401
A	401
A.1 ER con causa genética identificada	401
A.2 Agrupación de enfermedades por clasificación ICD-10	412
A.3 Estrategias de búsqueda por clasificación ICD-10	414
B	427
B.1 Temática	427
B.2 Instituciones	439
B.3 Colaboración	444
B.4 Dispersión	455
B.5 Productividad de los autores	484
C	491
C.1 Lista de abreviaturas y sus nombres desarrollados	491

Índice de Tablas

4.1	Categorías MeSH	57
4.2	Trabajos publicados por año sobre trastornos primarios de los músculos	73
4.3	Comparación entre métodos para obtener la similaridad de los nombres de autor	85
4.4	Estructura del fichero de autoridades	87
4.5	Variables y umbrales de similaridad en la fase 2.1	91
4.6	Variables y umbrales de similaridad en la fase 2.2	92
4.7	Ejemplo de instituciones normalizadas	99
5.1	Indicadores de Actividad Científica (Producción)	104
5.2	Indicadores de Actividad Científica (Consumo)	104
5.3	Indicadores de Impacto / Visibilidad	105
5.4	Indicadores multidimensionales y conexiones	105
5.5	Ejemplo de distribución por cuartiles corregidos (GENETICS & HEREDITARY, 2009	125
5.6	Ejemplo de matriz de doble entrada para MDS	131
5.7	Ejemplo de matriz de distancias para MDS	131
5.8	Paquetes estadísticos utilizados en R	144
6.1	Muestra de la distribución de trabajos por revista, fecha y base de datos	146
7.1	Producción y tasa de crecimiento por año (Pubmed y SCI)	153
7.2	Distribución de la producción por países en Pubmed y SCI	156
7.3	Tipología documental en SCI	158
7.4	Tipología documental Pubmed	159
7.5	Detalle de los diferentes tipos documentales en la base de datos Pubmed (U.S. NATIONAL LIBRARY OF MEDICINE, 2012)	160
7.6	Producción por clasificación ICD-10 y fecha. Pubmed	162
7.7	Producción por clasificación ICD-10 y fecha. SCI	163
7.8	Producción por categoría WOS y fecha (Más de 500 documentos)	164
7.9	Producción por tipos institucionales	165

7.10	Núcleo de instituciones en SCI	166
7.11	Principales instituciones productoras por país	170
7.12	Tabla de contingencia entre países y tipos institucionales	170
7.13	Producción por idiomas en Pubmed y SCI	172
7.14	Índice de colaboración por país y año	175
7.15	Índice de colaboración por tipo institucional y año	176
7.16	Índice de colaboración por categorías WOS y año	177
7.17	Media de coautorías por país y año en SCI	178
7.18	Media de coautorías por tipo institucional y año	179
7.19	Media de coautoría por categorías WOS y año	181
7.20	Porcentaje de documentos en colaboración nacional por tipo ins- titucional	183
7.21	Porcentaje de documentos en colaboración nacional por país	184
7.22	Porcentaje de documentos en colaboración internacional por tipo institucional	185
7.23	Porcentaje de documentos en colaboración internacional por país	186
7.24	Matriz de colaboración por países UE15	187
7.25	Núcleo de revistas en SCI	190
7.26	Núcleo de revistas en Pubmed	191
7.27	Concentración temática de las revistas SCI	194
7.28	Núcleo de autores en Pubmed	197
7.29	Núcleo de autores en SCI	197
7.30	Distribución de referencias por año	201
7.31	Vida media por año de publicación	203
7.32	Vida media por clasificación ICD y fecha	204
7.33	Índice de Price	204
7.34	Referencias por año y temática (más de 1000 referencias)	206
7.35	Índice de aislamiento por país	207
7.36	Núcleo de revistas referenciadas	208
7.37	Documentos en el primer cuartil por país	209
7.38	Documentos en el primer cuartil por clasificación ICD-10	211
7.39	Documentos en el primer cuartil por tipo institucional	212
7.40	Documentos en el primer cuartil por institución	213
7.41	Total de citas por fecha	214
7.42	Total de citas por código ICD-10	217
7.43	Total de citas por tipo institucional	218
7.44	Total de citas por institución del núcleo	219
7.45	Citas por documento por países	220
7.46	Citas por documento por clasificación ICD-10	222
7.47	Citas por documento por tipo institucional	224
7.48	Citas por documento por institución del núcleo	225
7.49	Porcentaje de documentos no citados por fecha	226
7.50	Porcentaje de documentos no citados por año	227
7.51	Porcentaje de documentos no citados por clasificación ICD-10	228
7.52	Porcentaje de documentos no citados por tipo institucional	229

7.53	Porcentaje de documentos no citados por institución del núcleo	230
7.54	Factor de impacto de las revistas del núcleo	234
7.55	Cuartil corregido de las revistas del núcleo	236
8.1	Variables para el estudio de regresión entre producción y gasto sanitario	241
8.2	Variables para el estudio de regresión entre producción y número de asociaciones de enfermos	250
9.1	Abreviaturas de las instituciones	259
10.1	Revistas del núcleo y sus abreviaturas. (SC)	286
10.2	Revistas del núcleo y sus abreviaturas (Pubmed)	287
A.1	Enfermedades raras con causa genética identificada (INERGEN)	412
A.2	Agrupación de enfermedades por código ICD-10	414
B.1	Producción por términos MeSH y fecha (Pubmed)	432
B.2	Producción por términos MeSH y fecha (SCI)	438
B.3	Distribución de las instituciones en el SCI	444
B.4	Media de coautoría por términos MeSH y año (SCI)	450
B.5	Media de coautoría por términos MeSH y año (Pubmed)	454
B.6	Distribución de Bradford en SCI	458
B.7	Distribución de Bradford en Pubmed	462
B.8	Distribución de Bradford en las referencias bibliográficas	483
B.9	Distribución de Lotka en Pubmed	487
B.10	Distribución de Lotka en SCI	490

Índice de figuras

1.1	Diagrama de la evolución de los centros españoles	19
4.1	Estructura de bases de datos (SCI)	71
4.2	Ejemplo del funcionamiento de la técnica de String Matching . . .	79
4.3	Proceso de normalización temática	102
5.1	Ejemplo de dendograma	133
5.2	Trabajos publicados por año (Trastornos primarios de los músculos)	143
7.1	Evolución de la producción anual (SCI y Pubmed)	153
7.2	Producción anual por país Pubmed	155
7.3	Producción anual por país SCI	157
7.4	Distribución de la producción por países en Pubmed y SCI	158
7.5	Mosaico de la distribución de tipos institucionales por países . . .	171
7.6	Dispersión por revistas en SCI	188
7.7	Dispersión por revistas en Pubmed	192
7.8	Dispersión por autores en Pubmed	196
7.9	Dispersión por autores en SCI	196
7.10	Función de densidad de la distribución de referencias por fecha . .	202
7.11	Distribución de Bradford sobre las referencias bibliográficas	209
7.12	Total de citas por país	215
7.13	Distribución de citas por documento	221
8.1	FAS y FAP del modelo ajustado	245
8.2	Distribución de los residuos studentizados	248
8.3	Distribución de los residuos frente a valores ajustados	248
8.4	Dispersión de la producción por país en función del gasto sanitario	249
8.5	Dispersión de la producción por país en función del número de asociaciones	252
9.1	Red de colaboración entre autores (SCI)	255
9.2	Red de colaboración entre autores (Pubmed)	256
9.3	Red de autores citados	257

9.4	Red de instituciones	259
9.5	Red de la colaboración entre países	260
9.6	Red de la colaboración entre países (EU15)	261
9.7	Red de trabajos codificados con ICD-10 (SCI)	262
9.8	Red de trabajos codificados con ICD-10 (Pubmed)	263
10.1	Cluster de autores (SCI)	266
10.2	Cluster de autores (Pubmed)	267
10.3	Análisis de correspondencias de autores frente a códigos ICD-10 (SCI)	269
10.4	Detalle: Análisis de correspondencias de autores frente a códigos ICD-10. (SCI)	270
10.5	Análisis de correspondencias de autores frente a códigos ICD-10 (Pubmed)	271
10.6	Análisis de clusters de instituciones	273
10.7	Análisis de correspondencias entre instituciones y clasificación ICD-10	274
10.8	Escalado multidimensional de los términos MeSH (SCI)	276
10.9	Escalado multidimensional de los términos MeSH (Pubmed)	277
10.10	Clusters de términos MeSH (SCI)	278
10.11	Clusters de términos MeSH (Pubmed)	279
10.12	MDS de países	281
10.13	Análisis de clusters de países	282
10.14	Gráfico de mosaico países-idiomas	283
10.15	Análisis de correspondencias de países y clasificación CIE	283
10.16	Análisis de correspondencias de revistas y clasificación CIE (SCI)	288
10.17	Detalle: Análisis de correspondencias de revistas y clasificación CIE (SCI)	289
10.18	Análisis de correspondencias de revistas y clasificación CIE (Pubmed)	290
10.19	Detalle: Análisis de correspondencias de revistas y clasificación CIE (Pubmed)	291
10.20	Análisis de correspondencias de revistas y países	293
10.21	Detalle: Análisis de correspondencias de revistas y países	294

Capítulo I

Introducción, hipótesis y objetivos

Introducción

El concepto de Enfermedades Raras (ER) se acuña por primera vez a mediados de los años 80 en los Estados Unidos de Norte América (EEUU) y siempre estrechamente relacionado con el concepto de medicamentos huérfanos. Ambos términos se desarrollan en paralelo y ambos se dirigen a dar solución a los problemas que tienen las enfermedades de baja prevalencia (POSADA DE LA PAZ et al., 2008).

Actualmente la definición más aceptada de “enfermedades raras” no se ciñe exclusivamente al número de casos, sino que incluye otras características que son comunes a estas enfermedades. La primera característica común a este conjunto de enfermedades hace referencia a su prevalencia, es decir, al número de casos estimado que tiene una enfermedad concreta. Las tasas de prevalencia para que una enfermedad sea considerada rara varía en función del país. En Estados Unidos se contabiliza de forma global, menos de 200.000 casos para todo el país, mientras que en la Unión Europea se habla de una prevalencia menor de 5 casos por cada 10.000 habitantes y en Japón de 4 por cada 10.000.

La siguiente característica que marca la definición de enfermedad rara es el hecho de que “se trata de enfermedades graves o muy graves, crónicas, degenerativas y, habitualmente mortales” (EURORDIS, 2005).

Esto, unido a que con frecuencia conllevan una discapacidad, supone que la calidad de vida de los pacientes se ve muy mermada y suele redundar en problemas

psicológicos tanto para el paciente como para su familia, y por tanto en una clara situación de dependencia.

Continuando con las características principales de las enfermedades raras, cabe mencionar que en su mayoría se manifiestan durante la infancia, aunque también pueden afectar a adultos; que se trata de enfermedades incurables y a menudo sin tratamiento eficaz, aunque en algunos casos existan tratamientos que permitan paliar la pérdida de calidad y aumentar la esperanza de vida de los pacientes.

Una vez determinadas las características que deben poseer las enfermedades raras para ser denominadas así, hay que precisar algunos conceptos relacionados con ellas y que pueden dar lugar a equívoco.

En primer lugar hay que explicar el término “enfermedades olvidadas”, pues a menudo suele confundirse con el de ER, y a continuación se definirá el concepto que engloba a ambas bajo el término de “enfermedades huérfanas” y su homólogo, los “medicamentos huérfanos”.

En contra de lo que ocurre con las ER, las enfermedades olvidadas son enfermedades habituales, perfectamente documentadas y generalmente contagiosas, que tienen en común con las ER el hecho de que la investigación que se realiza sobre ellas es escasa, pero se debe a motivos diferentes. En este caso la falta de investigación viene marcada por factores demográficos y sociales, puesto que se trata de enfermedades que afectan principalmente a países del tercer mundo o en vías de desarrollo y por tanto no suponen una prioridad para la industria farmacéutica, al considerarlo un mercado en el que puede obtener muy poca o ninguna rentabilidad. Es por ello que se hace necesaria: “una regulación económica y enfoques alternativos para crear incentivos que estimulen la investigación y la elaboración de tratamientos para luchar contra las enfermedades descuidadas, que son prevalentes en los países pobres” (EURORDIS, 2005).

Por todo lo anterior, los conceptos de “enfermedad rara” y “olvidada” pueden incluirse dentro de uno más amplio que hace referencia a ambas categorías. Se trata

del concepto de “enfermedades huérfanas” y se refiere a aquellas enfermedades “huérfanas” de investigación y por tanto de tratamiento.

Por tanto, esta orfandad, viene marcada por diferentes motivos. En el caso de las enfermedades raras se debe a su baja prevalencia, que dificulta enormemente la obtención de muestras biológicas, unido por supuesto a la baja rentabilidad de la investigación en este tipo de enfermedades. Sin embargo, en las enfermedades olvidadas la falta de investigación se debe exclusivamente a factores económicos y comerciales, que hacen poco rentable para la industria farmacéutica del primer mundo investigar en ellas, debido a que afectan principalmente a países que no cuentan con los medios suficientes para desarrollar ellos mismos estas investigaciones.

Por tanto, y aunque por motivos diferentes, la causa subyacente de la falta de investigación en estos campos es la misma; la investigación en ellos no es rentable.

En vista de todo ello, los responsables de las políticas sanitarias, tanto a nivel nacional como europeo, llegaron a la conclusión de que las enfermedades raras y su investigación son un problema acuciante para la sociedad, y como se verá en el siguiente punto, han desarrollado leyes, planes, proyectos, etc., con el objetivo de fomentar la investigación en este área.

Además, hay que tener en cuenta el destacado papel que juegan determinados factores sociales, como las asociaciones de enfermos, en la difusión de dichas patologías, pues muchas de las enfermedades, que por tasas de prevalencia entran dentro de la consideración de enfermedades raras, cuentan con una gran difusión en la sociedad. Es el caso de trastornos como la fibrosis quística, el autismo o el déficit de atención (RABEHARISOA y CALLON, 2002). Si bien este interés por parte de la sociedad se debe en muchas ocasiones a la difusión que se les da a estas enfermedades en los medios de comunicación, tanto tradicionales como “mass media” (SÁNCHEZ CASTILLO, 2012; HOLTZMAN et al., 2005; GARCÍA SÁNCHEZ y GARCÍA SÁNCHEZ, 2005). Ahora bien, esta visibilidad viene marcada en mu-

chas ocasiones por la actividad que se realiza desde las diferentes asociaciones de pacientes, de forma que las asociaciones más organizadas y que buscan activamente la presencia en los medios de comunicación, favorecen notablemente que se realice una mayor investigación sobre las enfermedades asociadas (ARRISCADO et al., 2007).

En este sentido, se considera importante desarrollar trabajos como el que aquí se presenta, que muestren cuál es el perfil de la actividad científica en este área, y cuáles son los factores, médicos y sociales, que la afectan.

1.1. Contexto histórico-social

En este apartado se explicarán dos de los conceptos más estrechamente relacionados con las ER, los medicamentos huérfanos y la llamada “paradoja de la rareza”. Al mismo tiempo, se muestra un análisis de cuáles son los principales problemas a los que se enfrenta la investigación en ER desde distintos puntos de vista: epidemiológico, de falta de información, financiero, etc.

1.1.1. Medicamentos huérfanos

Orphanet, el mayor portal europeo de información sobre enfermedades raras y medicamentos huérfanos, define los medicamentos huérfanos como:

aquellos fármacos que no son desarrollados por la industria farmacéutica por razones económicas pero que responden a necesidades de salud públicas (ORPHANET, 2011b).

Según la legislación de la Unión Europea:

“Un medicamento será declarado medicamento huérfano si su promotor puede demostrar que dicho producto: a) se destina al diagnóstico, prevención o tratamiento de una afección que ponga en peligro la vida o conlleve una incapacidad crónica y que no afecte a más de cinco personas por cada diez mil en la Comunidad en el momento de

presentar la solicitud; o se destina al diagnóstico, prevención o tratamiento, en la Comunidad, de una afección que ponga en peligro la vida o conlleve grave incapacidad, o de una afección grave y crónica, y que resulte improbable que, sin incentivos, la comercialización de dicho medicamento en la Comunidad genere suficientes beneficios para justificar la inversión necesaria; y b) que no existe ningún método satisfactorio autorizado en la Comunidad, de diagnóstico, prevención o tratamiento de dicha afección, o que, de existir, el medicamento aportará un beneficio considerable a quienes padecen dicha afección”.
(UNIÓN EUROPEA, 2000)

Además hay que tener en cuenta que, al igual que las enfermedades huérfanas, existen diferentes razones por las que un medicamento puede no ser rentable.

Orphanet distingue tres categorías en función de que sean medicamentos dirigidos a tratar enfermedades raras; medicamentos no desarrollados por estar orientados a un mercado importante pero poco rentable (enfermedades olvidadas); o medicamentos que han sido retirados del mercado por razones económicas o terapéuticas.
(ORPHANET, 2011b)

1.1.2. La paradoja de la rareza

Observando las categorías que establece Orphanet para los diferentes tipos de medicamentos huérfanos, se puede llegar a la conclusión de que, mientras las enfermedades olvidadas afectan a una gran parte de la población, las enfermedades raras afectan a muy pocas personas debido a su baja prevalencia. Sin embargo la realidad es muy diferente, pues en el caso de las enfermedades raras se da lo que se conoce como “paradoja de la rareza”, que se refiere a que el número de casos de una enfermedad rara puede ser, y de hecho, es muy bajo, pero si se tiene en cuenta el conjunto de enfermedades raras se observa que existe un número considerable de afectados. En el caso de la Unión Europea, aproximadamente 30 millones de personas presentan una enfermedad rara, lo que supone entre el 6 % y el 8 % de la población de la Unión Europea (EURORDIS, 2005), cifra equivalente a la suma de las poblaciones de los Países Bajos, Bélgica y Luxemburgo.

En el caso de España, la cifra de afectados por enfermedades raras ascendería a tres millones de personas, si bien para conocer el verdadero alcance de esta paradoja “se necesita conocer la extensión en la población de las ER, tener datos fiables acerca de la epidemiología de estas enfermedades y disponer de registros de estas que nos informen de la carga poblacional que suponen en su conjunto y para cada una de ellas” (PALAU, 2009).

En cuanto a la prevalencia en recién nacidos, hay que destacar que el 80% de las enfermedades raras tiene un origen genético identificado, con implicación de uno o varios genes. Pueden heredarse o derivarse de una mutación genética *ex novo* y afectan a una proporción comprendida entre el 3% y el 4% de los nacidos. De nuevo, si se trasladan estos porcentajes a datos brutos, en la UE, el total de nacimientos que se corresponden con ese 3% o 4%, supone que entre 141.900 y 189.200 niños padecerán una enfermedad rara cada año. (POSADA DE LA PAZ et al., 2008).

1.1.3. Problemática de la investigación en Enfermedades Raras (ER)

Para terminar de presentar el retrato de las enfermedades raras es necesario hablar de cuáles son las principales dificultades con las que se encuentra la investigación en enfermedades raras.

En primer lugar hay que enfrentarse con el problema epidemiológico, pues los datos epidemiológicos de que se dispone no son adecuados para la mayoría de las enfermedades raras debido a diversas causas. Por un lado ocurre que la mayoría de las enfermedades raras no se encuentran correctamente especificadas en la International Classification of Diseases o ICD (que se abordará más detenidamente en la metodología), pues aparecen muchas enfermedades que no se asignan a ningún código y códigos que incluyen muchas enfermedades raras diferentes.

Por otra parte, y relacionado con el problema de la clasificación, surgen las dificultades a la hora de contar con registros de pacientes con enfermedades raras,

pues son muy pocos los registros existentes y se encuentran dispersos (diferentes registros para cada grupo de patologías) y poco coordinados, sobre todo a la hora de clasificar las enfermedades pues cada uno de ellos utiliza diferentes clasificaciones (ICD, Mendelian Inheritance in Man, clasificaciones propias, etc.) dificultando que se comparta la información y por tanto la colaboración entre los diferentes registros.

Por último está el problema de determinar la prevalencia, pues la mayoría de estos desordenes no se detectan en el nacimiento, sino al año o dos años de vida, algunas veces en la adolescencia e incluso en la edad adulta. Por lo que son necesarios estudios que incluyan seguimientos desde el nacimiento para determinar su verdadera prevalencia (WORLD HEALTH ORGANIZATION, 2005). A todo lo anterior hay que sumar las muertes prematuras, que tienen su verdadera causa en enfermedades raras no diagnosticadas y por tanto, no computadas a efectos de los estudios de prevalencia.

En este sentido, las estrategias en enfermedades raras del sistema nacional de salud en nuestro país, pretenden implantar sistemas para la clasificación de las enfermedades raras diagnosticadas en los sistemas de Atención Primaria, permitiendo conformar unos registros de enfermedades raras y datos sobre la prevalencia de las mismas mucho más completos (MINISTERIO DE SANIDAD Y POLÍTICA SOCIAL, 2009).

Con respecto al problema epidemiológico con el que se enfrenta la investigación en ER, y señalado anteriormente, hay que destacar que se manifiesta fundamentalmente en la falta de conocimientos básicos acerca de los diferentes trastornos, pues de las cerca de 7.000 enfermedades raras descubiertas hasta el momento, son muy pocas las que se encuentran suficientemente documentadas, derivando en grandes dificultades diagnósticas, agravadas por la falta de experiencia en Atención Primaria a la hora de enfrentarse con estas patologías. Esto supone un retraso importante en el diagnóstico de las mismas, pues son estos profesionales

quienes se enfrentan en primer lugar con los síntomas iniciales de la enfermedad, y de su diagnóstico precoz depende en gran medida la evolución del afectado.

De nuevo en los planes estratégicos de enfermedades raras en España, figura como una de las recomendaciones más importantes “la formación específica en ER [...] para la detección precoz, atención y seguimiento adecuados de personas con ER, ya que aumenta la calidad de vida del paciente y su familia, evitando la visita a múltiples especialistas y pruebas diagnósticas innecesarias” (MINISTERIO DE SANIDAD Y POLÍTICA SOCIAL, 2009), así como fomentar la colaboración entre los diferentes centros e incorporar materias relativas a la Atención Primaria en los estudios universitarios, con el fin de que esta formación ayude al diagnóstico precoz de las enfermedades raras y desarrollar un enfoque sanitario más orientado al enfermo que a la enfermedad, para evitar el tránsito de los enfermos por los diferentes servicios especializados en busca del diagnóstico de la enfermedad.

Otro problema derivado de la falta de conocimientos acerca de la historia natural de estas patologías es que la investigación en enfermedades raras adolece de una importante carencia de muestras biológicas, ya sean humanas o animales, in vitro o in vivo, con las que llevar a cabo las investigaciones necesarias. Ello convierte este material en un producto muy costoso y difícil de obtener, contribuyendo por tanto a agravar el siguiente escollo con el que se encuentra la investigación en enfermedades raras (WORLD HEALTH ORGANIZATION, 2005).

Teniendo en cuenta la baja prevalencia (y por tanto el escaso mercado para la industria farmacéutica) de las enfermedades raras, la falta de información acerca de ellas y el escaso material biológico con el que trabajar, no es extraño que el desarrollo de productos farmacéuticos para estos trastornos sea más complejo, prolongado y costoso que en el caso de enfermedades de alta prevalencia (WORLD HEALTH ORGANIZATION, 2005). Por todo ello, la industria farmacéutica que se encarga de la producción de estos fármacos, no está interesada en invertir gran parte de sus recursos humanos y económicos en desarrollar productos que tienen un mercado muy limitado. A ello hay que añadir que, una vez desarrollado el

fármaco existen grandes dificultades para llevar a cabo los ensayos clínicos previos a su comercialización, precisamente debido al escaso número de pacientes y su difícil localización derivada de la falta de registros adecuados.

Como tercer punto, y dejando de lado los problemas propios de la investigación científica que se han comentado, existen algunas cuestiones de carácter más social y que conllevan igualmente serias dificultades para la investigación en enfermedades raras. En primer lugar, la falta de interés de la sociedad por estas patologías. Su misma denominación como “raras” lleva a pensar que se trata de enfermedades minoritarias, algo que ya ha sido comentado anteriormente y que es totalmente erróneo (30 millones de afectados sólo en la Unión Europea).

Además, al tratarse de enfermedades que requieren una terapia muy costosa y que el desarrollo de medicamentos para tratarlas lo es aún más, por lo que los sistemas sanitarios en muchas ocasiones no pueden hacer frente, con sus limitados presupuestos, a todos estos gastos, por lo que desde hace varios años se viene arrastrando la carencia de una adecuada política de apoyo y programas de investigación en el campo de las enfermedades raras.

Pese a todo lo anterior hay que indicar que se realizan esfuerzos para mejorar la situación. Durante mucho tiempo la falta de una infraestructura para la investigación y de un flujo de intercambio de información entre investigadores, ha sido predominante en el campo de las enfermedades raras. Sin embargo durante la última década se han desarrollado algunas organizaciones, como la National Organization of Rare Diseases (NORD) en EEUU, o las instituciones surgidas al amparo de los programas Marco de la Unión Europea (MOLINER y AYMÉ, 2010; MARTIN et al., 2010; MCCARTHY, 2010), entre las que se cuentan Orphanet y Eurordis, que han contribuido sobremanera al establecimiento de vías de comunicación entre los grupos de investigación que trabajan en el área de las enfermedades raras.

Estas instituciones han dado un nuevo impulso a la investigación que se realiza en ER y sobre todo a la visibilidad de la misma.

Por último, existe un problema que ya se ha mencionado anteriormente, y es la dificultad de calcular la verdadera prevalencia de estas enfermedades. En este sentido hay que distinguir entre prevalencia e incidencia, dos conceptos que a menudo se confunden. La prevalencia hace referencia al número de individuos que padecen una enfermedad concreta, en una área geográfica determinada, mientras que la incidencia mide el número de nuevos casos de una enfermedad concreta, diagnosticados en un área geográfica y un periodo de tiempo determinados.

Por tanto, aunque se trata de dos conceptos estrechamente relacionados, no poseen una relación directa, ya que una enfermedad con una baja prevalencia puede tener una incidencia alta y viceversa, especialmente en el caso de enfermedades causadas por factores exógenos (MALLADA, 1999). Ahora bien, a pesar de las diferencias entre ambos conceptos, en muchas ocasiones, cuando los datos de prevalencia no se encuentran disponibles, es posible estimarla utilizando los datos relativos a la incidencia de la enfermedad, e incluso al número de casos reportados a través de la literatura científica, si bien esta forma de estimar la prevalencia puede derivar en una sobreestimación de la misma (Odi, 2010).

1.2. Situación actual

Como ya se ha mencionado, en los últimos años las diferentes políticas de salud tanto nacionales como supranacionales, han establecido la investigación y atención a las enfermedades raras y quienes las padecen, como una prioridad dentro de los sistemas sanitarios. Esto sucede especialmente, desde que se promulgara la Decisión N° 1295/1999/CE del Parlamento Europeo y del Consejo, de 29 de abril de 1999, por la que se aprobaba un programa de acción comunitaria sobre las enfermedades poco comunes para el periodo 1999-2003. En el marco de la acción en el ámbito de la salud pública, se han desarrollado diferentes proyectos y políticas de actuación acerca de las enfermedades poco comunes en la UE. Entre todos estos proyectos cabe destacar el proyecto ORPHANET (2011a), que

a continuación se describirá más profundamente y que actualmente se encuentra en su cuarta fase de desarrollo, convirtiéndose en la mayor fuente de información sobre enfermedades raras a nivel europeo, pues cuenta con información integrada acerca de cada enfermedad.

1.2.1. Programas nacionales y supranacionales

A continuación se describen los desarrollos más importantes dentro de estas políticas sanitarias, tanto a nivel europeo como las realizadas en España, prestando especial atención a los proyectos financiados a través de los Programas Marco de la Unión Europea.

1.2.1.1. Europa

Desde la promulgación de las decisiones del Parlamento y Consejo de la Unión Europea en 1999 y 2000, estas han derivado en múltiples iniciativas y planes de acción relacionados con la investigación de enfermedades raras. Entre estas iniciativas destacan las surgidas al amparo de los Programas Marco, así como el EU Committee of Experts on Rare Diseases, sustituto del anterior Rare Diseases Task Force. Ambos se crearon con el objetivo de implementar y coordinar las actividades de la Unión Europea en relación con las enfermedades raras (MOLINER y AYMÉ, 2010).

Entre las principales actividades de este Comité, se encuentran la prevención de las enfermedades raras así como el análisis y la provisión de asistencia técnica en el desarrollo de medicamentos huérfanos que ayuden a prevenir y paliar el desarrollo de estas enfermedades. Además, el EU Committee of Experts on Rare Diseases participa activamente en el desarrollo de las nuevas versiones de diferentes sistemas de clasificación de enfermedades, como la Clasificación Internacional de Enfermedades (ICD10 e ICD11) y la Nomenclatura Sistematizada en Medicina - Términos Clínicos o Systematized Nomenclature of Medicine – Clinical

Terms (SNOMED-CT), que tiene el objetivo de desarrollar sistemas de clasificación que se adapten mejor a las necesidades de clasificación de las enfermedades raras.

En cuanto a los Programas Marco, desde su quinta edición (1998 - 2002), la Unión Europea ha financiado a través de ellos, un gran número de proyectos de investigación y colaboración entre investigadores especializados en el área de las enfermedades raras. Concretamente el 5º Programa Marco financió, con 64 millones de euros, 47 proyectos de investigación en el área, incluyendo un proyecto, el EuroBioBank, que tenía como finalidad el desarrollo de un banco de muestras biológicas para el estudio de estas enfermedades.

En el 6º Programa Marco (2002-2006) la financiación para proyectos relacionados con enfermedades raras fue aún mayor, con 59 proyectos financiados por un valor de 230 millones de euros, entre ellos múltiples proyectos relacionados con la investigación genética, que comenzaba a despuntar debido a la elaboración del mapa del genoma humano, cuya publicación completa en 2003 (COLLINS et al., 2003) dio lugar a múltiples proyectos que buscaban determinar las causas genéticas de muchas de las enfermedades catalogadas como raras.

Por último, en el 7º Programa Marco que concluirá en 2013, se han financiado (y se continúan financiando) muchos proyectos de investigación relacionados con las enfermedades raras, a pesar de que se ha minimizado su importancia respecto al programa anterior, que incluso contaba con un apartado específico de financiación dentro de la convocatoria de proyectos.

Actualmente la Unión Europea lleva a cabo diferentes políticas e iniciativas, para prevenir y tratar las enfermedades raras enmarcadas tanto en el planteamiento estratégico para la UE (2008-2013), como en el segundo Programa de acción comunitaria en el ámbito de la salud (2008-2013). Este último pivota sobre dos ejes principales de acción del plan de trabajo de la Dirección General de Sanidad y de los Consumidores para la aplicación del Programa de Salud Pública. Estos dos ejes son el intercambio de información mediante las redes europeas de información

sobre enfermedades poco comunes, y la elaboración de estrategias y mecanismos para el intercambio de información y la coordinación a escala comunitaria, para promover la continuidad del trabajo y la cooperación transnacional. (UNIÓN EUROPEA, 2007).

Entre los actuales proyectos de la Unión Europea sobre enfermedades raras se encuentran iniciativas orientadas a proporcionar información a los afectados y sus familias (proyectos RAPSODY, PARECELSUS, PARD o NEPHIRD. O el programa educativo orientado tanto a profesionales como a los pacientes y sus asociaciones. Asimismo, existen proyectos destinados a crear redes de actuación en el ámbito de las enfermedades poco comunes, tales como el proyecto EUROCAT de vigilancia de las anomalías congénitas en Europa. Y multitud de proyectos (ENERCA, SCN, EUROMUSCLENET, CAUSE, EDDNAL, etc.), orientados a crear redes europeas de información sobre enfermedades raras concretas, tales como anemias raras congénitas (ENERCA), la neutrocitopenia (SCN), enfermedades musculares (EUROMUSCLENET), etc. Del mismo modo, existen iniciativas para crear centros de referencia que sirvan como centros de investigación y de conocimiento, dedicados a actualizar y contribuir a los hallazgos científicos más recientes, tratar a pacientes de otros Estados miembros y asegurar la disponibilidad de servicios de tratamiento ulterior en caso necesario (HANNEMANN-WEBER et al., 2012).

Por otro lado, con la intención de fomentar la adopción de planes y de estrategias nacionales, antes de 2013 y con el fin de responder a las enfermedades raras mediante la mejora del reconocimiento y la visibilidad de esas enfermedades, se desarrolló el proyecto EUROPLAN. Éste está promoviendo la continuidad de las estrategias nacionales sobre enfermedades raras en aquellos países (Francia, España, etc.) que ya contaban con ellas, y en su creación ex novo en países (Bulgaria, Portugal, etc.) que aún no las habían desarrollado.

En cuanto a los medicamentos huérfanos, “la Agencia Europea de Medicamentos (EMA) y los Estados Miembros, han propuesto a la industria farmacéutica

incentivos para la investigación, el desarrollo y la comercialización de dicho tipo de medicamentos en campos como el cáncer, los desordenes del metabolismo, la inmunología y las afecciones cardiovasculares y respiratorias” (UNIÓN EUROPEA, 2006). Actualmente 270 nuevos medicamentos han obtenido la designación de huérfanos pero todavía deben ser sometidos a ensayos clínicos.

Para terminar, es destacable que el fomento de la investigación en enfermedades raras se encuentra presente desde el 5º Programa Marco (1998-2002) hasta el vigente 7º Programa (2007-2013), con propuestas tales como la creación de un marco reglamentario para los medicamentos obtenidos por ingeniería tisular o el Committee for Advanced Therapies (CAT), en aplicación de la legislación europea referente a la regulación de los medicamentos de terapia avanzada.

Sin embargo, a pesar de que la investigación en enfermedades raras se mantiene como una prioridad en la Unión Europea y en los sistemas nacionales de salud de los diferentes países que la componen, la actual crisis económica y financiera ha supuesto, en muchos países, importantes recortes en los presupuestos de investigación, desarrollo e innovación (I+D+i).

Estos recortes han sido de mayor magnitud en aquellos países más afectados por la crisis de deuda pública, tales como Grecia, Irlanda, Portugal, España e Italia, por lo que es de suponer que en los próximos años la investigación realizada en estos países, tanto en enfermedades raras como en otras áreas, se verá muy mermada, por lo que un trabajo de características similares al que se ha realizado en el periodo 2000-2009, obtendrá resultados muy diferentes en periodos de tiempo posteriores.

1.2.1.1.1. Orphanet:

Entre los proyectos más importantes que se han desarrollado en estos años cabe destacar la red ORPHANET (2011a). Se autodefine como un “portal de información de referencia en enfermedades raras y medicamentos huérfanos, dirigido a todos los públicos”, y su objetivo es proporcionar un sistema de información com-

pleto acerca del conjunto global de enfermedades raras. Incluye una lista completa de enfermedades con su ratio de prevalencia, clasificación (ICD y MIM), sinónimos, su descripción general, síntomas, causas, datos epidemiológicos, medidas preventivas, tratamiento, ensayos clínicos, laboratorios de diagnóstico, consultas especializadas, proyectos de investigación, recursos en la Web y mucha más información complementaria.

Este portal, desarrollado en colaboración por el Instituto Nacional Francés de la Salud y de la Investigación Médica (INSERM), el Ministerio de Sanidad Francés y la Comisión Europea, se encuentra disponible en seis idiomas (inglés, francés, español, alemán, italiano y portugués), aunque la información más completa se encuentra en inglés y francés.

Además el proyecto cuenta desde 2006 con su propia revista científica, la “Orphanet Journal of Rare Diseases”, de acceso abierto e indizada por algunas de las bases de dato bibliográficas más importantes como la Web of Science, Scopus, Embase o Pubmed y publica un gran número de trabajos, todos ellos relativos a la investigación sobre enfermedades raras, ya sean patologías concretas o grupos de ellas, así como terapias, procedimientos diagnósticos, medicamentos o cualquier otro aspectos relacionado con la investigación en este tipo de patologías (AYME et al., 2006).

Por tanto se puede concluir que la investigación en enfermedades raras es una prioridad sanitaria en la Unión Europea, y que suscita suficiente interés en la comunidad investigadora como para que se haya planteado la necesidad de llevar a cabo un estudio que muestre cuál es el perfil de dicha investigación y si responde estrictamente a las necesidades de los afectados o si se ve influido por factores sociales tales como las asociaciones de pacientes con mayor visibilidad.

1.2.1.2. España

Centrando el estudio en este país, hay que señalar que el interés en las enfermedades raras surgió a partir de la ingesta de aceite de colza desnaturalizado, que se produjo en España en los años 80, con el consecuente desarrollo de la Enfermedad del Aceite de Colza o Síndrome del Aceite Tóxico. Sin embargo no fue hasta 2001 cuando se creó el Centro de Investigación del Síndrome del Aceite Tóxico (MARTÍN-ARRIBAS et al., 2003), que más tarde se ocuparía también de otras enfermedades raras, derivando en la creación del Sistema de Información de Enfermedades Raras en Español (SIERE).

A raíz de las decisiones europeas de 1999 y 2000, surgen en España 12 Redes Temáticas de Investigación Cooperativa Sanitaria (RETICS), encontrando entre ellas el Instituto de Investigación de Enfermedades Raras de Base Genética (INERGEN, 2000). Todas ellas estaban dedicadas al estudio de determinadas enfermedades raras y al amparo del Instituto de Salud Carlos III (ISCIII, 1986). Entre la labor de estas redes cabe destacar las investigaciones realizadas sobre la incidencia de determinadas enfermedades raras a nivel autonómico (ARIZO et al., 2005; RAMOS ACEITERO et al., 2005; RAMALLE GOMARA et al., 2004; GUILLÉ, ENRIQUEZ et al., 2004).

Sin embargo, el tiempo ha ido en contra de las iniciativas creadas para incentivar la investigación en enfermedades raras. Así, el Centro de Investigación del Síndrome del Aceite Tóxico y las Enfermedades Raras se convirtió en el Instituto de Investigación de Enfermedades Raras y desaparecieron la mayor parte de las redes RETICS, quedando paralizada gran parte de la labor investigadora desarrollada hasta ese momento (MINISTERIO DE SANIDAD Y POLÍTICA SOCIAL, 2009).

En 2006, con el desarrollo de los Centros de Investigación Biomédica en Red (CIBER), se recupera la labor realizada por los RETICS, creándose el Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER, 2006), tal y como se muestra en el diagrama de la figura 1.1.

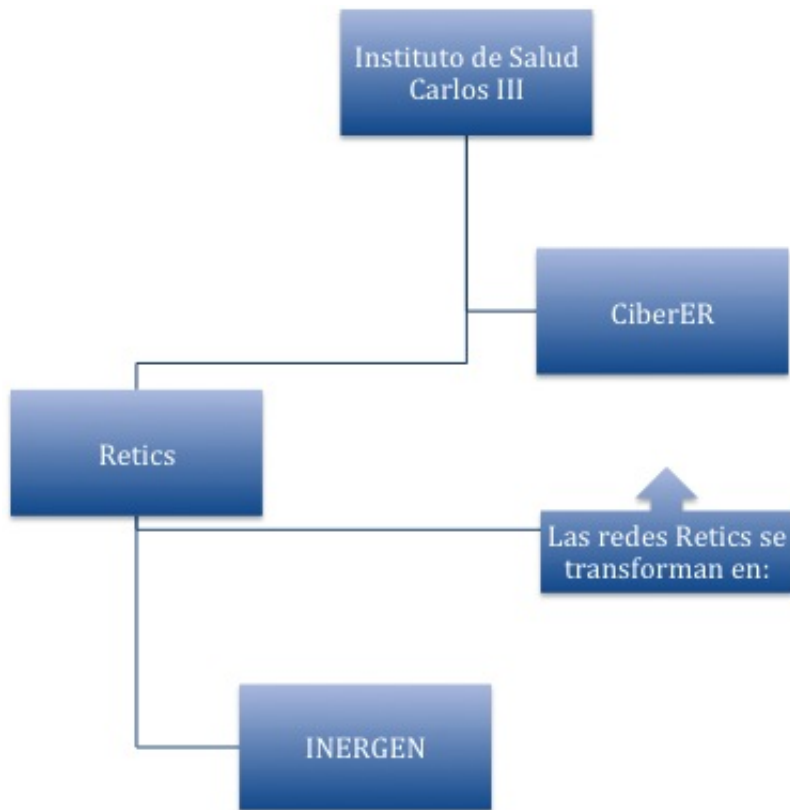


Figura 1.1: Diagrama de la evolución de los centros españoles

1.2.1.2.1. Instituto de Salud Carlos III:

El Instituto de Salud Carlos III (ISCIII, 1986) es el principal actor en el desarrollo de investigaciones relacionadas con las enfermedades raras en España. Este centro, creado en 1986, es el principal Organismo Público de Investigación relacionado con la financiación, gestión y ejecución la investigación biomédica en España. Así mismo es el responsable de la Acción Estratégica en Salud (AES) del Plan Nacional de I+D+i.

En cuanto a su papel en la investigación en enfermedades raras, el ISCIII es el responsable del Programa de Investigación en Enfermedades Raras y en su seno han surgido múltiples iniciativas relacionadas con esta investigación, como las mencionadas redes RETICS, el Instituto de Investigación de Enfermedades Raras o el CIBERER.

Además, el ISCIII ha desarrollado en los últimos años importantes iniciativas necesarias para cimentar las bases de la investigación biomédica en enfermedades raras. Tal es el caso del Registro de Enfermedades Raras, creado por el ISCIII junto con el Instituto de Investigación de Enfermedades Raras y el CIBERER. Este registro, establecido a nivel nacional, tiene como objetivo reunir y mantener actualizado un censo, fiable y lo mas completo posible, de pacientes que padezcan una determinada enfermedad considerada rara, para poder desarrollar investigación de mayor calidad y validez.

1.2.1.2.1.1. INERGEN:

Como ya se ha comentado, entre los centros españoles de investigación sobre enfermedades raras más importantes, se encuentran los RETICS. Entre ellos, destaca la labor realizada por el INERGEN, por ser el centro más relacionado con el desarrollo de esta tesis doctoral, y el que ha proporcionado el conjunto de enfermedades raras de base genética que ha sido utilizada como muestra.

El Instituto de Investigación de Enfermedades Raras de Base Genética surge en el año 2002, como una de las Redes Temáticas de Investigación Cooperativa (RETICS), y centrada en la “integración y coordinación de la investigación en la fisiopatología molecular, en nuevas tecnologías aplicables al diagnóstico, la validación e implementación de pruebas y su aplicación clínica, así como los avances en prevención y tratamiento” (CARRASCO MALLEN et al., 2005), y se encuentra estrechamente relacionado con la Red Epidemiológica de Investigación en Enfermedades Raras.

Por otro lado el INERGEN tenía como objetivo llamar la atención sobre las necesidades especiales tanto en la investigación como en los enfermos con enfermedades raras de base genética, a través de la formación de los diferentes agentes del Sistema Nacional de Salud. Para ello desarrolló, durante el tiempo que estuvo vigente, labores de formación en Genética Médica y Atención Primaria, así como proyectos de investigación relacionados con la carga genética de las enfermedades

raras. Entre ellos están los proyectos “Evaluación de los genes de susceptibilidad relacionados con el síndrome metabólico gestacional y el desarrollo a corto plazo tras el parto de diabetes mellitus y factores de riesgo vascular” o “Innovación e integración de la prevención y promoción de la salud en atención primaria”.

1.2.1.2.1.2. CiberER:

Como ya se ha comentado el CIBERER surge en 2006 con la finalidad de coordinar y potenciar la investigación biomédica sobre las enfermedades raras en el Estado. Esta institución está formada por 60 grupos de investigación diferentes, agrupados en siete programas de investigación: Medicina Genética, Medicina Metabólica Hereditaria, Medicina Mitocondrial, Medicina pediátrica y del desarrollo, Patología Neurosensorial, Medicina Endocrina y Cáncer Hereditario y Síndromes relacionados. Los objetivos comunes de todos ellos son:

1. Creación de base de datos y registros compartidos.
2. Dinamizar la actividad investigadora de los programas estimulando la colaboración científica.
3. Promover la divulgación y la participación en actividades docentes relacionadas con los programas.
4. Diseminar el conocimiento a otros profesionales y al público en general, dando a conocer la relevancia de la investigación y facilitando la transferencia de conocimiento al beneficio del paciente.
5. Promover la transferencia de los resultados de investigación y de tecnología hacia la sociedad y en especial al Sistema Nacional de Salud y al sector productivo farmacéutico y biotecnológico (CIBERER, 2006).

El CIBERER se encuentra incluido dentro del Instituto de Salud Carlos III y unido muy estrechamente al Consejo Superior de Investigaciones Científicas (CSIC), pues su razón social está ubicada en el Instituto de Biomedicina de Valencia

(IBV), que depende a su vez del CSIC y muchos de los grupos de investigación que componen el CIBERER pertenecen a su vez al CSIC.

Entre el trabajo desarrollado por el CIBERER destacan algunas iniciativas como las desarrollados por la unidad 741, que ha “contribuido significativamente a la implantación de tecnologías propias de la biología de sistemas (tanto bajo simulación informática como de forma experimental) en el estudio de las ERs y otras enfermedades emergentes, a través de distintos proyectos financiados por diversas instituciones públicas y privadas regionales, nacionales e internacionales” (JIMÉNEZ, 2010).

Además el CIBERER ha llevado a cabo una ingente labor comunicativa, junto con la Federación Española de Enfermedades Raras (FEDER), contribuyendo a un mejor conocimiento por parte de la opinión pública de las enfermedades raras y de la investigación que se realiza sobre ellas, nutriendo a los medios de comunicación de noticias acerca de la labor desarrollada, si bien no siempre ha alcanzado la repercusión que se buscaba o han sido malinterpretadas (BAÑON HERNÁNDEZ et al., 2011).

1.2.1.2.2. FEDER:

La Federación Española de Enfermedades Raras (FEDER) lidera el movimiento asociativo sobre enfermedades raras a nivel nacional. En 2005 la Federación contaba con aproximadamente 50 asociaciones y entidades, mientras que actualmente se compone de mas de 220 “que agrupan colectivos de afectados y/o interesados por enfermedades de baja incidencia en nuestro país, entre sus logros pueden citarse la celebración de numerosos actos de interés científico en favor de las enfermedades raras, así como la información y el apoyo emocional a pacientes afectados por estas enfermedades y sus familiares, favoreciendo el intercambio de experiencias entre las distintas asociaciones y el fomento del desarrollo y acceso de nuevos tratamientos para las enfermedades raras” (JUNTA DE EXTREMADURA, 2005). Actualmente incluso participa en la elaboración de las estrategias y planes

ministeriales en relación con las enfermedades raras en el Sistema Nacional de Salud (POSADA DE LA PAZ et al., 2008).

Cabría esperar que una federación tan importante, con peso en los planes estratégicos del Ministerio de Sanidad y con un proceso de expansión tan rápido, tuviera la capacidad de cohesionar a las asociaciones y organizaciones de enfermos con enfermedades raras, con el fin de unificar sus esfuerzos. Sin embargo, el movimiento asociativo en enfermedades raras, tanto en España como en el resto del mundo se encuentra todavía demasiado atomizado y disperso, si bien el proceso llevado a cabo va en la dirección correcta y ha servido para darle visibilidad internacional a las asociaciones españolas y fomentar su colaboración (AVELLANEDA et al., 2007).

1.2.2. Genética, células madre y Enfermedades Raras

“Hablar de genética en las enfermedades raras es referirse al conjunto amplio de enfermedades monogénicas, síndromes cromosómicos y malformaciones congénitas. Las enfermedades monogénicas son trastornos debidos a mutaciones que pueden afectar bien a alguno de los aproximadamente 25.000 genes del genoma nuclear que codifican proteínas y se transmiten según las leyes de la herencia de Mendel (de ahí que también se conozcan como enfermedades mendelianas), o en el pequeño genoma ubicado en la matriz de las mitocondrias con un patrón de transmisión específico conocido como herencia mitocondrial. Las aberraciones cromosómicas son el sustrato etiopatogénico de síndromes clásicos debidos a alteraciones numéricas de los cromosomas (síndrome de Down, síndrome de Turner, síndrome de Klinefelter), deleciones parciales, ejemplos de las cuales son el síndrome del maullido del gato, el síndrome de Williams y el síndrome de microdeleción 22q11, o a alteraciones estructurales de los cromosomas”
GONZÁLEZ-LAMUÑO y GARCÍA FUENTES (2008).

Además, la relación entre enfermedades raras y genética también se manifiesta en el otro sentido, es decir, que no solo ocurre que la mayor parte de las enfermedades raras tienen causa genética, si no que también la mayor parte de las enfermedades

con causa genética son trastornos de muy baja prevalencia (GUILLÉN-NAVARRO et al., 2011).

Esta relación entre enfermedades raras y enfermedades de base genética (que en muchos casos son sinónimos), ha causado que a lo largo del tiempo, las enfermedades más estudiadas desde el punto de vista genético hayan sido las enfermedades raras, pues en su mayor parte se trata de trastornos monogénicos, mientras que “la mayor parte de enfermedades comunes como diabetes, hipertensión, cáncer, etc., o características como talla, peso, color, son multifactoriales con la influencia de varios genes y del medio ambiente” (FUJITA, 2007).

Por otro lado, como se ha constatado en estudios anteriores, realizados sobre enfermedades concretas (SERRANO-LÓPEZ y MARTÍN-MORENO, 2011), el descubrimiento del mapa del Genoma Humano y su publicación completa en 2003 (COLLINS et al., 2003), ha marcado un antes y un después en la investigación de las enfermedades raras, debido a que ha permitido identificar muchas de las causas de estos trastornos, aislando los genes responsables de ellos y abriendo la puerta a los tratamientos génicos en el futuro.

Por tanto, al igual que ocurría en el trabajo mencionado anteriormente, en este se espera encontrar una fuerte relación entre la producción científica en el área de las enfermedades raras y los descubrimientos más importantes relativos a la Genética, tales como el Mapa del Genoma Humano o el descubrimiento de algunos de los genes causantes de los trastornos estudiados.

Otro campo de estudio que ha despertado gran interés sobre las enfermedades raras, es la investigación y desarrollo de terapias por medio de la manipulación de células madre. El desarrollo de terapias basadas en este tipo de células “reprogramables” abre la puerta a tratamientos mucho más eficaces en las enfermedades raras e incluso a la regeneración de tejidos inexistentes o desaparecidos por su causa (NECHAEVSKY, 2008; BUEREN, 2010).

Como ejemplo puede plantearse el caso de la Retinitis Pigmentaria, enfermedad rara que tal y como explica DE PABLO DÁVILA (2009), “está causada por una

larga serie de mutaciones en distintos genes, que comienza en la etapa juvenil y conduce a la ceguera”. Este mismo autor, junto con su grupo de investigación ha desarrollado una terapia, probada hasta la fecha exclusivamente en ratones, que permite la generación de una variante sintética de insulina, que retrasa significativamente la degeneración de los fotorreceptores y por tanto, de la ceguera.

Las asociaciones de pacientes han luchado activamente por conseguir que se apruebe un programa nacional de investigación con células madre (BAÑÓN HERNÁNDEZ, 2007), a pesar de que hasta la reforma de la Ley de Reproducción Asistida en 2003, en España tan solo estaba permitida la investigación con células madre procedentes del cordón umbilical. Sin embargo, en dicha reforma se permite el uso de células embrionarias procedentes de los procesos de reproducción asistida, siempre y cuando los donantes lo autoricen expresamente. A pesar de lo cual la legislación española está aún muy alejada de la más avanzada en este campo, la legislación británica, que permite incluso la clonación de embriones humanos con el fin de extraer células madre para la investigación.

1.3. La evaluación de la Ciencia y la Tecnología

La actividad científica y su evaluación son dos conceptos que han estado unidos desde la aparición de la primera. Mientras que la Ciencia cuenta con un método único, consolidado y aceptado por toda la comunidad científica, el método científico, sin el cual no existiría la propia Ciencia, a la hora de evaluar la actividad científica se dispone de diversos métodos, derivados de la aplicación del método científica en el área de las Ciencias Sociales, que buscan caracterizar y determinar la calidad de la investigación científica de forma objetiva.

Centrando el estudio en la actividad científica que da lugar a publicaciones, estos métodos se dividen en métodos cualitativos y métodos cuantitativos. Los primeros hacen referencia básicamente a la evaluación por expertos y tienen su máximo exponente en el “peer review” que es el método habitual para determinar la

calidad de las contribuciones científicas a priori, es decir, evalúan los trabajos científicos con el objetivo de determinar si poseen la calidad suficiente para ser publicados y “[...] ha demostrado ser hasta la fecha el método más apropiado para valorar el grado de desarrollo de un determinado campo de investigación y la calidad de las aportaciones concretas al área de los distintos científicos o grupos. No obstante el sistema presenta algunas limitaciones, entre las que se puede citar su carácter subjetivo [...], su elevado coste y su aplicación limitada a pequeñas unidades” (BORDONS y ZULUETA, 1999).

Por otro lado, los métodos cuantitativos de evaluación se aplican a posteriori y se basan en la utilización de una serie de indicadores que pueden definirse como: “Los datos numéricos extraídos de los documentos que publican los científicos o que utilizan los usuarios, y que permiten el análisis de las diferentes características de su actividad científica, vinculadas, tanto a su producción como a su consumo de información” (SANZ-CASADO y MARTÍN-MORENO, 1997).

Siguiendo esta definición puede establecerse que dichos indicadores permiten determinar cuáles son las características de la actividad científica publicada. Si esta definición se extrapola a los propios investigadores, también se podrán determinar sus características, sus necesidades y hábitos de información. Tienen la ventaja sobre los métodos cualitativos, de que permiten aplicarse sobre conjuntos de datos mucho mayores y, en muchas ocasiones, sobre poblaciones completas.

La aplicación de estos métodos cuantitativos es objeto de la bibliometría. Esta disciplina científica fue definida por primera vez por PRITCHARD (1969) como:

La aplicación de los métodos estadísticos y matemáticos dispuestos para definir los procesos de la comunicación escrita y la naturaleza y el desarrollo de las disciplinas científicas, mediante técnicas de recuento y análisis de dicha comunicación.

Si bien, la disciplina (y su definición) han evolucionado notablemente desde entonces, tal como señala JIMENEZ CONTRERAS (2000):

Existe un alto grado de consenso en torno a dos ideas que están presentes en prácticamente todas las definiciones: los métodos es-

estadístico-matemáticos, o más genéricamente cuantitativos, que se han de emplear y el objetivo: analizar conjuntos documentales, sus productores y consumidores.

WHITE y MCCAIN (1989) comparan la Bibliometría con la Demografía, estableciendo un símil entre las bases de datos bibliográficas que recogen los datos esenciales de las publicaciones y los censos demográficos que hacen lo propio con la población. Del mismo modo, mientras que los demógrafos no necesitan conocer a los individuos que componen la población para estudiarla en su conjunto, los bibliométricos pueden estudiar el conjunto de publicaciones científicas sin necesidad de haberlas leído y conocer su contenido en profundidad.

Además, se trata de una disciplina muy asentada y que incluso cuenta con un conjunto de revistas especializadas en publicar trabajos bibliométricos, tales como *Annual Review of Information Science and Technology* o *Scientometrics* y otras muchas revistas, relacionadas con áreas afines a la información y la documentación, que también publican trabajos bibliométricos, como por ejemplo *Journal of the American Society for Information Science*, *Canadian Journal of Information Science*, *Information Processing Management*, *Social Studies of Science* o *Research Policy* (WHITE y MCCAIN, 1989).

1.3.1. Los indicadores bibliométricos

Como se ha comentado, los métodos cuantitativos de evaluación de la actividad científica se basan en la aplicación de indicadores bibliométricos, que se explican a continuación.

1.3.1.1. Características de los indicadores

Las diferentes características de los indicadores de ciencia y tecnología se han tratado de forma exhaustiva, y tras el análisis de la literatura publicada sobre ellas, puede extraerse la conclusión de que la comunidad científica está de acuerdo en

que las características comunes a éstos indicadores son las de parcialidad, convergencia y relatividad, que enunciaron Martin e Irvine en sus trabajos (MARTIN, 1996; MARTIN y IRVINE, 1983).

La parcialidad y la convergencia son características relacionadas, que se refieren a la capacidad de los indicadores para medir cuestiones puntuales, de forma que cada uno de ellos cuantifica un aspecto concreto de la actividad científica, por lo que se podría pensar que no son válidos para su análisis global. Sin embargo, hay que tener en cuenta su capacidad de convergencia, mediante la cual un conjunto de indicadores parciales, orientados cada uno de ellos a rasgos puntuales, son capaces de medir, en su conjunto, características mucho más complejas de la actividad científica.

De esta forma:

los resultados basados en la convergencia de varios indicadores (preferiblemente incluyendo la evaluación por pares) son probablemente más fiables que unos [resultados] basados en un único indicador o tan sólo en la revisión por pares (MARTIN, 1996).

La tercera característica de los indicadores es que son relativos, es decir, que los resultados que se obtienen a partir de éstos sólo son aplicables a un ámbito determinado (un área científica, una institución, etc.) y no son extrapolables a otras áreas. En este sentido, uno de los ejemplos más ilustrativos de esta relatividad es el Factor de Impacto. Éste indicador, propuesto por GARFIELD (1972) en el contexto de las bases de datos del Institute for Scientific Information (ISI), está orientado específicamente para medir la repercusión que generan las revistas de Ciencia y Tecnología, por ello cuenta con una ventana de citación de dos años.

Si la ventana de citación es el periodo durante el cual se contabilizan el máximo de citas recibidas, no es de extrañar que con un periodo tan corto (2 años) el Factor de Impacto se revele insuficiente para su aplicación en el área de Ciencias Sociales y totalmente inútil en el área de Humanidades, dónde el periodo durante

el cual un trabajo publicado recibe el máximo de citas puede ser superior a diez años.

Debido a estas características de los indicadores bibliométricos, es imprescindible tomar ciertas medidas a la hora de su aplicación en cualquier estudio:

- En primer lugar es necesario determinar cuáles son los aspectos que se quieren medir, y en función de estos y del área científica que se vaya a estudiar, determinar los indicadores más apropiados para ello, teniendo en cuenta siempre que se debe aplicar una batería, pues como se ha observado, cada uno nos aportará una pequeña porción de resultados, que sólo en su conjunto tendrán sentido para medir aspectos complejos de la actividad científica.
- En segundo lugar, los resultados obtenidos a través de un grupo de indicadores bibliométricos deben ser cuidadosamente interpretados, pues sus valores absolutos no son indicativos por si mismos, sino que se tienen que comparar con los de otros grupos (OKUBO, 1997).
- Como tercer punto, a la hora de interpretar los resultados obtenidos, es determinante tener en cuenta el área de estudio y ceñirse a ella, pues los indicadores obtenidos han sido elegidos específicamente para ella y sus resultados en ningún caso pueden extrapolarse a otras áreas diferentes.

Como última característica, cabe señalar que debido al carácter social que poseen los estudios bibliométricos, con frecuencia las distribuciones estadísticas que se derivan de ellos no se adaptan a las tradicionales distribuciones gaussianas (normal, binomial, Poisson, etc.), si no que se asemejan más a las distribuciones zipfianas, estrechamente relacionadas con las leyes de potencia que rigen los estudios bibliométricos (leyes de Zipf, Bradford y Lotka principalmente) (WHITE y MCCAIN, 1989).

Por otro lado, ya en 1987 KING apuntó cual debía ser el futuro de los indicadores bibliométricos, indicando que debían ser elaborados de forma semi-automática,

rutinaria y con un coste relativamente bajo, además de ser generados de forma dinámica, en línea, y utilizando bases de datos creadas de forma consistente para este objetivo. Además, señalaba que los indicadores deben ser fiables, aplicables independientemente del campo de estudio, deben proveer información que no pueda ser obtenida de otra forma y deben ser también inteligibles para cualquier usuario, independientemente de si se trata de un experto en la materia o no. En este sentido, en España han surgido algunas iniciativas interesantes, entre las que se encuentran el SCImago Journal Country Rank (SJR) elaborado por el grupo SCIMAGO (2007) junto con la base de datos bibliográfica Scopus, que proporciona indicadores relativos a la actividad científica de los países, así como una herramienta para la evaluación de las revistas científicas, alternativa al extendido Journal Citations Report (JCR) de Thomson-Reuters.

Además, desde el último año, se cuenta con el Observatorio IUNE de la actividad científica de las universidades españolas, creado por la ALIANZA 4 UNIVERSIDADES (2012), con el objetivo de “desarrollar una batería de variables e indicadores de I+D+i que permitan determinar con precisión la actividad de las universidades públicas y privadas desde distintas vertientes, como son: la relativa al profesorado, al reconocimiento científico, a la actividad investigadora, a la innovación, a la capacidad competitiva, y a la actividad formativa de investigadores”. Cabe destacar que esta iniciativa no intenta asimilarse a los diferentes rankings relativos a las universidades, tanto nacionales como internacionales, que pretenden posicionar a las diferentes instituciones de educación superior en base a uno o varios indicadores unificados. Esta última metodología supone reducir en demasía las dimensiones del análisis, por lo que ha recibido severas críticas, mientras que el Observatorio IUNE tiene un objetivo muy diferente de acuerdo al cual ofrece una amplia batería de indicadores, “lo que permite evidenciar los puntos fuertes de cada una de las universidades y definir diferentes perfiles de instituciones en función de la dimensión que se quiera analizar” (SANZ-CASADO, 2012).

Por último, WHITE y MCCAIN (1989) en su trabajo señalaron la importancia de los estudios bibliométricos a través del concepto que denominaron “barrera de

la lectura”. Este concepto surge a partir de la limitación por parte de cualquier científico de conocer en su totalidad la literatura publicada en su área de interés, puesto que por mucho tiempo y esfuerzo que le dedique, dicha barrera siempre existirá y siempre será infranqueable para él. Sin embargo, los bibliómetras trabajan al otro lado de la barrera y son capaces de superarla, gracias a que su interés se centra en el análisis de los elementos formales de dicha literatura científica, buscando formas de describirla por medio de algoritmos a través de sistemas informáticos más que de sistemas del conocimiento humano, que permiten analizar las características bibliográficas de los trabajos científicos.

1.3.1.2. Limitaciones de los indicadores

Como se ha podido observar, los indicadores bibliométricos poseen algunas características que obligan a ser especialmente cuidadosos a la hora de aplicarlos, pero además, a lo largo de la historia de la Bibliometría, se han ido determinando las limitaciones en el uso de indicadores y su problemática.

La primera limitación a considerar ya se ha apuntado, y está estrechamente relacionada con las características de parcialidad, relatividad y convergencia de los indicadores. Se trata del principio estadístico de que los resultados obtenidos a través de cualquier indicador no tienen por que ser válidos para cada uno de los datos concretos utilizados para su cálculo, sino que su validez hace referencia al conjunto de los mismos. Por ésta razón OKUBO (1997) destaca que “el análisis debe incorporar un volumen de datos tan grande como sea posible, para permitir una compensación estadística para cualquier tendencia que afecte a cada pequeña entidad tomada por separado”, por lo que también será recomendable contar con una batería de indicadores que permitan caracterizar dicho conjunto de datos, tal y como señalan MORAVCSIK y PÉREZ ALVAREZ-OSSORIO (1989):

Como el sistema que estamos evaluando es siempre multidimensional y no puede caracterizarse sin ambigüedad mediante un indicador simple, y, también, porque cada dimensión no se puede caracterizar tampoco con un solo indicador, debemos esperar que el resultado de

todo proceso de evaluación sea un indicador compuesto, esto es, un conjunto de indicadores individuales. Únicamente todo el conjunto de éstos proporciona la información que buscamos, y debemos resignarnos a admitir que es imposible sustituir dicho conjunto de indicadores por uno sólo, de modo que un simple número expresase el resultado de la evaluación.

En éste y otros trabajos (OKUBO, 1997; SANCHO LOZANO, 1990) se destaca otra de las grandes limitaciones del uso de indicadores métricos para medir la actividad científica, y es que estos indicadores solo se aplican a la actividad científica que da lugar a publicaciones, pero no tienen en cuenta otras formas de comunicación científica. De nuevo OKUBO (1997), hace la siguiente observación: “La comunicación informal no ha sido incorporada y probablemente nunca lo será”. Sin embargo, trabajos más recientes han demostrado que se pueden realizar con éxito estudios que permitan medir la comunicación informal entre científicos. Concretamente se han llevado a cabo estudios de las relaciones informales entre científicos a través del análisis de enlaces hipertextuales (BORGMAN y FURNER, 2002).

En este sentido cabe destacar que la webmetría está abriendo nuevas vías de investigación, que permiten el estudio de las relaciones menos visibles entre científicos o instituciones, gracias a la cuantificación de enlaces hipertextuales que pueden considerarse una nueva forma de “cita bibliográfica”, que comienza a imponerse gracias a los nuevos medios de comunicación científica.

Además, se han propuesto indicadores para medir las relaciones entre instituciones académicas e instituciones comerciales, entre ellos, indicadores como el número de participaciones por parte de profesores universitarios en conferencias estrictamente profesionales, el número de apariciones de profesores universitarios en los medios de comunicación, el número de estudiantes de posgrado financiados directamente por instituciones privadas, el número de instituciones privadas que colaboran en proyectos de investigación dirigidos por instituciones académicas, etc., (MOLAS-GALLART et al., 2002). Todos ellos, combinados con indicadores de actividad científica, nos permiten medir la influencia de las relaciones informa-

les entre miembros de las instituciones académicas y las instituciones comerciales.

Para GÓMEZ CARIDAD y BORDONS (1996) uno de los principales inconvenientes de los indicadores bibliométricos reside en los sesgos de las bases de datos. Concretamente las autoras se refieren a las bases de datos SCI (Science Citation Index) y SSCI (Social Sciences Citation Index) de Thomson-Reuters (antes ISI), de las cuales mencionan tres grandes sesgos que condicionan su utilidad:

1. La falta de representación de países no anglosajones entre las revistas indizadas
2. La mayor representación (y por tanto visibilidad) de las revistas orientadas hacia la ciencia básica, en detrimento de las que recogen trabajos sobre ciencia aplicada
3. Recogen un mayor número de revistas que publican sobre temas de interés internacional, por lo que los temas de interés local están muy poco representados

Pese a que estos problemas de cobertura en las bases de datos bibliográficas se encuentran lejos de desaparecer, sus responsables están llevando a cabo esfuerzos para mejorar su alcance, incluyendo entre las publicaciones indizadas, revistas no anglosajonas, de ciencias aplicadas o que respondan a intereses más localistas. Si bien es cierto que esta respuesta ha llegado después de la aparición de productos altamente competitivos y respaldados por grandes multinacionales, como es el caso de Google Scholar o Scopus, que han hecho frente a la tradicional hegemonía de las bases de datos de Thomson-Reuters como principal herramienta bibliográfica, si bien diferentes estudios han demostrado que Scopus y la Web of Science son productos comparables y que ofrecen una cobertura similar (MEHO y YANG, 2007; BAR-ILAN et al., 2007; BAR-ILAN, 2008b), mientras que en el caso de Google Scholar los trabajos de TORRES-SALINAS et al. (2009), CABEZAS-CLAVIJO y TORRES-SALINAS (2012) han demostrado que se trata todavía de un producto inmaduro e incapaz de competir con Scopus o la Web of Science a la hora de lle-

var a cabo estudios para la evaluación de la actividad científica, desaconsejando totalmente su uso con este fin.

Por otro lado, y volviendo al trabajo de OKUBO (1997), cabe destacar la gran difusión que han experimentado los estudios métricos de la información en los últimos años, y cómo ha aumentado el número de trabajos bibliométricos e incluso de herramientas que permiten el cálculo de indicadores de forma automática. Esto ha derivado, en muchos casos, en mala praxis y en el uso de indicadores de forma aislada y descontextualizada. En este sentido debe hacerse mención al trabajo de LOPEZ PIÑERO y TERRADA (1992a) en el que destacan que:

La función de los indicadores de actividad científica no puede limitarse a amontonar datos estadísticos yuxtapuestos. Hay que integrarlos para conseguir explicaciones lo más sólidas que sea posible acerca de las actividades relacionadas con la ciencia, en general y en sus aspectos concretos.

Finalmente se debe destacar el papel que juegan los indicadores bibliométricos a la hora de desarrollar la política científica de países, instituciones investigadoras, universidades, etc. Y es que los indicadores bibliométricos se han convertido en las “reglas del juego” que permiten la toma de decisiones relacionada con la evaluación de la Ciencia, desde la aprobación y financiación de proyectos de investigación hasta la evaluación de los investigadores y docentes por parte de las diferentes agencias de acreditación, con el fin de determinar su valía para desempeñar su labor docente e investigadora. Este hecho se deriva de que los indicadores bibliométricos permiten cuantificar la actividad científica, su impacto y potencial calidad, haciéndolos muy útiles para la toma de decisiones, puesto que además suponen un alto retorno de la inversión (en forma de una mejor optimización de los recursos), son relativamente baratos de obtener y proporcionan una información objetiva de gran calidad acerca de la actividad científica (WHITE y MCCAIN, 1989).

Esta capacidad de los indicadores bibliométricos para servir a la toma de decisiones en materia de política científica los ha hecho muy útiles de cara a rentabilizar

la inversión de capital, especialmente en épocas de crisis económica como la ocurrida en los años 70 y 80 o la que actualmente se está desarrollando (MARTIN y IRVINE, 1983). Esta vertiente de los indicadores ha derivado en el peligro de su manipulación y tergiversación, sobre todo si la toma de decisiones se centra en un único indicador.

En este sentido MARTIN (1996) hace referencia al Principio de Indeterminación o Incertidumbre de Heisenberg y señala: *if you measure a research system, you disturb it*, indicando que la evaluación de un sistema de investigación, por medio de cualquier metodología, implica su perturbación, debido a que los diferentes actores (revistas, autores, universidades, etc.) tratan de adaptarse lo mejor posible al sistema de evaluación, desvirtuándolo en algunos casos. Tal es el caso del Factor de Impacto, utilizado a menudo de forma tendenciosa, irresponsable o simplemente inexperta (CAMÍ, 1997; ALFONSO et al., 2005) y que algunas veces también trata de ser manipulado por estos actores para su propio beneficio (NEUBERGER y COUNSELL, 2002; DONG et al., 2005).

Sin embargo, los estudios bibliométricos, realizados con una metodología sólida y objetiva y por expertos que dominen la aplicación e interpretación de los diferentes indicadores, constituye un área muy útil para evaluar la actividad científica en diferentes ámbitos, ya sea a nivel nacional, institucional o incluso de los grupos de investigación y los propios investigadores, realizándolos de una forma objetiva y teniendo en cuenta las fortalezas y limitaciones de los diferentes indicadores, con el fin de no incurrir en la “impactolatría” de la que habla CAMÍ (1997).

1.3.1.3. Tipos de indicadores

Los indicadores bibliométricos pueden dividirse en diferentes categorías según los criterios que se tengan en cuenta (objeto de estudio, metodología empleada, nivel de agregación, etc.), sin embargo lo más común es comenzar dividiéndolos en función de las variables que pueden medir de forma simultánea. Si se trata de indicadores orientados al estudio de una única variable, sin tener en cuenta las po-

sibles relaciones con otros indicadores, se consideran unidimensionales, mientras que los indicadores multidimensionales permiten trabajar de forma simultánea las distintas variables o las múltiples interrelaciones que pueden ser observadas en los documentos, o en los hábitos y necesidades de información de los usuarios (SANZ-CASADO y MARTÍN-MORENO, 1997).

1.3.1.3.1. Indicadores unidimensionales. Los indicadores de corte unidimensional están orientados a la observación y medición de una variable concreta, pero, como se ha comentado anteriormente, es necesario agrupar los indicadores de forma que su conjunto permita cuantificar aspectos complejos de la Ciencia. En este sentido, por un lado se encuentran quienes establecen tipologías de indicadores en función del objeto de estudio. Éste es el caso de SANCHO LOZANO (1990), que distingue entre Indicadores de calidad científica (opinión de expertos), de actividad científica, conexiones entre trabajos y autores, indicadores de impacto de los trabajos, de impacto de las fuentes o de asociaciones temáticas. Mientras que otros autores prefieren establecer una categoría intermedia en función del método empleado para obtener el valor del indicador. Así, VINKLER (1988), distingue entre indicadores de publicaciones e indicadores de citaciones, o lo que es lo mismo, indicadores de actividad científica e indicadores de impacto respectivamente.

Asimismo, los indicadores de actividad científica pueden dividirse en indicadores de producción, que se basan en el recuento de publicaciones, e indicadores de consumo, que tratan el análisis de las referencias bibliográficas incluidas en las publicaciones.

En este sentido los indicadores bibliométricos de producción se pueden definir como: “[...] medidas basadas habitualmente en *recuentos* de publicaciones, que persiguen cuantificar los *resultados científicos* atribuibles bien a unos *agentes*, bien a agregados significativos de esos agentes” (MALTRÁS, 2003).

Por otro lado, el estudio del uso o consumo de información que realizan los científicos, se lleva a cabo mediante el análisis de las referencias bibliográficas que inclu-

yen en sus trabajos, debido a que la inclusión de una referencia implica su utilidad para el autor citante, puesto que la ha usado, independientemente de que exista cierto número de referencias a otros trabajos que responden a motivos políticos, estratégicos o poco éticos (THORNE, 1977).

Por último, los indicadores de impacto son aquellos que permiten medir, a través del número de citas que reciben los trabajos publicados, la repercusión que tienen dichos trabajos en investigaciones posteriores. Estos indicadores se basan, por tanto, en la premisa de que el número de citas que recibe un trabajo es indicativo de su uso, utilidad y, por tanto, calidad científica, aunque, como se verá en el punto 5.1.2, existen numerosas discrepancias acerca de ello.

1.3.1.3.2. Indicadores multidimensionales. Al contrario que los indicadores unidimensionales, los indicadores multidimensionales, tienen como principal característica que:

Permiten tener en cuenta de forma simultánea las distintas variables o las múltiples interrelaciones que pueden ser observadas en los documentos, o en los hábitos y necesidades de información de los usuarios. [...] Hay diversos tipos de representaciones gráficas que se pueden obtener a partir del análisis multivariante. Probablemente las dos más utilizadas en los estudios de usuarios sean: el análisis de cluster y el escalado multidimensional. En ambos casos, se pueden representar individuos o variables cuya situación en el mapa va a depender de las similitudes que presenten. Para ello, antes de iniciarse el análisis se deben establecer los criterios que se van a aplicar en el estudio, como son: La selección de las variables que se van a utilizar para identificar a los grupos, y la selección de la medida de proximidad entre los individuos (SANZ-CASADO y MARTÍN-MORENO, 1997).

La estadística multivariada o multidimensional, es aquella que tiene en cuenta múltiples variables, de forma simultánea, con el fin de analizar y caracterizar la población de individuos u objetos analizados (HAIR et al., 1999). Por tanto, los indicadores multidimensionales son aquellos que se sirven de las técnicas derivadas de la estadística multidimensional o multivariada y pueden ser aplicados a cual-

quiera de las subdivisiones que se realizan sobre los indicadores unidimensionales: Producción, consumo e impacto.

En palabras de WHITE y MCCAIN (1989), los indicadores multidimensionales (refiriéndose en concreto al escalado multidimensional de autores), derivan comúnmente en “constelaciones” que muestran una vista aproximada y bastante inteligible de las dimensiones estudiadas. También indica que generalmente, la representación bidimensional suele ser suficiente para analizar la distribución y que, en muchas ocasiones, las diferencias encontradas en la distancia que separa los diferentes elementos del análisis, suelen corresponderse con la especialización o con las diferentes formas de abordar la investigación. Además, WHITE y MCCAIN (1989) indican que en múltiples ocasiones se encuentran diferentes factores sociales tales como el idioma, la nacionalidad, la ideología o la simple amistad entre científicos, que llevan a los autores a colaborar a pesar de que aparentemente no poseen intereses científicos en común.

1.3.2. Bibliometría en Ciencias de la Vida y de la Salud

En España, se han realizado estudios métricos para conocer la actividad investigadora en diversos campos científicos, así como aspectos relativos a la investigación científica, que son de gran interés y cada año se incrementan los trabajos que se publican en revistas internacionales. Los que más atraen la atención de los investigadores en bibliometría, son los que se realizan sobre ciencias puras y experimentales, que son analizadas en profundidad y desde distintas vertientes. En éste sentido, uno de los campos tradicionalmente más estudiado ha sido el de la medicina, en el cual se cuenta con multitud de trabajos relativos a la producción y consumo de información por parte del colectivo investigador, así como de las fuentes de información más utilizadas, que en este caso serían las publicaciones periódicas.

En este país existen varios grupos, tanto en universidades como en el Consejo Superior de Investigaciones Científicas (CSIC), que han realizado estudios bi-

biométricos en áreas médicas y biomédicas. En estos estudios es habitual utilizar las dos bases de datos (COSTAS et al., 2008) que también se utilizan en este trabajo. Sin embargo estas publicaciones habitualmente se centran en el estudio de áreas médicas o revistas, muy ocasionalmente en enfermedades concretas (SANZ-CASADO et al., 2007), y mucho menos sobre enfermedades raras, como es el caso del estudio que nos ocupa.

Por otro lado, lo más corriente es que en los estudios bibliométricos se calculen indicadores puramente bibliométricos o cienciométricos para definir cuantitativamente el perfil del área estudiada, pero es relativamente poco común que en algunos estudios se empleen además indicadores sociales y económicos para determinar la influencia de determinados factores no científicos en la investigación. En este sentido los indicadores de salud pública ofrecen una doble vía de estudio.

En primer lugar existen indicadores tales como las tasas de mortalidad y morbilidad que pueden indicar si una investigación intensiva en áreas concretas conlleva una reducción en la mortalidad y morbilidad por causa de las enfermedades estudiadas. En segundo lugar, estas mismas ratios de mortalidad y morbilidad, junto a indicadores de salud pública tales como el gasto sanitario o el número de investigadores que se encuentran trabajando en el sistema sanitario nacional, proporcionan datos acerca del grado de influencia de la mortalidad y morbilidad, si se investiga más en las enfermedades con tasas de morbilidad/mortalidad más altas, así como las fluctuaciones que experimenta dicha investigación en función de las políticas de salud pública, relacionadas con el gasto sanitario y los recursos humanos fundamentalmente (LASCURAIN-SÁNCHEZ et al., 2008).

Sin embargo en la literatura revisada apenas se han encontrado estudios que relacionen la presencia de las asociaciones de enfermos con la producción científica en las enfermedades que padecen (RANGNEKAR, 2005), aunque si existen trabajos que demuestran la influencia que ejercen las organizaciones de pacientes sobre la investigación científica (RABEHARISOA y CALLON, 2002; BOWER, 2005; ARRIS-

CASADO et al., 2007; BAGGOTT y FORSTER, 2008). En este sentido, si se demuestra que existe una correlación entre el número de asociaciones de enfermos que padecen una patología concreta y la producción científica en esta misma enfermedad, es posible determinar cual es el grado de influencia que ejercen las organizaciones de pacientes sobre la actividad científica de los investigadores.

1.3.2.1. Investigaciones bibliométricas en ER

Tal y como se ha mencionado anteriormente, el número de estudios bibliométricos sobre enfermedades raras es realmente pequeño y tan sólo se ha localizado una publicación que intenta abarcar un grupo de enfermedades heterogéneo (ESCUADERO GÓMEZ et al., 2005), pues el resto de estudios se centra en enfermedades concretas (AL-SHAHI et al., 2001; SANZ-CASADO et al., 2007; LÓPEZ-MUÑOZ et al., 2008; CAÑEDO ANDALIA, 2009; RIOU-GOTTA et al., 2009; FRIGOTTO y RICCABONI, 2011). Ahora bien, en los últimos años han aparecido algunos trabajos que estudian grupos de enfermedades raras mayores, tales como el realizado por HANNEMANN-WEBER et al. (2012), centrado en el análisis de los centros especializados en este tipo de enfermedades, o los trabajos sobre enfermedades neurodegenerativas realizados por BISHOP (2010), SERRANO-LÓPEZ y MARTÍN-MORENO (2011).

Sin embargo todos estos estudios tienen algo en común, y es que cada uno de ellos destaca la dificultad que entraña la investigación en este tipo de enfermedades, debido a los factores mencionados en el punto 1.1.3. También están de acuerdo en que realizar estudios bibliométricos que abarquen un gran espectro de enfermedades es realmente complicado debido a la heterogeneidad de las mismas.

El trabajo de ESCUDERO GÓMEZ et al. (2005) es el referente más directo para este proyecto, a pesar de las evidentes diferencias metodológicas, así como de cobertura temporal y geográfica. Sin embargo algunos de sus resultados deberán ser comparados con los que se obtengan en este trabajo, para determinar cuáles son las diferencias entre ambos periodos de estudio.

Entre los resultados más destacados, ESCUDERO GÓMEZ et al., determinaron que las instituciones más productivas eran los hospitales y que:

La revista que más artículos ha publicado es Anales Españoles de Pediatría, resultado que está en estrecha relación con el hecho de que gran parte de las enfermedades analizadas son diagnosticadas y tratadas en la infancia.

Por tanto, el trabajo de ESCUDERO GÓMEZ et al. será utilizado como referente para comparar nuestros resultados acerca de la afiliación institucional de los autores así como de la temática de las revistas más productivas, con los que obtuvieron dichos investigadores, con el objetivo de determinar si la investigación en enfermedades raras continúa siguiendo el mismo patrón o si ha cambiado.

Por último, ESCUDERO GÓMEZ et al. concluyen que:

este mismo análisis dentro de unos años debería ofrecer un panorama muy diferente y esperanzador y podrá servir de indicador de las políticas adoptadas a favor de la investigación en enfermedades raras

Efectivamente en este trabajo es esperable encontrar una situación más halagüeña y comprobar que las diferentes iniciativas a favor de la investigación en enfermedades raras llevadas a cabo por la Unión Europea, han tenido un efecto favorable en la cantidad y calidad de la investigación realizada.

1.3.2.2. Uso de indicadores de salud en estudios bibliométricos

Como se señaló en el punto 1.3.2, algunos estudios bibliométricos han tenido en cuenta la relación entre indicadores sociales y de salud pública y los indicadores bibliométricos tradicionales.

SOMOGYI y SCHUBERT (2005), en su trabajo sobre la diabetes mellitus, apuntan que pueden existir tres posibles patrones de relación entre producción científica y un indicador de salud, como es la prevalencia de casos en una enfermedad concreta. Estos tres patrones serían:

- (1) *A mayor prevalencia mayor intensidad en la investigación (Investigación bajo demanda)*
- (2) *La mayor intensidad en la investigación supone un descenso en la prevalencia (Investigación preventiva)*
- (3) *Sin correlación medible (Investigación y estado de salud no están directamente relacionados)*

SOMOGYI y SCHUBERT concluyen que en función del tipo de país estudiado (desarrollado, en vías de desarrollo, etc.), su sistema de salud, etc. el patrón de relación entre prevalencia y producción científica se ajustará a uno de los tres patrones posibles.

Otros estudios anteriores (RODRIGUES et al., 2000; LASCURAIN-SÁNCHEZ et al., 2008; ORTIZ et al., 2009), han relacionado la producción científica sobre enfermedades o áreas médicas con los datos de mortalidad y morbilidad. Estos trabajos han obtenido resultados similares. ORTIZ et al. (2009) determinaron que en Puerto Rico existía una correlación directa entre la mortalidad causada por el cáncer con el número de trabajos sobre cáncer publicados en las dos principales revistas médicas nacionales, a pesar de que durante el periodo 1960-1990 la mortalidad aumentara proporcionalmente más que la investigación.

LASCURAIN-SÁNCHEZ et al. (2008) determinaron que:

la incidencia de las dos principales causas de muerte en España – enfermedad cardiovascular y cáncer – no se reflejaba en el número de trabajos en Cardiología y Oncología [...] Sin embargo la Neurociencia inducía un gran interés en ambas (universidades y hospitales) comunidades de científicos

Además, en este trabajo se encontró que si que existía una relación más cercana entre la materia de estudio en las investigaciones españolas y las diferentes causas de enfermedad.

Por último, RODRIGUES et al. (2000) estudiaron la relación entre mortalidad, morbilidad y producción científica en Brasil, en tres grupos de enfermedades: cáncer, enfermedades cardiovasculares y malaria. Los resultados obtenidos mostraron una correlación directa entre los tres indicadores, siendo especialmente

destacada en el caso de la malaria que obtenía una proporción sensiblemente mayor de publicaciones por tasas de mortalidad y morbilidad.

A la vista de estos resultados se puede destacar que en aquellas enfermedades menos comunes (nerológicas o malaria) existe una mayor correlación entre el número de enfermos o fallecidos por su causa y la producción científica en dichas enfermedades, mientras que en las enfermedades más comunes (cáncer, enfermedades cardiovasculares, diabetes mellitus) esta relación es menos significativa. Este hecho también ha sido constatado por BISHOP (2010), encontrando en su análisis de las enfermedades nerológicas, que aquellas con una prevalencia más baja despertaban mayor interés en los investigadores que las enfermedades más extendidas.

HARVEY (1907), uno de los descubridores, tras Miguel Servet, de la circulación de la sangre, ya señalaba que las formas más raras de la enfermedad son las que más interés despiertan en los investigadores, puesto que son las que permiten descubrir comportamientos anómalos en el cuerpo humano, e indica que la mayor parte de los descubrimientos médicos se encuentran relacionados con el estudio de estas enfermedades “raras”.

Otro indicador muy utilizado es el gasto público en salud, ya sea destinado a equipamiento o a recursos humanos. En este sentido PESTAÑA y CERDÁN (2000) estudiaron el gasto en equipamiento, concretamente en la adquisición de instrumentos de resonancia magnética y el personal disponible en cada unidad de resonancia magnética en relación con la productividad de las instituciones, principalmente hospitales, universidades y CSIC. La conclusión a la que llegaron fue que:

el rendimiento científico y médico de los instrumentos de resonancia magnética resultaba de la combinación de variables incluyendo el número de instrumentos y algunos indicadores socio-económicos como el personal involucrado en I+D y la capacidad sanitaria reflejada en el personal involucrado en las tareas clínicas.

Por todo ello, se puede concluir que el gasto sanitario es uno de los indicadores que nos van a permitir medir la efectividad de las políticas de salud pública y su eficacia en la investigación en el grupo de enfermedades estudiadas.

Por último cabe destacar que en este trabajo se va a utilizar un indicador que hasta ahora apenas ha sido tenido en cuenta en los estudios que relacionan indicadores sociales e indicadores bibliométricos, y es la presencia de asociaciones de enfermos. En un área como las enfermedades raras, dónde la baja prevalencia hace que los lazos entre grupos de enfermos se estrechen y que se movilicen activamente para llamar la atención de la opinión pública sobre su enfermedad, las asociaciones de enfermos pueden desempeñar un papel fundamental a la hora de despertar el interés en su investigación, tanto por parte de las autoridades sanitarias como de los propios investigadores, por lo que se espera encontrar una correlación positiva entre el número de asociaciones existentes y la producción científica en cada país analizado.

1.4. Justificación científica

A lo largo de la introducción de este trabajo, se han señalado algunos de los motivos para llevar a cabo un estudio bibliométrico sobre esta temática concreta, con objeto de determinar cuáles son las características de la investigación en enfermedades raras (producción, consumo, redes, etc.) y cómo afectan determinados factores sociales como su visibilidad, el gasto sanitario o la presencia de asociaciones de enfermos, a dicha investigación. En este sentido es importante también estudiar si existe relación entre dichos factores sociales y el volumen de la investigación realizada sobre estas enfermedades, al tiempo que determinamos cuáles son los campos en los que más se investiga y que investigadores forman los frentes de investigación.

Entre los motivos que pueden señalarse para llevar a cabo este trabajo se encuentran las políticas sanitarias de la Unión Europea, así como las políticas nacionales

de los diferentes países que la componen, pues en ambos casos han señalado la investigación sobre enfermedades raras como una prioridad en la investigación en el área de Salud.

Del mismo modo, la paradoja de la rareza que se ha mencionado con anterioridad, supone también un aliciente a la hora de estudiar la actividad científica que generan este tipo de enfermedades, con el fin de comprobar si se corresponde con la baja prevalencia que en general poseen estas enfermedades o si, por el contrario, atraen el interés de la comunidad científica tanto o más que otras patologías más comunes.

También se ha considerado como un área de interés debido a la presencia de la Genética como área transversal, común a la práctica totalidad de enfermedades raras y muy especialmente a aquellas cuya causa genética se encuentra claramente identificada. Este motivo se ve acentuado debido al creciente interés por parte de la comunidad científica por la Genética, el descubrimiento de los genes responsables de determinadas enfermedades y las posibles terapias génicas a desarrollar, especialmente desde que la investigación sobre el mapa del genoma humano comenzó a dar sus frutos.

Finalmente, al margen del innegable interés que despierta la investigación sobre enfermedades raras, el desarrollo de este trabajo bibliométrico sirve también como justificación para el desarrollo de un sistema de normalización de datos, orientado fundamentalmente a la normalización de los campos de autor y dirección, así como a la asignación automática de términos de un lenguaje controlado sobre registros bibliográficos que no los incorporan a priori. Este sistema de normalización se ha probado y ha permitido el tratamiento de los datos utilizados en este trabajo, conformados por un volumen considerable de registros bibliográficos, demostrando su eficacia y fácil implementación en cualquier tipo de estudio bibliométrico.

Hipótesis

En función de las características de las enfermedades raras presentadas en la introducción, así como de la revisión realizada sobre trabajos anteriores de temática similar, se plantean las siguientes hipótesis:

- Se investiga muy poco sobre cada una de las enfermedades raras, pero en el conjunto de patologías raras, supone un frente de investigación muy importante.
- La investigación en enfermedades raras se ve influenciada por factores externos, tales como la financiación pública de I+D+i o la presencia de asociaciones de enfermos que impulsen la investigación sobre estas enfermedades.
- La Genética ejerce una gran influencia en la investigación realizada sobre enfermedades raras, y sirve como nexo de unión entre las mismas, especialmente a partir de la secuenciación del Genoma Humano.
- En las enfermedades raras es muy importante el diagnóstico precoz, por lo que la Pediatría será una de las temáticas más presentes en este tipo de investigación.

- Existen dificultades patentes para publicar trabajos sobre enfermedades raras en revistas de alto impacto, por lo que puede verse perjudicada su visibilidad.
- La colaboración es un factor de gran importancia en el tipo de trabajos analizados, ya que la baja prevalencia de este grupo de enfermedades hace necesaria la cooperación con otros centros investigadores.

Sección 3

Objetivos

Como objetivo principal se plantea analizar la actividad científica que se realiza sobre las enfermedades raras, en particular aquellas con causa genética identificada, y conocer a los autores de dicha actividad, así como sus hábitos de uso y consumo de información. Para conseguir dicho objetivo general, se plantean los siguientes objetivos específicos:

- Automatizar la normalización de los campos de autor e institución, así como desarrollar un sistema de asignación de términos que permita completar la descripción temática de los registros bibliográficos.
- Analizar la actividad científica que ha dado lugar a publicaciones, tanto en el ámbito nacional como internacional, así como su evolución temporal.
- Determinar cuáles son las publicaciones más importantes y su producción científica, realizando una comparación con las publicaciones más citadas y comprobando si existe relación entre ambas.
- Estudiar los hábitos de uso de información de los investigadores en éste área, para conocer las áreas de mayor interés para ellos y la inmediatez de la información que consumen.

- Conocer los patrones de citación de los investigadores, así como las relaciones que se establecen entre ellos por medio de los análisis de cocitaciones.
- Definir los patrones de producción científica en enfermedades raras por parte de los países pertenecientes a la Europa de los quince, su evolución y su caracterización por medio de indicadores bibliométricos unidimensionales y multidimensionales.
- Analizar las áreas de investigación más importantes en el desarrollo de la actividad científica sobre enfermedades raras, sus características y evolución.
- Determinar cuáles son los frentes de investigación en el área, estudiando a los autores que han publicado trabajos más citados y con mayor visibilidad y las instituciones en las que trabajan.
- Determinar los patrones de colaboración presentes entre los países, instituciones y autores que desarrollan su actividad científica en este área, determinando también el tamaño y dinámica de los grupos de investigación, a través de los indicadores tradicionales de colaboración así como la obtención de indicadores conexionistas.
- Plasmar, mediante el uso de indicadores multidimensionales, las relaciones entre los diferentes actores relacionados con la investigación en enfermedades raras, así como su distribución en torno a las diferentes clasificaciones temáticas de los trabajos publicados.
- Analizar cuáles son los factores sociales que suponen una mayor influencia sobre la actividad científica de estas enfermedades, principalmente a través del gasto sanitario que realizan los países analizados y de la presencia en los mismos de asociaciones de enfermos que ejerzan como impulsores de la investigación.

Capítulo II

Metodología

Sección 4

Fuentes, recogida y tratamiento de los datos

En este apartado se detalla el método que se ha seguido en el desarrollo de este trabajo para obtener los resultados propuestos en los objetivos.

4.1. Clasificación de las Enfermedades Raras

Para el desarrollo de este trabajo, era necesario contar en primer lugar con una clasificación de enfermedades, que permitiera determinar cuáles de ellas formará parte de la muestra. Con este objetivo, se presentan a continuación los principales sistemas de clasificación de enfermedades y los motivos para utilizar la International Classification of Diseases, en adelante ICD.

La clasificación de las enfermedades ha constituido siempre un pilar fundamental de la teoría y la práctica médicas. Para la medicina clínica conocer es diagnosticar. Y diagnosticar es encuadrar un padecimiento individual dentro de una clasificación preestablecida de las enfermedades posibles (SÁNCHEZ, 1991).

La clasificación de las enfermedades expresa el resultado de la observación directa de las mismas, así como del intento de dotarles de un orden que permita el

registro sistemático, el análisis, la interpretación y la comparación de datos de mortalidad y morbilidad recolectados en diferentes países o áreas, y en diferentes épocas. Precisamente por este afán sistematizador es interesante utilizar alguno de los múltiples sistemas de clasificación de enfermedades disponibles para determinar la muestra de enfermedades que comprende nuestro trabajo, facilitándonos además enormemente el diseño de las estrategias de búsqueda en las bases de datos bibliográficas y permitiéndonos establecer comparaciones entre la situación de los diferentes grupos de patologías estudiadas.

En este sentido, existen principalmente tres sistemas de clasificación de enfermedades, si bien uno de ellos no lo es realmente, pero ayuda a dicha clasificación. De entre ellos, ha sido necesario seleccionar el más adecuado para determinar los conjuntos de enfermedades sobre las que realizar nuestro estudio. Estos sistemas de clasificación son la International Classification of Diseases (ICD), el Medical Subject Headings (MeSH) y el catálogo Mendelian Inheritance in Man (MIM).

4.1.1. La Clasificación Internacional de Enfermedades (ICD)

Desde que en 1893 se adoptara la Clasificación Internacional de Causas de Muerte (WORLD HEALTH ORGANIZATION, 2009) se ha trabajado en mejorar el sistema e ir adaptándolo a los cambios y a la aparición de nuevas patologías. Actualmente la International Classification of Diseases (ICD) se encuentra disponible en su décima versión, actualizada en 2007. Este sistema de clasificación pasó a depender de la Organización Mundial de la Salud (World Health Organization) en 1949, al llegar a su sexta versión, y hoy en día, es la más utilizada en epidemiología, tanto a nivel de investigación como de la clasificación de historias clínicas o para generar las estadísticas de morbilidad y mortalidad.

Sin embargo la ICD presenta algunos problemas que dificultan su utilidad a la hora de clasificar determinadas enfermedades, entre ellas las raras. La primera dificultad encontrada es que:

no tiene un eje de clasificación definido. De hecho en algunos casos, el eje es etiológico, en otros es anatómico [...]

(RODRÍGUEZ JÚSTIZ et al., 2000).

Por esta razón es habitual que existan cambios significativos entre las diferentes versiones de la ICD, como ocurrió por ejemplo al incluir el SIDA en la ICD-10, con la consecuencia de que se viera incrementado en un 3,6 % el número de muertes clasificadas en la sección de “enfermedades infecciosas y parasitarias”, en detrimento de las “enfermedades endocrinas, nutricionales y metabólicas”, categoría en la que se incluían las muertes derivadas del VIH mientras se encontró vigente la ICD-9 (RUIZ et al., 2002).

Por otro lado el otro gran problema de la ICD reside en su extensión. A medida que la epidemiología avanza, se codifica un número mayor de enfermedades y síndromes, así como sus diferentes variantes y síntomas, por lo que la extensión de la ICD ha ido creciendo de forma notable pasando de 1.178 categorías en la ICD-9 a 2.032 en la ICD-10. A pesar de ello, no responde a las necesidades de las diferentes especialidades médicas y, a la vez, no es tampoco adecuada para la clasificación en atención primaria, dónde deben detectarse los primeros síntomas patológicos.

Por último RODRÍGUEZ JÚSTIZ et al. (2000), destacan que la ICD-10:

tiene importantes limitaciones para estudios epidemiológicos por seguir un esquema puramente 'biologicista', no ofreciéndose como instrumento para un análisis global de los problemas de salud.

Por todo lo anterior, actualmente se está trabajando en una reorganización de la ICD que acabe con estos problemas o al menos los reduzca al mínimo. Esta solución consiste en la división de la ICD en familias para las diferentes especialidades a partir de un núcleo común (core-ICD) que hiciera posible su vertebración. Estas familias incluirían una clasificación orientada a la Atención Primaria y otras más específicas por cada una de las grandes áreas médicas, permitiendo así resolver los problemas derivados del tamaño del sistema actual (demasiado extenso para la Atención Primaria y demasiado reducido para las diferentes especialidades).

4.1.2. Sistema de clasificación MIM (Mendelian Inheritance in Man)

Otro de los sistemas más importantes de clasificación de enfermedades es el denominado Mendelian Inheritance in Man (MIM). Este sistema, desarrollado por Victor McKusick consiste en un catálogo que codifica enfermedades con causa genética identificada, así como los genes implicados en ellas, sus funciones, variaciones (mutaciones y polimorfismos), genotipos, fenotipos y alelos. (PEARSON et al., 1994). La primera versión del catálogo se editó en 1966 bajo el título *Mendelian Inheritance in Man: Catalogs of Autosomal Dominant, Autosomal Recessive, and X-linked Phenotypes* (MCKUSICK, 1966) y tras sucesivas versiones acabó derivando en la que actualmente se conoce como *Mendelian Inheritance in Man: A Catalog of Human Genes and Genetic Disorders* (MCKUSICK, 1998).

Hoy en día, el catálogo se encuentra disponible en línea (OMIM) integrado en el sistema de bases de datos del National Center for Biotechnology Information, consiguiendo con ello un valor añadido de gran importancia, pues gracias a que sus más de 190 referencias se encuentran cruzadas con la base de datos Pubmed, se pueden realizar búsquedas en ella utilizando como descriptores los códigos del catálogo OMIM y de la misma forma, a partir de las referencias bibliográficas de Pubmed obtener los códigos MIM relacionados con ellas. Sin embargo el mayor problema que presenta éste catálogo es que:

Ningún gen opera en un vacío; cada gen interactúa directamente o indirectamente a través de las proteínas que codifica con muchos otros genes y sus productos. Esto explicaría las marcadas variaciones en los síntomas de pacientes con la misma enfermedad (PASQUALINI, 2001).

En este sentido, cuando se realiza una búsqueda a través del catálogo OMIM, éste incluye también los genes relacionados con las enfermedades incluidas en la estrategia de búsqueda, y por tanto, registros bibliográficos que no se encuentran relacionados directamente con las enfermedades que buscamos.

Por ello, como se verá más adelante, el catálogo MIM, aunque es de gran utilidad para la clasificación de enfermedades, puede no ser el sistema más adecuado para este trabajo.

4.1.3. Medical Subject Headings (MeSH)

Por último se debe mencionar el sistema de clasificación de la National Library of Medicine (NLM) el Medical Subject Headings (MeSH). Este sistema no es estrictamente un sistema de clasificación de enfermedades sino un tesoro diseñado para ser utilizado en la indización de publicaciones en ciencias de la vida y la salud, siendo su principal utilidad la de servir como lenguaje controlado para la indización de los contenidos recogidos en la base de datos Pubmed.

Puesto que el MeSH es un tesoro, se compone de una parte alfabética y una jerárquica, compuesta a su vez de 16 categorías principales, entre las que se incluye una dedicada específicamente a enfermedades (tabla 4.1).

Categorías MeSH
Anatomy
Organisms
Diseases
Chemicals and Drugs
Analytical, Diagnostic and Therapeutic Techniques and Equipment
Psychiatry and Psychology
Biological Sciences
Physical Sciences
Anthropology, Education, Sociology and Social Phenomena
Technology and Food and Beverages
Humanities
Information Science
Persons
Health Care
Publication Characteristics
Geographic Locations

Tabla 4.1: *Categorías MeSH*

Sin embargo a pesar de contar con una categoría diseñada especialmente para la indización de enfermedades, este sistema no es utilizado de forma habitual para clasificar tipologías de enfermedades, sino para clasificar la producción científica sobre dichas enfermedades.

Por lo tanto, aunque no es posible basarse en este sistema para establecer una muestra representativa de enfermedades raras sobre la que llevar a cabo este estudio, si que será de gran utilidad a la hora de realizar las estrategias de búsqueda necesarias para recuperar las publicaciones científicas sobre las que se basará nuestro trabajo.

4.1.4. La utilidad de las clasificaciones de enfermedades

Una vez presentadas las diferentes clasificaciones, sus puntos fuertes y sus debilidades, se puede determinar que el sistema que más se ajusta a nuestras necesidades es la ICD, puesto que, aunque no es el mejor sistema posible si es el más adecuado entre las opciones disponibles, pues a diferencia del MIM, no se corre el riesgo de mezclar enfermedades raras con enfermedades comunes (debido a la coocurrencia de genes) y, a diferencia del MeSH, si que es un sistema utilizado habitualmente para codificar enfermedades en los centros de salud y, por tanto utilizado también para calcular los indicadores de morbilidad y mortalidad que se utilizarán en este estudio.

Por otro lado, aunque no se vayan a utilizar los otros dos sistemas para determinar la muestra de enfermedades siempre se podrán obtener unos resultados más completos si se combinan los nombres de las enfermedades con el sistema MIM para abarcar todo el espectro de enfermedades.

Por último, se debe destacar que lo ideal para realizar las búsquedas documentales sería que todas las referencias que trataran sobre alguna enfermedad rara estuvieran clasificadas bajo el encabezamiento “rare diseases” del MeSH, pero hoy en

día es inviable, pues el número de registros indizados bajo este encabezamiento es realmente pequeño.

A continuación se va a describir la forma en que se va a emplear cada una de las clasificaciones mencionadas en este trabajo y cual es su utilidad.

4.1.4.1. Acotar por grupos de enfermedades

En primer lugar, se va a utilizar la ICD-10 para determinar la muestra, algo que se tratará en el punto 4.2. Sin embargo el uso de esta clasificación va más allá, en el sentido de que contar con una muestra de enfermedades clasificadas va a permitir estudiar por separado diferentes grupos de enfermedades, y comparar la investigación que se lleva a cabo en cada uno de ellos con todos los demás.

Para determinar estos grupos se van a utilizar las categorías de la ICD-10, de forma que permitan establecer grupos de enfermedades en función de su tipología como por ejemplo: Enfermedades endocrinas y nutricionales (clasificadas en el rango E16 a E56), Trastornos mentales y del comportamiento (rango F01 a F84), diferentes trastornos metabólicos, etc.

4.1.4.2. Mejorar la recuperación a través de la clasificación

En segundo lugar el sistema de clasificación MIM, a pesar de no ser adecuado para determinar la muestra y llevar a cabo las estrategias de búsqueda principales, si que resulta de gran utilidad para mejorarlas.

Concretamente la versión en línea del catálogo MIM (OMIM), permite realizar búsquedas por enfermedades y devuelve como resultado información acerca de la enfermedad, sus sinónimos, siglas, genes que intervienen en ella, etc., lo que resulta de gran utilidad para refinar las sentencias de búsqueda y conseguir una mayor precisión y exhaustividad en los resultados.

4.1.4.3. Comparar con la investigación global en Enfermedades Raras

Por último, ya se ha comentado que el Medical Subjects Headings (MeSH) cuenta con un término para “rare diseases” que sería el más adecuado para realizar las búsquedas siempre y cuando se utilizara de facto para la clasificación de registros bibliográficos. En el momento de realizar la descarga de datos (mayo de 2010), al realizar un búsqueda mediante este término en Pubmed se obtienen un total de 2.256 registros, un número demasiado pequeño como para obtener resultados fiables y relevantes. Sin embargo, si en el futuro este término fuese empleado por Pubmed para clasificar todos aquellos trabajos que tratasen al menos sobre una enfermedad rara, si que podría utilizarse para realizar una comparativa de la situación de las enfermedades seleccionadas en la muestra con el total de registros clasificados bajo el término “rare diseases” del MeSH. Esta comparativa mostraría las diferencias entre la investigación en enfermedades raras con causa genética identificada y aquellas que no la tienen, y es de suponer que existirá una mayor carga de este tipo de investigación en aquellas patologías que tienen un claro componente genético. Aunque esta comparativa está supeditada a que exista una actualización en los descriptores asignados a los registros de Pubmed.

4.2. Selección de la muestra de enfermedades

A la hora de desarrollar un análisis bibliométrico cómo el que se presenta, es de vital importancia determinar los documentos con los que se va a trabajar. En nuestro caso debe delimitarse la selección de datos en tres sentidos: Cobertura temporal, cobertura geográfica y cobertura temática.

Puesto que se pretende realizar un estudio que permita observar la evolución y tendencias de la investigación en enfermedades raras será necesario cubrir un periodo de tiempo lo suficientemente amplio. Bajo este punto de vista podría entenderse que la mejor opción sería realizar un estudio retrospectivo sobre la investigación en enfermedades raras utilizando la cobertura temporal lo más am-

plia posible dentro de los límites de las bases de datos bibliográficas. Sin embargo también se debe tener en cuenta que se pretende combinar los clásicos indicadores bibliométricos con indicadores sociales que permitan establecer un perfil claro, no sólo de la investigación en enfermedades raras sino también de sus motivaciones y de los factores sociales que más influyen en ella. Por tanto, y debido a que la disponibilidad de dichos indicadores sociales es más escasa, se decidió acotar el estudio geográficamente a la Unión Europea, y puesto que las primeras medidas para fomentar la investigación en enfermedades raras en la Unión Europea datan de 1999 (UNIÓN EUROPEA, 1999, 2000), se decidió utilizar como alcance temporal los trabajos publicados desde el año 2000 hasta el 2008 y como alcance geográfico la producción de los países que en ese momento componían la Unión Europea, es decir, el conjunto de países conocido como “la Europa de los 15” y que se compone de:

Bélgica	Alemania	España
Francia	Irlanda	Portugal
Italia	Reino Unido	Austria
Luxemburgo	Dinamarca	Finlandia
Países Bajos	Grecia	Suecia

En cuanto a la cobertura temática, como ya se ha mencionado, existe una gran cantidad de Enfermedades Raras (entre seis y siete mil, según ORPHANET (2011b)), por lo que es necesario seleccionar un número más reducido de ellas para llevar a cabo nuestro estudio. Debido a que una gran cantidad de patologías poseen una predisposición o base genética, una selección de las mismas, compuesta por aquellas enfermedades que cuenten con una causa genética totalmente identificada, se puede considerar muestra suficientemente representativa del conjunto de Enfermedades Raras. Este conjunto de enfermedades se ha obtenido a través del Instituto de Investigación de Enfermedades Raras de Base Genética (Ver apéndices, tabla A.1).

4.2.1. Instituto de Investigación de Enfermedades Raras de Base Genética (INERGEN)

INERGEN (2000) surge como una institución virtual que se propone reunir en una unidad operacional, a expertos relacionados con los distintos aspectos de las enfermedades raras de base genética. Pese a ello, actualmente el proyecto se encuentra interrumpido y su página web no es accesible. El instituto surgió dentro de las Redes Temáticas de Investigación Cooperativa Sanitaria (RETICS) como centro de referencia y de formación para cualquier grupo clínico o de investigación implicado en el manejo de enfermedades de origen genético.

4.2.1.1. Enfermedades con causa genética identificada

El Instituto de Investigación de Enfermedades Raras de Base Genética o INERGEN cuenta con una base de datos compuesta por aquellas Enfermedades Raras que poseen una causa genética identificada. Además, la mayor parte de las enfermedades contenidas en esta base de datos se encuentran codificadas mediante la Clasificación Internacional de Enfermedades en su décima versión (ICD-10) y el sistema MIM, por lo que se decidió tomar éstas 252 enfermedades como la base para nuestro trabajo (Ver apéndice A.1).

4.3. Fuentes de información

Tratándose de un estudio bibliométrico en el que el objetivo es comparar los datos de producción y consumo de información con el gasto en salud pública y la presencia de asociaciones de enfermos, para conocer tanto el perfil de la investigación como los factores sociales que más intervienen en ella, es necesario utilizar distintos tipos de fuentes de información:

4.3.1. Bases de datos bibliográficas

Se van a utilizar las bases de datos Pubmed y Science Citation Index (SCI) de Thomson Reuters. Pubmed, debido a que está especializada en medicina y porque recoge la práctica totalidad de los trabajos publicados en este área, y SCI debido a la necesidad de contar con datos acerca del consumo de información científica. Se descartó la utilización de bases de datos nacionales, como el Índice Médico Español (IME), debido en primer lugar a que se trata de un estudio de marcado carácter internacional, y en segundo lugar porque las bases de datos nacionales no ofrecen información lo suficientemente actualizada y completa como para ser de utilidad en este estudio.

4.3.2. Bases de datos estadísticas sobre salud

Para obtener los indicadores sociales y de salud pública van a utilizarse bases de datos estadísticas. Los datos se van a obtener principalmente de las bases de datos de indicadores de la Organización Mundial de la Salud (WHOSIS) y de la Unión Europea (Eurostat). De ellas se van a obtener indicadores tales como gasto sanitario per cápita, morbilidad, mortalidad, etc.

4.3.3. Registro de asociaciones de pacientes

Por último y como medida para determinar la capacidad de la movilización ciudadana para influir en los temas de investigación se va a utilizar el registro de asociaciones de pacientes disponible en ORPHANET (2011b), obteniendo el número de asociaciones de pacientes con enfermedades raras distribuidas por países, de forma que puedan compararse con la producción científica y determinar si influye en ella o no.

4.4. Método de recogida y tratamiento de los datos

Tradicionalmente la obtención y tratamiento de datos de forma masiva ha estado limitada por el acceso a equipos informáticos de gran capacidad que, por añadidura eran costosos, de difícil manejo y prácticamente innaccesibles. Sin embargo, con la aparición de los ordenadores personales y su evolución hacia equipos cada vez más potentes y baratos, la obtención y tratamiento de la información, incluso en grandes conjuntos de datos, se ha vuelto un procedimiento mucho más sencillo, barato y accesible.

Al trasladar este escenario al área informétrica, muestra que, inicialmente se realizaban los estudios bibliométricos prácticamente de forma artesanal, analizando manualmente las referencias bibliográficas. Posteriormente se comenzaron a realizar estudios con herramientas informáticas muy básicas, que generalmente, se correspondían con las que proporcionan los principales servicios de bases de datos bibliográficas, tales como el “Essential Indicators” o el “Citation Report” de Thomson Reuters. Estas herramientas realizan análisis muy limitados de la bibliografía, basados casi exclusivamente en el conteo de publicaciones en diferentes niveles de agregación y en el análisis de citas a un nivel muy básico. En muchos casos esta metodología es la única posibilidad que tienen investigadores, que no cuentan con formación y conocimientos informáticos, de desarrollar análisis de la literatura e información científica.

Sin embargo, cuando los investigadores cuentan con desarrollos informáticos, los análisis se vuelven mucho más potentes y son capaces de gestionar grandes conjuntos de datos que de otro modo, serían totalmente inmanejables. En este sentido, a finales de los años 90, KATZ y HICKS (1997) acuñaron el término “Desktop Scientometrics”, que traducido como “Bibliometría de escritorio” ha sido utilizado para referirse al conjunto de herramientas informáticas que permiten a los investigadores realizar estudios bibliométricos, con grandes conjuntos de datos, de forma sencilla y sin necesidad de poseer conocimientos informáticos avanzados.

KATZ y HICKS (1997) desarrollaron en sus investigaciones un método que combinaba, sobre una arquitectura de software basada en UNIX (Linux/GNU), dos herramientas fundamentales, en primer lugar el lenguaje de programación Perl (Practical Extraction and Report Language), que les permitía procesar los registros obtenidos de las bases de datos Thomson Reuters (entonces ISI). Este lenguaje de programación, combinado con el editor de textos programable “Emacs”, utilizado para normalizar las diferentes variantes de nombre en las instituciones analizadas, les permitió manipular, unificar y analizar un conjunto de 500.000 registros, publicados durante 14 años y obtenidos de la base de datos Science Citation Index. Este método ya había sido anticipado por BROOKS (1987) a través de su Brooks1987, que permitía obtener algunos indicadores básicos relativos a las distribuciones bibliométricas de Bradford y Lotka, así como los índices de concentración de las mismas, sin embargo este software, diseñado en Turbo Pascal, si que requería ciertos conocimientos informáticos para poder hacer uso de él y nunca alcanzó el volumen de registros analizados en el trabajo que llevaron a cabo KATZ y HICKS (1997).

Posteriormente, otros trabajos han profundizado en esta metodología. Tal es el caso de trabajos como el de SANZ-CASADO et al. (2002), que desarrollaron una metodología para el tratamiento de registros bibliográficos obtenidos de diversas bases de datos (ISI, Medline, EconLit, INSPEC, ICYT, IME, ISOC, etc.). Esta metodología, que desarrolla el concepto de Bibliometría de Escritorio, se basa en la utilización de gestores bibliográficos como Procite 5.0, para la obtención de los datos brutos procedentes de las diferentes bases de datos, de forma que pudieran ser procesados posteriormente mediante diferentes herramientas ofimáticas (Microsoft Word, Microsoft Excel, WordPerfect, etc.) y exportados a sistemas de análisis estadísticos como SPSS o Statgraphics, que permitiera la obtención de los indicadores bibliométricos pertinentes.

Por otro lado, el lenguaje Perl de cuyo uso se beneficiaron KATZ y HICKS (1997), ha servido para el desarrollo de metodologías más avanzadas como la de WOUTERS (1999), que desarrolló un conjunto de scripts de Perl que permitieran obte-

ner diferentes indicadores bibliométricos a partir de los registros obtenidos de la base de datos Science Citation Index.

Basándose en la metodología de KATZ y HICKS (1997) y WOUTERS (1999), GARCÍA GONZALEZ (2010) desarrolló en la realización de su tesis doctoral, un método de tratamiento de registros obtenidos de bases de datos bibliográficas, que con la ayuda del sistema de bases de datos MySQL, y Perl, permitiera diseñar, alimentar y gestionar una base de datos relacional a través de la que normalizar, extraer y procesar los datos bibliográficos, con el objetivo de calcular cada vez un número mayor de indicadores con un volumen también mayor de registros procesados.

Actualmente, esta metodología ha permitido el desarrollo de herramientas importantes para la evaluación de la actividad científica en España, especialmente el Observatorio IUNE, desarrollado por la ALIANZA 4 UNIVERSIDADES (2012), que maneja un volumen ingente de registros bibliográficos y que, además, permite combinar los datos puramente bibliométricos con otros indicadores sociales, aportándole un valor añadido al análisis de la actividad científica e investigadora (SANZ CASADO et al., 2011).

Por último, tomando como punto de partida la metodología de GARCÍA GONZALEZ (2010), el autor de este trabajo ha optimizado los scripts en Perl y ha introducido nuevos desarrollos con el objetivo de completar el trabajo de GARCÍA GONZALEZ (2010), incorporando la normalización de los registros, especialmente de los nombres de autor y de los centros de trabajo de los mismos.

4.4.1. Obtención de los datos

En el caso de las bases de datos bibliográficas, para obtener los datos, se deben diseñar sendas estrategias de búsqueda que permitan recuperar la información de la forma más exhaustiva posible, y procurando que la información obtenida de ambas bases de datos sea coherente. Para ello las estrategias de búsqueda deben

ser lo más completas posibles y al mismo tiempo similares entre si, pues de otro modo se estaría sesgando la información en cualquiera de las dos fuentes.

4.4.1.1. Estrategias de interrogación a las bases de datos bibliográficas

Tanto en el caso del SCI como de Pubmed se planteó la posibilidad de utilizar abreviaturas de las enfermedades acompañadas de los términos “disease*” o “disorder*” utilizando el operador AND. De esta forma podría asegurarse que las abreviaturas se encuentran relacionadas con la investigación de enfermedades. Sin embargo no excluyen totalmente el ruido, pues algunas de las abreviaturas no se refieren tanto a enfermedades como a los genes implicados en ellas y estos genes pueden estar relacionados con una multitud de enfermedades con una prevalencia más acusada. Por esta razón se ha considerado más adecuado limitar la búsqueda a los nombres desarrollados de las enfermedades para evitar así el ruido que puedan ocasionar las abreviaturas.

Puesto que las estrategias de búsqueda van a incorporar el nombre completo de las enfermedades, lo más correcto es dirigir dichas estrategias contra todos los campos de las bases de datos. Para ello se aplican las estrategias sobre el campo “TS” en el SCI y en “All Fields” en Pubmed.

Estas estrategias de búsqueda van a permitir descargar los registros en formato de texto plano para ambas bases de datos, pues es el que mejor se adapta al lenguaje de programación (Perl) que se utiliza para incorporar los registros a las bases de datos MySQL.

A dichas estrategias se les ha añadido en ambas bases de datos las limitaciones correspondientes para obtener los registros correspondientes al periodo de estudio (2000-2009) y, en el caso del SCI, aquellos que tienen la consideración de artículos de revista.

En el SCI se han incluido cada una de las enfermedades entre paréntesis y dentro de ellos los términos de búsqueda entre comillas dobles “ ”. El operador que separa

cada una de las enfermedades es “OR”, no habiendo sido necesario incorporar el operador “AND” para los términos de búsqueda dentro de los paréntesis, puesto que es el operador que el sistema utiliza por defecto. Al final de la estrategia se ha incluido el operador “AND” con el rango de fechas 2000-2009 en el campo de publicación.

Por otro lado, pese a que Pubmed ofrece el tesoro MeSH, se ha optado por no utilizarlo porque no cuenta con términos específicos para cada una de las enfermedades, sino que en muchos casos los términos MeSH incluyen tanto las enfermedades que son objeto de este estudio como otras patologías que no lo son, lo que podría comprometer la precisión de los datos recuperados.

La composición final de las estrategias de búsqueda utilizadas puede consultarse en el apéndice A.3, donde se encuentran detalladas en función de los códigos de clasificación ICD-10.

4.4.2. Gestión de las bases de datos: MySQL

Para mantener y gestionar nuestra propia base de datos es necesario contar con un sistema de gestión de bases de datos relacionales lo suficientemente potente como para manejar gran cantidad de registros por tabla y que al mismo tiempo fuera flexible (para programar sobre él un sistema de normalización) y lo más eficiente posible. Este es el principal motivo por el que se decidió escoger MySQL como sistema de gestión. A continuación se exponen cuáles son sus principales características y otras razones por las que se ha seleccionado como sistema de gestión de la base de datos conformada.

MySQL es un sistema de gestión de bases de datos relacionales, multiproceso y multiusuario, que permite que varios usuarios realicen diferentes operaciones sobre las bases de datos de forma simultánea. Existen multitud de sistemas de gestión de bases de datos relacionales, pero tal y como explica DuBois (2009),

MySQL posee ciertas funcionalidades que hace que lo convierte en el sistema más adecuado para este tipo de trabajos. Entre ellas se pueden señalar:

Velocidad. MySQL es realmente rápido comparado con otros sistemas de gestión de bases de datos relacionales.

Facilidad de uso. MySQL es un sistema de bases de datos de alto rendimiento pero relativamente simple y es mucho menos complejo de configurar y administrar que sistemas mayores.

Lenguaje de interrogación. MySQL implementa SQL (Structured Query Language), el lenguaje estándar de los sistemas de gestión de bases de datos modernos.

Flexibilidad. MySQL es un sistema multi-hilo, que permite que varios usuarios se conecten al mismo tiempo. Cada cliente puede utilizar diferentes bases de datos al mismo tiempo. Además, MySQL permite conectar a través de múltiples interfaces y lenguajes de programación como C, **Perl**, Java, PHP, Python y Ruby.

Conectividad y seguridad. MySQL funciona en red, por lo que las bases de datos pueden ser compartidas a través de Internet, y al mismo tiempo incorpora un control de acceso que permite que cada usuario acceda tan sólo a los datos sobre los que tiene permiso. Como capacidad de seguridad adicional, MySQL incorpora conexiones encriptadas utilizando el protocolo SSL (Secure Sockets Layer)

Portabilidad. MySQL funciona sobre una gran variedad de sistemas Unix y Linux, así como sobre otros sistemas operativos como Windows y NetWare. MySQL es sumamente escalable y funciona tanto en grandes servidores como en pequeños ordenadores personales.

Pequeño tamaño. MySQL tiene un tamaño modesto en el disco duro comparado con otros sistemas de gestión de bases de datos.

Disponibilidad y precio. MySQL es un proyecto Open Source disponible bajo varias licencias, entre ellas la GPL (General Public License), lo que significa que MySQL está disponible sin coste alguno para uso doméstico, aunque también cuenta con licencias comerciales que incluyen servicio de soporte técnico.

Distribución y código abierto. MySQL es fácil de obtener y de usar con un navegador Web. Si se quiere conocer su funcionamiento, algoritmos o mejorar alguna funcionalidad el código fuente está disponible, y si se encuentra algún fallo (bug) puede informarse a los desarrolladores sobre él.

Por último, citando a GILFILLAN (2003):

Lo que durante un tiempo se consideró como un sencillo juguete para su uso en sitios Web, se ha convertido en la actualidad en una solución viable y de misión crítica para la administración de datos. Antes, MySQL se consideraba como la opción ideal para sitios Web; sin embargo, ahora incorpora muchas de las funciones necesarias para otros entornos y conserva su gran velocidad. MySQL supera desde hace tiempo a muchas soluciones comerciales en velocidad y dispone de un sistema de permisos elegante y potente.

Por estos motivos, está justificada su utilización en este trabajo.

4.4.2.1. Volcado a la base de datos en MySQL

Puesto que la descarga de los datos se lleva a cabo en formato de texto (SCI) y XML (Pubmed) es necesario programar un sistema que permita incorporar dichos registros en la base de datos. Para ello van a utilizarse los scripts en lenguaje Perl, que se detallarán más en profundidad en el punto 4.4.3.2.

Los datos descargados del SCI se van a volcar a una base de datos compuesta por la estructura de tablas que se observa en la figura 4.1. En el caso de Pubmed la estructura es la misma, salvo por el hecho de que la tabla “tcitas” no existe, al no incluir los registros información acerca de los trabajos citados en cada registro

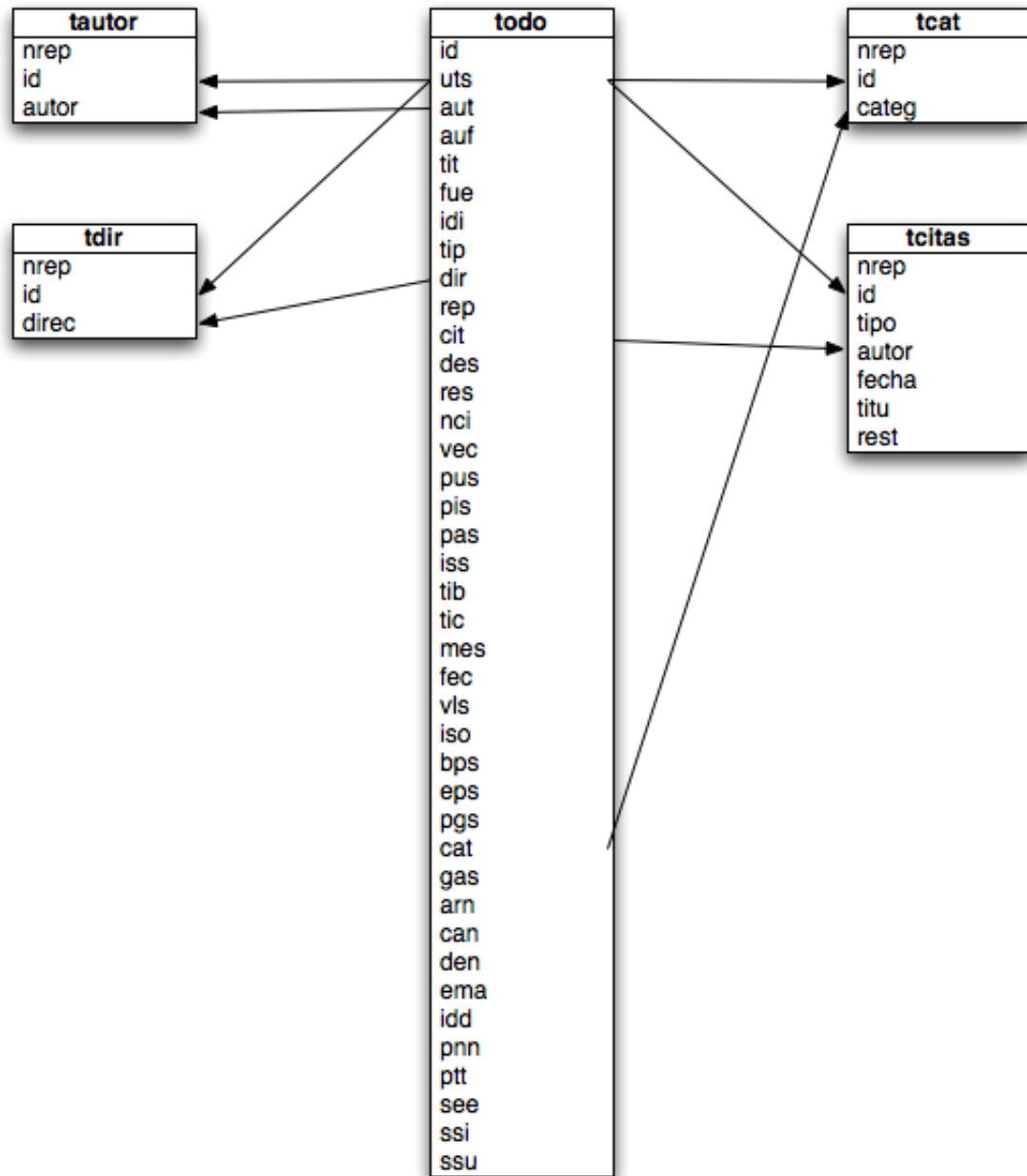


Figura 4.1: Estructura de bases de datos (SCI)

y variando ligeramente los nombres de los campos para adaptarlos a los nombres utilizados en Pubmed.

4.4.2.2. Extracción de datos mediante consultas SQL

El SQL es el lenguaje estándar ANSI/ISO de definición, manipulación y control de bases de datos relacionales. Es un lenguaje declarativo: solo hay que indicar que se quiere hacer. En cambio, en los lenguajes procedimentales es necesario especificar como hay que hacer cualquier acción sobre la base de datos. El SQL es un lenguaje muy parecido al lenguaje natural; concretamente, se parece al inglés, y es muy expresivo (MARTÍN ESCOFET, 2007).

La interrogación directa a la base de datos por medio de consultas SQL, ya sea a través de interface, por medio de scripts Perl o del paquete estadístico R (del que se hablará en el punto 5.2), es lo que va a permitir extraer los datos necesarios para calcular los indicadores bibliométricos.

Un ejemplo sencillo de consulta SQL sobre nuestra base de datos sería la siguiente:

```
SELECT COUNT( * ) AS `Filas` , `fec`  
FROM `todo`  
GROUP BY `fec`  
ORDER BY `fec` ASC
```

Al ejecutar esta consulta le indica a MySQL que cuente el número de filas para cada valor del campo “fec” en la tabla “todo”, es decir, que muestre la distribución de trabajos publicados por año de publicación en nuestra base de datos. En este caso al ejecutar la consulta sobre la base de datos construida a partir de los registros relativos a trastornos primarios de los músculos y que han sido recuperados a través de la estrategia G71 que se puede encontrar en el apéndice A.3, obteniendo la tabla 4.2.

Tabla 4.2: *Trabajos publicados por año sobre trastornos primarios de los músculos*

Año	Total de trabajos
2000	184
2001	185
2002	182
2003	179
2004	176
2005	168
2006	183
2007	182
2008	177
2009	208

De esta forma, a través de consultas SQL más elaboradas se pueden obtener tablas de datos más complejas, matrices y, como se mostrará más adelante, gráficos a partir de las propias consultas.

4.4.3. Tratamiento de los registros con scripts en Perl

En este apartado va a tratarse lenguaje de programación que ha sido utilizado para tratar los registros bibliográficos, construir la base de datos y normalizar los datos en ella.

4.4.3.1. Descripción del lenguaje Perl y su uso para el análisis de textos

Perl es lo que se conoce como lenguaje de programación interpretado, es decir, que no requiere de un compilador, y como tal, se basa en la construcción de scripts, es decir, trozos de código relativamente pequeños, escritos en un lenguaje interpretado (como Perl), que se almacenan en un archivo de texto plano y se ejecutan por medio del *shell*, terminal o consola de comandos del sistema operativo.

En el caso de Perl, se trata de un lenguaje de programación de alto nivel escrito por Larry Wall y desarrollado actualmente por una amplia comunidad de usuarios

que diseñan nuevos modos de implementación, módulos y desarrollos en este lenguaje.

Perl es bastante bueno en lo que se refiere a la manipulación de procesos, archivos y texto. Por esta razón, resulta especialmente apropiado para utilidades de sistema, herramientas de software, tareas de administración de sistema, acceso a bases de datos, programación gráfica, conexión a red y programación de aplicaciones basadas en la Web. Estas características lo convierten en un lenguaje atractivo para autores de scripts CGI, administradores de sistemas, matemáticos, periodistas y cualquier persona que necesite programar aplicaciones y utilidades rápidamente (BROWN, 2001).

Por estos motivos Perl ha sido utilizado de forma intensiva en el tratamiento de textos, minería de textos y se ha mostrado como un lenguaje flexible y adaptable a las necesidades de los programadores. Por otro lado Perl cuenta con una funcionalidad importante para el tratamiento de textos y son las denominadas “expresiones regulares”, las cuales permiten identificar en el texto patrones de gran utilidad para su tratamiento con diferentes objetivos (volcado a bases de datos, normalización, recuperación de información, etc.) (BILISOLY, 2008).

Por último, una razón de peso para elegir Perl como el lenguaje de programación utilizado para el procesamiento de datos en este trabajo, es el hecho de que cuenta con una gran cantidad de módulos bajo licencias libres (CPAN, 2009) disponibles para su implementación en los scripts que han sido desarrollados en este trabajo.

4.4.3.2. Scripts para SCI y Pubmed

Para trasladar los datos descargados de SCI y Pubmed a nuestra propia base de datos en MySQL, se utilizan scripts programados en lenguaje Perl, de forma que permitan leer cada registro de forma separada, identificar cada uno de los campos que contiene y asignar su contenido a los campos de nuestra base de datos en formato SQL. En la carga de los datos de los registros descargados del SCI se ha utilizado el script diseñado por GARCÍA GONZALEZ (2006) y que permite

realizar las operaciones que se mencionaron anteriormente. Sin embargo, a la hora de realizar la carga de los registros descargados de Pubmed éste script ha tenido que ser reprogramado para adaptarlo a las características propias de estas bases de datos y a los campos que en ellas se utilizan.

El principal problema encontrado a la hora de realizar este trabajo es el de los campos repetidos, pues algunos, como el de autor o el de encabezamientos de materia del Medical Subject Headings (MeSH), se encontraban repetidos, de forma que en cada registro existe un campo "AU" por cada autor o un campo "MH" por cada materia MeSH asignada. A ello hay que unir que en el caso de Pubmed cada autor aparece en dos campos diferentes, por un lado en el campo "AU" que incluye el nombre del autor con el nombre de pila abreviado en siglas y por otro el campo "FAU" que contiene el nombre desarrollado y que será el utilizado para llevar a cabo los diferentes indicadores asociados a los autores, pues permite normalizar sus nombres de forma más adecuada. A continuación se incluye un ejemplo de registro Pubmed en el que se observan estas dificultades:

```
PMID- 19009934
OWN - NLM
STAT- Pubmed
DA - 20081117
DCOM- 20081231
IS - 1865-9217 (Print)
VI - 102
IP - 1
DP - 2008
TI - [Rare diseases--acquaintance and uncertainty]
PG - 6-7
FAU - Lelgemann, Monika
AU - Lelgemann M
FAU - Francke, Robert
AU - Francke R
LA - ger
PT - Editorial
TT - Seltene Erkrankungen--
Umgang mit Unsicherheit.
```

PL - Netherlands
TA - Z Evid Fortbild Qual Gesundheitswes
JT - Zeitschrift fur Evidenz, Fortbildung und
Qualitat im Gesundheitswesen
JID - 101477604
SB - IM
MH - Age of Onset
MH - Child
MH - Chronic Disease
MH - Humans
MH - Quality Assurance, Health Care
MH - Rare Diseases/*therapy
MH - *Uncertainty
EDAT- 2008/11/18 09:00
MHDA- 2009/01/01 09:00
CRDT- 2008/11/18 09:00
PST - ppublish
SO - Z Evid Fortbild Qual Gesundheitswes.
2008;102(1):6-7.

Para resolver este doble problema se optó por descargar los registros en formato XML y utilizar un “parser” o analizador escrito en Perl, que permitiera extraer el contenido de cada campo y construir nuestra base de datos en formato SQL.

4.4.3.3. Scripts para la normalización

Además de la utilidad demostrada del lenguaje Perl para extraer los datos de registros en formato de texto, también ha permitido desarrollar un sistema de normalización para nuestras bases de datos. El sistema se encuentra en continua fase de desarrollo, aunque para el desarrollo de esta tesis doctoral se han implementado completamente los sistemas de normalización de autores, países y direcciones.

4.4.3.3.1. Normalización de autores Para conseguir el objetivo planteado, normalizar las diferentes variantes de firma de los autores, se ha experimentado con dos tipos diferentes de métodos estadísticos, el denominado “String Matching” y los “N-grams”. En ambos casos se trata de métodos orientados a obtener la

similaridad de cadenas de texto, en este caso nombres de autor. A continuación se presentan ambos métodos y se establece una comparativa entre ellos para determinar cuál es el que mejor se adapta a las necesidades planteadas.

En primer lugar se hace necesario definir qué es lo que se considera como similitud. Para ello se ha tenido en cuenta el trabajo de LIN (1998), en el cual se parte de tres intuiciones básicas:

Intuición 1: La similaridad entre A y B está relacionada con sus elementos comunes. Cuantos más elementos compartan mayor será su similaridad.

Intuición 2: La similaridad entre A y B está relacionada con sus diferencias. Cuantas más diferencias existan entre ellos menor será su similaridad.

Intuición 3: La similaridad máxima entre A y B es alcanzada cuando A y B son idénticos, no importan cuantos elementos en común compartan.

Este autor, define la similaridad como el ratio entre el total de información común y el total de información diferenciadora entre dos objetos. Por tanto en el caso de los nombres de autor se puede determinar que el grado de similaridad entre ellos puede establecerse a través del número de elementos comunes y el número de elementos diferentes que presentan.

Por esta razón van a utilizarse diferentes técnicas que difieren sustancialmente en la forma de determinar los elementos constituyentes de una línea de texto.

4.4.3.3.1.1. String Matching El método conocido como “String Matching” o “Approximate String Matching” consiste en hallar la distancia, $d(x, y)$, entre dos líneas de texto “ x ” e “ y ”. La forma de medir esta distancia es determinando el número mínimo de transformaciones que debe realizarse sobre x para transformarlo en y .

Considerando que las líneas de texto están compuestas por un número finito de caracteres (a, b, \dots, n), excluidos los caracteres no alfanuméricos, las operaciones a realizar serían un conjunto finito del estilo $d(x, y) = t$ donde x e y serían diferentes líneas de texto y t sería un número real positivo. Si las operaciones de $d(x, y)$ tienen el mismo “coste” que las operaciones para $d(y, x)$ entonces se trata de distancias simétricas (NAVARRO, 2001b). En nuestro caso los nombres de autor adoptan este tipo de estructura, por lo que para el conjunto de dos nombres de autor diferentes se obtendrán dos medidas idénticas de similaridad.

Esta técnica se compone de cuatro tipos de operaciones a realizar entre líneas o “strings” para determinar el “coste” de convertir una en otra:

- Inserción: Agrega un elemento a la línea
- Borrado: Elimina un elemento de la línea
- Sustitución: Sustituye un elemento de la línea por otro. Se contabiliza como una sola operación aunque se trate de la combinación de las operaciones de borrado e inserción
- Transposición: Se trata de intercambiar la posición de dos elementos contiguos de una línea. Se diferencia de la operación de sustitución en el hecho de que en este caso ambos elementos se encuentran ya en la línea pero en posiciones diferentes y adyacentes.

Un ejemplo claro de cómo funciona esta técnica se encuentra en el trabajo de HALL y DOWLING (1980). Cómo se puede observar en la figura 4.2, la distancia entre ambos nombres sería de 2 caracteres, la letra inicial y la “T” intercalada en el nombre que se encuentra en posición vertical (HILLIER).

En un principio se barajó la posibilidad de utilizar este método para obtener valores de similaridad de los nombres de autor que permitieran eliminar los autores duplicados. Sin embargo, después de realizar diversas pruebas con los módulos *String::Similarity* (LEHMANN, 2007) y *Text::Similarity* (PEDERSEN et al., 2008),

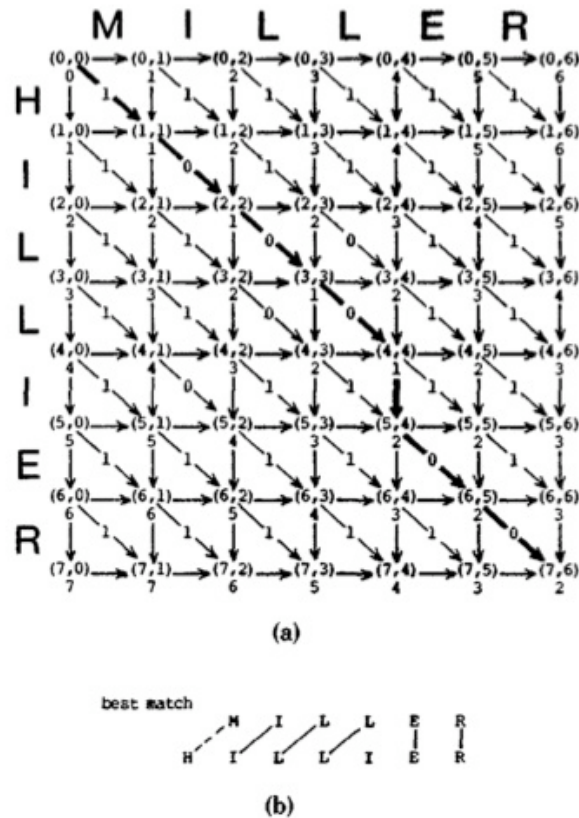


Figura 4.2: Ejemplo del funcionamiento de la técnica de String Matching

se observó que no se trataba de un método lo suficientemente eficaz por diversos motivos:

Como ya se ha explicado, esta técnica mide la “distancia de edición” o lo que es lo mismo, el número de cambios que deben realizarse en una línea de texto concreta para convertirla en cualquier otra de la lista con la que se compara. Por tanto se hace necesario definir unas estructuras fijas dentro de cada línea, que serán las que se intercambien para obtener la distancia entre líneas. En el caso de líneas de texto se trata de un método que puede resultar complicado de implementar, debido a que se hace necesario definir los componentes de las estructuras de comparación (en textos de gran longitud) y a que en textos cortos (como los nombres de autor) el número de combinaciones es muy pequeño y por tanto poco eficaz a la hora de compararlas.

Para poner a prueba el sistema se utilizaron dos tipos de estructuras. En primer lugar se programó un script en Perl que utilizara como elementos intercambiables las palabras que componían las diferentes líneas de texto. Este tipo de estructura funciona relativamente bien con textos de mayor extensión y puede ser muy útil en las direcciones institucionales o para realizar análisis de contenido, pero con textos de poca extensión (tales como los nombres de autor), no ofrece resultados eficaces.

Posteriormente se modificó el script para que utilizara los caracteres individuales de cada línea como elementos intercambiables. En un principio este método ofrece mejores resultados con textos cortos, como los nombres de autor, pero presenta algunos problemas. En primer lugar supone un conjunto demasiado amplio de elementos, por lo que el sistema es poco eficiente; y en segundo lugar el sistema no tiene en cuenta la posición de los caracteres (salvo para las operaciones de transposición), por lo que se obtenían altos índices de similitud entre nombres de autor con un número parecido de caracteres, sobretodo para aquellos nombres más cortos, como por ejemplo Chen, H ó Chen, T.

En este caso, se obtenía una similitud perfecta para el apellido (ningún cambio en cuatro caracteres) y del 50% para el nombre (un cambio en dos caracteres).

Por estos motivos se decidió utilizar otro método para determinar el grado de similitud entre los autores de la base de datos.

4.4.3.3.1.2. N-grams El siguiente método propuesto para hallar la similitud entre las diferentes formas que pueden adoptar los nombres de autor en las bases de datos bibliográficas se basa en el uso de N-grams. EGGHE (2000) define los N-grams como:

[...] palabras generalizadas consistentes en N símbolos consecutivos.
Los N-grams son un objeto de estudio muy importante que tiene aplicaciones en la indexación, recuperación de información, corrección de

errores, compresión de texto, identificación del lenguaje, clasificación por materias e incluso reconocimiento del habla.

Este método utiliza cadenas de texto de longitud fija y compara su frecuencia con la de esa misma cadena de caracteres en otro documento (MOREIRO, 2002). Existen diferentes tipos de n-grams en función del número de caracteres que los compongan. Así, pueden dividirse en bi-grams (dos n-grams), tri-grams (tres n-grams), quad-grams (cuatro n-grams), etc. Siguiendo el ejemplo de CAVNAR y TRENKLE (1994) estos diferentes tipos de n-grams dividirán una palabra concreta, por ejemplo “TEXT”, del siguiente modo:

bi-grams: _T, TE, EX, XT, T_

tri-grams: _TE, TEX, EXT, XT_, T__

quad-grams: _TEX, TEXT, EXT_, XT__, T___

Observando en que consiste esta metodología se puede deducir que debe tenerse muy en cuenta a la hora de desarrollar trabajos relacionados con la recuperación de información. Más aún considerando que su distribución se ajusta a la Ley de Zipf, según la cual la frecuencia de aparición de un elemento, en este caso un conjunto de caracteres dado, es inversamente proporcional a su rango o importancia en un contexto dado (EGGHE, 2000).

Este sistema, al igual que el “String Matching” puede ser utilizado con buenos resultados para comparar líneas de texto, con la salvedad de que los n-grams son mucho más flexibles, al permitir el ajuste manual del número de caracteres que componen la cadena a evaluar. Además, a diferencia del método anterior, tienen en cuenta la posición de los caracteres y, como veremos, también permiten especificar si se deben considerar las diferencias entre mayúsculas y minúsculas o no.

El principal problema que presenta este sistema reside en la cantidad de espacio de almacenamiento necesario para mantener el fichero inverso que genera el sistema (ROBERTSON y WILLETT, 1998). Sin embargo como se mostrará más

adelante, cuando se trate el método de implementación, en este caso no supone un problema, pues no es necesario construir un fichero inverso, puesto que los textos se comparan uno a uno.

Debido a la longitud de los textos que se van a utilizar para identificar autores duplicados (nombres de autor, direcciones institucionales y categorías ISI), se ha considerado que la mejor opción es utilizar n-grams de tres caracteres de longitud o trigramas, pues permiten hacer un análisis más eficiente que la comparación por caracteres (demasiado lenta). Al mismo tiempo aportan la capacidad de los n-grams de considerar conjuntos de caracteres, que en esta caso no deben ser demasiado amplios, pues se trata de textos de tamaño reducido.

Para el desarrollo de este método se ha utilizado el módulo de Perl *String::Trigram* (AHMED, 2003), que permite partir el texto en n-grams del tamaño que se desee (evidentemente deben ser mayores o iguales a dos) y, al mismo tiempo, establecer comparaciones entre dichos textos a partir de los trigramas obtenidos. Para implementar este módulo existían dos opciones.

El método orientado a objetos. Consiste en establecer una base de comparación que contenga un conjunto de caracteres suficientemente amplio como para establecer comparaciones con líneas de texto concretas. Esta forma de implementación sería la más adecuada para determinar si una línea de texto concreta se encuentra contenida en el “texto base”. Así un ejemplo de utilización de éste método sería su utilización para desarrollar un sistema anti-plagio que permitiera determinar si fragmentos de texto procedentes de un trabajo cualquiera se encuentran contenidos en un conjunto de textos más amplios, por ejemplo la Wikipedia.

El método funcional. En este sistema, que es el que ha sido utilizado en este trabajo, no es necesario establecer una base textual para realizar las comparaciones, sino que los textos se comparan uno a uno. De nuevo se trata del sistema más eficaz para comparar cadenas de textos de poca longitud, por no hablar de que

en este caso no tendría sentido establecer una base de texto, pues lo que se busca es comparar los datos de cada uno de los autores con todos los demás.

4.4.3.3.1.3. Comparación entre el método de String Matching y el de Trigramas Como ya se ha comentado, el método elegido para normalizar los nombres de autor, ha sido el de la comparación mediante n-grams de tres caracteres de tamaño o trigramas previa comparación con los métodos de string matching tanto para caracteres como para palabras. Una vez que se comprobó el funcionamiento de los tres métodos se realizaron pruebas sobre un pequeño grupo de autores para determinar el nivel de eficiencia, si permitían diferenciar entre elementos encontrados en el texto de forma consecutiva y si permitían ajustar el tamaño de dichos elementos.

Observando los tiempos de ejecución de los tres sistemas puede determinarse que, para obtener la similaridad entre nombres de autor el sistema más eficiente es el de string matching, basado en el análisis por palabras, mientras que el menos eficiente es el mismo string matching pero en su versión basada en el análisis de caracteres individuales. El método de comparación mediante trigramas ofreció unos valores medios en tiempo de ejecución que no llegaba a alcanzar los valores alcanzados por el método de análisis de palabras pero tampoco descendía a los del análisis de caracteres. Esto se debe a que, si se consideran como palabras diferentes las que se encuentran contenidas entre dos espacios y se eliminan los signos de puntuación, cada nombre de autor tiene generalmente dos o tres palabras, mientras que el número de caracteres a comparar es mucho mayor y el de trigramas algo menor que el de caracteres, por lo que los ciclos de procesador necesarios para llevar a cabo las comparaciones es mayor a medida que aumenta el número de elementos a tener en cuenta en cada línea de texto.

Por otro lado para las condiciones ya mencionadas, que el sistema tenga en cuenta los elementos consecutivos y que permita ajustar su tamaño, en los dos métodos basados en string matching, la respuesta es negativa para ambas condiciones,

mientras que para el sistema de n-grams se cumplen las dos. En este caso es indiferente el tamaño de los elementos a comparar, dado que en ningún caso la técnica de string matching va a permitir ajustar el tamaño de los elementos, pues en tal caso se acercaría más al método de los n-grams, y por su propia definición sólo puede tener en cuenta los elementos consecutivos adyacentes, mientras que en los n-grams el número de elementos consecutivos que se tienen en cuenta es de $N-1$.

Por último, tras observar los resultados conseguidos para el caso concreto de obtener la similaridad entre nombres de autor, debe señalarse que, globalmente, los menores niveles de efectividad los ha obtenido el sistema de string matching basado en palabras, mientras que el basado en caracteres ha obtenido unos niveles de eficacia aceptables a costa de la eficiencia y a pesar de obtener un mayor número de equiparaciones erróneas que el método basado en trigramas, que se destaca como la mejor alternativa para desarrollar un sistema de normalización de autores basado en la similaridad.

A continuación se muestra la tabla 4.3, en la que se pueden observar los criterios que se han tenido en cuenta para la elección y los valores observados para cada método.

Además, un motivo de peso para elegir el sistema de n-grams como método para obtener medidas de similaridad entre autores, es que se trata de un método que ya ha sido utilizado anteriormente con éxito para normalizar nombres propios (PFEIFER et al., 1996). Y a pesar de que en nuestro caso se deben tener en cuenta más factores a parte del nombre de autor, se trata de la medida más importante a tener en cuenta en nuestra metodología.

4.4.3.3.1.4. Procedimiento programado en Perl. En base a lo expuesto anteriormente se decidió, aplicando el método de la comparación por medio de trigramas, desarrollar un método para normalizar los autores de registros descargados de las bases de datos Thomson Reuters y Pubmed, con objeto de que permitan realizar

Criterio de evaluación	String Matching (caracteres)	String Matching (palabras)	Trigramas
Nivel de eficiencia	Baja	Alta	Media
Tipo de orientación	Textos cortos estructurados (números de teléfono)	Textos largos poco estructurados (resúmenes)	Textos cortos poco estructurados (nombres propios)
Diferencia entre elementos consecutivos	No	No	Si
Permite ajustar el tamaño de los elementos	No	No	Si
Eficacia global	Media	Baja	Alta

Tabla 4.3: Comparación entre métodos para obtener la similaridad de los nombres de autor

estudios bibliométricos más fiables, sin que los resultados se vean alterados por las variantes en los nombres de autor o por el hecho de que algunos autores firmen el mismo trabajo bajo nombres institucionales diferentes.

Debe destacarse que el sistema es mucho más limitado para la normalización de autores en Pubmed, debido a que muchos de los campos que sirven para llevar a cabo la normalización, como la dirección del autor o las categorías temáticas, no se encuentran presentes en esta base de datos, por lo que en el caso de Pubmed el sistema se ciñe a la comparación, mediante el método de n-grams, de tres campos (autor, coautores y revistas), siendo mucho más limitado que en el caso de las bases de datos Thomson-Reuters, que es el que se detalla a continuación.

PASO 1: PREPARACIÓN DE LA BASE DE DATOS.

En primer lugar y una vez establecida la herramienta a utilizar se hizo necesario determinar cuáles eran los campos que se iban a necesitar para conseguir la información más completa posible de cada autor para determinar similitudes entre

ellos. El primer problema encontrado es que se contaba con dos campos diferentes en los que existían nombres de autor. Por un lado el campo “author” contenía el nombre abreviado de los autores y estaba presente en todos los registros, mientras que en el campo “full author” se contaba con el nombre completo de los autores, pero sólo se encontraba disponible en los registros más recientes.

Para solventar este problema se decidió crear un sistema de tablas intermedias en nuestra base de datos, de forma que más adelante pudiera generarse un fichero único de autores que contuviera todas las formas posibles de cada autor junto con sus datos de identificación principales.

Para crear las tablas intermedias se procedió a separar los dos tipos de nombres de autor, de forma que por un lado se contara con una tabla con los autores procedentes del campo “author” asociados mediante el ID de cada registro con las instituciones del campo “reprint”, que era el único que permitía asociar algunos de estos nombres de autor con su dirección institucional. Mientras que por otro lado se generó una tabla con los nombres de autor completos, procedentes del campo “full author” así como sus afiliaciones institucionales procedentes del campo “C1”. Éste se caracteriza por incluir antes de cada dirección institucional, y entre corchetes, los nombres completos de todos los autores afiliados a ella.

Por último se creó una tabla que pudiera contener los registros conjuntos de ambas tablas intermedias y que permitiera también asociar a los autores con sus direcciones. Se puede observar la estructura detallada en la Tabla 4.4

PASO 2: IDENTIFICACIÓN DE AUTORES CON DIRECCIÓN ASOCIADA Y CONSTRUCCIÓN DE LA LISTA DE AUTORIDADES. Una vez determinada la estructura de las tablas que van a contener los nombres de autor con las direcciones asociadas, así como la tabla de autoridades, que servirá para hacer las comparaciones y detectar las variantes de nombres de autor junto con sus instituciones, el siguiente paso consiste en identificar, dentro de cada uno de los registros bibliográficos, a que institución pertenece cada uno de los autores firmantes y construir con esos

Campo	ID	proc	autor	dirección	categoría	coautores	revistas
Descripción del campo	Número de identificación de cada registro	Tabla intermedia de la que procede el registro	Nombre de autor procedente de los campos "author" y "full author"	dirección institucional	Categorías asignadas al registro	Coautores del registro bibliográfico	Revista en la que se ha publicado
Campo	vID	vproc	vautor	vdirección	vcategoría	vcoautores	vrevistas
Descripción del campo	Variantes de números de identificación	Tabla intermedia de la que procede el registro	Variantes de nombre de autor	Variantes de dirección institucional	Categorías ISI asignadas al registro	Variantes de los coautores	Variantes de las revistas

Tabla 4.4: Estructura del fichero de autoridades

datos la versión de la lista de autoridades que más tarde servirá para llevar a cabo el proceso de normalización.

Como ya se ha comentado anteriormente, la base de datos incluye dos campos que contienen nombres de autor (author y full author) así como dos campos que contienen direcciones institucionales (reprint y C1). Para asociar cada autor con su dirección correspondiente se siguió el siguiente procedimiento:

En primer lugar hay que tener en cuenta que los autores del campo “full author” aparecían en la dirección del campo “C1” entre los signos [y]. Un ejemplo sería la siguiente dirección:

```
[Gast, Marie-Christine W.; Beijnen, Jos H.]  
Slotervaart Hosp, Netherlands Canc Inst,  
Dept Pharm & Pharmacol, Amsterdam,  
Netherlands.
```

```
[van Tinteren, Harm] Antoni Van Leeuwenhoek  
Hosp, Netherlands Canc Inst, Dept Biometr,  
Amsterdam, Netherlands.
```

Esta dirección indica que “Gast, Marie-Christine W.” y “Beijnen, Jos H.” comparten la misma dirección institucional (Slotervaart Hosp, Netherlands Canc Inst, Dept Pharm & Pharmacol, Amsterdam, Netherlands) mientras que la institución de “van Tinteren, Harm” sería “Antoni Van Leeuwenhoek Hosp, Netherlands Canc Inst, Dept Biometr, Amsterdam, Netherlands”. En este caso es sencillo asociar cada autor con su dirección institucional, basta con indicarle a la base de datos que si el nombre de autor procedente del campo “full author” se encuentra contenido en el texto de una institución quiere decir que el autor pertenece a ella.

Sin embargo, aunque los registros más recientes cuentan con los nombres de autor completos y la dirección institucional recogida en este formato, los registros más

antiguos tan solo incluyen el nombre de autor abreviado y la dirección institucional del primer autor en el campo “reprint”.

Esto dificulta enormemente la labor de asociar nombres de autor y direcciones, y la mejor forma de conseguir una buena asociación para el primer autor, es siguiendo el mismo método que para el caso anterior, pues el formato de las direcciones en el campo “reprint” es:

```
Sekowska, M, UBLG 1 LF UK, Albertov 4, Prague,  
Czech Republic.
```

Como se puede observar la dirección incluye el nombre del primer autor firmante, en este caso “Sekowska, M”.

Para aplicar éste método de asociación e insertar los nombres de autor junto con sus direcciones identificadas y el resto de datos bibliográficos, se diseñó un script en Perl que asociaba al autor con una dirección institucional siempre y cuando dicha dirección contuviera el nombre del autor. De esta forma, para las direcciones completas se tienen en cuenta los nombres completos de los autores y para las direcciones del campo “reprint” se tienen en cuenta los nombres abreviados.

PASO 3: COMPARACIÓN ENTRE LOS COMPONENTES DE LA LISTA DE AUTORIDADES PARA LOCALIZAR DUPLICADOS. Una vez asignados a cada autor todas las posibles direcciones con las que aparece en los registros de la base de datos e incluidos en el fichero de autores preliminar, se siguió desarrollando el método de detección de duplicados, por medio del uso de n-grams de tres caracteres o trigramas, para obtener los valores de similaridad necesarios para determinar si dos o más nombres de autor se corresponden con un mismo autor.

Previamente debe especificarse que los umbrales de similaridad que se han establecido en este método, como los más óptimos para identificar duplicados, se han obtenido por medio de la observación directa de los registros. Utilizando un conjunto más reducido de registros y calculando sus similaridades por medio del mismo sistema, pudo determinarse cuáles eran los valores que debía alcanzar di-

cho cálculo para considerar dos variantes de firma como pertenecientes al mismo autor.

Debido a que se trata de un proceso en el que intervienen diferentes medidas de similaridad, y que para las diferentes combinaciones de cada una de ellas es necesario aplicar distintos umbrales de similaridad, ha sido necesario dividir el proceso en las fases que se explican a continuación.

Fase preliminar: La primera fase realiza tres funciones básicas. En primer lugar se llevan a cabo las consultas SQL sobre el listado de autoridades de la base de datos. Posteriormente se preparan los datos extraídos para su procesamiento (escapando caracteres, separando los valores por campos y renombrando las variables) y por último se calculan los valores de similaridad para cuatro variables diferentes: apellidos, nombres de pila, direcciones institucionales y coautores.

Fase 1. Similaridad entre apellidos y nombres: Una vez que los datos se encuentran listos para ser procesados se comienzan a incluir las estructuras de control o condicionales que determinan qué registros se considerarán duplicados y cuáles se mantendrán. Puesto que mantener la integridad de los datos es fundamental, se ha optado por no eliminar en ningún caso los registros duplicados de forma permanente, por lo que cada vez que un registro se identifica como duplicado, los datos que contienen, pasan a formar parte de los campos que incluyen todas las posibles variantes de autor, dirección, coautores, etc., del registro que se consideró como el más completo entre los que hacen referencia a un mismo autor.

En esta primera fase las condiciones que se establecen son muy simples. Basta con que la similaridad entre los apellidos de dos nombres de autor sea superior a 0,4 y que la similaridad entre sus nombres de pila sea superior a 0,5 para que se consideren como el mismo autor. Sin embargo para que a uno de ellos se le asigne la etiqueta de “duplicado” es necesario que la otra variante de firma sea más completa. En ese caso, para la variante más completa se incluirán en los

campos de variantes (vid, vautor, vdireccion, vcategoria, vrevistas y vcoautores), los datos del registro etiquetado como duplicado.

Fase 2. Similitud entre apellidos, nombres y direcciones: Esta segunda fase va orientada a detectar los duplicados que no se han detectado en la primera, o que habiéndose detectado, no se ha podido determinar cuál de ellos es el más completo (al procesarse únicamente los nombres de autor). Para ello van a utilizarse unos umbrales de similitud más bajos en los valores de similitud de apellidos y nombres que se filtrarán por medio de los valores obtenidos de la similitud entre direcciones.

Asimismo esta fase se encuentra dividida en dos partes. La primera comparte el objetivo de la fase 1, de detectar las variantes de firma que sean lo más parecidas posibles, mientras que la segunda parte permite identificar dos variantes de firma con apellidos prácticamente idénticos, pero en los que al menos uno de los nombres de pila está en forma de iniciales. Para conseguir esto se aumenta el umbral de similitud para los apellidos y se reduce casi al mínimo el del nombre de pila, al mismo tiempo que se exige como condición que la primera inicial del nombre sea idéntica. Además se reducen los umbrales de similitud de las direcciones, pues se encontraron autores que cuando firmaban con su nombre completo incluían la dirección institucional en inglés, mientras que cuando firmaban con sus iniciales, el nombre de la institución aparecía en su lengua materna. Así pues, para la fase 2 van a utilizarse los siguientes umbrales de similitud:

Variable	Umbrales
Apellido	0,4
Nombre	0,4
Dirección	0,2

Tabla 4.5: Variables y umbrales de similitud en la fase 2.1

Por otra parte, del mismo modo que en la fase 1 se determinaba cuál era el registro más completo a partir de la longitud de los nombres de autor, en este caso se utiliza el mismo criterio, salvo que ambos nombres de autor tengan la

Variable	Umbrales
Apellido	0.9
Nombre	0.05
Dirección	0.15

Tabla 4.6: Variables y umbrales de similaridad en la fase 2.2

misma longitud, en cuyo caso se toma como firma más completa aquella que incluya la dirección institucional de mayor longitud, sin olvidar que la dirección descartada se incluirá como variante, por lo que en cualquier caso seguirá asociada al nombre de autor normalizado.

Fase 3. Similaridad entre apellidos, nombres, direcciones y coautores:

Cómo último paso para determinar las variantes de firma, se ha diseñado una tercera fase orientada a detectar nombres de autor con variaciones significativas en las direcciones que no pudieron identificarse en las fases anteriores, debido a que siempre se ha utilizado como filtro el hecho de que el campo ID no fuera idéntico en ambos registros. De este modo se evita comparar el mismo registro consigo mismo. Pero por el contrario, en los casos de autores que firman con diferentes instituciones en un mismo trabajo van a encontrarse duplicados.

Por otro lado, algunos autores utilizan indistintamente su idioma materno y el inglés para reflejar su dirección institucional, por lo que un umbral de similaridad tan alto no permite identificarlos. En un principio se trató de ajustar el umbral de similaridad en las direcciones para detectar estos casos, sin embargo utilizando este método aparecerían un gran número de autores mal identificados. Por tanto se hacía necesario incluir una variable adicional que permitiera eliminar el filtro de ID's idénticas, sin que la precisión del sistema se viera afectada.

Después de realizar pruebas con diferentes variables (categorías, revistas y coautores) y teniendo en cuenta que las variables de categorías y revistas son demasiado parecidas cuando se trata de autores que trabajan en áreas muy cercanas, se optó por utilizar la variable coautores como filtro, otorgándole un umbral de 0,3 para filtrar los errores en la asignación. Por último, para asegurarnos de que en

esta última fase no se comparaban registros idénticos se incluyó la condición de que las direcciones de ambos registros fueran diferentes.

Fase 4. Eliminación de duplicados: Finalmente, para completar la construcción del fichero de autoridades se utiliza una consulta SQL a la base de datos para eliminar aquellos registros que han sido etiquetados como duplicados, tras lo cual se indica el tiempo de ejecución final del script y se da por terminado el proceso de normalización de la lista de autores con dirección identificada

PASO 4. OPTIMIZACIÓN DEL SISTEMA: El desarrollo del sistema de normalización por fases se hacía necesario debido a que, para realizar los ajustes en los umbrales de similaridad, era necesario saber en que punto del proceso se identificaba cada autor concreto. Por este motivo y a efectos de agilizar el proceso, una vez que se contó con una versión definitiva del código fuente por fases, se decidió implementar en un nuevo script que recogiera en una sola estructura condicional todas las condiciones que se había insertado en las diferentes fases del proceso.

Al mismo tiempo, se planteó la posibilidad de cambiar algunos de los procesos a la hora de realizar las consultas a la base de datos, para evitar la realización de tantas consultas como registros se encontraran en el fichero de autoridades. Pero esto suponía perder una funcionalidad importante, como es evitar la comparación de registros ya identificados como duplicados, y apenas suponía una mejora en cuanto a tiempo de ejecución y consumo de recursos.

Por este motivo, se optó por optimizar el sistema incluyendo dos nuevas condiciones, relativas a la similaridad de los nombres de autor antes de comenzar a calcular el resto de similaridades.

De esta forma, sólo para aquellos pares de autores que cuenten con un índice de similaridad en su apellido superior a 0,4 se calculará el índice de similaridad del nombre. Solo si éste es superior a 0,05 (el valor mínimo en las condiciones posteriores), entonces se calcularán los índices de similaridad para las direcciones y los coautores. Con este proceso se logró reducir el tiempo de ejecución del script

hasta una media aproximada de 2 horas para un fichero de autoridades compuesto por 4.393 registros.

Además, en la fase de optimización se detectaron ciertos errores a la hora de identificar las variantes de firma en las que en el apellido sólo variaba la inicial (por ejemplo Konrad y Conrad), por lo que se incluyó una nueva condición para que no fueran consideradas como variantes de firma de un mismo autor, salvo que las iniciales del nombre y del apellido fueran idénticas.

Por último se incluyó el módulo *Smart::Coments* (CONWAY, 2005), con el objetivo de proporcionar al usuario información acerca del estado del proceso, como la hora a la que se inició y a la que ha terminado o el número de registros que se están comparando en cada consulta a la base de datos.

Este sistema ha sido operativo para la normalización de los nombres de autor contenidos en el conjunto de datos utilizados para esta investigación (65.976 variantes de nombre diferentes), sin embargo, ha sido necesario mantener ejecutando el procedimiento durante más de 300 horas para completar la normalización, pues al aumentar el número de autores a comparar aumenta exponencialmente el tiempo necesario para su ejecución. Por lo tanto, debido a lo costoso del sistema (en términos de ciclos de procesador), se hace necesaria su optimización, con el objetivo de utilizarlo en trabajos futuros en los que el número de autores a normalizar sea mucho mayor.

Por último debe destacarse que el sistema ha demostrado su validez sobre un gran conjunto de autores internacionales, por tanto, si se quisiera aplicar a un conjunto de autores, una materia, revista, país o institución concreta, sería necesario ajustar los umbrales de similaridad para adaptarlos al conjunto de autores con el que se vaya a trabajar.

4.4.3.3.2. Normalización de países Para la identificar los países participantes en los trabajos analizados se han utilizado los campos de filiación institucional de los

autores. En el caso del SCI dicho campo recoge la filiación institucional de todos los autores, mientras que Pubmed solo recoge la filiación del primer autor.

En cualquier caso, el método aplicado para identificar el país de las filiaciones institucionales es por medio de consultas SQL, que sirviéndose de operadores booleanos, localizan el nombre del país en la filiación institucional de los autores el país y le asignan el acrónimo siguiendo la norma ISO 3166-1.

Como segundo paso, para asignar el país a las direcciones institucionales en las que no pudo localizarse en el primer paso, debido a que las direcciones se encontraban incompletas, se ha procedido a llevar a cabo un proceso similar al anterior pero, en lugar de utilizar los nombres de los países, se han utilizado los nombres de las principales ciudades de cada país. En este caso y debido al gran número de ciudades que esto supone y, por tanto, a la lentitud del proceso, se ha decidido limitar este proceso a los países de la Europa de los 15, que es en los que se centra este trabajo.

4.4.3.3. Normalización de direcciones Entre los resultados más interesantes de cualquier trabajo bibliométrico se encuentra el estudio de las instituciones que firman los trabajos, sin embargo, debido a las diferentes formas en que los autores consignan su lugar de trabajo, es un trabajo complicado y que suele conllevar una gran cantidad de tiempo dedicado a la normalización de los nombres institucionales.

4.4.3.3.1. Normalización de los tipos institucionales El primer paso que se ha aplicado para normalizar las instituciones pasa por identificar los diferentes tipos institucionales en los que pueden dividirse. En este caso se han utilizado fundamentalmente cuatro tipos institucionales, que recogen aproximadamente el 90 % de las instituciones participantes en los artículos analizados.

Los tipos institucionales son los siguientes: Universidades, Hospitales, Empresas y Fundaciones y Centros nacionales de investigación, que son las más institucio-

nes con una producción científica más significativa y ha sido utilizada de forma intensiva en estudios bibliométricos de reconocido prestigio, como el Observatorio IUNE de la Actividad Investigadora en la Universidad Española (SANZ CASADO et al., 2011).

Al igual que sucedía con los países, para identificar los tipos institucionales se han utilizado estrategias booleanas con los términos que más suelen aparecer en las direcciones en función del tipo al que correspondan.

4.4.3.3.2. Normalización de las instituciones En cuanto a la normalización de las instituciones concretas, ha tratado de realizarse de la forma más automática posible, pues el volumen de instituciones y de las formas diferentes en las que se consignan era inabarcable para un trabajo manual (más de 80.000 direcciones diferentes). Se ha programado un sistema que, mediante algoritmos de similaridad, permite identificar qué direcciones son lo suficientemente parecidas entre si como para considerar que se trata de la misma institución.

Como ya pudo apreciarse en el proceso de normalización de autores, el método implementado a través del módulo *String::Similarity* (LEHMANN, 2007) no resulta eficiente en cadenas de texto muy cortas, tales como los nombres de autor, pero si lo es en cadenas mayores y más estructuradas, como las direcciones de trabajo de los autores, por lo que ha sido el módulo utilizado para programar el sistema de normalización de instituciones.

Para realizar este trabajo se programó el siguiente script, que se ejecutó en dos tiempos diferentes. En un primer momento para identificar las direcciones completas más similares entre si, y posteriormente, para identificar la institución matriz, tomando tan solo la primera parte de las direcciones institucionales, pues es la que recoge la institución matriz, que viene seguida por los departamentos, secciones, etc. en los que trabajan los autores.

```
use String::Similarity::Group 'all';  
use DBI;  
use lib('../lib');
```



```

use CONECTAR;
use String::Trigram;
use Smart::Comments;
my @tipointst=("UNIV" ,"HOSPITAL" ,"HOSP-UNIV" ,"EMP-FDN" ,"NATL-RES");
print @tipointst;
foreach (@tipointst){
    print $_;
    my $tabla='isi_tdir_UE';
    my $dbh->{'mysql_auto_reconnect'} = 1;
    $dbh = Conectar->connect();
    my $sth =
        $dbh->prepare(" select _direc _FROM_ $tabla _WHERE_ pais
            != 'USA' _AND_ ( tipointst _=_ '$_' _OR_ tipointst='_' ) _
            GROUP_BY_ direc _ORDER_BY_ direc");
    $sth->execute();
    $array = $sth->fetchall_arrayref;
    my @elements;
    for my $lineas (@$array){ ### Evaluando %
        my $des = $lineas->[0];
        push @elements , $des;
    }
    my @groups = groups_lazy( 0.85, \@elements );
    foreach $x (0..@groups-1){ ### Agrupando %
        foreach $y (0..{@groups[$x]}-1){
            print " Normalizada: _$groups[$x][0]\n";
            $groups[$x][$y] =~ s/\'/\\'/g;
            $groups[$x][$y] =~ s/"/\\"/g;
            $dbh->do("UPDATE_ $tabla _SET_ direc2='$_groups
                [$x][0]' _WHERE_ direc='$_groups[$x][$y]'")
                ;
            print " Sin _normalizar: _$groups[$x][$y]\n";
        }
    }
}

```

De esta forma se agrupan las instituciones con una similaridad entre ellas mayor o igual al 85%. En la tabla 4.7 se muestran algunos ejemplos de instituciones normalizadas.

Firma original	Institución normalizada
Fdn Jimenez Diaz, Dept Neurol, E-28040 Madrid, Spain.	Fdn Jimenez Diaz
Fundac Jimez Diaz, Dept Genet, E-28040 Madrid, Spain.	Fdn Jimenez Diaz
Univ Complutense Madrid, Mi- tochondrial & Hereditary Metab Dis Unit, E-28040 Madrid, Spain.	Univ Complutense Madrid
Univ Complutense Madrid, Stat & Epidemiol Unit, E-28040 Ma- drid, Spain.	Univ Complutense Madrid
Univ Complutense, Dept Immu- nol, Fac Med, E-28040 Madrid, Spain.	Univ Complutense Madrid
Inserm, U29, Marseille, France.	Inserm
Inserm, Umr Neurol & Therapeut Expt S679, Paris, France.	Inserm
Inserm, U866, Ctr Rech, Lab Bio- chim Metab & Nutr, F-21000 Di- jon, France.	Inserm
Manchester Childrens Univ Hosp, Dept Resp Med, Manchester, Lancs, England.	Manchester Childrens Univ Hosp
Manchester Childrens Univ Hosp Nhs Trust, Manchester, Lancs, England.	Manchester Childrens Univ Hosp

Firma original	Institución normalizada
Manachester Childrens Univ Hosp, Nhs Trust, Dept Med Genet, Manchester, Lancs, England.	Manchester Childrens Univ Hosp
Appl Biosyst Inc, I-20052 Monza, Italy.	Appl Biosyst Inc
Appl. Biosyst Inc, Madrid, Spain.	Appl Biosyst Inc
Appl Biosyst Inc., Warrington, Cheshire, England.	Appl Biosyst Inc

Tabla 4.7: *Ejemplo de instituciones normalizadas*

Como se puede observar, el sistema permite agrupar, de forma masiva, las instituciones más similares entre si con un desviación mínima, que en un conjunto de datos limitado y revisado por el autor, utilizado como banco de pruebas, se encontraba alrededor del 1 %.

Debido a que el sistema está basado en algoritmos de similaridad y, al igual que ocurría en el método de normalización de autores, para conjuntos de datos más limitados, especialmente si es por razones geográficas, sería necesario ajustar los niveles de similaridad utilizados para minimizar el error en la asignación.

4.4.3.3.4. Normalización de las temáticas Debido a que la metodología comprende la utilización de dos bases de datos (SCI y Pubmed) de características muy diferentes, se hace imprescindible contar con un sistema que permita equiparar las materias de las que tratan los documentos en las dos bases de datos con el fin de hacer comparables los resultados obtenidos en ambas.

Puesto que Pubmed cuenta con un sistema de asignación de materias centrado específicamente en medicina (Medical Subject Headings) se decidió programar

un método que permitiera asignar los términos del MeSH a los documentos obtenidos del SCI. Para ello se construyó un sistema en dos fases que permitiera primero, asignar a través de las palabras clave los términos MeSH correspondientes a través del Unified Medical Language System (UMLS) y segundo, a través de algoritmos de similitud, asignar los términos MeSH a aquellas palabras clave que, sin contar con asignación en la primera fase, cuente con una similitud lo suficientemente alta como para asignarle un término MeSH sin riesgo de asignarlo erróneamente.

A continuación se describen ambos procesos de forma pormenorizada.

4.4.3.3.4.1. Fase 1: UMLS y términos MeSH El UMLS es un conjunto de software y ficheros que unifican diferentes vocabularios y estándares biomédicos con el fin de permitir la interoperabilidad entre sistemas informáticos y los diferentes vocabularios controlados que pueden utilizarse en las ciencias médicas.

El papel de este sistema en nuestro proceso de normalización viene marcado por la disponibilidad de un módulo programado en Perl que permite comparar términos no controlados con el macro-tesauro UMLS y obtener aquellos términos controlados que son equivalentes.

Aunque la base de datos que proporciona el UMLS es mucho mayor y cuenta con una gran variedad de lenguajes controlados (MeSH, SNOMED, RxNorm, LOINC, etc.) en nuestro trabajo se ha utilizado tan solo el lenguaje MeSH debido a que se contaba con la base de datos Pubmed normalizada con dicho lenguaje.

Para llevar a cabo este proceso, se comparó cada una de las palabras clave que los autores han asignado a sus propios trabajos en la base de datos SCI, con el macro-tesauro UMLS y, en el caso de que dicha palabra clave coincida con un término presente en el tesauro (ya sea un término aceptado o no) se le asigna el término aceptado que le corresponda. De este modo, salvo que la palabra clave no se encuentre en el tesauro (algo complicado, debido al enorme tamaño del macro-tesauro UMLS, a no ser que exista algún error tipográfico), se le asigna un

término aceptado proveniente del tesoro MeSH, permitiéndonos contar con un sistema de materias equivalente al utilizado en la base de datos Pubmed.

4.4.3.3.4.2. Fase 2: Asignación a través de algoritmos de similitud La segunda fase de la normalización consiste en comparar, por medio de un algoritmo de similitud basado en n-grams, las palabras clave que no han recibido asignación en la fase 1 con aquellas que si lo han hecho. El objetivo es subsanar fundamentalmente errores de tipo ortográfico y tipográfico, ya que estas palabras clave han sido asignadas manualmente por los autores.

Tras múltiples pruebas, se encontró que el porcentaje de similitud ideal era del 70 %, por lo que se decidió que si la similitud entre dos palabra clave supera ese umbral, se le asigna el término MeSH correspondiente, mientras que si no lo supera la palabra clave no recibe asignación. De este modo es posible asignar muchas de las diferentes variantes ortográficas de los términos.

En el gráfico 4.3 se encuentra descrito el proceso de normalización en sus dos fases. Utilizando este proceso de normalización, del total de documentos únicos obtenidos a través de la base de datos SCI (19.415), fue posible asignar al menos un término MeSH aceptado al 77,39 % (15.025), describiendo de forma más completa, el contenido del documento.

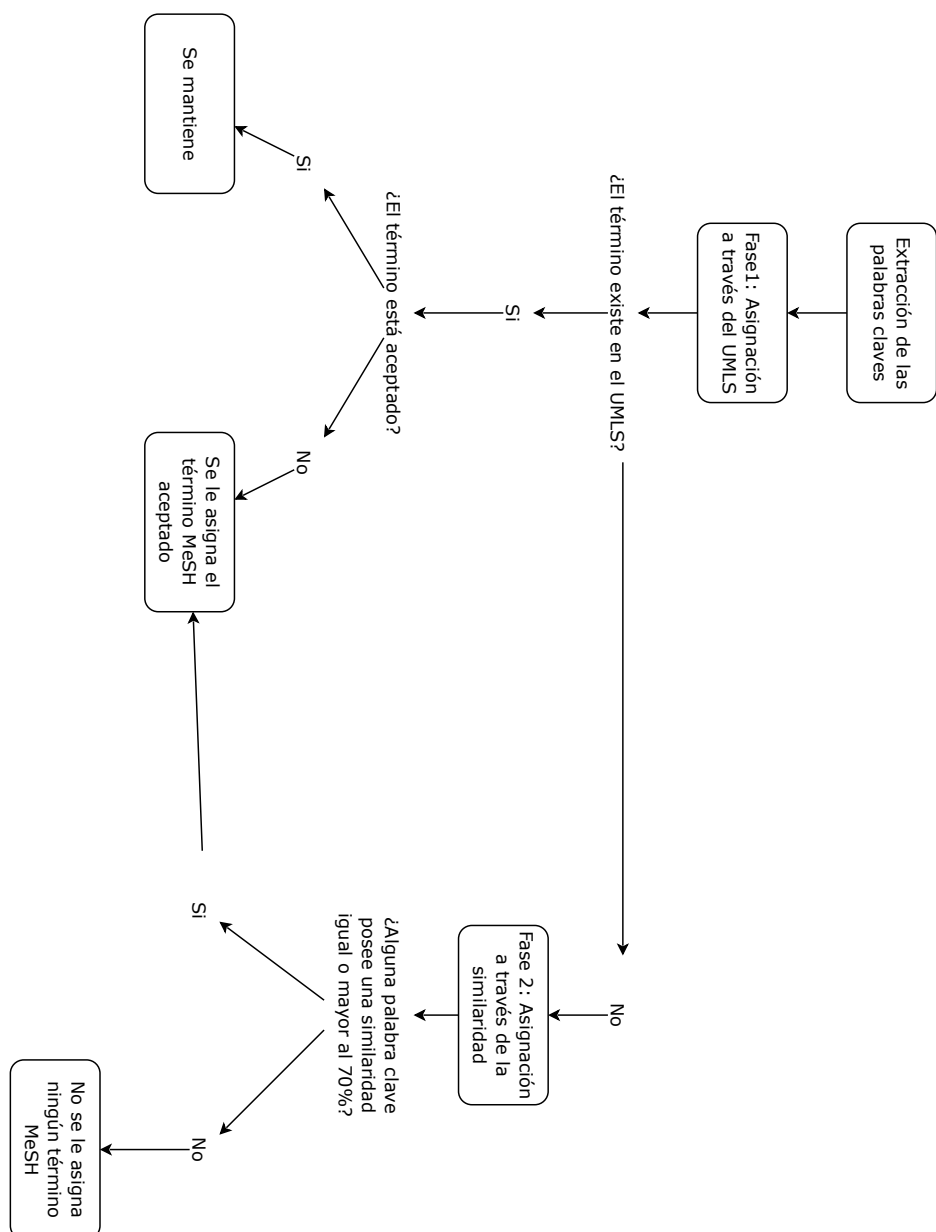


Figura 4.3: Proceso de normalización temática

Sección 5

Análisis estadístico

5.1. Definición de indicadores métricos

Una vez concluidos los procesos de carga y normalización de los datos, se ha establecido la siguiente categoría de indicadores, distinguiendo entre indicadores de Actividad Científica (tanto de producción como de consumo) e indicadores de Impacto / Visibilidad.

Cabe destacar que los, debido al volumen de datos con el que se ha trabajado, los siguientes indicadores han sido obtenidos sobre el conjunto general de datos, puesto que la extracción de los indicadores de forma agregada por enfermedades o países supone un trabajo mucho mayor que el planteado para esta tesis doctoral, si bien este autor pretende continuar con su desarrollo con el fin de caracterizar mucho mejor la actividad científica en torno a las enfermedades raras estudiadas en este trabajo.

5.1.1. Indicadores de Actividad Científica.

Como se observa en las tablas 5.1, 5.2 y 5.3, para calcular los indicadores seleccionados va a ser necesario realizar el recuento de la producción científica y

Tipo	Indicador/es	Nivel de agregación
Producción Anual	Recuentos de la producción anual, Evolución de la producción anual y Números índice	País y base de datos
Tipología documental	Recuentos y % por tipología	Base de datos
Temática	Recuentos y % por temática (Clasificación ICD-10, categorías WOS y términos MeSH)	Base de datos
Países	Recuentos y % por países	Base de datos
Instituciones	Recuentos, % de producción institucional y Tipo de institución	País
Idioma	Recuentos y % de idiomas en los que se publica	Base de datos
Colaboración	Índice de coautoría, Índice de colaboración y porcentaje de colaboración nacional e internacional	País, base de datos y temática
Dispersión	Ley de Bradford e Índice de concentración de Pratt	Base de datos
Productividad de los autores	Ley de Lotka, Índice de élite de Price e Índice de Yablonsky	Base de datos

Tabla 5.1: *Indicadores de Actividad Científica (Producción)*

Tipo	Indicador	Nivel de agregación
Obsolescencia	Vida Media, Índice de Price	Temática
Temática	Recuentos y % por temática (Clasificación ICD-10)	Ninguno
Aislamiento	Índice de aislamiento	País
Dispersión	Ley de Bradford	Ninguno

Tabla 5.2: *Indicadores de Actividad Científica (Consumo)*

en muchas casos aparecerán múltiples ocurrencias dentro de campos como el de autor, lugar de trabajo, país e idioma e incluso para diferentes categorías temáticas, que en algunos casos pueden solaparse. Por este motivo y antes de definir

Tipo	Indicador	Nivel de agregación
Impacto	Recuento de documentos publicados en revistas del primer cuartil (corregido)	País, temática, tipo institucional e institución
Inmediatez de las citaciones	Índice de inmediatez	Temática
Análisis de citas	Media de citas por documento, Recuento y % de citas y Documentos no citados	País y temática, tipo institucional e institución
Ranking de revistas	Posición de las revistas en las que se publica (FI) y distribución por cuartiles	Núcleo de producción SCI, país e idioma

Tabla 5.3: *Indicadores de Impacto / Visibilidad*

Indicador	Nivel de agregación
Escalado multidimensional (MDS)	Autores, instituciones, temáticas y países
Análisis de clusters	Autores, instituciones, temática y países
Análisis de correspondencias	Autores, instituciones, temáticas, países y publicaciones
Análisis de redes sociales	Autores, autores citados, instituciones y países

Tabla 5.4: *Indicadores multidimensionales y conexionistas*

los indicadores que se van a utilizar es necesario dejar claro que para todos estos recuentos se va a utilizar el método de recuento completo, es decir, atribuyéndole el trabajo de forma completa a cada ocurrencia dentro de un campo, frente a los tipos de recuentos fraccionado, para el primer autor y el recuento ponderado por el grado de participación de los autores (PRAVDIC y OLUIC-VUKOVIC, 1991). De esta forma los trabajos con varios autores contarán, para cada uno de ellos como si fuese el único autor del mismo.

El principal motivo para la elección de este tipo de asignación es por considerarla más justa para los autores, pues en un trabajo científico siempre es difícil determinar el grado de implicación de cada autor y mediante este método pueden

eliminarse ambigüedades respecto a la posición en que aparecen los autores, independientemente del criterio seguido para determinar el orden de firma. Además, este tipo de asignación sirve para fomentar la colaboración entre investigadores, pues considera el trabajo de cada uno por igual, algo que sólo puede derivar en beneficios para la investigación científica (NUDELMAN y LANDERS, 1972).

Por otro lado, el nivel de agregación correspondiente a las bases de datos, se debe a que las dos bases de datos que han servido para la recuperación de los registros bibliográficos (SCI y Pubmed), han sido gestionadas y utilizadas por separado, pues el solapamiento entre ellas no era lo suficientemente alto como para utilizarlas de forma conjunta. Por lo tanto, es necesario que los indicadores sean calculados de forma independiente para cada una de las bases de datos, salvo aquellos indicadores (especialmente los relativos al análisis de citas y referencias bibliográficas) que han sido calculados únicamente para los datos obtenidos del SCI.

5.1.1.1. Indicadores de Producción.

Entre todos los indicadores bibliométricos de producción que se pueden utilizar para determinar la actividad científica se han elegido los siguientes para nuestro estudio:

5.1.1.1.1. Producción anual. Este indicadores es probablemente el más sencillo de calcular y se compone básicamente de tres medidas.

5.1.1.1.1.1. Recuento de publicaciones Esta medida se basa en el número de trabajos publicados en cada uno de los años estudiados, utilizando para calcularlo las dos bases de datos seleccionadas (SCI y Pubmed) y la estrategia de búsqueda que previamente han sido diseñado para recuperar el conjunto de trabajos que serán objeto de este estudio.

Evolución de la producción anual. A partir del recuento de las publicaciones se va a determinar cómo se ha producido la evolución temporal en la producción de trabajos sobre Enfermedades Raras en el periodo estudiado y si ésta ha sido favorable o no.

5.1.1.1.1.2. Tasa de crecimiento Se trata de una medida muy presente en áreas como la Economía o la Demografía, dónde se utiliza para determinar el crecimiento relativo, ya sea de la población, en el caso de la Demografía, o de diferentes indicadores económicos, tales como los índices de precios, el Producto Interior Bruto (PIB), etc. En el caso de los estudios métricos de información, esta medida consiste en determinar el porcentaje en el que crece la producción en el año en el que se calcula respecto al año anterior, ya sea este crecimiento positivo o negativo. Por tanto permite observar las variaciones en el número de trabajos publicados a lo largo de los años en forma de porcentaje.

Para ello se utiliza la siguiente formula:

$$I_0^t = \frac{\gamma_i^t - \gamma_i^{t-1}}{\gamma_i^{t-1}} \times 100 \quad (5.1)$$

Donde:

γ_i^t = Valor de la variable en el año actual

γ_i^{t-1} = Valor de la variable en el año anterior

5.1.1.1.1.3. Tipología documental. Se trata también de un indicador básico, que consiste en el recuento de publicaciones en función de cuál sea la tipología documental de los trabajos. Para ello se utilizará como nivel de agregación la base de datos de procedencia de los registros (SCI o Pubmed). En este sentido se tendrán en cuenta tanto la tipología formal (artículo de revista, monografía, actas de congreso, etc.), cómo el tipo de estudio realizado que recoge Pubmed, y que resulta

de gran interés en Medicina, incluyendo categorías como estudios de caso, ensayos clínicos, estudios comparativos, etc.

5.1.1.1.2. Temática. Para calcular la distribución temática se van a realizar por separado las búsquedas sobre las bases de datos, agrupando las enfermedades por categorías de la ICD-10. De esta forma obtendremos, para cada base de datos, los recuentos y porcentajes de trabajos publicados para las enfermedades incluidas en cada categoría temática.

5.1.1.1.3. Países. Este indicador lo se calcula a partir de los resultados obtenidos en el recuento de publicaciones por año, obteniendo la suma de todos los trabajos publicados por año en cada país, gracias a la normalización que permite identificar los países de cada una de las filiaciones institucionales de los autores.

5.1.1.1.4. Instituciones. En las bases de datos utilizadas (SCI y Pubmed), los registros cuentan con un campo de dirección, que en el caso del SCI recoge el lugar de trabajo de cada uno de los autores y en el caso de Pubmed tan sólo del primero. Por tanto, debido a esta limitación de Pubmed, para calcular este indicador tan sólo se utilizarán los datos descargados del SCI. Este indicador va a calcularse en dos niveles de agregación diferentes:

Recuento de producción por institución. Esta medida se calcula mediante el recuento y la obtención de porcentajes de los trabajos publicados por los autores que trabajan en cada institución, obteniendo de esta forma las instituciones más productivas en cada país. Dicho esto, la distribución de publicaciones por institución se va a calcular mediante dos recuentos complementarios, uno general y otro con las instituciones más productivas por país.

Recuento de producción por tipo de institución. Para llevar a cabo estos recuentos es preciso realizar, como paso previo, un script en lenguaje Perl que permita asignarle a cada institución la categoría que le corresponda (Hospital,

Universidad, Centros de investigación, etc.), de forma que se obtengan el recuento y porcentaje de trabajos producidos en función del tipo de institución generadora para cada país.

5.1.1.1.5. Idiomas. Para determinar la distribución idiomática de los trabajos publicados, se llevarán a cabo los recuentos de trabajos publicados en cada idioma, utilizando el método de asignación completa en el caso de trabajos publicados en más de una lengua y se calculará la distribución idiomática para ambas bases de datos.

5.1.1.1.6. Colaboración. La colaboración entre investigadores es uno de los factores más importantes para determinar la calidad de la investigación realizada, pues generalmente, la colaboración entre autores, instituciones y países produce un estado de sinergia mediante el cual, el resultado obtenido de la colaboración es mayor que la suma de las aportaciones de cada uno de los agentes.

De este modo en un entorno altamente competitivo, como es hoy en día el mundo de la Ciencia, la colaboración es primordial para conseguir los mejores resultados posibles, pues los autores, instituciones y países que establecen acuerdos de colaboración obtienen una serie de ventajas mutuas como son: evitar la duplicidad de esfuerzos, compartir y optimizar los recursos disponibles, así como la transferencia de conocimientos y “know how” que se produce entre ellos (MALTRÁS, 2003).

Para determinar si se produce esta colaboración en el área estudiada, y si es así, establecer cuáles son sus características, van a aplicarse varios indicadores seleccionados del Diccionario de Bibliometría elaborado por SPINAK (1996). Llegados a este punto es necesario aclarar que todos los indicadores seleccionados se van a aplicar para dos niveles de agregación diferentes, por países y por temática.

5.1.1.1.6.1. Índice de coautoría Este indicador permite obtener la media de autores firmantes por trabajo y se calcula como:

$$ICoaut = \frac{\sum j_i n_j}{N} \quad (5.2)$$

Donde:

j_i = Número de autores por documento

n_j = Cantidad de documentos con j autores

N = Total de documentos

5.1.1.1.6.2. Índice de colaboración Al igual que el Índice de coautoría proporcionaba la media de autores firmantes por documento, el Índice de colaboración se basa en la media de instituciones firmantes por documento. A continuación se indica la forma de calcularlo, pero antes es necesario destacar que este indicador sólo va a aplicarse a los registros de la base de datos SCI, puesto que la otra base de datos utilizada (Pubmed) tan sólo proporciona el lugar de trabajo del primer autor.

$$ICol = \frac{\sum j_i n_j}{N} \quad (5.3)$$

Donde:

j_i = Número de instituciones firmantes por documento

n_j = Cantidad de documentos con j instituciones firmantes

N = Total de documentos

5.1.1.1.6.3. Porcentaje de documentos en colaboración nacional e internacional Este indicador muestra el porcentaje de documentos en colaboración respecto al total de documentos producidos, en dos niveles de agregación diferentes, por un lado los documentos en colaboración con instituciones del propio país y por otro aquellos en los que intervienen instituciones foráneas.

5.1.1.1.7. Dispersión. Al tratar sobre la distribución de trabajos científicos en las publicaciones dentro de un área concreta, el referente inmediato es BRADFORD (1934) y su Ley sobre la dispersión de los artículos de revista.

5.1.1.1.7.1. Ley de Bradford Su trabajo le llevó a enunciar ésta Ley en dos formulaciones. La formulación gráfica enunciada por (BRADFORD, 1934) demostraba que el logaritmo acumulado de revistas y el número de artículos publicados en ellas se ajustaba a una recta dividida en tres zonas diferenciadas, siendo la primera o “núcleo” dónde se concentra la mayor parte de los trabajos en un número relativamente pequeño de revistas. Esta zona puede identificarse claramente en la representación gráfica por ser la que se ajusta a una curva exponencial. La formulación verbal de la Ley de Bradford ha sido objeto de múltiples reinterpretaciones (BROOKES, 1969; LEIMKUHNER, 1967), debido a que no se ajusta exactamente a la formulación gráfica de la misma.

Para este trabajo va a seguirse el método propuesto en la formulación original de Bradford, consistente en ordenar las revistas científicas en orden decreciente de productividad y dividir las en tres zonas que contengan, de forma aproximada, el mismo número de artículos publicados y, de esta forma, hallar que revistas componen el núcleo que recoge la mayor parte de trabajos científicos sobre Enfermedades Raras en cada una de las bases de datos.

5.1.1.1.7.2. Concentración temática de las revistas De forma complementaria a la Ley de Bradford va a calcularse la concentración temática de las revistas para aquellas publicaciones que componen el núcleo de Bradford, con el fin de determinar si las publicaciones que recogen un mayor número de trabajos son revistas que publican exclusivamente sobre una temática o si se trata de revistas orientadas a múltiples áreas. Para ello obtenemos, de cada revista el número de categorías temáticas en las que está incluida. Es necesario señalar que para el cálculo de este indicador van a utilizarse exclusivamente los datos obtenidos del SCI, puesto que

es esta base de datos la que, a través de su herramienta “Journal Citation Report” (JCR) permite categorizar el área temática de las diferentes revistas.

5.1.1.1.8. Productividad de los autores.

La productividad suele medirse por la cantidad de publicaciones producidas por un autor, una institución o un país (SPINAK, 1996).

En nuestro caso, el estudio de la productividad de los autores que estudian las Enfermedades Raras con causa genética identificada, será de utilidad para determinar, en cada base de datos, cuáles son los autores e instituciones más productivos. Para ello van a utilizarse las siguientes medidas:

5.1.1.1.8.1. Ley de Lotka Como primera medida de la productividad de los autores, es imprescindible mencionar el trabajo de LOTKA (1926), en el que formuló su famosa ley según la cuál el número de autores A_n que publica n trabajos es inversamente proporcional al cuadrado de n :

$$A_n = \frac{A_1}{n^2} \quad (5.4)$$

Donde:

n = Número de observaciones

A_1 = Número de trabajos con un sólo autor

Para ello va a utilizarse la metodología propuesta por PAO (1982), que consiste en linealizar el modelo potencial de Lotka para determinar las constantes necesarias A y m , pues aunque Lotka las formuló como el número de trabajos con un sólo autor y “2” respectivamente, estos valores sólo son válidos para la formulación original, mientras que la modificación realizada por Pao para linealizar el modelo se ajusta más a la distribución real:

$$x^\beta y = c \quad (5.5)$$

Donde: c y β son constantes, y para $x = 1$ se deduce c

De esta forma el método propuesto por Pao pasa por recopilar los datos, realizar la distribución de frecuencias para $X = \log x$, $Y = \log y$ y $c = \log c$, siendo x el número de trabajos, e y el número de autores y tener una colección de datos con al menos 20 observaciones. De esta forma la formulación se corresponde con la ecuación de una recta dónde β es la pendiente de la misma. Posteriormente es necesario determinar β mediante el método de los mínimos cuadrados:

$$\beta = \frac{N \sum XY - \sum X \sum Y}{N \sum X^2 - (\sum X)^2} \quad (5.6)$$

Donde:

N = Número de rangos observados

X = Logaritmo del número de trabajos

Y = Logaritmo del número de autores

Una vez hallado β ya puede calcularse la constante c :

$$C = \frac{1}{\sum_1^{P-1} \frac{1}{X^\beta} + \frac{1}{\beta - 1} P^{\beta-1} + \frac{1}{2P^\beta} + \frac{\beta}{24} (P - 1)^{\beta-1}} \quad (5.7)$$

Donde:

P = Número de rangos observados.

Una vez completa la ecuación de la recta correspondiente al modelo de Lotka, es recomendable utilizar una prueba de contraste para determinar la bondad de ajuste entre la distribución observada de acumulación de frecuencias y los valores

teóricos, pudiendo utilizarse para este fin diferentes test de bondad de ajuste como el de Kolmogorov-Smirnov o la T de Student.

Índice de Élite de Price. Como medida para determinar el número de autores significativamente más productivos en un área científica, PRICE (1973) propone éste índice según el cual el número de autores más productivos se corresponde con la raíz cuadrada del total de autores:

$$EP = \sqrt{n} \quad (5.8)$$

Donde n es igual al número total de autores.

Ésta medida se justifica por el hecho de que:

El número de científicos se duplica cada 10 años, mientras que el de [científicos] notables lo hace solamente cada 20 (PRICE, 1973)

Índice de Yablonsky. Al igual que la Élite de Price, esta medida tiene el objetivo de determinar cuantos autores componen la élite o el conjunto de mayores productores en una materia determinada. Para su cálculo YABLONSKY (1980) establece que, los investigadores con un número de trabajos equivalente a la raíz cuadrada del número de autores con un solo trabajo pueden considerarse muy productivos, y matemáticamente se expresa como:

$$N = \sqrt{A_1} \quad (5.9)$$

Donde

N = Número de trabajos necesarios para que un autor sea considerado un gran productor.

A_1 = Número de autores con un sólo trabajo.

5.1.1.2. Indicadores de Consumo.

En nuestro caso es imprescindible mencionar que los indicadores de consumo van a calcularse exclusivamente con los datos obtenidos de la base de datos SCI, puesto que Pubmed no proporciona los datos de las referencias bibliográficas que contienen los artículos indizados.

5.1.1.2.1. Obsolescencia. Según LOPEZ PIÑERO y TERRADA (1992b) éste indicador, junto con el de aislamiento que se tratará a continuación, son los más importantes a la hora de estudiar el consumo de información. SPINAK (1996) define la obsolescencia como:

La tendencia de las publicaciones a ser citadas con mayor frecuencia poco tiempo después de ser publicadas, y esta frecuencia de uso decae rápidamente con el tiempo

Con el objetivo de cuantificar ésta obsolescencia van a calcularse las siguientes medidas utilizando temática y países como niveles de agregación:

Semiperiodo. La vida media, o más correctamente, el semiperiodo de BURTON y KEBLER (1960) se define como:

El tiempo durante el cual ha sido publicada la mitad de la literatura activa circulante sobre una materia.

Para calcular ésta medida se utiliza la mediana de las fechas de publicación de las referencias bibliográficas, aplicando la siguiente fórmula:

$$V\mu = k + \left(\frac{\frac{N}{2} - a_1}{a_2 - a_1} \right) \quad (5.10)$$

Donde:

k = Número entero de años necesarios para acumular la mitad de la literatura activa

N = Total de la literatura activa circulante en el periodo de estudio

a_1 = Valor % acumulado anterior a la mitad de la literatura circulante (a)

a_2 = Valor % acumulado posterior a la mitad de la literatura circulante (a)

Índice de Price. Éste indicador de obsolescencia de la literatura científica indica cómo de recientes son las citas en una disciplina científica y se calcula como el porcentaje de referencias bibliográficas que tienen 5 años o menos de antigüedad respecto al total de referencias (PRICE, 1973).

De este modo, aplicando los indicadores mencionados, es posible conocer cómo de volátil es la literatura científica en el ámbito de las Enfermedades Raras.

5.1.1.2.2. Aislamiento. La finalidad de éste indicador es medir la capacidad que tienen los investigadores de utilizar la literatura científica producida en países diferentes de los suyos propios. La forma de calcularlo es determinando los porcentajes de documentos referenciados que pertenecen al mismo país que la dirección institucional de los autores de los trabajos fuente. Por tanto, cuanto menor sea este porcentaje, menor será también el aislamiento de ese país en cada materia concreta.

5.1.1.2.3. Temática. Del mismo modo que para los indicadores de producción se calculó la frecuencia absoluta y el porcentaje de trabajos publicados en cada temática, en el caso del consumo de información va a calcularse este mismo indicador pero, en este caso, sobre las referencias bibliográficas incluidas en cada trabajo publicado, utilizando las categorías JCR en las que se incluyen las revistas citadas para determinar la temática de las mismas. De esta forma es posible determinar, cuáles son las temáticas que tienen más difusión y son más utilizadas por los investigadores en enfermedades raras.

5.1.2. Indicadores de Impacto / Visibilidad.

En el punto 1.3 se hizo mención a la evaluación por expertos y el “peer review”, y como BORDONS y ZULUETA (1999) opinaban que se trata del sistema más adecuado para evaluar científicos y grupos de investigación. Sin embargo, es habitual utilizar los indicadores bibliométricos llamados de Impacto y Visibilidad para evaluar los avances científicos, grupos de investigación, universidades o incluso a los propios científicos de forma individual, como en el caso del Índice H (HIRSCH, 2005).

Por lo tanto es posible plantearse la siguiente cuestión: ¿Cómo de válidos son estos indicadores para determinar la calidad de las aportaciones científicas?. Algunos autores ya han realizado trabajos, y llegado a conclusiones acerca de la validez de los indicadores bibliométricos. Es el caso de MEHO y SONNENWALD (2000), que determinaron que existe una gran correlación entre el número de citas recibidas y la evaluación por pares, en aquellos investigadores y grupos que se encontraban o bien en los puestos más altos en el “ranking” de su disciplina o en los más bajos, y la correlación era menor para las posiciones intermedias.

AKSNES (2006) llegó a las mismas conclusiones al comparar el recuento de citas de trabajos científicos con las opiniones de otros autores acerca de ellos, encontrando resultados muy similares para los mayores y menores productores.

En función de estos trabajos es posible deducir que, el análisis de citas y los indicadores bibliométricos destinados a medir el impacto de las contribuciones científicas, son totalmente válidos si nuestro objetivo es determinar cuáles son los investigadores, grupos, universidades, etc., más importantes en una disciplina científica, y se vuelven más imprecisos a la hora de tratar las contribuciones de calidad intermedia.

La repercusión de la actividad científica y su cuantificación, es uno de los factores que más interés ha despertado en el área de la bibliometría, debido a que permite cuantificar también la calidad de la investigación realizada. En esta línea se han desarrollado múltiples indicadores, destinados a medir el impacto de las publicaciones.

En un principio estos indicadores se servían del análisis de citas para determinar el impacto de publicaciones tales como las revistas científicas, pero su uso se ha ido modificando y hoy en día son una de las principales metodologías empleadas en la evaluación de la ciencia y los científicos y por tanto en su financiación, por lo que no siempre han sido vistos con buenos ojos por parte de la comunidad científica.

En este caso, la utilización de indicadores de impacto, va a ser de utilidad para conocer cuál es la repercusión de la investigación realizada en Enfermedades Raras. La forma más habitual de cuantificar esta repercusión es a través las revistas en las que se publican los trabajos científicos. Entre los indicadores que se pueden utilizar para este fin se encuentran los siguientes:

5.1.2.1. Factor de Impacto.

Éste indicador, propuesto por GARFIELD (1972) en el marco de las bases de datos Thomson-Reuters (originariamente ISI), es calculado anualmente para las revistas incluidas en el Journal Citation Reports (JCR) y además:

Se ha convertido en el producto bibliométrico más popular utilizado tanto en bibliometría como fuera de la comunidad científica. Sin embargo su popularidad se ha vuelto también peligrosa. El tratamiento arbitrario de los datos, la ponderación arbitraria de sus componentes e indicadores y el uso y aplicación desinformado y tendencioso de las citas a revistas ha dañado también la credibilidad de la investigación bibliométrica y la tecnología (GLÄNZEL y MOED, 2002).

En un principio, el Factor de Impacto tal y como lo planteó Garfield, se calcula de la siguiente forma:

$$FI = \frac{CIT(Y, Y - 1) + CIT(Y, Y - 2)}{PUB(Y - 1) + PUB(Y - 2)} \quad (5.11)$$

Donde:

Y = Año para el cual se va a calcular el FI

CIT = Se refiere a las citas recibidas

PUB = Se refiere a los trabajos publicados

De esta forma el cálculo del Factor de Impacto para una revista en el 2009 sería el siguiente:

$$FI = \frac{CIT(2009, 2008) + CIT(2009, 2007)}{PUB(2008) + PUB(2007)} \quad (5.12)$$

Es decir, que habría que dividir el número de citas que recibe la revista en 2009 por trabajos publicados en 2008 y 2007, entre el total de trabajos publicados por

dicha revista en 2008 y 2007. De esta forma, la media del Factor de Impacto obtenido por una revista a lo largo del tiempo es lo que se conoce como **Factor de Impacto Medio**.

Sobre esta formulación original se han formulado múltiples correcciones que tratan de corregir algunos de los sesgos que genera el factor de impacto.

Entre ellas, debido a que en las revistas no sólo se publican trabajos científicos, sino que en muchas ocasiones se incluyen editoriales, reseñas, etc. que normalmente no reciben citas, SCHUBERT y GLÄNZEL (1983) propusieron modificar el cálculo del Factor de Impacto para incluir tan sólo aquellas contribuciones que consideraran “ítems citables”, tales como artículos, revisiones, cartas, etc. Esta medida que enmienda dicha limitación se denomina Factor de Impacto Corregido.

Por otro lado, el Factor de impacto global, propuesto por ROUSSEAU (2002) se diferencia en que:

El primero (por el Factor de Impacto Medio) es una media de cocientes, mientras que el segundo (por el Factor de Impacto Global) es el cociente de las medias.

y se calcula mediante la aplicación de la siguiente fórmula:

$$FIG = \frac{\sum_{j=1}^n C_j}{\sum_{j=1}^n P_j} = \frac{\mu_c}{\mu_p} \quad (5.13)$$

Donde C se refiere al número de citas recibidas en un periodo concreto y P hace referencia al número de trabajos publicados por la revista en el mismo periodo. En resumen, el Factor de Impacto Global sería el cociente de la media de citas recibidas entre la media de trabajos publicados. De esta forma no se penaliza el hecho de que una revista cuente con un menor número de trabajos publicados, siempre y cuando el número de citas recibidas por ellos sea proporcional.

El Factor de Impacto Medio de las Revistas donde se producen las Citas (FIMRC) y el Factor de Impacto Ponderado fueron propuestos por BUELA-CASAL (2003).

En primer lugar el FIMRC:

Se refiere al promedio del Factor de Impacto de las revistas donde fue citada la revista. Se calcula multiplicando el Factor de Impacto de cada revista (en ese año) en la que se citan artículos de los años anteriores de la revista analizada por el número de artículos citados en cada revista, la suma total se divide por el número total de artículos citados.

De este modo aquellas revistas citadas por otras con un Factor de Impacto alto se verán recompensadas por ello, provocando un fenómeno similar al famoso *Page Rank* por el que Google posiciona los resultados de búsqueda. Este indicador, expresado matemáticamente, responde a la siguiente formulación:

$$FIMRC = \frac{\sum FI_n Cit}{AC} \quad (5.14)$$

Donde:

FI_n = Factor de Impacto de la revista citante

Cit = Número de citas recibidas de la revista n

AC = Total de trabajos citados

El principal problema de este indicador es el de la regresión a la media, y es que las revistas con un Factor e Impacto muy alto necesariamente recibirán sus citas de revistas con Factor de Impacto igual o inferior al suyo. Para corregir esta limitación el propio autor propone el Factor de Impacto Ponderado que se define como el promedio de ambas medidas, de forma que:

$$FIP = \frac{FI + FIMRC}{2} \quad (5.15)$$

Por todo lo expuesto, el Factor de Impacto, principalmente en su formulación original y a pesar de su aparente *simplicidad*, posee algunos inconvenientes que desaconsejan su uso como medida del impacto o visibilidad de los trabajos científicos. Según GLÄNZEL y MOED (2002) sus principales inconvenientes residen en que:

1. No se encuentra normalizado para las prácticas y tradiciones de citación en los diferentes campos y disciplinas científicas.
2. No tiene en cuenta el reconocimiento de la revista citante.
3. Se encuentra sesgado a favor de los trabajos más extensos, como por ejemplo las revisiones.
4. La frecuencia de citación está sujeta a desviaciones a lo largo del tiempo.
5. No cuenta con un indicador del grado de desviación estadística.
6. La media de tiempo en que un artículo de revista es citado desde su publicación no es siempre de dos años.
7. Una única medida no es suficiente para determinar el patrón de citación de las revistas científicas.
8. El concepto de “documento citable” no está correctamente definido.
9. Existen errores en su cálculo a la hora de identificar las revistas citadas, sobre todo en revistas de habla no inglesa.

Por estos motivos se considera más conveniente utilizar indicadores complementarios. Sin embargo, puesto que no es posible eliminar los sesgos que se derivan del uso de estas bases de datos, debido a que los productos de Thomson-Reuters son los que proporcionan un recuento de citas más completo para nuestro trabajo, si que es posible hacer un uso más consecuente de ellos y utilizar indicadores complementarios o reelaborados. Por este motivo, aunque se utiliza el Factor de Impacto, en su formulación original, para obtener una medida de la calidad de las revistas, a la hora de calcular la posición de las mismas dentro de las revistas de la misma área temática, se calcula un cuartil corregido, basado en una medida de ponderación diferente para determinar el mejor cuartil para cada revista.

También basados en el número de citas, existen otros indicadores, como el “Crown Field Indicator”, basado en el número de citas que reciben las publicaciones de un área concreta en un periodo de tiempo específico, comparado con la media

mundial de citas que reciben las publicaciones sobre los mismos tipos de documentos con la misma edad y temáticas, de forma agrupada. Sin embargo, este tipo de indicador queda fuera del objeto de estudio de este trabajo, puesto que el área a tratar es muy heterogénea y no puede encuadrarse dentro de una temática concreta con la que pueda establecerse una comparación a nivel de citas recibidas.

Lo mismo ocurre con otros indicadores como el Journal packet citation score (JCSm), puesto que, a pesar de su utilidad para calcular el impacto de las revistas en un área concreta, la heterogeneidad de los datos estudiados, conlleva que no pueda realizarse una comparación con las citas recibidas por las revistas de un área concreta.

5.1.2.2. Análisis de citas.

El análisis de citas ha recibido críticas de algunos autores, como por ejemplo MORAVCSIK y MURUGESAN (1975), MACROBERTS y MACROBERTS (1996) y CAMÍ (1997), debido sobre todo al sesgo de las citaciones, y al hecho de que éstas se utilicen para evaluar la calidad de los trabajos publicados. La mayor parte de las críticas hacia el análisis de citas se deben a que:

Los científicos no citan la mayor parte de sus influencias; la citación es sesgada, las fuentes secundarias sustituyen a las fuentes primarias, y las fuentes informales, que son la mayor parte, no son reconocidas (MACROBERTS y MACROBERTS, 1996).

Sin embargo otros trabajos, como los de ARENCIBIA y MOYA ANEGÓN (2008), GARFIELD y MERTON (1979), ponen de manifiesto que, el uso de las referencias bibliográficas para la elaboración de indicadores bibliométricos empleados en la evaluación de la Ciencia y la Tecnología es totalmente válido, pues tal como señala CAÑEDO ANDALIA (1999):

el número de citas que recibe un trabajo, no puede considerarse por sí solo como una medida de su calidad científica propiamente dicha, aunque la sistematicidad de la citación, cuando es producto de una

selección consciente por parte de los diferentes autores, puede indicar, además de su utilidad, el cumplimiento de ciertas normas generales de calidad científica exigidas, a menudo, por los investigadores como para considerar a los trabajos en realidad valiosos.

Asímismo BORNMANN et al. (2008) señalan que es necesario tener en cuenta una serie de recomendaciones a la hora de utilizar indicadores bibliométricos basados en el recuento de citas bibliográficas:

(1) El impacto científico de los grupos de investigación debe ser examinado de forma diferenciada, utilizando diagramas de caja así como curvas de Lorenz y el índice de Gini. (2) Las diferentes normas sobre referencias son utilizadas para evaluar el impacto de los grupos de investigación, por lo que las normas menos apropiadas deben ser examinadas de forma crítica, y (3) el hecho de que las citas estén influidas por muchos factores además de la calidad científica debe ser tomado en consideración en el análisis estadístico del recuento de citas

Por otro lado, es necesario señalar que el análisis de citas llevado a cabo en este trabajo es un análisis sincrónico, por lo tanto representa la situación de los trabajos citados en un momento concreto del tiempo, sin contemplar el análisis de las citas recibidas en diferentes momentos, puesto que se ha trabajado con el acumulado de citas recibidas por los trabajos en el momento de la descarga de los datos (mayo de 2010).

5.1.2.2.1. Media de citas por documento Se trata de un indicador básico del análisis de citas. Consiste en obtener la media de citas que recibe cada documento para contar con una medida de centralización que permita interpretar los datos que se obtendrán de los indicadores aplicados en el análisis de citas.

5.1.2.2.2. Recuentos y porcentajes por temática y país En este estudio, va a utilizar el análisis de citas y el recuento de las mismas sobre los documentos agrupados por temáticas y por países, y de este modo, determinar cuáles son los autores, instituciones y países que contribuyen de manera más significativa a la investigación en Enfermedades Raras.

5.1.2.2.3. Porcentaje de documentos no citados Igualmente, para los niveles de agregación de temática y país, va a calcularse este indicador, que consiste en obtener el porcentaje de documentos que no reciben ninguna cita, y que, junto a los indicadores anteriores, permite determinar el patrón de citación que siguen los investigadores sobre el grupo de enfermedades estudiadas.

5.1.2.3. Ranking de revistas.

BAR-ILAN (2008a) dedica una parte de su trabajo a la calidad de las revistas y menciona que:

Aunque hay una gran controversia sobre el uso de factores de impacto de revistas, parece existir una genuina necesidad de rankings de revistas. Este problema es especialmente acusado para áreas relativamente jóvenes y multidisciplinares.

Además realiza un análisis de los trabajos bibliométricos que han comparado la evaluación cualitativa de revistas con la posición de éstas en función de su Factor de Impacto, y en función de ellos, puede establecerse que, generalmente, existe correlación entre la evaluación por expertos y el posicionamiento de las revistas mediante el análisis de citas aunque debe ser utilizado con cautela.

Posición de las revistas en las que se publica: Anteriormente se mencionó que la medida que va a utilizarse para determinar la posición de las revistas dentro del área estudiada será el Factor de Impacto en su formulación original, junto con los cuartiles corregidos, pues permite establecer una lista de las revistas más utilizadas por los investigadores.

Al contrario que en el caso del Factor de Impacto, para el cálculo del cuartil al que pertenece cada revista no va a utilizarse la formulación original, sino que en lugar de establecer los cuatro cuartiles en función del número de revistas se hará en función del número de artículos publicados. Este tipo de asignación es mucho más restrictiva y establece un primer cuartil con un menor número de revistas y los cuartiles más bajos poseen un mayor número de revistas.

A modo de ejemplo, la categoría “GENETICS & HEREDITY” del Journal Citation Report contiene en 2009, 144 revistas. Con la asignación tradicional hecha por Thomson Reuters, cada cuartil estaría compuesto por 36 revistas. Sin embargo, aplicando esta nueva asignación, más restrictiva, los cuartiles no poseen el mismo número de revistas, sino el mismo número de artículos, de forma que, estando la categoría “GENETICS & HEREDITARY” compuesta por 144 revistas que producen en 2009 15.990 artículos, cada cuartil estaría compuesto por las revistas que acumulen 3.997,5 artículos ordenadas por Factor de Impacto de mayor a menor, se obtendría la distribución de revistas por cuartil que se muestra en la tabla 5.5.

Cuartil	Nº de revistas
1Q	21
2Q	28
3Q	39
4Q	56

Tabla 5.5: Ejemplo de distribución por cuartiles corregidos (*GENETICS & HEREDITARY*, 2009)

5.1.3. Indicadores conexionistas

Estos indicadores están basados en la metodología de *Análisis de Redes Sociales* o *Social Network Analysis* (SNA), una técnica de representación gráfica de relaciones, entre elementos o individuos, basada en la teoría matemática de grafos. Apoyándonos en el trabajo de MOYA-ANEGÓN et al. (2006) en el que se refiere a los procesos autoorganizativos como un ejemplo de método de análisis basado en la conexión entre elementos, parece adecuado emplear el término de *indicadores conexionistas* para referirse también a aquellos en los que la técnica empleada es el SNA.

Los resultados presentes en este trabajo hacen uso del SNA como principal técnica para representar las conexiones entre trabajos en diferentes niveles de agregación. Sin embargo esta técnica tiene muchas variantes en función de cuáles sean los algoritmos utilizados para su representación así como las medidas asociadas a ellas. En este sentido, LEYDESDORFF y VAUGHAN (2006) consideran que el coeficiente r de Pearson no es apropiado para matrices simétricas tales como las empleadas en el análisis de cocitaciones, aunque si lo son para matrices asimétricas como las de citación. Por ese motivo, proponen el uso de la transformación logarítmica de los datos. Por otro lado, para la representación gráfica de las redes se ha optado por utilizar el método propuesto por FRUCHTERMAN y REINGOLD (1991), que tiene como objetivo dibujar sobre un plano los nodos y sus relaciones, en un grafo que represente las atracciones y repulsiones entre ellos, siguiendo para ello el modelo de física de partículas para calcular la distancia óptima entre los diferentes vértices.

En este caso, dicha técnica ha sido empleada para obtener representaciones gráficas de las relaciones de colaboración y cocitación entre autores, instituciones y países, pero también para obtener imágenes representativas de las relaciones entre los términos descriptivos de los trabajos, a través de la metodología desarrollada para asignar a cada uno de los trabajos los términos del tesoro MeSH más adecuados para la temática que tratan cada uno de ellos.

5.1.3.1. Colaboración

Por sus características, el indicador al que más fácilmente se puede aplicar el análisis de redes sociales, es la colaboración, en diferentes niveles de agregación, principalmente autores, países e instituciones. Para construir las matrices iniciales obtenemos, de cada documento, los elementos (autores, países o instituciones) dependiendo del nivel de agregación, elaborando una matriz simétrica de doble entrada, con un umbral mínimo de trabajos requeridos para que el elemento se tenga en cuenta.

De esta forma, se dispone de una matriz que incluye el número de trabajos en común que tienen entre si los elementos seleccionados. Como siguiente paso en el procedimiento, se calculan las distancias entre los vértices del grafo en función del modelo planteado por FRUCHTERMAN y REINGOLD (1991), de forma que se obtienen las coordenadas de los elementos en un plano bidimensional, en el que se representan los diferentes nodos, junto con sus relaciones y la fuerza de las mismas, representadas tanto por el grosor de la línea que los une, como por la distancia que los separa. Del mismo modo, el tamaño de los vértices corresponde al valor del “degree”, una medida del SNA que equivale el número de nodos que apuntan a un nodo concreto, de forma que, cuanto mayor sea el número de relaciones de un nodo con los demás, mayor será su degree y por tanto, mayor será su tamaño en la visualización.

5.1.3.2. Cocitación

Se entiende por cocitación el hecho de que dos o más trabajos sean citados de forma simultánea por un tercer documento. Esta medida enunciada por SMALL (1973) y MARSHAKOVA (1973) de forma casi simultánea, parte de la hipótesis de que cuanto mayor sea el número de veces que dos trabajos son citados de forma continua, mayor es la posibilidad de que su contenido esté relacionado. Esta medida puede calcularse también para autores, revistas, instituciones, etc.

Al igual que en los indicadores de impacto, el cálculo de estos indicadores se lleva a cabo mediante matrices en las que se recoge el número de citas que recibe un trabajo, autor, revista, etc. junto con otros de la población estudiada. En este caso va a utilizarse el análisis de redes sociales para representar estas relaciones. Aplicando el mismo principio de los trabajos cocitados a los autores, es de suponer que cuanto mayor sea el índice de cocitación de dos autores mayor será la posibilidad de que su investigación se encuentre relacionada y de que, por tanto, se encuentren trabajando en el mismo área o en áreas afines

Por otro lado, el análisis de cocitaciones debe tener en cuenta que no siempre son citados de forma conjunta los autores que trabajan en las mismas especialidades, entre otras causas, debido a las diferentes escuelas científicas y de pensamiento que pueden existir en las disciplinas estudiadas. En este sentido, NADEL (1983) señala como ejemplo las teorías de Thomas S. Khun y Karl Popper acerca de la inconmensurabilidad de las teorías científicas. Así, Khun defiende que los científicos tienden a subestimar e ignorar las teorías científicas “rivales”, con las que no están de acuerdo, mientras que Popper defiende que las teorías científicas deben ser juzgadas juntas y enfrentadas contra los hechos empíricos.

De este modo, NADEL (1983) indica que si Khun está en lo cierto, las teorías rivales nunca serán cocitadas, mientras que si es Popper el que tiene razón, sí que lo serán. Para determinarlo, NADEL (1983) analizó la literatura científica relativa a la unificación de la super-conductividad y el electromagnetismo débil, encontrando evidencias que apoyaban ambas teorías. Así, las teorías rivales son citadas juntas a menudo durante la primera fase de su desarrollo, cuando son introducidas en el acervo científico y apenas existe cocitación entre ellas cuando, en una segunda fase, ya se encuentran en posiciones firmemente establecidas.

Además, a la hora de determinar la correlación entre autores mediante matrices de cocitaciones, es necesario tener en cuenta que:

En los últimos años ha habido una importante discusión sobre la medida de similaridad más apropiada para el análisis de cocitaciones, hay colegas a favor de la r de Pearson y también quienes critican su uso (BAR-ILAN, 2008a).

En este caso, al igual que con el análisis de redes sociales de colaboración, va a utilizarse el modelo de FRUCHTERMAN y REINGOLD (1991), que va a permitir representar la cocitación de autores en un gráfico bidimensional en el que pueda observarse tanto la distribución general, como el número de citaciones en común entre los diferentes nodos, a través del degree de cada uno de ellos y, por medio de la distancia que los separa y del grosor de las líneas que los unen, el nivel de relación entre nodos concretos.

5.1.3.3. Temáticas

El análisis de redes sociales en el caso de las temáticas va a ser elaborado de forma muy similar al análisis de colaboraciones y cocitaciones, con la diferencia de que las temáticas han sido asignados a los documentos a través de los códigos ICD-10 asignados a las enfermedades de las que trata cada documento. De este modo y al igual que en los casos anteriores, se construye una matriz de doble entrada simétrica, con los diferentes elementos (en este caso códigos ICD-10) que serán los nodos de la red social. Finalmente se calcula la distancia entre ellos a través del método descrito y se obtiene un gráfico en dos dimensiones con la representación de los términos y las relaciones entre ellos.

5.1.4. Indicadores multidimensionales

Los indicadores bibliométricos multidimensionales beben directamente de las técnicas matemáticas de análisis multivariante y multivariable, por lo que para empezar a desarrollar los indicadores primero deben definirse las técnicas empleadas en ellos. El análisis multivariante “En un sentido amplio, se refiere a todos los métodos estadísticos que analizan simultáneamente medidas múltiples de cada individuo u objeto sometido a investigación” (HAIR et al., 1999).

Estas variables de estudio pueden ser cuantitativas, cualitativas o un conjunto de ambas y, en función de sus características será más recomendable el uso de una técnica u otra. En cualquier caso, el uso de estas técnicas implica complejos procesos matemáticos que solo pueden abordarse mediante el empleo del cálculo matricial. Esta complejidad ha hecho que su uso solo fuera posible a partir del desarrollo de la informática y de los procesos de cálculo. Hoy en día, estos procesos están al alcance de cualquier grupo de investigación, lo cual ha llevado a su generalización a través del desarrollo software específico que permitiera su aplicación. Lo más importante de este desarrollo es que permite a investigadores, que no tienen porque ser expertos en matemáticas ni en el desarrollo de aplicaciones

informáticas, realizar los cálculos y las representaciones necesarias para emplear estas técnicas (CARRASCO y HERNÁN, 1993). Buena prueba de esta aplicación son los trabajos de GARCÍA-ZORITA (2000) y GARCÍA GONZALEZ (2010). El primero desarrollando un método para la aplicación de herramientas bibliográficas a la obtención de indicadores bibliométricos de forma semi-automática, mientras que el segundo desarrolló un método para la implementación de los registros bibliográficos completos en bases de datos ad hoc, para la posterior obtención de indicadores bibliométricos de cualquier tipo, ya sean univariantes o multivariantes, desarrollando de forma más completa el trabajo iniciado por GARCÍA-ZORITA (2000).

Entre las técnicas de análisis multivariante o multidimensional que van a emplearse se encuentran el escalado multidimensional (no-métrico), el análisis de clusters y el análisis de correspondencias, que se describen a continuación.

5.1.4.1. Escalados multidimensionales (MDS)

El escalado multidimensional o MDS (Multidimensional Scaling) es una técnica multivariante que consiste en situar sobre un plano bidimensional a los individuos (entendiendo por individuos los objetos de estudio: países, autores, instituciones, etc.) atendiendo a las distancias entre ellos, estando estas distancias calculadas en base a una variable común a todos los individuos, como puede ser el volumen de trabajos en colaboración, de forma que exista una mayor distancia entre aquellos individuos con un menor número de trabajos en colaboración entre ellos y se encontrarán más cercanos entre sí aquellos individuos que colaboren más estrechamente.

Por lo tanto este análisis multidimensional tiene como datos fuente una matriz de doble entrada que contiene las distancias o similitudes entre los diferentes individuos del estudio. Esta medida de distancia entre los individuos suele ser habitualmente la distancia euclídea o la de Pearson, que es una modificación de

la anterior, aunque pueden utilizarse otras medidas de distancia como el índice de Jaccard o el de Salton (GARCÍA-ZORITA, 2000).

Un ejemplo de matriz para el escalado multidimensional sería la siguiente:

	AUT	BEL	DEU	DNK	ESP	FIN	FRA	GBR	GRC	IRL
AUT	0	22	150	4	20	4	32	69	6	4
BEL	22	0	139	27	45	9	153	136	5	14
DEU	150	139	0	99	101	35	251	440	24	20
DNK	4	27	99	0	15	10	39	82	4	10
ESP	20	45	101	15	0	13	88	110	8	11
FIN	4	9	35	10	13	0	30	52	4	5
FRA	32	153	251	39	88	30	0	297	12	27
GBR	69	136	440	82	110	52	297	0	31	92
GRC	6	5	24	4	8	4	12	31	0	0
IRL	4	14	20	10	11	5	27	92	0	0

Tabla 5.6: Ejemplo de matriz de doble entrada para MDS

Cuando se calculan las distancias euclídeas entre los elementos de dicha matriz se obtiene la tabla 5.7

	AUT	BEL	DEU	DNK	ESP	FIN	FRA	GBR	GRC	IRL
AUT	0.00	212.68	616.33	83.52	114.57	132.80	356.98	552.92	136.81	136.52
BEL	212.68	0.00	456.00	206.79	162.60	243.17	316.36	467.83	271.56	249.70
DEU	616.33	456.00	0.00	589.59	525.52	610.22	437.38	647.39	642.13	599.27
DNK	83.52	206.79	589.59	0.00	114.79	75.44	368.55	577.20	115.38	104.13
ESP	114.57	162.60	525.52	114.79	0.00	149.45	324.57	524.84	166.46	145.46
FIN	132.80	243.17	610.22	75.44	149.45	0.00	423.43	623.47	66.45	74.99
FRA	356.98	316.36	437.38	368.55	324.57	423.43	0.00	488.21	448.23	417.41
GBR	552.92	467.83	647.39	577.20	524.84	623.47	488.21	0.00	649.10	652.23
GRC	136.81	271.56	642.13	115.38	166.46	66.45	448.23	649.10	0.00	64.70
IRL	136.52	249.70	599.27	104.13	145.46	74.99	417.41	652.23	64.70	0.00

Tabla 5.7: Ejemplo de matriz de distancias para MDS

Para llevar a cabo este tipo de análisis, como el de clusters, es requisito indispensable que la matriz de datos sea simétrica, es decir, que cada objeto representado en una dimensión de la matriz tenga su par en la otra, siendo la diagonal siempre cero.

La técnica más empleada en el área de la Bibliometría para calcular los MDS es la del Escalado Multidimensional Clásico No-Métrico y los mayores exponentes en su creación son SHEPARD (1962b,a) y KRUSKAL (1964). El primero por definir en sus trabajos para “Psykometrika” el cálculo de este tipo de análisis, y el segundo por perfeccionar el algoritmo e introducir el concepto de “Stress”, que indica la

bondad de ajuste del MDS sobre el conjunto real de datos, de forma que a menor Stress mejor ajustado estará el análisis.

Ambos métodos pretenden a partir de una matriz de distancias - disimilaridades o similaridades-, no necesariamente euclídeas, encontrar una configuración de coordenadas $x_{i,j}$ de modo que las distancias $d_{i,j}$ entre todos los pares de coordenadas calculados estén en relación monótona con las disimilaridades iniciales (GARCÍA-ZORITA, 2000).

KRUSKAL (1964) determinó diferentes umbrales para facilitar la interpretación de este valor, que por encima del 20% (o 0,2) indica un muy mal ajuste y va descendiendo hasta el 0% que indicaría un ajuste perfecto.

5.1.4.2. Análisis de clusters

El análisis de conglomerados (clusters) tiene por objeto agrupar elementos en grupos homogéneos en función de las similitudes o similitudes entre ellos. Normalmente se agrupan las observaciones, pero el análisis de conglomerados puede también aplicarse para agrupar variables (PEÑA, 2002).

Por tanto se trata de una técnica de clasificación de datos, al igual que el Análisis Discriminante, con la salvedad de que en el Análisis Discriminante es necesario conocer los grupos a priori, mientras que el Análisis de Clusters permite determinarlos a posteriori en función de la distribución de los propios datos, de forma que los individuos que pertenecen a un mismo cluster tienen más elementos en común, y por tanto son más similares entre si, que los individuos de otros clusters (CARRASCO y HERNÁN, 1993; HAIR et al., 1999).

En este caso el Análisis de Clusters va a utilizarse con el fin de encontrar agrupaciones basadas en la producción conjunta de los diferentes objetos de estudio (autores, instituciones, países, temáticas), de forma que haga posible localizar las estructuras subyacentes, tales como grupos de investigación, instituciones en

estrecha colaboración o grupos de términos que definen áreas concretas o que son significativamente diferentes del resto.

Los conglomerados pueden formarse mediante la aplicación de diversos métodos. En este caso se ha optado por utilizar el que se conoce como Análisis de Cluster Jerárquico, que se basa en una matriz de distancias como las que se utilizan para el MDS (tablas 5.6 y 5.7), con una ligera variación, introducida por WARD JR (1963) que define una medida global de la heterogeneidad del conglomerado de individuos agrupados. Esta medida es la W , que se determina como la suma de las distancias euclídeas al cuadrado entre cada elemento del grupo y la media del grupo (PEÑA, 2002):

$$W = \sum_g \sum_{i \in g} (x_{ig} - \bar{x}_g)' (x_{ig} - \bar{x}_g) \quad (5.16)$$

La representación gráfica de este análisis sirve para identificar la estructura subyacente en el conjunto de datos y a determinar los grupos en los que puede dividirse. La forma más habitual de representar el Análisis de Clusters es el dendograma. Se trata de una representación jerárquica en la que cada elemento de análisis se encuentra situado en el eje de las abscisas y parten de él las líneas que permiten agrupar los diferentes individuos en función de la distancia que los separa (en el eje de las ordenadas).

Un ejemplo de este tipo de representación se encuentra en PEÑA (2002):

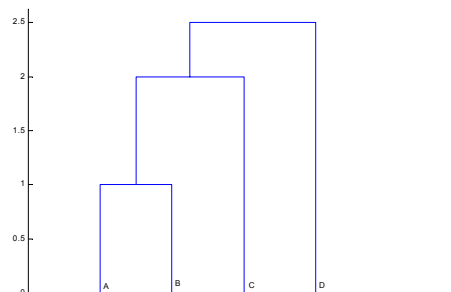


Figura 5.1: Ejemplo de dendograma

5.1.4.3. Análisis de correspondencias

El Análisis de Correspondencias (AC), tal y como lo define GREENACRE (2008), “es una generalización de una representación gráfica con la que todos estamos familiarizados, el diagrama de dispersión. Un diagrama de dispersión representa los datos en forma de puntos con relación a dos ejes de coordenadas perpendiculares: el eje horizontal, eje de las x , y el eje vertical, eje de las y . [...] nos permite visualizar tablas de datos en subespacios de baja dimensionalidad que explican de forma óptima la inercia.” Se trata de un método de análisis multivariante que permite poner en relación unas variables cualitativas (ordinales o categóricas) a través de un número menor de variables ficticias (generalmente dos o tres). Dado que se trata de variables cualitativas, el punto de partida para el AC será siempre una tabla de contingencia, en la que se presenten las frecuencias absolutas entre las distintas categorías de las variables cualitativas estudiadas (GARCÍA-ZORITA, 2000).

El primer paso para aplicar el AC sobre una tabla de contingencia consiste en determinar si las diferencias observadas en la distribución de frecuencias se deben al azar o si son producto de una estructura en los datos. Para ello, previamente al cálculo del AC se debe realizar siempre un test de significación, en este caso, debido a la naturaleza cualitativa de los datos, un test de χ^2 , que indique si existe diferencia significativa en la distribución de frecuencias.

Una vez realizado el test de significación y comprobado que efectivamente las diferencias encontradas no se deben al azar es necesario obtener de la tabla de contingencia original dos valores fundamentales para el AC, las inercias y las distancias.

Las inercias se calculan como la suma de los cuadrados de los valores singulares, mientras que las distancias son el resultado de aplicar el χ^2 entre los perfiles de las filas y los de las columnas (GREENACRE, 2008).

“Como en el resto de análisis factoriales, los factores son extraídos en función de su orden decreciente de inercia, de modo que el primero es el que explica una mayor proporción de la varianza del sistema” (GARCÍA-ZORITA, 2000). Teniendo en cuenta este dato, es necesario ser muy cuidadoso a la hora de determinar que factores o dimensiones serán las que se utilicen para la representación gráfica del análisis, pues puede variar significativamente en función de las que se escojan. Sin embargo, en la práctica, la elección de factores suele ser sencilla, debido a que si el ACS ha extraído más de dos factores, estos suelen ordenarse por porcentaje de la variabilidad explicada, así que habitualmente es suficiente con utilizar los dos (o tres en el caso de una representación tridimensional) factores extraídos en primer lugar.

Una vez determinada la matriz de distancias y las inercias se procede a plasmar en un plano, generalmente bidimensional (aunque pueden hacerse representaciones en tres dimensiones), los puntos correspondientes a las diferentes categorías de cada variable. En este caso un ejemplo de AC sería el que puede calcularse entre autores y temáticas, que una vez plasmados gráficamente indicarán, mediante la distancia que separa a los autores de cada temática, la preferencia de publicación de cada uno de ellos, de forma que cuanto menor sea la distancia que separa a un autor de una temática, mayor será la preferencia de dicho autor por ella.

5.2. Análisis estadístico y uso del paquete estadístico R

Los métodos estadísticos suponen una parte fundamental en el desarrollo del método científico, siendo indispensables para calcular los indicadores bibliométricos propuestos. Estos métodos estadísticos se dividen en dos grupos.

5.2.1. Estadística descriptiva

Una definición bastante completa de estadística descriptiva es la que proponen EGGHE y ROUSSEAU (1990):

El término “estadística descriptiva” se refiere al conjunto de métodos, procedimientos y técnicas usados para representar, resumir o comunicar de otra forma las características esenciales de un conjunto de datos brutos [...] La aplicación de las técnicas de estadística descriptiva permite realizar inferencias estadísticas.

Los métodos y técnicas tradicionales de la estadística descriptiva son las tablas o distribuciones de frecuencias, las matrices de doble entrada, las diferentes representaciones gráficas (diagramas, histogramas, etc.), las medidas tanto de centralización como de dispersión (media, moda, mediana, desviación típica, etc.) así como diferentes coeficientes que ayudan a resumir y aclarar diversos aspectos de los datos estudiados.

La estadística descriptiva, como su propio nombre indica, permite describir, esto es, extraer las características principales, de una o varias variables, cuantitativas o cualitativas, así como transformar dichas variables de forma que puedan convertirse estas características principales en coeficientes y representaciones que aporten una información mucho más clara, completa y accesible que la presentación de los datos en bruto.

5.2.2. Estadística inferencial

Por otro lado la inferencia estadística se ocupa de:

inferir las características de una población que no es completamente observable analizando una parte de ella, llamada muestra (PEÑA y ROMO, 1999)

Este método está estrechamente relacionado con la probabilidad, puesto que las hipótesis estadísticas nunca son completamente ciertas o completamente falsas,

sino que la inferencia estadística indica la probabilidad de que dichas hipótesis sean ciertas o falsas (EGGHE y ROUSSEAU, 1990)

La estadística inferencial proporciona las herramientas para determinar la muestra más adecuada, tanto en alcance como en tamaño, sobre una población concreta, y por otro lado se utiliza para comprobar las hipótesis propuestas. Tal y como explica MARTÍN-MORENO (1999):

Su empleo permite realizar estimaciones, esto es inferir características propias de la población a partir de las muestras; establecer comparaciones entre éstas, a través de pruebas de homogeneidad y heterogeneidad, y relacionar diferentes variables mediante el establecimiento de leyes que describen el comportamiento de una variable cuando se producen modificaciones en otra.

Esto se consigue aplicando métodos como la t de student, el χ^2 de Pearson o la prueba de Kolmogorov-Smirnov con el objetivo de medir de forma cuantitativa si existe una discrepancia significativa entre los datos observados (a través de la muestra) y lo previsto en la hipótesis a contrastar.

Generalmente las técnicas relacionadas con la estadística inferencial se utilizan para inferir características de una población determinada a través de una muestra obtenida mediante un muestreo aleatorio simple, proporcional a la población estudiada. Sin embargo, en bibliometría, en muchas ocasiones y gracias a las técnicas de tratamiento masivo de datos, se trabaja con poblaciones completas o muy cercanas a estas.

Concretamente en este trabajo se seleccionaron 252 enfermedades raras que poseen una causa genética identificada sobre el total de enfermedades raras tipificadas, que se encuentra entre las 6.000 y las 7.000. Estudios previos han demostrado que existe una estrecha relación entre las enfermedades raras y la genética, pues se sospecha que la mayor parte de este tipo de dolencias poseen causa genética, a pesar de que sea un conjunto pequeño, pero creciente, en el que esta causa ha podido ser identificada (FISCHER et al., 2005; STOLK, 2006).

Precisamente debido a esta característica concerniente a las enfermedades raras, se considera que la muestra escogida es representativa de la población total de enfermedades raras y por tanto, sus resultados son inferibles al resto de la población, salvo en el caso de las temáticas, pues el hecho de que las enfermedades de la muestra posean causa genética identifica hace que su perfil temático esté mucho más cerca de la genética de lo que cabría esperar en una muestra aleatoria.

5.2.3. Regresión lineal

En este caso los estudios de regresión van a basarse fundamentalmente en el uso simultáneo de dos tipos de variables. Por un lado aquellas que se refieren a los datos obtenidos de las bases de datos bibliográficas y por otro aquellas variables que se derivan de los datos estadísticos sobre gasto en salud pública, tal y como se ha hecho en estudios similares (LASCURAIN-SÁNCHEZ et al., 2008). Estos datos se han obtenido de los servicios estadísticos de la Unión Europea (Eurostat) y de la Organización Mundial de la Salud (WHOSIS).

Asimismo, puesto que ORPHANET (2011b) cuenta con un registro de asociaciones de pacientes por países, es posible utilizar dichos datos para comparar la investigación que se realiza en enfermedades raras con el número de asociaciones de pacientes en cada país, determinando que países cuentan con un mayor número de asociaciones, y si esto se corresponde con una investigación más numerosa por parte de los mismo. De esta forma es posible averiguar en que grado intervienen las asociaciones de enfermos en el número de investigaciones que se realizan.

El método empleado para analizar el grado de relación entre las dos variables es la regresión lineal, de forma que en aquellas variables relacionadas entre si de forma lineal los puntos de la dispersión tienden a agruparse en torno a un recta (PEÑA, 2001).

El diagrama de dispersión, en el que se representan la nube de puntos y la recta de regresión, indica la fuerza de la relación entre las variables. El grado de correlación

también puede determinarse por medio del coeficiente de determinación R^2 , un estadístico que, tomando un valor entre -1 y 1, indica si existe relación entre las variables y la fuerza de la misma, de forma que cuanto más cercano a la unidad (positiva o negativa) sea el valor del estadístico, mayor será la fuerza de la relación.

Una vez determinadas las variables con las que se va a trabajar y construido el modelo, generalmente un modelo de regresión lineal simple o múltiple en función del número de variables a estudiar, se debe utilizar un método estadístico que permita estimar los valores del modelo. Habitualmente el método a emplear es el de los mínimos cuadrados ordinarios (MCO). Sin embargo, en ocasiones es necesario emplear otros métodos, debido fundamentalmente a que el modelo no cumple alguna de las hipótesis básicas (linealidad, normalidad, independencia y homocedasticidad).

En estos casos es necesario recurrir a otros métodos, concretamente cuando no se cumple la hipótesis de independencia porque se trate de variables autocorrelacionadas, como es el caso de la producción científica de cada país por año, se puede utilizar el método de los mínimos cuadrados generales o ponderados (MCG o MCP) (CORDY y GRIFFITH, 1993).

Para comprobar si los estimadores son correctos es necesario comprobar si se cumplen las hipótesis mencionadas mediante la diagnosis del modelo. Para llevar a cabo dicha diagnosis puede emplearse el método elaborado por PEÑA y SLATE (2006), que realiza de forma automática las pruebas pertinentes para las cuatro hipótesis básicas. Estas hipótesis pueden comprobarse también con otros métodos estadísticos y gráficos que se describen a continuación.

Hipótesis de linealidad: Para comprobarlo se tiene en cuenta fundamentalmente la nube de puntos de las variables, si dicha nube de puntos sigue una distribución lineal quiere decir que se cumple la hipótesis.

Hipótesis de normalidad: Esta hipótesis se cumple cuando las variables siguen una distribución normal. Se puede verificar tanto de forma gráfica (a

través del histograma de las variables y de los residuos), como a través de contrastes como el de Kolmogorov-Smirnov.

Hipótesis de independencia: Se basa en el supuesto de que ambas son variables aleatorias independientes, es decir que el valor anterior no afecta al siguiente. Pueden utilizarse múltiples métodos gráficos para comprobar la independencia, por ejemplo la nube de puntos de los residuos del modelo frente al índice temporal. También pueden emplearse contrastes de hipótesis como el de Durbin-Watson. En el caso de que se rechace la hipótesis de independencia, lo que indica es que existe autocorrelación en el modelo por diferentes motivos, por ejemplo porque se trata de una serie temporal, en cuyo caso puede tratar de ajustarse el modelo utilizando el método de los mínimos cuadrados generales, tal y como se indicó anteriormente.

Hipótesis de homocedasticidad: También llamada *Hipótesis de la igualdad de varianzas*. Esta hipótesis indica que los residuos del modelo son matemáticamente despreciables y que tienen todos ellos la misma varianza, por lo que no afectan a los estimadores del modelo planteado. La forma de comprobarla se basa fundamentalmente en el gráfico de residuos frente a valores previstos, que debe presentar un aspecto de nube de puntos sin ningún tipo de estructura, ante lo cual es importante descartar cualquier tipo de efecto de tipo “embudo”.

Una vez ajustado el modelo de regresión, el siguiente paso natural es interpretarlo, determinando si realmente existe relación entre las variables, cuáles de ellas son significativas (a través del p-valor) y en que grado se produce dicha relación.

5.2.4. El paquete estadístico R

Para llevar a cabo el análisis estadístico de los datos se cuentan con múltiples opciones comerciales (SPSS, Statgraphics, XLStat, etc.), sin embargo, como ha podido apreciarse durante todo el capítulo de metodología, se ha utilizado soft-

ware libre, pues cuenta con un beneficio intrínseco a su definición: Este tipo de software no se mantiene por medio de un único equipo de desarrollo, a merced de intereses comerciales, sino que permite que cualquiera tenga acceso al código fuente con el fin de modificarlo, mejorarlo y redistribuirlo. Debido a esta estructura de desarrollo, los proyectos basados en el software libre cuentan con toda una comunidad de desarrollo que complementa el proyecto original, lo mejora y le añade funcionalidades.

Por este motivo se ha decidido utilizar diversas herramientas de software libre para la extracción y manipulación de datos (Perl y MySQL) y va a seguirse la misma filosofía para el análisis estadístico.

El software que va a utilizarse para ello es el paquete estadístico denominado “R” (BATES et al., 2009).

R es un lenguaje de programación y un entorno para análisis estadístico y la realización de gráficos. Debido a su naturaleza es fácilmente adaptable a una gran variedad de tareas. Fue inicialmente escrito por Robert Gentleman y Ross Ihaka del Departamento de Estadística de la Universidad de Auckland en Nueva Zelanda. [...] R abarca una amplia gama de técnicas estadísticas que van desde los modelos lineales a las más modernas técnicas de clasificación pasando por los test clásicos y el análisis de series temporales. Proporciona una amplia gama de gráficos que además son fácilmente adaptables y extensibles. La calidad de los gráficos producidos y la posibilidad de incluir en ellos símbolos y fórmulas matemáticas, posibilitan su inclusión en publicaciones que suelen requerir gráficos de alta calidad (ARRIAZA GOMEZ et al., 2008).

Como puede observarse, R es un entorno muy potente, que cuenta con un importante equipo de desarrollo y una licencia libre que permite incluir multitud de modificaciones, añadidos y paquetes adicionales, convirtiéndolo en una de las herramientas más potentes para el análisis estadístico.

Dos de estos paquetes adicionales son DBI (JAMES, 2007) y RMySQL (JAMES et al., 2009), y sirven para establecer la conexión entre R y las bases de datos presentes en MySQL, permitiendo la posibilidad de obtener las matrices directamente de las bases de datos. Por ejemplo, si se introducen las siguientes instrucciones

en R, se establece conexión con la base de datos como si de un interfaz web se tratase:

```
> library(RMySQL)
> con = dbConnect(dbDriver("MySQL"),
  dbname = "base-de-datos", user="usuario",
  password= "contraseña", host="servidor")
```

A continuación puede realizarse cualquier consulta sobre ella. Siguiendo con el ejemplo del punto 4.4.2.2, ejecutando la siguiente instrucción:

```
x<-dbGetQuery(conn,
  "SELECT fec, COUNT(DISTINCT uts) as total
  FROM `isi_todo_UE`
  WHERE CIE LIKE 'G71'
  GROUP BY fec")
```

se obtienen los datos de la tabla 4.2 (página 73), y si a continuación se introduce la siguiente instrucción:

```
> pie(test$total, test$fec)
```

se genera un sencillo gráfico de tarta como el de la figura 5.2

Como se puede observar resulta sumamente sencillo generar gráficos en R a partir de consultas SQL sobre las bases de datos en MySQL. Entre los gráficos que puede manejar R se encuentran los gráficos clásicos que incluye cualquier software de hojas de cálculo (gráficos de barras, de tarta, de dispersión, etc.) así como otros propios del software estadístico (diagramas de caja o de bigotes, histogramas, etc.) y, por último, también permite generar gráficos propios del análisis multidimensional como el escalado multidimensional, análisis de clusters y de correspondencias. Del mismo modo, también pueden obtenerse fácilmente medidas básicas (de centralización, de dispersión, etc.) sobre las variables con las que se trabaja y someterlas a diferentes pruebas de contraste de hipótesis, con el fin de determinar si existen diferencias significativas entre ellas o entre las variables

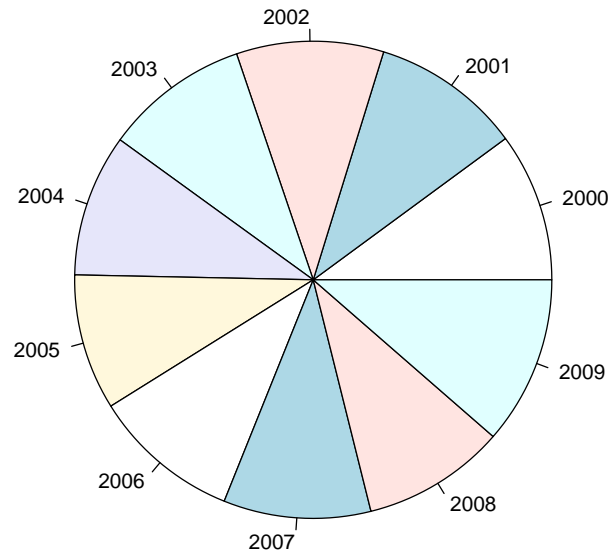


Figura 5.2: Trabajos publicados por año (*Trastornos primarios de los músculos*)

escogidas y las poblaciones estudiadas. Los paquetes estadísticos que serán utilizados para generar los resultados del siguiente capítulo están recogidos en la tabla 5.8.

Paquete	Creadores	Finalidad
RMySQL	JAMES et al. (2009)	Conexión y extracción de registros de la base de datos en MySQL
car	WEISBERG y FOX (2010)	Cálculo de regresiones lineales
ca	GREENACRE (2010)	Análisis de correspondencias
vegan	OKSANEN et al. (2010)	Ordenación de variables cualitativas mediante múltiples análisis, entre ellos el Análisis de Correspondencias
MASS	VENABLES y RIPLEY (2002)	Múltiples cálculos de estadística univariante
proxy	MEYER y BUCHTA (2009)	Medidas de distancias, similitudes y disimilitudes
vcd	ZEILEIS et al. (2007)	Análisis de mosaico y de asociación entre variables cualitativas con test de hipótesis incluidos
igraph	CSARDI y NEPUSZ (2006)	Construcción y visualización de Análisis de Redes Sociales

Continúa en la página siguiente

Paquete	Creadores	Finalidad
gvlma	PEÑA y SLATE (2006)	Test de hipótesis para modelos de regresión
dynlm	ZEILEIS (2009)	Cálculo de mínimos cuadrados generalizados para modelos de regresión
cluster	MAECHLER y ROUSSEEUW (2005)	Análisis de clusters o conglomerados

Tabla 5.8: Paquetes estadísticos utilizados en R

En vista de todas las posibilidades que ofrece, se ha considerado que la elección de R como paquete estadístico para el análisis de los datos, está totalmente justificada como la mejor opción posible.

Sección 6

Estudio de solapamiento entre Pubmed y SCI

Con los datos extraídos de ambas bases de datos se han construido dos bases de datos independientes para tratar los registros descargados. Sin embargo, puesto que se ha constatado que existe cierto solapamiento entre ellas (COSTAS et al., 2008), se ha procedido a cuantificar dicho solapamiento con el fin de determinar si se continúa trabajando con dos bases de datos o si es más conveniente construir una sola base de datos que contenga los registros únicos de cada una de ellas.

6.1. Solapamiento de las publicaciones

Para llevar a cabo este contraste se realiza un conteo del número de trabajos publicados por revista y fecha de publicación en ambas bases de datos y se unen a través del ISSN con el fin de obtener un tabla con cuatro columnas: Revista (ISSN), fecha de publicación, producción SCI y producción Pubmed. Un ejemplo de la misma puede observarse en la tabla 6.1.

El test de normalidad de Shapiro-Wilk para ambas variables arroja p-valores inferiores a $2,2e-16$, por lo que se concluye que ninguna de las dos variables sigue

ISSN	Fecha	Total SCI	Total Pubmed
00029297	2000	32	24
00029297	2001	38	27
00219258	2001	42	24
00219258	2002	42	25
00219258	2003	48	33
00219258	2004	49	37
00219258	2005	35	24
00219258	2009	42	20
01418955	2000	35	38
01418955	2001	34	45
01418955	2003	30	33
01418955	2005	50	58
01418955	2006	45	21
09608966	2002	40	44
09646906	2000	44	32
09646906	2001	30	30
09646906	2002	36	28
09646906	2003	31	30
09646906	2008	34	1
15691993	2007	40	39
15691993	2008	63	56
15691993	2009	42	9

Tabla 6.1: Muestra de la distribución de trabajos por revista, fecha y base de datos

una distribución normal. Por tanto es necesario utilizar técnicas no paramétricas para determinar si hay diferencias significativas entre las distribuciones de ambas bases de datos.

Sobre la distribución completa, con 3959 filas, se ha aplicado el contraste de hipótesis de Kolmogorov-Smirnov, que es sensible a cualquier tipo de diferencia entre las dos distribuciones (media, varianza, etc.), obteniendo los siguientes valores:

$D = 0.0321$, $p\text{-value} = 0.03402$

Estos valores indican una diferencia significativa entre ambas distribuciones, por lo que se puede concluir que no provienen de la misma población.

6.2. Solapamiento de los términos MeSH

Por último, para determinar si existe solapamiento a nivel de temáticas de los documentos, se ha elaborado una matriz similar a la de las revistas, pero con los términos MeSH de ambas bases de datos, obtenidos en Pubmed a través de su propia clasificación y en el SCI mediante el método de asignación descrito en el apartado metodológico.

De esta forma, se obtiene una tabla con los términos MeSH y su frecuencia de aparición en ambas bases de datos. Lo primero que se debe destacar es que el número de términos obtenidos en el SCI es mucho mayor, debido a que la técnica para su asignación parte de las palabras clave, que son más numerosas que los términos MeSH asignados a los documentos de la base de datos Medline / Pubmed. Concretamente, en los datos obtenidos del SCI son 6.251 los términos diferentes, mientras que en Pubmed son solo 3.793. Por otro lado, debido a las diferentes forma de asignar los términos MeSH (asignación manual en Pubmed y automática a través de las palabras clave en SCI), el número de términos coincidentes en ambas bases de datos es muy pequeño, tan solo 358.

Este dato ya sirve como indicativo de las diferencias entre ambas bases de datos, pero para hacer una comprobación más exhaustiva se han sometido las parejas de términos a diferentes test. En primer lugar se ha aplicado los test de correlación de Spearman y Pearson, arrojando en ambos casos, valores de correlación pequeños (0,254 en el caso de Pearson y 0,301 en el del Spearman), indicando una pobre correlación entre ambas variables. Posteriormente se ha aplicado de nuevo el test de Kolmogorov-Smirnov, con el fin de detectar finalmente si existen diferencias significativas entre las dos distribuciones, obteniendo los siguientes resultados:

$$D = 0.0285, \text{ p-value} = 0.04361$$

De nuevo los datos indican una diferencia entre ambas distribuciones, permitiéndonos afirmar que no provienen de la misma población.

Por tanto, a la vista de los resultados obtenidos en ambos casos (publicaciones y términos MeSH), debe rechazarse la hipótesis de que las dos distribuciones provienen de la misma población y concluir que hay diferencia significativa entre ellas, por lo que se ha decidido mantener ambas bases de datos separadas y utilizarlas de forma complementaria para obtener los indicadores que pueden observarse a lo largo del capítulo de resultados.

Capítulo III

Resultados

Indicadores unidimensionales

Los indicadores unidimensionales son aquellos que se centran en el análisis de un solo aspecto de los datos, aunque generalmente se abordan desde diferentes niveles de agregación, con el fin de aportar información desde diferentes puntos de vista.

Entre los indicadores unidimensionales que se han calculado se encuentran los indicadores de actividad científica, divididos en dos secciones (producción y consumo) y los indicadores de impacto y visibilidad.

7.1. Indicadores de actividad científica (Producción)

Los indicadores relativos a la actividad científica desde el punto de vista de la producción, se centran en analizar los trabajos publicados en la temática y el periodo determinado, desde diferentes puntos de vista, como son los países, autores, revistas, etc., que producen los trabajos, así como ciertas características de los mismos: temáticas, idiomas, tipologías, etc.

7.1.1. Producción anual

El primer aspecto por el que es recomendable comenzar al abordar un trabajo bibliométrico es por la caracterización general de los datos, es decir, de que datos se dispone para obtener la batería de indicadores y cuál es su distribución. En este caso, el volumen total de datos asciende a 19.415 registros procedentes de la base de datos SCI y 20.572 de Pubmed, distribuidos por fecha de publicación tal y como muestra la tabla 7.1.

7.1.1.1. Publicaciones por año

Para calcular el número de publicaciones por año se ha obtenido el total de documentos únicos publicados anualmente, tanto en las bases de datos Pubmed como SCI, que responden a las estrategias de búsqueda planteadas. Con los datos obtenidos se han calculado el gráfico 7.1 y la tabla 7.1, que contienen respectivamente la evolución y tasa de crecimiento anual de la publicación de artículos de revista, tal y como se estableció en la metodología, tanto para Pubmed como para SCI.

El gráfico 7.1 muestra una progresión lineal ascendente en el número de publicaciones por año. Esta progresión es continua, con pequeños altibajos, salvo en el año 2005, en el que se produce un aumento notable en la producción en ambas bases de datos, recuperando posteriormente su evolución anterior.

7.1.1.2. Producción y tasa de crecimiento

La tasa de crecimiento (TC), que se puede observar en la siguiente tabla (7.1) se ha calculado obteniendo el porcentaje de variación en la producción de trabajos respecto al año anterior, ya sea esta positiva o negativa. Como se puede observar, el año con una mayor tasa de crecimiento positiva es 2005, siendo ésta del del 12,89% en Pubmed y del 11,88% en el SCI, cayendo la producción científica un 7,88% y un 3,83% respectivamente al año siguiente.

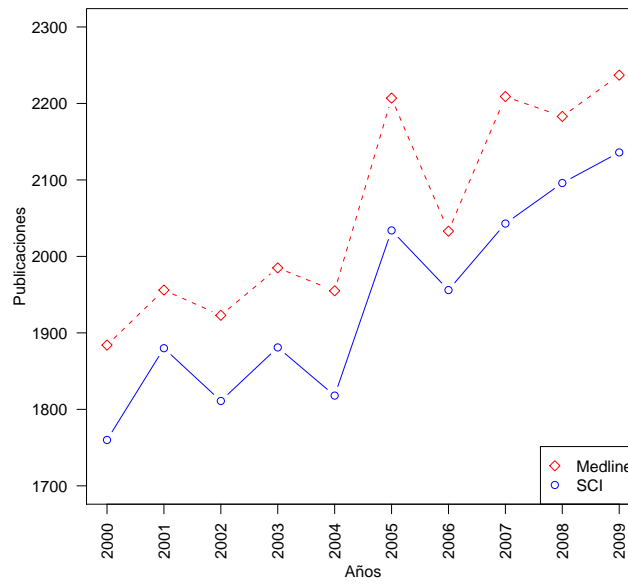


Figura 7.1: Evolución de la producción anual (SCI y Pubmed)

Año	Pubmed	SCI	TC Pubmed	TC SCI
2000	1884	1760	0	0
2001	1956	1880	3,82	6,82
2002	1923	1811	-1,69	-3,67
2003	1985	1881	3,22	3,87
2004	1955	1818	-1,51	-3,35
2005	2207	2034	12,89	11,88
2006	2033	1956	-7,88	-3,83
2007	2209	2043	8,66	4,45
2008	2183	2096	-1,18	2,59
2009	2237	2136	2,47	1,91

Tabla 7.1: Producción y tasa de crecimiento por año (Pubmed y SCI)

7.1.2. Países

Una vez obtenida la distribución general de los datos, se va a analizar su distribución en torno a los países productores de los mismos, a través del análisis de las direcciones institucionales de los autores.

7.1.2.1. Producción anual por país

En los siguientes gráficos (7.2 y 7.3) se observa la distribución de la producción por países en Pubmed y SCI respectivamente. A la hora de interpretar estos resultados hay que tener en cuenta que SCI recoge todas y cada una de las filiaciones institucionales de los autores, mientras que Pubmed solo incluye en sus registros la dirección institucional del primer autor firmante.

En ambos gráficos ocurre el mismo fenómeno, una serie de países (Gran Bretaña, Francia, Alemania e Irlanda) se sitúan entre los más productivos y, aun existiendo diferencias entre ellos, destacan sobre el resto de los países, mientras que la mayor parte de los países restantes se encuentran en la parte inferior de los gráficos, con una producción mucho menor y muy similar entre ellos. Por último, en ambos gráficos destacan las posiciones de España y Holanda, que sin llegar a situarse entre los más productivos, tienen un lugar destacado entre el resto de países.

7.1.2.1.1. Distribución por países y base de datos A continuación se muestra la distribución de la producción por países de forma conjunta en las bases de datos SCI y Pubmed. En la tabla 7.2 aparecen los valores absolutos, mientras que en el gráfico 7.4 se observan las diferencias entre dicha distribución, siendo este gráfico necesario para observar, mediante prueba estadística, las diferencias entre la producción de cada país en las diferentes bases de datos. El degradado de colores es representativo de estas discrepancias, de forma que los tonos azules indican una sobrerrepresentación y los tonos rojos una infrarepresentación para una supuesta distribución equidistante.

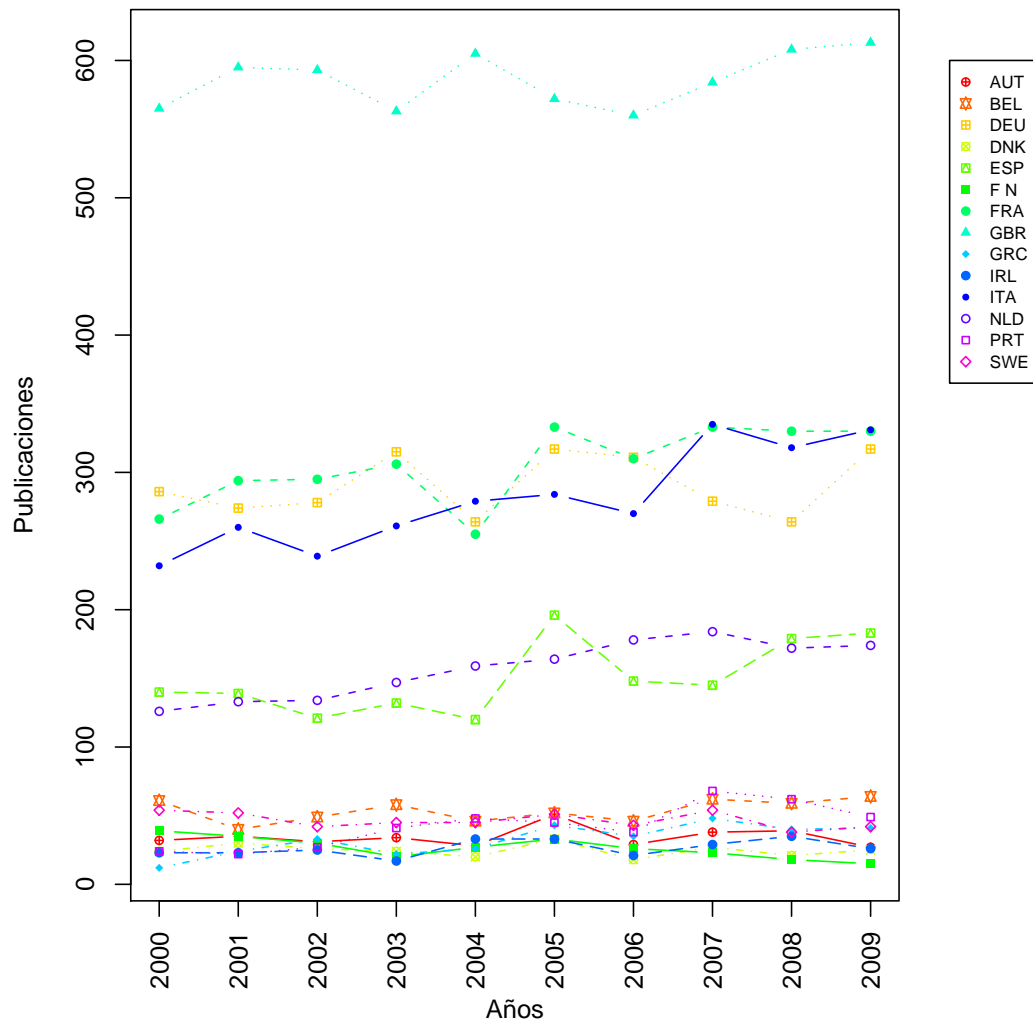


Figura 7.2: Producción anual por país Pubmed

Tal y como muestra la tabla 7.2, el número de documentos por país es mayor en Pubmed en los casos de Gran Bretaña, España, Bélgica y Portugal, mientras que en el resto de países es menor que la producción en SCI, a pesar de que la producción global es mayor en Pubmed con 20.572 documentos, frente a los 19.415 de SCI.

País	Pubmed	SCI
GBR	5858	4759
FRA	3052	3627

Continúa en la página siguiente...

País	Pubmed	SCI
DEU	2905	4230
ITA	2809	3052
NLD	1571	1957
ESP	1503	1347
BEL	537	11
SWE	466	832
PRT	424	321
AUT	344	521
GRC	324	341
FIN	266	419
IRL	265	422
DNK	248	577
LUX	0	3

Tabla 7.2: *Distribución de la producción por países en Pubmed y SCI*

7.1.3. Tipología documental

La tipología documental indica el formato, y en el caso de Pubmed, algunos aspectos formales del contenido, de los trabajos analizados. El estudio se realiza sobre los 19.415 documentos de SCI y los 20.572 de Pubmed. Con la diferencia de que la tipología documental de SCI es excluyente (un documento no puede tener más de una tipología), y la de Pubmed no lo es.

7.1.3.1. SCI

En el caso de SCI se ha trabajado únicamente con los documentos clasificados como artículos, ya sean artículos propiamente dichos, actas de congresos o capítu-

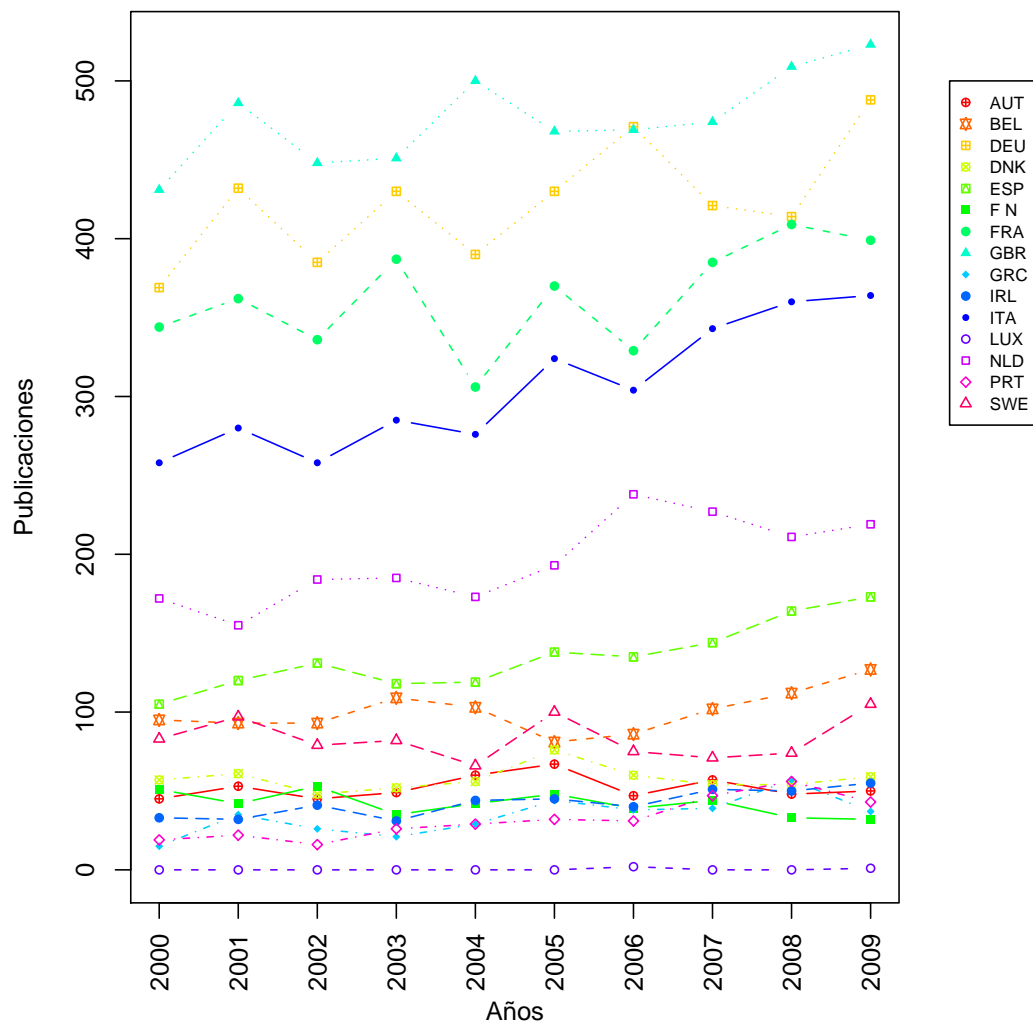


Figura 7.3: Producción anual por país SCI

los de libros publicados como artículos en revistas indizadas en el SCI, y cuya distribución se puede observar en la tabla 7.3

7.1.3.2. Pubmed

La tipología documental en Pubmed difiere notablemente de lo habitual, pues no hace referencia únicamente al formato de los trabajos (artículos de revista, actas de congreso, etc.) sino también a su contenido, de forma que ofrece información acerca del enfoque del trabajo e incluso de su metodología.

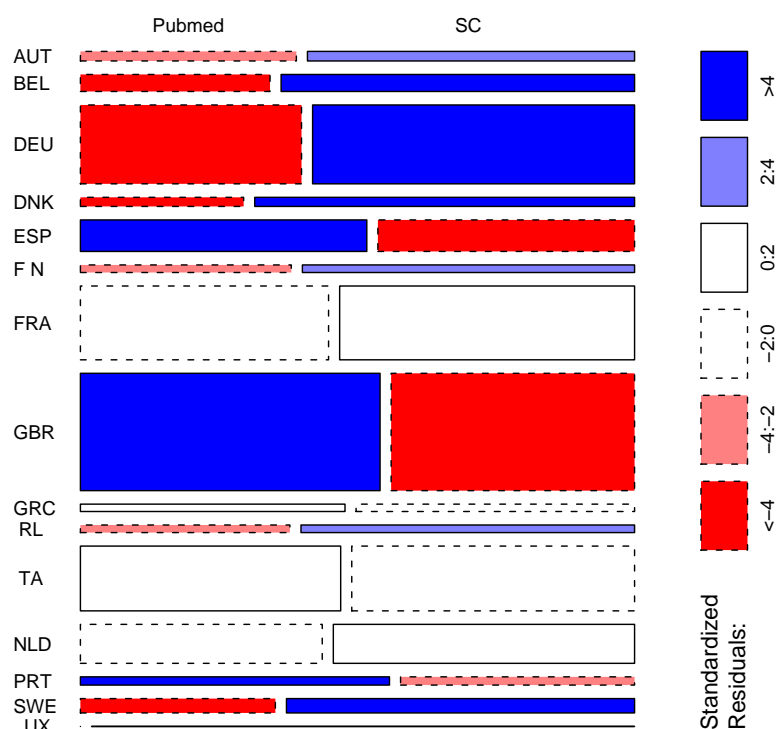


Figura 7.4: Distribución de la producción por países en Pubmed y SCI

Tipología	Documentos	Porcentaje
Article	17894	92,16 %
Article & Proceedings	1520	7,83 %
Article & Book Chapter	1	0,005 %

Tabla 7.3: Tipología documental en SCI

En la tabla 7.4 se observa como se distribuyen estas tipologías de documentos. Para elaborar este indicador se han tomado los tipos documentales con mayor producción, aquellos que tienen más de 100 documentos en todo el periodo, por considerarlos los más representativos.

Como se puede observar, tanto en SCI como en Pubmed predominan los artículos de revista, si bien, en el caso de Pubmed llama la atención el número (3808) de *case reports* presentes.

Tipología	Documentos	Porcentaje
Journal Article	20471	99,51 %
Case Reports	3808	18,51 %
Review	2688	13,07 %
Comparative Study	1725	8,38 %
English Abstract	1476	7,17 %
Clinical Trial	460	2,24 %
Multicenter Study	349	1,70 %
Evaluation Studies	255	1,24 %
Randomized Controlled Trial	249	1,21 %
In Vitro	178	0,86 %
Comment	120	0,58 %
Otros	518	2,52 %

Tabla 7.4: *Tipología documental Pubmed*

A continuación, en la tabla 7.5, se recoge una explicación sobre las características de los diferentes tipos documentales.

Tipología	Explicación
Journal article	Artículos de revista
Review	Artículo o libro publicado después de revisar el material publicado sobre una materia
Case reports	Presentaciones clínicas que seguidas de estudios de evaluación que eventualmente conducen a un diagnóstico.
Clinical trial	Estudio clínico de la seguridad, eficacia, o pauta de dosificación óptima de uno o más diagnósticos, medicamentos, profilácticos, dispositivos o técnicas en seres humanos seleccionados de acuerdo a criterios predeterminados de elegibilidad y vigilados cuidadosamente para detectar efectos favorables y desfavorables.
Comparative study	Comparación de los resultados, respuestas, etc. Para las diferentes técnicas, métodos terapéuticos u otros insumos.
English abstract	Resúmenes en inglés sobre artículos publicados en otros idiomas
Evaluation studies	Trabajo consistente en determinar la efectividad o utilidad de un proceso, personal o equipamiento
In vitro	Trabajo realizado con tejidos extirpados
Randomized control	Trabajo consistente en un ensayo clínico que incluye al menos un grupo experimental y un grupo control
Multicenter study	Trabajo consistente en un estudio controlado llevado a cabo por diferentes instituciones cooperantes

Tabla 7.5: *Detalle de los diferentes tipos documentales en la base de datos Pubmed (U.S. NATIONAL LIBRARY OF MEDICINE, 2012)*

7.1.4. Temática

Para estudiar la distribución de los trabajos por temática se han utilizado diferentes niveles de agregación. En primer lugar se ha calculado, para ambas bases de datos, la distribución de la producción científica en función de la Clasificación Internacional de Enfermedades en su décima versión, o ICD-10 (tablas 7.6 y 7.7). Tanto en Pubmed como en SCI, esta distribución muestra un claro predominio del conjunto de enfermedades clasificadas en el grupo E83-E88, que contiene entre otras la fibrosis quística. También destaca el hecho de que, en las dos bases de datos, solo cuatro grupos sobrepasan los 1000 documentos en todo el periodo (E83-E88, Q80-Q89, G71 y E75).

A continuación se ha calculado la producción por año para las categorías WOS con 500 o más documentos en todo el periodo (7.8), destacando sobre el resto las áreas de Genética, Bioquímica y biología molecular y Neurociencias y Neurología (con más de 3000 documentos en todo el periodo).

Por último se ha obtenido la distribución de la producción por términos MeSH para ambas bases de datos (Ver apéndices: tablas B.1 y B.2), recogiendo aquellos términos que aparecen en, la menos, 50 documentos. En el caso de Pubmed los registros ya vienen clasificados con estos encabezamientos de materia, mientras que para poder aplicarlos en el SCI se ha utilizado el método descrito en el capítulo metodológico.

En este caso el término que destaca en las dos bases de datos es la fibrosis quística, que es la primera enfermedad cuando se ordenan los términos por frecuencia, pues, salvo esta excepción, los términos que ocupan las primeras posiciones (sobre todo en Pubmed) hacen referencia a conceptos íntimamente relacionados con la Genética.

Clasificación ICD-10	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Total
E83-E88	665	666	714	724	694	765	652	734	753	774	7141
Q80-Q89	217	221	186	211	209	270	237	288	285	279	2403
G71	150	171	161	147	158	175	162	172	180	181	1657
E75	97	163	139	129	123	175	177	195	166	186	1550
E72	79	80	64	80	73	114	80	112	76	70	828
E74	62	51	95	74	79	81	62	78	93	73	748
Q04-Q61	72	66	69	74	75	66	85	75	70	75	727
G10-G12	70	73	46	69	83	81	68	63	80	54	687
Q93-Q99	56	63	46	55	65	63	54	49	73	86	610
F01-F84	45	66	45	53	49	57	75	76	55	63	584
G60	56	58	49	58	52	43	77	59	66	64	582
E70	72	42	70	61	50	70	48	47	56	51	567
M87	46	47	54	58	47	63	71	54	49	69	558
H31-H90	55	48	50	54	45	55	52	60	46	45	510
E71	35	39	38	36	41	57	40	40	50	36	412
E16-E56	31	33	33	30	25	26	33	45	31	36	323
E78	30	38	28	36	32	33	24	27	16	26	290
E77	32	29	28	29	32	37	21	16	25	29	278
E79	28	24	26	27	41	20	29	36	17	25	273
G24-G40	19	13	22	31	33	17	24	20	20	32	231
E76	12	19	18	15	18	21	20	37	27	33	220
Q77-Q78	20	17	23	14	10	14	20	24	24	26	192

Tabla 7.6: Producción por clasificación ICD-10 y fecha. Pubmed

Clasificación ICD-10	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Total
E83-E88	728	767	750	770	706	867	769	826	895	914	7992
Q80-Q89	249	226	198	241	234	265	261	284	286	263	2507
G71	184	185	182	179	176	168	183	182	177	208	1824
E75	69	89	114	115	109	153	127	142	137	132	1187
E72	93	102	89	83	89	73	96	83	90	59	857
F01-F84	43	69	49	55	65	72	76	72	79	66	646
Q04-Q61	60	60	64	59	53	76	62	63	46	67	610
G10-G12	61	53	51	55	66	58	63	67	73	55	602
E74	47	47	75	57	62	64	53	57	56	52	570
E78	36	61	60	57	49	43	58	46	43	39	492
H31-H90	41	60	43	56	39	47	40	51	42	53	472
E71	45	42	32	49	50	58	51	48	46	47	468
E16-E56	41	40	55	43	49	41	41	43	45	61	459
E79	33	22	33	35	44	31	38	41	28	25	330
Q93-Q99	28	30	31	24	24	31	22	32	41	45	308
Q77-Q78	19	27	28	23	19	22	26	32	30	30	256
E70	28	17	22	17	22	19	26	24	23	22	220
G60	26	27	18	20	27	18	26	18	13	22	215
E77	18	31	19	14	23	21	16	17	18	28	205
G24-G40	11	17	13	20	23	13	27	20	13	25	182
E76	10	9	18	14	13	18	20	23	26	30	181
M87	1	4	3	6	7	13	14	2	9	8	67

Tabla 7.7: Producción por clasificación ICD-10 y fecha. SCI

Categoría WOS	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Total
Genetics & Heredity	392	386	361	365	396	396	390	366	371	369	3792
Biochemistry & Molecular Biology	307	312	317	323	348	320	307	295	331	357	3217
Neurosciences & Neurology	247	287	309	291	315	314	313	345	316	313	3050
Pediatrics	176	191	196	214	180	213	207	212	180	196	1965
Respiratory System	123	115	136	127	122	161	167	170	208	175	1504
Endocrinology & Metabolism	111	112	122	120	116	173	151	143	148	151	1347
Cell Biology	118	116	104	114	103	112	107	141	110	129	1154
Research & Experimental Medicine	99	106	92	101	108	97	98	95	125	102	1023
General & Internal Medicine	103	103	102	74	77	68	73	95	70	75	840
Microbiology	45	67	54	82	74	95	60	87	80	97	741
Gastroenterology & Hepatology	68	75	79	75	54	66	92	55	63	68	695
Pharmacology & Pharmacy	41	48	49	46	47	58	76	75	82	86	608
Biotechnology & Applied Microbiology	67	57	54	50	58	41	41	58	62	66	554
Surgery	44	43	42	50	47	60	45	57	72	69	529
Cardiovascular System & Cardiology	44	40	42	35	50	70	51	60	57	71	520
Hematology	40	40	56	58	50	73	47	41	54	52	511
Biophysics	60	60	46	53	47	40	46	52	41	57	502

Tabla 7.8: Producción por categoría WOS y fecha (Más de 500 documentos)

7.1.5. Instituciones

Tal y como se ha apuntado en la metodología, para el tratamiento de las instituciones se han utilizado exclusivamente los datos obtenidos del SCI, puesto que Pubmed solo ofrece información acerca de la institución del primer autor y, por otro lado, la firma institucional que recoge Pubmed no cuenta con una estructura normalizada que permita aplicar métodos para el tratamiento masivo de los datos, tal y como se hizo con las direcciones del SCI, para identificar la institución matriz dentro de cada firma institucional.

Los principales tipos institucionales identificados son los siguientes: Hospitales, universidades, centros nacionales de investigación y empresas y fundaciones, que se corresponden con las siguientes etiquetas: HOSP, UNIV, NATL-RES Y EMP-FDN.

En la tabla 7.9 aparecen los principales tipos institucionales junto con su producción a lo largo del periodo de estudio, destacando notablemente los centros hospitalarios.

Tipo institucional	Total docs.
HOSP	13427
UNIV	6803
NATL-RES	6768
EMP-FDN	307

Tabla 7.9: *Producción por tipos institucionales*

Siguiendo el método descrito en la metodología para normalizar las instituciones, el número de instituciones representadas se reduce a 8.974 instituciones diferentes, pertenecientes a países de la Europa de los 15.

El núcleo de instituciones que se pueden observar en la tabla 7.10 ha sido obtenido aplicando el método de LOTKA (1926) a la distribución de instituciones (Ver apéndices: tabla B.3). De esta forma se ha obtenido el núcleo de instituciones que

Instituciones	País	Trabajos
Univ Paris	Fra	598
Katholieke Univ Leuven	Bel	375
Univ Amsterdam	Nld	358
Univ Milan	Ita	307
Univ Oxford	Gbr	306
Univ Munich	Deu	296
Cnrs	Fra	289
Hop Necker Enfants Malad	Fra	288
Inserm	Fra	278
Univ Cambridge	Gbr	259
Univ London Imperial Coll Sci Technol & Med	Gbr	230
Leiden Univ	Nld	230
Univ Heidelberg	Deu	202
Univ Tubingen	Deu	201
Ucl	Gbr	200
Univ Roma La Sapienza	Ita	199
Hannover Med Sch	Deu	197
Inst Child Hlth	Gbr	192
Univ Helsinki	Fin	187

Tabla 7.10: *Núcleo de instituciones en SCI*

se pueden observar en la tabla 7.10, y que acumulan el 33,67 % de la producción. Por otro lado hay que señalar que estas son las instituciones que han servido para el cálculo de los indicadores multidimensionales, relativos a la filiación institucional. Otro aspecto importante, relacionado con las instituciones, es el país en que radican, tal y como se puede ver en la tabla 7.11, que recoge las instituciones más productivas por país.

Institución	País	Total docs.
Univ Oxford	Gbr	306
Univ Cambridge	Gbr	259
Univ London Imperial Coll Sci Technol & Med	Gbr	230
Ucl	Gbr	198
Inst Child Hlth	Gbr	178
Univ Edinburgh	Gbr	151
Royal Brompton Hosp	Gbr	146
Univ Manchester	Gbr	133
Kings Coll London	Gbr	113
Univ Birmingham	Gbr	112
Univ Munich	Deu	296

Continúa en la página siguiente. . .

Institución	País	Total docs.
Univ Heidelberg	Deu	202
Univ Tübingen	Deu	201
Hannover Med Sch	Deu	197
Univ Bonn	Deu	186
Tech Univ Munich	Deu	155
Univ Göttingen	Deu	144
Univ Freiburg	Deu	129
Univ Düsseldorf	Deu	128
Univ Hamburg	Deu	118
Univ Paris	Fra	598
Cnrs	Fra	289
Hop Necker Enfants Malad	Fra	288
Inserm	Fra	278
Hop Robert Debre	Fra	134
Inst Pasteur	Fra	110
Univ Lyon 1	Fra	100
Hop Debrousse	Fra	95
Hop La Pitié Salpêtrière	Fra	91
Hop Henri Mondor	Fra	90
Univ Milan	Ita	307
Univ Roma La Sapienza	Ita	199
Univ Siena	Ita	163
Univ Padua	Ita	152
Cnr	Ita	151
Univ Florence	Ita	145
Univ Naples Federico 2	Ita	141
Univ Verona	Ita	139
Univ Genoa	Ita	137
Univ Pavia	Ita	126
Univ Amsterdam	Nld	358
Leiden Univ	Nld	230
Erasmus Mc	Nld	143
Erasmus Univ	Nld	139
Radboud Univ Nijmegen Med Ctr	Nld	120
Univ Utrecht	Nld	116
Univ Groningen	Nld	106
Univ Groningen Hosp	Nld	90
Univ Med Ctr Utrecht	Nld	82
Acad Med Ctr	Nld	77
Univ Barcelona	Esp	134
Csic	Esp	80
Univ Autónoma Madrid	Esp	71
Hosp Clin Barcelona	Esp	52

Continúa en la página siguiente...

Institución	País	Total docs.
Hosp Ramon & Cajal	Esp	46
Univ Autonoma Barcelona	Esp	43
Hosp 12 Octubre	Esp	43
Univ Pompeu Fabra	Esp	39
Hosp Duran & Reynals	Esp	36
Fdn Jimenez Diaz	Esp	34
Karolinska Inst	Swe	177
Univ Gothenburg	Swe	136
Karolinska Univ Hosp	Swe	121
Lund Univ	Swe	74
Uppsala Univ	Swe	61
Univ Uppsala Hosp	Swe	58
Huddinge Univ Hosp	Swe	50
Sahlgrens Univ Hosp	Swe	46
Umea Univ	Swe	36
Univ Uppsala	Swe	35
Univ Copenhagen	Dnk	160
Rigshosp	Dnk	101
Univ Copenhagen Hosp	Dnk	87
Aarhus Univ Hosp	Dnk	81
Univ Aarhus	Dnk	76
Tech Univ Denmark	Dnk	54
John F Kennedy Inst	Dnk	31
Copenhagen Univ Hosp	Dnk	31
Natl Univ Hosp Copenhagen	Dnk	20
Odense Univ Hosp	Dnk	20
Med Univ Vienna	Aut	125
Innsbruck Med Univ	Aut	86
Univ Vienna	Aut	73
Graz Univ	Aut	29
Innsbruck Univ	Aut	27
Med Univ Graz	Aut	21
Krankenanstalt Rudolfstiftung Wien	Aut	20
Karl Franzens Univ Graz	Aut	15
Graz Tech Univ	Aut	12
Univ Hosp Vienna	Aut	11
Queens Univ Belfast	Irl	82
Belfast City Hosp	Irl	71
Trinity Coll Dublin	Irl	47
Our Ladys Hosp Sick Children	Irl	36
Beaumont Hosp	Irl	36
Univ Coll Dublin	Irl	34
Univ Ulster	Irl	25

Continúa en la página siguiente. . .

Institución	País	Total docs.
St Vincent Univ Hosp	Irl	24
Royal Coll Surgeons Ireland	Irl	23
Natl Univ Ireland Univ Coll Cork	Irl	17
Univ Helsinki	Fin	187
Univ Turku	Fin	91
Univ Helsinki Hosp	Fin	68
Univ Oulu	Fin	54
Tampere Univ Hosp	Fin	46
Univ Kuopio	Fin	39
Family Fed Finland	Fin	31
Turku Univ Hosp	Fin	30
Oulu Univ Hosp	Fin	29
Univ Tampere	Fin	29
Univ Athens	Grc	131
Aristotle Univ Thessaloniki	Grc	53
Aghia Sophia Childrens Hosp	Grc	37
Univ Crete	Grc	21
Univ Ioannina	Grc	20
Inst Child Hlth	Grc	15
Univ Patras	Grc	15
Democritus Univ Thrace	Grc	15
Henry Dunant Hosp	Grc	14
Aghia Sofia Childrens Hosp	Grc	13
Univ Porto	Prt	97
Univ Lisbon	Prt	38
Hosp Santa Maria	Prt	30
Univ Minho	Prt	24
Ibmc	Prt	23
Univ Coimbra	Prt	23
Hosp Sao Joao	Prt	16
Natl Inst Hlth	Prt	15
Genet Med Cent Jacinto Magalhaes	Prt	12
Abel Salazar Inst Biomed Sci	Prt	12
Katholieke Univ Leuven	Bel	373
Free Univ Brussels	Bel	106
State Univ Ghent	Bel	96
State Univ Ghent Hosp	Bel	71
Univ Hosp Gasthuisberg	Bel	50
Univ Ghent	Bel	44
Univ Antwerp Hosp	Bel	33
Univ Antwerp	Bel	32
Ghent Univ Hosp	Bel	30
Vrije Univ Brussels	Bel	20

Continúa en la página siguiente...

Institución	País	Total docs.
Ctr Hosp Luxembourg	Lux	2
Clin Pediat Luxembourg	Lux	1
Clin St Joseph	Lux	1

Tabla 7.11: Principales instituciones productoras por país

En cuanto a la distribución de los tipos institucionales por países, se encuentra en la tabla 7.12, en la que se recogen los 15 países de la UE, y la presencia que tienen en ellos dichos tipos institucionales.

País	EMP-FDN	HOSPITAL	NATL-RES	UNIV
AUT	0	748	258	142
BEL	3	1187	160	836
DEU	16	4703	3086	2505
DNK	1	777	330	229
ESP	111	2604	518	482
FIN	3	879	233	382
FRA	53	7998	1089	1257
GBR	86	7054	2350	2190
GRC	4	687	113	140
IRL	0	711	114	175
ITA	130	3153	2283	3450
LUX	1	5	0	0
NLD	31	4367	347	404
PRT	2	423	350	210
SWE	3	1010	569	287

Tabla 7.12: Tabla de contingencia entre países y tipos institucionales

El test del χ^2 , aplicado a la tabla de contingencia entre tipos institucionales y países (tabla 7.12), muestra un resultado de 18097,14 en el estadístico, con un p-valor inferior a 2,2e-16, por lo que se puede concluir que existe relación de dependencia entre las variables.

Dicha relación de dependencia es observable en el gráfico 7.5, junto con el contraste de hipótesis correspondiente, mostrando en que valores se encuentran discrepancias sobre una distribución teórica equidistante, En este sentido destaca la

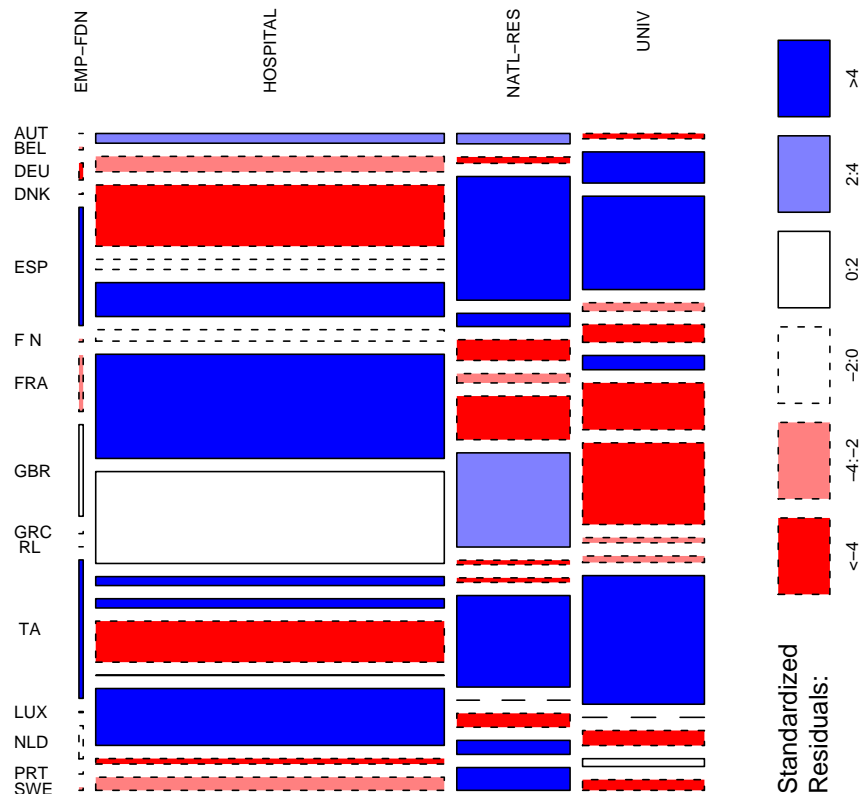


Figura 7.5: Mosaico de la distribución de tipos institucionales por países

sobrerrepresentación de los hospitales en tres países (Francia, España y Holanda), mientras que en otros países como Alemania e Italia, se observa que los centros nacionales de investigación y las universidades tienden a mostrar una mayor producción. En cuanto a la producción en el resto de los países, su distribución es muy cercana a la equidistancia. Por último, el país con mayor producción en todo el periodo (Gran Bretaña) muestra una distribución muy cercana a la equidistante teórica, salvo en el caso de las universidades, que se encuentran infrarepresentadas.

7.1.6. Idiomas

En la tabla 7.13, que muestra la distribución idiomática de los documentos en ambas bases de datos, se observa que el idioma predominante es el inglés, seguido del francés, el alemán y el español, que es el tercer idioma en Pubmed y el cuarto

en SCI, puesto que no se tiene en cuenta como idioma la etiqueta “unspecific” que figura en las bases de datos de Thomsom Reuters cuando sus sistema no son capaces de identificar el idioma del documento.

En el análisis de datos multivariantes, en el gráfico 10.14, se puede observar también como se distribuyen los idiomas por países.

Idioma	Pubmed	SCI	% Pubmed	% SCI
English	19448	17845	91,00 %	93,98 %
French	865	493	4,04 %	2,59 %
Spanish	496	60	2,32 %	0,31 %
German	302	220	1,41 %	1,15 %
Dutch	110	2	0,51 %	0,01 %
Italian	106	8	0,49 %	0,04 %
Portuguese	18	10	0,084 %	0,05 %
Japanese	7	0	0,03 %	0 %
Swedish	6	0	0,02 %	0 %
Danish	3	1	0,01 %	0,005 %
Finnish	3	0	0,01 %	0 %
Chinese	2	0	0,009 %	0 %
Bulgarian	1	0	0,004 %	0 %
Hungarian	1	1	0,004 %	0,005 %
Romanian	1	1	0,004 %	0,005 %
Polish	1	0	0,004 %	0 %
Serbian	1	0	0,004 %	0 %
Unspecific	0	346	0 %	1,82 %
Turkish	0	1	0 %	0,005 %

Tabla 7.13: Producción por idiomas en Pubmed y SCI

7.1.7. Colaboración

En esta sección se han obtenido diversos indicadores relativos al perfil de colaboración presente en los trabajos analizados. Para ello se ha calculado el índice de colaboración por países, tipos institucionales y categorías WOS (tablas 7.14, 7.15 y 7.16) y los índices de coautoría agregados por país, tipo institucional, áreas WOS y términos MeSH (tablas 7.17, 7.18, 7.19 y tablas B.4 y B.5 en los apéndices). Por otro lado, en las tablas 7.20, 7.21, 7.22 y 7.23 se encuentra la evolución de la colaboración nacional e internacional, tanto por tipo institucional como por país. Cabe destacar que, debido a la limitación de Pubmed de recoger únicamente la institución del primer autor, estos indicadores de colaboración sólo han sido obtenidos para los trabajos publicados en el SCI, que si ofrece las direcciones institucionales de todos los autores firmantes.

7.1.7.1. Índice de colaboración

Este indicador muestra la media del número de instituciones firmantes por país (tabla 7.14), por tipos institucionales (tabla 7.15) y por categorías WOS (tabla 7.16). Se ha claculado únicamente para los datos obtenidos del SCI, puesto que Pubmed solo recoge la firma institucional del primer autor.

En la tabla de índices de colaboración por países destacan como países más colaboradores Finlandia, Portugal y Dinamarca, con más de 4,5 instituciones por trabajo como promedio, mientras que Gran Bretaña y Alemania, a pesar de ser los países con mayor producción son también, junto con Luxemburgo, los que poseen un índice de colaboración más bajo.

En cuanto al índice de colaboración por tipos institucionales todos los valores son muy similares, situándose los hospitales en primer lugar y las empresas y fundaciones en el último.

Por último, en los valores agregados por categorías WOS, el dato más destacable es el índice de colaboración del área de Genética (tabla 7.16), que es notablemente superior al del resto de categorías.

7.1.7.2. Coautoría

En cuanto al número de firmas por documento, la tabla 7.17 hace referencia al índice de coautoría desglosado por país y año de los documentos y la tabla 7.18 a la media de coautorías por tipo institucional.

Entre los países, aquellos con una media de coautorías más alta son Luxemburgo (poco significativo debido a su volumen de documentos), Italia, Bélgica y Portugal, todos ellos con una media de más de 8 autores por documento.

En cuanto a los valores de coautorías agregados por tipos institucionales, al igual que ocurría con el índice de colaboración, son muy similares entre si, aunque en este caso las empresas se sitúan en la primera posición mientras que en el índice de colaboración lo hacían en la última.

Por último, las tablas 7.19, B.4 y B.5 muestran la distribución de la coautoría por año y áreas WOS (más de 500 documentos en todo el periodo), y por año y por los términos MeSH que recogen al menos el 50% de la producción. En el caso de las áreas WOS, solo se muestran para la base de datos SCI, mientras que los términos MeSH pueden utilizarse en ambos casos, gracias al script de asignación de términos MeSH a través de las keywords, que se desarrolló en la metodología.

En las áreas WOS también se sitúa la genética como el área con un mayor índice de coautoría, superando en más de un autor al resto de áreas.

Al analizar las tablas de coautorías por términos MeSH se confirma que los términos con un mayor índice de coautoría hacen referencia a conceptos relacionados con la genética. Destaca además que la fibrosis quística, que aparecía en la primera posición en el análisis de producción posee un índice de coautoría muy bajo

País	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Promedio
FIN	4,39	5,07	4,91	5,83	5,71	6,23	5,56	6,66	6,21	5,59	5,62
PRT	6,32	4,91	3,69	3,85	4,45	4,62	4,94	4,7	5,36	3,84	4,67
DNK	3,54	4,64	3,56	4,4	4,12	4,83	4,28	5,07	5,43	5,2	4,51
BEL	4,03	3,92	3,72	4,31	3,89	4,74	4,55	4,57	5,41	5,32	4,45
AUT	4,02	3,77	2,98	3,69	4,43	4,31	4,28	3,58	5,85	7,04	4,39
NLD	3,6	4,14	3,81	3,83	4,02	4,03	3,99	4,65	5,06	5,84	4,30
SWE	4,12	3,96	3,58	3,49	4,44	3,7	4,63	4,28	4,58	5,2	4,20
FRA	3,5	3,73	3,65	3,66	3,83	3,7	4,02	4,29	4,92	5,61	4,09
ITA	3,67	3,69	3,76	3,86	3,76	4,05	4,02	4,11	4,64	4,89	4,04
ESP	3,77	3,27	3,78	3,75	3,94	3,79	4,09	4,04	4,26	4,12	3,88
IRL	3,24	2,81	2,9	4,84	4	3,36	4,65	4,08	3,3	5,22	3,84
GRC	6,2	2,83	3,46	3,14	4,21	3,27	3,89	2,62	3,53	4,76	3,79
DEU	3,32	3,32	3,36	3,37	3,6	3,76	3,67	4,04	4,19	4,49	3,71
GBR	3,06	3,29	3,25	3,34	3,49	3,54	3,56	3,95	4,05	4,46	3,60
LUX	0	0	0	0	0	0	4	0	0	11	1,5

Tabla 7.14: Índice de colaboración por país y año

Tipo institucional	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Promedio
HOSPITAL	2,71	2,76	2,86	2,86	3,08	3,12	3,13	3,26	3,47	3,68	3,12
UNIV	2,64	2,71	2,76	2,85	2,86	3,05	3,16	3,24	3,38	3,21	3,02
NATL-RES	2,84	2,77	2,65	2,96	2,9	3,12	3,06	3,16	3,06	3,40	3,01
EMP-FDN	2	4	2,65	3,18	2,62	2,95	2,52	3,09	3,63	3,17	2,99

Tabla 7.15: Índice de colaboración por tipo institucional y año

Categoría WOS	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Promedio
Genetics & Hereditary	3,83	4,01	3,73	4,06	4,09	4,22	4,24	4,9	5,24	5,54	4,39
Neurosciences & Neurology	3,12	3,14	3,14	3,38	3,6	3,58	3,78	3,74	4,5	4,49	3,65
Gastroenterology & Hepatology	3,03	2,93	3,18	3,32	3,39	3,11	3,88	3,3	4,15	3,84	3,41
Biochemistry & Molecular Biology	3,12	3,14	2,98	3,16	3,14	3,49	3,44	3,85	3,97	3,79	3,41
Endocrinology & Metabolism	3,2	3,15	3,12	3,06	2,96	3,35	3,19	3,57	3,85	4,51	3,40
Cell Biology	2,87	2,82	2,65	2,9	2,75	3,19	3,26	3,76	3,9	3,79	3,19
Pediatrics	2,74	2,73	2,78	2,73	3,02	3,1	3,15	3,22	3,34	3,6	3,04
General & Internal Medicine	2,5	2,63	2,55	2,99	3,47	3,31	2,74	3,14	3,38	3,47	3,02
Microbiology	2,49	2,64	2,7	2,72	2,62	3,41	3,17	3,29	3,14	3,49	2,97
Respiratory System	2,49	2,63	2,63	2,47	3,04	2,9	3,12	3,09	3,54	3,61	2,95

Tabla 7.16: Índice de colaboración por categorías WOS y año

País	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Promedio
LUX							8,5			15	11,75
ITA	7,21	8,2	7,72	8,03	7,86	8,56	7,98	8,6	8,96	9,53	8,26
BEL	7,92	7,47	7,34	8,01	6,98	8,93	7,63	8,4	10,4	9,11	8,22
PRT	10,53	9	7,5	7,42	7,48	7,19	8,55	8,64	8,29	6,79	8,14
FIN	5,73	6,98	7,62	7,51	7,33	8,98	7,03	9,57	8,79	7,91	7,74
NLD	7,07	7,32	7,2	7,06	7,58	7,49	7,27	8,28	8,7	9,42	7,74
AUT	7,44	6,75	6,16	6,98	7,95	7,57	7,23	6,54	9,81	10,3	7,67
FRA	7,17	7,26	7	6,85	7,35	7,07	7,31	7,41	8,22	8,84	7,45
GRC	10,6	6,37	6,77	6,67	8,79	5,93	7,34	6,41	7,04	8,41	7,43
ESP	7,08	6,86	7,34	7,21	7,21	7,63	7,73	7,64	7,71	7,68	7,41
DNK	5,72	7,54	6,29	7,42	6,75	7,22	6,88	8,65	8,02	8,59	7,31
IRL	6,33	6,12	5,95	8,16	7,66	6,27	7,53	7,43	6,88	8,6	7,09
DEU	6,53	6,57	6,64	6,65	6,62	7,27	6,87	7,44	7,57	7,9	7,01
SWE	6,73	5,87	6,92	6	6,12	6,09	6,99	6,85	6,95	8,22	6,67
GBR	5,68	6,09	6,21	6,06	6,18	6,36	6,12	6,78	7,14	7,67	6,43

Tabla 7.17: Media de coautorías por país y año en SCI

Tipo institucional	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Promedio
EMP-FDN	5,23	7,30	6,25	6,27	6,62	7,61	6,05	7,14	8,18	5,96	6,65
NATL-RES	6,01	5,98	5,77	6,14	5,97	6,24	6,30	6,46	6,33	6,94	6,24
HOSPITAL	5,68	5,80	5,92	5,88	6,11	6,34	6,21	6,28	6,63	6,89	6,20
UNIV	5,2996	5,38	5,44	5,81	5,71	6,36	6,11	6,44	6,74	6,45	6,04

Tabla 7.18: *Media de coautorías por tipo institucional y año*

(5,84 en SCI y 5,44 en Pubmed), frente a los primeros términos que superan ampliamente los 8 autores por documento.

7.1.7.3. Colaboración nacional

En las tablas 7.20 y 7.21 se muestran los valores porcentuales de colaboraciones sobre el total de documentos por tipos de institución y por país respectivamente.

Para calcular el porcentaje de documentos en colaboración nacional se ha tenido en cuenta toda la producción de los 15 países en la que existe colaboración entre instituciones del mismo país. Posteriormente se ha calculado el porcentaje que esta producción representa sobre el total de documentos, con valores agregados por tipos institucionales y por países.

En los tipos institucionales se observa que las empresas y fundaciones son las que poseen un mayor porcentaje de documentos en colaboración nacional, mientras que los centros de investigación nacionales colaboran menos con instituciones del propio país.

En cuanto a los valores de colaboración nacional agregados por país, están en torno al 30 %, destacando los valores de Italia, Francia y España, por encontrarse entre los más altos.

7.1.7.4. Colaboración internacional

Al igual que en las tablas de colaboración nacional, las tablas 7.22 y 7.23 muestran el porcentaje de colaboraciones sobre el total de documentos por tipos de institución y por país.

En los valores agregados por tipos institucionales los centros nacionales de investigación aparecen como los que más colaboran con instituciones internacionales,

Categoría WOS	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Promedio
Genetics & Heredity	7,42	7,63	7,33	7,82	7,73	7,85	7,77	9,08	9,22	9,76	8,16
Biochemistry & Molecular Biology	6,36	6,74	6,55	6,76	6,48	7,23	7,09	7,65	7,53	7,73	7,01
Gastroenterology & Hepatology	6,1	6,28	6,28	6,83	6,06	6,55	7,88	6,51	7,88	7,12	6,75
Neurosciences & Neurology	5,81	6,37	6,02	6,44	6,73	6,36	6,64	6,86	7,68	7,79	6,67
Cell Biology	5,81	6,32	6,32	5,81	5,91	6,67	6,8	7,61	7,43	7,67	6,63
Endocrinology & Metabolism	5,71	6,01	6,7	5,87	6,03	6,67	6,21	6,5	7,06	7,4	6,42
Microbiology	5,18	5,64	5,5	5,7	5,71	6,79	6,02	6,41	6,54	6,78	6,03
Respiratory System	5,3	5,39	5,4	5,06	5,66	5,87	5,6	5,47	6,52	6,16	5,64
Pediatrics	5,34	5,01	5,19	4,92	5,47	5,61	5,71	5,93	5,76	6,11	5,50
General & Internal Medicine	4,81	5,43	4,77	5,19	6,19	6,12	4,94	5,26	5,81	6,03	5,45

Tabla 7.19: Media de coautoría por categorías WOS y año

mientras que los hospitales apenas lo hacen, con un valor medio del 24,6 % de los documentos en colaboración con centros internacionales.

En cuanto a los países, los valores generales, igual que en la colaboración nacional, se encuentran en torno al 30 %, siendo ligeramente más bajos y situándose Bélgica y Portugal entre los que destacan por tener un porcentaje de colaboración internacional mayor que la media.

Por último, la tabla 7.23 muestra la matriz de colaboraciones entre países. Se trata de una matriz simétrica que muestra las colaboraciones internacionales, por lo que la diagonal no refleja la colaboración nacional, si no que es igual a cero. Esta matriz es la que ha sido utilizada para calcular los indicadores multidimensionales y conexionistas relativos a las redes de colaboración entre países.

Tipo institucional	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Promedio
EMP-FDN	30 %	38,89 %	30 %	48 %	46,67 %	48,72 %	47,22 %	42,11 %	38,46 %	44,19 %	42,13 %
HOSPITAL	38,30 %	39,26 %	38,53 %	36,45 %	38,63 %	39,07 %	37,64 %	36,95 %	36,74 %	34,19 %	37,46 %
UNIV	37,32 %	38,10 %	35,90 %	34,35 %	36,01 %	38,94 %	36,32 %	35,82 %	34,31 %	33,17 %	35,83 %
NATL-RES	38,34 %	36,17 %	37,18 %	36,26 %	34,63 %	37,90 %	35,83 %	37,01 %	33,20 %	33,09 %	35,81 %

Tabla 7.20: Porcentaje de documentos en colaboración nacional por tipo institucional

País	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Promedio
ITA	44,96%	46,79%	47,29%	46,32%	50%	47,22%	48,36%	48,69%	46,67%	46,98%	47,33%
ESP	39,05%	47,50%	43,51%	39,83%	42,86%	55,80%	45,93%	45,83%	45,12%	48,55%	45,40%
FRA	42,44%	43,92%	42,56%	40,05%	49,02%	47,30%	46,20%	48,05%	49,39%	44,36%	45,33%
GRC	46,67%	42,86%	26,92%	38,10%	24,14%	45,45%	47,37%	48,72%	45,61%	37,84%	40,37%
FIN	49,02%	35,71%	45,28%	34,29%	38,10%	35,42%	43,59%	36,36%	45,45%	34,38%	39,76%
NLD	31,98%	36,13%	29,89%	37,30%	32,95%	40,93%	36,13%	41,85%	36,97%	33,79%	35,79%
DEU	38,75%	28,01%	32,99%	33,72%	33,33%	35,12%	32,70%	38,48%	30,68%	32,79%	33,66%
IRL	30,30%	21,88%	29,27%	25,81%	25%	48,89%	50%	37,25%	40%	27,27%	33,57%
DNK	31,58%	36,07%	33,33%	36,54%	35,71%	34,21%	31,67%	37,04%	24,07%	20,34%	32,06%
PRT	21,05%	22,73%	37,50%	19,23%	27,59%	40,62%	38,71%	31,91%	32,14%	41,86%	31,33%
GBR	33,64%	31,89%	34,82%	26,83%	34%	33,12%	26,65%	30,38%	32,42%	26,39%	31,01%
SWE	30,12%	40,21%	34,18%	31,71%	27,27%	34%	29,33%	29,58%	24,32%	28,57%	30,93%
BEL	22,11%	22,58%	19,35%	22,02%	26,21%	19,75%	26,74%	31,37%	30,36%	25,98%	24,65%
AUT	35,56%	16,98%	22,22%	20,41%	15%	23,88%	23,40%	26,32%	20,83%	16%	22,06%
LUX	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Tabla 7.21: *Porcentaje de documentos en colaboración nacional por país*

Tipo institucional	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Promedio
NATL-RES	32,28 %	36,57 %	32,08 %	36 %	36,43 %	36,65 %	35,39 %	32,03 %	36,44 %	36,37 %	35,11 %
UNIV	31,70 %	34,97 %	31,25 %	33,16 %	35,5 %	31,91 %	34,53 %	29,40 %	32,27 %	35,25 %	32,98 %
EMP-FDN	45 %	44,44 %	35 %	28 %	33,33 %	20,51 %	22,22 %	26,32 %	37,18 %	26,74 %	30,2 %
HOSPITAL	23,54 %	24,21 %	23 %	23,76 %	25,97 %	24,78 %	26,86 %	22,61 %	24,62 %	26,27 %	24,6 %

Tabla 7.22: Porcentaje de documentos en colaboración internacional por tipo institucional

País	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Promedio
BEL	58,95 %	68,82 %	60,22 %	60,55 %	66,02 %	71,60 %	60,47 %	54,90 %	57,14 %	64,57 %	62,32 %
AUT	40 %	60,38 %	55,56 %	53,06 %	73,33 %	59,70 %	63,83 %	56,14 %	66,67 %	82 %	61,07 %
PRT	68,42 %	72,73 %	50 %	61,54 %	68,97 %	56,25 %	61,29 %	61,70 %	62,50 %	44,19 %	60,76 %
DNK	56,14 %	57,38 %	52,08 %	53,85 %	57,14 %	59,21 %	61,67 %	59,26 %	62,96 %	69,49 %	58,92 %
SWE	53,01 %	50,52 %	54,43 %	50 %	63,64 %	54 %	54,67 %	57,75 %	68,92 %	65,71 %	57,26 %
NLD	54,65 %	52,90 %	51,09 %	52,43 %	54,34 %	50,26 %	51,68 %	51,10 %	58,77 %	61,64 %	53,89 %
FIN	43,14 %	52,38 %	47,17 %	57,14 %	57,14 %	56,25 %	51,28 %	61,36 %	51,52 %	56,25 %	53,36 %
DEU	39,57 %	48,84 %	45,45 %	44,19 %	48,21 %	49,30 %	50,74 %	45,84 %	51,21 %	52,25 %	47,56 %
GBR	41,30 %	44,44 %	41,96 %	48,56 %	45,60 %	45,09 %	50,11 %	51,05 %	50,88 %	55,26 %	47,43 %
IRL	48,48 %	43,75 %	34,15 %	45,16 %	56,82 %	37,78 %	40 %	41,18 %	46 %	56,36 %	44,97 %
GRC	40 %	37,14 %	46,15 %	47,62 %	65,52 %	50 %	42,11 %	23,08 %	35,09 %	51,35 %	43,81 %
FRA	34,88 %	39,78 %	37,80 %	37,21 %	37,91 %	37,03 %	41,03 %	36,88 %	41,56 %	45,86 %	39 %
ITA	36,05 %	36,07 %	34,50 %	38,60 %	35,51 %	37,65 %	42,43 %	37,03 %	42,50 %	43,41 %	38,37 %
ESP	37,14 %	27,50 %	38,93 %	31,36 %	40,34 %	31,16 %	41,48 %	40,97 %	38,41 %	41,04 %	36,83 %
LUX	0 %	0 %	0 %	0 %	0 %	0 %	100 %	0 %	0 %	100 %	20 %

Tabla 7.23: *Porcentaje de documentos en colaboración internacional por país*

Países	AUT	BEL	DEU	DNK	ESP	FIN	FRA	GBR	GRC	IRL	ITA	LUX	NLD	PRT	SWE
AUT	0	22	150	4	20	4	32	69	6	4	47	0	25	5	7
BEL	22	0	139	27	45	9	153	136	5	14	87	2	172	16	27
DEU	150	139	0	99	101	35	251	440	24	20	225	0	305	36	93
DNK	4	27	99	0	15	10	39	82	4	10	21	0	42	8	62
ESP	20	45	101	15	0	13	88	110	8	11	102	1	48	30	19
FIN	4	9	35	10	13	0	30	52	4	5	21	0	36	3	62
FRA	32	153	251	39	88	30	0	297	12	27	190	0	150	29	55
GBR	69	136	440	82	110	52	297	0	31	92	294	1	196	40	119
GRC	6	5	24	4	8	4	12	31	0	0	26	1	21	1	6
IRL	4	14	20	10	11	5	27	92	0	0	17	0	22	2	3
ITA	47	87	225	21	102	21	190	294	26	17	0	1	138	22	46
LUX	0	2	0	0	1	0	0	1	1	0	1	0	1	0	0
NLD	25	172	305	42	48	36	150	196	21	22	138	1	0	20	44
PRT	5	16	36	8	30	3	29	40	1	2	22	0	20	0	8
SWE	7	27	93	62	19	62	55	119	6	3	46	0	44	8	0

Tabla 7.24: Matriz de colaboración por países UE15

7.1.8. Dispersión

La distribución de trabajos por revista en el SCI (gráfico 7.6) muestra la clásica curva de distribución de Bradford en la que pueden verse los tres estadios diferentes (núcleo, recta y saturación). La tabla completa, compuesta por 2.083 revistas, puede consultarse en los apéndices (Tabla B.6).

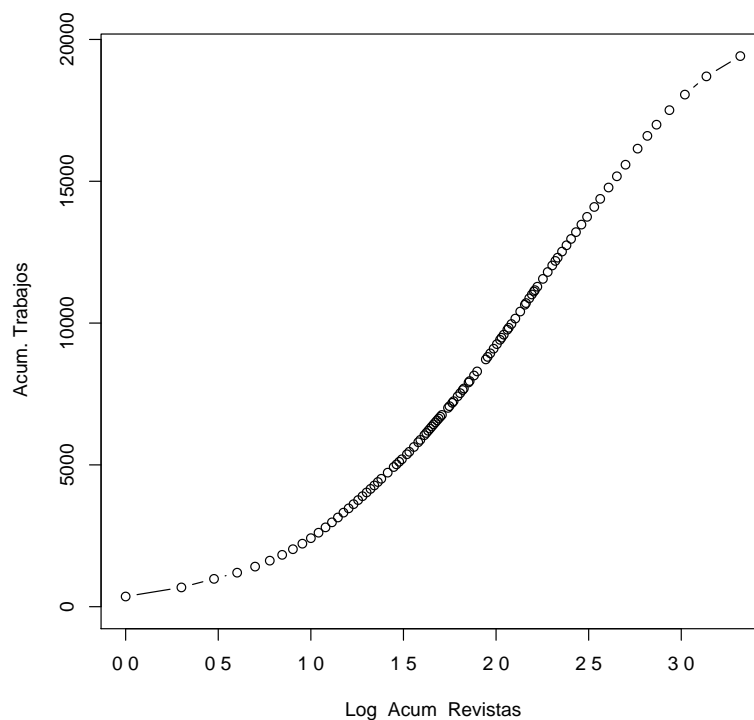


Figura 7.6: *Dispersión por revistas en SCI*

El núcleo observado en el gráfico 7.6 se ha determinado dividiendo las revistas en tres zonas, que acumulan aproximadamente la misma cantidad de artículos. Las revistas que componen dicho núcleo se muestran en la tabla 7.25. Este núcleo está formado por 46 revistas, que acumulan el 33,12% de la producción y en él destacan, situándose entre las primeras, las revistas de Genética.

Revista	Total de artículos
Journal Of Biological Chemistry	359
Journal Of Inherited Metabolic Disease	321
Human Molecular Genetics	302

Continúa en la página siguiente. . .

Revista	Total de artículos
Molecular Genetics And Metabolism	218
European Journal Of Human Genetics	214
American Journal Of Medical Genetics Part A	210
American Journal Of Human Genetics	205
Pediatric Pulmonology	199
Neurology	195
Journal Of Medical Genetics	194
European Respiratory Journal	193
Human Mutation	188
Neuromuscular Disorders	177
Journal Of Cystic Fibrosis	172
Journal Of Clinical Endocrinology & Metabolism	170
Biochemical And Biophysical Research Communications	151
Archives De Pediatrie	148
American Journal Of Respiratory And Critical Care Medicine	144
European Journal Of Pediatrics	137
Human Genetics	132
Thorax	125
Proceedings Of The National Academy Of Sciences Of The United States Of America	122
Acta Paediatrica	121
Prenatal Diagnosis	114
Journal Of Clinical Microbiology	108
Clinical Genetics	108
Revista De Neurologia	99
Journal Of Pediatrics	99
Biochemical Journal	94
Chest	91
Clinical Chemistry	89
Blood	88
Pediatric Research	88
American Journal Of Medical Genetics	87
Hormone Research	85
Febs Letters	85
Revue Des Maladies Respiratoires	84
Journal Of Lipid Research	84
Archives Of Disease In Childhood	83
Blood Cells Molecules And Diseases	82
Brain	82
Neuropediatrics	79
Respiratory Medicine	78
Annals Of Neurology	77

Continúa en la página siguiente. . .

Revista	Total de artículos
Archives Of Neurology	75
Antimicrobial Agents And Chemotherapy	74

Tabla 7.25: *Núcleo de revistas en SCI*

Replicando este procedimiento para obtener la distribución de trabajos por publicación en Pubmed, se obtienen los resultados que se observan en el gráfico 7.7 la tabla 7.26 y el apéndice B.7, que recoge la distribución completa, con 2325 revistas. En este caso son 51 revistas las que acumulan el 33,10% de los trabajos publicados y de nuevo destacan las revistas relacionadas con la Genética, ocupando los primeros puestos del núcleo.

Revista	Total de artículos
Journal Of Inherited Metabolic Disease	424
Human Mutation	305
Human Genetics	286
Journal Of Cystic Fibrosis : Official Journal Of The European Cystic Fibrosis Society	282
Neurology	254
Human Molecular Genetics	248
The Journal Of Biological Chemistry	247
Molecular Genetics And Metabolism	211
Neuromuscular Disorders : Nmd	196
European Journal Of Human Genetics : Ejhg	195
American Journal Of Medical Genetics. Part A	188
Pediatric Pulmonology	158
Archives De Pediatrie : Organe Officiel De La Societe Francaise De Pediatrie	157
American Journal Of Human Genetics	155
Biochimica Et Biophysica Acta	145
The European Respiratory Journal : Official Journal Of The European Society For Clinical Respiratory Physiology	144
European Journal Of Pediatrics	143
Journal Of Medical Genetics	131
The Journal Of Clinical Endocrinology And Metabolism	127
Revista De Neurologia	124
Thorax	114

Continúa en la página siguiente. . .

Revista	Total de artículos
Clinical Genetics	112
Archives Of Disease In Childhood	110
Movement Disorders : Official Journal Of The Movement Disorder Society	107
The Journal Of Pediatrics	103
Brain : A Journal Of Neurology	98
Prenatal Diagnosis	97
Nederlands Tijdschrift Voor Geneeskunde	96
Acta Paediatrica (Oslo, Norway : 1992)	93
Archives Of Neurology	93
Cochrane Database Of Systematic Reviews (Online)	93
Proceedings Of The National Academy Of Sciences Of The United States Of America	93
Biochemical And Biophysical Research Communications	91
Blood Cells, Molecules & Diseases	91
Clinical Chemistry	88
Annals Of Neurology	87
Journal Of Pediatric Surgery	87
American Journal Of Respiratory And Critical Care Medicine	86
Neuropediatrics	83
Journal Of The Royal Society Of Medicine	79
Muscle & Nerve	79
Genetic Testing	77
Chest	76
Pediatric Research	76
American Journal Of Medical Genetics	72
Journal Of Neurology, Neurosurgery, And Psychiatry	70
Hormone Research	69
European Journal Of Paediatric Neurology : Ejpn : Official Journal Of The European Paediatric Neurology Society	68
The Biochemical Journal	68
Journal Of Neurology	67
La Revue De Medecine Interne / Fondée ... Par La Société Nationale Française De Médecine Interne	67

Tabla 7.26: Núcleo de revistas en Pubmed

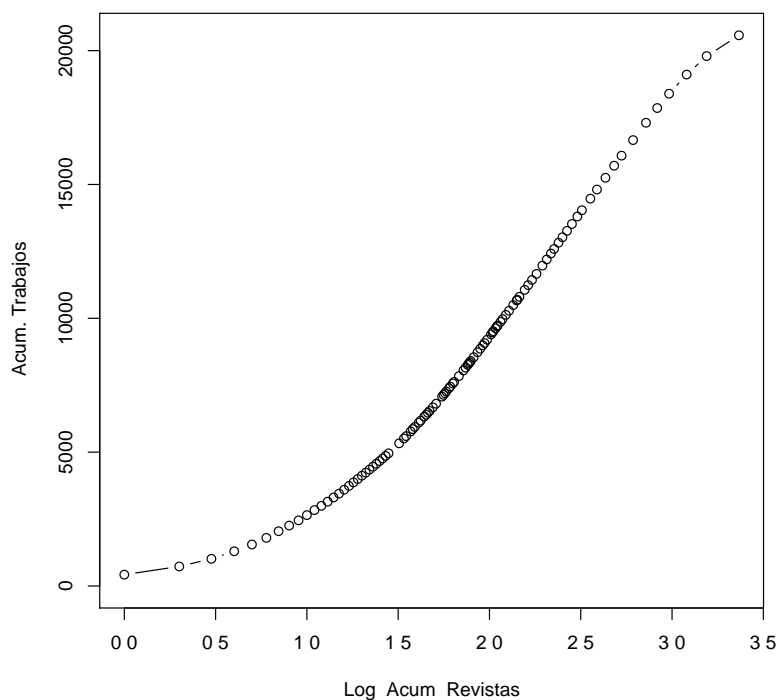


Figura 7.7: *Dispersión por revistas en Pubmed*

7.1.8.1. Concentración temática de las revistas

Para cada revista del núcleo en la base de datos SCI, se calcula el número de categorías temáticas en las que se encuentra recogida (Tabla 7.27), utilizando las áreas del Journal Citation Report (JCR), con el fin de examinar la concentración temática de dichas revistas y determinar si se trata de revistas muy especializadas o si, por el contrario son revistas con un alto grado de multidisciplinariedad.

Como se puede observar tan solo 12 de las 46 revistas pertenecen a más de un área JCR, por lo que el grado de multidisciplinariedad en ellas es mínimo.

Revista	Número de categorías
Febs Letters	3
Molecular Genetics And Metabolism	3
American Journal Of Respiratory And Critical Care Medicine	2
Antimicrobial Agents And Chemotherapy	2

Continúa en la página siguiente...

Revista	Número de categorías
Biochemical And Biophysical Research Communications	2
European Journal Of Human Genetics	2
Human Molecular Genetics	2
Journal Of Inherited Metabolic Disease	2
Neuropediatrics	2
Pediatric Pulmonology	2
Prenatal Diagnosis	2
Respiratory Medicine	2
Acta Paediatrica	1
American Journal Of Human Genetics	1
American Journal Of Medical Genetics	1
American Journal Of Medical Genetics Part A	1
Annals Of Neurology	1
Archives De Pediatrie	1
Archives Of Disease In Childhood	1
Archives Of Neurology	1
Biochemical Journal	1
Blood	1
Blood Cells Molecules And Diseases	1
Brain	1
Chest	1
Clinical Chemistry	1
Clinical Genetics	1
European Journal Of Pediatrics	1
European Respiratory Journal	1
Hormone Research	1
Human Genetics	1
Human Mutation	1
Journal Of Biological Chemistry	1
Journal Of Clinical Endocrinology & Metabolism	1
Journal Of Clinical Microbiology	1
Journal Of Cystic Fibrosis	1
Journal Of Lipid Research	1
Journal Of Medical Genetics	1
Journal Of Pediatrics	1

Continúa en la página siguiente...

Revista	Número de categorías
Neurology	1
Neuromuscular Disorders	1
Pediatric Research	1
Proceedings Of The National Academy Of Sciences Of The United States Of America	1
Revista De Neurologia	1
Revue Des Maladies Respiratoires	1
Thorax	1
Promedio	1,30

Tabla 7.27: *Concentración temática de las revistas SCI*

7.1.9. Productividad de los autores

Los indicadores relativos a la productividad de los autores analizan la actividad científica que desarrollan los investigadores, y tienen como objetivo determinar cuáles son los autores más productivos, es decir, aquellos que acumulan la mayor parte de la producción científica en un área temática determinada y que por tanto, forman el núcleo más importante de investigación en dicha área. En este caso se han utilizado indicadores relativos a la distribución de los trabajos por autor, siguiendo la metodología propuesta por Lotka, así como diferentes medidas, orientadas a medir la concentración de la producción científica por parte de un núcleo de autores determinados.

7.1.9.1. Distribución de Lotka

En cuanto a los autores, también siguen la clásica distribución de Lotka, según la cual “En toda comunidad de autores científicos, el número (A) de aquellos que han publicado una cantidad determinada de trabajos (n), en el curso de varios años de actividad, es decir (An) autores, es igual a la cantidad de los que han

publicado un solo trabajo (A_1) en el mismo período de tiempo, dividida por el cuadrado de n ", por lo que un pequeño número de autores acumulan la mayor parte de los trabajos.

$$A_n = \frac{A_1}{n^2} \quad (7.1)$$

Si bien debemos indicar que en este caso no se ha aplicado la representación gráfica habitual en dicha distribución, sino que se ha representado por medio de la gráfica de dispersión entre el acumulado del número de autores, frente al acumulado del número de trabajos, más propia de las distribuciones de Pareto que de la de Lotka, pero que en este caso se ajusta perfectamente y nos permite observar además las diferencias entre la distribución teórica de Lotka y la observada. Por otro lado, dicha distribución teórica ha sido calculada a partir del número de autores con un solo trabajo, en cada una de las distribuciones observadas, en las dos bases de datos utilizadas.

En este caso, como se puede observar en las tablas 7.28 y 7.29 y las tablas B.9 y B.10 de los apéndices, 16 autores acumulan el 34,09% de los trabajos publicados en Pubmed y 11 autores el 35,24% de la producción en el SCI. Además, ambos gráficos (7.8 y 7.9) muestran que esta distribución (línea roja) se ajusta casi a la perfección a la distribución teórica (línea azul).

Observando el núcleo de autores en ambas bases de datos, se aprecia que son cuatro los autores (Vandamme, Hoiby, Tummler y Hoffmann), que aparecen en el núcleo del SCI pero no en el de Pubmed, que es mucho más numeroso, con 16 grandes productores, frente a los 11 que aparecen en SCI.

En los gráficos 10.3 y 10.5 del apartado de indicadores multidimensionales se puede observar como se distribuyen estos autores en los diferentes grupos de enfermedades.

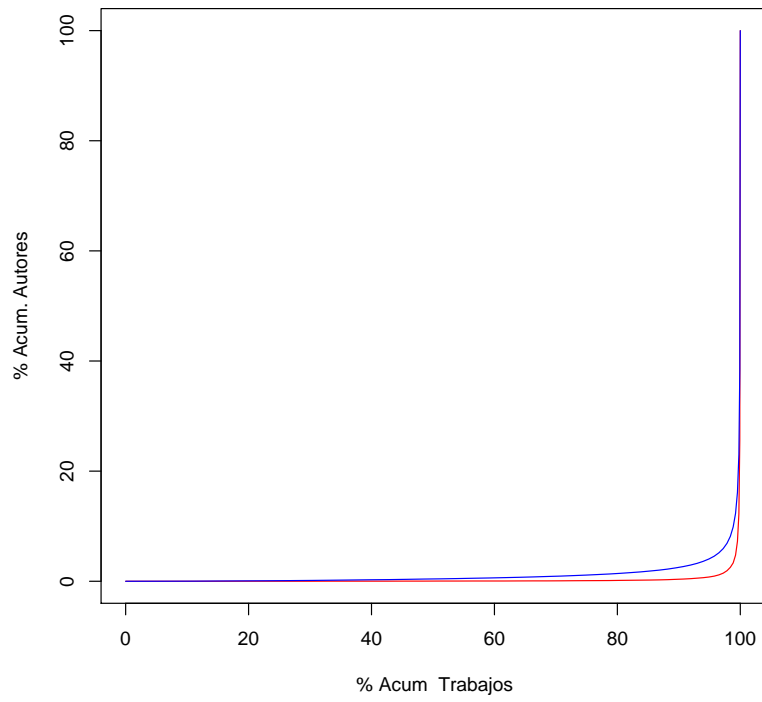


Figura 7.8: *Dispersión por autores en Pubmed*

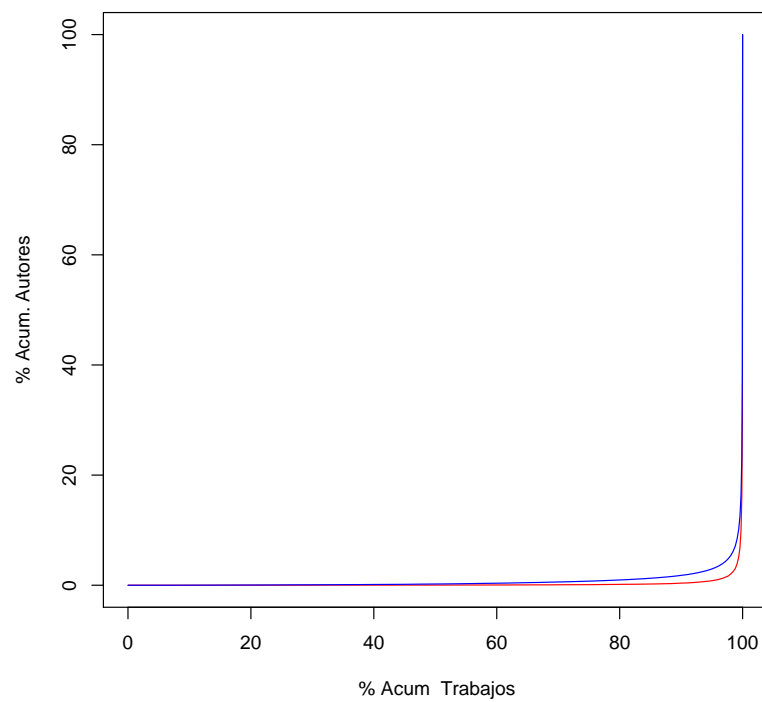


Figura 7.9: *Dispersión por autores en SCI*

Autor	Total Docs.
Wanders, R	130
Wajner, M	99
Bush, A	88
Mehta, A	86
Munnich, A	86
Elborn, J	80
Smeitink, J	79
Ferec, C	79
Beck, M	79
Waterham, H	74
Muntoni, F	73
Conway, S	71
Aerts, J	70
Bisceglia, L	69
Dimauro, S	67
Van Den Heuvel, L	66

Tabla 7.28: Núcleo de autores en Pubmed

Autor	Total docs.
Wanders, Rja	151
Munnich, A	89
Ferec, C	82
Bush, A	79
Beck, M	75
Vandamme, P	74
Hoiby, N	73
Hoffmann, Gf	73
Elborn, Js	71
Waterham, Hr	69
Tummler, B	63

Tabla 7.29: Núcleo de autores en SCI

A ambas distribuciones se les ha aplicado una prueba de bondad de ajuste sobre las distribuciones teóricas, obteniendo unos p-valores de 0,54 en el caso del SCI y de 0,47 en el caso de Pubmed, por lo que no se puede rechazar la hipótesis nula de que las distribuciones de autores observadas y las distribuciones teóricas provienen de la misma población y, por tanto, no existe diferencia significativa entre ellas.

SCI: $t = -0,6083$, $df = 127,843$, $p\text{-value} = 0,544$

Pubmed: $t = -0,7157$, $df = 97,808$, $p\text{-value} = 0,4759$

7.1.9.2. Índice de élite de Price

Según el índice de élite de Price, el número de autores significativamente más productivos se corresponde con la raíz cuadrada del total de autores.

En el caso del SCI, el total de autores es 65.976 y 256 el número de autores que componen la élite, acumulando el 90,15 % de la producción.

Por otro lado en Pubmed el total de autores son 55.433, y son 238 el número de autores que componen la élite y que acumulan el 89,47 % de los trabajos publicados.

Estos datos indican que la distribución de trabajos por autor es muy similar en ambas bases de datos, con un pequeño número de autores que acumulan la mayor parte de los trabajos indizados, tanto en Pubmed como en el SCI.

7.1.9.3. Índice de Yablonsky

Tal y como se estableció en el capítulo metodológico, para calcular el Índice de YABLONSKY (1980) se seleccionan los autores con un número de trabajos equivalente a la raíz cuadrada del número de autores con un solo trabajo. En el caso de Pubmed este número sería de 191,15 trabajos y en el caso de SCI de 213,84 trabajos.

Sin embargo, observando las distribuciones de trabajos por autor en Pubmed y SCI (ver apéndices: B.9 y B.10), queda patente que ningún autor supera esos umbrales de producción, probablemente debido a la presencia de numerosos autores “transeúntes”, que publican un solo trabajo, así como a la heterogeneidad de las enfermedades estudiadas, que conlleva una falta de autores especializados, que si se encontraría en un análisis pormenorizado de cada una de las enfermedades.

7.2. Indicadores de actividad científica (Consumo)

Los siguientes indicadores, relativos al consumo de información científica por parte de los investigadores en el área de estudio, han sido calculados a partir de los datos que ofrece la base de datos Thomson Science Citation Index, pues la base de datos Pubmed no ofrece información acerca de las referencias bibliográficas de los trabajos que recoge.

7.2.1. Distribución de las referencias

En la tabla 7.30 se puede ver el número total de referencias bibliográficas por año y en el gráfico 7.10 la función de densidad de esta distribución, que muestra como se va desplazando la ventana de citación a lo largo del tiempo.

La tabla muestra un aumento progresivo del número de referencias bibliográficas, alcanzando su punto culminante entre los años 1999 y 2001, momento a partir del cual comienzan a decrecer el número de referencias, indicando una ventana de citación relativamente amplia.

Esta ventana de citación puede observarse en el gráfico 7.10, en el que se muestran diez funciones de densidad de forma conjunta, una por cada año de los trabajos citantes en el periodo de estudio (2000 a 2009), por lo que se puede observar cómo va evolucionando la ventana de citación a lo largo del tiempo, volviéndose cada vez menos apuntada e indicando un aumento en la vida media de las referencias bibliográficas.

Año	Total de referencias
Anteriores a 1940	1410
1940	42
1941	43
1942	42
1943	57
1944	46
1945	40
1946	58

Continúa en la página siguiente. . .

Año	Total de referencias
1947	52
1948	90
1949	145
1950	167
1951	444
1952	205
1953	191
1954	213
1955	245
1956	308
1957	262
1958	353
1959	431
1960	253
1961	472
1962	475
1963	624
1964	637
1965	729
1966	824
1967	1024
1968	1116
1969	1149
1970	1361
1971	1293
1972	1421
1973	1488
1974	1615
1975	1806
1976	2462
1977	2344
1978	2714
1979	2890
1980	3553
1981	4071
1982	4319
1983	4858
1984	5771
1985	6192
1986	6979
1987	8453
1988	9987
1989	11837
1990	13840

Continúa en la página siguiente...

Año	Total de referencias
1991	15775
1992	18657
1993	23307
1994	25128
1995	31965
1996	34266
1997	40744
1998	45293
1999	50770
2000	49222
2001	48117
2002	41782
2003	36753
2004	32368
2005	26766
2006	20528
2007	13834
2008	7115
2009	1172
2010	7

Tabla 7.30: *Distribución de referencias por año*

7.2.2. Obsolescencia

El objetivo de analizar la obsolescencia de las referencias utilizadas es determinar la antigüedad media de los documentos consultados y utilizados por los investigadores en sus trabajos y de este modo, conocer cuánto tiempo tardan en perder validez los trabajos científicos en el área de estudio. Para ello se han utilizado dos medidas fundamentales, la vida media o semiperiodo de Burton y Kebler y el Índice de Price, ambas medidas indican, de forma diferente, la obsolescencia de la literatura científica.

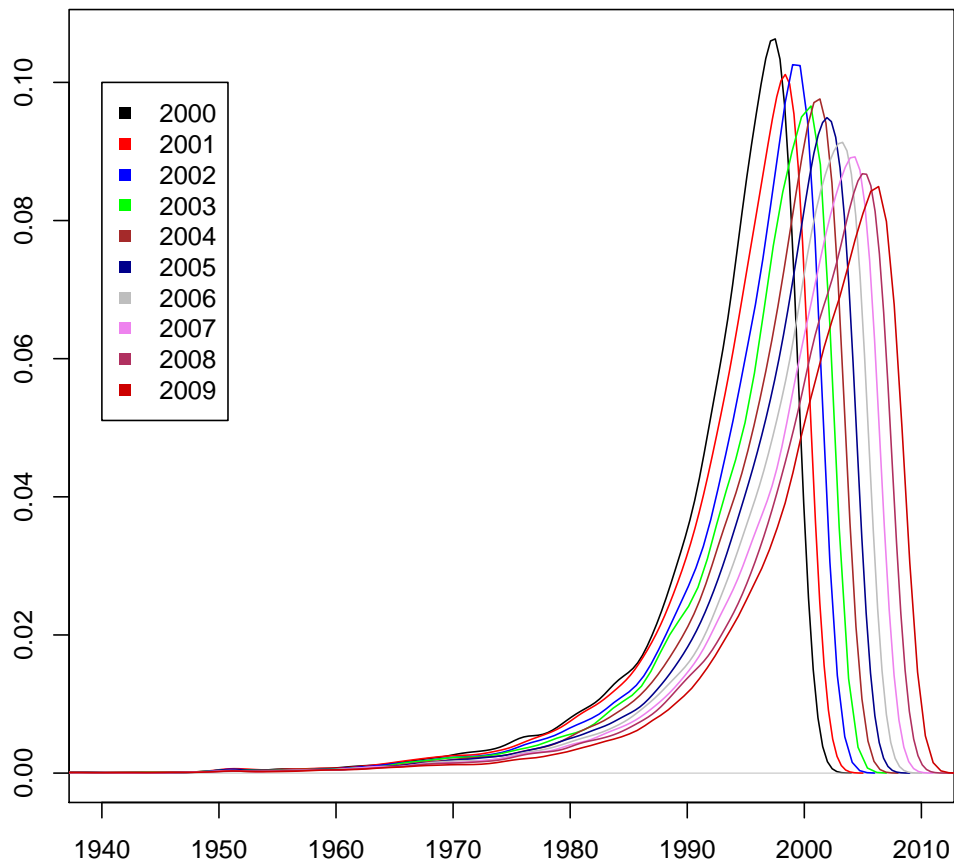


Figura 7.10: *Función de densidad de la distribución de referencias por fecha*

7.2.2.1. Semiperiodo de Burton y Kehler

El semiperiodo de BURTON y KEHLER (1960) se calcula como una mediana de las fechas de las referencias bibliográficas producidas por los documentos de cada año. En este caso se ha calculado tanto para el conjunto de documentos tabla (7.31), como para cada grupo de documentos clasificado según la ICD-10 (tabla 7.32). Para ello se ha utilizado el indicador de vida media por año de publicación de los documentos, tal y como lo enunciaron Burton y Kehler.

En el primer caso la vida media evoluciona desde los 5 años en los que se sitúa en el año 2000 hasta los 7 años de promedio en el 2009. Esta evolución se reproduce con ligeras variaciones en el cálculo agrupado por la Clasificación Internacional de Enfermedades, indicando, tal y como se vio anteriormente, un aumento en la ventana de citación de las enfermedades estudiadas.

Año de publicación	Vida Media
2000	5
2001	6
2002	5
2003	6
2004	6
2005	6
2006	6
2007	6
2008	7
2009	7

Tabla 7.31: *Vida media por año de publicación*

7.2.2.2. Índice de Price

En cuanto al Índice de Price, los resultados de la tabla 7.33 confirman los valores obtenidos en el cálculo de la vida media, mostrando una literatura que, desde el 2000, ve reducido en un 10 % el porcentaje de referencias con 5 años o menos de antigüedad.

Clas. ICD-10	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Promedio
E16-E56	6	7	6	7	7	8	7	7	8	8	7,1
E74	6	8	7	7	7	7	7	7	7	8	7,1
E79	5	9	6	6	7	7	7	7	8	7	6,9
E70	6	6	8	6	7	6	8	7	7	7	6,8
Q77-Q78	5	6	6	6,5	6	7	8	8	8	7	6,75
E77	6	6	6	6	6	8	7	8	7	7	6,7
E72	6	7	6	7	6	7	6	7	7	7	6,6
G60	6	6	6	6	6	7	7	6	9	7	6,6
E75	7	7	7	7	7	6	6	6	6	6	6,5
E76	6	6	6	7	7	7	6	7	6	5	6,3
G71	6	6	6	6	6	6	6,5	6	6	7	6,15
E71	5	5	6	5	7	6	7	6	7	6	6
E83-E88	5	6	5	6	6	6	6	6	7	7	6
Q80-Q89	5	5	5	6	6	6	6	7	7	6,5	5,95
H31-H90	5	6	5	6	6	6	6	6	6	7	5,9
G10-G12	4	5	5	6	5	6	6	7	7	6	5,7
G24-G40	5	4	4	6	6	5	5	6	6	6	5,3
E78	5	5	4	4	4	5	5	6	7	7	5,2
Q93-Q99	5	5	5	5	5	5	5	5	6	5	5,1
Q04-Q61	4	4	5	5	5	5	5	5	6	6	5
F01-F84	5	4	4	4	5	5	5	5	5	6	4,8
M87	5	6	5	4	4	4	4	4	5	7	4,8

Tabla 7.32: *Vida media por clasificación ICD y fecha*

Año	Índice de Price
2000	52,42 %
2001	49,48 %
2002	50,92 %
2003	48,89 %
2004	48,17 %
2005	46,83 %
2006	46,06 %
2007	44,53 %
2008	43,11 %
2009	42,55 %

Tabla 7.33: *Índice de Price*

7.2.3. Temática

La distribución temática de las referencias bibliográficas se ha obtenido a través de las categorías JCR de las revistas citadas (tabla 7.34). Se han tomado únicamente aquellas categorías JCR que acumulan 1.000 o más referencias en todo el periodo. Entre ellas, destacan tres áreas con más de 20.000 referencias, son las áreas de Bioquímica y Biología Molecular, Genética y Neurología Clínica, aunque también destaca el hecho de que la cuarta categoría sea la de *Multidisciplinary Sciences*.

7.2.4. Aislamiento

Con el índice de aislamiento se mide la capacidad que tienen los autores de los diferentes países para consumir literatura publicada en países diferentes al propio. Se cuantifica como el porcentaje de referencias bibliográficas publicadas en revistas de la propia nacionalidad. Así, en la tabla 7.35, se puede observar que Gran Bretaña es el país que posee un mayor índice de aislamiento con un 24.59 % de trabajos referenciados publicados en revistas nacionales, muy por encima del resto de países que se sitúan entre el 0 % y el 5,96 %.

7.2.5. Dispersión

En el gráfico 7.11 y la tabla B.8 se muestra la distribución de referencias bibliográficas por revista, cumpliéndose la distribución de Bradford, tal y como puede observarse en el gráfico.

El total de revistas citadas fue de 24693, mientras que el núcleo se encuentra formado por 29 títulos que acumulan el 33,42 % de las citas que reciben las revistas en todo el periodo. Estas revistas se muestran en la tabla 7.36, siendo en su mayoría revistas generalistas de medicina (Lancet, New England Journal of Medicine, etc.), revistas de genética (Nat Genet., American Journal of Human

Area JCR	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	total
Biochemistry & Molecular Biology	3590	5414	6062	6545	7399	7832	8145	8642	8750	9657	72236
Genetics & Heredity	2670	3464	3397	3889	4046	4425	4376	4745	4667	5010	40689
Clinical Neurology	1060	1544	1639	1970	2136	2343	2705	3252	3055	3437	23141
Multidisciplinary Sciences	992	1424	1590	1783	1961	2016	2053	2322	2294	2855	19290
Medicine, General & Internal	747	944	1183	1365	1375	1642	1735	1784	1714	1916	14405
Neurosciences	543	784	877	1164	1319	1466	1641	1889	2155	2319	14157
Endocrinology & Metabolism	454	546	1042	1011	1202	1421	1818	1840	1882	2266	13482
Pediatrics	515	759	888	1194	1040	1254	1554	1815	1751	1849	12619
Gastroenterology & Hepatology	636	814	1169	1100	1004	1096	1108	1117	1220	1269	10533
Cell Biology	389	494	641	684	803	1022	1143	1375	1402	1744	9697
Microbiology	219	466	619	839	828	1091	962	1315	1486	1821	9646
Hematology	387	491	693	809	845	1170	872	927	1098	1311	8603
Immunology	277	365	585	833	679	995	887	1033	1161	1312	8127
Cardiac & Cardiovascular Systems	290	353	427	562	716	984	907	1004	1263	1182	7688
Respiratory System	286	300	459	560	533	709	934	924	1173	1344	7222
Biotechnology & Applied Microbiology	356	382	492	548	661	587	573	779	942	880	6200
Oncology	226	230	357	448	449	670	791	737	834	852	5594
Medicine, Research & Experimental	264	305	452	448	463	594	518	611	646	696	4997
Physiology	221	274	325	417	440	582	496	528	600	712	4595
Critical Care Medicine	5	59	149	285	344	523	691	611	715	681	4063
Pharmacology & Pharmacy	88	126	200	313	312	444	542	575	634	793	4027
Biochemical Research Methods	176	189	198	289	338	422	508	460	631	768	3979
Developmental Biology	149	227	341	403	373	415	583	471	491	491	3944
Urology & Nephrology	146	173	288	346	346	419	291	441	457	448	3356
Medical Laboratory Technology	146	166	185	247	228	320	351	340	476	450	2909
Radiology; Nuclear Medicine & Medical Imaging	28	178	104	175	207	276	360	437	413	495	2673
Infectious Diseases	89	98	204	205	155	276	295	353	521	461	2657
Pathology	130	151	171	260	220	276	309	308	370	373	2568
Ophthalmology	96	119	121	181	200	238	293	346	402	520	2516
Nutrition & Dietetics	105	120	149	180	186	271	274	328	312	353	2278
Obstetrics & Gynecology	224	97	190	146	219	289	167	266	328	308	2234
Emergency Medicine & Critical Care	205	191	267	223	200	173	202	143	170	131	1905
Rheumatology	42	63	83	159	160	138	255	233	261	228	1622
Plant Sciences	50	112	108	130	131	95	168	130	168	194	1286
Surgery	33	60	67	78	114	133	152	148	208	189	1182
Allergy	91	28	110	71	95	127	193	148	192	123	1178
Peripheral Vascular Disease	35	51	82	93	113	114	138	133	125	147	1031
Virology	75	65	88	84	186	112	85	108	128	98	1029
Psychiatry	25	73	33	100	68	103	138	149	173	156	1018

Tabla 7.34: Referencias por año y temática (más de 1000 referencias)

Pais	Índice de aislamiento
GBR	24,59 %
NLD	5,96 %
ESP	2,05 %
DEU	1,97 %
FRA	1,25 %
ITA	0,71 %
IRL	0,63 %
AUT	0,40 %
DNK	0,39 %
BEL	0,12 %
GRC	0,11 %
PRT	0,02 %
SWE	0,01 %
FIN	0 %
LUX	0 %

Tabla 7.35: Índice de aislamiento por país

Genetics, etc.) y revistas multidisciplinares como Science o Nature, por lo que se corresponde con la distribución de citas por temática, tal y como se mostraba en la tabla 7.34.

Revista	Referencias
J Biol Chem	23013
P Natl Acad Sci Usa	16396
Nat Genet	15209
Am J Hum Genet	13541
Hum Mol Genet	12029
Nature	10276
New Engl J Med	9187
Science	9139
Lancet	8503
J Clin Invest	7857
Cell	7788
Neurology	7568
Am J Resp Crit Care	7154
Am J Med Genet	6461
J Med Genet	6347
J Pediatr	6199
Ann Neurol	5674
J Inherit Metab Dis	5193
Hum Genet	5157

Revista	Referencias
Hum Mutat	5052
J Clin Endocr Metab	5022
Blood	4657
Eur Respir J	4381
Gastroenterology	4343
Biochem Bioph Res Co	4259
Thorax	4109
Chest	4025
J Cell Biol	3777
Pediatr Pulm	3680

Tabla 7.36: Núcleo de revistas referenciadas

7.3. Indicadores de Impacto / Visibilidad

En cuanto a los indicadores relativos al impacto y la visibilidad, también se han calculado a partir de los datos obtenidos del SCI, puesto que otorgan información acerca del número de citas que reciben los trabajos científicos, así como el Journal Citation Report (JCR) de Thomson Reuters, que ofrece información acerca del impacto de las revistas en las que se publican dichos trabajos.

7.3.1. Impacto (Documentos en el primer cuartil)

Basándonos en la información que proporciona el JCR, se ha calculado el número de documentos en el primer cuartil por país (tabla 7.37), por clasificación ICD-10 (tabla 7.38), por tipo institucional (tabla 7.39) y por institución (tabla 7.40)

En la tabla 7.37 se muestra la evolución del número de documentos publicados en revistas del primer cuartil por país. En las posiciones más destacadas se encuentran: Reino Unido, Alemania, Francia, Italia y Holanda, con más de mil documentos publicados en el primer cuartil, mientras que el resto de países se distribuyen de forma más uniforme, entre las 200 y las 569 publicaciones en el primer

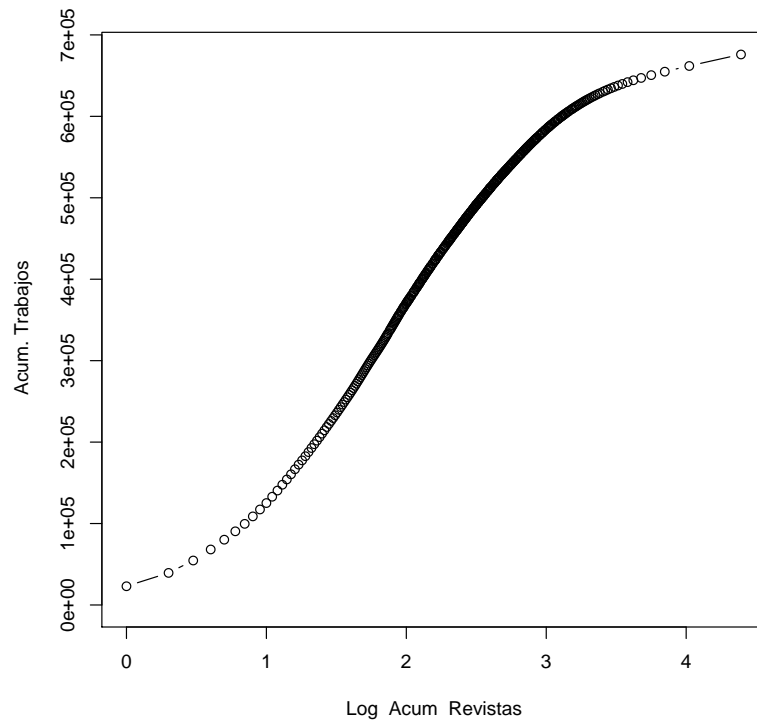


Figura 7.11: *Distribución de Bradford sobre las referencias bibliográficas*

País	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Total
GBR	272	296	272	280	306	276	266	255	274	309	2806
DEU	181	217	191	227	204	232	243	221	204	255	2175
FRA	154	184	168	182	179	191	156	190	187	219	1810
ITA	118	142	129	142	139	171	151	166	160	185	1503
NLD	90	97	98	105	114	112	122	133	116	125	1112
BEL	58	50	58	55	52	56	47	53	63	77	569
ESP	44	34	53	49	53	65	57	65	65	79	564
SWE	41	51	45	45	43	55	50	37	42	57	466
DNK	25	40	28	31	26	36	26	29	31	29	301
AUT	18	23	22	18	31	28	20	30	30	29	249
FIN	32	26	32	24	28	26	20	23	22	12	245
IRL	22	18	22	14	21	23	21	23	24	24	212
PRT	11	8	7	8	17	16	10	30	16	17	140
GRC	5	12	11	8	15	15	7	11	14	18	116
LUX	0	0	0	0	0	0	1	0	0	1	2

Tabla 7.37: *Documentos en el primer cuartil por país*

cuartil, siendo tan solo tres los países con menos de 200 trabajos publicados en revistas del primer cuartil (Portugal, Grecia y Luxemburgo).

En cuanto a la distribución de trabajos en el primer cuartil por clasificación ICD-10, es menos uniforme que en el caso de los países, puesto que en este caso el intervalo de la distribución va de los 30 trabajos del grupo M87 (osteonecrosis) a los 3978 del grupo E83-E88 (otros desórdenes metabólicos), siendo tan solo los grupos de enfermedades E83-E88, G71 y Q80-Q89 (otros desórdenes metabólicos, trastornos primarios musculares y distrofias congénitas y cromosómicas respectivamente), los que se sitúan con más de 1000 trabajos en el primer cuartil, a una distancia notable del resto de enfermedades, todas ellas con menos de 500 trabajos en revistas del primer cuartil.

En la tabla 7.39 se observa la producción en el primer cuartil por tipo de institución. Como se esperaba, la mayor producción recae en los hospitales y universidades, a pesar de lo cual destaca el total de documentos en el primer cuartil de los centros nacionales de investigación sobre el total de documentos que publican estos centros (tabla 7.9), pues pasan a ocupar el segundo puesto en el número de publicaciones en el primer cuartil, siendo el impacto de sus publicaciones mucho mayor de lo esperado (χ^2 de 117,83 y p-valor inferior a $2,2e-16$). Entre las instituciones destacan las tres que poseen más de 200 documentos en el primer cuartil (Universidad de París, Universidad Católica de Lovaina y Universidad de Oxford).

7.3.2. Análisis de citas

El cálculo de los siguientes indicadores se lleva a cabo a través del campo “TC” de los registros descargados del SCI. Este campo contiene el acumulado de citas que ha recibido un trabajo concreto hasta la fecha en la que se consulta su referencia bibliográfica, por lo tanto se trata de un análisis sincrónico, que ofrece una instantánea sobre el volumen de citas recibidas por los trabajos, en un momento

Clas. CIE	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Total
E83-E88	389	388	392	375	385	471	379	384	387	428	3978
Q80-Q89	105	103	91	121	102	125	130	130	121	121	1149
G71	106	109	120	111	110	97	105	102	113	136	1109
E75	29	30	43	49	54	58	51	48	57	66	485
G10-G12	36	24	27	32	47	36	43	41	44	33	363
E72	36	43	35	32	39	29	32	37	44	27	354
Q04-Q61	31	34	38	38	32	42	29	34	24	42	344
F01-F84	19	34	27	29	42	36	34	40	43	35	339
E78	22	34	34	40	32	34	41	26	20	20	303
H31-H90	22	30	17	28	19	24	18	31	25	29	243
E74	19	15	25	24	30	23	20	21	24	22	223
E71	22	19	14	22	24	32	20	24	20	21	218
E16-E56	15	23	28	20	17	18	15	22	18	21	197
Q93-Q99	12	15	15	12	13	19	8	15	20	26	155
E79	12	12	18	17	11	15	20	18	10	10	143
Q77-Q78	9	14	15	7	9	9	11	16	8	15	113
G60	14	17	9	11	15	9	13	7	4	13	112
E70	11	9	9	11	12	11	14	10	12	12	111
G24-G40	8	12	4	10	15	9	13	11	6	15	103
E77	10	16	10	7	12	9	8	6	5	14	97
E76	3	5	4	3	6	7	6	9	6	11	60
M87	1	2	1	1	3	6	4	1	7	4	30

Tabla 7.38: Documentos en el primer cuartil por clasificación ICD-10

Tipo institucional	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Total
HOSPITAL	572	623	631	623	652	718	663	685	626	721	6514
NATL-RES	329	371	321	370	351	431	408	406	416	489	3892
UNIV	308	318	317	319	367	397	381	409	444	508	3768
EMP-FDN	12	9	8	8	9	15	20	29	26	34	170

Tabla 7.39: Documentos en el primer cuartil por tipo institucional

Institución	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Total
Univ Paris	19	21	22	23	27	38	34	54	63	79	380
Katholieke Univ Leuven	23	21	23	28	16	17	20	26	31	34	239
Univ Oxford	18	28	30	22	29	24	14	20	21	24	230
Chrs	8	13	12	22	18	21	17	24	28	30	193
Inserm	5	7	12	13	16	24	9	31	24	48	189
Univ Amsterdam	19	15	15	23	21	22	25	18	12	16	186
Univ Cambridge	20	25	16	17	17	20	18	9	16	21	179
Univ Munich	15	18	16	13	16	27	19	24	13	16	177
Univ Milan	19	15	10	11	11	15	18	26	16	21	162
Univ London Imperial Coll Sci Technol & Med	13	16	13	21	11	11	14	8	24	25	156
Hop Necker Enfants Malad	10	19	14	17	14	17	11	14	14	19	149
Leiden Univ	5	10	12	10	16	20	20	18	16	14	141
Univ Helsinki	17	13	14	13	16	15	10	8	13	8	127
Ucl	11	11	7	9	12	14	12	14	16	18	124
Univ Tubingen	9	13	8	17	7	11	11	14	19	7	116
Inst Child Hlth	14	12	9	7	13	11	11	7	10	18	112
Univ Roma La Sapienza	3	6	7	12	10	15	15	11	7	15	101
Hannover Med Sch	4	8	5	12	10	6	11	13	8	23	100
Univ Heidelberg	3	7	13	11	7	18	9	14	8	9	99

Tabla 7.40: Documentos en el primer cuartil por institución

concreto del tiempo y en diferentes niveles de agregación (países, temáticas, tipos institucionales e instituciones).

En este caso por tanto el cálculo ha sido realizado con el acumulado de citas que han recibido los trabajos hasta la fecha de descarga de los registros (octubre de 2011).

7.3.2.1. Total de citas

En la siguiente tabla (7.41) se puede ver como se distribuye el total de citas por fecha para las enfermedades estudiadas. Siguiendo la distribución clásica, los documentos con una ventana de citación mayor reciben un mayor número de citas, salvo en el caso de los trabajos publicados en 2001, que reciben un mayor número de citas del que cabría esperar en este tipo de distribuciones.

Año	Total citas
2000	57518
2001	61546
2002	53780
2003	47510
2004	45608
2005	42124
2006	34288
2007	29171
2008	24410
2009	15647

Tabla 7.41: *Total de citas por fecha*

En el gráfico 7.12 puede observarse esta misma distribución pero desglosada por países. En este caso se encuentra claramente el mismo pico de citas en 2001, así como un valle en el número de citas en 2004 y 2005, sobre todo en algunos países concretos como Reino Unido, Francia e Italia, y en menor medida en España, Dinamarca y Austria.

En cuanto a la clasificación ICD-10, la distribución del número de citas por enfermedades codificadas se encuentra reflejada en la tabla 7.42. En este caso se

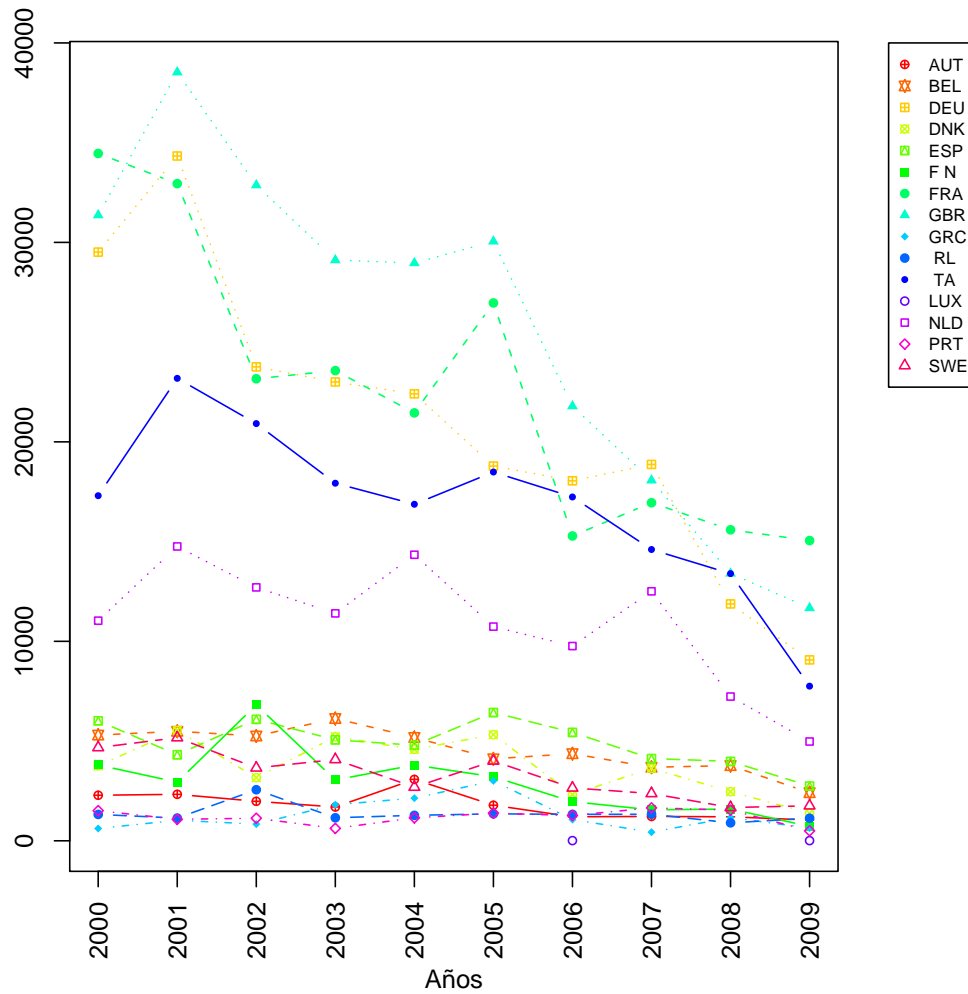


Figura 7.12: Total de citas por país

puede observar que, tal y como ocurre en la distribución de trabajos en el primer cuartil (tabla 7.38), los grupos E83-E88, G71 y Q80-Q89, relacionados con los desórdenes metabólicos, trastornos primarios musculares y distrofias congénitas y cromosómicas respectivamente, se posicionan en primer lugar en cuanto al número de citas recibidas, siendo los únicos con más de 40.000 en todo el periodo, aunque el E83-E88 destaca sobre todos los demás con 158.576 citas en todo el periodo.

De nuevo, al igual que ocurría con los indicadores de impacto, destaca la posición de los centros nacionales de investigación, que a pesar de no ocupar la primera posición en cuanto al total de citas, el total de citas que acumulan es superior a lo

que cabría esperar observando su producción global. En cuanto a las instituciones, de nuevo la Universidad de París y la de Oxford ocupan la primera y segunda posición respectivamente. Sin embargo, las universidades de Amsterdam y Cambridge, cuyo número de documentos en el primer cuartil no llamaba la atención, escalan hasta las cinco primeras posiciones en el recuento de citas.

7.3.2.2. Citas por documento

Con el fin de relativizar el total de citas que reciben los trabajos se ha procedido a dividir el número de citas entre el número de documentos publicados, obteniendo por tanto el número de citas por documento y midiendo de este modo, el impacto de las publicaciones en la comunidad científica.

En el gráfico 7.13 se puede ver la evolución de las citas por documento a lo largo de los diez años de estudio. Lo más notable de esta figura es que, como se ha anticipado en la distribución del número de citas por año, se producen dos hitos, en 2001 y en 2004, que rompen la progresión natural descendente del número de citas por documento, pues en 2001 no solo no decrece como cabría esperar sino que aumenta el número de citas por documento y en 2004 la caída es mucho menor de lo esperado.

Al trasladar esta relativización a la distribución de citas por país se observa (tabla 7.45) que el panorama cambia sustancialmente respecto al monto total de citas por país (gráfico 7.12), situándose en primer lugar dos países que se encontraban en una situación mucho más discreta en cuanto al total de citas. Estos países son Dinamarca, con 29,27 citas por documento, y Bélgica, con 25,38 citas por documento. Con respecto al número de citas por documento que recibe España, hay que señalar que ocupa el último puesto, sin contar a Luxemburgo, dada su escasa presencia.

En cuanto a la distribución de citas por documento, agregado por clasificación ICD-10 (tabla 7.46), es destacable que los grupos de enfermedades que aparecen

Clas. ICD-10	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Total
E83-E88	23747	24753	21226	17696	16573	17450	11518	10693	9000	5920	158576
Q80-Q89	6696	5857	5720	6214	4722	5251	4943	3703	3455	1641	48202
G71	6008	6196	6615	5216	4427	3639	3302	3192	2436	1753	42784
E75	1305	2269	2047	2661	2944	2336	2245	2188	1605	1227	20827
E78	2178	3844	2791	2246	1434	983	1465	751	408	356	16456
Q04-Q61	1974	2695	2205	2390	1649	1918	1289	923	541	580	16164
G10-G12	3349	1399	1607	1764	1958	1361	1412	1141	1167	418	15576
F01-F84	1968	2376	1312	1169	1840	1911	1414	1018	955	514	14477
E72	1814	2328	1779	1504	2222	776	1066	640	763	272	13164
H31-H90	1551	1599	1316	1110	779	735	649	1089	520	319	9667
E74	894	535	1079	824	1328	899	604	685	397	492	7737
E16-E56	919	889	1259	830	746	507	613	546	343	314	6966
Q93-Q99	784	1497	917	678	533	756	340	501	672	285	6963
E71	781	826	700	760	898	764	650	444	438	421	6682
E79	901	633	712	525	837	326	484	411	598	97	5524
Q77-Q78	399	827	928	363	461	401	399	265	208	239	4490
E70	637	633	284	396	401	399	291	175	192	156	3564
G60	599	635	249	377	528	186	326	143	53	138	3234
E77	554	392	542	155	399	272	291	153	99	128	2985
G24-G40	320	717	186	386	368	192	269	214	115	136	2903
E76	119	143	197	145	428	317	508	292	302	189	2640
M87	21	503	109	101	133	745	210	4	143	52	2021

Tabla 7.42: Total de citas por código ICD-10

Tipo institucional	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Total
HOSPITAL	34174	38151	31990	29202	27705	26795	21994	18642	14520	10023	253196
NATL-RES	22859	23846	19987	17936	17102	17054	14406	11951	9285	6604	161030
UNIV	19901	20521	17912	15851	16990	14304	13094	12853	9903	6704	148033
EMP-FDN	655	353	719	407	355	571	704	859	1133	419	6175

Tabla 7.43: Total de citas por tipo institucional

Institución	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Total
Univ Paris	1035	880	1388	959	813	1294	1492	1353	1458	1381	12053
Univ Oxford	903	1974	1501	1070	1011	913	409	642	432	582	9437
Univ Amsterdam	1224	1491	968	1062	835	729	1032	799	342	293	8775
Univ Cambridge	983	1575	1076	708	1241	869	864	659	445	275	8695
Katholieke Univ Leuven	1193	856	1146	1412	622	865	708	943	502	398	8645
Cnrs	1122	869	1108	976	1039	920	542	768	606	484	8434
Ucl	1200	827	779	647	1133	743	490	391	534	465	7209
Univ London Imperial Coll Sci Technol & Med	1009	1153	774	1172	495	471	552	261	410	424	6721
Univ Milan	842	938	480	823	533	514	697	1180	428	257	6692
Univ Munich	693	1519	668	522	825	611	527	821	224	191	6601
Univ Tubingen	739	818	983	787	537	696	522	576	476	89	6223
Hop Necker Enfants Malad	768	936	620	928	741	716	279	355	390	427	6160
Leiden Univ	159	666	1089	466	755	665	734	854	326	315	6029
Hannover Med Sch	1849	557	238	695	589	264	258	385	173	297	5305
Inserm	257	307	404	418	636	976	294	465	514	856	5127
Univ Helsinki	584	639	985	666	562	549	209	132	382	103	4811
Inst Child Hlth	612	833	769	296	548	557	441	225	220	250	4751
Univ Heidelberg	194	669	636	515	224	486	332	415	241	110	3822
Univ Roma La Sapienza	91	168	584	398	405	395	541	375	135	177	3269

Tabla 7.44: Total de citas por institución del núcleo

País	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Promedio
DNK	34,6	45,54	33,6	44,54	37,29	26,05	18,97	24,24	17,07	10,78	29,27
BEL	37,35	35,85	35,13	25,54	24,34	29,95	19,42	18,77	17,46	10,01	25,38
GBR	36,44	39,58	34,55	30,69	28,41	23,91	19,49	16,65	12,17	9,19	25,11
NLD	28,92	43,54	32,98	25,08	27,16	22,42	17,04	18,81	13,02	8,78	23,78
DEU	41,37	36,58	31,47	26,23	25,78	19,58	17,95	16,46	12,31	8,03	23,58
FIN	28,76	26,88	38,87	35,43	19,86	24,73	12,26	13,16	16,33	6,41	22,27
FRA	32,98	37,98	26,79	22,21	26,61	18,44	17,43	13,04	10,33	8,34	21,41
ITA	25,1	34,11	26,97	26,21	23,76	21,59	18,92	14,1	11	6,72	20,85
SWE	30,64	31,85	26,39	23,54	22,09	20,32	17,83	15,73	10,89	7,45	20,67
IRL	26,12	24,81	36,12	23,16	18,36	18,58	23,5	12,47	11,06	10,27	20,45
AUT	27,62	26,58	28,44	23,61	27,5	14,69	16,85	12,86	12,62	9,32	20,01
GRC	22,4	14,83	18,62	43,9	37,07	24,91	8,87	5,54	7,6	7	19,07
PRT	33,58	25,05	30,69	13,42	18,52	15,25	16,42	13,47	10,43	4,79	18,16
ESP	34,02	16,37	21,92	22,47	21,89	18,33	15,84	11,56	9,42	5,42	17,72
LUX	0	0	0	0	0	0	4,5	0	0	5	4,75

Tabla 7.45: Citas por documento por países

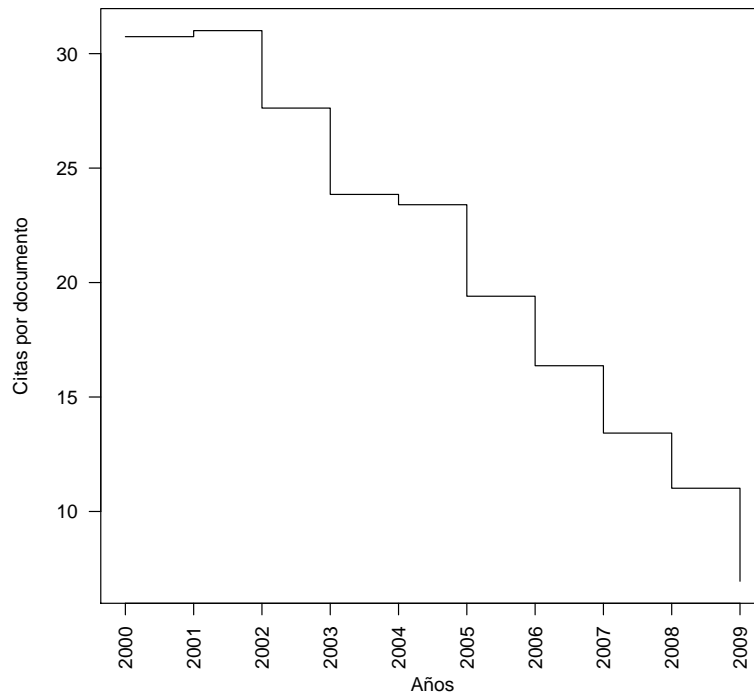


Figura 7.13: *Distribución de citas por documento*

con un mayor promedio de citas por documento no son aquellos que se identificaron como los que más citas recibían y que tenían más documentos en el primer cuartil. Estos grupos que destacaban en los indicadores anteriores se sitúan en torno a las 20 citas por documento, mientras que los grupos con un mayor número de citas por documento sobrepasan las 30 (E78 y M87).

De nuevo, al igual que en la tabla ocurría con el total de citas por tipo institucional, en la tabla 7.47, que recoge las citas por documento, llama la atención la posición de los centros nacionales de investigación, que se sitúan como los que mejor promedio de citas por documento poseen en todo el periodo, indicando un alto impacto en sus publicaciones.

Respecto a las instituciones, destaca en la primera posición la Escuela de Medicina de Hannover, con un promedio de 43,65 citas por documento. Este promedio, muy por encima del resto de las instituciones, se debe a que en el año 2000, dicha institución cuenta con un documento que ha acumulado 1635 citas y, por

Clas. ICD-10	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Promedio
E78	60,50	63,02	46,52	39,40	29,27	22,86	25,26	16,33	9,49	9,13	32,18
M87	21	125,75	36,33	16,83	19	57,31	15	2	15,89	6,50	31,56
Q04-Q61	32,90	44,92	34,45	40,51	31,11	25,24	20,79	14,65	11,76	8,66	26,50
G10-G12	54,90	26,40	31,51	32,07	29,67	23,47	22,41	17,03	15,99	7,60	26,10
G71	32,65	33,49	36,35	29,14	25,15	21,66	18,04	17,54	13,76	8,43	23,62
Q93-Q99	28	49,90	29,58	28,25	22,21	24,39	15,45	15,66	16,39	6,33	23,62
F01-F84	45,77	34,43	26,78	21,25	28,31	26,54	18,61	14,14	12,09	7,79	23,57
H31-H90	37,83	26,65	30,60	19,82	19,97	15,64	16,23	21,35	12,38	6,02	20,65
E83-E88	32,62	32,27	28,30	22,98	23,47	20,13	14,98	12,95	10,06	6,48	20,42
Q80-Q89	26,89	25,92	28,89	25,78	20,18	19,82	18,94	13,04	12,08	6,24	19,78
E75	18,91	25,49	17,96	23,14	27,01	15,27	17,68	15,41	11,72	9,30	18,19
Q77-Q78	21	30,63	33,14	15,78	24,26	18,23	15,35	8,28	6,93	7,97	18,16
G24-G40	29,09	42,18	14,31	19,30	16	14,77	9,96	10,70	8,85	5,44	17,06
E79	27,30	28,77	21,58	15	19,02	10,52	12,74	10,02	21,36	3,88	17,02
E70	22,75	37,24	12,91	23,29	18,23	21	11,19	7,29	8,35	7,09	16,93
E76	11,90	15,89	10,94	10,36	32,92	17,61	25,40	12,70	11,62	6,30	15,56
E16-E56	22,41	22,23	22,89	19,30	15,22	12,37	14,95	12,70	7,62	5,15	15,48
E77	30,78	12,65	28,53	11,07	17,35	12,95	18,19	9	5,50	4,57	15,06
E72	19,51	22,82	19,99	18,12	24,97	10,63	11,10	7,71	8,48	4,61	14,79
E71	17,36	19,67	21,88	15,51	17,96	13,17	12,75	9,25	9,52	8,96	14,60
G60	23,04	23,52	13,83	18,85	19,56	10,33	12,54	7,94	4,08	6,27	14
E74	19,02	11,38	14,39	14,46	21,42	14,05	11,40	12,02	7,09	9,46	13,47

Tabla 7.46: Citas por documento por clasificación ICD-10

tanto, aumenta notablemente el cómputo de citas por documento para dicha institución.

7.3.2.3. Documentos no citados

Para terminar con el análisis de citas se ha calculado el porcentaje de documentos no citados, es decir, con cero citas en el campo “TC” del SCI, que como se comentó anteriormente, contiene el acumulado de citas desde que se publicó el documento.

En primer lugar, la distribución del porcentaje de documentos no citados por año (tabla 7.49) muestra un ligero incremento a lo largo del tiempo en el número de documentos no citados, salvo en el caso de 2009, que cuenta con un porcentaje mayor, debido a que la ventana de citación de los documentos publicados en 2009 no ha sido suficiente como para acumular el número de citas habitual en este área.

El porcentaje de documentos no citados por país (tabla 7.50) confirma los resultados que se observaban en la distribución de citas por documento, y es que Dinamarca, que es la que tiene un mejor ratio de citas por documento, también es el país con un menor porcentaje de documentos no citados, indicándonos que su producción posee un gran impacto en el resto de la comunidad científica. También en este aspecto llama la atención la posición de España, con el mayor porcentaje (9,44 %) de documentos que no reciben ninguna cita.

En cuanto al porcentaje de documentos no citados por clasificación temática, el grupo E78, que se situaba como el que mayor citas por documento recibe, también es el que poseen un menor porcentaje de documentos sin citar (3,68 %).

Por último, el porcentaje de documentos no citados en función del tipo institucional (tabla 7.52) y las instituciones del núcleo (tabla 7.53), muestran que los resultados se corresponden con los tipos institucionales con un mayor número

Tipo institucional	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Promedio
NATL-RES	38,48	38,11	36,79	29,40	30,31	24,72	20,81	18,44	13,09	8,78	17,87
EMP-FDN	37,33	23,65	43,85	17,64	25,13	18,08	25,20	21,75	17,47	5,94	16,46
UNIV	34,55	33,28	32,11	27,44	27,76	21,47	18,93	17,68	12,70	7,97	16,28
HOSPITAL	30,92	32,85	26,59	24,62	22,23	21,43	15,98	14,48	10,94	8,10	14,47

Tabla 7.47: Citas por documento por tipo institucional

Institución	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Promedio
Hannover Med Sch	264,14	34,81	18,31	31,59	21,04	20,31	11,73	16,74	9,11	8,74	43,65
Ucl	66,67	55,13	38,95	35,94	53,95	35,38	27,22	19,55	22,25	18,6	37,37
Univ Cambridge	37,81	46,32	48,91	33,71	51,71	26,33	27,87	41,19	17,8	10,19	34,18
Chrs	66	48,28	52,76	32,53	45,17	30,67	19,36	20,76	14,43	11,26	34,12
Univ Tubingen	38,89	38,95	51,74	32,79	33,56	34,8	34,8	28,8	16,41	4,94	31,57
Univ London Imperial Coll Sci Technol & Med	63,06	50,13	33,65	39,07	29,12	31,4	27,6	16,31	12,42	11,46	31,42
Univ Oxford	41,05	58,06	37,52	41,15	28,08	28,53	20,45	22,93	14,4	15,32	30,75
Leiden Univ	22,71	41,62	54,45	25,89	35,95	21,45	22,94	30,5	10,52	12,12	27,82
Univ Amsterdam	31,38	53,25	32,27	24,7	25,3	19,7	21,06	24,21	10,69	8,62	25,12
Univ Helsinki	27,81	33,63	49,25	41,62	23,42	21,12	13,06	11	20,11	7,36	24,84
Inst Child Hlth	34	34,71	40,47	21,14	23,83	30,94	18,38	20,45	11	11,9	24,68
Katholieke Univ Leuven	30,59	22,53	38,2	32,09	20,73	32,04	22,12	23,57	11,16	7,96	24,1
Univ Paris	33,39	25,14	38,56	28,21	27,1	20,22	24,87	16,5	13,25	11,91	23,91
Univ Milan	26,31	34,74	17,78	32,92	20,5	17,13	24,03	28,78	11,89	7,56	22,16
Univ Munich	25,67	42,19	25,69	19,33	28,45	16,97	17,57	27,37	8,3	6,82	21,84
Hop Necker Enfants Malad	32	29,25	21,38	28,12	28,5	26,52	11,16	11,83	13,45	12,94	21,52
Inserm	28,56	19,19	25,25	20,9	28,91	27,89	18,38	12,24	11,42	14,03	20,68
Univ Heidelberg	16,17	41,81	26,5	25,75	16	18	15,09	16,6	10,04	6,11	19,21
Univ Roma La Sapienza	10,11	15,27	30,74	20,95	17,61	18,81	20,81	15	7,5	6,32	16,31

Tabla 7.48: Citas por documento por institución del núcleo

Año	Total	Porcentaje
2000	91	5,17 %
2001	107	5,69 %
2002	73	4,03 %
2003	117	6,22 %
2004	71	3,91 %
2005	79	3,88 %
2006	104	5,32 %
2007	147	7,20 %
2008	203	9,69 %
2009	290	13,58 %

Tabla 7.49: *Porcentaje de documentos no citados por fecha*

de citas, que ya observó en el recuento de citas por tipos institucionales, siendo los centros nacionales de investigación los que destacaban sobre el resto, con el menor porcentaje de documentos sin citas y destacando entre ellos el CNRS y el INSERM, por encontrarse entre los cinco primeros.

País	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Promedio
LUX							0 %			0 %	0 %
DNK	1,75 %	1,64 %	2,08 %	3,85 %	1,79 %	1,32 %	1,67 %	1,85 %	1,85 %	1,69 %	1,95 %
SWE	1,20 %	5,15 %	2,53 %	1,22 %	1,52 %	2 %	1,33 %	0 %	2,70 %	5,71 %	2,34 %
NLD	3,49 %	0 %	1,09 %	0,54 %	1,73 %	1,04 %	0,84 %	3,52 %	6,16 %	8,22 %	2,66 %
FIN	0 %	7,14 %	0 %	0 %	0 %	8,33 %	2,56 %	0 %	9,09 %	6,25 %	3,34 %
BEL	0 %	1,08 %	2,15 %	7,34 %	1,94 %	0 %	2,33 %	7,84 %	5,36 %	6,30 %	3,43 %
GBR	2,55 %	3,09 %	2,90 %	1,77 %	2,60 %	2,35 %	4,69 %	7,81 %	6,09 %	9,37 %	4,32 %
PRT	5,26 %	0 %	0 %	0 %	3,45 %	0 %	3,23 %	2,13 %	7,14 %	25,58 %	4,68 %
ITA	5,81 %	5 %	2,71 %	2,46 %	2,54 %	2,47 %	3,62 %	5,54 %	10,83 %	10,16 %	5,11 %
IRL	0 %	0 %	4,88 %	0 %	6,82 %	4,44 %	5 %	7,84 %	12 %	14,55 %	5,55 %
AUT	2,22 %	3,77 %	2,22 %	10,20 %	6,67 %	7,46 %	8,51 %	7,02 %	4,17 %	8 %	6,02 %
DEU	7,05 %	5,32 %	4,16 %	5,12 %	4,87 %	4,19 %	5,73 %	5,23 %	9,90 %	12,70 %	6,43 %
GRC	13,33 %	2,86 %	0 %	9,52 %	0 %	4,55 %	13,16 %	10,26 %	7,02 %	13,51 %	7,42 %
FRA	5,52 %	8,84 %	5,06 %	14,73 %	5,88 %	7,03 %	6,38 %	8,83 %	9,05 %	16,29 %	8,76 %
ESP	9,52 %	11,67 %	8,40 %	7,63 %	5,04 %	2,17 %	6,67 %	8,33 %	15,85 %	19,08 %	9,44 %

Tabla 7.50: Porcentaje de documentos no citados por año

Clas. ICD-10	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Promedio
E78	0%	3,28%	5%	3,51%	0%	0%	1,72%	6,52%	11,63%	5,13%	3,68%
G24-G40	0%	0%	0%	0%	8,70%	0%	3,70%	0%	7,69%	20%	4,01%
G10-G12	3,28%	5,66%	3,92%	1,82%	1,52%	3,45%	6,35%	2,99%	5,48%	7,27%	4,17%
E72	3,23%	3,92%	2,25%	1,20%	2,25%	2,74%	6,25%	3,61%	10%	10,17%	4,56%
E16-E56	0%	5%	1,82%	4,65%	2,04%	2,44%	0%	6,98%	8,89%	14,75%	4,66%
G71	3,80%	2,70%	2,20%	5,03%	5,11%	0,60%	4,37%	3,30%	6,21%	13,46%	4,68%
Q93-Q99	7,14%	6,67%	0%	8,33%	0%	3,23%	0%	3,12%	14,63%	4,44%	4,76%
Q04-Q61	8,33%	3,33%	0%	6,78%	1,89%	5,26%	4,84%	3,17%	4,35%	13,43%	5,14%
H31-H90	2,44%	6,67%	4,65%	5,36%	2,56%	4,26%	5%	1,96%	4,76%	15,09%	5,28%
F01-F84	4,65%	17,39%	4,08%	5,45%	3,08%	0%	3,95%	6,94%	5,06%	6,06%	5,67%
M87	0%	0%	0%	0%	0%	0%	0%	50%	11,11%	0%	6,11%
G60	0%	3,70%	0%	0%	0%	0%	11,54%	11,11%	7,69%	27,27%	6,13%
E76	10%	0%	5,56%	0%	0%	5,56%	5%	17,39%	11,54%	6,67%	6,17%
E70	0%	5,88%	0%	0%	4,55%	5,26%	3,85%	4,17%	13,04%	31,82%	6,86%
E75	8,70%	3,37%	6,14%	6,09%	2,75%	3,27%	6,30%	12,68%	9,49%	9,85%	6,86%
E83-E88	6,59%	6,26%	4,80%	8,70%	4,67%	4,15%	6,11%	8,11%	8,83%	12,58%	7,08%
Q80-Q89	3,61%	5,75%	5,05%	4,15%	4,27%	6,04%	3,83%	6,69%	13,29%	19,77%	7,25%
E71	8,89%	7,14%	3,12%	10,20%	2%	3,45%	1,96%	14,58%	8,70%	14,89%	7,49%
E74	12,77%	4,26%	6,67%	7,02%	4,84%	4,69%	3,77%	10,53%	10,71%	9,62%	7,49%
Q77-Q78	0%	3,70%	7,14%	0%	0%	9,09%	19,23%	9,38%	13,33%	16,67%	7,85%
E79	6,06%	0%	0%	5,71%	9,09%	9,68%	2,63%	2,44%	35,71%	8%	7,93%
E77	5,56%	6,45%	0%	0%	0%	9,52%	0%	5,88%	16,67%	39,29%	8,34%

Tabla 7.51: Porcentaje de documentos no citados por clasificación ICD-10

Tipo institucional	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Promedio
NATL-RES	2,90 %	2,20 %	2,03 %	2,53 %	2,44 %	2,36 %	2,95 %	2,70 %	5,44 %	9,50 %	3,51 %
UNIV	3,94 %	3,81 %	2,20 %	2,69 %	2,08 %	2,22 %	4,29 %	4,27 %	6,21 %	9,38 %	4,11 %
EMP-FDN	6,25 %	6,67 %	7,14 %	13,64 %	0,00 %	0,00 %	3,70 %	4,76 %	6,45 %	18,18 %	6,68 %
HOSPITAL	5,00 %	6,72 %	4,41 %	7,40 %	4,92 %	4,27 %	5,45 %	8,16 %	11,10 %	14,50 %	7,19 %

Tabla 7.52: *Porcentaje de documentos no citados por tipo institucional*

Institución	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Promedio
CNRS	0%	0%	0%	0%	0%	3,33%	0%	0%	0%	4,65%	0,80
Univ Cambridge	0%	2,94%	0%	0%	0%	0%	0%	6,25%	4%	0%	1,32%
Univ Oxford	0%	0%	2,50%	0%	5,56%	0%	0%	0%	0%	5,26%	1,33%
UCL	0%	0%	5%	0%	0%	0%	0%	10%	0%	4%	1,90%
INSEERM	0%	0%	0%	10%	0%	2,86%	0%	2,63%	2,22%	8,20%	2,59%
Leiden Univ	0%	0%	0%	0%	0%	3,23%	3,12%	0%	12,90%	7,69%	2,69%
Univ London											
Imperial Coll											
Sci Technol & Med	0%	0%	0%	3,33%	0%	0%	5%	6,25%	6,06%	8,11%	2,88%
Katholieke Univ Leuven	0%	0%	6,67%	2,27%	3,33%	0%	3,12%	2,50%	4,44%	8%	3,03%
Univ Amsterdam	5,13%	0%	3,33%	2,33%	0%	0%	2,04%	3,03%	9,38%	5,88%	3,11%
Univ Milan	0%	3,70%	0%	0%	0%	0%	0%	4,88%	8,33%	14,71%	3,16%
Univ Paris	3,23%	8,57%	0%	2,94%	0%	1,56%	3,33%	2,44%	5,45%	7,76%	3,53%
Inst Child Hlth	5,56%	0%	5,26%	7,14%	8,70%	0%	4,17%	0%	5%	4,76%	4,06%
Univ Munich	7,41%	8,33%	3,85%	0%	3,45%	2,78%	6,67%	0%	7,41%	10,71%	5,06%
Univ Roma La Sapienza	0%	9,09%	0%	5,26%	0%	4,76%	7,69%	4%	11,11%	10,71%	5,26%
Royal Brompton Hosp	0%	9,09%	0%	0%	4,35%	0%	11,11%	21,43%	0%	13,04%	5,90%
Hannover Med Sch	0%	6,25%	0%	0%	7,14%	0%	13,64%	8,70%	15,79%	11,76%	6,33%
Univ Heidelberg	16,67%	0%	4,17%	15%	0%	0%	4,55%	4%	16,67%	11,11%	7,22%
Hop Necker Enfants Malad	8,33%	12,50%	13,79%	15,15%	0%	14,81%	8%	13,33%	6,90%	9,09%	10,19%

Tabla 7.53: Porcentaje de documentos no citados por institución del núcleo

7.3.3. Ranking de revistas

Para concluir con los indicadores relativos al impacto de las publicaciones, se han elaborado dos rankings de las revistas del núcleo en función del factor de impacto (tabla 7.54) y del cuartil corregido (tabla 7.55) que ocupan en las diversas áreas JCR en las que se encuentran clasificadas. Para calcular este último indicador se ha tenido en cuenta, exclusivamente, el área JCR en la que la revista a obtenido el cuartil más alto en todo el periodo de estudio, con el fin de comparar entre ellas siempre la mejor posición posible.

En el caso de la evolución del Factor de impacto por año, llaman la atención dos revistas, en primer lugar la *American Journal Of Medical Genetics*, cuyo cambio de nombre ocasiona que en 2003 deje de tener Factor de impacto, para recuperarlo posteriormente con su nuevo nombre (*American Journal Of Medical Genetics Part A*), a pesar de lo cual no vuelve a obtener un Factor de impacto equivalente hasta tres años después.

Por otro lado, el análisis de los cuartiles en los que se encuentran las diferentes revistas que componen el núcleo (tabla 7.55) muestra discrepancias entre los cuartiles en función, tanto del idioma como del lugar de publicación de las revistas. Cuando se lleva a cabo un contraste de independencia entre las variables idioma y cuartil se obtiene un χ^2 de 46, con un p-valor de 2,961e-08, y cuando se efectúa ese mismo contraste sobre las variables cuartil y país de publicación de la revista se obtiene un valor de 60,43 en el estadístico y un p-valor de 1,09e-05. Por lo tanto, en ambos casos, se rechaza la hipótesis nula de independencia entre las variables.

Revista	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	FI Medio
American Journal Of Human Genetics	10,35	10,54	10,65	11,6	12,34	12,65	12,63	11,09	10,15	12,3	11,43
Proceedings Of The National Academy Of Sciences Of The United States Of America	10,79	10,9	10,7	10,27	10,45	10,23	9,64	9,6	9,38	9,43	10,14
Blood	8,98	9,27	9,63	10,12	9,78	10,13	10,37	10,9	10,43	10,55	10,02
Annals Of Neurology	8,48	8,48	8,6	7,72	8,1	7,57	8,05	8,81	9,94	9,32	8,51
American Journal Of Respiratory And Critical Care Medicine	5,44	5,96	6,57	8,88	8,12	8,69	9,09	9,07	9,79	10,69	8,23
Human Molecular Genetics	9,05	9,32	8,73	8,6	7,8	7,76	8,1	7,81	7,25	7,39	8,18
Brain	7,3	7,41	7,12	7,97	8,2	7,54	7,62	8,57	9,6	9,49	8,08
Human Mutation	3,67	6,13	6,89	6,33	6,84	7,92	6,47	6,27	7,03	6,89	6,44
Journal Of Biological Chemistry	7,37	7,26	6,7	6,48	6,36	5,85	5,81	5,58	5,52	5,33	6,23
Neurology	4,78	5,21	5,34	5,68	5,97	4,95	5,69	6,01	7,04	8,17	5,88
Journal Of Clinical Endocrinology & Metabolism	5,45	5,16	5,2	5,87	5,78	6,02	5,8	5,49	6,33	6,2	5,73
Clinical Chemistry	4,26	4,37	4,79	5,54	6,5	7,72	5,45	4,8	5,58	6,26	5,53
Thorax	3,98	4,09	4,08	4,19	5,04	6,15	6,06	6,23	7,07	7,04	5,39
Journal Of Medical Genetics	3,29	5,1	7,77		4,11	4,33	5,09	5,54	5,71	5,75	5,19
Archives Of Neurology	4,39	4,05	4,34	4,68	4,83	4,9	5,2	5,78	5,87	6,31	5,04
Antimicrobial Agents And Chemotherapy	3,95	4,56	4,21	4,25	4,22	4,38	4,15	4,39	4,72	4,8	4,36
Biochemical Journal	4,28	4,33	4,59	4,1	4,28	4,22	4,1	4,01	4,37	5,16	4,34
Journal Of Lipid Research	3,7	4,14	3,73	3,89	4,16	3,91	4,36	4,34	4,41	4,92	4,16
European Respiratory Journal	2,59	2,99	2,93	3	3,1	3,95	5,08	5,35	5,54	5,53	4,01
Human Genetics	3,42	3,21	3,43	4,02	4,33	4,33	3,66	3,97	4,04	4,52	3,89

Continúa en la página siguiente...

Revista	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	FI Medio
Chest	2,45	2,48	2,97	3,26	3,12	4,01	3,92	4,14	5,15	6,36	3,79
Journal Of Clinical Microbiology	3,5	3,96	3,56	3,49	3,44	3,54	3,44	3,71	3,94	4,16	3,67
Journal Of Pediatrics	3,47	3,54	3,22	2,91	3,12	3,84	3,99	4,02	4,12	4,09	3,63
Febs Letters	3,44	3,64	3,91	3,61	3,84	3,42	3,37	3,26	3,26	3,54	3,53
European Journal Of Human Genetics	3,17	3,17	3,14	3,67	2,74	3,25	3,7	4	3,92	3,56	3,43
Pediatric Research	2,79	3,29	3,38	3,06	2,88		2,62	2,84	2,6	2,61	2,90
Biochemical And Biophysical Research Communications	3,06	2,95	2,94	2,84	2,9	3	2,85	2,75	2,65	2,55	2,85
Neuromuscular Disorders	2,72	2,55	2,59	2,89	3,04	3,34	2,62	2,67	2,93	2,98	2,83
Clinical Genetics	1,64	2,4	2,24	2,02	2,37	3,28	3,14	3,18	3,21	3,3	2,68
Molecular Genetics And Metabolism	2,36	2,35	2,48	2,04	2,5	2,68	2,37	2,55	2,63	2,9	2,49
Blood Cells Molecules And Diseases	2,9	1,7	1,77	1,99	2,55	2,43	2,68	2,56	2,75	2,9	2,42
American Journal Of Medical Genetics	2,48	2,38	2,33	0							2,39
Archives Of Disease In Childhood	1,87	2,13	2,1	1,72	1,66	1,79	2,09	2,79	2,83	2,66	2,16
American Journal Of Medical Genetics Part A					0,81	1,91	2,06	2,44	2,56	2,4	2,03
Journal Of Inherited Metabolic Disease	1,31	1,79	1,62	1,8	1,59	1,72	1,57	1,67	2,69	3,6	1,94
Journal Of Cystic Fibrosis											
Pediatric Pulmonology	1,54	1,74	1,74	1,92	1,66	1,59	1,97	2,27	1,55	2,24	1,90
Respiratory Medicine	1,25	1,38	1,5	1,42	1,51	1,66	2,09	2,23	1,88	1,82	1,81
Prenatal Diagnosis	1,65	1,79	1,8	1,48	1,61	1,64	1,51	1,32	2,34	2,33	1,77
Hormone Research	1,3	1,12	1,23	1,59	1,56	1,39	1,39	2,02	1,6	1,71	1,61
Acta Paediatrica	1,31	1,58	1,26	1,13	1,14	1,28	1,3	1,41	2,29	1,73	1,56

Continúa en la página siguiente...

Revista	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	FI Medio
Neuropediatrics	1,6	1,48	0,96	1,38	1,62	1,38	1,37	1,23	1,22		1,36
European Journal Of Pediatrics	1,32	1,22	1,22	1,16	1,37	1,38	1,14	1,28	1,42	1,63	1,31
Revista De Neurologia	0,26	0,26	0,29	0,2	0,21	0,39	0,53	0,74	1,08	1,23	0,52
Revue Des Maladies Respiratoires	0,19	0,26	0,3	0,66	1,16	0,57	0,44	0,59	0,37	0,53	0,51
Archives De Pediatrie	0,3	0,4	0,38	0,24	0,31	0,27	0,26	0,3	0,36	0,41	0,32

Tabla 7.54: Factor de impacto de las revistas del núcleo

Revista	Idioma	País	Mejor cuartil
Journal Of Clinical Endocrinology & Metabolism	English	Usa	1Q
American Journal Of Human Genetics	English	Usa	1Q
American Journal Of Respiratory And Critical Care Medicine	English	Usa	1Q
Annals Of Neurology	English	Usa	1Q
Antimicrobial Agents And Chemotherapy	English	Usa	1Q
Archives Of Disease In Childhood	English	England	1Q
Archives Of Neurology	English	Usa	1Q
Blood	English	Usa	1Q
Brain	English	England	1Q
Chest	English	Usa	1Q
Clinical Chemistry	English	Usa	1Q
European Respiratory Journal	English	England	1Q
Febs Letters	English	Netherlands	1Q
Human Molecular Genetics	English	England	1Q
Human Mutation	English	Usa	1Q
Journal Of Biological Chemistry	English	Usa	1Q
Journal Of Medical Genetics	English	England	1Q
Journal Of Pediatrics	English	Usa	1Q
Neurology	English	Usa	1Q
Neuromuscular Disorders	English	England	1Q
Pediatric Pulmonology	English	Usa	1Q
Pediatric Research	English	Usa	1Q
Thorax	English	England	1Q
Acta Paediatrica	English	Norway	2Q
American Journal Of Medical Genetics	English	Usa	2Q
Biochemical And Biophysical Research Communications	English	Usa	2Q
Biochemical Journal	English	England	2Q
Clinical Genetics	English	England	2Q
European Journal Of Human Genetics	English	England	2Q
European Journal Of Pediatrics	English	Usa	2Q
Human Genetics	English	Usa	2Q
Journal Of Clinical Microbiology	English	Usa	2Q
Journal Of Inherited Metabolic Disease	English	Netherlands	2Q
Journal Of Lipid Research	English	Usa	2Q
Molecular Genetics And Metabolism	English	Usa	2Q
Neuropediatrics	English	Germany	2Q
Prenatal Diagnosis	English	England	2Q

Continúa en la página siguiente...

Revista	Idioma	País	Mejor cuartil
Proceedings Of The National Academy Of Sciences Of The United States Of America	English	Usa	2Q
American Journal Of Medical Genetics Part A	English	Usa	3Q
Blood Cells Molecules And Diseases	English	Usa	3Q
Hormone Research	English	Switzerland	3Q
Journal Of Cystic Fibrosis	English	Netherlands	3Q
Respiratory Medicine	English	England	3Q
Archives De Pediatrie	French	France	4Q
Revista De Neurologia	Spanish	Spain	4Q
Revue Des Maladies Respiratoires	French	France	4Q

Tabla 7.55: Cuartil corregido de las revistas del núcleo

Sección 8

Análisis de regresión

8.1. Producción por país y porcentaje del gasto sanitario respecto al PIB

Para la elaboración de este análisis de regresión se cuenta con las siguientes variables:

País	Año	Producción SCI	Gasto sanitario (% sobre el PIB)
AUT	2000	58	9,95 %
AUT	2001	68	10,09 %
AUT	2002	67	10,11 %
AUT	2003	66	10,29 %
AUT	2004	87	10,41 %
AUT	2005	92	10,40 %
AUT	2006	69	10,26 %
AUT	2007	84	10,27 %
AUT	2008	82	10,44 %
AUT	2009	84	11,08 %
BEL	2000	118	8,98 %
BEL	2001	109	9,11 %
BEL	2002	113	9,31 %
BEL	2003	135	10,11 %
BEL	2004	123	10,50 %

Continúa en la página siguiente...

País	Año	Producción SCI	Gasto sanitario (% sobre el PIB)
BEL	2005	108	10,27 %
BEL	2006	115	9,92 %
BEL	2007	143	10,78 %
BEL	2008	155	11,12 %
BEL	2009	186	11,80 %
DEU	2000	460	10,29 %
DEU	2001	530	10,44 %
DEU	2002	502	10,64 %
DEU	2003	522	10,80 %
DEU	2004	514	10,58 %
DEU	2005	562	10,69 %
DEU	2006	614	10,54 %
DEU	2007	609	10,42 %
DEU	2008	588	10,61 %
DEU	2009	666	11,37 %
DNK	2000	69	8,26 %
DNK	2001	72	8,56 %
DNK	2002	61	8,76 %
DNK	2003	69	9,30 %
DNK	2004	70	9,45 %
DNK	2005	93	9,49 %
DNK	2006	70	9,60 %
DNK	2007	73	9,74 %
DNK	2008	76	9,92 %
DNK	2009	76	11,19 %
ESP	2000	142	7,21 %
ESP	2001	146	7,24 %
ESP	2002	170	7,25 %
ESP	2003	147	8,15 %
ESP	2004	152	8,19 %
ESP	2005	185	8,28 %
ESP	2006	191	8,36 %
ESP	2007	212	8,44 %
ESP	2008	242	8,97 %
ESP	2009	256	9,67 %
FIN	2000	65	7,23 %
FIN	2001	48	7,42 %
FIN	2002	58	7,79 %
FIN	2003	46	8,16 %
FIN	2004	52	8,24 %
FIN	2005	51	8,44 %
FIN	2006	49	8,39 %

Continúa en la página siguiente...

País	Año	Producción SCI	Gasto sanitario (% sobre el PIB)
FIN	2007	46	8,18 %
FIN	2008	46	8,79 %
FIN	2009	43	9,69 %
FRA	2000	430	10,07 %
FRA	2001	446	10,20 %
FRA	2002	434	10,52 %
FRA	2003	490	10,89 %
FRA	2004	398	11,01 %
FRA	2005	492	11,10 %
FRA	2006	420	11,09 %
FRA	2007	516	11,03 %
FRA	2008	546	11,15 %
FRA	2009	535	11,72 %
GBR	2000	612	7,04 %
GBR	2001	675	7,27 %
GBR	2002	649	7,59 %
GBR	2003	653	7,79 %
GBR	2004	679	8,02 %
GBR	2005	713	8,25 %
GBR	2006	695	8,45 %
GBR	2007	768	8,40 %
GBR	2008	792	8,68 %
GBR	2009	813	9,36 %
GRC	2000	16	7,88 %
GRC	2001	38	8,80 %
GRC	2002	32	9,10 %
GRC	2003	29	8,93 %
GRC	2004	41	8,68 %
GRC	2005	67	9,57 %
GRC	2006	53	9,69 %
GRC	2007	58	9,71 %
GRC	2008	90	10,10 %
GRC	2009	66	10,62 %
IRL	2000	44	6,13 %
IRL	2001	37	6,76 %
IRL	2002	52	7,09 %
IRL	2003	43	7,40 %
IRL	2004	61	7,61 %
IRL	2005	56	7,53 %
IRL	2006	57	7,47 %
IRL	2007	87	7,54 %
IRL	2008	71	8,74 %

Continúa en la página siguiente...

País	Año	Producción SCI	Gasto sanitario (% sobre el PIB)
IRL	2009	87	9,74 %
ITA	2000	312	8,06 %
ITA	2001	333	8,21 %
ITA	2002	337	8,34 %
ITA	2003	370	8,35 %
ITA	2004	355	8,65 %
ITA	2005	415	8,59 %
ITA	2006	412	8,67 %
ITA	2007	517	8,37 %
ITA	2008	523	8,71 %
ITA	2009	536	9,51 %
LUX	2002	1	8,26 %
LUX	2006	2	7,75 %
LUX	2009	1	7,78 %
NLD	2000	208	7,96 %
NLD	2001	195	8,30 %
NLD	2002	237	8,87 %
NLD	2003	234	9,77 %
NLD	2004	232	9,97 %
NLD	2005	261	9,83 %
NLD	2006	308	9,72 %
NLD	2007	302	9,68 %
NLD	2008	291	9,86 %
NLD	2009	318	10,78 %
PRT	2000	21	8,52 %
PRT	2001	29	8,50 %
PRT	2002	25	8,69 %
PRT	2003	32	9,40 %
PRT	2004	33	9,66 %
PRT	2005	43	9,86 %
PRT	2006	44	9,63 %
PRT	2007	77	10,09 %
PRT	2008	80	10,22 %
PRT	2009	64	10,98 %
SWE	2000	92	8,18 %
SWE	2001	107	8,89 %
SWE	2002	96	9,21 %
SWE	2003	98	9,30 %
SWE	2004	82	9,07 %
SWE	2005	114	9,04 %
SWE	2006	93	8,93 %
SWE	2007	99	8,92 %

Continúa en la página siguiente...

País	Año	Producción SCI	Gasto sanitario (% sobre el PIB)
SWE	2008	94	9,22 %
SWE	2009	125	9,77 %

Tabla 8.1: Variables para el estudio de regresión entre producción y gasto sanitario

Es necesario destacar que la producción científica utilizada para este estudio ha sido la del SCI, debido a que esta base de datos es la que mejor representa la producción por países debido a que recoge todas las instituciones firmantes en cada trabajo.

En primer lugar se ajustó un modelo de regresión simple por el método "stepwise" de forma que el sistema aportara la información necesaria sobre las variables que influyen significativamente en el modelo, con el fin de determinar si el país es una variable significativa o no.

Start: AIC=1539.06

prod ~ 1

	Df	Sum of Sq	RSS	AIC
+ dpais	14	6442287	217078	1077
+ gasto	1	197337	6462028	1537
<none>			6659365	1539

Step: AIC=1077.5

prod ~ dpais

	Df	Sum of Sq	RSS	AIC
+ gasto	1	59077	158001	1034
<none>			217078	1077
- dpais	14	6442287	6659365	1539

Step: AIC=1034.07

```
prod ~ dpais + gasto
```

	Df	Sum of Sq	RSS	AIC
+ gasto:dpais	14	58238	99763	996
<none>			158001	1034
- gasto	1	59077	217078	1077
- dpais	14	6304027	6462028	1537

```
Step: AIC=996.32
```

```
prod ~ dpais + gasto + dpais:gasto
```

	Df	Sum of Sq	RSS	AIC
<none>			99763	996
- dpais:gasto	14	58238	158001	1034

```
Call:
```

```
lm(formula = prod ~ dpais + gasto + dpais:gasto)
```

Como se puede observar en el output de R, el mejor modelo dentro de los posibles sería el siguiente:

$$Prod = \beta_0 + Dpais + Gasto + (DpaisGasto) + \varepsilon \quad (8.1)$$

Dónde:

Prod = Número de artículos producidos.

Dpais = País productor de los documentos (tratado como variable “dummy”).

Gasto = Porcentaje del gasto en salud sobre el PIB del país.

β_0 = Coeficiente de regresión estimado.

ε = Residuos del modelo.

Aplicando este modelo, que contiene las variables país, gasto y la interacción entre ellas, se obtienen los siguientes resultados:

Residuals:

Min	1Q	Median	3Q	Max
-80,5821	-9,3431	-0,7713	8,0158	131,8525

Coefficients:

	Estimate	Std, Error	t value	Pr(> t)
(Intercept)	-1,661e+02	3,338e+02	-0,498	0,61974
dpaisBEL	5,814e+01	3,520e+02	0,165	0,86910
dpaisDEU	-6,040e+02	4,884e+02	-1,237	0,21878
dpaisDNK	2,055e+02	3,528e+02	0,583	0,56127
dpaisESP	-1,468e+01	3,492e+02	-0,042	0,96654
dpaisFIN	2,707e+02	3,538e+02	0,765	0,44570
dpaisFRA	-1,278e+01	3,999e+02	-0,032	0,97455
dpaisGBR	1,741e+02	3,533e+02	0,493	0,62306
dpaisGRC	-6,671e+00	3,537e+02	-0,019	0,98499
dpaisIRL	1,253e+02	3,422e+02	0,366	0,71500
dpaisITA	-6,782e+02	3,951e+02	-1,717	0,08878 ,
dpaisLUX	1,761e+02	6,713e+02	0,262	0,79352
dpaisNLD	4,902e+01	3,518e+02	0,139	0,88942
dpaisPRT	3,281e+00	3,540e+02	09	0,99262
dpaisSWE	1,277e+02	4,017e+02	0,318	0,75109
gasto	2,341e+01	3,230e+01	0,725	0,47014
dpaisBEL:gasto	-6,108e-03	3,410e+01	-00179	0,99986
dpaisDEU:gasto	1,013e+02	4,654e+01	2,177	0,03156 *
dpaisDNK:gasto	-1,986e+01	3,448e+01	-0,576	0,56579
dpaisESP:gasto	2,124e+01	3,463e+01	0,613	0,54084
dpaisFIN:gasto	-30e+01	3,528e+01	-0,850	0,39704
dpaisFRA:gasto	3,631e+01	3,811e+01	0,953	0,34272
dpaisGBR:gasto	6,279e+01	3,531e+01	1,778	0,07808 ,
dpaisGRC:gasto	4,180e-01	3,465e+01	0,012	0,99040

```

dpaisIRL:gasto -1,021e+01  3,377e+01  -0,302  0,76297
dpaisITA:gasto  1,235e+02  4,067e+01   3,036  0297 **
dpaisLUX:gasto -2,450e+01  8,021e+01  -0,306  0,76054
dpaisNLD:gasto  1,625e+01  3,435e+01   0,473  0,63716
dpaisPRT:gasto -1,679e+00  3,456e+01  -0,049  0,96134
dpaisSWE:gasto -8,125e+00  4,065e+01  -0,200  0,84192

```

```

Signif, codes:  0 '***' 0,001 '**'
0,01 '*' 0,05 '.' 0,1 ' ' 1

```

Residual standard error: 29,71 on 113 degrees of freedom

Multiple R-squared: 0,985, Adjusted R-squared: 0,9812

F-statistic: 256,2 on 29 and 113 DF, p-value: < 2,2e-16

Aparentemente el modelo ajusta correctamente con un coeficiente de determinación ajustado de 0,981, pero al realizar la diagnosis y comprobar la hipótesis de independencia con el estadístico de Durbin-Watson se obtiene el siguiente resultado:

```

lag Autocorrelation D-W Statistic p-value
  1      0,2694504      1,458345      0
Alternative hypothesis: rho != 0

```

Como se puede ver el valor del estadístico Durbin-Watson está cercano al rango de aceptación, que tradicionalmente se sitúa entre 1,5 y 2,5, sin embargo se debe rechazar la hipótesis de independencia. Este rechazo se produce debido a que los datos tienen una componente temporal, que no ha sido tomada en cuenta en el modelo, puesto que hubiera sido necesario un análisis de series temporales que, por otro lado, no es necesario para comprobar si existe correlación entre las variables estudiadas.

Sin embargo, corregir la influencia del factor tiempo en el modelo de linealidad es importante debido a que, de otra forma, los valores de los estimadores no

serían del todo correctos. Para corregir el modelo se ha utilizado el método de los mínimos cuadrados generalizados, a través del paquete "dynlm" en R. Este método puede aplicarse gracias a que la serie temporal del modelo ajustado es un modelo autorregresivo AR2, tal y como se puede observar en los gráficos 8.1, que presentan las funciones de autocorrelación simples (FAS) y parciales (FAP).

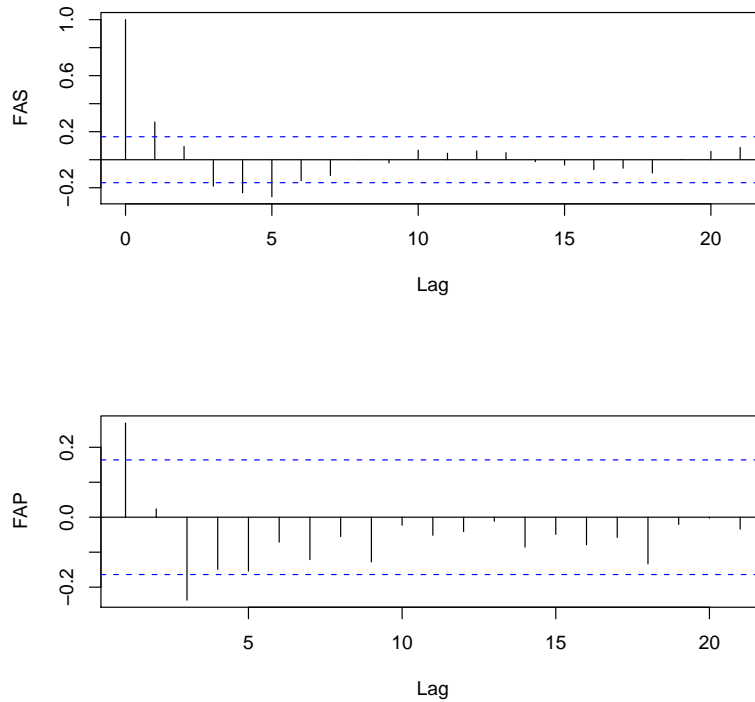


Figura 8.1: FAS y FAP del modelo ajustado

Por tanto, una vez incorporada la corrección sobre el modelo lineal múltiple original, se obtiene el siguiente modelo:

$$Prod = \underset{(61,333)}{\beta_0} + Dpais + 2,612TGasto + (DpaisTGasto) + \varepsilon \quad (8.2)$$

Dónde:

Prod = Número de artículos producidos.

Dpais = País productor de los documentos (tratado como variable "dummy").

TGasto = Porcentaje del gasto en salud sobre el PIB del país con las correcciones aplicadas para evitar la autocorrelación.

β_0 = Coeficiente de regresión estimado.

ε = Residuos del modelo

Que arroja los siguientes resultados de regresión:

Residuals:

Min	1Q	Median	3Q	Max
-68,3273	-6,7939	-0,7091	11,9455	41,8606

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	61,333	130	4,718	6,87e-06 ***
dpaisBEL	-25,618	35,494	-0,722	0,471931
dpaisDEU	13,339	55,313	0,241	0,809868
dpaisDNK	-26,515	75,744	-0,350	0,726945
dpaisESP	-448,679	96,399	-4,654	8,91e-06 ***
dpaisFIN	77,867	117,159	0,665	0,507647
dpaisFRA	-360,358	137,977	-2,612	0,010232 *
dpaisGBR	-904,412	158,830	-5,694	9,95e-08 ***
dpaisGRC	-552,279	179,706	-3,073	02654 **
dpaisIRL	-486,279	200,598	-2,424	0,016930 *
dpaisITA	-2484,127	221,501	-11,215	< 2e-16 ***
dpaisLUX	-600	1507,200	-0,040	0,968316
dpaisNLD	-1410,024	248,687	-5,670	1,11e-07 ***
dpaisPRT	-834,261	269,606	-3,094	02486 **
dpaisSWE	-161,109	290,530	-0,555	0,580309
trend(gasto)	2,612	2,095	1,247	0,215066
dpaisBEL:trend(gasto)	3,503	2,963	1,182	0,239581
dpaisDEU:trend(gasto)	16,291	2,963	5,498	2,41e-07 ***
dpaisDNK:trend(gasto)	-1,539	2,963	-0,520	0,604398

dpaisESP:trend(gasto)	9,952	2,963	3,359	01068	**
dpaisFIN:trend(gasto)	-4,212	2,963	-1,422	0,157901	
dpaisFRA:trend(gasto)	9,139	2,963	3,085	02563	**
dpaisGBR:trend(gasto)	17,891	2,963	6,038	2,03e-08	***
dpaisGRC:trend(gasto)	3,703	2,963	1,250	0,213967	
dpaisIRL:trend(gasto)	2,461	2,963	0,830	0,408035	
dpaisITA:trend(gasto)	24,248	2,963	8,184	4,60e-13	***
dpaisLUX:trend(gasto)	-2,612	13,618	-0,192	0,848237	
dpaisNLD:trend(gasto)	10,952	2,963	3,696	00339	***
dpaisPRT:trend(gasto)	3,752	2,963	1,266	0,208070	
dpaisSWE:trend(gasto)	-1,170	2,963	-0,395	0,693756	

Signif, codes: 0 '***' 0,001 '**'

0,01 '*' 0,05 '.' 0,1 ' ' 1

Residual standard error: 19,03 on 113 degrees of freedom

Multiple R-squared: 0,9939, Adjusted R-squared: 0,9923

F-statistic: 630,2 on 29 and 113 DF, p-value: < 2,2e-16

Siguiendo este modelo se puede determinar que existe relación significativa entre la producción científica en ER y las variables porcentaje del gasto sanitario en salud sobre el PIB y el país, así como la interacción entre ellas. El grado de relación entre las variables responde a un coeficiente de determinación ajustado de 0,992, con un p-valor inferior a 2,2e-16.

Al realizar la diagnosis del modelo final el estadístico Durbin-Watson devuelve un valor de 2,295, estando por tanto en la región de aceptación, por lo que se acepta la hipótesis de independencia. En cuanto a la hipótesis de normalidad de los residuos, en el histograma de los residuos studentizados (gráfico 8.2) se confirma que se distribuyen de forma normal. Por último, la hipótesis de homocedasticidad se puede corroborarla a través del gráfico 8.3, que contiene la distribución de los residuos frente a los valores ajustados, y que como vemos, no contiene ningún

tipo de estructura que induzca a pensar en la existencia de heterocedasticidad en los datos..

Por lo tanto se puede determinar que se cumplen las cuatro hipótesis básicas del modelo de regresión lineal: linealidad, independencia, normalidad y homocedasticidad.

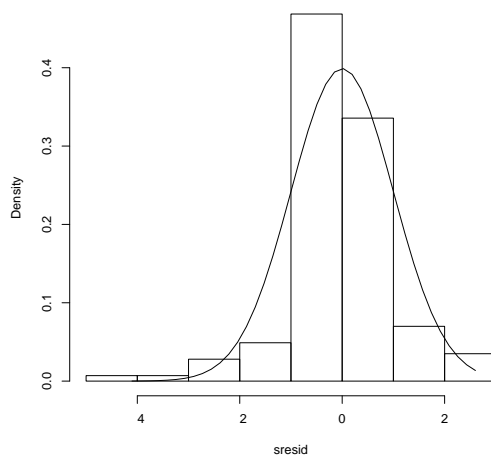


Figura 8.2: *Distribución de los residuos studentizados*

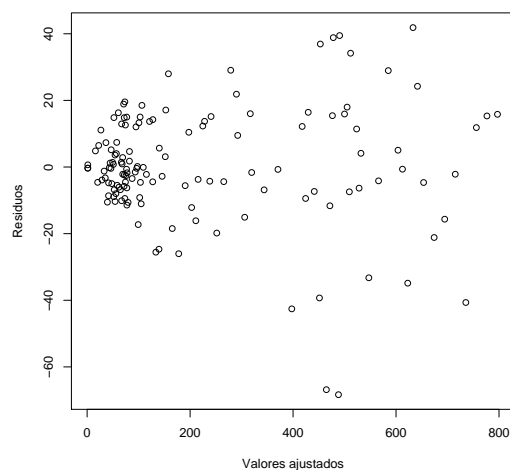


Figura 8.3: *Distribución de los residuos frente a valores ajustados*

A continuación se presenta el gráfico 8.4, que muestra la dispersión relativa al modelo de regresión ajustado. En él se puede observar la distribución de la producción por país en función del gasto sanitario y comprobar la linealidad del modelo presentado.

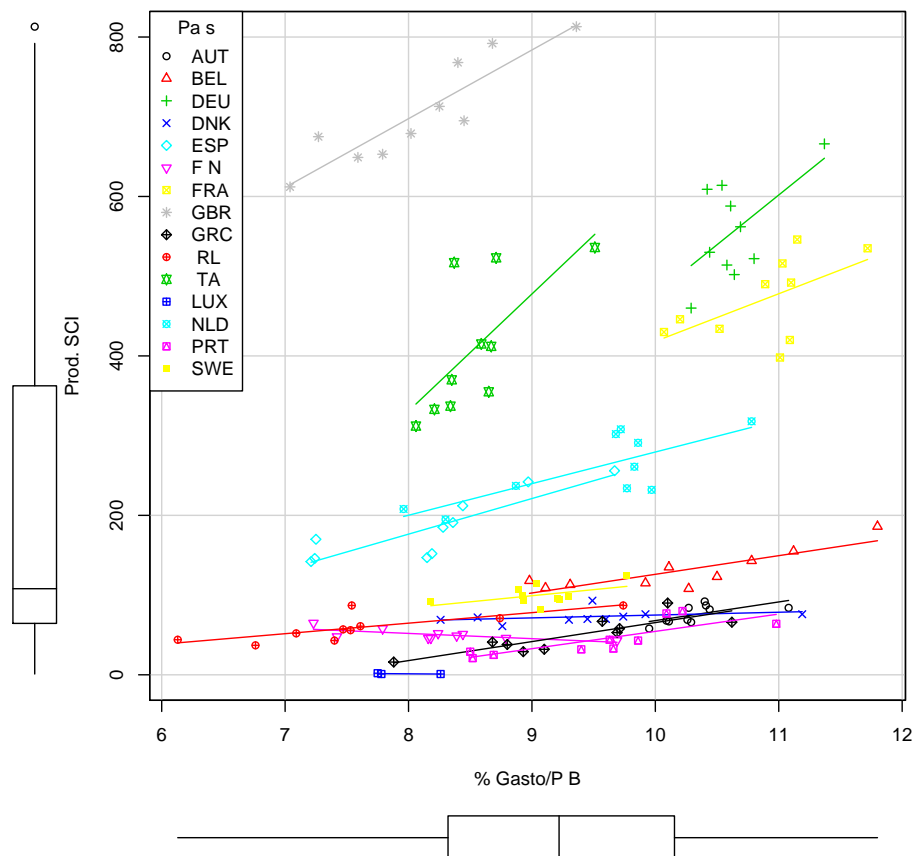


Figura 8.4: Dispersión de la producción por país en función del gasto sanitario

8.2. Producción frente a número de asociaciones de enfermos por país

Los datos con los que se cuenta para realizar el siguiente estudio de regresión son los que se muestran en la tabla 8.2

País	Producción SCI	Asociaciones de enfermos
AUT	757	73
BEL	1305	156
DEU	5567	272
DNK	729	39
ESP	1843	254
FIN	504	38
FRA	4707	385
GBR	7049	313
GRC	490	28
IRL	595	48
ITA	4110	244
LUX	4	11
NLD	2586	118
PRT	448	70
SWE	1000	41

Tabla 8.2: Variables para el estudio de regresión entre producción y número de asociaciones de enfermos

De nuevo se utiliza la producción de la base de datos SCI debido a la mejor representación de la producción por países. En este caso se cuenta tan solo con una variable por lo que se ha optado por utilizar el método de regresión lineal simple propuesto por PEÑA y SLATE (2006) para ajustar el modelo de regresión, obteniendo los siguientes resultados.

$$\text{Modelo: } PROD = \underset{(-92,486)}{\beta_0} + 15,828NUMAS + \epsilon$$

Donde:

PROD = Número de artículos.

NUMAS = Número de asociaciones.

β_0 = Coeficiente de regresión estimado.

ϵ = Residuos del modelo ajustado.

```
lm(formula = cor2$prod ~ cor2$numas)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-2084,920	-436,742	-4,992	391,943	2187,207

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-92,486	432,364	-0,214	0,834
cor2\$numas	15,828	2,367	6,688	1,5e-05 ***

Signif. codes: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

Residual standard error: 1083 on 13 degrees of freedom

Multiple R-squared: 0,7748, Adjusted R-squared: 0,7575

F-statistic: 44,73 on 1 and 13 DF, p-value: 1,499e-05

ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS

USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:

Level of Significance = 0,05

Call:

```
gvlma(x = fit)
```

	Value	p-value	Decision
Global Stat	1,142861	0,8874	Assumptions acceptable.
Skewness	0,009394	0,9228	Assumptions acceptable.
Kurtosis	0,029683	0,8632	Assumptions acceptable.

Link Function 0,057671 0,8102 Assumptions acceptable.
Heteroscedasticity 1,046113 0,3064 Assumptions acceptable.

Como se puede observar el modelo de regresión cumple con todas las hipótesis exigibles (linealidad, normalidad, independencia y homocedasticidad) con unos p-valores aceptables. Además, las variables presentan un alto grado de relación, con un R^2 de 0,757, por lo que se puede concluir que existe una relación directa entre ambas variables de forma que el número de asociaciones de enfermos influye positivamente en el número de publicaciones científicas.

Para concluir, en el siguiente gráfico (8.5) se observa dicha relación.

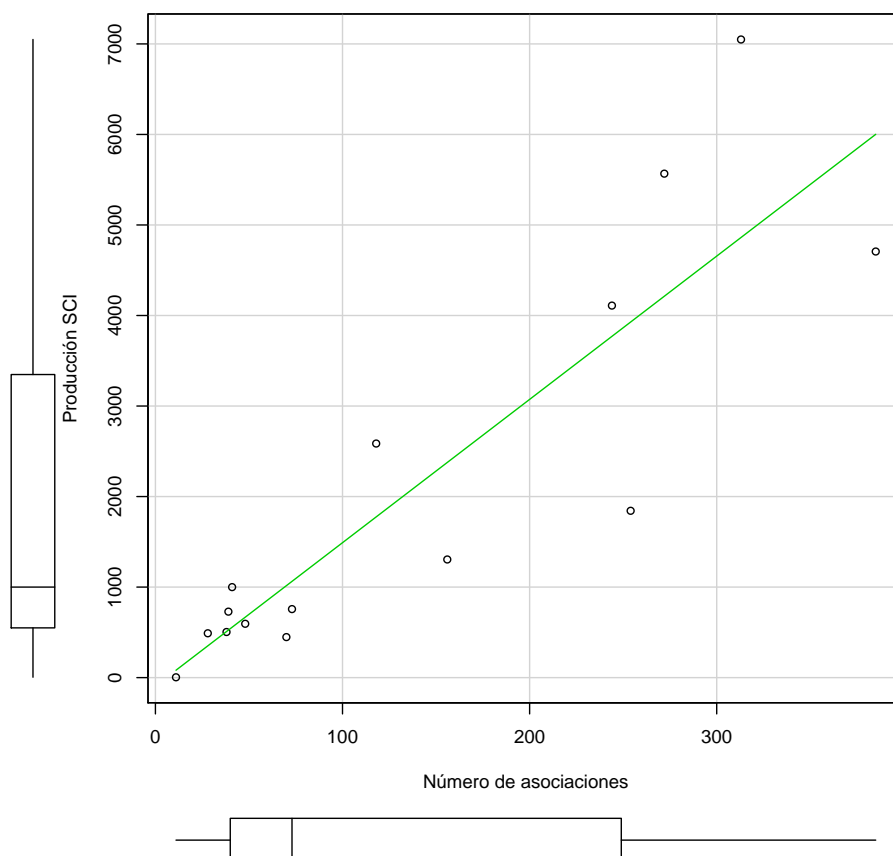


Figura 8.5: *Dispersión de la producción por país en función del número de asociaciones*

Sección 9

Indicadores conexionistas

Los indicadores conexionistas son aquellos que permiten mostrar gráficamente las posibles relaciones entre los datos en diferentes niveles de agregación, como pueden ser autores, temáticas, instituciones, etc. La ventaja de la utilización de este tipo de indicadores es que permiten, de un solo vistazo, comprender como se distribuyen los diferentes elementos de los datos obtenidos y cuál es su relación con los demás.

Para esta sección se han utilizado como niveles de agregación la colaboración entre autores, entre instituciones y entre países, así como las cocitaciones de autores y las materias que tienen documentos en común.

En estos grafos, el tamaño de cada nodo se corresponde con el número de conexiones que posee con el resto de nodos, mientras que el tamaño de las líneas que unen los nodos responde al número de relaciones que conectan dos nodos concretos.

9.1. Autores

Los siguientes indicadores conexionistas, relativos a la colaboración entre autores, se basan en una matriz realizada con los autores resultantes del Índice de Élite

de Price, que acumulan en el SCI el 90,15 % de la producción y el 89,47 % en el caso de Pubmed, frente al resto de autores que tienen al menos una colaboración con alguno de ellos. De esta forma, se construye una matriz de colaboración entre los autores y se obtiene un grafo basado en el número de trabajos en común que tienen los autores.

En los grafos de colaboración entre autores, tanto en el SCI como en Pubmed, se produce una agrupación en torno a algunos autores concretos, cuyo degree o número de conexiones con el resto de nodos es significativamente mayor que el del resto, estos autores acumulan a su alrededor un núcleo de autores con los que colaboran activamente y que apenas publican con otros grandes productores.

Sin embargo se dan algunos casos significativos en ambos grafos. Así, en el caso del SCI, los nodos situados entre la pareja Waterham-Wanders y el resto de núcleos, o en el caso de Pubmed, los nodos situados entre Wajner y Beck o los que unen a Bush y Elborn, existiendo conexiones entre ellos por medio de autores muy concretos.

Por otro lado se observan relaciones muy estrechas entre algunos autores que son a su vez grandes productores, es el caso de Wanders, Waterham y Hoffmann en ambos grafos, o en el caso del SCI, Hoiby, Tummler y Vandamme, a los que también podría añadirse Ferec, que se encuentra relacionado con ellos a través de un grupo de medianos productores. En el caso de Pubmed, además del caso que ya se ha mencionado de Waterham y Wanders también se encuentran otros, como el de Van Den Heuvel y Smeitink, que cuentan con un gran número de colaboraciones entre si.

En el grafo de cocitación, construido con los autores que han recibido al menos 100 citas, se muestran dos grandes grupos de autores que son citados juntos habitualmente, uno en la parte inferior de la visualización y otro en el lado derecho, ambos unidos a través de un solo autor, Ganzt, que es citado conjuntamente con autores de ambos núcleos. También aparecen grupos de autores menos numerosos que también son cocitados habitualmente, como es el caso de Eng, Desnik y

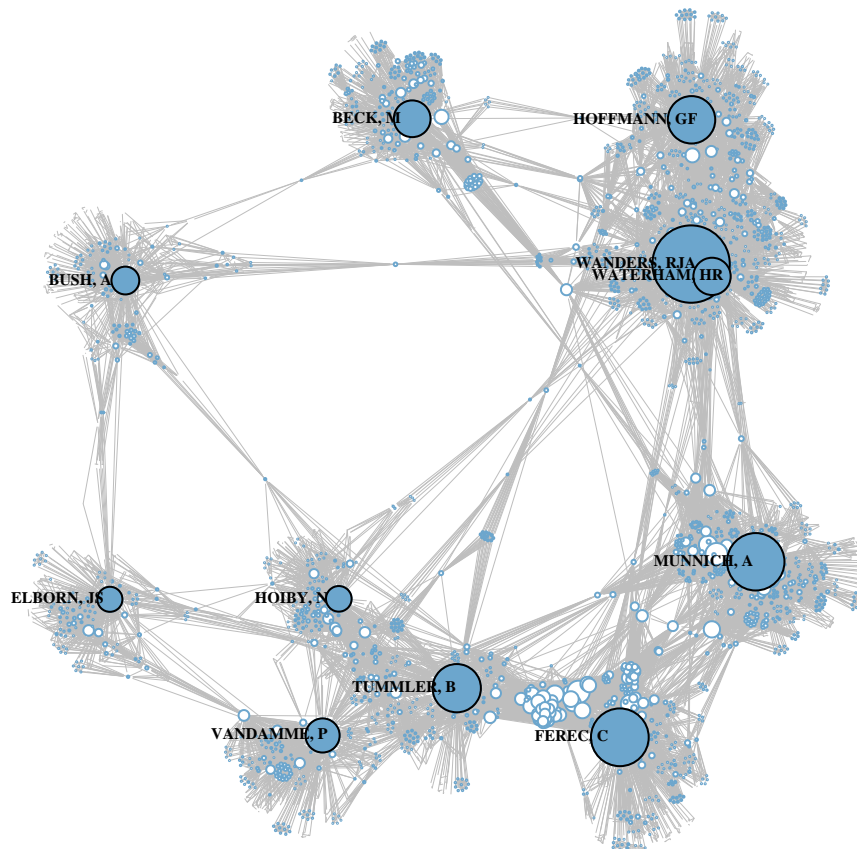


Figura 9.1: Red de colaboración entre autores (SCI)

Schiffmann, el grupo situado en centro del grafo, compuesto por Brook-Wilson, Bodzioch y Rust, a la izquierda el núcleo que componen Joutel, Dichangs y Charbriat y en la parte superior los compuestos por Amir y Hagberg y Koenig, Ervasti, Hoffmann y Blake.

Como se puede observar a través de los tres grafos, se trata de una red de autores muy cohesionada, con grandes núcleos de cocitación, lo que hace suponer que los autores trabajan en temáticas muy concretas, pero relacionadas entre si.

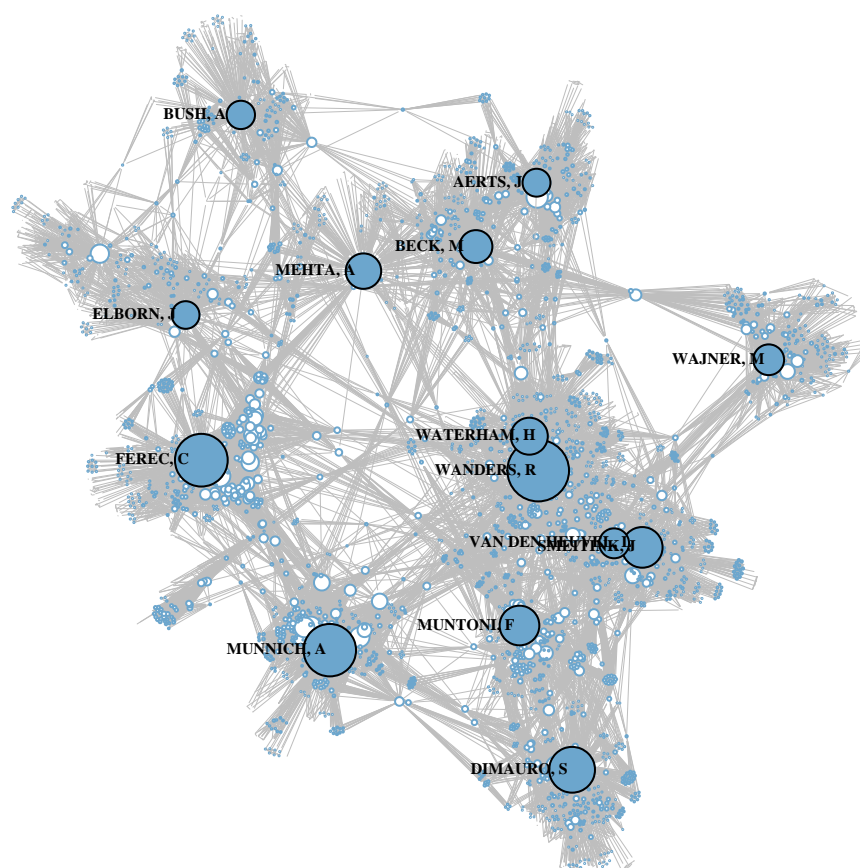


Figura 9.2: Red de colaboración entre autores (Pubmed)

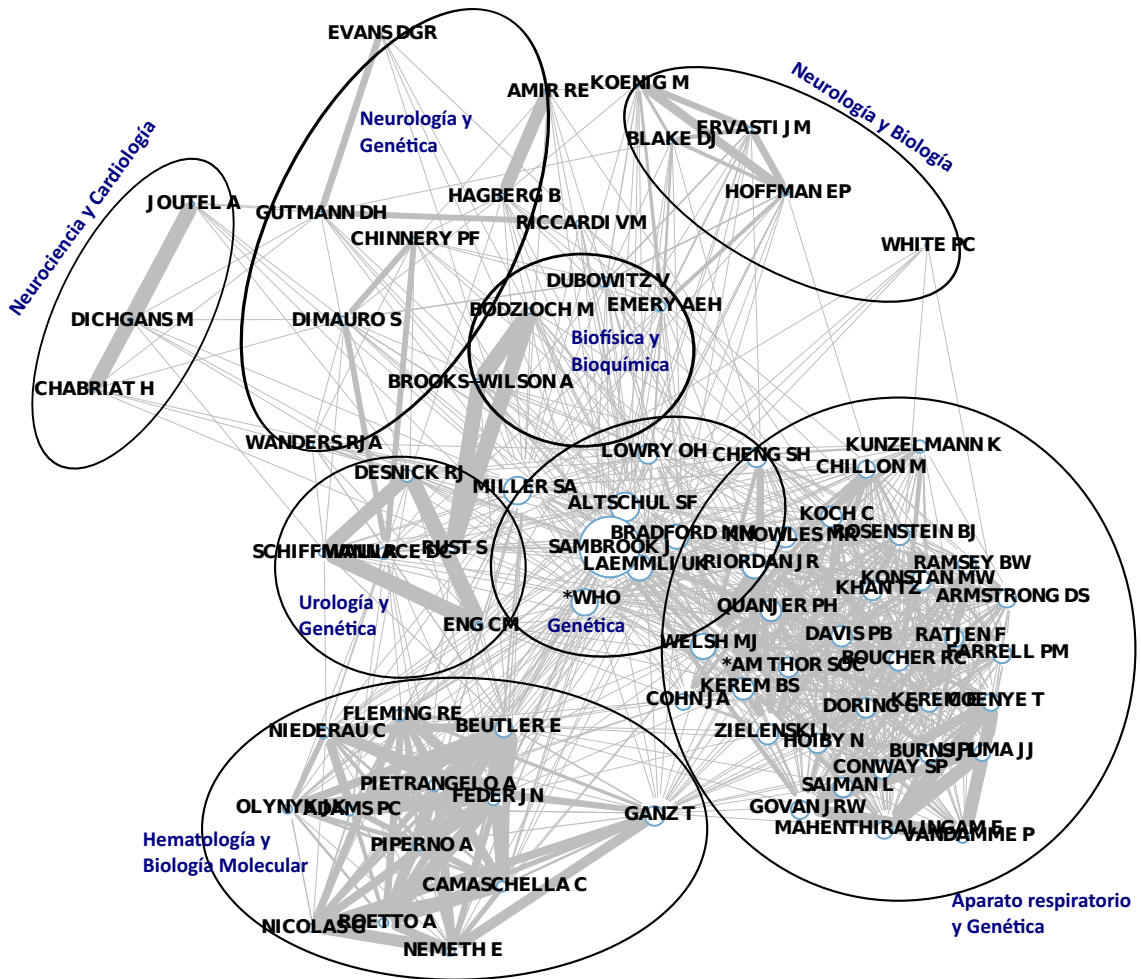


Figura 9.3: Red de autores citados

9.2. Instituciones

Al igual que en el caso de los autores, el análisis de redes sociales con las instituciones productoras permite encontrar las estructuras de colaboración que las rigen. En este caso el grafo de instituciones, realizado con la matriz de colaboraciones entre las instituciones identificadas en la base de datos SCI, se muestra aun más cohesionado que el de autores, ya que las instituciones más productivas forman tres grandes grupos, claramente visibles en el grafo. El primero, formado por la Universidad de Roma y la de Milán se encuentra en la parte superior, en la parte inferior derecha aparecen las siguientes instituciones francesas: INSERM, CNRS, Hospital Necker, Hospital Robert Degre y Universidad de París, y el último núcleo, situado en el centro del gráfico lo conforman el resto de instituciones. Cabe destacar que los dos primeros tienen una clara relación geográfica, pues el primero lo conforman instituciones italianas y el segundo instituciones francesas, mientras que en el tercer grupo, aparecen instituciones alemanas, inglesas, holandesas, etc., sin una estructura geográfica identificable. Se debe mencionar que para la elaboración de este grafo se han etiquetado las instituciones con una mayor número de colaboraciones, incluyendo aquellas que conforman el núcleo de instituciones más productoras así como 2 instituciones más (el Hospital Universitario de Oulu y el Hospital Robert Degre), que cuentan con un grado de colaboración muy alto y superan el umbral establecido para el etiquetado.

Institución	Abreviatura
Cnrs	Cnrs
Hannover Med Sch	Hams
Hop Necker Enfants Malad	Hnem
Inserm	Inse
Inst Child Hlth	Inch
Katholieke Univ Leuven	Kaul
Leiden Univ	Leiu
Ucl	Ucl
Univ Amsterdam	Unia
Univ Cambridge	Unic
Univ Heidelberg	Univhei
Univ Helsinki	Univhel
Univ London Imperial Coll Sci Technol & Med	Ulicst&M
Univ Milan	Univmi
Univ Munich	Univmu
Univ Oxford	Unio
Univ Paris	Unip
Univ Roma La Sapienza	Urlls
Univ Tubingen	Unit
Univ Hosp Oulu	Unih
Hosp Robert Debre	Hord

Tabla 9.1: Abreviaturas de las instituciones

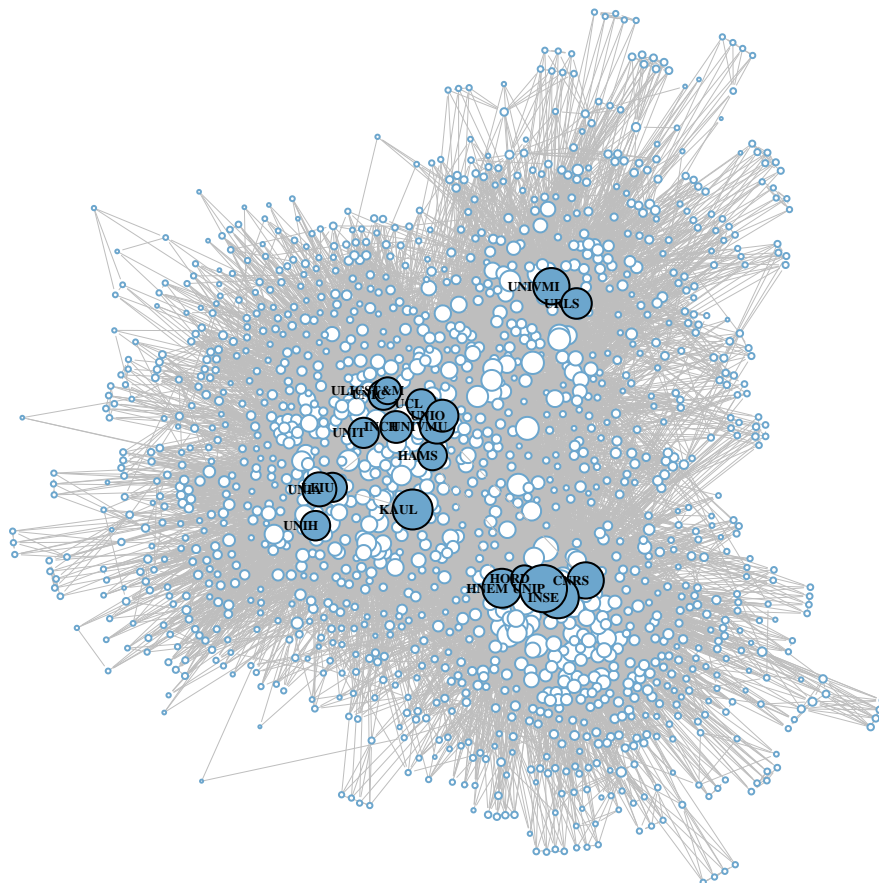


Figura 9.4: Red de instituciones

9.3. Países

En el primer grafo (9.5), correspondiente a la colaboración de los países de la Europa de los 15 con el resto del mundo, se puede observar que existe una gran colaboración entre estos 15 países, pero también con otros países entre los que destacan Estados Unidos (USA), China (CHN), Brasil (BRA), Argentina (ARG), Canadá (CAN), Turquía (TUR), Polonia (POL) o Israel (ISR).

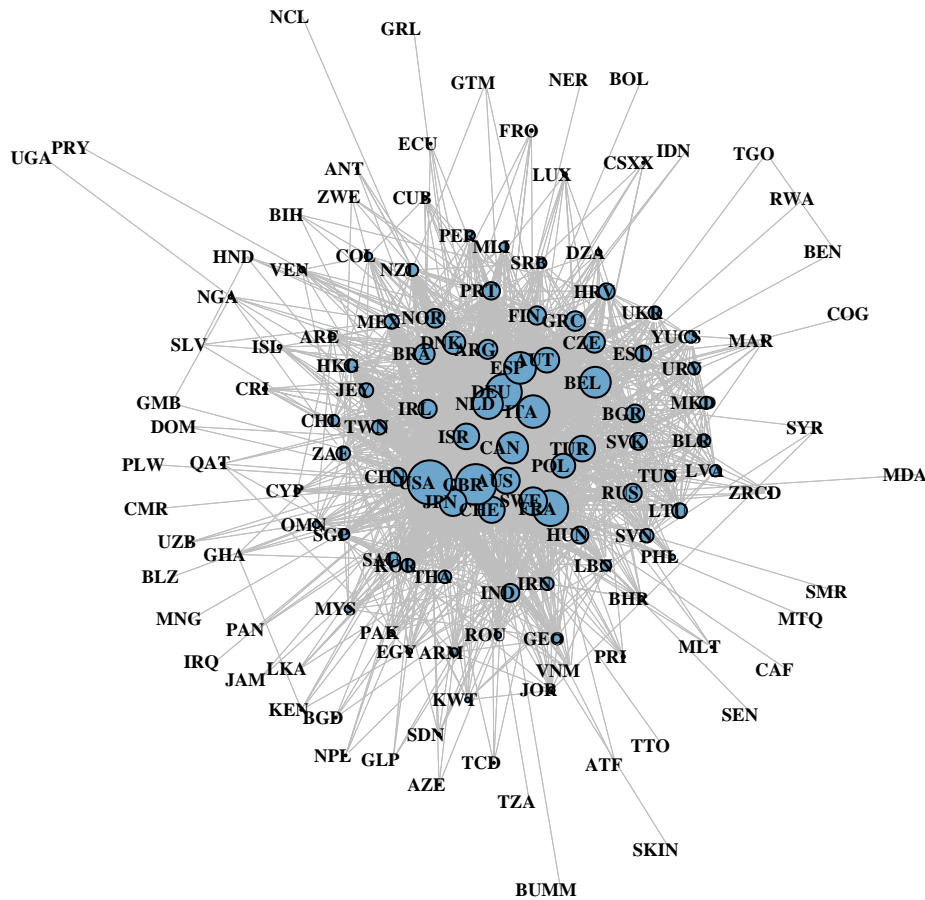


Figura 9.5: Red de la colaboración entre países

Por otro lado, en la red social del gráfico 9.6, conformado exclusivamente con las colaboraciones entre los países de la Europa de los 15, aparecen algunas diferencias en los perfiles de colaboración, especialmente entre los de Gran Bretaña y Alemania, puesto que el primero colabora más con países menos productivos como Irlanda, Suecia y Finlandia, mientras que Alemania solo tiene un nivel de

9.4. Temáticas

Al igual que las redes sociales relativas a la colaboración, las relativas a las temáticas han sido calculadas a partir del número de documentos clasificados con al menos dos temáticas (1.125 documentos en el SCI y 668 en Pubmed), y a partir de ellos se ha obtenido la matriz con las temáticas que dichos documentos comparten.

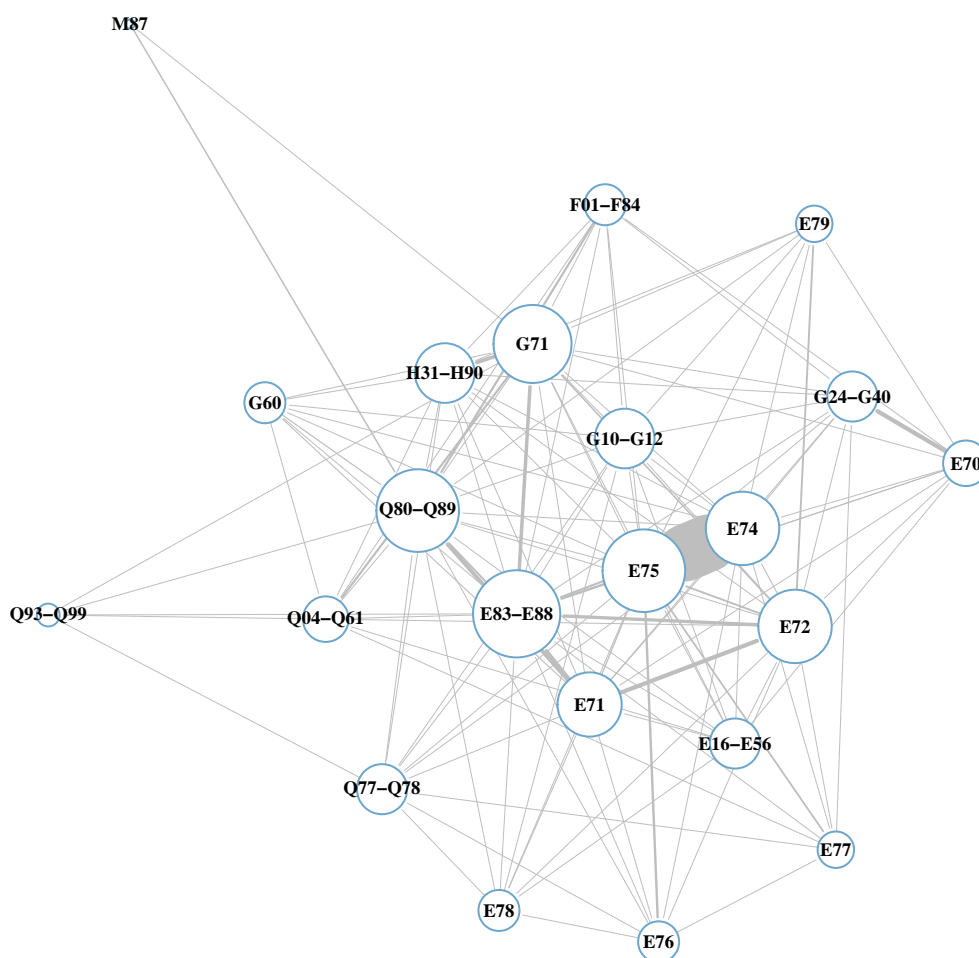


Figura 9.7: Red de trabajos codificados con ICD-10 (SCI)

Como se puede observar, los grafos obtenidos en ambas bases de datos son muy similares, destacando en los dos casos las posiciones centrales de los nodos E75, E83-E88 y Q80-Q89. Como diferencias entre ambos la más notable es que, mientras en el SCI aparecen un gran número de documentos compartidos entre los

nodos E75 y E74, en Pubmed dicha relación es mucho menor, estando el nodo E75 más relacionado con el E77 y el E72. Sin embargo, el número de documentos que comparten temática es mucho menor en Pubmed, por lo que el análisis es más fiable sobre los datos del SCI.

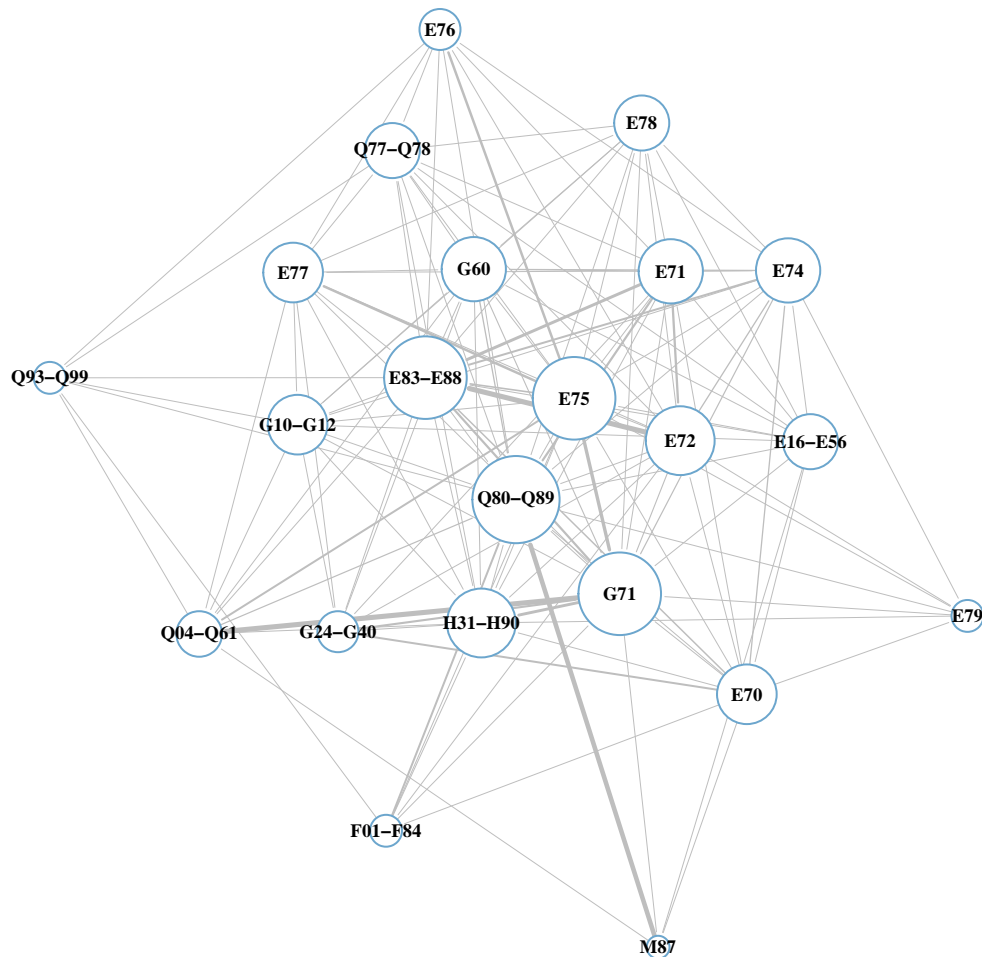


Figura 9.8: Red de trabajos codificados con ICD-10 (Pubmed)

Sección 10

Indicadores multidimensionales

Los indicadores multidimensionales permiten caracterizar la producción y hábitos de consumo de información científica, de forma simultánea y desde diferentes puntos de vista o dimensiones. En este caso se han dividido en función de la dimensión principal de cada análisis de forma que se encuentran agrupados en cinco grandes grupos: Autores, instituciones, temáticas, países y publicaciones.

10.1. Autores

En esta sección se han desarrollado dos tipos de análisis. Por un lado, a partir de las matrices de colaboración entre autores que posean colaboraciones con al menos otros 100 autores en todo el periodo de estudio (límite impuesto con el fin de eliminar a los autores transeúntes), se han calculado los análisis de clusters de los gráficos 10.1 y 10.2. Por otro lado, a partir del mismo conjunto de autores se han construido los análisis de correspondencias de autores frente a códigos de clasificación ICD-10, de forma que permitan observar gráficamente la distribución temática de los autores.

Como ya pudo observarse en los indicadores conexionistas, es fácil suponer que los autores están trabajando en áreas muy diferentes. El análisis de cluster confirma

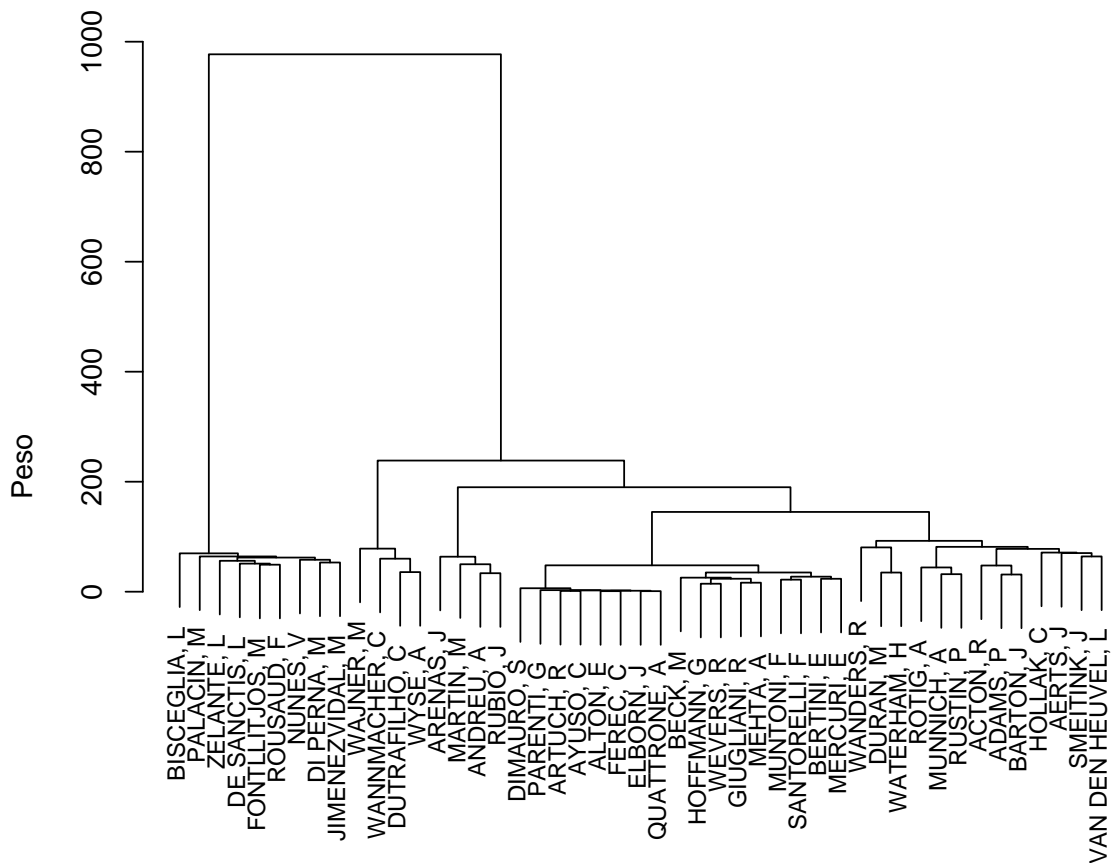


Figura 10.2: Cluster de autores (Pubmed)

autores y las clasificaciones ICD-10 en las que publican y por lo tanto, es correcto calcular los análisis de correspondencias que se muestran a continuación.

Estos análisis de los autores frente a las clasificaciones ICD-10 en las bases de datos SCI y Pubmed, (gráficos 10.3 y 10.5) cuentan con un porcentaje de variabilidad explicada del 37,4 % y el 35,1 % respectivamente a través de los dos componentes mostrados en los análisis. Esta variabilidad puede considerarse aceptable, principalmente porque gran parte de la variabilidad en las matrices de autores viene marcada por el número de colaboraciones entre ellos, tal y como se muestra en los grafos de colaboración y en los análisis de clusters analizados anteriormente.

En cuanto a la distribución de los autores alrededor de los códigos ICD-10, en ambas bases de datos se percibe una gran especialización de los mismos en torno a

materias concretas, si bien algunas de ellas se encuentran claramente relacionadas. Tanto en el caso de la base de datos SCI (gráfico 10.3) como de Pubmed (gráfico 10.5), se observa un claro sesgo de los autores hacia temáticas concretas.

En ambos análisis de correspondencias (gráfico ampliado 10.4 y 10.5) se puede observar que el grupo de autores que se comporta de forma diferente, es el que se sitúa en torno al grupo de enfermedades E83-E88. Este comportamiento es observable por su distanciamiento respecto a la primera dimensión analizada en el caso de SCI y en ambas dimensiones en el caso de Pubmed. Cabe resaltar que entre las enfermedades incluidas en este grupo se encuentra la Fibrosis Quística, que como ya se ha mencionado anteriormente, se trata de una de las enfermedades raras con mayor producción científica.

Por otro lado, respecto a la segunda dimensión en SCI, presenta un gran distanciamiento del eje central de los grupos Q93-Q99 (anormalidades cromosómicas) y F01-F84 (trastornos mentales y del comportamiento), mientras que en el caso de Pubmed, además del grupo E83-E88 mencionado anteriormente, los grupos de autores que más se distancian del eje central son aquellos que tratan enfermedades metabólicas relativas a los aminoácidos (E72), lípidos (E75) y glucasaminoglucanos (E76).

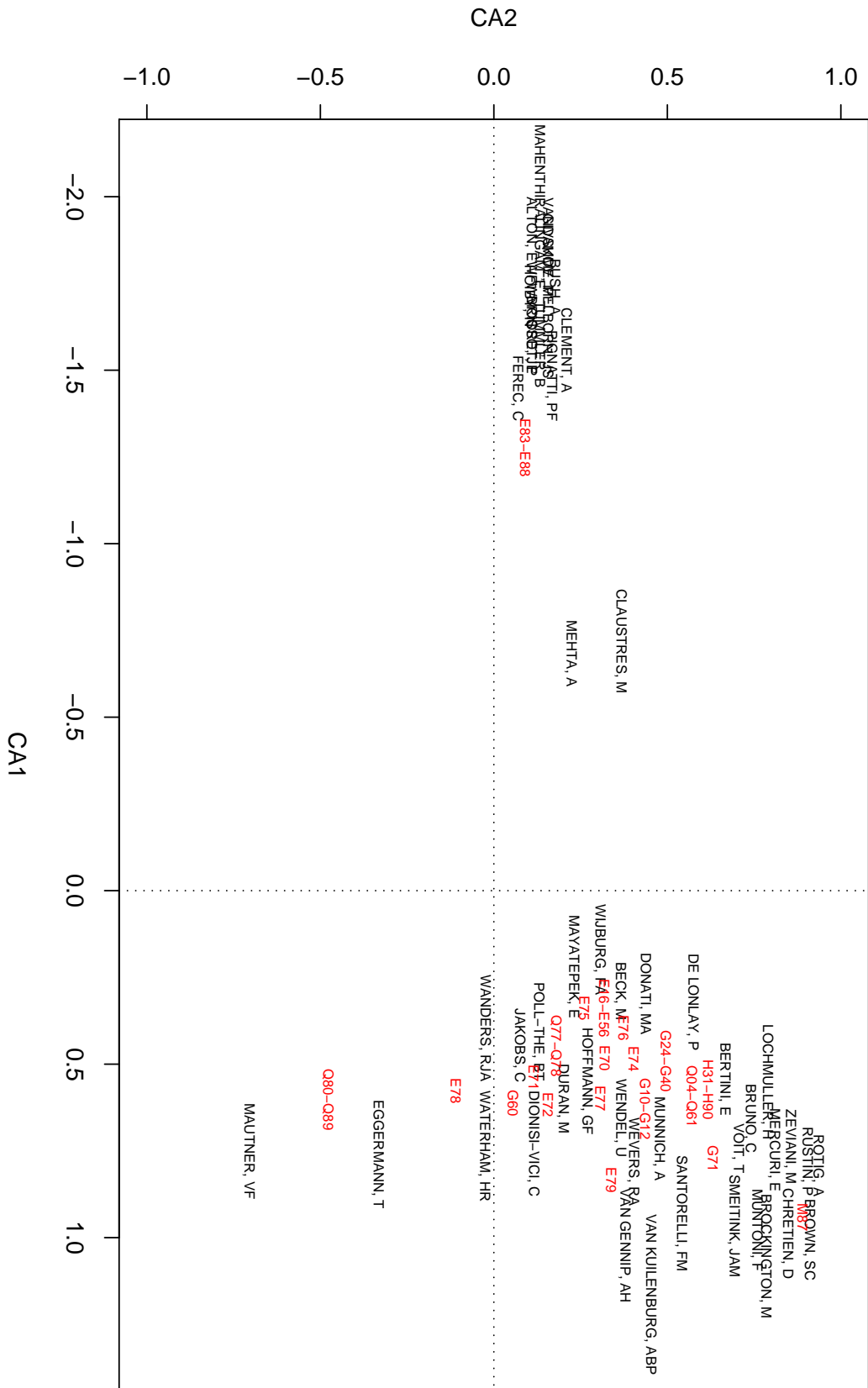


Figura 10.4: Detalle: Analisis de correspondencias de autores frente a códigos ICD-10. (SCI)

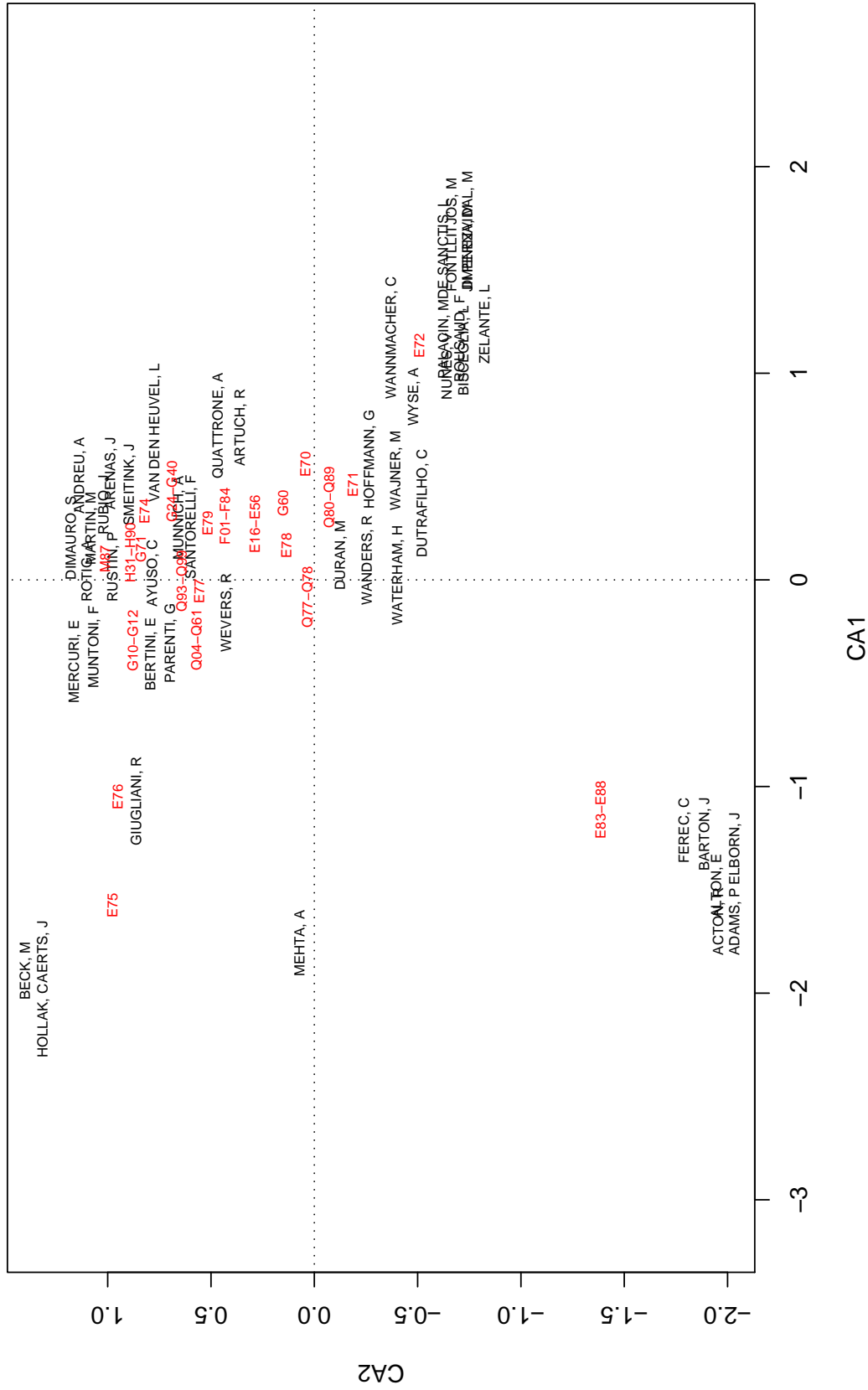


Figura 10.5: Análisis de correspondencias de autores frente a códigos ICD-10 (Pubmed)

10.2. Instituciones

Para el análisis multidimensional de los datos de la producción por institución se ha utilizado la base de datos SCI, dado que Pubmed solo recoge la filiación institucional del primer autor y, por tanto, no permite realizar las matrices de co-ocurrencias necesarias para los análisis multidimensionales.

El análisis de la colaboración, realizado anteriormente, por medio del análisis de redes sociales, presentaba una distribución llamativa, con la mayor parte de las instituciones agrupadas en el centro del grafo y dos núcleos de instituciones, francesas por un lado e italianas por otro, separados del resto y bien identificados.

Cuando se procede a realizar el análisis de clusters (gráfico 10.6), se observa que la agrupación encontrada en el análisis de las redes sociales se reproduce, especialmente entre los cuatro centros franceses, que colaboran estrechamente entre ellos, haciendo que posean un perfil muy diferente del resto de instituciones, que presentan una colaboración mucho más dispersa entre ellos.

Por otro lado, el análisis de clusters muestra una estructura de colaboración entre las instituciones del Reino Unido (UCL, Univ. Cambridge, Univ. London, Inst. Child Hlth. y Univ. Oxford). Esta estructura, que pasaba desapercibida en el análisis de redes sociales, muestra una ligera diferenciación, presente a un nivel muy bajo, entre las instituciones del Reino Unido y el resto de instituciones que conforman el cluster principal de la distribución.

Por último la distribución del análisis de correspondencias entre centros de trabajo de los autores y clasificación ICD-10 (gráfico 10.7) también muestra cierta dispersión en cuanto a las temáticas de investigación de las diferentes instituciones. También destaca el hecho de que las instituciones que presentaban unos perfiles más diferentes de colaboración se sitúan muy cerca del eje central, por lo que no se encuentran sesgadas hacia ninguna materia concreta.

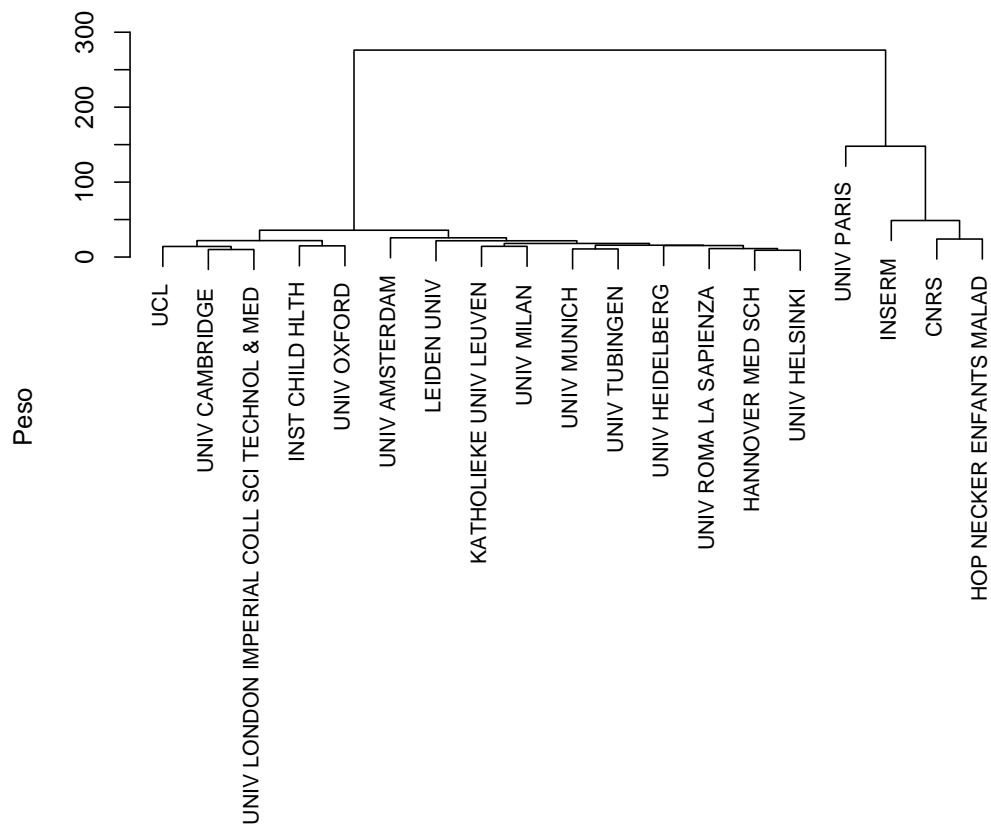


Figura 10.6: *Análisis de clusters de instituciones*

Dicho análisis cuenta con un 51,6% de la variabilidad explicada, que como en el caso de los autores, se debe a que el resto de la variabilidad queda explicada por los patrones de colaboración. Del mismo modo, previamente al análisis de correspondencias se realizó un test de significación mediante la prueba estadística del χ^2 , devolviendo un p-valor de 2,2e-16, por lo que se puede determinar que existe una clara relación de dependencia entre las variables.

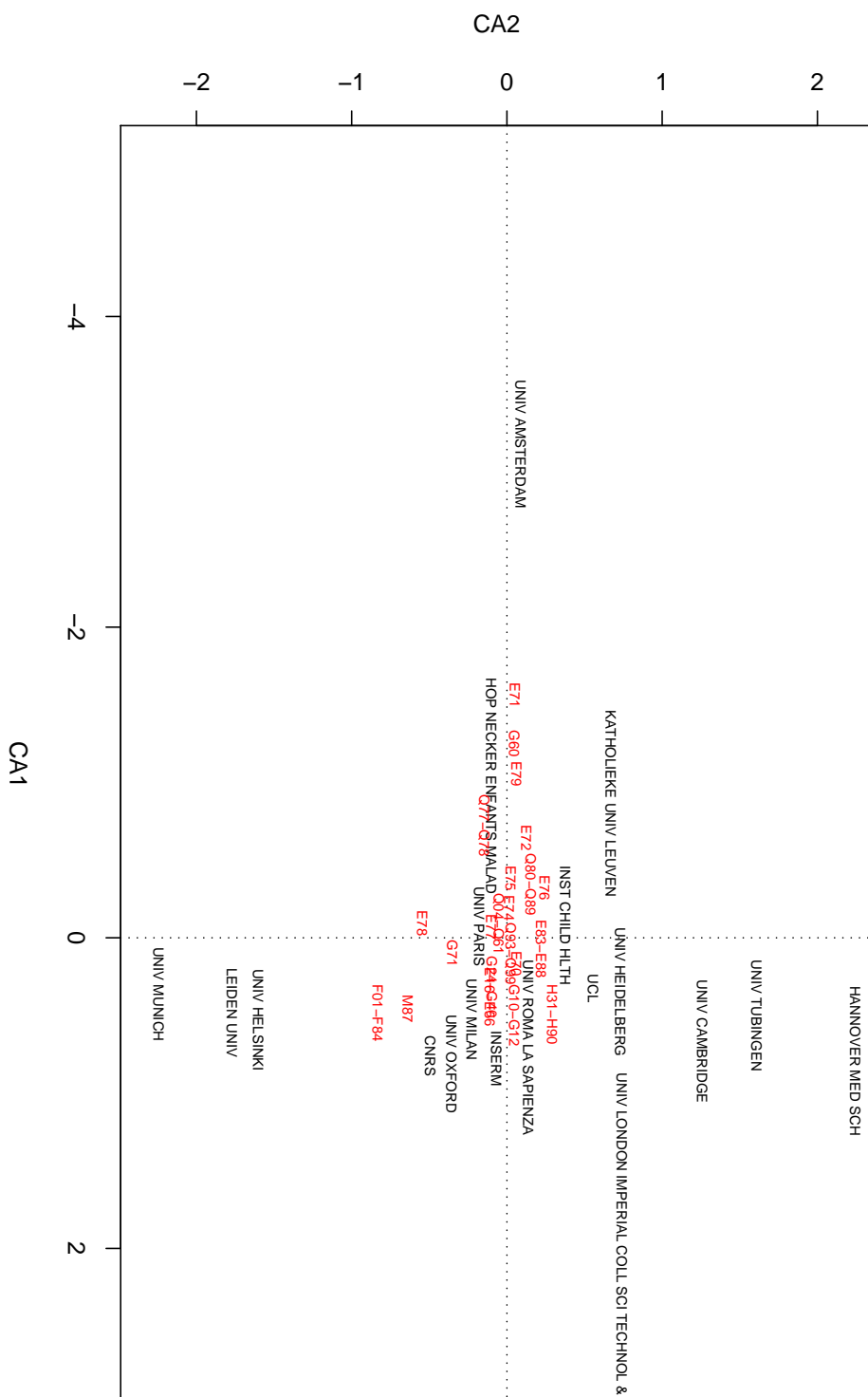


Figura 10.7: Análisis de correspondencias entre instituciones y clasificación ICD-10

10.3. Temáticas

Como primera aproximación a la distribución de las diferentes materias en el conjunto de datos estudiados se ha realizado un escalado multidimensional con la matriz de términos MeSH contenidos en las bases de datos (gráficos 10.8 y 10.9), asignados a través del método descrito en el apartado de metodología.

Para realizar la matriz se han utilizado los términos MeSH que aparecen en, al menos 50 (SCI) o 60 (Pubmed) documentos, y que acumulan al menos el 50 % de la producción científica en el conjunto de datos estudiados.

Como se puede observar la mayor parte de los términos se encuentran acumulados en torno al eje central, mientras que algunos términos se distancian de él en uno o ambos ejes. Es el caso de la Fibrosis quística, que al ser el término con mayor producción ejerce una influencia significativa (con un p-valor menor a $2,2e-16$) sobre la medida de las distancias entre los términos MeSH.

Otro elemento que muestra la influencia de la Fibrosis quística sobre la matriz de distancias es el hecho de que, si se elimina de dicha matriz, el valor del stress se reduce hasta el 10,91 %, un valor mucho más aceptable.

Por último al aplicarle la técnica de clusters (gráficos 10.10 y 10.11), también se observa que la Fibrosis quística se comporta de forma diferente al resto de los términos MeSH, con menos producción, separándose rápidamente de ellos. Este hecho se repite también en el caso de los términos asociados al área de Genética, que se separan rápidamente del resto, indicando un comportamiento claramente diferenciador, especialmente en el caso de Pubmed, en el que el dendograma se divide en dos grandes clusters, con los términos relativos a la Genética y a la Fibrosis quística situados a la derecha y el resto de términos en el cluster de la izquierda. Por otro lado, cabe destacar el hecho de que el término que antes se separa de los clusters principales sea “Base Sequence”, seguido de la Fibrosis quística.

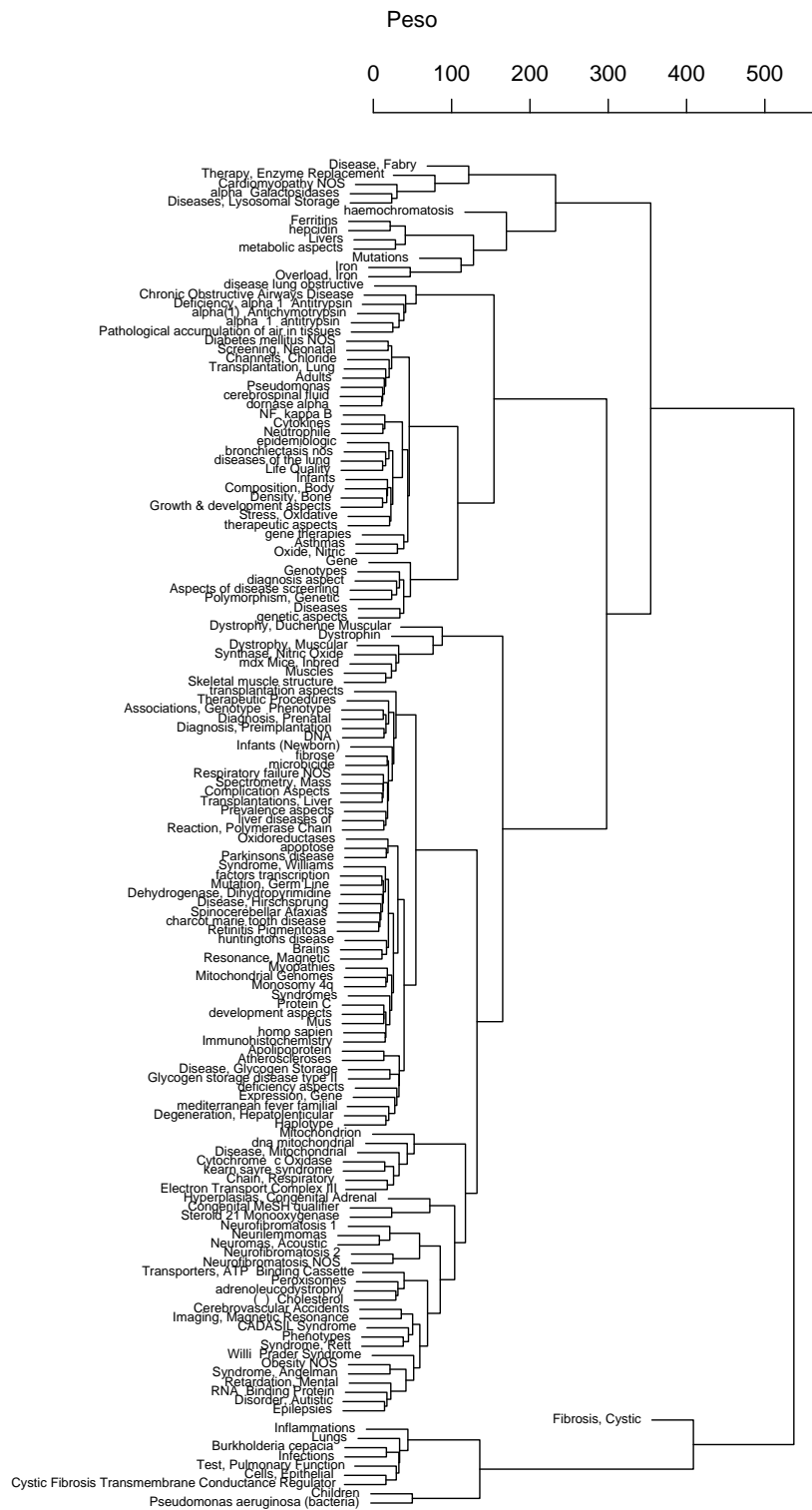


Figura 10.10: Clusters de términos MeSH (SCI)

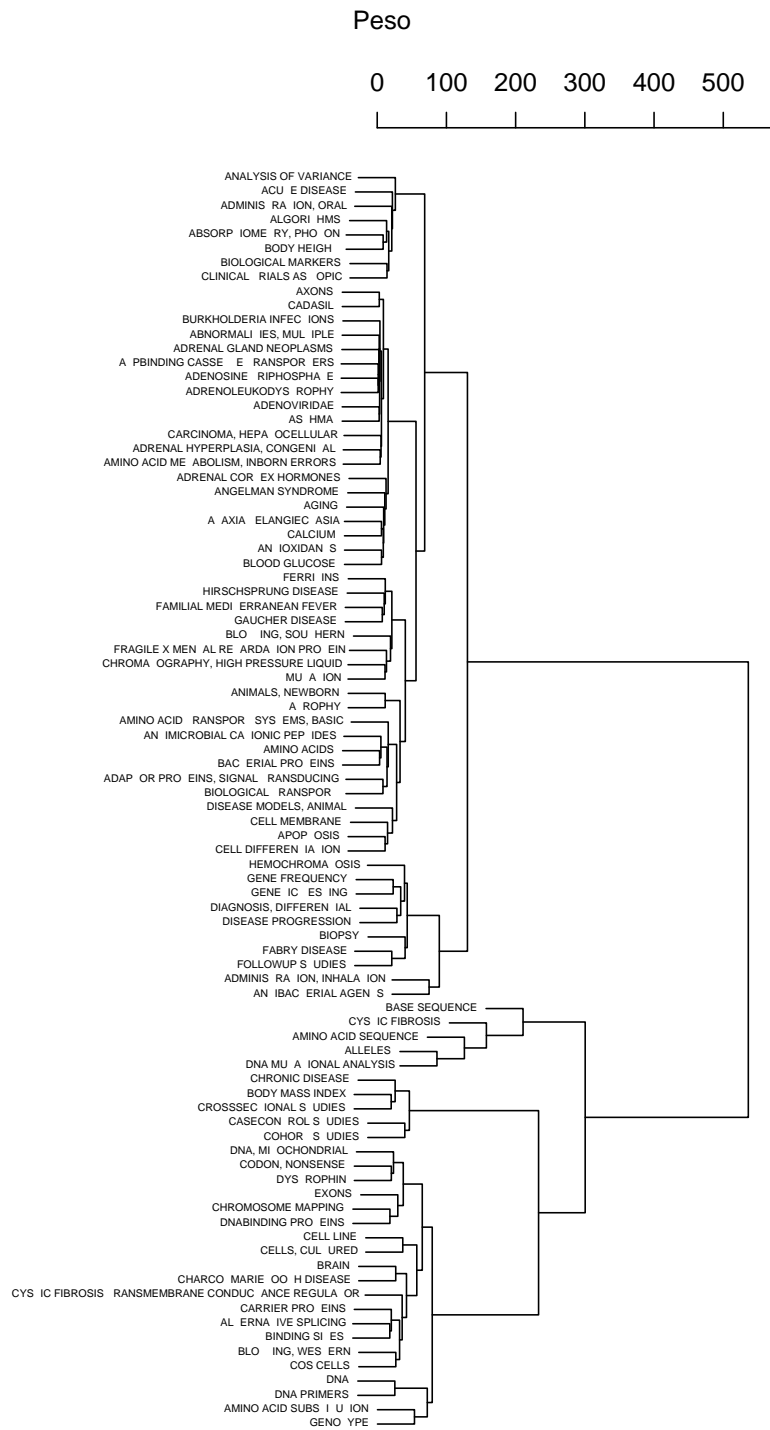


Figura 10.11: Clusters de términos MeSH (Pubmed)

10.4. Países

Al igual que en el caso de las instituciones, se han utilizado exclusivamente los datos de SCI debido a las limitaciones de Pubmed.

En el escalado multidimensional (gráfico 10.12) se observa que los países se distribuyen en dos grupos bien diferenciados respecto a la primera dimensión, que por lo que se ha observado en los indicadores conexionistas, debe corresponder al volumen de publicaciones de cada país, mientras que la distribución sobre la segunda dimensión posiciona a los países con respecto a sus perfiles de publicación tal y como se puede observar en la matriz de colaboraciones de la tabla 7.24.

Esta distribución se confirma en el análisis de clusters (gráfico 10.13), en el que se observa que existen dos grupos bien diferenciados, correspondientes a la distinción sobre la primera dimensión del escalado, y respecto a la segunda dimensión también muestra una rápida separación entre los dos países con perfiles más opuestos (Gran Bretaña y Alemania).

En el gráfico 10.14 se muestra la distribución idiomática por país, siendo fundamental para su interpretación los niveles de significación adjuntos al gráfico, gracias a los cuales se pueden apreciar las diferencias entre los diferentes idiomas de publicación según los países productores y si hay diferencia significativa con el resto o no. En este caso los valores que más destacan son los del idioma “UNSPECIFIC”, que es el que la Web of Science asigna al documento cuando no es capaz de determinar el idioma, y como vemos, es especialmente significativo en los documentos alemanes y españoles. También destacan los idiomas específicos de cada país, especialmente el caso de Francia, pues el Francés se encuentra muy sobrerrepresentado en detrimento del Inglés, que cuenta con una producción mucho menor de la esperada en el caso de documentos firmados por instituciones francesas.

Por último el gráfico 10.15 muestra la distribución de los países con la clasificación ICD-10 de las enfermedades a través de la técnica de análisis de correspondencias.

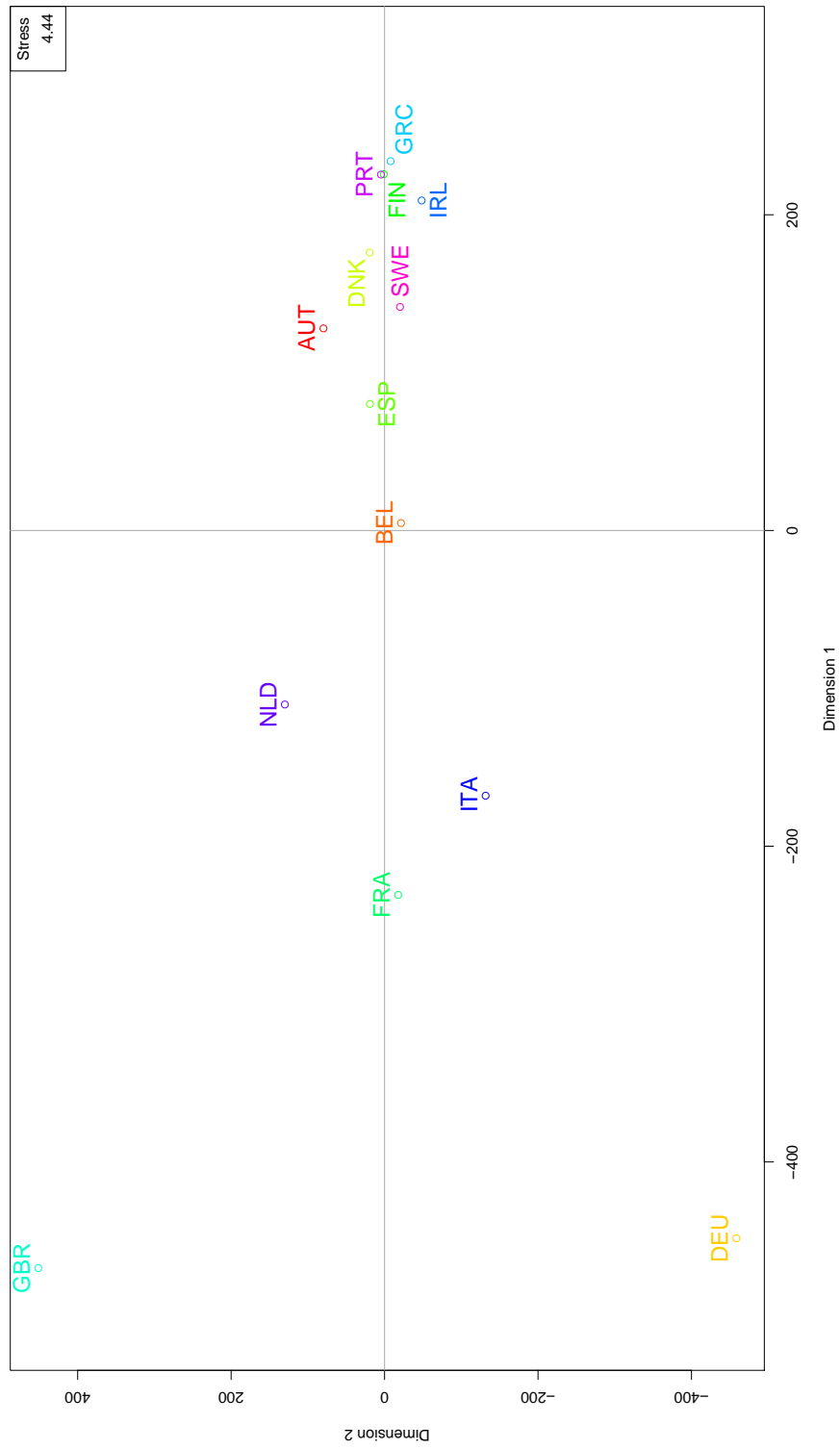


Figura 10.12: MDS de países

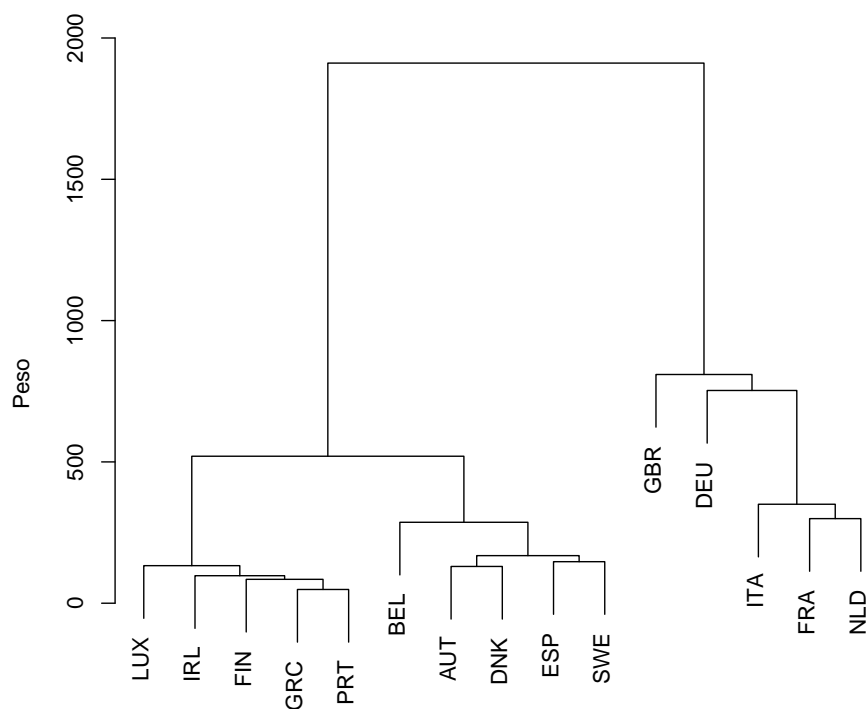


Figura 10.13: *Análisis de clusters de países*

De nuevo la prueba estadística arroja un p-valor aceptable para cualquier nivel de significación ($2,2e-16$) y la variabilidad explicada es del 49,8%. La representación gráfica muestra que la mayoría de los países se distribuyen de forma bastante uniforme con respecto a las temáticas, salvo en el caso de Irlanda, con un claro sesgo hacia la clasificación E83-E88 y el de Finlandia que se aproxima más a las clasificaciones F01-F84 y G71. También existe un sesgo destacable, pero mucho menor, de Luxemburgo (aunque por el volumen de documentos no es significativo) y Holanda hacia el conjunto de las clasificaciones “E”.

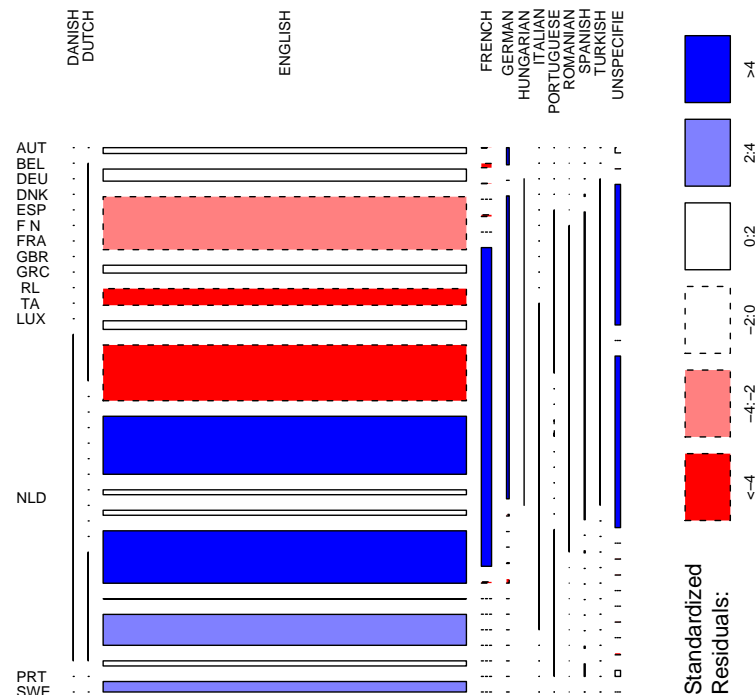


Figura 10.14: Gráfico de mosaico países-idiomas

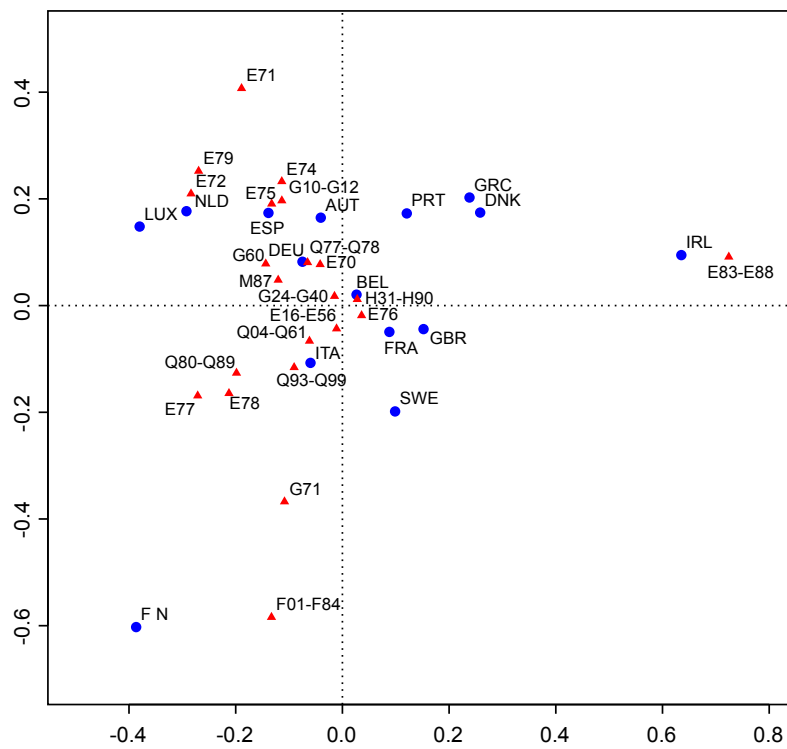


Figura 10.15: Análisis de correspondencias de países y clasificación CIE

10.5. Publicaciones

Para concluir con el análisis multidimensional de los datos se han llevado a cabo análisis de correspondencias de las revistas del núcleo, tanto en el SCI como en Pubmed, sobre la clasificación ICD-10 (gráficos 10.16 y 10.18) y sobre los países de las instituciones que publican en dichas revistas (gráfico 10.20), si bien en este caso solo se han considerado las direcciones del SCI debido a las limitaciones de Pubmed comentadas anteriormente.

Para llevar a cabo los análisis de correspondencias y con el objetivo de que fueran más legibles, se identificó cada revista con una abreviatura. Dicha combinación de revista-abreviatura puede consultarse en las tablas 10.1 y 10.2.

Revista	Abreviatura
Acta Paediatrica	Actp
American Journal Of Human Genetics	Ajohg
American Journal Of Medical Genetics	Ajomg
American Journal Of Medical Genetics Part A	Ajomgpa
American Journal Of Respiratory And Critical Care Medicine	Ajoraccm
Annals Of Neurology	Anon
Antimicrobial Agents And Chemotherapy	Aaac
Archives De Pediatrie	Ardp
Archives Of Disease In Childhood	Aodic
Archives Of Neurology	Aron
Biochemical And Biophysical Research Communications	Babrc
Biochemical Journal	Bioj
Blood	Bloo
Blood Cells Molecules And Diseases	Bcmad
Brain	Brai

Continúa en la página siguiente...

Revista	Abreviatura
Chest	Ches
Clinical Chemistry	Clic
Clinical Genetics	Clig
European Journal Of Human Genetics	Ejohg
European Journal Of Pediatrics	Ejop
European Respiratory Journal	Eurj
Febs Letters	Febl
Hormone Research	Horr
Human Genetics	Humag
Human Molecular Genetics	Hummg
Human Mutation	Humm
Journal Of Biological Chemistry	Jobc
Journal Of Clinical Endocrinology & Metabolism	Joce&M
Journal Of Clinical Microbiology	Jocm
Journal Of Cystic Fibrosis	Jocf
Journal Of Inherited Metabolic Disease	Joimd
Journal Of Lipid Research	Jolr
Journal Of Medical Genetics	Jomg
Journal Of Pediatrics	Joop
Molecular Genetics And Metabolism	Mgam
Neurology	Neurol
Neuromuscular Disorders	Neud
Neuropediatrics	Neurop
Pediatric Pulmonology	Pedp
Pediatric Research	Pedr
Prenatal Diagnosis	Pred
Proceedings Of The National Academy Of Sciences Of The United States Of America	Potnaosotusoa
Respiratory Medicine	Resm
Revista De Neurologia	Redn
Revue Des Maladies Respiratoires	Rdmr

Continúa en la página siguiente...

Revista	Abreviatura
Thorax	Thor

Tabla 10.1: *Revistas del núcleo y sus abreviaturas. (SC)*

Revista	Abreviatura
Acta Paediatr	Actp
Am J Hum Genet	Ajhg
Am J Med Genet A	Ajmga
Am J Respir Crit Care Med	Ajrccm
Ann Neurol	Annn
Arch Dis Child	Arde
Arch Neurol	Arcn
Arch Pediatr	Arcp
Biochem Biophys Res Commun	Bbrc
Biochim Biophys Acta	Biba
Blood Cells Mol Dis	Bcmd
Brain	Brai
Chest	Ches
Clin Chem	Clic
Clin Genet	Clig
Cochrane Database Syst Rev	Cdsr
Eur J Hum Genet	Ejhg
Eur J Pediatr	Eujp
Eur Respir J	Eurj
Genet Test	Gent
Hum Genet	Humge
Hum Mol Genet	Hummg
Hum Mutat	Humm
J Biol Chem	Jbic
J Clin Endocrinol Metab	Jcem
J Cyst Fibros	Jcyf
J Inherit Metab Dis	Jimd
J Med Genet	Jmeg
J Pediatr	Jped
J Pediatr Surg	Jpes
J R Soc Med	Jrsm
Mol Genet Metab	Mogm
Mov Disord	Movd

Continúa en la página siguiente...

Revista	Abreviatura
Muscle Nerve	Musn
Ned Tijdschr Geneeskd	Netg
Neurology	Neurol
Neuromuscul Disord	Neud
Neuropediatrics	Neurop
Pediatr Pulmonol	Pedp
Pediatr Res	Pedr
Prenat Diagn	Pred
Proc Natl Acad Sci U S A	Pnasusa
Rev Neurol	Revn
Thorax	Thor

Tabla 10.2: *Revistas del núcleo y sus abreviaturas (Pubmed)*

En ambas bases de datos las revistas se distribuyen de forma similar, concentrándose la mayor parte de ellas en torno al eje central, sin gran diferencia entre sus perfiles de publicación respecto a la clasificación ICD-10, pero con algunas revistas que se alejan de dicho núcleo y cuentan con un perfil más cercano a dos clasificaciones concretas (M87 y E16-E56) que responden a trastornos osteopáticos y trastornos hormonales respectivamente.

Sin embargo, cuando se practica una ampliación sobre el eje central de los análisis de correspondencias (gráficos 10.17 y 10.19) se puede observar mucho mejor la distribución de las revistas frente a las temáticas y percibir más claramente el sesgo de algunas revistas hacia el área E83-E88 en ambas bases de datos.

En cuanto a los lugares de publicación por países, el test del χ^2 como medida de la dependencia entre el país de publicación de las revistas y el país de la filiación de los autores que arroja un valor de 7198,4 y un p-valor inferior a 2,2e-16, por lo que se puede rechazar la hipótesis de independencia y concluir que existe relación de dependencia entre ambas variables.

Una vez realizada la comprobación estadística, el análisis de correspondencias (gráficos 10.20 y 10.21), muestra que la mayor parte de los países productores (en rojo) tienen un perfil de publicación que se centra fundamentalmente en revistas

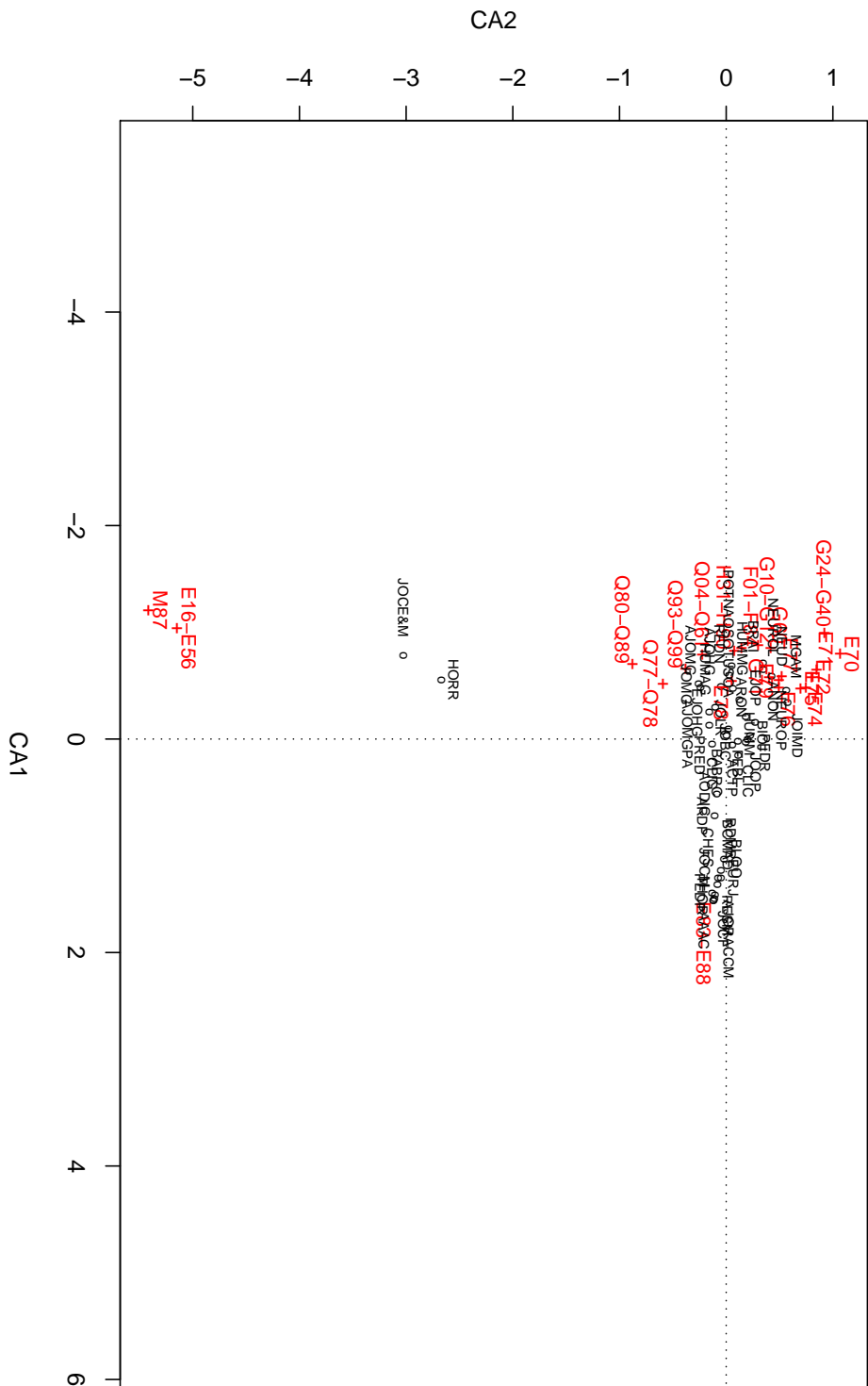


Figura 10.16: Análisis de correspondencias de revistas y clasificación CIE (SCI)

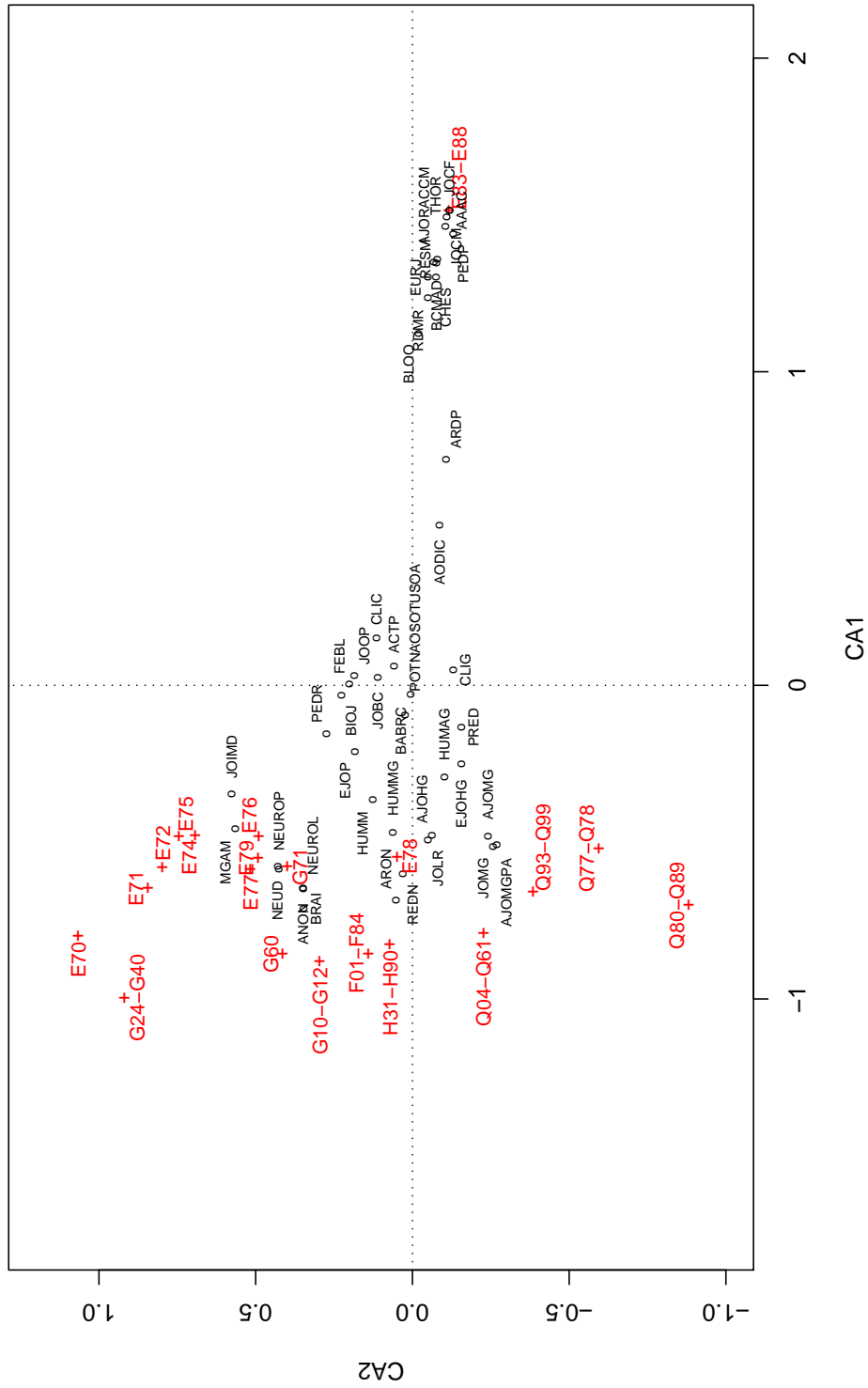


Figura 10.17: Detalle: Análisis de correspondencias de revistas y clasificación CIE (SCI)

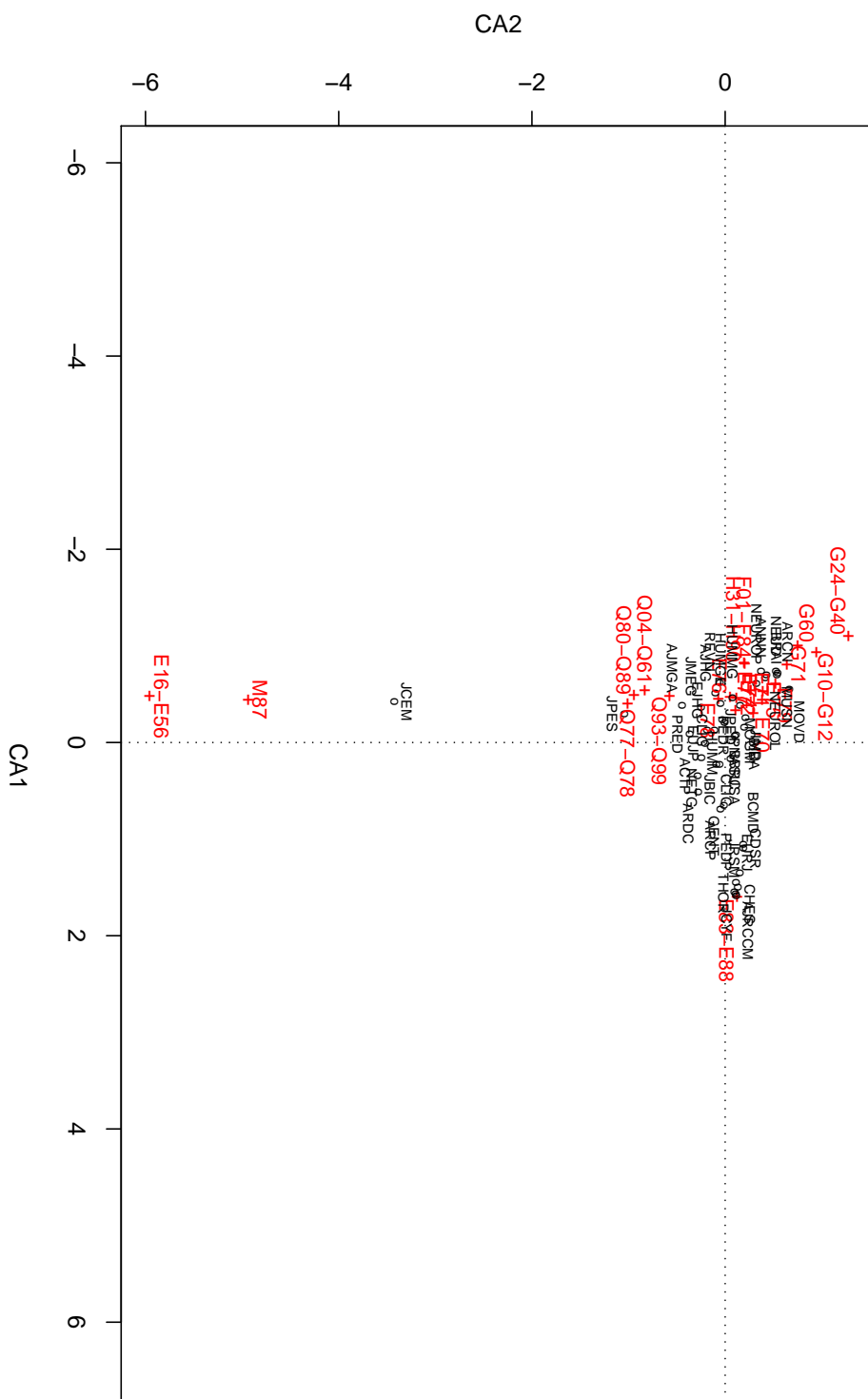


Figura 10.18: Análisis de correspondencias de revistas y clasificación CIE (Pubmed)

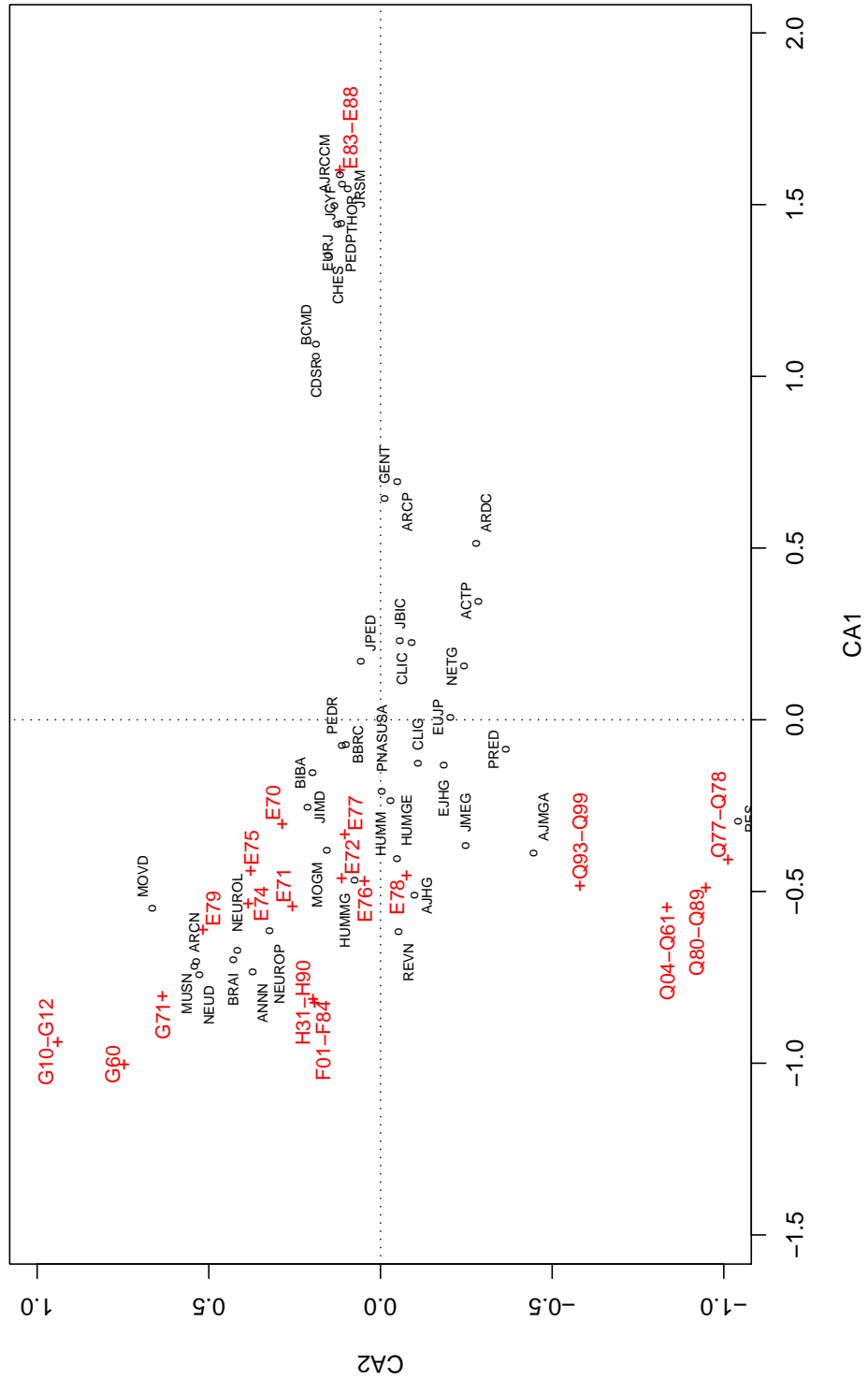


Figura 10.19: Detalle: Análisis de correspondencias de revistas y clasificación CIE (Pubmed)

estadounidenses o inglesas, mientras que solo dos países (España y Francia) y en menor medida Portugal, se acercan más a un perfil de publicación nacional, salvo en el caso de Portugal, que al no contar con ninguna revista nacional en el núcleo, se asemeja más al perfil de publicación de España.

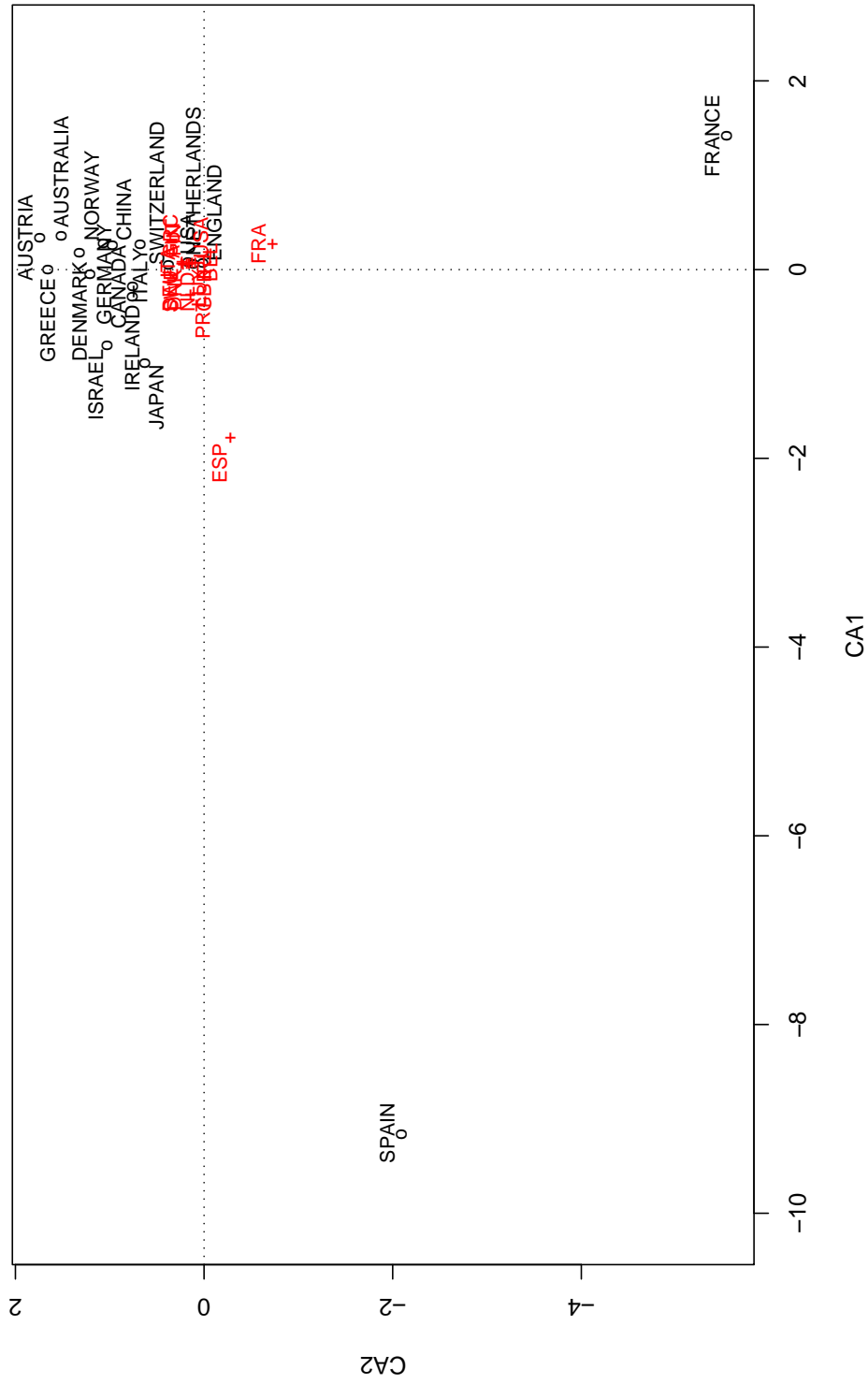


Figura 10.20: Análisis de correspondencias de revistas y países

Capítulo IV

Discusión y conclusiones

Sección 11

Discusión

En el capítulo de resultados se han expuesto los diferentes indicadores, unidimensionales y multidimensionales, que permiten caracterizar la investigación científica realizada en torno a las 252 enfermedades estudiadas. Cabe destacar que se trata de un conjunto de enfermedades relativamente limitado, pues el total de enfermedades raras abarca entre seis y siete mil desórdenes (ORPHANET, 2011a).

Sin embargo, debido a la naturaleza de estas enfermedades y al hecho de que la mayor parte de las mismas poseen una causa genética (FISCHER et al., 2005; STOLK, 2006), se ha considerado que el conjunto estudiado, compuesto por enfermedades cuya causa genética ha sido identificada (INERGEN, 2000), es lo suficientemente representativo como para extrapolar gran parte de los resultados obtenidos al conjunto de la investigación realizada sobre enfermedades raras en la Europa de los quince, salvo en el caso de la distribución temática, como se verá más adelante.

Para llevar a cabo el análisis de los resultados es conveniente, debido al carácter general de los mismos, comenzar por una caracterización global de los indicadores obtenidos tanto a nivel de producción como de consumo e impacto para posteriormente descender en el análisis de los indicadores agregados por países, temáticas, fuentes, instituciones y autores.

Tanto la producción anual como su tasa de crecimiento, recogidas por el gráfico 7.1 y la tabla 7.1 respectivamente, muestran una tendencia claramente creciente. Ahora bien, en ambas bases de datos, la producción de 2005 experimenta un aumento significativamente mayor, para luego volver a tener una tendencia creciente y ligeramente más apuntada. Esta situación, observada en el año 2005, puede estar indicando que algo antes de esta fecha, para que se pueda recoger posteriormente y de forma simultánea en ambas bases de datos, se ha producido un hito importante en la investigación sobre enfermedades raras, que conlleva un aumento en la actividad investigadora. En vista de los resultados que arrojan estudios anteriores (SERRANO-LÓPEZ y MARTÍN-MORENO, 2011; HEEMSTRA y VAN WEELY, 2009; LUENGO et al., 2001), se puede determinar que dicho aumento en la actividad científica, observable tanto gráficamente como en las tasas de crecimiento, se debe fundamentalmente a la influencia que la secuenciación del genoma humano ha ejercido sobre la investigación realizada en estas enfermedades, pues se estima que el 80 % de ellas poseen una causa genética (MINISTERIO DE SANIDAD Y POLÍTICA SOCIAL, 2009).

Precisamente, esta secuenciación del genoma humano, ha popularizado la investigación genética, en busca de las cadenas de ADN y los genes que influyen en las enfermedades humanas, ya sean raras o no y ha abierto las puertas a investigaciones con un desarrollo más profundo, cuyo mayor exponente es la reciente aparición del banco de datos ENCODE, que ha permitido crear un catálogo con todos los elementos funcionales que contiene el genoma humano, alcanzando un volumen de datos próximo a los 15 terabytes de información genómica (ECKER et al., 2012; SANYAL et al., 2012; GERSTEIN et al., 2012; NEPH et al., 2012; THURMAN et al., 2012; DJEBALI et al., 2012; DUNHAM et al., 2012).

En cuanto a los tipos documentales que prefieren utilizar los autores para publicar sus trabajos, las tablas 7.3 y 7.4, debido a que se ha trabajado únicamente con artículos de revista, la proporción de trabajos publicados además como actas de congresos o capítulos de libros es muy pequeña. Por otro lado, en el caso de Pubmed, la tipología documental no se centra únicamente en aspectos formales si

no, también de contenido, por lo que se puede observar que la mayor parte de los trabajos se han publicado como artículos de revista, pero también que el volumen de “estudios de caso” (case reports), es mucho menor de lo habitual en el área de Ciencias de la Vida (GLÄNZEL y SCHOEPFLIN, 1999) y más cercana al porcentaje de estudios de caso presentes en las ciencias experimentales, como muestra el trabajo de MARTÍN-MORENO y SANZ-CASADO (1996), lo cual puede ser debido a la relación, mencionada anteriormente, entre las enfermedades estudiadas y la investigación genética.

Para terminar con el análisis general de la producción científica se va a tratar la distribución idiomática de las publicaciones. En primer lugar, tal y como es habitual en la literatura científica y mucho más en las ciencias puras (VILLAR, 1988), la mayor parte de los documentos (un 91 % en SCI y un 93,98 % en Pubmed) se encuentran escritos en inglés. Entre los efectos que esto supone, tal y como NAVARRO (2001a) describe en su trabajo, se encuentran la adopción de anglicismos por parte de las comunidades científicas de habla no inglesa; la simplificación de la comunicación internacional, facilitando la labor de intercambio de información entre científicos de cualquier área geográfica; el sesgo hacia el inglés en la publicación de trabajos científicos, de forma que los trabajos publicados en idiomas diferentes al inglés alcanzan una repercusión mucho menor y en ocasiones pasan por desconocidos descubrimientos hechos mucho tiempo atrás. Tal es el caso del “músculo esfenomandibular”, que era desconocido por la comunidad científica angloparlante debido a que solo se encontraba recogido en los tratados anatómicos en francés y alemán (GROSCURTH, 1996).

Volviendo a la distribución de idiomas en los registros analizados, se observa que los idiomas predominantes, aparte del inglés, son el francés, el español y el alemán, si bien el número de trabajos en estos idiomas es mucho mayor en Pubmed que en SCI debido a que en esta última base de datos el sesgo idiomático es mucho mayor y recoge fundamentalmente literatura científica publicada en inglés, incluyendo un conjunto comparativamente pequeño de revistas en otros idiomas. Por todo ello, se ha encontrado que siendo el español el tercer idioma en Pubmed y el

cuarto en SCI, existe una diferencia de 436 trabajos entre ambas bases de datos, suponiendo el español el 2,32 % de la producción en Pubmed y solo el 0,31 % en SCI. Este efecto, que se produce a una escala mucho menor con el resto de idiomas recogidos, indica una preferencia de los investigadores de habla hispana de publicar en revistas en su idioma, estando estas recogidas por Pubmed pero no por la Web of Science.

Por otro lado, la distribución de las referencias bibliográficas por año (tabla 7.30), muestra un aumento progresivo desde la fecha de la primera referencia, siguiendo una progresión exponencial, como se observa en las funciones de densidad por año de publicación. Es destacable que a pesar de que el periodo de estudio abarca las publicaciones científicas publicadas entre 2000 y 2009, aparecen referencias bibliográficas con fecha 2010, este hecho se produce porque la base de datos Web of Science indiza algunas revistas cuando aún no han sido publicados los números correspondientes, pero se encuentran disponibles en línea, por lo tanto, es posible que un trabajo publicado al final de 2009 pueda hacer referencia a otro trabajo que, pese a que no se ha publicado en papel, ya está disponible en línea y por tanto su fecha de publicación corresponderá a 2010. Tal es el caso de los siete trabajos que aparecen en la tabla de frecuencias 7.30.

En cuanto a la distribución de las referencias por año de publicación (figura 7.10), muestra la tendencia clásica en el área de medicina, con una gran acumulación de referencias en los años inmediatamente anteriores a la fecha de publicación, descendiendo rápidamente y produciendo el efecto conocido como “larga cola”.

La obsolescencia general, tanto en el cálculo de la vida media como en el del Índice de Price, muestra un envejecimiento progresivo de la literatura científica circulante. La vida media aumenta desde los cinco años en los registros publicados en el año 2000 hasta los siete años en los documentos publicados diez años después, en el 2009.

Presumiblemente, este aumento de la vida media se debe a la confluencia de dos patrones de investigación diferentes, que reflejan obsolescencias muy distintas, la

Genética y la Medicina en todas sus variantes. La Genética es un área con una obsolescencia muy alta y por tanto con una vida media muy corta, en torno a los cinco años, en la que la literatura científica envejece muy rápidamente, tal y como demuestran los trabajos de STINSON y LANCASTER (1987), MARTÍN-MORENO y SANZ-CASADO (1996), SANZ-CASADO y MARTÍN-MORENO (1997). Por otro lado, los trabajos relativos a las ciencias médicas poseen generalmente una vida media mucho mayor, que varía desde los seis hasta los diez años (LOPEZ PIÑERO y TERRADA, 1992b).

Este patrón de obsolescencia en las referencias bibliográficas se refleja también en el Índice de Price, que en este caso oscila entre el 52,42 % en el año 2000 hasta el 42,55 % en el 2009, indicando un aumento en el número de referencias con más de cinco años de antigüedad, hecho que está unido tanto a la diversidad de temáticas que se han mencionado anteriormente, como a la presencia de trabajos publicados entre los años 2001 y 2003, relativos a las causas genéticas de múltiples enfermedades como el Síndrome de Rett o el CADASIL (SERRANO-LÓPEZ y MARTÍN-MORENO, 2011), que provocan un efecto de envejecimiento en la literatura científica al acumular un gran número de referencias bibliográficas.

Para terminar con la caracterización general del área estudiada, se han examinado los resultados relativos a las citas que reciben los trabajos analizados. Estas citas, cuya distribución por fecha se observa en la tabla 7.41, muestran una tendencia decreciente, desde el primer año estudiado (2000) hasta el 2009. Si bien esta tendencia es la habitual se observa que en el año 2001 esta se rompe, al encontrarse un número de citas sensiblemente mayor a lo esperado, pasando de 57.518 citas en el 2000 a 61.546 en el 2001. Ahora bien, debido a que los artículos han tenido menos tiempo para ser citados, lo lógico sería que decreciera el número de citas, como efectivamente ocurre en los años posteriores. Esta ruptura en la tendencia de la serie temporal se repite en el gráfico 7.13, en el que se observa un aumento en el año 2001 en el promedio de citas por documento, respecto al año anterior y un fenómeno similar en el año 2004, cuando, a pesar de no superar el promedio

del año anterior si que se produce un descenso menos acusado de lo que cabría esperar al observar la tendencia decreciente de la serie.

En trabajos anteriores (SERRANO-LÓPEZ y MARTÍN-MORENO, 2011), se ha observado que estas rupturas en la tendencia de la distribución de citas por fecha se producen debido al efecto acumulativo de algunos artículos, que reciben un número exorbitado de citas debido a que se convierten en hitos que marcan un antes y un después en la investigación de determinadas enfermedades. Tal es el caso, por ejemplo de la Fibrosis Quística, que se encuadra en el grupo de enfermedades E83-E88 y que contiene algunos de los artículos con un mayor número de citas recibidas, entre ellos algunos trabajos relativos fundamentalmente al desarrollo de tratamientos más efectivos y al descubrimiento de los genes responsables de la enfermedad así como de posibles terapias génicas (PAI y NAHATA, 2001; WESTERMAN et al., 2004; CONWAY, 2004; HANSEN et al., 2005), muchos de los cuales fueron publicados en 2001 y 2004.

Este efecto acumulativo de algunos trabajos se comprueba al observar la distribución de documentos no citados por fecha (tabla 7.49), pues se observa que el porcentaje de documentos no citados se mantiene relativamente estable, salvo en los últimos años, que aumenta debido a la corta ventana de citación. A la vista de estos resultados se puede determinar que el hecho de que los documentos de 2001 y 2004 reciban una mayor número de citas no se debe a que estos años en concreto presenten un comportamiento anormal, sino a que durante estos periodos de tiempo se publicaron determinados trabajos que reciben un número de citas anormalmente alto.

Tradicionalmente se han relacionado determinados factores como el gasto en I+D+i, el capital humano dedicado a la investigación, la financiación de proyectos, etc. con el desarrollo de la investigación científica (SADOSKY, 1975; RIP, 1992; BLANCO MERLO y IRANZO, 2000; MUÑOZ, 2010), por lo que es conveniente la evaluación de la investigación realizada con dichos fondos y esfuerzo con el

objetivo de rendir cuentas a los actores sociales que la hacen posible (CAMÍ, 2001; LASCURAIN SÁNCHEZ, 2006; GÓMEZ CARIDAD et al., 2006).

Esta justificación de la inversión realizada cobra especial relevancia en el ámbito de las Ciencias de la Vida y de la Salud, dónde la inversión llevada a cabo por los países supone un porcentaje importante del Producto Interior Bruto (PIB) de cada uno de ellos (tabla 8.1). Por ello es necesario establecer prioridades en cuanto a los objetivos a abordar en esta investigación (CAMÍ, 2003) y más concretamente en aquellos puntos que inciden con más fuerza en la sociedad y en la calidad de vida de los ciudadanos. Entre ellas, por ejemplo están las enfermedades raras, de ahí que se hayan establecido diferentes políticas, tanto a nivel nacional como supranacional, con el fin de priorizar la investigación en este tipo de enfermedades, tal y como han señalado diferentes autores, entre ellos POSADA DE LA PAZ et al. (2002), WORLD HEALTH ORGANIZATION (2005), MINISTERIO DE SANIDAD Y POLÍTICA SOCIAL (2009), FIELD y BOAT (2011), FORMAN et al. (2012) y la propia UNIÓN EUROPEA (1999, 2000, 2007).

Volviendo a los actores que influyen en la investigación en Ciencias de la Salud, para este caso concreto, se han analizado dos aspectos determinados. Por un lado la incidencia del gasto general en salud, que realiza cada país, sobre su propia producción científica y por otro lado la influencia que poseen las asociaciones de pacientes y enfermos sobre dicha investigación. Para llevar a cabo estos análisis se han empleado técnicas de regresión tal y como han sido descritas en el apartado metodológico.

En el primer punto, la incidencia del porcentaje del PIB que destina cada país al gasto sanitario, frente a la producción científica, los resultados muestran una clara correlación entre el gasto sanitario y la producción científica, con un coeficiente de determinación (R^2) de 0,992 y un p-valor más que aceptable. Cabe destacar que estos resultados han sido obtenidos mediante la utilización del método de los mínimos cuadrados generalizados, en lugar de los mínimos cuadrados ordinarios,

que se utilizan habitualmente, debido al componente temporal presente en los datos.

Esta relación ha sido estudiada con anterioridad en múltiples trabajos en los que, al igual que en este, se encontró una relación directa con el gasto en salud y en algunos casos el PIB de los diferentes países, y la producción científica (CLARKE et al., 2007; HAINES y REDMAN, 2010; SPITZMUELLER et al., 2010; KING, 2004), entre dichos trabajos, cabe destacar el de LASCURAIN-SÁNCHEZ et al. (2008), debido a que en él, se relacionan numerosos factores como la morbilidad, mortalidad y el gasto en medicamentos, con la producción científica, siendo este último el menos correlacionado, a pesar de que efectivamente se encuentra una relación positiva.

En el caso del trabajo que se está discutiendo, los resultados muestran que existe una correlación positiva, año a año, entre el gasto sanitario y la producción científica en enfermedades raras. Ahora bien, esta relación cobra verdadera importancia cuando se introduce el país como variable cualitativa de la recta de regresión. Tal y como puede observarse en el gráfico 8.4, la relación entre las variables producción científica y gasto sanitario se hace mucho más patente cuando se tiene en cuenta la interacción entre esta última y el país estudiado, ya que cada uno de ellos presentan características diferentes que les otorgan un perfil diferenciador y por tanto en muchos casos no son comparables entre sí (KING, 2004). A pesar de lo cual se pueden establecer al menos tres perfiles diferentes de países con un volumen de producción y un porcentaje del PIB dedicado al sector sanitario similares. Entre estos perfiles se encuentran los grandes productores (Reino Unido, Francia, Alemania y Italia), que presentan una gran correlación entre las variables; a los medianos productores, que muestran unas rectas de regresión prácticamente idénticas (Holanda y España); y al resto de países de la Europa de los quince, que se pueden considerar como pequeños productores y en los que la relación entre producción y gasto es menor.

En cuanto al segundo factor social que se ha estudiado, la incidencia de la presencia de asociaciones de pacientes y enfermos en cada país estudiado sobre la producción científica en enfermedades raras, los resultados muestran una relación directa entre ambas variables, con un coeficiente de determinación ajustado del 0,757, indicando un ajuste de la recta de regresión más que aceptable, con una diagnosis correcta de los resultados obtenidos, a pesar de que se cuenta con tan solo quince puntos en el modelo de regresión (uno por cada país estudiado).

De nuevo se trata de una relación que ha sido estudiada con anterioridad, aunque con menor profusión que la relación gasto-producción y desde luego no en el marco de las enfermedades raras. Los autores de estos trabajos, han encontrado evidencias de una alta visibilidad de estas asociaciones en la producción científica (RABEHARISOA y CALLON, 2002; RANGNEKAR, 2005), así como “colaboraciones relevantes entre estas asociaciones y los organismos de investigación, en especial en las enfermedades raras y en las enfermedades con causa genética, que no tienen terapias disponibles, y en las que gran parte del conocimiento especializado sobre la enfermedad depende crucialmente de la información producida a partir de los relatos y registros de la experiencia vital con la enfermedad o de cuidar a aquellos que están afectados por ella” (ARRISCADO et al., 2007). En cuanto a la producción de medicamentos, tan importantes en el tratamiento y mejora de la calidad de vida de los que padecen enfermedades raras, los estudios de BOWER (1996, 2005) sugieren una relación activa entre los pacientes y las compañías farmacéuticas responsables de la producción de los medicamentos, cobrando un gran impacto en el proceso de innovación.

En el caso de las enfermedades raras, debido al volumen de enfermedades y al número de asociaciones de pacientes que estas generan, la presencia de las asociaciones como lobby de presión en el desarrollo de la investigación es mucho mayor, más si cabe al tener en cuenta que gran parte de estas enfermedades se encuentran estrechamente relacionadas con la genética, que es un campo de interés creciente desde que comenzó el desarrollo del Proyecto Genoma Humano (COLLINS et al., 2003). Por tanto no es de extrañar que los resultados obtenidos al poner en rela-

ción la presencia de asociaciones y la producción científica de cada país, muestren una correlación directa de tendencia creciente, de forma que a mayor número de asociaciones, mayor es la producción científica, si bien en esta correlación probablemente podrían encontrarse otro tipo de factores, la relación encontrada es clara y la diagnosis efectuada sobre la misma arroja resultados correctos.

Cabe destacar que, como se puede observar en la recta de regresión, cuanto mayor es el número de asociaciones (véase el caso de Francia) o mayor es la producción científica (por ejemplo la de Gran Bretaña), más se distancian estos puntos del modelo propuesto, debido a que son muchos otros factores los que influyen en ambas variables, sin embargo el modelo funciona extremadamente bien en el caso de los pequeños y medianos productores.

Por último señalar que, dado que la información sobre el número de asociaciones de pacientes y enfermos ha sido obtenido de ORPHANET (2011a), una base de datos que tiene su sede en Francia, se trata de una variable que cuenta con cierto sesgo hacia este país, haciendo que el número de asociaciones francesas sea superior al de Reino Unido a pesar de que su producción científica sea menor.

11.1. Países

Continuando con el análisis de los indicadores agregados por países, que ya se han introducido con las regresiones lineales, se va a comparar la producción de los países estudiados con otros estudios similares realizados tanto a nivel global en ciencias de la salud, como en el caso de algunos grupos de enfermedades en particular.

En términos generales, la producción de la Europa de los quince (UE-15) es comparable a la de Estados Unidos de América (USA) e incluso, en el último periodo estudiado por KING (2004), correspondiente a los años 1997 a 2001, dicha producción supera a la de USA, acumulando el 37,12 % de la producción mundial frente al 34,86 % de USA.

En estudios posteriores, CHINCHILLA-RODRÍGUEZ (2010) y ALBARRÁN y CRESPO (2010), mostraron una diferencia aún mayor entre la UE15 y USA, debido a que a la Unión Europea le ha afectado menos que a Estados Unidos el aumento de producción en los países asiáticos, especialmente China (ZHOU y LEYDESDORFF, 2006; LEYDESDORFF, 2012). En el área de Ciencias de la Vida y la Salud, las mayores diferencias entre los países de la Unión Europea y Estados Unidos estriban en que los centros de investigación de Estados Unidos están mucho más relacionados con la industria farmacológica así como con las empresas dedicadas a la biología molecular. Por ello su investigación tiene unos objetivos mucho más orientados a fines comerciales y tiene un mayor carácter de innovación que de investigación básica, al contrario que lo que ocurre en Europa, dónde la mayor parte de esta investigación la desarrollan las universidades y los centros nacionales de investigación. Por esta razón sus objetivos en la UE-15, están mucho más orientados al desarrollo de investigaciones básicas relativas al área de la biología y la biología molecular (OWEN-SMITH y RICCABONI, 2002).

Descendiendo a nivel de la producción científica en cada país, se observa que Gran Bretaña es, con diferencia, el país más productivo, seguido de Francia y Alemania (alternándose dependiendo de la base de datos en la que se mida), Italia, Holanda y España, seguidos del resto de países de la UE-15 que tienen menos de 1.000 documentos en todo el periodo (Bélgica, Suecia, Portugal, Austria, Grecia, Finlandia, Irlanda, Dinamarca y Luxemburgo). Esta estructura es la que aparece habitualmente en los estudios de la producción científica de los países europeos en Ciencias de la Salud efectuado en periodos similares de tiempo (CLARKE et al., 2007; DELNOIJ y GROENEWEGEN, 2007).

En cuanto a la evolución temporal de la producción por países, como se puede observar en las figuras 7.2 y 7.3, los países estudiados se distribuyen en tres grupos claramente definidos en función de su nivel de producción a lo largo de los diez años estudiados. Estos grupos se componen de países con una gran producción a lo largo de todo el periodo (Gran Bretaña, Francia, Alemania e Italia), países con una producción muy pequeña (Bélgica, Suecia, Portugal, Austria, Grecia,

Finlandia, Irlanda, Dinamarca y Luxemburgo) y dos países (España y Holanda) que se encuentran a media camino entre un grupo y el otro, aunque su tendencia creciente hace que al inicio del periodo estudiado (año 2000) se encuentren más cerca del grupo de los pequeños productores pero al final del mismo (año 2009) se distancien lo suficiente como para situarse a la cola del grupo de los grandes productores.

Este perfil de producción está en consonancia con otros trabajos relativos al análisis bibliométrico de enfermedades raras concretas, como por ejemplo los de SANZ-CASADO et al. (2007), SERRANO-LÓPEZ y MARTÍN-MORENO (2011) ó LÓPEZ-MUÑOZ et al. (2006) que muestran una producción emergente en el estudio de enfermedades raras en países como España, Holanda, Japón, Suiza o Brasil, pero difiere notablemente del perfil de publicación en Medicina General, dónde no se produce este agrupamiento, siendo el Reino Unido el principal productor seguido, a una distancia notable, de Holanda, Francia, Alemania e Italia (MAN et al., 2003).

Por tanto, el perfil de publicación por países es muy similar a otros estudios realizados en áreas y enfermedades concretas, pero difiere en gran medida de los perfiles de producción encontrados en el área de Medicina General, dónde el agrupamiento por países mencionado anteriormente no se produce en absoluto, apareciendo Reino Unido como el gran productor de información científica, muy alejado de los demás países.

En cuanto a la producción de cada país en las dos bases de datos estudiadas (SCI y Pubmed), se producen dos efectos relativos a las particularidades de cada base de datos (PESTAÑA, 1997; COSTAS et al., 2008), tal y como se muestra en la figura 9.5. Por un lado, las limitaciones de Pubmed a la hora de recoger las filiaciones institucionales de los autores, hace que las estrategias de búsqueda recuperen solo aquellos documentos que están firmados por autores de instituciones de un país únicamente. Esto significa que en dicha base de datos solo se recuperan aquellos documentos en los que las instituciones europeas aparecen como primer

firmante. Por esta razón cabría esperar una mayor producción de estos países en SCI, ya que esta base de datos si recoge la filiación completa de los autores y por tanto contendría algunos que nunca se recuperarían en Pubmed. Sin embargo, la producción de los diferentes países en SCI y Pubmed sigue una distribución relativamente equidistante, presentando las mayores discrepancias en el caso de Luxemburgo, debido a su escasísima producción, Bélgica, Dinamarca y Portugal. Por otro lado, la distribución de Reino unido presenta cierto sesgo hacia Pubmed, debido principalmente a que los autores de este país aparecen como primer firmante en un gran número de trabajos escritos en colaboración con otros países, principalmente Irlanda.

Este hecho se debe al segundo efecto que se produce en esta distribución, y es que a pesar de recoger solo la filiación del primer autor firmante, Pubmed recoge un conjunto de revistas nacionales mucho mayor que SCI, por lo tanto, la proporción de países que no aparecen como primer firmante en trabajos publicados en revistas internacionales se compensa con la producción de los mismos en revistas nacionales e incluso, en algunos casos como el de España, Gran Bretaña o Portugal, la proporción de documentos publicados en revistas nacionales y, por tanto su presencia en Pubmed, es relativamente mayor que en las revistas internacionales que recoge la Web of Science.

En la distribución de la producción por idiomas ya se observó que la mayor parte de la producción científica se encontraba en inglés, seguido de francés, alemán y español. En la figura 10.14 se puede ver esta distribución más detallada en un gráfico de mosaico, en el que se muestra la distribución idiomática por países. En este gráfico, además de observar la proporción de documentos en cada idioma en función del país productor, se cuenta con el complemento de la distribución por colores, que indica la sobrerrepresentación (en azul) o infrarrepresentación (en rojo) de los diferentes idiomas en cada país, de forma que cuando más intenso es el color que representa la combinación país-idioma, mayor es la discrepancia de la misma respecto a una teórica distribución equidistante entre países e idiomas.

En este caso, se puede observar que, como era de esperar, los idiomas locales están mucho más representados en los países en los que se hablan de forma nativa, destacando el caso de Francia, con una proporción de documentos en francés mayor que la de otros países y sus idiomas locales. Por otro lado la categoría que asigna la Web of Science a los idiomas sin identificar representa una proporción mayor de lo esperado en dos países: España y Alemania, por lo que se puede deducir que se trata de idiomas locales, correspondientes a los de las diferentes comunidades autónomas españolas y a los cantones alemanes, que no han sido correctamente identificados.

En cuanto a los índices de colaboración que muestran los países, tanto a nivel institucional (tabla 7.14) como entre autores (tabla 7.17) se observa un promedio de colaboraciones mayor en los países menos productivos, tales como Finlandia, Portugal, Dinamarca o Bélgica, mientras que algunos de los mayores productores (Gran Bretaña y Alemania) poseen los índices de colaboración y coautoría más bajos. Este comportamiento ya fue descrito por LUUKKONEN et al. (1992), que llegó a la conclusión de que el menor desarrollo tecnológico de los países (especialmente en áreas experimentales como la biología, bioquímica o la genética), provoca una búsqueda de colaboración por parte de éstos, con países con una mayor capacidad productiva y un mayor desarrollo tecnológico, con el fin de poder desarrollar investigaciones que de otro modo serían inviables.

Este puede ser el caso de Francia, España y Holanda, tres países que poseen valores intermedios tanto de colaboración institucional como de coautoría, a pesar de que el primero es el segundo con mayor producción y los otros dos países son productores emergentes, por lo que se puede señalar, al igual que se hace en otros trabajos (NAVARRO y MARTÍN, 2004; BARJAK y ROBINSON, 2008a), que la colaboración internacional, especialmente en el caso de las enfermedades raras, ayuda a los países a convertirse en grandes productores, mientras que aquellos que han alcanzado este estatus tienden a disminuir su proporción de documentos en colaboración.

En cuanto a los patrones de colaboración, las tablas 7.21 y 7.23 muestran que los países con mayores tasas de colaboración presentan un mayor porcentaje de documentos en colaboración internacional, mientras que aquellos que colaboran menos mantienen un equilibrio entre el número de documentos en colaboración nacional y el de documentos en colaboración internacional.

Llama la atención el caso de España, puesto que tratándose de uno de los países con mayor proyección en cuanto a producción científica (junto con Holanda) y teniendo unos índices de colaboración y de coautoría aceptables (3,88 instituciones y 7,41 autores por documento), se sitúa en una posición más baja de lo esperado en el porcentaje de documentos en colaboración internacional, colaborando fundamentalmente con instituciones nacionales. Aplicando el modelo de LUUKKONEN et al. (1992), este perfil de colaboración supone que España ha alcanzado un nivel de desarrollo tecnológico que capacita al país para realizar sus propias investigaciones y no depende tanto de la colaboración internacional. Ahora bien, ante la presencia emergente de países como Holanda, Bélgica, Austria, Portugal o Dinamarca, España corre el riesgo de perder la tendencia creciente en cuanto a producción científica en las enfermedades estudiadas en detrimento de los países mencionados.

Estos perfiles de colaboración pueden observarse claramente a través de los indicadores conexionistas (figuras 9.5 y 9.6, del escalado multidimensional (figura 10.12) y del análisis de clusters (figura 10.13). Los indicadores conexionistas muestran, en primer lugar una gran colaboración entre los países del núcleo duro investigador en Europa, que coinciden prácticamente con los quince países estudiados, con la excepción de Luxemburgo, que tiene una producción muy pequeña. Entre los países que más colaboran con los países estudiados, USA se encuentra en el primer lugar, y su nodo es el más grande debido a que colabora prácticamente con todos los nodos de la red, y después se encuentran países como China (CHN), Brasil (BRA), Argentina (ARG), Canadá (CAN), Turquía (TUR), Polonia (POL) o Israel (ISR), que también poseen una colaboración intensiva con los países de la Europa de los quince.

Como se indicó en el capítulo de resultados y como puede observarse en el escalado multidimensional (figura 10.12), los perfiles de colaboración sitúan a Gran Bretaña y a Alemania prácticamente solapados en la primera dimensión del análisis y completamente alejados respecto a la segunda. Estas posiciones se explican por el hecho de que ambos son los países más productivos y que más colaboran, pero poseen perfiles de colaboración muy diferentes, tal y como se muestra en el análisis de clusters y en la segunda red social de países (figura 9.6), que muestra la colaboración entre los quince países estudiados. Como puede observarse, Gran Bretaña y Alemania se separan rápidamente en el análisis de cluster, indicando perfiles muy distintos, y observando las relaciones de cada uno con sus respectivos nodos, se observa que la diferencia entre ambos es clara, puesto que Alemania colabora fundamentalmente con los países más productivos y solo colabora con uno de los países de menor producción (Austria), mientras que Gran Bretaña colabora mucho más con países como Irlanda, Suecia y Finlandia. Pero además, en éste se produce el efecto maestro-pupilo (descrito por ZITT (2000)) con Irlanda mientras que Alemania colabora menos con los países de la Europa de los quince, salvo con Austria, y tiene un alto índice de colaboración con los países de Europa del Este (no incluidos en este trabajo), por lo que tal y como indican GLÄNZEL y SCHUBERT (2004), puede ser considerada como “la puerta del oeste”.

Este análisis de la colaboración, especialmente a nivel internacional, coincide con los resultados obtenidos a través del índice de aislamiento (tabla 7.35), que muestran unos valores medios de aislamiento en los países estudiados, salvo en el caso de Gran Bretaña, que presenta un aislamiento mucho mayor debido principalmente a dos factores: En primer lugar, una gran parte de las revistas internacionales de medicina son de nacionalidad británica, por lo que los investigadores del Reino Unido no necesitan publicar en revistas extranjeras para que su trabajo tenga difusión, sino que pueden publicar en revistas nacionales. Por otro lado, relacionando este indicador con la colaboración internacional, se observa que una gran parte de los documentos que publican los investigadores británicos lo hacen en colaboración con instituciones irlandesas y es común que dichos trabajos aparez-

can publicados preferentemente en revistas del Reino Unido, tal y como muestran los estudios realizados por JONES et al. (2004, 2006, 2007) a través de encuestas sobre los colectivos británicos de Pedriatría, Cirugía y Psiquiatría.

Concluyendo con el análisis de la producción, la relación entre países y temáticas, representadas por la clasificación ICD-10, (figura 10.15), muestra sesgos muy claros de determinados países hacia algunos grupos de enfermedades. Tal es el caso de Irlanda que produce en el grupo E83-E88 (diferentes tipos de desórdenes metabólicos) mucho más que los demás países o el de Finlandia, mucho más cercana al grupo F01-F84 (trastornos mentales y del comportamiento) que el resto, si bien el núcleo de grandes productores (Reino Unido, Alemania, Italia y Francia) se encuentran muy cercanos al eje central de análisis y por tanto no presentan sesgo hacia ninguna enfermedad en concreto.

Para terminar, con el análisis de los quince países abordados en este trabajo, se va a estudiar su distribución de citas y, por tanto, el impacto que genera la investigación sobre enfermedades raras que se lleva a cabo en cada uno de ellos. En primer lugar, la distribución del total de citas (figura 7.12) y total de citas por documento (tabla 7.45) en valores agregados por país, muestran unos valores que contrastan con los datos analizados de producción por países, ya que indican que los países más productivos no se corresponden con los que producen un mayor impacto en sus investigaciones.

En este caso, hay que señalar que países como Dinamarca o Bélgica, cuya producción no es significativa, reciben un mayor número de citas por documento que los grandes productores (Gran Bretaña, Alemania, Francia e Italia). Sin embargo, en el caso de los países con una producción emergente (Holanda y España), se observan dos perfiles muy diferentes, puesto que Holanda escala posiciones en el número de citas por documento y se sitúa a la altura de Alemania y Reino Unido, mientras que España se sitúa a la cola, con un promedio de citas por documento inferior al de países como Portugal o Grecia.

El mismo perfil se reproduce en el porcentaje de documentos no citados (tabla 7.50), con la salvedad de que los grandes productores descienden aún más, produciendo un mayor número de documentos sin citas que los países nórdicos, como Dinamarca, Suecia, Holanda o Finlandia. Este perfil de publicación, con un bajo número de publicaciones, comparado con el de otros países europeos como Reino Unido, Alemania o Francia, pero con un mayor número de citas por documento, está en consonancia con los resultados encontrados por INGWERSEN y WORMELL (1999) y GLÄNZEL (2000), en los que destacaban la evolución de los países escandinavos, especialmente en las áreas relacionadas con las Ciencias de la Vida y la Salud, produciendo un número relativamente pequeño de documentos pero siendo estos de gran calidad y generando un número significativo de citas y por tanto un gran impacto sobre la comunidad científica.

Por otro lado, llama la atención el caso de Francia, que desciende hasta un porcentaje de documentos sin citas (8,76%), muy cercano al de España (9,44%), que es el último país de la lista. Este patrón concuerda con el trabajo previo realizado por LEEUWEN et al. (2001), en el que se concluía que los países con mayor número de revistas en idiomas locales (como Francia o España), sufren un sesgo significativo a la hora de recibir citaciones, debido a que los sistemas de fomento de la actividad científica promueven la publicación en revistas de impacto, es decir, que estén incluidas en el Journal Citation Reports de Thomson Reuters (MUÑOZ, 1998; JIMÉNEZ-CONTRERAS, 2002; WEINGART, 2005).

11.2. Temática

Antes de comenzar con la discusión de los resultados relativos a las temáticas (en diferentes niveles de agregación), es necesario señalar que este apartado cuenta con un sesgo hacia el área de la Genética, debido a que la muestra escogida para estudiar la actividad científica en el área de las enfermedades raras está es-

pecialmente marcada por la investigación genética, pues todas las enfermedades estudiadas poseen una causa genética claramente identificada.

Dicho esto, los primeros resultados relativos a la temática de los trabajos publicados, se encuentran en las tablas 7.6 y 7.7, que muestran la producción científica agregada por las diferentes clasificaciones ICD-10. En estas tablas, se observa que, en ambas bases de datos, las enfermedades con una mayor producción científica son las relativas a los desórdenes metabólicos (E83-E88), incluida la fibrosis quística, que como ya se indicó anteriormente, acapara gran parte de la producción científica, seguida de las dismorfias congénitas (Q80-Q89), los trastornos primarios de los músculos (G71), los desórdenes del metabolismo de esfingolípidos y otros desórdenes de almacenamiento lipídico (E75) y los desórdenes del metabolismo de los aminoácidos (E72). La mayor parte de estas enfermedades se encuentran relacionadas con los trastornos metabólicos, que tal y como otros estudios han puesto de manifiesto (HEEMSTRA, 2009; SANZ-CASADO et al., 2007), despiertan especialmente el interés de los investigadores en Ciencias de la Vida y la Salud.

La producción sobre el resto de enfermedades varía entre las dos bases de datos, en ambas tiene un volumen similar (algo mayor en el caso de Pubmed debido a su mayor cobertura), salvo en el caso de la Osteonecrosis (M87), que en el caso de Pubmed se sitúa en torno a la media de producción, con 567 documentos, mientras que en el SCI está en la cola de la actividad científica, con tan solo 67 documentos. Este comportamiento en el caso de la Osteonecrosis, se debe a que gran parte de los trabajos relativos a esta enfermedad se publican en revistas de Odontología, que no están especialmente bien representadas en la Web of Science (77 revistas en 2010), frente a las áreas con mayor producción, que como muestra la tabla 7.8, se corresponden con la Genética (156 revistas), la Biología Molecular (286 revistas) o la Neurología y las Neurociencias (239 revistas).

Por otro lado la distribución de términos MeSH en ambas bases de datos (tablas anexas B.1 y B.2) muestra una presencia destacada de términos relativos a la

investigación genética y a diversas enfermedades relacionadas con el metabolismo, aunque el término más destacado en la distribución de ambas bases de datos es, sin duda, la fibrosis quística, pues esta enfermedad es una de las enfermedades raras con una prevalencia mayor (de uno a cinco casos por cada 10.000 habitantes), pero no lo suficiente como para dejar de considerarla rara (ORPHANET, 2011a; HEEMSTRA, 2009).

En cuanto a la colaboración entre instituciones (tabla 7.16) y autores (tabla 7.19), las temáticas en las que existen unos mayores índices de colaboración son la Genética, Neurociencia-Neurología, Gastroenterología-Hepatología y Bioquímica-Biología Molecular. Estas áreas tradicionalmente han tenido unos índices de colaboración relativamente altos en el conjunto de las Ciencias de la Vida y de la Salud (GÓMEZ, 1995), pero desde que la Genética cobró todavía más importancia, debido al desarrollo del mapa del genoma humano, dado a conocer en 2003, la colaboración entre genetistas y científicos que estudian enfermedades concretas se ha disparado (HANNEMANN-WEBER et al., 2012).

A través de las redes sociales de las figuras 9.7 y 9.8 se puede observar como se produce esta colaboración entre las diferentes enfermedades clasificadas por medio de la ICD-10. Como ya se pudo apreciar en el apartado de resultados, la red social realizada sobre la base de datos SCI es mucho más representativa que la realizada sobre Pubmed, debido a que el número de documentos que comparten temática es mucho mayor, mostrando relaciones muy interesantes, unas esperables, como por ejemplo la que se produce entre los distintos trastornos metabólicos (E70 a E88) y también otras, a primera vista sorprendentes, como por ejemplo la estrecha relación existente entre los desórdenes del metabolismo de los aminoácidos aromáticos (E70) y las enfermedades degenerativas del sistema nervioso (G24-G40). Sin embargo dichas relaciones poseen una clara explicación, debido a que los aminoácidos ejercen como unos de los principales neurotransmisores del sistema nervioso, por lo que los problemas en su metabolización afectan irremediablemente al desarrollo del sistema nervioso (FLOREZ y PAZOS, 1987; GARCÍA SILVA, 2000).

Por otro lado, observando la distribución de los términos MeSH en el apartado de indicadores multidimensionales, tanto en el caso de Pubmed como del SCI y tanto en los escalados multidimensionales (figuras 10.8 y 10.9) como en los análisis de clusters (figuras 10.10 y 10.11), se puede observar que los términos se encuentran muy agrupados entre sí, muy cerca unos de otros, permitiendo detectar comportamiento singulares. Concretamente este es el caso de los términos relativos a la Fibrosis Quística, por lo motivos que ya se han mencionado anteriormente, o de la Hemocromatosis y los términos relacionados con esta enfermedad, como el hierro (Iron) o el exceso del mismo (Iron Overload). En este caso, la explicación más razonable para este fenómeno, es el hecho de que tanto la Diabetes Mellitus como la Osteonecrosis están estrechamente relacionadas entre sí, así como con el exceso de hierro (SHLOSSMAN et al., 1990; FERNANDEZ-REAL et al., 2002; HERRING y SHAH, 2006; KOSTOFF et al., 2009; FAVUS, 2007; KHAMAISI et al., 2007), presentando por tanto un comportamiento diferente al resto de enfermedades raras, que normalmente no suelen encontrarse muy relacionadas con enfermedades tan comunes y con un bagaje investigador tan importante como el de la Diabetes.

En cuanto a las referencias bibliográficas que utilizan los investigadores que trabajan en el ámbito de las enfermedades raras, las áreas más destacadas (clasificadas a través del JCR en la tabla 7.34) se corresponden con las mismas áreas que aparecían como las más productivas (Bioquímica-Biología Molecular, Genética y Neurología), destacando también la posición de la categoría “Multidisciplinary”, dentro de la cual se encuentran la revistas “Nature” y “Science”, dos de las revistas que publican un mayor número de trabajos relacionados con la Genética, entre ellos los que dieron a conocer el mapa del genoma humano (COLLINS et al., 2003).

Por último, los indicadores relativos al impacto de las publicaciones (documentos en el primer cuartil, total de citas, citas por documento y porcentaje de documentos no citados), muestran una clara correspondencia entre las enfermedades con mayor producción y las que más impacto generan en la comunidad científica,

si bien destaca sobremanera el caso de la Osteonecrosis (M87), que a pesar de ser la enfermedad con menos producción en el SCI en todo el periodo (tan solo 67 trabajos), se encuentra entre las que más citas por documento reciben (tabla 7.46), reforzando la hipótesis de que su escasa presencia se debe a que gran parte de los trabajos sobre esta enfermedad se publican en revistas de Odontología, con escasa cobertura por parte de la Web of Science, lo cual no significa (como demuestra el número de citas por documento), que dicha enfermedad no genere interés e impacto en la comunidad científica.

11.3. Fuentes

En el caso de las fuentes que los investigadores utilizan para publicar sus trabajos sobre enfermedades raras, la distribución se ajusta claramente a la propuesta teórica de Bradford, tal y como puede observarse en los gráficos 7.6 y 7.7 relativos a la distribución de trabajos por revista en las bases de datos SCI y Pubmed, así como en sus respectivas tablas 7.25, 7.26, B.6 y B.7, que recogen la distribución de las revistas y el núcleo principal de las mismas que acumula más del 33% de la producción científica.

Este núcleo de revistas se compone en el caso de la base de datos de Thomson-Reuters de 46 revistas que acumulan el 33,12% de la producción científica, mientras que en el caso de Pubmed son 51 revistas las que forman el núcleo, recogiendo el 33,1% de los trabajos publicados. Al igual que en otros trabajos en los que se estudió la distribución de los trabajos científicos en revistas de Medicina (LOPEZ PIÑERO y TERRADA, 1992b,c; ESCUDERO GÓMEZ et al., 2005; CAMÍ et al., 1997; DALPÉ, 2002), en ambas bases de datos se cumple claramente la distribución de Bradford y se trata de un grupo muy pequeño de revistas, pues el conjunto total de revistas asciende a 2.083 revistas en los datos obtenidos de SCI y a 2.325 en el caso de Pubmed.

Posteriormente se llevó a cabo un análisis de la concentración temática en las revistas del núcleo de la base de datos SCI. Este análisis, visible en la tabla 7.27, muestra un alto índice de concentración en las revistas, indicando una gran especialización de las mismas, ya que tan solo 12 revistas están incluidas en dos o más categorías JCR.

Esta gran especialización de las revistas, unida el hecho de que estas 46 revistas acumulan más de un tercio de la producción científica sobre enfermedades raras, indica que se trata de un área cada vez más consolidada, en la que se está produciendo un fenómeno similar al descrito por TORTOSA GIL (1989) en la Psicología, mostrando un patrón de publicación cada vez más especializado y, en este caso, orientado fundamentalmente a las revistas de Genética. Ahora bien, en el núcleo también aparecen muchas revistas dedicadas al estudio del aparato respiratorio, debido a la influencia de la Fibrosis Quística, a la neurología y neurociencia, a las enfermedades metabólicas y a la Pediatría, estando esta última área presente debido a que la mayoría de las enfermedades raras son diagnosticadas en las primeras etapas del desarrollo humano (ESCUDERO GÓMEZ et al., 2005).

Entre las revistas que conforman el núcleo en SCI, destaca una revista española, Revista de Neurología, con 99 artículos en todo el periodo, situándose entre las 30 primeras en cuanto al número de trabajos publicados y destacando por su posición entre una gran mayoría de revistas anglosajonas

Debido al proceso de especialización en el que se encuentra sumida el área de la investigación en enfermedades raras, la distribución de Bradford aplicada a las revistas referenciadas, muestra una clara preferencia por parte de los investigadores a la hora de utilizar revistas mayoritariamente de Genética o revistas generalistas relativas a las áreas médicas, así como revistas multidisciplinares, como es el caso de Nature o, en menor medida, Science. Por otro lado, la preferencia en las revistas referenciadas es muy clara, puesto que se cumple sobradamente la Ley de Bradford y ésta es apreciable tanto en su representación gráfica como en la tabla B.8, que contiene el número de referencias por revista. Éstas permiten determinar

la existencia de un núcleo de revistas, formado por 29 publicaciones que recogen el 33,42% de las referencias bibliográficas contenidas en los trabajos analizados, sobre un total de 676.147 referencias analizadas y recogidas en 24.693 títulos de revista.

Como se puede observar, el núcleo de revistas referenciadas es realmente pequeño, comparado con el volumen total de referencias, a pesar de lo cual no difiere sustancialmente de los patrones habituales en el área de las Ciencias de la Vida y la Salud (OJASOO y DORÉ, 1999; GLÄNZEL y SCHOEPFLIN, 1999), especialmente en lo que a la Genética y a las enfermedades raras se refiere (ESCUADERO GÓMEZ et al., 2005).

Cabe destacar que el núcleo de revistas referenciadas no coincide exactamente con el de las revistas en las que se publican la mayor parte de los trabajos, si no que los autores publican la mayoría de sus trabajos en revistas de menor impacto que las que referencian, como se verá posteriormente en el análisis del factor de impacto de las revistas del núcleo. Estos datos pueden considerarse coincidentes con la denuncia que realizan RUANO-RAVINA y PÉREZ-RÍOS (2012), acerca de las dificultades para publicar trabajos sobre enfermedades raras en revistas de primer orden, debido a que estas los rechazan para evitar que dichos trabajos devalúen el factor de impacto de la revista.

La posición de estas revistas (que forman el núcleo en la distribución de SCI), se ha relativizado a través del Factor de Impacto (FI), obtenido a través del Journal Citations Report. En la tabla 7.54 se puede observar como se posicionan dichas revistas, de mayor a menor Factor de Impacto Medio. Entre las revistas mejor posicionadas se encuentran publicaciones relativas al área de Genética y revistas generalistas de Medicina. Todas ellas tienen en común que no son aquellas revistas que ocupan las primeras posiciones en la distribución de trabajos por revista (a pesar de pertenecer al núcleo), y sin embargo sí que se encuentran contenidas en el núcleo de referencias por revista, apoyando de nuevo la hipótesis de RUANO-

RAVINA y PÉREZ-RÍOS (2012) relativa a la dificultad de publicar trabajos sobre enfermedades raras en revistas de alto impacto.

Por otro lado, se ha estudiado la distribución por cuartiles de estas revistas. La tabla 7.55 muestra las principales revistas en las que se publican la mayor parte de los trabajos (SCI), así como el idioma en el que se encuentran dichos trabajos, el país de origen de la revista, así como el mejor cuartil posible dentro del periodo de estudio. En este punto se debe destacar que el cuartil no ha sido calculado de la misma forma que lo presenta Thomson-Reuters en la Web of Science, ya que en este caso se ha tenido en cuenta el número de trabajos publicados por año, y no el número de revistas, tal y como propone la formulación original. Como ya se explicó en la metodología, este tipo de asignación es mucho más restrictiva y establece un primer cuartil más pequeño, en el que se recogen tan solo las grandes publicaciones dentro del área concreta, mientras que los cuartiles más bajos recogen un mayor número de revistas que en la formulación original.

La distribución de las revistas por cuartil muestra que en el primer cuartil existen una gran preponderancia de revistas estadounidenses y británicas, encontrándose tan solo una revista de otra nacionalidad, holandesa. En el segundo y tercer cuartil se encuentran más revistas holandesas, tres en total, junto a tres revistas de diferente nacionalidad cada una, alemana, noruega y suiza, todas ellas con el inglés como idioma de publicación. Hasta el cuarto cuartil no aparecen revistas de un idioma diferente. Se trata de dos revistas francesas (Archives de Peadiatrie y Revue des maladies respiratoires) y una española (Revista de Neurología). Las tres revistas publican en sus respectivos idiomas, por lo que se ha pensado que podría existir relación entre el cuartil en el que se encuentran las revistas y las variables relativas al país de publicación y al idioma de los textos. Para comprobarlo se aplicaron sendos contrastes del χ^2 , demostrando que, efectivamente el cuartil de posicionamiento de las revistas presentaba una clara relación de dependencia respecto a las variables nacionalidad e idioma de las publicaciones, tal y como muestran estudios similares tanto a nivel de Ciencias de la Vida y de la Salud

como a nivel general de las publicaciones en la Web of Science (DE GRANDA ORIVE, 2003; BORDONS, 2004; CASTILLO-ESPARCIA, 2012)

En el análisis de las temáticas, ya se observó que determinadas clasificaciones, se comportaban de forma sensiblemente diferente a las demás, tal era el caso de la Fibrosis quística (E83-E88) y la Osteonecrosis (M87). En el caso de esta última, este comportamiento se produce de nuevo en la distribución de los trabajos por revista y clasificación ICD (figuras 10.16, 10.17, 10.18 y 10.19), como se puede observar en los análisis de correspondencias realizados sobre ambas bases de datos. Concretamente, el análisis muestra un comportamiento singular de las revistas “Journal of Clinical Endocrinology Metabolism” y “Hormone Research”, que se sitúan en una posición más cercana a la de las clasificaciones M87 y E16-E59, que al conjunto de revistas de la mitad superior, mucho más cercanas entre si. También cabe destacar que este efecto se produce respecto a la segunda dimensión del análisis. Respecto a las ampliaciones de ambos análisis, se puede observar como la principal diferencia en la primera dimensión, que es la que más porcentaje de la variabilidad explica, se muestra en la clasificación E83-E88, que como ya se ha explicado, corresponde mayoritariamente a la Fibrosis quística. Ésta atrae hacia ella a un gran número de revistas, principalmente relativas a la Pediatría, Genética y al Aparato Respiratorio, que recogen la practica totalidad de los trabajos sobre esta enfermedad, como muestran los trabajos de SHAW (1990b,a, 1991b,a) y SHAW y WOOD (1991).

Por último, para terminar con el análisis de las fuentes utilizadas para publicar la investigación sobre enfermedades raras, se ha realizado un análisis de correspondencias, sobre los datos obtenidos de SCI, entre los países que realizan los trabajos y las nacionalidades de las revistas en las que estos se publican (figura 10.20 y su ampliación en la figura 10.21).

En este análisis se observa como, respecto a la primera dimensión, España es el único país que se comporta de forma diferente al resto, pues mientras la mayoría de los países se encuentran agrupados en torno a las revistas de nacionalidad

anglosajona (principalmente USA, Reino Unido y en menor medida los Países Bajos), España se acerca mucho más a las revistas nacionales, separándose del resto.

El mismo fenómeno ocurre con Francia en la segunda dimensión, con la salvedad de que el porcentaje de variabilidad explicada por la segunda dimensión es mucho menor, por lo que la diferencia con el resto de países es también menor. En el detalle se puede observar más detenidamente como las revistas que acaparan la mayor parte de la producción son estadounidenses, británicas u holandesas. Así mismo se observa que Portugal se separa mínimamente de la distribución del resto de países, acercándose tímidamente a las publicaciones de nacionalidad española, probablemente debida a la proximidad geográfica e idiomática.

Este perfil concuerda con el trabajo de VAZQUEZ VALERO y GONZALO BALMISA (1997), cuyos resultados muestran que las revistas españolas aparecen como las segundas más solicitadas, muy cerca de las estadounidenses (24,02 % de las españolas frente a 24,17 % de las estadounidenses), aunque la tendencia a publicar en revistas nacionales ha ido disminuyendo con el paso del tiempo (GÓMEZ CARIDAD et al., 2006).

11.4. Instituciones

La producción en el caso de las instituciones muestra un claro predominio de los hospitales, que se sitúan como los principales productores de información científica, duplicando el número de artículos publicados por las universidades o los centros nacionales de investigación. En cuanto a los centros de investigación de naturaleza privada, como empresas y fundaciones, la representación es mínima, pues tan solo 307 documentos cuentan con la participación de este tipo de centros.

La presencia dominante de los hospitales como grandes productores de información concuerda con la distribución habitual en medicina y biomedicina mostrada

por GÓMEZ CARIDAD et al. (2004). Sin embargo, el comportamiento del sector privado en el caso de las enfermedades raras difiere notablemente del que presentan en otros ámbitos de la investigación biomédica. Por ejemplo en el caso de la innovación en biotecnología, tal y como señalan MCMILLAN et al. (2000), el sector privado posee un gran peso. Mientras que otros autores (DRESSER, 2006; COCKBURN, 1996), que estudiaron las implicaciones de la industria privada en el desarrollo de fármacos y en los ensayos clínicos, concluyeron que los intereses de la industria farmacéutica están lejos del desarrollo de los denominados “medicamentos huérfanos”, debido a que su desarrollo supone una gran inversión en recursos humanos, tecnológicos y monetarios, así como un proceso largo y complicado para conseguir la aprobación y comercialización de los mismos. Esto hace que la mayor parte de los desarrollos relativos a medicamentos huérfanos provengan de spin-offs de universidades u hospitales dedicados a la investigación clínica (HEEMSTRA, 2009).

Por lo tanto, el desarrollo de este tipo de medicamentos no despierta el interés de las empresas, pues los beneficios que pueden obtener de su comercialización en muchas ocasiones no superan los gastos generados en su desarrollo. Por tanto, solo se implican en su desarrollo cuando se deriva de la colaboración con el sector público. De ahí el alto porcentaje de documentos producidos por empresas y fundaciones en colaboración, tanto nacional como internacional, con hospitales, universidades y centros nacionales de investigación (tablas 7.20 y 7.22).

En cuanto al impacto que genera la investigación por parte de los diferentes centros, de nuevo son los hospitales los que producen un mayor número de documentos en revistas del primer cuartil, seguidos por los centros nacionales de investigación, las universidades y las empresas y fundaciones. En todos los casos el número de documentos en el primer cuartil supone más del 50% de los documentos publicados por las diferentes instituciones, por lo que el impacto de estas publicaciones se puede considerar bastante alto.

Por otro lado, el análisis de citas muestra que, aunque los hospitales, debido al gran número de documentos que publican, son los que reciben un mayor volumen total de citas (253.196 en todo el periodo), los centros nacionales de investigación destacan por su posición, por delante de las universidades con casi 13.000 citas más, a pesar de que ambos poseen un volumen de documentos publicados muy similar.

Al relativizar estas cifras, dividiendo el número de citas entre el número total de documentos publicados por cada tipo institucional, el cambio es mucho mayor. Ahora aparecen los centros nacionales como los que más citas por documento reciben, muy por delante de los hospitales, que se sitúan como los que menos citas por documento generan, mientras que las empresas y fundaciones poseen un volumen de citas por documento muy similar al de las universidades.

Este tipo de distribución contradice los resultados publicados por ROMERO et al. (2009), en los que mostraban que los ensayos clínicos, realizados mayoritariamente en hospitales, reciben significativamente más citas que otros tipos de publicaciones, por lo que sería de esperar que los hospitales contaran con un volumen de citas por documento mucho mayor y no que se sitúen a la cola de la distribución. Sin embargo, en este caso y debido a las particulares características que presentan las enfermedades raras, los ensayos clínicos no son uno de los tipos documentales con mayor producción, pues representan tan solo el 2,24 % de los trabajos estudiados, por lo que no se produce el efecto de acumulación de citas que indica ROMERO et al. (2009), y los hospitales reciben muchas menos citas de las que podrían producirse en la investigación sobre otro tipo de enfermedades, en detrimento de los centros nacionales de investigación.

Una vez caracterizada la actividad científica en torno a los tipos institucionales, se va a descender a analizar los datos de producción, consumo e impacto de la actividad científica generada por las principales instituciones responsables de la mayor parte de la producción científica en el ámbito de las enfermedades raras.

En primer lugar, para determinar cuáles son estas instituciones se aplicó el método de normalización descrito en el apartado metodológico para reducir las filiaciones institucionales a la institución matriz o institución principal en todos los casos, renunciando, debido a su complejidad en un conjunto tan vasto de documentos, a realizar un análisis pormenorizado a nivel de departamentos.

Una vez conformado el conjunto de instituciones presentes en los trabajos de investigación, se procedió a someterlas a una distribución de tipo Lotka, con el fin de determinar cuáles son aquellos centros que acumulan un mayor porcentaje de documentos publicados, dando como resultado las instituciones de la tabla 7.10, que recoge aquellos centros de investigación que acumulan el 33,67% de los trabajos publicados en el SCI sobre enfermedades raras.

Lo primero que se observa, en las instituciones que conforman el núcleo con mayor producción, es que pertenecen fundamentalmente a tres países: Gran Bretaña, Alemania y Francia, si bien también aparecen algunas instituciones de Italia y Holanda y tan solo una de Bélgica y Finlandia. Cabe destacar que los tres primeros países forman parte de los denominados “grandes productores”, sin embargo Italia, tiene muy poca representación (tan solo 2 instituciones) entre los centros más productivos, al igual que Holanda, que se sitúa entre los países con un mayor crecimiento a lo largo de todo el periodo, junto con España, que sin embargo no aparece representada entre las instituciones con mayor producción.

Con respecto a la presencia de una institución belga, hay que señalar que Bélgica cuenta con la segunda institución más productiva, la Universidad Católica de Lovaina, por detrás de la Universidad de París, que se sitúa como la más productiva. Ello podría ser consecuencia de que Bélgica se sitúa entre los países con mayor impacto en la investigación que realizan, pese a no contar con una gran cantidad de instituciones que realicen investigación sobre enfermedades raras al tratarse de un país pequeño, Ahora bien, aunque esto puede significar que el número de instituciones belgas que desarrollan su labor investigadora en enfermedades raras es pequeño, ello no implica que alguna de las que lo hacen, como la Universi-

dad de Lovaina, pueda ser una de las más importantes y activas en este ámbito, presente en multitud de congresos y conferencias internacionales y participante en el proyecto europeo EURO-CDG (E-RARE, 2011), destinado a estudiar los defectos congénitos del metabolismo.

En cuanto a la presencia de Finlandia, a través de la Universidad de Helsinki, se debe a que esta universidad se encuentra muy involucrada en el estudio y mapeo de los genes relativos a las enfermedades raras en poblaciones aisladas (DE LA CHAPELLE, 1993; PELTONEN et al., 1995).

En ambos casos, tanto la presencia de Bélgica como de Finlandia se debe a las instituciones señaladas en cada uno, lo que se confirma cuando se observa la distribución de la producción científica entre las principales instituciones de los quince países estudiados (tabla 7.11), pues en ambos países tan solo destaca la producción de las instituciones mencionadas (Universidades de Lovaina y Helsinki), mientras que el resto de instituciones poseen una producción muy pequeña.

Por otro lado en la misma tabla se puede observar que la producción de las instituciones holandesas se encuentra mucho más concentrada que la de España, el otro país con una producción creciente a lo largo del periodo estudiado. Ello podría ser una de las causas por las que la producción española, a pesar de ser muy similar a la holandesa tiene mucho menos impacto. El hecho es que en España la producción se encuentra dispersa entre las universidades y los hospitales, mientras que en Holanda las universidades asumen el grueso de la investigación, por delante de los hospitales, tal y como ocurre en los países con mayor producción (Reino Unido, Alemania, Francia e Italia). Por todo ello, en este sentido el perfil de Holanda, de nuevo se asemeja más al de los grandes productores, mientras que el de España se acerca más al de los pequeños productores.

Para analizar con más detalle esta distribución de los tipos institucionales por país, puede observarse la figura 7.5, que en un gráfico de mosaico muestra la distribución de los diferentes tipos institucionales por país, sobre una supuesta distribución equidistante. En ella se observa que la producción de empresas y

fundaciones es muy baja en todos los países, mientras que en España, Francia, Reino Unido e Italia es relativamente mayor que en el resto de países. En función del test de independencia realizado sobre los residuos estandarizados, los hospitales son los más productivos en todos los países, sin embargo España, Francia, Grecia, Irlanda y Holanda poseen una producción significativamente mayor que en el resto de países. En el caso de España es especialmente notable debido a que la presencia de las universidades es mucho más baja de lo esperado, tal y como se vio en el análisis de las principales instituciones productoras por países. Por otro lado los centros nacionales de investigación y las universidades destacan especialmente en el caso de Alemania e Italia. Por último cabe señalar que la distribución del Reino Unido es la que más se asemeja a una teórica distribución equidistante, con una presencia balanceada de los cuatro tipos de instituciones.

En cuanto a las 19 instituciones concretas que conforman el núcleo con mayor producción, cabe destacar que, tal y como se recoge en la tabla 7.10, solo dos de ellas son hospitales, tres si se incluye la Escuela de Medicina de Hannover, que ejerce como universidad y centro hospitalario. En este sentido, es notable el predominio de las universidades como los mayores centros productivos de información científica, por delante de los hospitales, a pesar de que éstos cuentan con una producción agregada mucho mayor. Es decir, que mientras que los hospitales, como tipo institucional acumulan la mayor parte de los trabajos científicos, las universidades cuentan con los centros más importantes, junto con los centros nacionales de investigación franceses, tanto a nivel general (CNRS), como el centro especializado en investigación médica (INSERM), que se sitúan entre los diez centros más productivos dentro de la distribución de trabajos por institución. Ello es debido a que Francia alberga algunos de los centros más importantes en cuanto al estudio de las enfermedades raras y cuenta con un destacado programa de investigación sobre enfermedades desarrollado entre 2004 y 2008, que ha dado lugar a múltiples publicaciones científicas (AYMÉ, 2006). Pero además participa en muchos de los proyectos de investigación relacionados con las enfermedades

raras dentro del sexto programa MARCO de la Unión Europea, con 198 proyectos en total, solo por detrás de Alemania, con 203 (ORPHANET, 2010).

En cuanto al impacto de las publicaciones en las instituciones del núcleo, como se puede observar en la tabla 7.40, las instituciones francesas aparecen incluso mejor posicionadas que en el caso de la producción, puesto que tanto el CNRS como el INSERM escalan posiciones y se sitúan, junto con la Universidad de París, la Universidad Católica de Lovaina y la Universidad de Oxford, entre los cinco centros con un mayor número de publicaciones en el primer cuartil del Journal Citations Report (JCR). De esta forma se confirma que las instituciones francesas se encuentran entre las más punteras en el ámbito de las enfermedades raras. Sin embargo, a pesar de que las instituciones francesas son las que más documentos publican en revistas de alto impacto, no reciben un monto de citas equivalente, si no que, tal y como muestra la tabla 7.48, con distribución de citas por documento para las instituciones del núcleo, tan solo el CNRS se mantiene entre las cinco primeras instituciones con un mayor número de citas por documento, por detrás de la Escuela de Medicina de Hannover, la Universidad de Londres (UCL) y la Universidad de Cambridge. En cambio la Universidad de París y el INSERM, que se encontraban entre los centros más productivos y con mayor número de publicaciones en revistas de impacto, pasan a formar parte del último tercio de instituciones en cuanto al número de citas que reciben por documento publicado.

La explicación de este efecto se encuentra en el ingente volumen de publicaciones que realizan, especialmente la Universidad de París, si bien solo unas cuantas de ese gran volumen de publicaciones recibe un considerable número de citas. Por ello la Universidad de París se encuentra en un punto intermedio en cuanto al porcentaje de documentos no citados (con un promedio del 3,53%). Sin embargo, en el caso del INSERM el promedio de documentos no citados es comparativamente menor (un 2,59%), por lo que se debe concluir que el hecho de que descienda tanto en el número de citas por documento se debe a que las publicaciones de

este centro no generan un gran interés en la comunidad científica a pesar de ser extremadamente productivo.

Tal y como mostraron OWEN-SMITH y RICCABONI (2002), la colaboración entre las diferentes instituciones en el área de las ciencias de la vida y la salud, especialmente entre los grandes actores en este ámbito (universidades, hospitales, centros nacionales de investigación y empresas), es un factor determinante a la hora del desarrollo de investigaciones que deriven en publicaciones científicas. En este sentido, para estudiar dichas relaciones se ha elaborado un análisis de redes sociales basado en la matriz de colaboraciones entre los diferentes centros identificados. Esta red social, representada en la figura 9.4, muestra tres agrupamientos de instituciones muy claros, dos de ellos muy cohesionados y un tercer un poco más disperso. Los dos primeros corresponden a la colaboración entre los centros de investigación franceses (Universidad de París, CNRS, INSERM, Hospitale Necker Enfants Maladies y Hospital Robert Debre) y el segundo a la colaboración entre las instituciones italianas (Universidad la Sapiencia de Roma y Universidad de Milán).

Esta distribución indica una fuerte carga del factor geográfico a la hora de determinar que instituciones colaboran entre si. En cuanto al tercer núcleo, más disperso y compuesto por el resto de instituciones, hay que señalar que tan solo algunas instituciones como la Universidad de Helsinki, la de Amsterdam, la de Leiden y el Hospital Universitario de Oulu, se separan del resto. Ello indica un perfil de colaboración ligeramente diferente al que muestran el resto de instituciones, especialmente la Universidad de Helsinki, que como se señaló anteriormente, se distingue del resto de instituciones debido a su perfil orientado a la investigación sobre los genes relativos a las enfermedades raras en poblaciones aisladas (DE LA CHAPPELLE, 1993; PELTONEN et al., 1995).

Este análisis de la colaboración entre instituciones se complementa con el análisis de clusters realizado entre las instituciones que forman parte del núcleo más productivo y que muestra una clara segregación a nivel geográfico, especialmente

en cuanto a las instituciones francesas, que son las primeras en separarse del resto, que forman un grupo más unificado, tal y como se observa en la figura 10.6.

En cuanto a la distribución temática, presente en el análisis de correspondencias de la figura 10.7, entre las instituciones del núcleo y las clasificaciones temáticas de las enfermedades, muestra de nuevo una clara relación entre las instituciones francesas, que se encuentran muy cercanas entre si en las proyecciones de las dos dimensiones del análisis. La proyección sobre la primera dimensión muestra unos perfiles muy parecidos en todas las instituciones, salvo en el caso de la Universidad de Amsterdam, que se diferencia claramente del resto, y se encuentra sesgada hacia determinados trastornos metabólicos (E70 y E79) así como del sistema nervioso periférico (G60). Ya se ha mencionado que existe una gran relación entre los trastornos metabólicos y los del sistema nervioso, pues muchos aminoácidos ejercen como neurotransmisores del sistema nervioso. Por tanto su asimilación defectuosa afecta necesariamente al mismo, por lo que no es de extrañar que una institución que se centre en el análisis de las enfermedades del sistema nervioso se ocupe también de los trastornos relativos al metabolismo humano.

Sobre la segunda dimensión del análisis se produce una mayor discrepancia, pues aparecen dos grupos de instituciones que se separan notablemente del eje central. En primer lugar las universidades de Helsinki, Leiden y Munich, se sitúan en la parte baja del análisis, más orientadas al estudio de la Osteonecrosis (M87) y de los trastornos mentales y del comportamiento (F01-F84). En ambos grupos el número de enfermedades es muy reducido: En el caso del M87 tan solo la Osteonecrosis, se encuentra dentro de este grupo, mientras que el grupo F01-F84 se encuentra compuesto por dos enfermedades, el Síndrome de Rett y el CADA-SIL, por lo que aquellas instituciones que traten estos trastornos con frecuencia, necesariamente van a encontrarse separadas del resto en alguna de las dimensiones del análisis, en este caso en la segunda. Por otro lado, estos dos últimos desórdenes poseen una importante carga de investigación genética (SERRANO-LÓPEZ y MARTÍN-MORENO, 2011) y como ya se ha mencionado, la Universidad de Helsinki cuenta con importantes trabajos en este área.

Por otro lado, el la Escuela de Medicina de Hannover, así como las universidades de Tübingen y Cambridge se sitúan en el otro extremo (superior) del análisis, indicando un interés muy bajo o nulo, de estas instituciones por los grupos de enfermedades situados en el otro extremo de la segunda dimensión del análisis.

11.5. Autores

Los datos de la productividad de los autores en ambas bases de datos cumplen sobradamente con la distribución propuesta por LOTKA (1926). En el caso de Pubmed 16 autores acumulan el 34,09% de la producción y en el caso del SCI son 11 autores los que acumulan el 35,24% de los trabajos. De estos autores que componen el “núcleo” de la distribución, tan solo siete aparecen en los núcleos de ambas bases de datos, debido a que el volumen de autores en la distribución de ambas distribuciones varía notablemente, con aproximadamente 10.000 autores de diferencia entre ambas (65.976 en el SCI y 55.433 en Pubmed). Esta diferencia se debe a que Pubmed es una base de datos especializada en medicina y biomedicina, mientras que la Web of Science es multidisciplinar, por lo que en el monto total de trabajos aparecen artículos que tratan las enfermedades raras de forma tangencial y no necesariamente en revistas de medicina o biomedicina, por lo que no se encuentran recogidas en Pubmed.

Por otro lado, el Índice de élite de Price muestra valores muy similares en ambas bases de datos, situándose los valores del índice entre 235 y 257 autores dependiendo de la base de datos, por lo que se puede concluir que, a pesar de las diferencias entre ambas bases de datos, el volumen de autores que acumulan el 90% de los trabajos es muy similar, aunque existan diferencias en el núcleo duro de autores mencionado anteriormente.

El otro método empleado para determinar la élite de autores, el Índice de Yablonsky, indica que los autores que conforman este núcleo están relacionados con la raíz cuadrada del número de autores con un solo trabajo. Sin embargo, debido

al gran volumen de autores “transeúntes” en ambas distribuciones, el corte en el Índice de Yablonsky es tan alto que ningún autor lo supera (191 trabajos en Pubmed y 213 en el SCI). Esto es debido a que en Ciencias de la Vida y la Salud, y especialmente en el área de Biomedicina, el número de autores transeúntes suele ser extremadamente alto, estando alrededor del 65 % en trabajos similares (ABAD-GARCÍA et al., 2009). En este caso, en el SCI es del 69,31 % y en Pubmed del 65,91 %, estando por tanto dentro de los valores normales descritos por SCHUBERT y GLÄNZEL (1991) y conformes a la distribución teórica de LOTKA (1926).

En cuanto a los índices de coautoría, agregados por países, instituciones y temáticas, muestran un rango que oscila entre los 6 y los 8 autores por documento, en función, fundamentalmente de la temática de los documentos y de los países de los autores. Ello se debe a que en áreas como la Genética o la Bioquímica y Biología Molecular, los autores que firman los trabajos son más numerosos que en áreas más especializadas, debido a que se trata de áreas interdisciplinarias (NEWMAN, 2004; BELLANCA, 2009). Por otro lado, los países influyen notablemente en los índices de coautoría debido a que, como se comentó anteriormente, el factor de desarrollo tecnológico de los países, tal y como describió LUUKKONEN et al. (1992), afecta a su grado de colaboración, produciendo mayores índices de colaboración y coautoría en aquellos países con menor desarrollo tecnológico, pues buscan colaborar con instituciones de otros países que posean los medios para poder desarrollar las investigaciones propuestas.

Descendiendo a nivel de autores concretos se observa que las categorías temáticas en las que publican los autores del núcleo la mayor parte de sus trabajos, corresponden a revistas de Genética y Bioquímica - Biología Molecular mayoritariamente, seguidas de publicaciones sobre Microbiología, Sistema Respiratorio, Pediatría y, en menor medida, Neurología y Neurociencia. Queda patente que la carga principal de la investigación en este tipo de enfermedades reside en áreas multidisciplinarias, como la Genética o la Biología Molecular, en las que publican los grandes productores, mientras que los pequeños productores son los que

producen trabajos más especializados, y probablemente a este hecho sea debido el gran número de autores transeúntes que aparece en esta distribución. Por otro lado, dicha concentración de los autores alrededor de las grandes áreas de investigación concuerda con las conclusiones de MORIN (2004), que indican que los trabajos sobre enfermedades raras se encuentran muy dispersos en múltiples campos, debido a su diferente naturaleza, salvo en el caso de enfermedades con un núcleo o área común, como las que se han estudiado en este trabajo o como las enfermedades producidas por priones, tal y como ocurre en el trabajo de SANZ-CASADO et al. (2007)

En relación con esta especialización de los autores no se debe pasar por alto que las investigaciones sobre enfermedades de baja prevalencia se encuentran con serios problemas a la hora de ser difundidas en revistas, especialmente en revistas internacionales en las que el Factor de Impacto es muy influyente, tal y como se comentó anteriormente a raíz del trabajo de RUANO-RAVINA y PÉREZ-RÍOS (2012) y además generalmente no poseen la suficiente importancia como para contar con sus propias revistas especializadas (MORIN, 2004).

En cuanto a la forma en la que estos autores colaboran entre si, a través del análisis de redes sociales se pueden identificar algunos patrones realmente esclarecedores. En primer lugar, tal y como ya se adelantó en el capítulo de resultados, aparecen relaciones muy pronunciadas entre determinados autores, como Waterham y Wanders en el caso del SCI o Wajner y Beck y Van Den Heuvel y Smeitink en el de Pubmed. Analizando con más detalle estas relaciones y teniendo en cuenta la especialidad en la que trabaja cada uno de estos investigadores (a través de las categorías temáticas de la Web of Science), se aprecia que las relaciones más estrechas como el binomio Wanders-Wateham o el grupo Munnich-Ferec-Tummler, responde a que todos ellos trabajan fundamentalmente en Genética, Biología Molecular y Microbiología. A este último grupo se pueden añadir los nodos representados por Vandamme y Hoiby, autores especializados en Microbiología y que se encuentran unidos al grupo de autores relacionados con la Genética a través del nodo que identifica Tummler y que aúna ambas categorías.

En cuanto a los nodos más aislados, representados por Beck, Bush y Elborn, el análisis de las temáticas en las que trabajan cada uno de ellos, indica que Bush y Elborn se centran en la investigación sobre el Sistema Respiratorio. De ahí que exista cierta relación entre ellos, y que se encuentren unidos a través de nodos específicos a los autores que trabajan en Genética. En cambio, el nodo representado por Beck aún a un notable grupo de autores que trabajan fundamentalmente en Pediatría, y a su vez se encuentran unidos a los nodos relacionados con la Genética a través de autores concretos que sirven de nexo de unión entre ellos.

Por último los nodos representados por Hoffmann, se encuentran estrechamente unidos a Waterham y Wanders, debido a que sus áreas de especialización están relacionadas con la Genética por un lado y la Neurociencia por otro.

Estos resultados confirman de nuevo la hipótesis inicial de que existe una notable relación entre la Genética y el estudio de diferentes enfermedades, especialmente desde que se comenzó a trabajar en el mapa del genoma humano (DALPÉ, 2002; BARJAK y ROBINSON, 2008b). Por otro lado esta relación es mucho más estrecha en el caso de las enfermedades raras (HEEMSTRA, 2009; SERRANO-LÓPEZ y MARTÍN-MORENO, 2011), y mucho más en cuanto al hecho de que todas las enfermedades estudiadas en este trabajo poseen una causa genética identificada, por lo que está mucho más claro que en otros casos que la Genética actúa como nexo de unión entre ellas.

En cuanto a la red de cocitaciones, la figura 9.3, muestra también una serie de relaciones muy interesantes, y de nuevo cobran sentido cuando se ponen en relación con las áreas en las que trabajan los diferentes autores citados.

Así, se ha encontrado que los principales grupos de nodos corresponden a agrupaciones temáticas de los autores. En este sentido los autores del grupo situado en la parte inferior-izquierda corresponden a aquellos que son citados en trabajos relativos a Hematología e Inmunología, mientras que los que están en la parte inferior derecha se corresponde con los trabajos relativos a las enfermedades del sistema respiratorio y fundamentalmente a la Fibrosis Quística, motivo por el

cual “Ganz T” se encuentra situado entre ambos grupos, pues trabaja en ambas áreas.

Un poco más arriba, en el centro del grafo, se encuentra el grupo de autores con los nodos más grandes de todo el análisis, indicando que son los que poseen un mayor número de cocitaciones y no solo entre ellos, sino con todo el grupo de autores. Este grupo de autores tiene la Genética como especialidad, por lo que su situación y relación con el resto de autores indican, de nuevo, que la Genética es el nexo de unión entre las diferentes enfermedades raras aquí estudiadas. Junto a este grupo se sitúa otro formado principalmente por Eng, Desnick y Schiffmann, autores que trabajan fundamentalmente en el área de Urología, aunque también poseen una gran relación con la Genética.

Por último, el resto de autores del grafo se encuadran en diferentes campos de la Neurociencia y la Neurología y se pueden dividir en función de como son citados por sus diferentes especialidades. Así a la derecha se encuentran los autores más relacionados con la Biología, en el centro los que poseen un mayor número de relaciones con la Genética, en la parte inferior los que se encuadran dentro de la Biofísica y Bioquímica y finalmente, a la izquierda, un grupo de autores, muy cocitados entre ellos, y relacionados con la Cardiología.

A la vista de la distribución, se pueden deducir varias cuestiones. Por un lado que el análisis recoge las áreas fundamentales en cuanto a producción científica dentro de la investigación en enfermedades raras. Dicha producción está compuesta por las áreas relativas a la Genética, Aparato Respiratorio, Hematología e Inmunología, Neurología y Neurociencia y, en menor medida, Urología, actuando la Genética y la Biología en sus diferentes ramas como áreas interdisciplinarias, presentes en prácticamente todos los frentes de investigación y especialmente en la Neurología (PORTER et al., 2007). Mientras que por otro lado, analizando la cohesión de la red de cocitaciones, la conclusión a la que se llega es que las áreas relativas al Aparato Respiratorio y la Hematología e Inmunología son las redes que más cohesionadas se encuentran, mientras que la Neurología y Neurociencia

se encuentra mucho más dispersa, debido a que se trata de un área más interdisciplinar (BURRIGHT et al., 2005). Ello hace que las redes de cocitación se distancien más entre ellas, a pesar de que los autores que son citados juntos, lo son con mucha frecuencia. Esta frecuencia viene indicada por el grosor de la línea que une los diferentes nodos.

Finalmente los análisis multidimensionales, en este caso análisis de clusters y análisis de correspondencias muestran otros factores que afectan a la distribución de los autores.

El análisis de clusters realizado, al igual que el análisis de redes sociales, sobre la base de las colaboraciones entre los diferentes autores, muestran como se agrupan los autores con al menos 100 colaboraciones en todo el periodo. En el caso del SCI los autores se separan rápidamente en dos clusters muy diferenciados, el de la izquierda se compone de autores que trabajan casi exclusivamente en Genética, mientras que el de la derecha, mucho más numeroso, se compone de autores que, a pesar de trabajar en muchos casos en Genética, combinan esta temática con otras especialidades. En el segundo cluster los autores van separándose en función de su especialización. Así los autores de la derecha, centrados en Genética y Biología Molecular son los primeros en separarse, pues apenas comparten especialidades con el resto de investigadores, mientras que el resto de autores del cluster se van agrupando en función del grado de colaboración existente entre ellos, marcado por el interés común en ciertos grupos de enfermedades, como se verá en el análisis de correspondencias.

En cuanto a Pubmed, la división en los diferentes clusters se produce también muy rápidamente, pero a diferencia del SCI es necesario descender más en el análisis para poder diferenciar los distintos clusters que se forman a un nivel más bajo del análisis. Por otro lado las agrupaciones de autores en este caso son más homogéneas que en el caso del SCI. Este comportamiento se debe a que la cobertura de Pubmed en revistas biomédicas es mucho mayor que en otras bases de datos como la Web of Science, Scopus o Google Scholar (FALAGAS et al., 2008),

por lo tanto la especialidad biomédica les otorga a los autores una cohesión en cuanto a su colaboración que no existe en los datos obtenidos a través de la Web of Science. Este hecho también explica las diferencias encontradas en los autores más productivos en ambas bases de datos, puesto que los especialistas en Biomedicina aparecen más fácilmente en el núcleo de autores en Pubmed.

Por último, para analizar la distribución de los autores en función de las enfermedades concretas en las que trabajan, se han realizado sendos análisis de correspondencias para ambas bases de datos. En ambos análisis, a pesar de mostrar distribuciones diferentes, subyacen estructuras de distribución muy similares:

Lo primero que se puede destacar es que autores como Waterham y Wanders ocupan situaciones centrales, muy cerca del eje que forman las dos dimensiones analizadas, indicando que no se encuentran especializados en ninguna enfermedad concreta, debido a que son especialistas en Genética y Biología Molecular, áreas interdisciplinarias que intervienen en múltiples enfermedades. Sin embargo, se sitúan especialmente cerca de las enfermedades clasificadas en los diferentes grupos “E”, relativos a los distintos desórdenes del metabolismo.

En cuanto a las enfermedades que se comportan de una forma diferente al resto, el primer grupo a destacar es el E83-E88. Como ya se ha mencionado anteriormente se trata de un grupo muy especial, pues incluye la Fibrosis Quística, enfermedad rara pero con un frente de investigación muy importante que se comporta prácticamente como una enfermedad común. En el caso del SCI los autores que investigan sobre este grupo de enfermedades, se separan del grupo principal de autores únicamente sobre la primera dimensión del análisis, mientras que en Pubmed el distanciamiento es mucho mayor y se produce en ambas dimensiones, si bien se debe destacar que el análisis de correspondencias en el caso de Pubmed muestra una mayor dispersión que en el SCI.

En cuanto a la segunda dimensión de ambos análisis, se observa una clara separación de los autores que trabajan en el grupo M87, relativo a la Osteonecrosis.

Este grupo, del que ya se han mencionado algunas de sus características, cuenta con una producción muy baja, especialmente en el SCI, debido a que los trabajos relativos a esta enfermedad se publican fundamentalmente en revistas de Odontología, que no poseen una gran cobertura en el SCI, mientras que en Pubmed su representación es mayor. A pesar de ello, en ambos casos se distancia del eje central en base a la segunda dimensión, indicando que los autores que trabajan sobre esta enfermedad se encuentran muy especializados y apenas poseen relación con los investigadores que trabajan sobre otros grupos de enfermedades raras.

Por último, respecto a la segunda dimensión del análisis efectuado sobre los datos obtenidos del SCI, se observa un claro distanciamiento de los autores que investigan sobre los grupos Q93-Q99 (Anormalidades cromosómicas no clasificadas en otra parte) y F01-F84 (trastornos mentales y del comportamiento). Esta estructura, que en Pubmed es mucho menos acusada, puede deberse a que se trata de grupos de enfermedades muy heterogéneos y compuestos por un pequeño número de desórdenes (7 en el caso del Q93-Q99 y 2 en el grupo F01-F84), por lo que la especialización de los autores en estas enfermedades es mucho mayor que en las que componen el núcleo central de la distribución y que se centran en los trastornos metabólicos (AL-SHAHI et al., 2001; BISHOP, 2010; SERRANO-LÓPEZ y MARTÍN-MORENO, 2011).

Sección 12

Conclusions

Like the discussion, this section, that contains the conclusions of the work, begins with the general conclusions to, later, descend to the particular conclusions about the methodology, countries, subjects, sources, institutions and authors.

- There is a strong influence of Genetics over the research field of rare diseases, especially since the sequencing of the human genome map, something visible through the production peaks observed in the evolution of scientific activity.
- The influence of Genetics on the investigation, has changed the pattern of the research in many ways, especially in the distributions of the work at the level of document types, languages and obsolescence, closer to Experimental Sciences than to Life and Health Sciences.
- Observed values of obsolescence are closer to Experimental Sciences than Life Sciences and Health, showing an increase in the half-life, as we moved forward in the years of study.
- Working on the human genome and the discovery of genes that cause different diseases have a cumulative effect of citation, disrupting the distribution of citations, especially in the case of some diseases such as cystic fibrosis.

- Both the analysis of the regression between scientific activity and health expenditure by country as well as the number of patient associations, showed strong direct correlations, indicating a strong relationship between the variables, so a possible causal relationship between them can be guessed.
- In the case of the analysis of the regression between scientific activity and health expenditure by country, it is particularly relevant the interaction between health spending and country variables, appearing as the most statistically significant variable.
- The relationship between scientific activity and the number of patient associations shows a high correlation, due to the big influence of patient associations on the development of research.

12.1. Methodology

- The normalization system developed here, allows to automatize the processing of data at different levels, related to the normalization of addresses and author names.
- The procedure programmed for the automatic assignment of MeSH thesaurus terms, allows to add these terms to records from the Web of Science, which originally does not incorporate this feature.
- The development of an automated procedure for obtaining indicators, from relational databases, allows to program it, obtaining the indicators from statistical functions and replicating it in other bibliometric works.

12.2. Countries

- The countries studied in this work show a clear differentiation by production, leading to the division between large producers, small producers and emerging producers.
- Those countries with a highest interest are the emerging countries, that at the beginning of the studied period are closer to small farmers, but at the end thereof are closer to the top producing countries. This is the case of the Netherlands and Spain.
- We observed a clear preference of researchers in some countries (especially Spain and France) to publish in national journals, and in local languages. This preference causes, especially in the case of Spain, that the research achieves a lower impact than it could have if it was published in international journals of high impact.
- The countries with a consolidated scientific activity have indexes of collaboration lower than countries with lower research potential, because they have the resources to carry out their scientific work.
- Countries with a lower technology development show a greater collaboration, because they require such resources to carry out their research.
- Collaboration is one of the indicators that shows much better the evolution of countries, such as the Netherlands and Spain, whom despite being at the beginning of the study period near to the small producers, later achieve much more prominent positions regarding the scientific output they produce and the impact of it, achieving almost the position of the large producers.
- In the case of Spain, international collaboration has greatly declined in comparison to the other emerging country (Netherlands), so it is expected that in later periods of the study, scientific production of Spain will be at a

standstill or even diminish, since it had not yet achieved a position strong enough to perform without international collaboration.

- We observed a significant difference in the collaboration profiles of the two countries with the highest production (UK and Germany). UK collaborates actively with the least productive countries and shows a master-ward relationship with Ireland. On the other hand Germany collaborates more with the countries of Eastern and Austria.
- The profile of collaboration teacher-ward in the UK, causes an isolation index higher than other countries, publishing more in British journals, especially in the case of papers in collaboration with Ireland.
- Some of the less productive countries tend to publish on specific themes, as in the case of Ireland and the research on cystic fibrosis, or Finland and mental and behavioral disorders, while big producers did not show significant biases to any particular area.
- There is no correspondence between the productivity of the countries and the impact of their research. Countries such as Belgium and the Scandinavian countries, small producers, receive a largest number of citations per paper than the big producers. In the case of emerging countries, the Netherlands is the most cited country.

12.3. Subjects

- The diseases with most interest to researchers include cystic fibrosis and metabolic diseases.
- The Web of Science suffers from a lack of journals in Dentistry, so one of the disorders related to this area, Osteonecrosis, is under-represented in this database, even though the number of papers about this diseases in PubMed is near to the average.

- The literature on Osteonecrosis is one of the most cited, and has more interest among the scientific community.
- The relationship of rare diseases with Genetics increased the degree of collaboration in all areas. Found interesting relationships between different disease groups have been found. On one hand, those which share a common interest (such as metabolic disease), and secondly, between disease groups apparently unrelated, as in the case of nervous system diseases and disorders affecting the metabolism of amino acids.
- Pediatrics has a very important role in the research on rare diseases that have been studied, because most of these disorders are diagnosed in the early stages of human development.

12.4. Sources

- The journals in which researchers tend to publish that have been studied in this work, fit very well the distribution of Bradford, showing a very small number of journals that collect most of the papers. These journals have a greater specialization, oriented especially to the publication in the areas of Genetics, Neurology and Neuroscience, and Pediatrics and Metabolic Diseases of the Respiratory System.
- Referenced journals also have the effect of specialization, belonging most of them to the area of Genetics, general journals of Medicine and of course, Science and Nature as multidisciplinary journals. The core of these journals is composed of a really small number of publications (only 29), relating to such issues.
- There is a clear correspondence between the subjects of the most productive journals and those most cited, with the Genetics, Molecular Biology and

Neurology topping the list, along with the journal category "Multidisciplinary", including Nature and Science journals.

- Cores of referenced journals and publishing journals do not match exactly, because the system of quality assessment of journals can make that those with a highest impact factor, often refuse to publish works about rare diseases.
- The evaluation system of scientific journals, through the JCR Impact Factor, produces most of the work appearing in British and American journals and of course in English.
- The nationality and language of the journals affect their position in the JCR, so publications of nationalities different from British or American, have less representation, only being three journals published in languages different than English in the last quartile of their respective subject distributions.
- Groups of diseases E83-E88 (cystic fibrosis), M87 (Osteonecrosis) and E16-E59 (endocrine and nutritional), have a different behavior, due to the special characteristics of these diseases, especially observable in cystic fibrosis, which attracts a large group of journals related to the study of the respiratory system, Genetics and Pediatrics.
- Most researchers prefer to publish in US or UK journals, while Spain and France researchers show a more biased profile to publish in national journals and in their own languages, although the trend to publish in national journals is declining.

12.5. Institutions

- The prevailing institutional type over the studied works are hospitals, participating in most of the papers on rare diseases, while private institutions

(mainly pharmaceutical companies) have a very small percentage of participation.

- Among the most productive institutions are the French centers, especially the CNRS, INSERM and the University of Paris, which however did not receive an equivalent volume of citations.
- The most productive institutions are British, German and French, while the rest of the countries have very low representation. However, in Belgium, the Catholic University of Leuven, appears as the second most productive institutions, and one of the most important institutions in the area of rare diseases.
- In Netherlands, scientific production is much more concentrated than in Spain, especially in universities, while in Spain the volume of works is spread out between universities and hospitals, closer to the profile of the small producers, while the Netherlands is closer to the big producers.
- The presence in Spain of companies and foundations is closer to the profile of the large producers, so we can conclude that Spain is in the process of becoming one of them.
- The works that achieve a greatest impact come from national research centers, which have accumulated a number of citations per document much higher than other institutions.
- In the area of rare diseases there are not many clinical trials, so the documental type that usually brings more citations to hospitals, is decreasing and attracting fewer citations to these institutions.
- The volume of publications is not indicative of their quality, especially for INSERM, which is among the centers with a higher percent of uncitedness works and fewer papers published in high impact journals.

- The institutions have a clear preference to collaborate with geographically closer centers, so the distance between different research centers is a key factor in their cooperation, especially in this area.
- The preferences of research centers by specific subjects vary according to the size of the centers. While large centers do not specialize in any particular subject (especially those with the highest burden of genetic research), small centers conduct most of their research on specific diseases.

12.6. Authors

- The volume of authors in the core of both databases is very similar.
- It exists a large number of passing authors, according to the data observed in the Lotka distribution and Price Elite Index.
- The coauthorship index is between 6 to 8 authors per paper, mainly depending on the subject area, country and institutional types.
- The researchers on rare diseases published most of their papers in the area of Genetics and Biochemistry-Molecular Biology, Microbiology, Respiratory System, and to a lesser extent Pediatrics, Neurology and Neuroscience.
- Most of the authors that are considered big producers are related to the Genetics and Molecular Biology fields, while small producers are focused on more specialized areas.
- Diseases with a higher degree of specialized authors are related to Cystic Fibrosis, Osteonecrosis, and to a lesser extent to the mental and behavioral disorders, indicating a greater difference between these diseases and other rare diseases, more linked between them.
- The relationships between authors are very narrow, establishing very strong networks, especially among those who work more actively in Genetics, Mo-

lecular Biology, Microbiology, Pediatrics and Respiratory System. However, researchers in the field of Neurology and Neuroscience have more diffuse relationships.

- The distribution of authors and their areas of interest are very similar for the citing and cited authors, with the exception that, among the cited authors, there are some secondary areas with less weight, like the Hematology, Immunology and Urology.
- The Genetics and Molecular Biology act like a link between the authors, both cited as citing, working as interdisciplinary areas that are related to virtually all research on rare diseases, regardless of the area of specialization.

Conclusiones

Del mismo modo que en el apartado de discusión de los resultados, en este apartado relativo a las conclusiones finales del trabajo, se va a comenzar destacando las conclusiones generales, para posteriormente descender a las conclusiones particulares de las secciones destacadas, referentes a la metodología, así como a los países, temáticas, fuentes, instituciones y autores.

- Existe una fuerte influencia del área Genética sobre la investigación en enfermedades raras, especialmente desde la secuenciación del mapa del genoma humano, visible a través de los picos de producción observados en la evolución de la actividad científica.
- La influencia de la Genética sobre la investigación realizada, ha modificado los patrones de investigación en muchos sentidos, especialmente en cuanto a las distribuciones de los trabajos a nivel de tipologías documentales, idiomas y en la obsolescencia de los trabajos, acercándose más a las Ciencias Experimentales que a las Ciencias de la Vida y la Salud.
- Los valores de obsolescencia observados se acercan más a las Ciencias Experimentales que a las Ciencias de la Vida y la Salud, encontrando un aumento de la vida media, conforme se avanza en los años estudiados.

- Los trabajos relativos al genoma humano y al descubrimiento de los genes causantes de las distintas enfermedades obtienen un efecto acumulativo de citas, viéndose perturbada por ellos la distribución de citas, especialmente en el caso de algunas enfermedades como la fibrosis quística.
- El estudio de regresión entre actividad científica y gasto sanitario por país, así como con el número de asociaciones de enfermos, ha mostrado en ambos casos fuertes correlaciones directas, indicando una gran relación entre las variables, por lo que se puede intuir una posible causalidad entre ellas.
- En el caso del estudio de la correlación entre la actividad científica y el gasto sanitario por país, cobra especial relevancia la interacción entre las variables “gasto sanitario” y “país”, mostrándose como la variable más significativas estadísticamente.
- La relación entre la actividad científica y el número de asociaciones de enfermos muestra una gran correlación, debido a la influencia de las asociaciones de enfermos en el desarrollo de las diferentes investigaciones.

13.1. Metodología

- El sistema de normalización desarrollado ha permitido automatizar el tratamiento de los datos en diferentes niveles, tanto de normalización de las direcciones como de los nombres de autor.
- El procedimiento programado para la asignación automática de los términos del tesoro MeSH, permite agregar estos términos a registros procedentes de la Web of Science, que inicialmente no los incorporan.
- El desarrollo de un procedimiento para la obtención automatizada de indicadores, a partir de bases de datos relacionales, permite programar la obtención de los mismos a partir de funciones estadísticas y replicarlos en cualquier otro trabajo bibliométrico.

13.2. Países

- Los países estudiados muestran una clara diferenciación según su producción, dando lugar a grandes productores, pequeños productores y productores emergentes.
- El grupo de países de mayor interés es el de los países emergentes, que al inicio del periodo estudiado se encuentran más cerca de los pequeños productores, pero al final del mismo se han acercado mucho al de los países con mayor producción. Tal es el caso de Holanda y España.
- Se ha observado una clara preferencia por parte de los investigadores de algunos países (especialmente España y Francia), para publicar en revistas nacionales, en sus idiomas locales. Esta preferencia provoca, sobre todo en el caso de España, que las investigaciones realizadas alcancen un impacto mucho menor del que podrían tener si se publicaran en revistas internacionales de alto impacto.
- Los países con una actividad científica más consolidada poseen unos índices de colaboración mucho menores que los países con un potencial investigador menor, debido a que poseen los medios necesarios para realizar su actividad científica.
- Los países con menor desarrollo tecnológico muestran una mayor colaboración, porque requieren de dichos medios para llevar a cabo sus investigaciones.
- La colaboración es uno de los indicadores que mejor refleja la evolución de países como Holanda y España, que pese a encontrarse, al inicio del periodo de estudio, cerca de los pequeños productores, alcanzan posteriormente unas posiciones mucho más destacadas en cuanto a la producción científica que elaboran y al impacto de la misma, ascendiendo prácticamente hasta la posición de los grandes productores.

- En el caso de España, la colaboración internacional ha decaído en gran medida respecto a la del otro país emergente (Holanda), por lo que es de prever que en periodos de estudio posteriores, la producción científica de nuestro país se vea estancada o incluso mermada respecto a la actual, puesto que aún no ha alcanzado una posición lo suficientemente fuerte como para prescindir de la colaboración internacional.
- Se ha observado una gran diferencia en los perfiles de colaboración de los dos países con mayor producción (Reino Unido y Alemania). Reino Unido colabora activamente con los países menos productivos y muestra una relación maestro-pupilo con Irlanda. En cambio Alemania colabora más con los países del Este y Austria.
- El perfil de colaboración maestro-pupilo del Reino Unido, provoca que tenga un índice de aislamiento mayor que otros países, publicando fundamentalmente en revistas británicas, especialmente en el caso de trabajos en colaboración con Irlanda.
- Algunos de los países menos productivos tienden a publicar sobre temáticas concretas, tal es el caso de Irlanda y la investigación en fibrosis quística, o el de Finlandia y los desórdenes mentales y del comportamiento, mientras que los grandes productores no muestran sesgos significativos hacia ningún área concreta.
- No existe correspondencia entre la productividad de los países y el impacto en sus investigaciones. Países como Bélgica y los países escandinavos, con una producción pequeña, reciben un mayor número de citas por documento que aquellos que pertenecen a los grandes productores. En el caso de los países emergentes, Holanda es el que más citas recibe.

13.3. Temática

- Las enfermedades que despiertan mayor interés en los investigadores son la fibrosis quística y las enfermedades metabólicas.
- La Web of Science adolece de una falta de revistas especializadas en Odontología, por lo que uno de los trastornos relacionados con éste área, la Osteonecrosis, apenas se encuentra representado en esta base de datos, a pesar de que en Pubmed el número de artículos que tratan esta enfermedad se encuentra dentro de la media.
- La literatura sobre Osteonecrosis se encuentra entre las que más citas recibe, y tiene más interés en la comunidad científica.
- La relación de estas enfermedades con la Genética ha hecho que aumente el grado de colaboración en todas las áreas. Se han encontrado relaciones muy interesantes entre diferentes grupos de enfermedades. Por un lado, aquellos que comparten un interés común (como pueden ser las enfermedades metabólicas), y por otro, entre grupos de enfermedades aparentemente desconectados, como es el caso de las enfermedades del sistema nervioso y las que afectan a la metabolización de los aminoácidos.
- La Pediatría tiene un peso muy importante en la investigación en las enfermedades raras estudiadas, debido a que la mayor parte de estos trastornos son diagnosticados en las primeras fases del desarrollo humano.

13.4. Fuentes

- Las revistas en las que publican los investigadores sobre el tema de estudio, se ajustan perfectamente a la distribución de Bradford, encontrando un número muy pequeño de revistas que recogen la mayor parte de los trabajos. Dichas revistas poseen una mayor especialización, orientada fundamental-

mente a la publicación en las áreas de Genética, Neurología y Neurociencia, Pediatría y Enfermedades Metabólicas y del Sistema Respiratorio.

- En las revistas referenciadas, también se produce el efecto de especialización, perteneciendo la mayor parte de ellas, al área de Genética, revistas generalistas en Medicina y por supuesto, Science y Nature dentro de las revistas multidisciplinarias. El núcleo de dichas revistas está compuesto por un número realmente pequeño de publicaciones (tan solo 29), relativas a dichas temáticas.
- Existe una clara correspondencia entre las temáticas de las revistas en las que más se produce y aquellas más citadas, con la Genética, Biología Molecular y Neurología encabezando la lista, junto con la categoría de revistas “Multidisciplinarias”, entre las que se encuentran Nature y Science.
- Los núcleos de revistas referenciadas y de revistas dónde se publica no coinciden exactamente, debido a que el sistema de evaluación de la calidad de las revistas puede hacer que las de mayor Factor de Impacto, rechacen a menudo los trabajos relativos a las enfermedades raras.
- El sistema de evaluación de las revistas científicas, por medio del Factor de Impacto del JCR, produce que la mayor parte de los trabajos aparezcan publicados en revistas británicas y estadounidenses y por supuesto en inglés.
- La nacionalidad y el idioma de las revistas afecta a su posición en el JCR, por lo que las publicaciones de nacionalidades diferentes a la británica o estadounidense, tienen menor representación y solo aparecen tres revistas en lenguas diferentes del inglés, en el último cuartil de sus respectivas distribuciones temáticas.
- Los grupos de enfermedades E83-E88 (fibrosis quística), M87 (Osteonecrosis) y E16-E59 (enfermedades endocrinas y nutricionales), poseen un comportamiento diferenciado, debido a las características especiales de estas enfermedades, observables sobre todo en la fibrosis quística, que atrae un

nutrido grupo de revistas relativas al estudio del aparato respiratorio, la Genética y la Pediatría.

- La mayor parte de los investigadores prefieren publicar en revistas estadounidenses o británicas, mientras que los de España y Francia muestran un perfil más sesgado hacia la publicación en revistas nacionales y en sus propios idiomas, a pesar de que la tendencia a publicar en revistas nacionales es cada vez menor.

13.5. Instituciones

- El tipo institucional predominante en los trabajos estudiados son los hospitales, que participan en la mayor parte de los artículos relativos a enfermedades raras, mientras que los centros privados (fundamentalmente laboratorios farmacéuticos) poseen un porcentaje de participación muy pequeño.
- Entre las instituciones más productivas destacan los centros franceses, especialmente el CNRS, INSERM y la Universidad de París, que sin embargo no reciben un volumen de citas equivalente.
- Las instituciones más productivas son británicas, alemanas y francesas, mientras que el resto de los países cuentan con una representación muy baja. Sin embargo, en Bélgica, la Universidad Católica de Lovaina, aparece como la segunda más productiva, y una de las instituciones más importantes en el área de las enfermedades raras.
- En Holanda la producción científica se encuentra mucho más concentrada que en España, especialmente en las universidades, mientras que en España el volumen de trabajos se encuentra muy disperso entre universidades y hospitales, acercándose más al perfil de los pequeños productores, mientras que el de Holanda se asemeja más al de los grandes productores.

- La presencia en España de empresas y fundaciones se asemeja más al perfil de los grandes productores, por lo que se puede concluir que está en proceso de convertirse en uno de ellos.
- Los trabajos que alcanzan una mayor repercusión proceden de los centros nacionales de investigación, que acumulan un número de citas por documento mucho mayor que el resto de instituciones.
- En el área de las enfermedades raras no abundan los ensayos clínicos, por lo que el tipo documental que más citas les suele reportar a los hospitales, se ve muy mermado y repercute en un menor número de citas para este tipo de instituciones.
- El volumen de publicaciones no es un indicativo de su calidad, especialmente en el caso del INSERM, que se sitúa entre los centros que poseen una mayor producción no citada y que publican menos trabajos en revistas de alto impacto.
- Las instituciones muestran una clara preferencia para colaborar con los centros geográficamente próximos, por lo que la distancia entre los diferentes centros investigadores es un factor clave para su colaboración, especialmente en este área.
- Las preferencias de los centros de investigación por las temáticas concretas varían según el tamaño de los centros. Mientras los grandes centros no se especializan en ninguna materia concreta (especialmente aquellos con mayor carga de investigación genética), los pequeños realizan la mayor parte de su investigación en torno a enfermedades concretas.

13.6. Autores

- El volumen de autores que conforma el núcleo en ambas bases de datos es muy similar.

- Existe en un gran número de autores transeúntes, de acuerdo con los datos observados en la distribución de Lotka y el índice de élite de Price.
- El índice de coautoría oscila entre 6 y 8 autores por trabajo, dependiendo fundamentalmente del área temática, así como de los países y tipos institucionales.
- Los investigadores en enfermedades raras publican la mayor parte de sus trabajos en revistas del área de Genética y Bioquímica-Biología Molecular, Microbiología, Sistema Respiratorio, Pediatría y en menor medida Neurología y Neurociencia.
- La mayor parte de los autores considerados grandes productores están relacionados con la Genética y la Biología Molecular, mientras que los pequeños productores se centran en áreas más especializadas.
- Las enfermedades con un mayor grado de autores especializados son la Fibrosis Quística, la Osteonecrosis y en menor medida los trastornos mentales y del comportamiento, indicando una mayor diferencia entre estas enfermedades y el resto de enfermedades raras, más relacionadas entre si.
- Las relaciones entre los autores son muy estrechas, formando redes muy cohesionadas, especialmente entre aquellos que trabajan de forma más activa en Genética, Biología Molecular, Microbiología, Sistema Respiratorio y Pediatría. Sin embargo, los investigadores en el área de Neurología y Neurociencia poseen unas relaciones más difusas.
- La distribución de autores y sus áreas de interés son muy similares, tanto para los autores citados como los productores, con la salvedad de que, entre los autores citados, aparecen algunas áreas secundarias con cierto peso, como la Hematología e Inmunología y la Urología.
- La Genética y la Biología Molecular sirven como nexo de unión entre los autores, tanto citados como citantes, actuando como áreas interdisciplinares que se encuentran relacionadas con prácticamente todas las investigacio-

nes realizadas sobre enfermedades raras, independientemente del área de especialización.

Recomendaciones

A raíz del trabajo realizado, se pueden ofrecer una serie de recomendaciones orientadas a la mejora de algunos aspectos concretos de la investigación en enfermedades raras:

Es necesario que se continúen desarrollando planes de acción, tanto nacionales como supranacionales, que financien la investigación en enfermedades raras, pues el gasto sanitario influye notablemente en la producción científica en enfermedades raras.

Las asociaciones de pacientes mejoran notablemente la visibilidad, y por tanto, la investigación que se realiza sobre las enfermedades concretas, por lo que es recomendable que los enfermos se organicen en este tipo de instituciones.

En el caso de los países emergentes y especialmente en el de España, es necesario que continúen colaborando con aquellos países más productivos y con una mayor capacidad investigadora, hasta consolidar su posición, debido a que la actual situación requiere una colaboración muy activa entre los diferentes actores de la comunidad investigadora, pues de lo contrario corren el riesgo de que su investigación se vea estancada.

Es necesario hacer hincapié en el diagnóstico precoz de las enfermedades raras y en la investigación de los primeros síntomas de las mismas, especialmente en

las edades más tempranas, pues el diagnóstico precoz de las enfermedades raras supone un número notablemente mayor de oportunidades de tratamiento y supervivencia de los enfermos.

Es primordial conseguir que las revistas científicas perciban la necesidad de publicar trabajos relativos a la investigación en enfermedades raras, puesto que en la mayor parte de los casos estos trabajos no alcanzan una mayor visibilidad debido a las reticencias por parte de las revistas para publicarlos.

En la misma línea, es necesario dotar de mayor visibilidad a la investigación que se realiza en estas áreas, para evitar el efecto negativo que pueden producir en las revistas y contribuir a una mayor difusión de las investigaciones realizadas en estas áreas.

En cuanto a la colaboración entre autores, es prácticamente un requisito imprescindible en este tipo de enfermedades, especialmente cuando se lleva a cabo con autores de áreas interdisciplinarias como la Genética o la Biología Molecular, que le otorgan un valor añadido a la investigación realizada y la dotan de mayor visibilidad.

Capítulo V

Bibliografía y anexos

Bibliografía

(2010). «Prevalence of rare diseases: Bibliographic data». *Informe técnico*, INSERM. Consultado el 11 de agosto de 2012.

http://www.orpha.net/orphacom/cahiers/docs/GB/Prevalence_of_rare_diseases_by_alphabetical_list.pdf

ABAD-GARCÍA, F; GONZÁLEZ-TERUEL, A; CORT ZAMAKOLA, J; GIMENEZ MARTINEZ, F; SANZ-CASADO, E; LASCURAIN SÁNCHEZ, M; GARCÍA-ZORITA, C y CHISBERT PERALES, M (2009). *Producción científica de la Comunitat Valenciana en materias de biomedicina y ciencias de la salud a través de las bases de datos del Institute of Scientific Information (ISI) periodo 2004-2006 y evolución 2000-2006*. Generalitat Valenciana, Conselleria de Sanitat, Direcció General de Ordenació, Evaluació e Investigació Sanitaria, Valencia.

AHMED, T (2003). «String::Trigram - Find similar strings by trigram (or 1, 2, 4, etc.-gram) method». Consultado el 30 de mayo de 2012.

<http://search.cpan.org/~tareka/String-Trigram-0.11/Trigram.pm>

AKSNES, DW (2006). «Citation rates and perceptions of scientific contribution». *Journal of the American Society for Information Science and Technology*, 57(2), pp. 169–185.

- AL-SHAHI, R; WILL, RG y WARLOW, CP (2001). «Amount of research interest in rare and common neurological conditions: bibliometric study». *British Medical Journal*, 323.
- ALBARRÁN, P y CRESPO, J (2010). «A comparison of the scientific performance of the US and the European Union at the turn of the 21st century». *Scientometrics*, 85, pp. 329–344.
- ALFONSO, F; BERMEJO, J y SEGOVIA, J (2005). «Impactología, impactitis, impactoterapia». *Revista Española de Cardiología*, 58(10), pp. 1239–1245.
- ALIANZA 4 UNIVERSIDADES (2012). «Observatorio IUNE». Consultado el 18 de septiembre de 2012.
<http://iune.es/>
- ARENCIBIA, JR y MOYA ANEGÓN, F (2008). «La evaluación de la investigación científica: una aproximación teórica desde la cienciometría». *Acimed*, 17(4), pp. 1–27.
- ARIZO, V; BOTELLO, P; ABELLA, C; MARTÍNEZ, M; GARCÍA, M y AMADOR, A (2005). *Aproximación a las Enfermedades Raras en la Comunidad de Valencia en el periodo 1999-2003*. Servicio de Estudios Epidemiológicos y Estadísticas Sanitarias. Área de Epidemiología. Dirección General de salud Pública, Valencia.
- ARRIAZA GOMEZ, AJ; FERNÁNDEZ PALACÍN, F; LÓPEZ SÁNCHEZ, MA; MUÑOZ MÁRQUEZ, M; PÉREZ PLAZA, S y SÁNCHEZ NAVAS, A (2008). *Estadística básica con R y R-commander*. Universidad de Cádiz, Cádiz. ISBN 978-84-9828-186-6.
- ARRISCADO, JN; MATIAS, M y FILIPE, AM (2007). «As organizações de pacientes como atores emergentes no espaço da saúde: o caso de Portugal». *Revista Eletrônica de Comunicação Informação & Inovação em Saúde*, 1(1), pp. 107–110.

- AVELLANEDA, A; IZQUIERDO, M; TORRENT-FARNELL, J y RAMÓN, J (2007). «Enfermedades raras: enfermedades crónicas que requieren un nuevo enfoque sociosanitario». *Anales del Sistema Sanitario de Navarra*, 30(2), pp. 177–190.
- AYMÉ, S (2006). «Maladies rares: 67 centres de référence pour une révolution culturelle ?» *La Presse Médicale*, 35(3), pp. 456–460.
- AYME, S; DALLAPICCOLA, B y DONNAI, D (2006). «Orphanet Journal of Rare Diseases: Launch Editorial». *Orphanet Journal of Rare Diseases*, 1(1), p. 1.
- BAÑÓN HERNÁNDEZ, A (2007). «Las enfermedades raras y su representación discursiva: Propuestas para un análisis crítico». *Discurso & Sociedad*, 1(2), pp. 188–229.
- BAÑÓN HERNÁNDEZ, AM; FORNIELES ALCARAZ, J; SOLVES ALMELA, JA y RIUS SANCHIS, I (Eds.) (2011). *Desafíos y estrategias comunicativas de las enfermedades raras: La investigación médica como referente*. Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Valencia. ISBN 978-84-694-0596-3.
- BAGGOTT, R y FORSTER, R (2008). «Health consumer and patients' organizations in Europe: towards a comparative analysis.» *Health expectations: an international journal of public participation in health care and health policy*, 11(1), pp. 85–94.
- BAR-ILAN, J (2008a). «Informetrics at the beginning of the 21st century a review». *Journal of Informetrics*, 2(1), pp. 1–52.
- (2008b). «Which h-index?—A comparison of WoS, Scopus and Google Scholar». *Scientometrics*, pp. 257–271.
- BAR-ILAN, J; LEVENE, M y LIN, A (2007). «Some measures for comparing citation databases». *Journal of Informetrics*, 1(1), pp. 26–34.
- BARJAK, F y ROBINSON, S (2008a). «International collaboration, mobility and team diversity in the life sciences: impact on research performance». *Social*

- Geography Discussions*, (3), pp. 23–36.
- (2008b). «International collaboration, mobility and team diversity in the life sciences: impact on research performance». *Social Geography*, 3(1), pp. 23–36.
- BATES, D; CHAMBERS, J; DALGAARD, P; FALCON, S; GENTLEMAN, R; HORNIK, K; IACUS, S; IHAKA, R; LEISCH, F; LUMLEY, T; MAECHLER, M; MURDOCH, D; MURRELL, P; RIPLEY, B; SARKAR, D; LANG, DT; TIERNEY, L; URBANEK, S; SCHWARTE, H y MASAROTTO, G (2009). «The R Project for Statistical Computing». Consultado el 2 de marzo de 2012.
<http://www.r-project.org/>
- BELLANCA, L (2009). «Measuring interdisciplinary research: analysis of co-authorship for research staff at the University of York». *Bioscience Horizons*, 2(2), pp. 99–112.
- BILISOLY, R (2008). *Practical text mining with Perl*. Wiley Publishing, New Jersey.
- BISHOP, D (2010). «Which Neurodevelopmental Disorders Get Researched and Why?» *PLoS One*, 5(11).
- BLANCO MERLO, JR y IRANZO, JM (2000). «Ambivalencia e incertidumbre en las relaciones entre ciencia y sociedad». *Papers: revista de sociología*, (61), pp. 89–112.
- BORDONS, M (2004). «Hacia el reconocimiento internacional de las publicaciones científicas españolas». *Revista Española de Cardiología*, 57(9), pp. 799–802.
- BORDONS, M y ZULUETA, MA (1999). «Evaluación de la actividad científica a través de indicadores bibliométricos». *Revista Española de Cardiología*, 52(10), pp. 790–800.
- BORGMAN, CL y FURNER, J (2002). «Scholarly communication and bibliometrics». *Annual Review of Information Science and Technology*, 36, pp. 3–72.

- BORNMANN, L; MUTZ, R; NEUHAUS, C y DANIEL, HD (2008). «Citation counts for research evaluation: standards of good practice for analyzing bibliometric data and presenting and interpreting results». *Ethics in Science and Environmental Politics*, 8, pp. 93–102.
- BOWER, DJ (1996). «User-producer Interaction and the Case of Biomedical Innovation». *Journal of Industry Studies*, 3(1), pp. 21–34.
- (2005). «From the ‘rhetoric of hope’ to the ‘patient-active paradigm’—strategic positioning of pharmaceutical and biotechnology companies». *Technology Analysis & Strategic Management*, 17(2), pp. 183–204.
- BRADFORD, SC (1934). «Sources of information on specific subjects». *Engineering*, 137.
- BROOKES, BC (1969). «Bradford’s law and the bibliography of science». *Nature*, 224(5223), pp. 953–956.
- BROOKS, T (1987). «The Bibliometrics Toolbox. Version 2.8». Consultado el 15 de marzo de 2012.
<http://staff.washington.edu/tbrooks/vita/vita.html>
- BROWN, MC (2001). *Perl: manual de referencia*. McGraw-Hill Interamericana de España, Madrid. ISBN 84-481-2884-2.
- BUELA-CASAL, G (2003). «Evaluación de la calidad de los artículos y de las revistas científicas: propuestas del factor de impacto ponderado y de un índice de calidad». *Psicothema*, 15(1), pp. 23–35.
- BUEREN, J (2010). «Avances de la terapia génica en el tratamiento de enfermedades monogénicas». En: *9ª edición del curso de Biotecnología Aplicada a la Salud Humana*, capítulo 6. CIEMAT y Amgen, Madrid.
- BURRIGHT, MA; BELLARDO HAHN, T y ANTONISSE, MJ (2005). «Understanding Information Use in a Multidisciplinary Field: A Local Cita-

- tion Analysis of Neuroscience Research». *College & Research Libraries*, 66(3), pp. 198–211.
- BURTON, RE y KEBLER, RW (1960). «The half-life of some scientific and technical literatures». *American Documentation*, 11(1), pp. 18–22.
- CAÑEDO ANDALIA, R (1999). «Los análisis de citas en la evaluación de los trabajos científicos y las publicaciones seriadas». *Acimed*, 7(1), pp. 30–39.
- (2009). «Búsqueda bibliográfica, investigación métrica e inteligencia: el caso de la ataxia espinocerebelosa tipo 2 en Cuba». *Acimed*, 19(2).
- CABEZAS-CLAVIJO, A y TORRES-SALINAS, D (2012). «Google Scholar Citations y la emergencia de nuevos actores en la evaluación de la investigación». *Anuario ThinkEPI*, 6, pp. 147–153.
- CAMÍ, J (1997). «Impactología: diagnóstico y tratamiento». *Medicina Clínica*, 109(13), pp. 515–524.
- (2001). «Evaluación de la investigación biomédica». *Medicina Clínica*, 117(13), pp. 510–513.
- (2003). «Priorización de la investigación biomédica: implicaciones para la investigación en salud pública y servicios sanitarios». *Revista de Calidad Asistencial*, 488, pp. 54–90.
- CAMÍ, J; ZULUETA, M y FERNÁNDEZ, M (1997). «Producción científica española en biomedicina y ciencias de la salud durante el período 1990-1993 (Science Citation Index y Social Science Citation Index) y comparación con el período 1986-1989». *Medicina Clínica*, (109), pp. 481–496.
- CARRASCO, J y HERNÁN, MA (1993). *Estadística multivariante en las ciencias de la vida: fundamentos, métodos y aplicación*. Editorial Ciencia 3, Madrid. ISBN 84-86204-54-2.

- CARRASCO MALLEN, M; ANTIÑOLO GIL, G; MARRUGAT DE LA IGLESIA, J y CALDERÓN GÓMEZ, C (2005). «La colaboración interredes: una herramienta para la excelencia». *Atención primaria*, 36(Suplemento 1), pp. 101–112.
- CASTILLO-ESPARCIA, A (2012). «La investigación en Comunicación. Análisis bibliométrico de las revistas de mayor impacto del ISI». *Revista Latina de Comunicación Social*, (67), pp. 248–270.
- CAVNAR, WB y TRENKLE, JM (1994). «N-Gram-Based Text Categorization». En: *Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval*, pp. 161–175. Las Vegas, US.
- CHINCHILLA-RODRÍGUEZ, Z (2010). «Indicadores bibliométricos de España en el Mundo 2008». Una version de este trabajo ha sido publicada en Informe Cotec 2010: Tecnología e innovación en España (pp. 41-47). Madrid: Fundación Cotec.
- CIBERER (2006). «Centro de Investigación Biomédica en Red de Enfermedades Raras». Consultado el 8 de agosto de 2012.
<http://www.ciberer.es/>
- CLARKE, A; GATINEAU, M; GRIMAUD, O; ROYER-DEVAUX, S; WYN-ROBERTS, N; LE BIS, I y LEWISON, G (2007). «A bibliometric overview of public health research in Europe.» *European Journal of Public Health*, 17(0), pp. 43–9.
- COCKBURN, I (1996). «Public-private interaction in pharmaceutical research». *Proceedings of the National Academy of Sciences*, 93(23), pp. 12725–12730.
- COLLINS, F; MORGAN, M y PATRINOS, A (2003). «The Human Genome Project: lessons from large-scale biology». *Science*, 300(5617), p. 286.
- CONWAY, D (2005). «Smart::Comments - Comments that do more than just sit there». Consultado el 19 de abril de 2012.
<http://search.cpan.org/~dconway/Smart-Comments-v1.0.3/lib/Smart/Comments.pm>

- CONWAY, SP (2004). «Effect of oral bisphosphonates on bone mineral density and body composition in adult patients with cystic fibrosis: a pilot study». *Thorax*, 59(8), pp. 699–703.
- CORDY, C y GRIFFITH, D (1993). «Efficiency of least squares estimators in the presence of spatial autocorrelation». *Communications in Statistics-Simulation and Computation*, 22(4), pp. 1161–1179.
- COSTAS, R; MORENO, L y BORDONS, M (2008). «Solapamiento y singularidad de MEDLINE, WoS e IME para el análisis de la actividad científica de una región en Ciencias de la Salud». *Revista Española de Documentación Científica*, 31(3).
- CPAN (2009). «The CPAN Search Site - search.cpan.org». Consultado el 5 de marzo de 2012.
<http://search.cpan.org/>
- CSARDI, G y NEPUSZ, T (2006). «The igraph software package for complex network research». *InterJournal, Complex Systems*, p. 1695.
- DALPÉ, R (2002). «Bibliometric analysis of biotechnology». *Scientometrics*, 55(2), pp. 189–213.
- DE GRANDA ORIVE, J (2003). «Algunas reflexiones y consideraciones sobre el factor de impacto». *Archivos de bronconeumología*, 39(9), pp. 409–417.
- DE LA CHAPELLE, A (1993). «Disease gene mapping in isolated human populations: the example of Finland.» *Journal of medical genetics*, (30), pp. 857–865.
- DE PABLO DÁVILA, F (2009). «Células madre neurales, neurogénesis y neuroprotección». En: *Monografías de la Real Academia Nacional de Farmacia. Monografía XXVII: Células Madre y Terapia regenerativa*, Real Academia Nacional de Farmacia, Madrid.
- DELNOLJ, DMJ y GROENEWEGEN, PP (2007). «Health services and systems research in Europe: overview of the literature 1995–2005». *European Journal of Public Health*, 17(Supplement 1), p. 2007.

DJEBALI, S; DAVIS, CA; MERKEL, A; DOBIN, A; LASSMANN, T; MORTAZAVI, A; TANZER, A; LAGARDE, J; LIN, W; SCHLESINGER, F; XUE, C; MARINOV, GK; KHATUN, J; WILLIAMS, BA; ZALESKI, C; ROZOWSKY, J; RÖDER, M; KOKOCINSKI, F; ABDELHAMID, RF; ALIOTO, T; ANTOSHECHKIN, I; BAER, MT; BAR, NS; BATUT, P; BELL, K; BELL, I; CHAKRABORTTY, S; CHEN, X; CHRAST, J; CURADO, J; DERRIEN, T; DRENKOW, J; DUMAIS, E; DUMAIS, J; DUTTAGUPTA, R; FALCONNET, E; FASTUCA, M; FEJES-TOTH, K; FERREIRA, P; FOISSAC, S; FULLWOOD, MJ; GAO, H; GONZALEZ, D; GORDON, A; GUNAWARDENA, H; HOWALD, C; JHA, S; JOHNSON, R; KAPRANOV, P; KING, B; KINGSWOOD, C; LUO, OJ; PARK, E; PERSAUD, K; PREALL, JB; RIBECA, P; RISK, B; ROBYR, D; SAMMETH, M; SCHAFFER, L; SEE, LH; SHAHAB, A; SKANCKE, J; SUZUKI, AM; TAKAHASHI, H; TILGNER, H; TROUT, D; WALTERS, N; WANG, H; WROBEL, J; YU, Y; RUAN, X; HAYASHIZAKI, Y; HARROW, J; GERSTEIN, M; HUBBARD, T; REYMOND, A; ANTONARAKIS, SE; HANNON, G; GIDDINGS, MC; RUAN, Y; WOLD, B; CARNINCI, P; GUIGÓ, R y GINGERAS, TR (2012). «Landscape of transcription in human cells». *Nature*, 489(7414), pp. 101–108.

DONG, P; LOH, M y MONDRY, A (2005). «The impact factor revisited.» *Biomedical digital libraries*, 2, p. 7.

DRESSER, R (2006). «Private-sector research ethics: marketing or good conflicts management? The 2005 John J. Conley Lecture on Medical Ethics.» *Theoretical medicine and bioethics*, 27(2), pp. 115–39.

DUBOIS, P (2009). *MySQL*. Developer's library. Addison-Wesley, Upper Saddle River, NJ, 4th ededición. ISBN 0-672-32938-7.

DUNHAM, I; KUNDAJE, A; ALDRED, SF; COLLINS, PJ; DAVIS, CA; DOYLE, F; EPSTEIN, CB; FRIETZE, S; HARROW, J; KAUL, R; KHATUN, J; LAJOIE, BR; LANDT, SG; LEE, BK; PAULI, F; ROSENBLOOM, KR; SABO, P; SAFI, A; SANYAL, A; SHORESH, N; SIMON, JM; SONG, L; TRINKLEIN, ND; ALTSHULER, RC; BIRNEY, E; BROWN, JB; CHENG, C; DJEBALI, S; DONG, X;

ERNST, J; FUREY, TS; GERSTEIN, M; GIARDINE, B; GREVEN, M; HARDISON, RC; HARRIS, RS; HERRERO, J; HOFFMAN, MM; IYER, S; KELLIS, M; KHERADPOUR, P; LASSMANN, T; LI, Q; LIN, X; MARINOV, GK; MERKEL, A; MORTAZAVI, A; PARKER, SCJ; REDDY, TE; ROZOWSKY, J; SCHLESINGER, F; THURMAN, RE; WANG, J; WARD, LD; WHITFIELD, TW; WILDER, SP; WU, W; XI, HS; YIP, KY; ZHUANG, J; BERNSTEIN, BE; GREEN, ED; GUNTER, C; SNYDER, M; PAZIN, MJ; LOWDON, RF; DILLON, LAL; ADAMS, LB; KELLY, CJ; ZHANG, J; WEXLER, JR; GOOD, PJ; FEINGOLD, EA; CRAWFORD, GE; DEKKER, J; ELNITSKI, L; FARNHAM, PJ; GIDDINGS, MC; GINGERAS, TR; GUIGÓ, R; HUBBARD, TJ; JAMES KENT, W; LIEB, JD; MARGULIES, EH; MYERS, RM; STAMATOYANNOPOULOS, JA; TENENBAUM (2012). «An integrated encyclopedia of DNA elements in the human genome». *Nature*, 489(7414), pp. 57–74.

E-RARE (2011). «EURO-CDG». Consultado el 10 de julio de 2012.

<http://e-rare.eu/content/euro-cdg>

ECKER, JR; BICKMORE, WA; BARROSO, I; PRITCHARD, JK; GILAD, Y y SEGAL, E (2012). «Genomics: ENCODE explained». *Nature*, 489(7414), pp. 52–55.

EGGHE, L (2000). «The distribution of N-grams». *Scientometrics*, 47(2), pp. 237–252.

EGGHE, L y ROUSSEAU, R (1990). *Introduction to informetrics: Quantitative Methods in Library, Documentation and Information Science*. Elsevier Science Publishers, Amsterdam. ISBN 0-444-88493-9.

ESCUDERO GÓMEZ, C; MILLÁN SANTOS, I y POSADA DE LA PAZ, M (2005). «Análisis de la producción científica española sobre enfermedades raras: 1990-2000». *Medicina Clínica*, 125(9), pp. 329–332.

EURORDIS (2005). «What is a rare disease?». Consultado el 24 de agosto de 2012.

http://www.eurordis.org/article.php3?id_article=252

- FALAGAS, ME; PITSOUNI, EI; MALIETZIS, GA y PAPPAS, G (2008). «Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses.» *FASEB Journal: Official publication of the Federation of American Societies for Experimental Biology*, 22(2), pp. 338–42.
- FAVUS, MJ (2007). «Diabetes and the risk of osteonecrosis of the jaw.» *The Journal of clinical endocrinology and metabolism*, 92(3), pp. 817–8.
- FERNANDEZ-REAL, JM; LOPEZ-BERMEJO, A y RICART, W (2002). «Cross-Talk Between Iron Metabolism and Diabetes.» *Diabetes*, 51(8), pp. 2348–2354.
- FIELD, MJ y BOAT, TF (Eds.) (2011). *Rare Diseases and Orphan Products: Accelerating Research and Development*. The National Academies Press, Washington, D.C. ISBN 978-0-309-15806-0.
- FISCHER, A; BORENSZTEIN, P y ROUSSEL, C (2005). «The European rare diseases therapeutic initiative.» *PLoS medicine*, 2(9).
- FLOREZ, J y PAZOS, A (1987). «Neurotransmisión en el sistema nervioso central». En: J Flórez; J Armijo y A Mediavilla (Eds.), *Farmacología Humana*, capítulo 24, pp. 409–433. Masson, Barcelona.
- FORMAN, J; TARUSCIO, D; LLERA, VA; BARRERA, LA; COTÉ, TR; EDFJÄLL, C; GAVHED, D; HAFFNER, ME; NISHIMURA, Y; POSADA, M; TAMBUYZER, E; GROFT, SC y HENTER, JI (2012). «The need for worldwide policy and action plans for rare diseases.» *Acta paediatrica*.
- FRIGOTTO, M y RICCABONI, M (2011). «A few special cases: scientific creativity and network dynamics in the field of rare diseases.» *Scientometrics*, 89(1), pp. 397–420.
- FRUCHTERMAN, TMJ y REINGOLD, EM (1991). «Graph drawing by force-directed placement.» *Software: Practice and Experience*, 21(11), pp. 1129–1164.
- FUJITA, R (2007). «Genómica y su aplicación en producción animal.» *Archivos Latinoamericanos de Produccion Animal*, 15(Suplemento 1), pp. 67–68.

- GARCÍA GONZALEZ, PE (2006). «Tratamiento automatizado de datos bibliográficos mediante uso del lenguaje perl, y desarrollo de una interfaz web para su análisis bibliométrico descriptivo». Director: José Carlos García Zorita.
- (2010). *Diseño, desarrollo y aplicación de un método para el análisis y tratamiento de la información con fines métricos*. Tesis doctoral, Universidad Carlos III de Madrid, Departamento de Biblioteconomía y Documentación. Director: José Carlos García Zorita.
- GARCÍA SÁNCHEZ, JE y GARCÍA SÁNCHEZ, E (2005). «Rare diseases in the cinema». *Journal of Medicine and Movies*, (1), pp. 93–94.
- GARCÍA SILVA, M (2000). «Errores congénitos del metabolismo con repercusión sobre el sistema nervioso del recién nacido. ¿Cuándo y cómo investigarlos?». *Revista de Neurología*, 31(5).
- GARCÍA-ZORITA, C (2000). *La actividad científica de los economistas españoles en función del ámbito nacional o internacional de sus publicaciones: estudio comparativo basado en un análisis bibliométrico durante el periodo 1986/95*. Tesis doctoral, Universidad Carlos III de Madrid, Departamento de Biblioteconomía y Documentación. Director: Elias Sanz Casado.
- GARFIELD, E (1972). «Citation Analysis as a Tool in Journal Evaluation Journals can be ranked by frequency and impact of citations for science policy studies». *Science*, 178(4060), pp. 471–479.
- GARFIELD, E y MERTON, RK (1979). *Citation indexing: Its theory and application in science, technology, and humanities*. Wiley, New York. ISBN 0-471-02559-3.
- GERSTEIN, MB; KUNDAJE, A; HARIHARAN, M; LANDT, SG; YAN, KK; CHENG, C; MU, XJ; KHURANA, E; ROZOWSKY, J; ALEXANDER, R; MIN, R; ALVES, P; ABYZOV, A; ADDLEMAN, N; BHARDWAJ, N; BOYLE, AP; CAYTING, P; CHAROS, A; CHEN, DZ; CHENG, Y; CLARKE, D; EASTMAN, C; EUSKIRCHEN, G; FRIETZE, S; FU, Y; GERTZ, J; GRUBERT, F; HARMANCI, A; JAIN, P;

- KASOWSKI, M; LACROUTE, P; LENG, J; LIAN, J; MONAHAN, H; O'GEEN, H; OUYANG, Z; PARTRIDGE, EC; PATACSIL, D; PAULI, F; RAHA, D; RAMIREZ, L; REDDY, TE; REED, B; SHI, M; SLIFER, T; WANG, J; WU, L; YANG, X; YIP, KY; ZILBERMAN-SCHAPIRA, G; BATZOGLOU, S; SIDOW, A; FARNHAM, PJ; MYERS, RM; WEISSMAN, SM y SNYDER, M (2012). «Architecture of the human regulatory network derived from ENCODE data». *Nature*, 489(7414), pp. 91–100.
- GILFILLAN, I (2003). *La biblia de MySQL*. Anaya Multimedia, Madrid. ISBN 978-84-415-1558-1.
- GLÄNZEL, W (2000). «Science in Scandinavia: A bibliometric approach». *Scientometrics*, 48(2), pp. 121–150.
- GLÄNZEL, W y MOED, HF (2002). «Journal impact measures in bibliometric research». *Scientometrics*, 53(2), pp. 171–193.
- GLÄNZEL, W y SCHUBERT, A (2004). «Analyzing Scientific Collaboration through Co-Authorship». En: *Handbook of quantitative science and technology research. The use of publication and patent statistics in studies on S&T systems*, pp. 257–276. Kluwer Academic Publishers, Dordrecht. ISBN 978-1-4020-2702-4.
- GLÄNZEL, W y SCHOEPFLIN, U (1999). «A bibliometric study of reference literature in the sciences and social sciences». *Information Processing & Management*, 35(1), pp. 31–44.
- GÓMEZ, I (1995). «Collaboration patterns of Spanish scientific publications in different research areas and disciplines». En: *Proceedings of the fifth biennial international conference of the International Society for Scientometrics and Infometrics*, ETATS-UNIS, River Forest.
- GÓMEZ CARIDAD, I y BORDONS, M (1996). «Limitaciones en el uso de los indicadores bibliométricos para la evaluación científica». *Política científica*, 46, pp. 21–26.

GÓMEZ CARIDAD, I; FERNÁNDEZ MUÑOZ, M; BORDONS GANGAS, M y MORILLO ARIZA, F (2004). «La producción científica española en Medicina en los años 1994-1999». *Revista Clínica Española*, 204(2), pp. 75–88.

GÓMEZ CARIDAD, I; SANCHO LOZANO, R; BORDONS, M y FERNÁNDEZ MUÑOZ, M (2006). «La I+D en España a través de publicaciones y patentes». En: J Sebastián y E Muñoz (Eds.), *Radiografía de la investigación pública en España*, Biblioteca Nueva, Madrid. ISBN 84-9742-540-5.

GONZÁLEZ-LAMUÑO, D y GARCÍA FUENTES, M (2008). «Enfermedades de base genética Genetically based diseases». *Anales del Sistema Sanitario de Navarra*, 31(Suplemento 2), pp. 105–126.

GREENACRE, M (2008). *La práctica del análisis de correspondencias*. Fundación BBVA, Bilbao. ISBN 978-84-96515-71-0.

— (2010). «ca: Simple, Multiple and Joint Correspondence Analysis. R package version 0.33». Consultado el 7 de marzo de 2012.

<http://cran.r-project.org/web/packages/car/index.html>

GROSCURTH, P (1996). «New muscle — old story?» *Nature Medicine*, 2(11), pp. 1162–1163.

GUILLÉ, ENRIQUEZ, J; DURÁN PLÁ, E; PASTOR GRACIA, MA y RODRÍGUEZ ROMERO, E (2004). *Aproximación a la Situación de las Enfermedades Raras en Andalucía*. Consejería de Salud; RepiER, Sevilla.

GUILLÉN-NAVARRO, E; BALLESTA-MARTÍNEZ, MJ y LÓPEZ-GONZÁLEZ, V (2011). «Genética y enfermedad. Concepto de genética médica». *Revista de Nefrología*, 2(1), pp. 3–10.

HAINES, M y REDMAN, S (2010). «Putting science to work for health care reform: how much research is available to support improvements to our hospitals?» *The Medical Journal of Australia*, 192(11), pp. 646–650.

- HAIR, J; ANDERSON, R; TATHAM, R y BLACK, W (1999). *Análisis multivariante*. Prentice Hall Iberia, Madrid. ISBN 84-8322-035-0.
- HALL, PAV y DOWLING, GR (1980). «Approximate string matching». *ACM Computing Surveys (CSUR)*, 12(4), pp. 381–402.
- HANNEMANN-WEBER, H; KESSEL, M y SCHULTZ, C (2012). «Research performance of centers of expertise for rare diseases—The influence of network integration, internal resource access and operational experience». *Health Policy*, 105(2-3), pp. 138–145.
- HANSEN, CR; PRESSLER, T; KOCH, C y HØIBY, N (2005). «Long-term azitromycin treatment of cystic fibrosis patients with chronic *Pseudomonas aeruginosa* infection; an observational cohort study.» *Journal of Cystic Fibrosis: Official Journal of the European Cystic Fibrosis Society*, 4(1), pp. 35–40.
- HARVEY, W (1907). *The circulation of the blood: and other writings*. J.M. Dent & Sons; E.P. Dutton, London: New York.
- HEEMSTRA, H (2009). *From research on rare diseases to new orphan drug development*. Tesis doctoral, Utrecht University. Directores: H.G.M. Leufkens y H.A. Büller.
- HEEMSTRA, H y VAN WEELY, S (2009). «Translation of rare disease research into orphan drug development: disease matters». *Drug discovery today*, 14(23-24).
- HERRING, ME y SHAH, SK (2006). «Periodontal Disease and Control of Diabetes Mellitus». *J Am Osteopath Assoc*, 106(7), pp. 416–421.
- HIRSCH, JE (2005). «An index to quantify an individual's scientific research output». *Proceedings of the National Academy of Sciences*, 102(46), pp. 16569–16572.
- HOLTZMAN, NA; BERNHARDT, BA; MOUNTCASTLE-SHAH, E; RODGERS, JE; TAMBOR, E y GELLER, G (2005). «The quality of media reports on discoveries related to human genetic diseases.» *Community Genetics*, 8(3), pp. 133–44.

- INERGEN (2000). «Instituto de Investigación de Enfermedades Raras de Base Genética». Consultado el 31 de enero de 2009.
<http://inergen.retics.net/inergen/home.aspx>
- INGWERSEN, P y WORMELL, I (1999). «Publication behaviour and international impact: Scandinavian clinical and social medicine, 1988–96». *Scientometrics*, 46(3), pp. 487–499.
- ISCI III (1986). «Instituto de Salud Carlos III». Consultado el 12 de septiembre de 2012.
<http://www.isciii.es/>
- JAMES, DA (2007). «{CRAN} - Package {DBI}». Consultado el 30 de abril de 2012.
<http://cran.r-project.org/web/packages/DBI/index.html>
- JAMES, DA; S., D y HORNER, J (2009). «{CRAN} - Package {RMySQL}». Consultado el 30 de mayo de 2012.
<http://cran.r-project.org/web/packages/RMySQL/>
- JIMÉNEZ, FS (2010). «Una propuesta de la biología de sistemas para el estudio de enfermedades raras». *Uciencia*, 4, pp. 44–47.
- JIMENEZ CONTRERAS, E (2000). «Los métodos bibliométricos». *Cuadernos de Documentación Multimedia*, (10), pp. 757–771.
- JIMÉNEZ-CONTRERAS, E (2002). «Impact-factor rewards affect Spanish research». *Nature*, 417, p. 898.
- JONES, T; HANNEY, S y BUXTON, MJ (2006). «The journals of importance to UK clinicians: a questionnaire survey of surgeons». *BMC Medical Informatics and Decision Making*, 6(24).
- (2007). «The information sources and journals consulted or read by UK paediatricians to inform their clinical practice and those which they consider important». *BMC Pediatrics*, 7(1).

- JONES, T; HANNEY, S; BUXTON, MJ y BURNS, T (2004). «What British psychiatrists read Questionnaire survey of journal usage among clinicians». *The British Journal of Psychiatry*, (185), pp. 251–257.
- JUNTA DE EXTREMADURA (2005). «Enfermedades raras en Extremadura». Mérida: Junta de Extremadura, Consejería de Sanidad y Consumo, p. 174.
- KATZ, J y HICKS, D (1997). «Desktop scientometrics». *Scientometrics*, 38(1), pp. 141–153.
- KHAMAISSI, M; REGEV, E; YAROM, N; AVNI, B; LEITERSDORF, E; RAZ, I y ELAD, S (2007). «Possible association between diabetes and bisphosphonate-related jaw osteonecrosis». *The Journal of Clinical Endocrinology and Metabolism*, 92(3), pp. 1172–5.
- KING, D (2004). «The scientific impact of nations». *Nature*, 430, pp. 311–316.
- KING, J (1987). «A review of bibliometric and other science indicators and their role in research evaluation». *Journal of Information Science*, 13(5), pp. 261–276.
- KOSTOFF, R; BLOCK, J y SOLKA, J (2009). «Literature-Related Discovery». *Annual Review of Information Science and Technology*, 43(1), pp. 1–71.
- KRUSKAL, JB (1964). «Nonmetric multidimensional scaling». *Psychometrika*, 29(3), pp. 1–27.
- LASCURAIN-SÁNCHEZ, ML; GARCÍA-ZORITA, C; MARTÍN-MORENO, C; SUÁREZ-BALSEIRO, C y SANZ-CASADO, E (2008). «Impact of health science research on the Spanish health system, based on bibliometric and healthcare indicators». *Scientometrics*, 77(1), pp. 131–146.
- LASCURAIN SÁNCHEZ, ML (2006). «La evaluación de la actividad científica mediante indicadores bibliométricos». *Bibliotecas: Revista de la Escuela de Bibliotecología, Documentación e Información*, 24(1 y 2), pp. 9–26.

- LEEUWEN, TV; MOED, H y TIJSSEN, R (2001). «Language biases in the coverage of the Science Citation Index and its consequences for international comparisons of national research performance». *Scientometrics*, 51(1), pp. 335–346.
- LEHMANN, M (2007). «String::Similarity - calculate the similarity of two strings». Consultado el 27 de marzo de 2012.
<http://search.cpan.org/~mlehmann/String-Similarity-1.03/Similarity.pm>
- LEIMKUHLER, FF (1967). «The Bradford Distribution.» *Journal of Documentation*, 23(3), pp. 197–207.
- LEYDESDORFF, L (2012). «World Shares of Publications of the USA, EU-27, and China Compared and Predicted using the New Interface of the Web-of-Science versus Scopus». *El Profesional de la Información*, 21(1), pp. 43–49.
- LEYDESDORFF, L y VAUGHAN, L (2006). «Co-occurrence matrices and their applications in information science: Extending ACA to the Web environment». *Journal of the American Society for Information Science and Technology*, 57(12), pp. 1616–1628.
- LIN, D (1998). «An information-theoretic definition of similarity». En: *Proceedings of the Fifteenth International Conference on Machine Learning*, pp. 296–304.
- LÓPEZ-MUÑOZ, F; ALAMO, C; QUINTERO-GUTIÉRREZ, FJ y GARCÍA-GARCÍA, P (2008). «A bibliometric study of international scientific productivity in attention-deficit hyperactivity disorder covering the period 1980–2005». *European Child & Adolescent Psychiatry*, 17(6), pp. 381–391.
- LÓPEZ-MUÑOZ, F; VIETA, E; RUBIO, G; GARCÍA-GARCÍA, P y ALAMO, C (2006). «Bipolar disorder as an emerging pathology in the scientific literature: a bibliometric approach.» *Journal of affective disorders*, 92(2-3), pp. 161–70.
- LOPEZ PIÑERO, JM y TERRADA, ML (1992a). «Los indicadores bibliométricos y la evaluación de la actividad médico-científica. (I) Usos y abusos de la

- bibliometria». *Medicina Clínica*, 98(2), pp. 64–68.
- (1992b). «Los indicadores bibliometricos y la evaluacion de la actividad medico-cientifica. (III). Los indicadores de produccion, circulacion y dispersion, consumo de la informacion y repercusion». *Medicina Clínica*, 98(4), pp. 142–148.
- (1992c). «Los indicadores bibliometricos y la evaluacion de la actividad medico-cientifica. (IV). La aplicacion de los indicadores». *Medicina Clínica*, 98(10), pp. 384–388.
- LOTKA, AJ (1926). «The frequency distribution of scientific productivity». *Journal of the Washington academy of sciences*, 16(12), pp. 317–323.
- LUENGO, S; ARANDA, MT y LA FUENTE, M (2001). *Enfermedades Raras: Situación y demandas sociosanitarias*. Ministerio de Trabajo y Asuntos Sociales. Coleccion estudios e Informes. Serie estudios no 5., Madrid. ISBN 84-8446-038-X.
- LUUKKONEN, T; PERSSON, O y SIVERTSEN, G (1992). «Understanding patterns of international scientific collaboration». *Science, Technology & Human Values*, 17(1), pp. 101–126.
- MACROBERTS, MH y MACROBERTS, BR (1996). «Problems of citation analysis». *Scientometrics*, 36(3), pp. 435–444.
- MAECHLER, M y ROUSSEEUW, P (2005). «Cluster analysis basics and extensions». Consultado el 22 de marzo de 2012.
<http://cran.r-project.org/web/packages/cluster/>
- MALLADA, J (1999). «Epidemiología de la esclerosis múltiple en España. Datos de prevalencia e incidencia». *Revista de Neurologia*, (29), pp. 864–867.
- MALTRÁS, B (2003). *Los indicadores bibliométricos*. Trea, Gijón, Asturias.
- MAN, JP; WEINKAUF, JG; TSANG, M y SIN, JHDD (2003). «Why do Some Countries Publish More Than Others? An International Comparison of Re-

- search Funding, English Proficiency and Publication Output in Highly Ranked General Medical Journals». *European Journal of Epidemiology*, 19(8), pp. 811–817.
- MARSHAKOVA, IV (1973). «System of document connections based on references». *Scientific and Technical Information Serial of VINITI*, 6(2), pp. 3–8.
- MARTIN, BR (1996). «The use of multiple indicators in the assessment of basic research». *Scientometrics*, 36(3), pp. 343–362.
- MARTIN, BR y IRVINE, J (1983). «Assessing basic research: Some partial indicators of scientific progress in radio astronomy». *Research Policy*, 12(2), pp. 61–90.
- MARTIN, N; DOULET, N; HIVERT, V y AYMÉ, S (2010). «Rare diseases research in Europe: an overview based on data from the Orphanet database». *Orphanet Journal of Rare Diseases*, 5(Suplemento 1), p. 5.
- MARTÍN-ARRIBAS, M; POSADA, M; TERRACINI, B; CARBALLO, F y ABAITUA, I (2003). «Revisión de los aspectos éticos en la investigación biomédica: la experiencia del Comité de Ética del Centro de Investigación sobre el Síndrome del Aceite Tóxico y Enfermedades Raras (CISATER)». *Gaceta Sanitaria*, 17(6), pp. 512–514.
- MARTÍN ESCOFET, C (2007). *El lenguaje SQL*. UOC OpenCourseWare.
- MARTÍN-MORENO, C (1999). *Hábitos y necesidades de la información de los científicos experimentales encuadrados en el área de ciencias de la vida*. Tesis doctoral, Universidad Carlos III de Madrid, Departamento de Biblioteconomía y Documentación. Director: Elias Sanz Casado.
- MARTÍN-MORENO, C y SANZ-CASADO, E (1996). «Producción científica española en el área de Genética». *Revista Española de Documentación Científica*, 19(4), pp. 377–391.

- MCCARTHY, M (2010). «Who supports health research in Europe?» *European Journal of Public Health*, 20(1), pp. 3–5.
- MCKUSICK, VA (1966). *Mendelian Inheritance in Man: Catalogs of Autosomal Dominant, Autosomal Recessive, and X-linked Phenotypes*. Johns Hopkins University Press, Baltimore.
- (1998). *Mendelian Inheritance in Man: A Catalog of Human Genes and Genetic Disorders*. Johns Hopkins University Press, Baltimore.
- MCMILLAN, G; NARIN, F y DEEDS, DL (2000). «An analysis of the critical role of public science in innovation: the case of biotechnology». *Research Policy*, 29(1), pp. 1–8.
- MEHO, LI y SONNENWALD, DH (2000). «Citation ranking versus peer evaluation of senior faculty research performance: A case study of Kurdish scholarship». *Journal of the American Society for Information Science*, 51(2).
- MEHO, LI y YANG, K (2007). «A new era in citation and bibliometric analyses: Web of Science, Scopus, and Google Scholar». *Journal of American Society for Information Science and Technology*, 58(13), pp. 2105–2125.
- MEYER, D y BUCHTA, C (2009). «proxy: Distance and Similarity Measures». Consultado el 4 de marzo de 2012.
<http://cran.r-project.org/web/packages/proxy/index.html>
- MINISTERIO DE SANIDAD Y POLÍTICA SOCIAL (2009). *Estrategia en Enfermedades Raras del Sistema Nacional de Salud*. Consejo Interterritorial del Sistema Nacional de Salud. Ministerio de Sanidad y Política Social, Madrid.
- MOLAS-GALLART, J; SALTER, A; PATEL, P; SCOTT, A y DURAN, X (2002). «Measuring Third Stream Activities». *Informe técnico*, SPRU, University of Sussex.
- MOLINER, A y AYMÉ, S (2010). «The European Union Committee of Experts on Rare Diseases (EUCERD): a new committee to help the European Commission

- advance in the field of rare disease policy». *Orphanet Journal of Rare Diseases*, 5(Suplemento 1), p. 28.
- MORAVCSIK, MJ y MURUGESAN, P (1975). «Some results on the function and quality of citations». *Social Studies of Science*, 5(1), p. 86.
- MORAVCSIK, MJ y PÉREZ ALVAREZ-OSSORIO, JR (1989). «¿Cómo evaluar la ciencia y a los científicos?» *Revista Española de Documentación Científica*, 12(3), pp. 313–325.
- MOREIRO, JA (2002). «Aplicaciones al análisis automático del contenido provenientes de la teoría matemática de la información.» *Anales de Documentación*, 5(5), pp. 273–286.
- MORIN, A (2004). «Intensive use of correspondence analysis for information retrieval». En: *26th Int. Conf. Information Technology Interfaces IT*, pp. 255–258. Cavtat, Croacia.
- MOYA-ANEGÓN, F; HERRERO-SOLANA, V y JIMÉNEZ-CONTRERAS, E (2006). «A connectionist and multivariate approach to science maps: the SOM, clustering and MDS applied to library and information science research». *Journal of Information Science*, 32(1), pp. 63–77.
- MUÑOZ, E (1998). «Research and innovation policies in the new global economy: An international comparative analysis». En: P Laredo y P Mustar (Eds.), *Evolution des Systemes de Recherche et d'Innovation des Pays Industrialisés*, pp. 359–397. Edward Elgar, Paris.
- (2010). «Gobernanza, ciencia, tecnología y política: trayectoria y evolución». *Arbor*, CLXXXI(715), pp. 287–300.
- NADEL, E (1983). «Commitment and Co-Citation: An Indicator of Incommensurability in Patterns of Formal Communication». *Social Studies of Science*, 13, pp. 255–282.

- NAVARRO, A y MARTÍN, M (2004). «Scientific production and international collaboration in occupational health, 1992-2001». *Scandinavian Journal of Work Environmental & Health*, 30(3), pp. 223–233.
- NAVARRO, FA (2001a). «El inglés, idioma internacional de la medicina». *Panace*, 2(3), pp. 35–51.
- NAVARRO, G (2001b). «A guided tour to approximate string matching». *ACM Computing Surveys (CSUR)*, 33(1), pp. 31–88.
- NECHAEVSKY, MDR (2008). «Avances en medicina regenerativa y terapia génica en dermatología». *Piel*, 23(10), pp. 537–539.
- NEPH, S; VIERSTRA, J; STERGACHIS, AB; REYNOLDS, AP; HAUGEN, E; VERNOT, B; THURMAN, RE; JOHN, S; SANDSTROM, R; JOHNSON, AK; MAURANO, MT; HUMBERT, R; RYNES, E; WANG, H; VONG, S; LEE, K; BATES, D; DIEGEL, M; ROACH, V; DUNN, D; NERI, J; SCHAFER, A; HANSEN, RS; KUTYAVIN, T; GISTE, E; WEAVER, M; CANFIELD, T; SABO, P; ZHANG, M; BALASUNDARAM, G; BYRON, R; MACCOSS, MJ; AKEY, JM; BENDER, MA; GROUDINE, M; KAUL, R y STAMATOYANNOPOULOS, JA (2012). «An expansive human regulatory lexicon encoded in transcription factor footprints». *Nature*, 489(7414), pp. 83–90.
- NEUBERGER, J y COUNSELL, C (2002). «Impact factors: uses and abuses». *European Journal of Gastroenterology & Hepatology*, 14, pp. 209–211.
- NEWMAN, MEJ (2004). «Coauthorship networks and patterns of scientific collaboration.» *Proceedings of the National Academy of Sciences of the United States of America*, 101 Suppl(suppl_1), pp. 5200–5.
- NUDELMAN, A y LANDERS, C (1972). «The failure of 100 divided by 3 to equal 33-1/3». *The American Sociologist*, 7(9), pp. 9–11.
- OJASOO, T y DORÉ, JC (1999). «Citation bias in medical journals». *Scientometrics*, 45(1), pp. 81–94.

- OKSANEN, J; BLANCHET, FG; KINDT, R; LEGENDRE, P; O'HARA, RB; SIMPSON, GL; SOLYMOS, P; STEVENS, MHH y WAGNER, H (2010). «vegan: Community Ecology Package». Consultado el 26 de marzo de 2012.
<http://cran.r-project.org/package=vegan>
- OKUBO, Y (1997). *Bibliometric indicators and analysis of research systems : methods and examples*. OECD, Paris.
- ORPHANET (2010). «European collaborative research projects funded by DG Research and by E-Rare in the field of rare diseases and European clinical networks funded by DG Sanco and contributing to clinical research in the field of rare diseases». Consultado el 6 de marzo de 2012.
<http://www.orpha.net/orphacom/cahiers/docs/GB/Networks.pdf>
- (2011a). «Orphanet». Consultado el 10 de julio de 2012.
<http://www.orpha.net/consor/cgi-bin/index.php>
- (2011b). «Orphanet: Sobre los medicamentos huérfanos». Consultado el 10 de julio de 2012.
http://www.orpha.net/consor/cgi-bin/Education_AboutOrphanDrugs.php?lng=ES
- ORTIZ, AP; CALO, WA; SUÁREZ-BALSEIRO, C; MAURA-SARDO, M y SUÁREZ, E (2009). «Bibliometric assessment of cancer research in Puerto Rico, 1903-2005». *Revista Panamericana de Salud Pública*, 25, pp. 353–361.
- OWEN-SMITH, J y RICCABONI, M (2002). «A comparison of US and European university-industry relations in the life sciences». *Management Science*, 48(1), pp. 24–43.
- PAI, VB y NAHATA, MC (2001). «Efficacy and safety of aerosolized tobramycin in cystic fibrosis». *Pediatric Pulmonology*, 32(4), pp. 314–327.
- PALAU, F (2009). «Enfermedades raras, un paradigma emergente en la medicina del siglo XXI». *Medicina Clínica*, 134(4), pp. 161–168.

- PAO, ML (1982). «Lotka's test». *Collection management*, 4(1), pp. 111–124.
- PASQUALINI, CD (2001). «El genoma humano». *Medicina Buenos Aires*, 61, pp. 243–246.
- PEÑA, D (2001). *Fundamentos de estadística*. Alianza editorial, Madrid. ISBN 84-206-8696-4.
- (2002). *Análisis de Datos Multivariantes*. Mc Graw-Hill Interamericana de España, Madrid, España. ISBN 84-481-3610-1.
- PEÑA, D y ROMO, J (1999). *Introducción a la Estadística para las Ciencias Sociales*. McGraw Hill Interamericana de España, New York. ISBN 84-481-1617-8.
- PEÑA, EA y SLATE, EH (2006). «Global Validation of Linear Model Assumptions.» *Journal of the American Statistical Association*, 101(473), p. 341.
- PEARSON, P; FRANCOMANO, C; FOSTER, P; BOCCHINI, C; LI, P y MCKUSICK, V (1994). «The status of Online Mendelian Inheritance in Man (OMIM)». *Nucleic Acids Research*, 22, p. 3470.
- PEDERSEN, T; PATWARDHAN, S y MICHELIZZI, J (2008). «Text::Similarity - Measure the pair-wise Similarity of Files or Strings». Consultado el 24 de marzo de 2012.
http://search.cpan.org/~tpederse/Text-Similarity-0.06/bin/text_similarity.pl
- PELTONEN, L; PEKKARINEN, P y AALTONEN, J (1995). «Messages from an isolate: lessons from the Finnish gene pool.» *Biological Chemistry Hoppe-Seyler*, 376(12), pp. 697–704.
- PESTAÑA, A (1997). «El MedLine como fuente de informacion bibliométrica de la producción española en biomedicina y ciencias médicas. Comparación con el Science Citation Index.» *Medicina Clínica*, 109(13), pp. 506–511.

- PESTAÑA, A y CERDÁN, S (2000). «Spanish scientific productivity and equipment in magnetic resonance from a regional and European perspective». *Scientometrics*, 49(2), pp. 215–231.
- PFEIFER, U; POERSCH, T y FUHR, N (1996). «Retrieval effectiveness of proper name search methods». *Information Processing and Management*, 32(6), pp. 667–679.
- PORTER, AL; COHEN, AS; DAVID ROESSNER, J y PERREAULT, M (2007). «Measuring researcher interdisciplinarity». *Scientometrics*, 72(1), pp. 117–147.
- POSADA DE LA PAZ, M; IZQUIERDO MARTÍNEZ, M; FERRARI, MJ; AVELLANEDA FERNÁNDEZ, A; DE ANDRÉS COPEA, P y MARTÍN ARRIBAS, C (2002). «Plan de Acción de la Unión Europea y del Estado Español sobre enfermedades de Baja Prevalencia». En: *Symposium sobre Enfermedades de Baja Prevalencia en la Edad Pediátrica*, Santander.
- POSADA DE LA PAZ, M; MARTIN-ARRIBAS, C; RAMIREZ, A; VILLAVERDE, A y ABITUA, I (2008). «Enfermedades raras. Concepto, epidemiología y situación actual en España». *Anales del Sistema Sanitario de Navarra*, 31(Suplemento 2), pp. 9–20.
- PRAVDIC, N y OLUIC-VUKOVIC, V (1991). «Distribution of scientific productivity: Ambiguities in the assignment of author rank». *Scientometrics*, 20(1), pp. 131–144.
- PRICE, DJS (1973). *Hacia una ciencia de la ciencia*. Ariel, Barcelona. ISBN 84-344-0739-6.
- PRITCHARD, A (1969). «Statistical bibliography or bibliometrics». *Journal of documentation*, 25(4), pp. 348–349.
- RABEHARISOA, V y CALLON, M (2002). «The involvement of patient's associations in research». *International Social Science Journal*, 54(171), pp. 57–63.

- RAMALLE GOMARA, E; GONZÁLEZ MARTÍNEZ, M y PERUCHA GONZÁLEZ, M (2004). *Enfermedades Raras en La Rioja en el periodo 1999-2002: Análisis a partir del Conjunto Mínimo Básico de Datos*. Consejería de Salud, Logroño.
- RAMOS ACEITERO, J; SÁNCHEZ CANCHO, J; ÁLVAREZ DÍAS, M; GARCÍA BAZAGA, M; GARCÍA RAMOS, P y ZAMBRANOS CASIMIRO, M (2005). *Enfermedades raras en Extremadura*. Junta de Extremadura. Consejería de Sanidad y Consumo Dirección General de Consumo y Salud Comunitaria, Mérida.
- RANGNEKAR, D (2005). «Analysing the bibliometric presence of the Multiple Sclerosis Society». En: *ASLIB Proceedings New Information Perspectives*, volumen 57, pp. 247–260.
- RIOU-GOTTA, MO; FOURNIER, E; DANZON, A; PELLETIER, F; LEVANG, J; MERMET, I; BLANC, D; HUMBERT, P y AUBIN, F (2009). «Rare skin cancer: a population-based cancer registry descriptive study of 151 consecutive cases diagnosed between 1980 and 2004.» *Acta oncologica (Stockholm, Sweden)*, 48(4), pp. 605–609.
- RIP, A (1992). «Science and technology as dancing partners». En: P Kroes y M Bakker (Eds.), *Technological Development and Science in the Industrial Age*, pp. 231–270. Kluwer Academic. ISBN 978-0-7923-1898-9.
- ROBERTSON, AM y WILLETT, P (1998). «Applications of n-grams in textual information systems». *Journal of Documentation*, 54(1), pp. 48–69.
- RODRIGUES, PS; FONSECA, L y CHAIMOVICH, H (2000). «Mapping cancer, cardiovascular and malaria research in Brazil». *Brazilian Journal of Medical and Biological Research*, 33, pp. 853–867.
- RODRÍGUEZ JÚSTIZ, F; FERNÁNDEZ NÚÑEZ, A y BALY GIL, A (2000). «Apreciaciones sobre la clasificación internacional de enfermedades». *Revista cubana de higiene y epidemiología*, 38(3), pp. 215–219.
- ROMERO, AG; CORTÉS, JN; ESCUDERO, C; LÓPEZ, JAF y MORENO, JAC (2009). «Measuring the influence of clinical trials citations on several biblio-

- metric indicators». *Scientometrics*, 80(3), pp. 747–760.
- ROUSSEAU, R (2002). «Journal Evaluation: Technical and Practical Issues». *Library Trends*, 50(3), pp. 418–439.
- RUANO-RAVINA, A y PÉREZ-RÍOS, M (2012). «Regarding a case report: Rare diseases and bibliometric impact factor». *Journal of Clinical Epidemiology*, 65(8), pp. 915–917.
- RUIZ, M; CIRERA SUAREZ, L; PEREZ, G; BORRELL, C; AUDICA, C y MORENO, C (2002). «Comparabilidad entre la novena y la décima revisión de la Clasificación Internacional de Enfermedades aplicada a la codificación de la causa de muerte en España». *Gaceta Sanitaria*, 16(6), pp. 526–532.
- SADOSKY, C (1975). *Investigacion científica y dependencia*. Universidad Pedagógica y Tecnológica de Colombia Secretaria de Investigaciones y Extension Universitaria, Tunja. ISBN 978-950-29-1013-0.
- SÁNCHEZ, MA (1991). «La clasificación de las enfermedades: funciones actuales y fundamentos históricos». *Medicina Clínica*, 96, pp. 703–706.
- SÁNCHEZ CASTILLO, S (2012). «Representación social de las enfermedades raras en la prensa española». *Aposta. Revista de Ciencias Sociales*, (54), pp. 1–31.
- SANCHO LOZANO, R (1990). «Indicadores bibliométricos utilizados en la evaluación de la ciencia y la tecnología : revisión bibliográfica». *Revista Española de Documentación Científica*, 13(3/4).
- SANYAL, A; LAJOIE, BR; JAIN, G y DEKKER, J (2012). «The long-range interaction landscape of gene promoters». *Nature*, 489(7414), pp. 109–113.
- SANZ-CASADO, E (2012). «Lanzamiento del Observatorio IUNE, una herramienta para el seguimiento de la actividad científica de las universidades españolas». *Revista Española de Documentación Científica*, 35(3), pp. 503–505.
- SANZ CASADO, E; DE FILIPPO, D; GARCÍA-ZORITA, C y EFRAÍN GARCÍA, P (2011). «Observatorio IUNE: una nueva herramienta para el seguimiento de la

- actividad investigadora del sistema universitario español». *Bordón. Revista de Pedagogía*, 63(2), pp. 101–116.
- SANZ-CASADO, E y MARTÍN-MORENO, C (1997). «Técnicas bibliométricas aplicadas a los estudios de usuarios». *Revista General de Información y Documentación*, 7(2), pp. 41–68.
- SANZ-CASADO, E; SUAREZ-BALSEIRO, C; GARCÍA-ZORITA, C; MARTÍN-MORENO, C y LASCURAIN-SÁNCHEZ, M (2002). «Metric studies of information: An approach towards a practical teaching method». *Education for information*, 20(2), pp. 133–144.
- SANZ-CASADO, E; SUÁREZ-BALSEIRO, C; IRIBARREN-MAESTRO, I; PAU, MRDS y DE PEDRO-CUESTA, J (2007). «Bibliometric mapping of scientific research on prion diseases, 1973–2002». *Information Processing and Management*, 43(1), pp. 273–284.
- SCHUBERT, A y GLÄNZEL, W (1983). «Statistical reliability of comparisons based on the citation impact of scientific publications». *Scientometrics*, 5(1), pp. 59–73.
- (1991). «Publication dynamics: Models and indicators». *Scientometrics*, 20(1), pp. 317–331.
- SCIMAGO (2007). «SJR — SCImago Journal & Country Rank». Consultado el 8 de junio de 2012.
<http://www.scimagojr.com/>
- SERRANO-LÓPEZ, AE y MARTÍN-MORENO, C (2011). «Cadasil e Síndrome de Rett: Estudo de caso de dois doenças raras neurológicas.» *Ponto de Acesso*, 5(3), pp. 130–148.
- SHAW, W (1990a). «Subject indexing and citation indexing— part II: An evaluation and comparison». *Information Processing & Management*, 26(6), pp. 705–718.

- (1990b). «Subject indexing and citation indexing—part I: Clustering structure in the cystic fibrosis document collection». *Information Processing & Management*, 26(6), pp. 693–703.
- (1991a). «Subject and Citation Indexing. Part I: The clustering structure of composite representations in the cystic fibrosis document collection». *Journal of the American Society for Information Science*, 42(9), pp. 669–675.
- (1991b). «Subject and citation indexing. Part II: The optimal, clusterbased retrieval performance of composite representations». *Journal of the American Society for Information Science*, 42(9), pp. 676–684.
- SHAW, W y WOOD, J (1991). «The cystic fibrosis database: Content and research opportunities». *Library and Information Science Research*, 12(4), pp. 347–366.
- SHEPARD, R (1962a). «The analysis of proximities: Multidimensional scaling with an unknown distance function. I.» *Psychometrika*, 27(2), pp. 125–140.
- (1962b). «The analysis of proximities: Multidimensional scaling with an unknown distance function. II.» *Psychometrika*, 27(3), pp. 219–246.
- SHLOSSMAN, M; KNOWLER, WC; PETTITT, DJ y GENCO, RJ (1990). «Type 2 diabetes mellitus and periodontal disease.» *Journal of the American Dental Association (1939)*, 121(4), pp. 532–536.
- SMALL, H (1973). «Co-citation in the scientific literature: A new measure of the relationship between two documents». *Journal of the American Society for information Science*, 24(4), p. 265–269.
- SOMOGYI, A y SCHUBERT, A (2005). «Short communication: Correlation between national bibliometric and health indicators: The case of diabetes». *Scientometrics*, 62(2), pp. 285–292.
- SPINAK, E (1996). *Diccionario enciclopédico de bibliometría, ciencia métrica e informática*. UNESCO, Caracas. ISBN 92-9143-007-2.

- SPITZMUELLER, D; HODLER, J; SEIFERT, B y ZANETTI, M (2010). «Radiological research activity 1998-2007: relationship to gross domestic product, health expenditure and public expenditure on education.» *Insights Into Imaging*, 1(4), pp. 269–280.
- STINSON, ER y LANCASTER, F (1987). «Synchronous versus diachronous methods in the measurement of obsolescence by citation studies». *Journal of Information Science*, 13(2), pp. 65–74.
- STOLK, P (2006). «Rare essentials: drugs for rare diseases as essential medicines». *Bulletin of the World health Organization*, 84(9), pp. 745–751.
- THORNE, FC (1977). «The citation index: Another case of spurious validity». *Journal of Clinical Psychology*, 33(4), pp. 1157–1161.
- THURMAN, RE; RYNES, E; HUMBERT, R; VIERSTRA, J; MAURANO, MT; HAUGEN, E; SHEFFIELD, NC; STERGACHIS, AB; WANG, H; VERNOT, B; GARG, K; JOHN, S; SANDSTROM, R; BATES, D; BOATMAN, L; CANFIELD, TK; DIEGEL, M; DUNN, D; EBERSOL, AK; FRUM, T; GISTE, E; JOHNSON, AK; JOHNSON, EM; KUTYAVIN, T; LAJOIE, B; LEE, BK; LEE, K; LONDON, D; LOTAKIS, D; NEPH, S; NERI, F; NGUYEN, ED; QU, H; REYNOLDS, AP; ROACH, V; SAFI, A; SANCHEZ, ME; SANYAL, A; SHAFER, A; SIMON, JM; SONG, L; VONG, S; WEAVER, M; YAN, Y; ZHANG, Z; ZHANG, Z; LENHARD, B; TEWARI, M; DORSCHNER, MO; HANSEN, RS; NAVAS, PA; STAMATOYANNOPOULOS, G; IYER, VR; LIEB, JD; SUNYAEV, SR; AKEY, JM; SABO, PJ; KAUL, R; FUREY, TS; DEKKER, J; CRAWFORD, GE y STAMATOYANNOPOULOS, JA (2012). «The accessible chromatin landscape of the human genome». *Nature*, 489(7414), pp. 75–82.
- TORRES-SALINAS, D; RUIZ-PÉREZ, R y DELGADO-LÓPEZ-CÓZAR, E (2009). «Google Scholar como herramienta para la evaluación científica». *El Profesional de la Información*, 18(5), pp. 501–510.

- TORTOSA GIL, F (1989). «La psicología en España a través de algunas de sus revistas». *Papeles del Psicólogo: Revista del Colegio Oficial de Psicólogos*, (36,37), pp. 79–82.
- UNIÓN EUROPEA (1999). «Decisión n 1295/1999/CE del Parlamento Europeo y del Consejo, de 29 de abril de 1999, por la que se aprueba un programa de acción comunitaria sobre las enfermedades poco comunes en el marco de la acción en el ámbito de la salud pública (1999-2003)».
- (2000). «Reglamento (CE) nº 141/2000 del Parlamento Europeo y del Consejo, de 16 de diciembre de 1999, sobre los medicamentos huérfanos».
- (2006). «Commision Staff Working Document on the experience acquired as a result of the application of Regulation (EC) No 141/2000 on orphan medicinal products and account of the public health benefits obtained».
- (2007). «Decisión N° 1350/2007/CE del Parlamento Europeo y del Consejo de 23 de octubre de 2007 por la que se establece el segundo Programa de acción comunitaria en el ámbito de la salud (2008-2013)».
- U.S. NATIONAL LIBRARY OF MEDICINE (2012). «Publication Characteristics (Publication Types) - Scope Notes - 2012». Consultado el 24 de marzo de 2012. <http://www.nlm.nih.gov/mesh/pubtypes.html>
- VAZQUEZ VALERO, M y GONZALO BALMISA, Y (1997). «Utilización de las revistas españolas a través de su demanda en su servicio de suministro de documentos». *Boletín de la ANABAD*, 47(2), pp. 193–204.
- VENABLES, W y RIPLEY, B (2002). *Modern applied statistics with S*. Springer, New York. ISBN 0-387-95457-0.
- VILLAR, J (1988). «El inglés, idioma internacional en medicina». *Medicina Clínica*, 91, pp. 23–24.
- VINKLER, P (1988). «An attempt of surveying and classifying bibliometric indicators for scientometric purposes». *Scientometrics*, 13(5), pp. 239–259.

- WARD JR, J (1963). «Hierarchical grouping to optimize an objective function». *Journal of the American Statistical Association*, 58(301), pp. 236–244.
- WEINGART, P (2005). «Impact of bibliometrics upon the science system: Inadvertent consequences?» *Scientometrics*, 62(1), pp. 117–131.
- WEISBERG, S y FOX, J (2010). *An R companion to applied regression*. Sage Publications, California. ISBN 978-1-4129-7514-8.
- WESTERMAN, E; LE BRUN, P; TOUW, D; FRIJLINK, H y HEIJERMAN, H (2004). «Effect of nebulized colistin sulphate and colistin sulphomethate on lung function in patients with cystic fibrosis: a pilot study». *Journal of Cystic Fibrosis*, 3(1), pp. 23–28.
- WHITE, HD y MCCAIN, KW (1989). «Bibliometrics». *Annual Review of Information Science and Technology*, 24, pp. 119–186.
- WORLD HEALTH ORGANIZATION (2005). «Policy, Access and Rational Use of Drugs». *Informe técnico*, World Health Organization. Consultado el 2 de marzo de 2012.
<http://archives.who.int/prioritymeds/report/index.htm>
- (2009). «History of the development of the ICD». Consultado el 2 de julio de 2012.
<http://www.who.int/entity/classifications/icd/en/HistoryOfICD.pdf>
- WOUTERS, P (1999). *The citation culture*. Tesis doctoral, University of Amsterdam. Directores: S. S. Blume y R. P. Hagendijk.
- YABLONSKY, A (1980). «On fundamental regularities of the distribution of scientific productivity». *Scientometrics*, 2(1), pp. 3–34.
- ZEILEIS, A (2009). «dynlm: Dynamic Linear Regression». Consultado el 22 de marzo de 2012.
<http://cran.r-project.org/package=dynlm>

- ZEILEIS, A; MEYER, D y HORNIK, K (2007). «Residual-based shadings for visualizing (conditional) independence». *Journal of Computational and Graphical Statistics*, 16(3), pp. 507–525.
- ZHOU, P y LEYDESDORFF, L (2006). «The emergence of China as a leading nation in science». *Research Policy*, 35(1), pp. 83–104.
- ZITT, M (2000). «Shadows of the past in international cooperation: Collaboration profiles of the top five producers of science». *Scientometrics*, 47(3), pp. 627–657.

Apendices

Apéndice **A**

A.1. ER con causa genética identificada

Enfermedad	Código ICD-10
Hiperinsulinismo con hiperamonemia	E161
Kallman syndrome (KS)	E230
Adrenal hyperplasia, congenital, due to 21-hydroxylase deficien	E250
Ataxia por deficiencia de vitamina E	E560
Fenilcetonuria (PAH): Fenilalanina hidroxilasa, Deficiencia de: Fenilalanina monooxigenasa, Deficiencia de	E700
6-piruvoil-tetrahydropterina sintasa, Deficiencia de	E701
Dihidropterina reductasa (DHPR), Deficiencia de	E701
GTP ciclohrolasa I, Deficiencia de	E701
Alcaptonuria: Homogentisico dioxigenasa, Deficiencia de	E702
Hawkinsinuria	E702
Tirosina hidroxilasa, Deficiencia de	E702
Tirosinemia tipo I: Fumaril acetoacetasa, Deficiencia de	E702
Tirosinemia tipo II: Tirosina aminotransferasa hepática, Deficiencia de	E702
Tirosinemia tipo III: Hidroxifenilpiruvato dioxigenasa, Deficiencia de	E702

Enfermedad	Código ICD-10
Descarboxilasa de aminoácidos aromáticos, Deficiencia de	E708
Jarabe de Arce, Enfermedad del (MSUD)	E710
Acidemia isovalérica: IsovalerilCoA deshidrogenasa, Deficiencia de	E711
Acidemia metilmalónica: Cob(I)alamina Adenosiltransferasa, Deficiencia de (CblB)	E711
Acidemia metilmalónica: MetilmalonilCoA mutasa, Deficiencia de (MCM)	E711
Acidemia metilmalónica: MMAA, Deficiencia de (CblA)	E711
Acidemia propiónica: PropionilCoA carboxilasa, Deficiencia de (PCC)	E711
Aciduria 3-hidroxi-3-metilglutárica: 3-hidroxi-3-metilglutarilCoA liasa, Deficiencia de (HL)	E711
Aciduria metilglutacónica tipo I: 3-metilglutaconilCoA hidratasa, Deficiencia de	E711
Aciduria metilglutacónica tipo III: síndrome de Costeff	E711
Aciduria metilglutacónica tipo IV:	E711
3-OH acilCoA deshidrogenasa, Deficiencia de (LCHAD)	E713
Acil CoA deshidrogenasa de ácidos grasos de cadena larga, Deficiencia de (VLCAD)	E713
AcilCoA deshidrogenasa de ácidos grasos de cadena corta, Deficiencia de (SCAD)	E713
Acil-CoA oxidasa peroxisomal, Deficiencia de	E713
Adrenoleucodistrofia ligada al cromosoma X (ALD)	E713
Adrenoleucodistrofia neonatal	E713
Proteína D-bifuncional peroxisomal, Deficiencia de	E713
Proteína trifuncional mitochondrial, Deficiencia de	E713

Enfermedad	Código ICD-10
Cistinosis	E720
Cistinuria tipo I	E720
Cistinuria tipo no-I	E720
Acidemia metilmalónica con homocistinuria CblD	E721
Acidemia metilmalónica con homocistinuria CblF	E721
Acidemia metilmalónica con homocistinuria: CblC	E721
Cofactor molibdeno, Deficiencia del: sulfito oxidasa y xantina oxidasa, Deficiencia combinada de	E721
Homocistinuria por deficiencia de 5-10 metiléntetrahidrofolato reductasa (MTHFR)	E721
Homocistinuria por deficiencia de metionina sintasa: cblG	E721
Homocistinuria por deficiencia en cistationina beta sintasa	E721
Homocistinuria por deficiencia de metionina sintasa reductasa: CblE	E721
Sulfito oxidasa, Deficiencia de	E721
Aciduria Argininsuccínica: Arginino succínico liasa, Deficiencia de (ASL)	E722
Argininemia: Arginasa, Deficiencia de	E722
Carbamil fosfato sintetasa, Deficiencia de (CPS)	E722
Citrulinemia tipo II: Citrina, Deficiencia de	E722
Citrulinemia tipo II: Citrina, Deficiencia de	E722
Citrulinemia: Arginino succínico sintetasa, Deficiencia de (ASS)	E722
Ornitina oxo-ácido deshidrogenasa, deficiencia de. Gira-ta atrofia	E722
Ornitina transcarbamilasa, Deficiencia de (OTC)	E722

Enfermedad	Código ICD-10
a-cetoglutarato deshidrogenasa, Deficiencia de	E723
Aciduria glutárica tipo I (GA1)	E723
Aciduria L-2-hidroxi-glutárica	E723
AcilCoA deshidrogenasa de ácidos grasos, Deficiencia múltiple de (MADD): aciduria glutárica Tipo II	E723
Glutation sintetasa, Deficiencia de	E723
Hiperornitinemia-Hiperammonemia-homocitrulinuria (HHH)	E724
Hiperglicinemia no cetósica (HG NK)	E725
Aciduria malónica: MalonilCoA descarboxilasa, Deficiencia de	E728
Aciduria mevalónica: Mevalonato kinasa, Deficiencia de	E728
Arginina:glicina amidino transferasa, Deficiencia de (AGAT)	E728
b-cetotiolasa, Deficiencia de	E728
beta-ureidopropionasa, Deficiencia de	E728
Intolerancia protéica con lisinuria	E728
N-acetilglutamato sintetasa, Deficiencia de (NAGS)	E728
Transportador de creatina, Deficiencia del	E728
Transporte de carnitina, Deficiencia de	E728
Glucógeno fosforilasa quinasa de hígado y músculo, Deficiencia de	E740
Glucógeno fosforilasa quinasa de músculo, Deficiencia de	E740
Glucogenosis Tipo IV: Enzima ramificante, Deficiencia de	E740
Glucogenosis Tipo 0	E740
Glucogenosis Tipo Ia: Actividad hidrolítica de la glucosa-6-fosfatasa, Deficiencia de la	E740

Enfermedad	Código ICD-10
Glucogenosis Tipo Ib	E740
Glucogenosis Tipo II: a-glucosidasa ácida, Deficiencia de	E740
Glucogenosis Tipo III: Amilo 1,6-glucosidasa, Deficiencia de	E740
Glucogenosis Tipo V: Glucógeno fosforilasa, Deficiencia de	E740
Glucogenosis Tipo VI: Glucógeno fosforilasa en hígado, Deficiencia de	E740
Glucogenosis Tipo VII: 6-fosfofructo 1-quinasa de músculo, Deficiencia de	E740
Glucogenosis Tipo VIII: Glucógeno fosforilasa quinasa en hígado, Deficiencia de	E740
Fructoquinasa, Deficiencia de	E741
Fructosa 1,6-difosfato aldolasa B: intolerancia hereditaria a la fructosa, Deficiencia de	E741
Galactosemia por deficiencia de epimerasa	E742
Galactosemia por deficiencia de galactoquinasa	E742
Galactosemia por deficiencia de galactosa 1-P uridiltransferasa	E742
Glucogenosis Tipo II: a-glucosidasa ácida, Deficiencia de	E744
Hiperoxaluria tipo I	E748
Piruvato carboxilasa, Deficiencia de (PC)	E748
Piruvato deshidrogenasa, Deficiencia del complejo E1a (PDH)	E748
Sandhoff, Enfermedad de: Gangliosidosis GM2 Tipo 0, (HEXB)	E750
Tay- Sachs, Enfermedad de y variante B1: Gangliosidosis GM2 Tipo I (HEXA)	E750

Enfermedad	Código ICD-10
Gangliosidosis GM1	E751
Aciduria N-acetilaspártica: Canavan, Enfermedad de	E752
Deficiencia múltiple en sulfatasas: Austin, Enfermedad de. Mucopolipidosis	E752
Enfermedad de Gaucher Tipo II (GD2)	E752
Fabry, Enfermedad de	E752
Gaucher disease (GBA) - Type I	E752
Gaucher disease, type III (GD3)	E752
Krabbe, Enfermedad de	E752
Leucodistrofia metacromática (MLD)	E752
Niemann-Pick disease, type C	E752
Niemann-Pick, Enfermedad de Tipo A y Tipo B	E752
Defectos en la O-Glicosilación: Muscle-eye-brain disease (MEB)	E754
Lipofuscinosis neuronal ceroide (CLN 1): Palmitoil proteína tioesterasa, Deficiencia de	E754
Lipofuscinosis neuronal ceroide (CLN 2): Tripeptidil peptidasa I, Deficiencia de	E754
Wolman, Enfermedad de y Enf. por acúmulo de ésteres de colesterol (CESD)	E755
Xantomatosis cerebrotendinosa	E755
Mucopolisacaridosis (MPS) IH/S: Hurler-Scheie	E760
Mucopolisacaridosis (MPS) IS: Scheie	E760
Mucopolisacaridosis(MPS) II: Hunter (MPS2)	E761
Mucopolisacaridosis (MPS) IIIA: SanfilippoA	E762
Mucopolisacaridosis (MPS) IIIB: SanfilippoB	E762
Mucopolisacaridosis (MPS) IIIC: SanfilippoC	E762
Mucopolisacaridosis (MPS) IIID: SanfilippoD	E762

Enfermedad	Código ICD-10
Lecitina:colesterol aciltransferasa, Deficiencia de	E786
Tangier, Enfermedad de	E786
Retención de quilomicrones	E788
Sitosterolemia	E788
3-hidroxi-3-metilglutarilCoA sintasa, Deficiencia de	E789
Latosterolosis	E789
Hipoxantina –guanina fosforribosiltransferasa, Deficiencia de	E791
Adenilosuccinato liasa, Deficiencia de	E798
Adenina fosforribosiltransferasa, Deficiencia de	E798
Dihidropirimidina deshidrogenasa, Deficiencia de	E798
Dihidropirimidinasa, Deficiencia de	E798
Fosforribosilpirofosfato sintasa, Superactividad de	E798
Mioadelinato desaminasa, Deficiencia de	E798
Timidina fosforilasa, Deficiencia de (MNGIE)	E798
Xantina oxidasa, Deficiencia de	E798
Menkes, Enfermedad de	E830
Wilson, Enfermedad de (WND, WD)	E830
Hemochromatosis (HFE)	E831
Cystic fibrosis (CF)	E84
Familial Mediterranean Fever (FMF)	E850
Amyloidosis I or Andrade type	E859
Dihidrolipoildeshidrogenasa, Deficiencia de	E872
Alfa-1-antitripsina deficiency	E880
Acil CoA deshidrogenasa de ácidos grasos de cadena media, Deficiencia de (MCAD)	E888
Deficiencia en biotinidasa	E888
Holocarboxilasa sintetasa, Deficiencia de (HCS)	E888

Enfermedad	Código ICD-10
Pterina carbinolamina deshidratasa, Deficiencia de	F011
Rett Syndrome (RTT)	F842
Huntington chorea (HD)	G10
Friedreich ataxia 1 (FRDA1)	G111
Machado-Joseph Disease (MJD)	G118
Spinocerebellar Ataxia 1 (SCA1)	G118
Spinocerebellar Ataxia 2 (SCA2)	G118
Spinocerebellar ataxia 6 (SCA6)	G118
Spinocerebellar ataxia 7 (SCA7)	G118
Spinocerebellar atrophy, type 8 (SCA8)	G118
Dentatorubral-pallidolulsian atrophy (DRPLA)	G119
Spinal muscular atrophy type 1 (SMA1)	G120
Spinal muscular atrophy type 3 (SMA3)	G121
Spinal muscular atrophy, adult (SMA4)	G121
Torsion Dystonia 1 Gene (DYT1)	G241
Dystonia, progressive, with diurnal variation	G248
Sepiapterina reductasa, Deficiencia de	G248
Almacenamiento del ácido siálico, Enfermedad por formas clínicas: infantil y enfermedad de Salla	G318
Encefalopatías mitocondriales: MERFF	G403
Charcot-Tooth disease type 1A (CMT1A)	G600
Refsum clasica, Enfermedad de	G601
Refsum infantil, Enfermead de	G601
Encefalopatías mitocondriales: NARP	G608
Neuropathy, hereditary with liability to pressure palsies (HNPP)	G608
Defectos en la O-Glicosilación: Fukuyama distrofia muscular congénita	G710

Enfermedad	Código ICD-10
Duchenne muscular dystrophy (DMD)	G710
Dystrophia myotonica 1 (DM)	G711
Complejo I de la cadena respiratoria, Deficiencia de	G713
Complejo II de la cadena respiratoria, Deficiencia de	G713
Complejo III de la cadena respiratoria, Deficiencia de	G713
Complejo IV de la cadena respiratoria, Deficiencia de	G713
Complejo V de la cadena respiratoria, Deficiencia de	G713
Depleción de DNA mitocondrial: Deoxiguanosinquinasa, Deficiencia de (DGUOK)	G713
Encefalopatías mitocondriales: LHON	G713
Encefalopatías mitocondriales: MELAS	G713
Encefalopatías mitocondriales: síndrome de Pearson	G713
Fumarasa, Deficiencia de	G713
Choroideremia (CHM)	H312
Retinoschisis, X-linked juvenile (RS1)	H331
Retinal degeneration, slow (RDS)	H355
Retinitis pigmentosa 2 (RP2)	H355
Retinitis pigmentosa 3 (RP3)	H355
Retinitis pigmentosa 6 (RP6)	H355
Rhodopsin (RHO)	H355
Usher syndrome, type IIA (USH2A)	H355
Encefalopatías mitocondriales: KSS	H498
Deafness, sensorineural, autosomal-mitochondrial type	H905
Familial pheochromocytoma	M870
Miller-Dieker syndrome (MDS)	Q043
Defectos en la O-Glicosilación: Sínd. De Walker-Warburg	Q078
Norrie disease (NPD)	Q138

Enfermedad	Código ICD-10
Hirschprung disease (HSCR)	Q431
Polycystic Kidney Disease 2 (PKD2)	Q612
Polycystic kidney Disease I (PKD1)	Q612
Thanatophoric dysplasia (TD)	Q771
Condrodisplasia rizomélica punctata Tipo I (Pex7P)	Q773
Condrodisplasia rizomélica punctata TipoII: DHAP acil-CoA transferasa, Deficiencia de	Q773
Condrodisplasia rizomélica punctata TipoIII: alkil-DHAP sintasa, Deficiencia de	Q773
Síndrome de Conradi-Hünemann	Q773
Achondroplasia (ACH)	Q774
Hypochondroplasia (HCH)	Q774
Picnodisostosis: Catepsina K, Deficiencia de	Q788
Ictiosis ligada al X: Arilsulfatasa C, Deficiencia de. Esteroidesulfatasa, Deficiencia de	Q801
Neurofibromatosis, type 1 (NF1)	Q850
Neurofibromatosis, type 2 (NF2)	Q850
Von Hippel-Lindau syndrome (VHL)	Q858
Angelman syndrome (AS)	Q870
Langer mesomelic dysplasia	Q871
Prader-Willi syndrome (PWS)	Q871
Silver-Russel Syndrome (RSS)	Q871
Síndrome de Smith-Lemli-Opitz	Q871
Síndrome de Alport ligada al X	Q878
Síndrome de CHILD	Q878
Síndrome Zellweger	Q878
Smith-Magenis syndrome (SMS)	Q878
Williams-Beuren syndrome (WBS)	Q878

Enfermedad	Código ICD-10
Wolfram, Síndrome de	Q878
Desmosterolosis	Q897
Wolf-Hirshhorn syndrome (WHS)	Q933
Cri-du-chat syndrome	Q934
Azoospermia factor 1 (AZF1)	Q988
Deleted in azoospermia (DAZ)	Q988
Fragile site mental retardation 1 (FMR1)	Q992
Fragile site, folic acid type, rare, Fra(X)(q28) (FRAXE)	Q992
Sex determining region Y (SRY)	Q998

Tabla A.1: *Enfermedades raras con causa genética identificada (INERGEN)*

A.2. Agrupación de enfermedades por clasificación ICD-10

Clasificación ICD-10	Agrupación de enfermedades
E16-E56	Enfermedades endocrinas y nutricionales
E70	Desórdenes del metabolismo de aminoácidos aromáticos
E71	Desórdenes del metabolismo de la cadena de aminoácidos y del metabolismo de ácidos grasos
E72	Otros desórdenes del metabolismo de aminoácidos
E74	Otros desórdenes del metabolismo de carbohidratos
E75	Desórdenes del metabolismo de esfingolípidos y otros desórdenes de almacenamiento lipídico
E76	Desórdenes del metabolismo de glucosaminoglucanos
E77	Desórdenes del metabolismo de glicoproteínas
E78	Desórdenes del metabolismo de lipoproteínas y otras lipidemias
E79	Desórdenes del metabolismo de la purina y la pirimidina

Clasificación ICD-10	Agrupación de enfermedades
E83-E88	Desórdenes del metabolismo de minerales; Fibrosis quística; Amiloidosis; Agotamiento volumétrico; Otros desórdenes de fluidos, electrolitos y del equilibrio ácido-base; Otros desórdenes metabólicos
F01-F84	Trastornos mentales y del comportamiento
G10-G12	Atrofas sistémicas con afección primaria del sistema nervioso central
G24-G40	Patología extrapiramidal y del movimiento; Otras enfermedades degenerativas del sistema nervioso; Enfermedades desmielinizantes del sistema nervioso central; Patología episódica o paroxística
G60	Polineuropatías y otras patologías del sistema nervioso periférico
G71	Trastornos primarios de los músculos
H31-H90	Enfermedades del ojo y sus anexos; Enfermedades del oído y de la apófisis mastoides
M87	Osteonecrosis
Q04-Q61	Enfermedades congénitas del sistema nervioso; Enfermedades congénitas de los ojos, cara y cuello; Enfermedades congénitas del sistema circulatorio; Enfermedades congénitas del aparato respiratorio; Malformaciones de labios, boca y paladar; Otras malformaciones del tracto digestivo; Malformaciones de los órganos genitales; Malformaciones de los órganos urinarios
Q77-Q78	Osteocondrodisplasia con defecto del crecimiento de los huesos largos y de la columna vertebral y otras osteocondrodisplasias
Q80-Q89	Otras dismorfias congénitas

Clasificación ICD-10	Agrupación de enfermedades
Q93-Q99	Anormalidades cromosómicas no clasificadas en otra parte

Tabla A.2: *Agrupación de enfermedades por código ICD-10*

A.3. Estrategias de búsqueda por clasificación ICD-10

E16-E56 ((“hyperinsulinism hyperammonemia” OR (“hyperinsulinemic hypoglycemia” AND familial AND 6)) OR (“kallman syndrome”) OR ((“adrenal hyperplasia” congenital deficiency) OR (“adrenal hyperplasia iii”) OR (“cyp21 deficiency”) OR (“congenital adrenal hyperplasia 1”)) OR ((ataxia “vitamin e deficiency”) OR (“friedreich-like ataxia”)))

E70 (((“6 pyruvoyl tetrahydropterin synthase” deficiency) OR (hyperphenylalaninemia “tetrahydrobiopterin-deficient”) OR (“pts deficiency”)) OR ((“dihydropteridine reductase” deficiency) OR (“dhpr deficiency”) OR (“quinoid dihydropteridine reductase deficiency”) OR (“qdpr deficiency”)) OR (“gtp cyclohydrolase i” deficiency) OR (alkaptonuria OR “homogentisic acid oxidase deficiency”) OR (hawkinsinuria) OR ((“tyrosine hydroxylase” deficiency) OR (parkinsonism infantile recessive autosomal) OR (dystonia dopa-responsive autosomal recessive)) OR ((tyrosinemia AND (“type i” OR “type ii” OR “type iii”)) OR (“hepatorenal tyrosinemia”) OR (“fumarylacetoacetase deficiency”) OR (“fah deficiency”)) OR ((“dopa decarboxylase” OR “aromatic l-amino acid decarboxylase”) deficiency) OR (“aadc deficiency” OR “dcc deficiency”))

E71 (((“maple syrup urine disease”) OR (“branched-chain ketoaciduria”) OR (“branched-chain alpha-keto acid dehydrogenase deficiency”) OR (“bckd deficiency”) OR (“keto acid decarboxylase deficiency”)) OR (“isovaleric acidemia”) OR (“isovaleric acid coa dehydrogenase deficiency”) OR (“methylmalonic aciduria”) OR ((“propionic acidemia”) OR (“propionyl-coa car-

boxylase deficiency”) OR (“pcc deficiency”) OR (“glycinemia ketotic”) OR (“hyperglycinemia with ketoacidosis AND leukopenia”) OR (“ketotic hyperglycinemia”) OR (“methylglutaconic aciduria” AND (“type i” OR “type iii” OR “type iv”)) OR (“costeff syndrome”) OR (“optic atrophy plus syndrome”) OR (“iraqi-jewish optic atrophy plus”) OR (“optic atrophy 3” autosomal recessive)) OR (“acyl-coa dehydrogenase” AND (“long-chain” OR “short chain”) deficiency) OR (“peroxisomal acyl-coa oxidase deficiency”) OR (“straight-chain acyl-coa oxidase deficiency”) OR (“pseudoneonatal adrenoleukodystrophy”) OR (adrenoleukodystrophy x-linked) OR (adrenoleukodystrophy neonatal) OR (“d-bifunctional protein” deficiency) OR (“dbp deficiency”) OR (“pbfe deficiency”) OR (“peroxisomal bifunctional enzyme deficiency”))

E72 (“cystinosis”) OR (“cystinuria”) OR (homocystinuria AND (cbld OR cblf OR cblc)) OR (“molybdenum cofactor deficiency”) OR (homocystinuria (“methylenetetrahydrofolate reductase” OR methylcobalamin OR “cystathionine beta” OR “methionine synthase reductase”) deficiency) OR (sulfocysteinuria OR (“sulfite oxidase deficiency”) OR (sulfocysteinuria)) OR (“argininosuccinic aciduria”) OR (“argininosuccinate lyase deficiency”) OR (“asl deficiency”) OR (“argininosuccinic acid lyase deficiency”) OR (argininemia) OR (“arginase deficiency”) OR (hyperargininemia) OR (“arg1 deficiency”) OR (“carbamoyl phosphate synthetase” deficiency) OR (citrullinemia “type ii”) OR (citrullinemia) OR (“ornithine aminotransferase” deficiency) OR (“gyrate atrophy of choroid AND retina”) OR (“gyrate atrophy”) OR (“ornithine keto acid aminotransferase deficiency”) OR (“okt deficiency”) OR (“ornithine-delta-aminotransferase deficiency”) OR (“ornithine transcarbamylase” deficiency) OR (“ornithine carbamoyltransferase deficiency”) OR (“otc deficiency”) OR (“alpha-ketoglutarate dehydrogenase” deficiency) OR (“oxoglutaric aciduria”) OR (“glutaric acidemia type i” OR “glutaric acidemia type 1” OR “glutaric aciduria i”)) OR (“l-2-hydroxyglutaric aciduria” OR “l-2-hydroxyglutaric acidemia”))

OR (“multiple acyl-coa dehydrogenase” deficiency) OR (“glutaric aciduria ii”) OR (“glutaric aciduria type ii”) OR (“ethyl-malonic-adipicaciduria”) OR (“glutaric aciduria type 2”) OR (“glutaric aciduria ii”)) OR (“glutathione synthetase” deficiency) OR (“pyroglutamic aciduria”) OR (((hhh) OR (“hyperornithinemia-hyper -ammonemia-homocitrullinuria”)) syndrome) OR (“ornithine translocase deficiency”) OR (“glycine decarboxylase”) OR (“glycine cleavage system p protein”) OR (“glycine dehydrogenase”) OR (“malonyl-coa decarboxylase” deficiency) OR (“mevalonic aciduria”) OR (“mevalonate kinase deficiency”) OR (“l-arginine” “glycine amidinotransferase” deficiency) OR (“alpha-methylacetoacetic aciduria”) OR (“beta-ketothiolase” deficiency) OR (“mitochondrial acetoacetyl-coa thiolase deficiency”)) OR (“beta-ureidopropionase” deficiency) OR (“lysinuric protein intolerance”) OR (“dibasic amino aciduria ii”) OR (“n-acetylglutamate synthase” deficiency) OR (“hyperammonemia due to n-acetylglutamate synthetase deficiency”) OR (“n-acetylglutamate synthetase deficiency”) OR (“nags deficiency”) OR (“creatine deficiency syndrome”) OR (“carnitine deficiency” “systemic primary”) OR (“systemic carnitine deficiency”) OR (“carnitine deficiency”) OR (“carnitine uptake defect”))

E74 ((“glycogen storage disease” ix) OR (“phosphorylase kinase” muscle deficiency) OR (“glycogen storage disease” (“type iv” OR “type 4”)) OR (“glycogen storage disease 0”) OR (“glycogen storage disease” (“type i” OR “type 1”)) OR (“glycogen storage disease” (“type ib” OR “type 1b”)) OR (“glycogen storage disease” (“type ii” OR “type 2”)) OR (“glycogen storage disease” (“type iii” OR “type 3”)) OR (“glycogen storage disease” (“type v” OR “type 5”)) OR (“glycogen storage disease” (“type vi” OR “type 6”)) OR (“glycogen storage disease” (“type vii” OR “type 7”)) OR (“glycogen storage disease” (“type viii” OR “type 8”)) OR (“fructose-1” “6-bisphosphatase” deficiency) OR (“fructose intolerance” hereditary) OR (“galactose epimerase” deficiency) OR (galactokinase deficiency) OR (galactosemia) OR (“galt deficiency”) OR (galactosemia deficiency) OR (“acid

alpha-glucosidase”) OR (hyperoxaluria primary (“type i” OR “type 1”))
OR (“pyruvate carboxylase” deficiency) OR (“pyruvate decarboxylase” deficiency) OR (ataxia intermittent “abnormal pyruvate metabolism”) OR (“pyruvate dehydrogenase deficiency”) OR (“pdh deficiency”) OR (“ataxia with lactic acidosis i”))

E75 ((“sandhoff disease”) OR (“gm2-gangliosidosis type ii”) OR (“hexosaminidases a AND b deficiency”) OR (“tay-sachs disease”) OR (“hexosaminidase a deficiency”) OR (“hexa deficiency”) OR (“gm1-gangliosidosis”) OR (“beta-galactosidase-1 deficiency”) OR (“canavan disease”) OR (“canavan-van bogaert-bertrand disease”) OR (“spongy degeneration of central nervous system”) OR (“aspartoacylase deficiency”) OR (“aspa deficiency”) OR (“asp deficiency”) OR (“aminoacylase 2 deficiency”) OR (“acy2 deficiency”) OR (“multiple sulfatase deficiency”) OR (mucosulfatidosis) OR (“sulfatidosis juvenile austin type”) OR (“gaucher disease” (“type ii” OR “type 2”)) OR (“fabry disease”) OR (“angiokeratoma corporis diffusum”) OR (“anderson-fabry disease”) OR (“hereditary dystopic lipidosis”) OR (“alpha-galactosidase a deficiency”) OR (“ceramide trihexosidase deficiency”) OR (“gla deficiency”)) OR ((“gaucher disease” (“type i” OR “type 1”)) OR (“glucocerebrosidase deficiency”) OR (“acid beta-glucosidase deficiency”) OR (“gba deficiency”) OR (“gaucher disease” (“type iii” OR “type 3”)) OR (“krabbe disease”) OR (“globoid cell leukodystrophy”) OR (“globoid cell leukoencephalopathy”) OR (“galactosylceramide beta-galactosidase deficiency”) OR (“galactocerebrosidase deficiency”) OR (“galc deficiency”) OR (“metachromatic leukodystrophy”) OR (“metachromatic leukoencephalopathy”) OR (“cerebral sclerosis” diffuse “metachromatic form”) OR (“sulfatide lipidosis”) OR (“arylsulfatase a deficiency”) OR (“arsa deficiency”) OR (“cerebroside sulfatase deficiency”) OR (“niemann-pick disease” (“type a” OR “type b” OR “type c”)) OR (“sphingomyelin lipidosis”) OR (“sphingomyelinase deficiency”) OR (“neurovisceral storage disease with vertical supranuclear ophthalmoplegia”) OR (“muscle-eye-brain disease”) OR (“ce-

roid lipofuscinosis” neuronal (1 OR 2)) OR (“ceroid lipofuscinosis” neuronal (1 OR 2)) OR (“jansky-bielschowsky disease”) OR (“wolman disease”) OR (“cerebrotendinous xanthomatosis”) OR (“cerebral cholesterinosis”))

E76 (mucopolysaccharidosis (“type ih” OR “type is” OR “type ii” OR “type 2” OR “type iiiia” OR “type 3a” OR “type iiib” OR “type 3b” OR “type iiic” OR “type 3c” OR “type iiid” OR “type 3d” OR “type iva” OR “type 4a” OR “type ivb” OR “type 4b” OR “type vi” OR “type 6”) or (“hunter syndrome”) OR (“iduronate 2-sulfatase deficiency”) OR (“ids deficiency”) OR (“sids deficiency”) OR (“sulfoiduronate sulfatase deficiency”) OR (“hurler syndrome”) OR (“scheie syndrome”) OR (“sanfilippo syndrome a”) OR (“heparan sulfate sulfatase deficiency”) OR (“sulfaamidase deficiency”) OR (“sanfilippo syndrome b”) OR (“n-acetyl-alpha-d-glucosaminidase deficiency”) OR (“naglu deficiency”) OR (“sanfilippo syndrome c”) OR (“acetyl-coa:alpha-glucosaminide n-acetyltransferase deficiency”) OR (“sanfilippo syndrome d”) OR (“n-acetylglucosamine-6-sulfatase deficiency”) OR (“maroteaux-lamy syndrome”) OR (“arylsulfatase b deficiency”) OR (“arsb deficiency”) OR (“n-acetylgalactosamine-4-sulfatase deficiency”))

E77 ((mucopolipidosis (“type ii” OR “type 2” OR “type iii” OR “type 3”)) OR (“i-cell disease”) OR (“mucopolipidosis iii”) OR (“pseudo-hurler polydystrophy”) OR (“alpha mannosidosis”) OR (“lysosomal alpha-d-mannosidase deficiency”) OR (“alpha-mannosidase b deficiency”) OR (aspartylglucosaminuria) OR (“glycosylasparaginase deficiency”) OR (“aspartylglucosaminidase deficiency”) OR (“aga deficiency”) OR (“glycoasparaginase”) OR (“aspartylglycosaminuria”) OR (“beta mannosidosis”) OR (“lysosomal beta-mannosidase deficiency”) OR (“beta-mannosidase deficiency”)) OR ((fucosidosis) OR (“alpha-l-fucosidase deficiency”) OR (“neuraminidase deficiency”) OR “sialidosis type ii” OR “mucopolipidosis i” OR “lipomucopolysaccharidosis” OR “sialidase deficiency” OR “glycoprotein neuraminidase deficiency” OR “neug deficiency” OR “neuraminidase 1 deficiency” OR “neu

deficiency” OR “neu1 deficiency”) OR (“n-acetyl-alpha-d-galactosaminidase” OR naga) OR (“congenital disorder of glycosylation” AND (“type ia” OR “type 1a” OR “type ib” OR “type 1b” OR “type ic” OR “type 1c” OR “type id” OR “type 1d” OR “type ie” OR “type 1e” OR “type ig” OR “type 1g” OR “type ih” OR “type 1h” OR “type ii” OR “type 1i” OR “type ii a” OR “type 2 a” OR “type ii d” OR “type 2 d” OR “type ij” OR “type 1j” OR “type x” OR “type 10”)) OR (“jaeken syndrome”) OR (“phosphomannomutase 2 deficiency”) OR (“mannosephosphate isomerase deficiency”) OR (“mpi deficiency”) OR (“protein-losing enteropathy-hepatic fibrosis syndrome”) OR (“saguenay-lac saint-jean syndrome”) OR (“slsj syndrome”) OR (galactosialidosis))

E78 ((abetalipoproteinemia OR acanthocytosis OR “bassen-kornzweig syndrome” OR “apolipoprotein b deficiency” OR “microsomal triglyceride transfer protein deficiency” OR “mtp deficiency”) OR (hypobetalipoproteinemia) OR (“lecithin:cholesterol acyltransferase” deficiency) OR (“lcat deficiency” OR “norum disease”) OR (“tanger disease”) OR (“high density lipoprotein deficiency” “type 1”) OR (“high density lipoprotein deficiency” “tanger type”) OR (analphalipoproteinemia) OR (“chylomicron retention”) OR (sitosterolemia OR phytosterolemia) OR (“3-hydroxy-3-methylglutaryl coa synthase” deficiency) OR (“mitochondrial hmg-coa synthase deficiency” OR “hmgcs2 deficiency”) OR (lathosterolosis OR “sterol c5-desaturase deficiency” OR “sc5d deficiency”))

E79 ((“hypoxanthine guanine phosphoribosyltransferase” deficiency) OR (“lesch-nyhan syndrome”) OR (adenylosuccinase deficiency) OR (“adenylosuccinate lyase deficiency”) OR (“adenine phosphoribosyltransferase” deficiency) OR (“dihydropyrimidine dehydrogenase” deficiency) OR (“dpd deficiency” OR “dpyd deficiency” OR “thymine-uraciluria” OR “pyrimidinemia”) OR (dihydropyrimidinase deficiency) OR (“dihydropyrimidinuria”) OR (“dyps deficiency” OR “dph deficiency”) OR (“phosphoribosylpyrophosphate synthetase” superactivity) OR (“prps1 superactivity”) OR (“adenosine monop-

hosphate deaminase” deficiency) OR (“amp deaminase”) OR (“mitochondrial neurogastrointestinal encephalopathy” syndrome) OR (“myoneurogastrointestinal encephalopathy syndrome” OR “polip syndrome” OR “polyneuropathy ophthalmoplegia, leukoencephalopathy AND intestinal pseudo-obstruction”) OR (“thymidine phosphorylase” deficiency) OR (xanthinuria OR “xdh deficiency” OR “xanthine oxidase deficiency” OR (“xanthine dehydrogenase” deficiency)))

E83-E88 ((“menkes disease” OR “menkes syndrome” OR “kinky hair disease” OR “steely hair disease” OR “cooper transport disease”) OR (“wilson disease” OR “hepatolenticular degeneration”) OR (hemochromatosis) OR (“cystic fibrosis” OR “mucoviscidosis”) OR (“familial mediterranean fever”) OR (polyserositis (recurrent OR familial)) OR (amyloidosis (“type i” OR “type 1”)) OR (“dihydrolipoamide dehydrogenase” deficiency) OR (“alpha-1-antitrypsin” deficiency) OR (“acyl-coa dehydrogenase” “medium chain” deficiency) OR (“acadm deficiency” OR “mcad deficiency” OR “mcahd deficiency” OR “carnitine deficiency secondary to medium-chain acyl-coa dehydrogenase deficiency”) OR (biotinidase deficiency) OR (“multiple carboxylase deficiency late-onset” OR “multiple carboxylase deficiency, juvenile-onset” OR “btd deficeincy”) OR (“holocarboxylase synthetase” deficiency) OR (“hlcs deficiency” OR (“multiple carboxylase deficiency” neonatal) OR (“multiple carboxylase deficiency” “early onset”)))

F01-F84 ((cadasil) OR (“cerebral arteriopathy autosomal dominant with subcortical infarcts AND leukoencephalopathy”) OR (“dementia, hereditary multi-infarct type”) OR (casil) OR (“rett syndrome”) OR (“autism dementia ataxia AND loss of purposeful hand use”))

G10-G12 ((“huntington disease” OR “huntington chorea”) OR (“friedreich ataxia” (“type i” OR “type 1”)) OR (“machado-joseph disease”) OR (“spinocerebellar ataxia 3” OR “spinocerebelar atrophy 3” OR “azorean neurlogic disease” OR “spinotype atrphy” OR “nigrospinodentatal degeneration”))

("spinocerebellar ataxia" ("type i" OR "type 1" OR "type ii" OR "type 2" OR "type vi" OR "type 6" OR "type vii" OR "type 7")) OR ((("spinocerebellar ataxia" OR "spinocerebellar atrophy") ("type 8" OR "type viii")) OR ("dentatorubral-pallidoluysian atrophy") OR ("spinal muscular atrophy" ("type i" OR "type 1" OR "type iii" OR "type 3" OR "type iv" OR "type 4")) OR ("spinocerebellar atrophy i" OR "olivopontocerebellar atrophy i" OR "cerebelloparenchymal disorder i" OR "menzel type opca" OR "olivopontocerebellar atrophy iv" OR "schut-haymaker type opca")) OR ((("spinocerebellar atrophy ii" OR "olivopontocerebellar atrophy holguin type" OR "olivopontocerebellar atrophy ii" OR "spinocerebellar ataxia cuban type" OR "cerebellar degeneration with slow eye movements" OR "wadia-swami syndrome" OR "spinocerebellar degeneration with slow eye movements") OR ("norwegian infantile onset ataxia") OR ("olivopontocerebellar atrophy iii" OR "opca with retinal degeneration" OR "opca with macular degeneration AND external ophthalmoplegia" OR "autosomal dominant cerebellar ataxia type ii") OR ("werdnig-hoffmann disease") OR ("kugelberg-welander syndrome") OR ("adult spinal muscular atrophy"))

G24-G40 ((("torsin-a" OR torsina OR "tor1a") OR ((dystonia "dopa-responsive" OR (dystonia progressive "with diurnal variation")) OR ("dystonia-parkinsonism with diurnal fluctuation" OR "segawa syndrome, autosomal dominant" OR "dopa-responsive dystonia") OR ("sepiapterin reductase" deficiency) OR ("spr deficiency") OR ("infantile sialic acid storage" OR "n-acetylneuraminic acid storage disease" OR "nana storage disease" OR (sialuria infantil)) OR ("merrf syndrome" OR ("myoclonic epilepsy associated with ragged-red fibers"))))

G60 ((("charcot-marie-tooth disease" ("type 1 a" OR "type i a" OR "type 1a" OR "typeia" OR cmt1a)) OR ("hereditary motor AND sensory neuropathy ia") OR ("charcot-marie-tooth neuropathy" ("type 1 a" OR "type i a" OR "type 1a" OR "typeia")) OR ("refsum disease" adult*) OR ("phytanic acid oxidase deficiency") OR ("heredopathia atactica polyneuritiformis") OR

(“hereditary motor AND sensory neuropathy iv”) OR (“refsum disease” infantil*) OR (“infantile phytanic acid storage disease”) OR (“narp syndrome”) OR (“neuropathy, ataxia, AND retinitis pigmentosa”) OR ((neuropathy hereditary “with liability to pressure palsies”) OR (polyneuropathy familial recurrent) OR (tomaculous neuropathy)))

G71 ((fukuyama “congenital muscular dystrophy”) OR (“muscular dystrophy” “congenital progressive” “with mental retardation”) OR (“muscular dystrophy” congenital “with central nervous system involvement”) OR (“muscular dystrophy” congenital “fukuyama type”) OR (“cerebromuscular dystrophy” “fukuyama type”) OR (“micropolygyria with muscular dystrophy”) OR (“muscular dystrophy” duchenne) OR (“duchenne muscular dystrophy”) OR (“steinert disease” OR ((“dystrophia myotonica” OR “myotonic dystrophy”) (“type i” OR “type 1”)))) OR (mitochondrial deficiency (“complex i” OR “complex 1” OR “complex ii” OR “complex 2” OR “complex iii” OR “complex 3” OR “complex iv” OR “complex 4” OR “complex v” OR “complex 5”)) OR (“nadh-coenzyme q reductase deficiency” OR “nadh:q(1) oxidoreductase deficiency” OR “deficiency of mitochondrial nadh dehydrogenase component of complex i”) OR (“succinate coq reductase deficiency”) OR (“cytochrome c oxidase deficiency” OR “cox deficiency”) OR (mitochondrial “dna depletion syndrome”) OR (“mitochondrial dna depletion myopathy” “autosomal recessive”) OR (“leber optic atrophy” OR “leber hereditary optic neuropathy”) OR (“melas syndrome”) OR (“mitochondrial myopathy encephalopathy lactic acidosis AND stroke-like episodes”) OR (“pearson marrow-pancreas syndrome”) OR (“sideroblastic anemia with marrow cell vacuolization AND exocrine pancreatic dysfunction”) OR (“fumarase deficiency” OR “fumaric aciduria”))

H31-H90 ((choroideremia) OR (“tapetochoroidal dystrophy” progressive) OR (“choroidal sclerosis”) OR (“retinoschisis” “x-linked” juvenile) OR (“retinoschisis”) OR (“retinal degeneration slow”) OR (“retinitis pigmentosa” (“type 2” OR “type ii” OR “type 3” OR “type iii” OR “type 6” OR “type

vi”)) OR (“choroidoretinal degeneration with retinal reflex in heterozygous women”) OR (rhodopsin “retinitis pigmentosa”) OR (“retinitis pigmentosa 4”) OR (“usher syndrome” (“type iia” OR “type 2a” OR “type ii a” OR “type 2 a”)) OR (“kearns-sayre syndrome” OR “ophthalmoplegia, pigmentary degeneration of retina, AND cardiomyopathy” OR “oculocranosomatic syndrome” OR “ophthalmoplegia-plus syndrome” OR “mitochondrial cytopathy” OR “ophthalmoplegia, progressive external, with ragged-red fibers” OR “chronic progressive external ophthalmoplegia with myopathy” OR “cpeo with myopathy” OR “cpeo with ragged-red fibers”) OR (deafness sensorineural autosomal mitochondrial))

M87 (“pheochromocytoma familial” OR “familial pheochromocytoma” OR “paragangliomas 4” OR “carotid body tumors AND multiple extraadrenal pheochromocytomas” OR “pheochromocytoma, extraadrenal, AND cervical paraganglioma” OR “paragangliomas, hereditary extraadrenal” OR “pheochromocytoma, familial extraadrenal”)

Q04-Q61 (“miller-dieker” syndrome) OR (“walker-warburg syndrome”) OR (“hydrocephalus, agyria, AND retinal dysplasia” OR “hard syndrome” OR “hard +/- e syndrome” OR “warburg syndrome” OR “chemke syndrome” OR “pagon syndrome” OR “cerebroocular dysgenesis” OR “cerebroocular dysplasia-muscular dystrophy syndrome” OR “cod-md syndrome”) OR (“norrie disease” OR “atrophia bulborum hereditaria” OR “episkopi blindness”) OR (“hirschsprung disease” OR “aganglionic megacolon”) OR (“polycystic kidney disease 2” OR “polycystic kidney disease 1” OR “polycystic kidneys” OR “polycystin 1”))

Q77-Q78 (“thanatophoric dysplasia” OR “thanatophoric dwarfism”) OR (“rhizomelic chondrodysplasia punctata” (“type 1” OR “type i” OR “type 2” OR “type ii” OR “type 3” OR “type iii”)) OR (“chondrodysplasia punctata” “rhizomelic form”) OR (“chondrodystrophia calcificans punctata”) OR (“conradi-hunermann syndrome” OR “happle syndrome” OR “conradi-

hunermann-happle syndrome”) OR (“chondrodysplasia punctata 2” “x-linked dominant”) OR (achondroplasia) OR (hypochondroplasia) OR (pycnodysostosis) OR (pyknodysostosis))

Q80-Q89 ((“ichthyosis x-linked” OR “x-linked ichthyosis” OR “steroid sulfatase deficiency” OR “sts deficiency” OR “placental steroid sulfatase deficiency” OR “steroid sulfatase deficiency disease”) OR (neurofibromatosis (“type i” OR “type 1” OR “type ii” OR “type 2”)) OR (“von recklinghausen disease”) OR (“acoustic schwannomas” bilateral) OR (“bilateral acoustic neurofibromatosis”) OR (“acoustic neurinoma” bilateral) OR (“von hippel-lindau syndrome”) OR (“angelman syndrome” OR “happy puppet syndrome”) OR (“langer mesomelic dysplasia” OR (dyschondrosteosis homozygous)) OR (dyschondrosteosis homozygous) OR (“mesomelic dwarfism of the hypoplastic ulna, fibula, AND mandible type”) OR (“prader-willi syndrome” OR “prader-labhart-willi syndrome”) OR (“silver-russell syndrome” OR “russell-silver syndrome” OR “silver-russell dwarfism”) OR (“smith-lemliopitz syndrome” OR “slo syndrome” OR “rsh syndrome” OR “rutledge lethal multiple congenital anomaly syndrome” OR “polydactyly, sex reversal, renal hypoplasia, AND unilobar lung” OR “lethal acrodysgenital syndrome”) OR (x-linked “alport syndrome” OR “x-linked alport syndrome”) OR (“nephropathy AND deafness” x-linked) OR (“congenital hemidysplasia with ichthyosiform erythroderma AND limb defects” OR (“child syndrome” hemidysplasia)) OR (“zellweger syndrome” OR “cerebrohepatorenal syndrome” OR “chr syndrome”) OR (“smith-magenis syndrome” OR “chromosome 17p11.2 deletion syndrome”) OR (“williams-beuren syndrome” OR “chromosome 7q11.23 deletion syndrome” OR “williams syndrome”) OR (“wolfram syndrome” OR “diabetes insipidus AND mellitus with optic atrophy AND deafness”) OR (desmosterolosis))

Q93-Q99 ((“wolf-hirschhorn syndrome” OR “chromosome 4p16.3 deletion syndrome” OR “pitt-rogers-danks syndrome” OR “pitt syndrome”) OR (“cri-du-chat syndrome” OR “cat cry syndrome” OR “chromosome 5p deletion

syndrome”) OR ((azoospermia azf1) OR (azoospermia (“factor i” OR “factor 1”))) OR (“deleted in azoospermia”) OR (“fragile site” “mental retardation” rare*) OR (“fragile x mental retardation protein”) OR (“fragile site” “mental retardation 1”) OR (“sex-determining region y” OR “testis-determining factor”))

Apéndice **B**

B.1. Temática

Términos MeSH	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Total
Cystic Fibrosis	111	126	121	161	137	144	109	122	133	140	1304
Base Sequence	65	81	54	67	60	58	47	47	52	60	591
Brain	53	62	41	61	55	59	58	47	54	56	546
Amino Acid Sequence	63	77	61	51	56	48	39	58	46	45	544
Alleles	56	64	46	37	42	45	43	24	30	24	411
Dna Mutational Analysis	35	43	37	29	33	51	43	53	42	44	410
Antibacterial Agents	27	34	39	41	24	45	39	42	34	60	385
Amino Acid Substitution	54	14	28	25	25	31	23	42	23	28	293
Casecontrol Studies	14	26	23	35	32	30	28	30	34	40	292
Abnormalities, Multiple	30	34	27	22	28	26	18	30	29	27	271
Adrenal Gland Neoplasms	18	24	25	23	24	29	44	27	26	31	271
Charcotmarietooth Disease	30	29	20	17	21	21	33	29	27	33	260
Biopsy	37	24	22	19	25	33	23	25	27	17	252
Cell Line	25	22	25	18	22	31	17	24	33	32	249
Biological Markers	23	6	14	14	26	25	29	35	25	26	223
Cells, Cultured	18	14	17	19	21	22	22	18	32	25	208
Diagnosis, Differential	22	22	29	24	17	17	18	19	22	18	208
Fabry Disease	5	22	22	20	16	17	34	32	15	21	204
Cohort Studies	9	13	10	11	21	23	20	27	36	30	200
Hemochromatosis	20	16	26	25	22	24	10	21	9	13	186
Atbinding Cassette Transporters	19	20	20	20	19	30	11	18	11	9	177
Dna	27	30	24	11	11	10	13	16	15	17	174
Adrenal Hyperplasia, Congenital	5	16	17	14	10	16	18	33	15	23	167
Blotting, Western	14	14	18	21	20	20	23	9	14	14	167
Carrier Proteins	27	25	26	28	18	15	7	6	7	7	166
Adenosine Triphosphatases	9	16	18	18	21	20	15	16	15	17	165

Continúa en la página siguiente...

Términos MeSH	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Total
Dna, Mitochondrial	30	16	16	20	17	20	6	17	12	11	165
Genotype	21	15	15	21	14	20	12	12	16	15	161
Analysis Of Variance	10	11	12	13	14	25	18	18	18	15	154
Ataxia Telangiectasia	26	21	9	15	20	15	10	11	16	9	152
Cystic Fibrosis Transmembrane Conductance Regulator	14	19	22	11	19	17	12	10	17	7	148
Administration, Inhalation	5	18	16	19	8	15	11	15	13	15	135
Aging	5	18	11	16	8	18	15	16	12	15	134
Angelman Syndrome	7	15	13	11	15	20	10	11	16	16	134
Disease Progression	5	13	13	15	6	12	12	17	17	22	132
Chronic Disease	10	15	24	20	8	8	10	11	8	16	130
Acute Disease	13	16	12	21	14	14	8	10	8	8	124
Adenosine Triphosphate	9	17	9	8	8	12	22	12	12	13	122
Dnabinding Proteins	24	24	17	16	11	7	4	4	9	5	121
Alternative Splicing	17	16	11	21	8	10	6	9	7	10	115
Disease Models, Animal	10	12	10	4	10	10	12	14	14	19	115
Apoptosis	17	8	5	5	12	13	15	10	11	18	114
Followup Studies	7	12	15	7	9	13	13	14	12	10	112
Exons	12	13	10	8	11	9	14	14	6	13	110
Crosssectional Studies	5	11	3	11	11	16	6	12	16	17	108
Binding Sites	5	14	15	11	17	12	8	6	9	9	106
Chromosome Mapping	23	16	13	11	14	7	9	4	6	3	106
Amino Acid Metabolism, Inborn Errors	13	6	5	9	7	9	17	9	18	9	102
Antimicrobial Cationic Peptides	1	1	1	8	16	18	8	16	15	18	102
Cos Cells	7	10	14	16	15	8	6	10	10	6	102
Body Mass Index	9	10	6	8	8	6	9	7	20	18	101

Continúa en la página siguiente...

Términos MeSH	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Total
Dystrophin	4	11	13	7	10	7	11	7	12	18	100
Calcium	10	11	4	5	10	9	18	9	7	12	95
Administration, Oral	4	10	11	15	8	10	10	7	8	9	92
Dna Primers	17	10	4	6	21	12	6	5	4	7	92
Animals, Newborn	7	7	11	4	6	9	7	14	16	9	90
Amino Acid Transport Systems, Basic	2	1	0	0	1	44	9	23	3	5	88
Antioxidants	4	5	6	9	8	12	8	12	15	7	86
Asthma	15	8	13	5	5	6	8	8	8	9	85
Adrenoleukodystrophy	7	14	8	10	6	5	8	7	14	3	82
Gaucher Disease	4	10	6	10	4	9	11	15	6	7	82
Fragile X Mental Retardation Protein	7	15	3	8	10	7	8	6	6	10	80
Acylcoa Dehydrogenase	5	5	4	9	4	8	9	14	14	7	79
Achondroplasia	5	8	9	5	2	6	12	12	9	9	77
Algorithms	8	6	6	4	9	8	9	7	7	12	76
Cell Membrane	7	3	6	9	15	4	7	11	6	7	75
Chromatography, High Pressure Liquid	12	10	7	9	11	9	7	4	2	4	75
Adenoviridae	9	10	13	7	3	8	8	5	8	2	73
Biological Transport	8	7	11	4	10	3	3	8	7	12	73
Codon, Nonsense	4	10	4	4	4	15	6	9	9	8	73
Cell Differentiation	4	10	4	7	6	5	8	9	4	15	72
Genetic Testing	7	5	3	9	11	6	4	8	6	12	71
Bacterial Proteins	4	5	6	5	6	6	5	9	11	13	70
Amino Acids	10	3	4	8	7	7	8	8	5	9	69
Burkholderia Infections	6	9	6	4	10	6	9	8	6	5	69
Clinical Trials As Topic	3	8	8	6	4	8	5	13	6	8	69
Atrophy	4	6	7	7	8	7	11	9	5	4	68

Continúa en la página siguiente...

Términos MeSH	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Total
Blotting, Southern	13	6	9	11	9	8	2	3	1	6	68
Ferritins	8	3	12	9	6	8	7	5	5	5	68
Cadasil	0	0	0	1	6	10	11	21	11	7	67
Familial Mediterranean Fever	9	12	6	10	6	7	3	5	3	6	67
Absorptiometry, Photon	3	4	5	8	12	4	7	6	10	7	66
Body Height	6	7	8	10	6	6	2	5	5	11	66
Carcinoma, Hepatocellular	11	3	8	10	7	10	4	5	2	5	65
Gene Frequency	5	11	6	9	7	7	10	6	2	2	65
Cytoskeletal Proteins	11	7	12	15	5	3	3	0	5	3	64
Adaptor Proteins, Signal Transducing	3	3	1	4	10	7	5	8	10	12	63
Adrenal Cortex Hormones	5	2	6	12	5	10	2	5	5	11	63
Hirschsprung Disease	10	9	7	8	2	6	6	8	5	2	63
1Deoxynojirimycin	1	2	2	4	5	4	8	12	11	13	62
Autistic Disorder	3	8	4	8	4	8	9	2	6	10	62
Axons	4	7	7	12	5	4	4	9	7	3	62
Ataxia	6	2	7	5	4	9	3	7	7	11	61
Blood Glucose	1	3	8	4	6	6	6	11	7	9	61
Mutation	8	8	7	1	5	2	9	6	6	9	61
Amp Deaminase	4	4	2	10	5	5	8	9	6	7	60
Amyloidosis	7	10	2	6	5	6	8	2	3	11	60
Cell Nucleus	12	8	6	4	6	5	1	3	5	10	60
Arginine	6	1	7	5	0	8	14	10	6	2	59
Muscular Dystrophy, Duchenne	5	4	4	2	11	6	9	8	5	5	59
Antigens, Cd	2	6	10	9	10	4	3	5	5	4	58
Cognition Disorders	5	6	5	6	3	6	3	7	8	9	58
Genetic Predisposition To Disease	2	4	4	5	3	10	3	7	8	12	58

Continúa en la página siguiente...

Términos MeSH	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Total
Bronchiectasis	4	0	10	3	7	6	8	5	6	8	57
Chromosome Aberrations	13	8	3	8	7	3	5	4	2	4	57
Cardiomyopathies	5	6	5	6	6	6	6	6	8	2	56
Chlorides	4	4	8	5	8	6	5	6	8	1	55
Chromosome Deletion	5	6	4	4	0	2	7	12	6	7	53
Molecular Sequence Data	7	4	5	3	2	7	6	5	4	10	53
Antiinflammatory Agents	6	3	2	8	3	2	9	8	2	9	52
Bone Density	2	5	6	4	7	4	5	6	8	5	52
Bronchi	3	4	4	7	4	4	3	11	9	3	52
Mice	9	6	3	2	5	6	4	7	5	5	52

Tabla B.1: Producción por términos MeSH y fecha (Pubmed)

Términos MeSH	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Total
Fibrosis, Cystic	233	229	171	216	170	214	245	277	256	284	2295
Mutations	95	70	88	57	74	98	66	49	58	75	730
Children	58	39	72	50	59	77	42	48	57	37	539
haemochromatosis	60	60	81	67	45	61	34	28	25	50	511
Congenital MeSH qualifier	24	9	39	29	141	29	42	28	51	56	448
Therapy, Enzyme Replacement	0	0	8	17	25	61	108	67	58	93	437
Diseases	35	46	46	32	50	41	65	28	48	44	435
Gene	73	22	40	27	31	52	34	29	41	32	381
Glycogen storage disease type II	4	8	8	20	30	20	57	91	69	69	376
Willi-Prader Syndrome	29	45	33	22	31	24	43	45	62	25	359
deficiency aspects	29	45	47	30	41	35	49	38	18	21	353
disease lung obstructive	18	20	40	36	29	42	33	31	40	52	341
Neurofibromatosis 2	54	13	42	38	24	31	24	35	38	25	324
mediterranean fever familial	26	21	21	22	25	33	38	23	59	50	318
diagnosis aspect	47	33	31	25	19	37	36	15	34	40	317
adrenoleucodystrophy	32	24	11	30	31	35	14	62	34	34	307
Dystrophy, Duchenne Muscular	17	26	39	21	27	19	29	30	22	34	264
kearn sayre syndrome	22	30	30	16	9	45	34	18	1	52	257
Disease, Glycogen Storage	14	20	52	36	28	13	8	28	33	20	252
Syndrome, Rett	11	41	33	12	22	25	35	17	19	32	247
Diseases, Lysosomal Storage	7	16	7	17	20	12	48	52	29	35	243
genetic aspects	14	22	24	36	14	41	23	19	18	31	242
Transporters, ATP-Binding Cassette	24	53	25	35	15	23	19	15	17	15	241
Disease, Fabry	5	9	19	19	17	29	46	27	37	27	235
Mitochondrion	30	21	24	16	24	21	19	27	21	29	232
Retardation, Mental	9	30	19	16	13	29	27	26	28	29	226

Continúa en la página siguiente...

Términos MeSH	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Total
Iron	22	23	23	18	18	17	18	27	20	37	223
Oxidoreductases	35	46	26	31	5	26	14	14	11	15	223
Cytochrome-c Oxidase	43	27	36	40	25	3	12	4	9	23	222
Imaging, Magnetic Resonance	11	28	11	24	21	17	29	28	28	24	221
Neurofibromatosis 1	18	16	21	21	21	22	26	20	27	24	216
Dystrophy, Muscular	14	17	24	14	37	25	15	31	16	22	215
dna mitochondrial	18	19	24	10	16	32	16	30	19	28	212
Burkholderia cepacia	24	37	23	36	12	26	6	9	19	19	211
Diagnosis, Prenatal	38	22	31	10	15	24	34	7	17	10	208
Pseudomonas aeruginosa (bacteria)	15	11	14	22	10	30	21	24	29	29	205
Steroid 21 Monooxygenase	13	24	26	11	22	23	22	22	11	27	201
Inflammations	21	13	17	18	15	13	18	30	27	28	200
metabolic aspects	22	14	12	15	23	14	26	10	27	36	199
Cells, Epithelial	18	20	19	19	22	24	21	17	16	10	186
Overload, Iron	17	12	24	23	10	21	22	10	18	17	174
alpha-1-antitrypsin	26	18	28	14	11	19	17	3	23	14	173
Dystrophin	18	23	25	22	18	11	15	21	8	8	169
Neurofibromatosis NOS	23	12	11	10	13	23	19	31	9	17	168
gene therapies	29	26	18	27	14	10	7	11	16	6	164
huntingtons disease	15	10	7	6	16	17	23	31	22	17	164
Polymorphism, Genetic	12	19	15	18	16	11	18	17	14	21	161
Syndromes	11	27	8	14	14	20	11	20	20	14	159
CADASIL Syndrome	11	10	13	9	16	16	22	20	20	14	151
therapeutic aspects	6	14	25	15	15	5	15	25	15	16	151
Expression, Gene	15	7	17	12	14	9	24	17	21	14	150
Stress, Oxidative	6	15	21	19	12	13	15	15	17	17	150

Continúa en la página siguiente...

Términos MeSH	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Total
Hyperplasias, Congenital Adrenal	15	11	13	10	14	17	14	9	16	30	149
Cystic Fibrosis Transmembrane Conductance Regulator	19	21	11	11	13	16	10	14	15	14	144
Density, Bone	5	9	1	9	22	25	13	32	12	15	143
Livers	7	26	13	15	17	18	10	12	11	10	139
Syndrome, Williams	13	13	8	12	19	12	16	14	24	8	139
Cardiomyopathy NOS	7	11	15	10	11	13	11	13	32	15	138
Chain, Respiratory	7	17	22	9	5	29	8	3	18	19	137
transplantation aspects	22	21	12	10	15	13	8	2	19	14	136
Screening, Neonatal	13	13	5	18	7	13	9	20	5	32	135
Synthase, Nitric Oxide	8	20	17	19	12	11	10	11	15	10	133
Syndrome, Angelman	10	15	12	8	15	27	7	2	13	23	132
Test, Pulmonary Function	8	13	8	8	12	14	15	17	18	14	127
Disease, Hirschsprung	9	10	18	13	5	17	15	13	7	19	126
Lungs	12	7	11	10	3	13	18	15	20	17	126
Peroxisomes	14	13	4	16	17	9	7	17	13	16	126
microbicide	6	4	10	6	26	16	13	12	17	15	125
Skeletal muscle structure	11	8	14	9	12	17	10	24	6	14	125
Disease, Mitochondrial	4	10	7	13	7	9	18	30	6	16	120
Transplantations, Liver	11	7	11	16	11	18	13	10	5	18	120
Resonance, Magnetic	20	13	9	15	8	10	8	14	7	14	118
(-)-Cholesterol	13	24	14	10	8	4	20	7	8	9	117
alpha(1)-Antichymotrypsin	8	11	17	6	10	7	16	10	18	11	114
dornase alpha	8	23	7	12	12	8	12	21	2	9	114
Neuromas, Acoustic	14	17	10	7	11	11	4	10	12	17	113
Degeneration, Hepatolenticular	9	13	4	8	11	22	8	13	17	6	111

Continúa en la página siguiente...

Términos MeSH	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Total
Diabetes mellitus NOS	11	18	6	22	9	14	3	8	9	10	110
Phenotypes	11	8	6	13	7	15	14	10	13	13	110
Apolipoprotein	25	11	25	8	7	6	9	8	5	5	109
Deficiency, alpha 1-Antitrypsin	21	8	13	10	3	15	9	6	9	15	109
Aspects of disease screening	4	11	5	19	9	29	10	2	12	7	108
Oxide, Nitric	16	15	15	12	5	10	7	9	10	9	108
Infections	4	14	5	9	12	11	11	10	17	14	107
NF-kappa B	5	14	4	5	0	2	16	22	17	21	106
Therapeutic Procedures	3	7	15	12	2	13	11	27	9	7	106
alpha-Galactosidase	5	24	12	5	14	6	12	11	8	8	105
Dehydrogenase, Dihydropyrimidine	20	11	12	8	8	10	9	15	9	3	105
homo sapien	16	5	11	8	6	18	13	8	10	10	105
Myopathies	5	9	18	7	7	9	7	16	18	5	101
RNA-Binding Protein	5	22	14	17	5	7	8	12	6	5	101
Infants (Newborn)	10	4	8	8	5	24	7	2	24	5	97
Muscles	7	7	17	10	12	7	13	7	9	8	97
apoptose	11	11	9	4	10	8	17	13	5	8	96
Protein C	9	4	6	12	16	7	16	8	6	12	96
Genotypes	10	10	6	9	9	11	9	13	8	8	93
Monosomy 4q	6	5	4	8	18	14	13	10	11	4	93
Channels, Chloride	15	13	5	7	13	9	2	13	10	5	92
Complication Aspects	10	8	8	4	7	20	5	15	8	7	92
Asthmas	17	8	12	4	5	8	15	8	9	5	91
Chronic Obstructive Airways Disease	5	7	13	5	5	12	7	10	14	12	90
fibrose	6	14	4	9	8	7	7	16	9	10	90
Retinitis Pigmentosa	5	9	12	13	6	4	13	8	8	11	89

Continúa en la página siguiente...

Términos MeSH	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Total
cerebrospinal fluid	9	4	1	2	9	7	3	33	11	9	88
development aspects	10	6	2	11	8	7	15	12	3	14	88
DNA	11	9	6	14	4	4	10	11	8	11	88
Growth & development aspects	1	8	11	9	5	11	7	16	8	12	88
Infants	7	3	10	7	9	12	5	10	19	6	88
Mutation, Germ Line	5	4	24	14	4	11	13	6	5	2	88
factors transcription	3	4	12	9	6	12	11	17	5	8	87
epidemiologic	5	8	9	9	3	11	6	10	11	14	86
Spinocerebellar Ataxias	1	9	14	13	10	7	2	6	17	7	86
Electron Transport Complex III	7	6	9	4	8	5	4	8	21	13	85
Mus	11	7	9	9	2	6	13	11	12	5	85
Parkinsons disease	8	6	2	10	9	2	15	7	14	11	84
Reaction, Polymerase Chain	11	12	23	4	1	9	2	8	10	3	83
Adults	6	4	3	3	3	5	9	20	12	17	82
Composition, Body	13	8	3	13	9	2	8	6	12	8	82
hepcidin	0	0	0	1	9	11	7	10	20	24	82
Spectrometry, Mass	4	9	5	11	9	8	7	8	10	8	79
Associations, Genotype-Phenotype	10	5	9	10	11	8	3	4	5	12	77
Brains	14	7	4	9	10	8	13	6	3	3	77
Haplotype	2	2	0	6	15	13	11	9	9	8	75
Transplantation, Lung	6	7	3	3	7	6	4	6	10	23	75
Ferritins	2	7	9	8	13	10	4	6	6	9	74
Immunohistochemistry	10	9	12	7	3	3	10	6	5	7	72
Prevalence aspects	9	6	2	10	11	8	2	6	13	5	72
charcot marie tooth disease	14	4	5	15	9	0	7	1	4	12	71
Epilepsies	6	7	8	4	5	10	9	5	9	8	71

Continúa en la página siguiente...

Términos MeSH	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Total
Neutrophile	14	5	9	9	6	6	5	6	10	1	71
diseases of the lung	12	0	5	5	7	7	9	9	10	6	70
Mitochondrial Genomes	9	10	3	7	5	2	5	7	12	10	70
Life Quality	7	5	3	3	4	7	8	12	8	12	69
Diagnosis, Preimplantation	11	10	8	8	5	4	2	6	4	9	67
Respiratory failure NOS	6	2	4	10	6	4	19	3	7	5	66
Obesity NOS	8	5	4	4	7	2	11	11	3	10	65
bronchiectasis nos	3	3	6	5	6	8	9	7	6	11	64
Cerebrovascular Accidents	4	4	9	4	6	9	7	2	10	9	64
Cytokines	2	5	8	12	3	6	6	5	9	7	63
liver diseases of	7	1	7	7	2	7	5	9	11	7	63
mdx Mice, Inbred	6	8	14	3	2	1	9	10	7	3	63
Atheroscleroses	4	8	17	3	6	4	6	5	5	4	62
Pseudomonas	4	5	7	4	2	4	9	6	6	13	60
Neurilemmomas	5	0	6	5	5	8	3	8	9	9	58
Disorder, Autistic	0	7	3	4	6	5	7	5	11	8	56
Pathological accumulation of air in tissues	2	5	5	3	4	5	4	9	7	8	52

Tabla B.2: Producción por términos MeSH y fecha (SCI)

B.2. Instituciones

Inst.	trab.	nt.acum	ni.acum	nt.p	ni.p	d.teo	it.acum	it.p
0	0	0	0	0	0	0	0	0
1	598	598	1	4.06	0.01	0	0	0
1	375	973	2	6.60	0.03	0	0	0
1	358	1331	3	9.03	0.04	0	0	0
1	307	1638	4	11.11	0.05	0.01	0.01	0
1	306	1944	5	13.19	0.07	0.01	0.02	0
1	296	2240	6	15.20	0.08	0.01	0.03	0
1	289	2529	7	17.16	0.09	0.01	0.04	0
1	288	2817	8	19.12	0.11	0.01	0.05	0.01
1	278	3095	9	21	0.12	0.01	0.06	0.01
1	259	3354	10	22.76	0.13	0.01	0.07	0.01
2	230	3584	12	24.32	0.16	0.01	0.08	0.01
1	202	3786	13	25.69	0.17	0.01	0.09	0.01
1	201	3987	14	27.05	0.19	0.01	0.10	0.01
1	200	4187	15	28.41	0.20	0.01	0.11	0.01
1	199	4386	16	29.76	0.21	0.02	0.13	0.01
1	197	4583	17	31.10	0.23	0.02	0.15	0.02
1	192	4775	18	32.40	0.24	0.02	0.17	0.02
1	187	4962	19	33.67	0.25	0.02	0.19	0.02
1	186	5148	20	34.93	0.27	0.02	0.21	0.02
1	177	5325	21	36.13	0.28	0.02	0.23	0.02
1	167	5492	22	37.27	0.29	0.02	0.25	0.03
1	163	5655	23	38.37	0.31	0.02	0.27	0.03
1	161	5816	24	39.47	0.32	0.02	0.29	0.03
1	159	5975	25	40.54	0.33	0.02	0.31	0.03
1	155	6130	26	41.60	0.35	0.02	0.33	0.03
1	152	6282	27	42.63	0.36	0.03	0.36	0.04

Continúa en la página siguiente...

Inst.	trab.	nt.acum	ni.acum	nt.p	ni.p	d.teo	it.acum	it.p
1	151	6433	28	43.65	0.37	0.03	0.39	0.04
1	146	6579	29	44.64	0.38	0.03	0.42	0.04
1	145	6724	30	45.63	0.40	0.03	0.45	0.05
1	144	6868	31	46.60	0.41	0.03	0.48	0.05
1	143	7011	32	47.57	0.42	0.03	0.51	0.05
1	141	7152	33	48.53	0.44	0.03	0.54	0.06
1	140	7292	34	49.48	0.45	0.03	0.57	0.06
1	139	7431	35	50.42	0.46	0.03	0.60	0.06
1	137	7568	36	51.35	0.48	0.03	0.63	0.06
1	136	7704	37	52.28	0.49	0.03	0.66	0.07
2	134	7838	39	53.19	0.52	0.03	0.69	0.07
1	133	7971	40	54.09	0.53	0.03	0.72	0.07
1	131	8102	41	54.98	0.54	0.03	0.75	0.08
1	129	8231	42	55.85	0.56	0.04	0.79	0.08
1	128	8359	43	56.72	0.57	0.04	0.83	0.08
2	126	8485	45	57.58	0.60	0.04	0.87	0.09
1	125	8610	46	58.42	0.61	0.04	0.91	0.09
3	121	8731	49	59.25	0.65	0.04	0.95	0.10
1	119	8850	50	60.05	0.66	0.04	0.99	0.10
1	118	8968	51	60.85	0.68	0.04	1.03	0.11
1	117	9085	52	61.65	0.69	0.04	1.07	0.11
1	116	9201	53	62.43	0.70	0.04	1.11	0.11
1	115	9316	54	63.22	0.72	0.05	1.16	0.12
1	113	9429	55	63.98	0.73	0.05	1.21	0.12
1	112	9541	56	64.74	0.74	0.05	1.26	0.13
2	111	9652	58	65.50	0.77	0.05	1.31	0.13
1	110	9762	59	66.24	0.78	0.05	1.36	0.14
2	106	9868	61	66.96	0.81	0.05	1.41	0.14

Continúa en la página siguiente...

Inst.	trab.	nt.acum	ni.acum	nt.p	ni.p	d.teo	it.acum	it.p
1	104	9972	62	67.67	0.82	0.06	1.47	0.15
1	101	10073	63	68.35	0.84	0.06	1.53	0.16
1	100	10173	64	69.03	0.85	0.06	1.59	0.16
2	98	10271	66	69.70	0.88	0.06	1.65	0.17
3	97	10368	69	70.35	0.92	0.06	1.71	0.17
1	96	10464	70	71	0.93	0.06	1.77	0.18
1	95	10559	71	71.65	0.94	0.07	1.84	0.19
2	94	10653	73	72.29	0.97	0.07	1.91	0.20
2	93	10746	75	72.92	1	0.07	1.98	0.20
1	92	10838	76	73.54	1.01	0.07	2.05	0.21
4	91	10929	80	74.16	1.06	0.07	2.12	0.22
3	90	11019	83	74.77	1.10	0.07	2.19	0.22
1	89	11108	84	75.37	1.11	0.08	2.27	0.23
2	87	11195	86	75.97	1.14	0.08	2.35	0.24
1	86	11281	87	76.55	1.15	0.08	2.43	0.25
2	85	11366	89	77.13	1.18	0.08	2.51	0.26
1	84	11450	90	77.70	1.19	0.08	2.59	0.26
2	83	11533	92	78.26	1.22	0.09	2.68	0.27
4	82	11615	96	78.82	1.27	0.09	2.77	0.28
1	81	11696	97	79.36	1.29	0.09	2.86	0.29
1	80	11776	98	79.91	1.30	0.09	2.95	0.30
2	79	11855	100	80.44	1.33	0.10	3.05	0.31
1	77	11932	101	80.97	1.34	0.10	3.15	0.32
2	76	12008	103	81.48	1.37	0.10	3.25	0.33
2	75	12083	105	81.99	1.39	0.11	3.36	0.34
1	74	12157	106	82.49	1.41	0.11	3.47	0.35
2	73	12230	108	82.99	1.43	0.11	3.58	0.37
1	72	12302	109	83.48	1.45	0.12	3.70	0.38

Continúa en la página siguiente...

Inst.	trab.	nt.acum	ni.acum	nt.p	ni.p	d.teo	it.acum	it.p
5	71	12373	114	83.96	1.51	0.12	3.82	0.39
1	70	12443	115	84.43	1.53	0.12	3.94	0.40
2	69	12512	117	84.90	1.55	0.13	4.07	0.42
2	68	12580	119	85.36	1.58	0.13	4.20	0.43
2	67	12647	121	85.82	1.61	0.13	4.33	0.44
2	66	12713	123	86.27	1.63	0.14	4.47	0.46
1	64	12777	124	86.70	1.65	0.15	4.62	0.47
3	63	12840	127	87.13	1.69	0.15	4.77	0.49
3	62	12902	130	87.55	1.73	0.16	4.93	0.50
1	61	12963	131	87.96	1.74	0.16	5.09	0.52
5	60	13023	136	88.37	1.80	0.17	5.26	0.54
2	59	13082	138	88.77	1.83	0.17	5.43	0.55
5	58	13140	143	89.16	1.90	0.18	5.61	0.57
1	57	13197	144	89.55	1.91	0.18	5.79	0.59
1	55	13252	145	89.92	1.92	0.20	5.99	0.61
7	54	13306	152	90.29	2.02	0.21	6.20	0.63
6	53	13359	158	90.65	2.10	0.21	6.41	0.65
5	52	13411	163	91	2.16	0.22	6.63	0.68
2	51	13462	165	91.35	2.19	0.23	6.86	0.70
9	50	13512	174	91.69	2.31	0.24	7.10	0.73
3	49	13561	177	92.02	2.35	0.25	7.35	0.75
1	48	13609	178	92.35	2.36	0.26	7.61	0.78
5	47	13656	183	92.66	2.43	0.27	7.88	0.81
5	46	13702	188	92.98	2.50	0.28	8.16	0.83
5	45	13747	193	93.28	2.56	0.30	8.46	0.86
5	44	13791	198	93.58	2.63	0.31	8.77	0.90
6	43	13834	204	93.87	2.71	0.32	9.09	0.93
6	42	13876	210	94.16	2.79	0.34	9.43	0.96

Continúa en la página siguiente...

Inst.	trab.	nt.acum	ni.acum	nt.p	ni.p	d.teo	it.acum	it.p
3	41	13917	213	94.44	2.83	0.36	9.79	1
2	40	13957	215	94.71	2.85	0.37	10.16	1.04
9	39	13996	224	94.97	2.97	0.39	10.55	1.08
4	38	14034	228	95.23	3.03	0.41	10.96	1.12
8	37	14071	236	95.48	3.13	0.44	11.40	1.16
8	36	14107	244	95.73	3.24	0.46	11.86	1.21
4	35	14142	248	95.96	3.29	0.49	12.35	1.26
8	34	14176	256	96.19	3.40	0.52	12.87	1.32
3	33	14209	259	96.42	3.44	0.55	13.42	1.37
6	32	14241	265	96.63	3.52	0.58	14	1.43
10	31	14272	275	96.84	3.65	0.62	14.62	1.49
7	30	14302	282	97.05	3.74	0.66	15.28	1.56
12	29	14331	294	97.25	3.90	0.71	15.99	1.63
10	28	14359	304	97.44	4.03	0.76	16.75	1.71
6	27	14386	310	97.62	4.11	0.82	17.57	1.80
3	26	14412	313	97.79	4.15	0.88	18.45	1.89
13	25	14437	326	97.96	4.33	0.96	19.41	1.98
10	24	14461	336	98.13	4.46	1.04	20.45	2.09
15	23	14484	351	98.28	4.66	1.13	21.58	2.20
16	22	14506	367	98.43	4.87	1.24	22.82	2.33
14	21	14527	381	98.58	5.06	1.36	24.18	2.47
20	20	14547	401	98.71	5.32	1.50	25.68	2.62
25	19	14566	426	98.84	5.65	1.66	27.34	2.79
13	18	14584	439	98.96	5.83	1.85	29.19	2.98
26	17	14601	465	99.08	6.17	2.07	31.26	3.19
32	16	14617	497	99.19	6.60	2.34	33.60	3.43
31	15	14632	528	99.29	7.01	2.66	36.26	3.70
31	14	14646	559	99.38	7.42	3.05	39.31	4.02

Continúa en la página siguiente. . .

Inst.	trab.	nt.acum	ni.acum	nt.p	ni.p	d.teo	it.acum	it.p
41	13	14659	600	99.47	7.96	3.54	42.85	4.38
46	12	14671	646	99.55	8.57	4.15	47	4.80
63	11	14682	709	99.63	9.41	4.94	51.94	5.31
55	10	14692	764	99.69	10.14	5.98	57.92	5.92
80	9	14701	844	99.76	11.20	7.38	65.30	6.67
82	8	14709	926	99.81	12.29	9.34	74.64	7.63
108	7	14716	1034	99.86	13.72	12.20	86.84	8.87
159	6	14722	1193	99.90	15.83	16.61	103.45	10.57
202	5	14727	1395	99.93	18.51	23.92	127.37	13.01
313	4	14731	1708	99.96	22.67	37.38	164.75	16.83
504	3	14734	2212	99.98	29.36	66.44	231.19	23.62
1103	2	14736	3315	99.99	43.99	149.50	380.69	38.90
4220	1	14737	7535	100	100	598	978.69	100

Tabla B.3: *Distribución de las instituciones en el SCI*

B.3. Colaboración

Término MeSH	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Promedio
Syndrome, Angelman	6,17	7,25	4,11	7	5	6,67	7,29	53	4,5	7,08	10,81
Haplotype	5	9	6,22	7,8	10	9,89	8,43	10,83	8,56	8,8	8,70
Associations, Genotype-Phenotype	9,29	10,2	6,22	8,75	8,2	9	6,33	9,75	8,2	10,17	8,61
Spinocerebellar Ataxias	11	7	8,25	7,22	9	7,2	10,5	9	7,14	7,14	8,35
Transplantations, Liver	8,71	9,25	6,6	7,29	5,75	5,17	9,2	9,17	10,4	10,75	8,23
Apolipoprotein	10,12	5,4	7,38	6,57	7,6	6,5	10,5	6,33	8	11,67	8,01
NF-kappa B	7	7,75	7	5,67		10	8,9	6,09	9,12	9,54	7,90
Protein C	6,17	3,67	18,5	7,38	6,14	6,8	8,9	7,5	6,6	6,89	7,86
Disorder, Autistic		7,71	11	7,75	8	5,33	6,43	8,6	7,27	8,12	7,80
Cytochrome-c Oxidase	6,5	7,06	6,5	6,33	5,7	9,5	12,33	8,5	8	7,1	7,75
RNA-Binding Protein	8,33	10,82	6,14	5	5	9	13,5	9	4,33	5,67	7,68
Polymorphism, Genetic	6,92	8,23	6,4	7,1	7	8,27	6,69	8	9,14	8,94	7,67
Retardation, Mental	7,8	5,71	7,14	7,42	6,88	10,13	5,84	7,62	8,44	8,56	7,55
Atherosclerosis	9,25	8,12	5	9	7,83	5,25	7	5,4	7,8	9,5	7,42
Oxidoreductases	6,62	6,12	6,2	5,57	9	9,2	7,5	7,38	6,71	9,14	7,34
Mutations	7	7,57	6,64	7,98	5,98	7,38	7,78	7,86	6,72	8,51	7,34
cerebrospinal fluid	9,75	6,5	5	5	7,14	5,67	8,67	7	7,9	10,71	7,33
(-)-Cholesterol	7,4	7,05	7,8	7,67	7	5	8,19	5,14	8,5	9,56	7,33
Glycogen storage disease type II	5	5	7	8,5	10,33	7	8,29	6,15	7,07	8,58	7,29
Dehydrogenase, Dihydropyrimidine	7,86	7,5	5,33	6,75	5,67	8	7,4	5,73	7,67	11	7,29
Synthase, Nitric Oxide	5,2	5,2	6,3	7,22	5,71	6	10,5	9,17	9,62	7,5	7,24
Mitochondrial Genomes	7,17	6,7	3	7,5	7	9,5	7,2	7,57	8,25	8	7,19
Phenotypes	5,5	7,14	7	6	6,2	7,1	6,38	8,86	6,78	10,82	7,18
huntingtons disease	6	6,71	4,2	9,5	9,92	6,55	7	5,89	9,25	6,73	7,18
Monosomy 4q	5,8	8	4,75	6,4	7,11	5,67	9,6	7,67	7,11	9,5	7,16
Screening, Neonatal	6	7,8	7,4	5,91	7,29	7,5	7,33	7,71	8	6,64	7,16
Diagnosis, Preimplantation	11,38	5,71	7	7,38	7,6	7,75	4,5	6,17	8,67	5,33	7,15

Continúa en la página siguiente...

Término MeSH	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Promedio
Transporters, ATP-Binding Cassette	7,3	9,63	7,18	6,65	7,15	6,29	5,77	7,75	6,71	6,57	7,10
liver diseases of	4,86	8	5,6	5,5	14,5	5,6	5,4	6	8,25	7,2	7,09
Overload, Iron	4,77	6,91	6,77	6,61	7,89	7,28	6,89	6,43	9	8,29	7,08
Parkinsons disease	5,17	5	6	6,67	4,8	9	7,5	6,71	8,6	10,83	7,03
Mitochondrion	7,23	5,94	5,2	5,62	6,2	7,71	6,59	6,52	11	7,81	6,98
Syndrome, Rett	6,17	8,39	6,38	4,7	6,13	7,06	7,18	8,19	7,85	7,76	6,98
Therapy, Enzyme Replacement			5	6,4	8,33	6,73	7,22	7,1	8,19	6,62	6,95
Mutation, Germ Line	7,33	7,25	8,1	6,56	6,5	7,86	6,86	5,5	8	5,5	6,95
Neurofibromatosis 2	6,06	7,1	5,78	6,55	7,1	6,67	8,73	6,95	6,61	7,44	6,90
adrenoleucodystrophy	5	6,36	5,6	8,17	7	7,67	6,5	7,07	8,7	6,8	6,89
Disease, Hirschsprung	8	5,71	6,5	5,8	5,4	9,67	6,45	8,5	7,5	5,27	6,88
dna mitochondrial	7,07	6	5,33	5,29	7,13	7,06	7,56	8,12	6,69	8,41	6,87
Expression, Gene	5,43	6,83	5,38	5,6	6,55	6,78	8	8,23	7,4	8,09	6,83
Electron Transport Complex III	3,25	4,6	4	10,5	8,33	9	7	8,25	6,14	7,17	6,82
Resonance, Magnetic	6,6	7	5,6	5,75	8,4	9,14	6,6	6,33	5,2	7,57	6,82
Cardiomyopathy NOS	6,86	7,12	5,11	7	6	8,22	5,71	5,8	7,94	8,33	6,81
genetic aspects	4,08	6,83	6,56	7,6	8,42	4,96	5,67	6,28	9,71	7,88	6,80
Gene	6,15	6,31	6,5	6,19	7,87	7,52	6,31	8,21	6,92	6	6,80
Disease, Glycogen Storage	5	6,8	4,62	8,64	8	8	7,5	8	6,22	5,2	6,80
Mus	6,12	6	4,86	6,38	9,5	6,33	5,89	6,25	6,64	10	6,80
Peroxisomes	4,38	5,8	11	6,43	5,17	8	6	6,75	7,9	6,4	6,78
hepcidin				8	7,75	6,45	6,14	5,7	7,68	5,75	6,78
apoptose	7,1	7	4,11	5,75	7,8	4,88	6	6,3	9,8	8,75	6,75
Diseases, Lysosomal Storage	7,25	4,45	7,5	7,08	7,38	7,8	7,3	6,65	6	6,05	6,75
metabolic aspects	5,92	6,9	6,6	6,43	5,38	5,75	6,89	6,5	8,83	8,21	6,74
homo sapien	4,54	8	5	5,5	6	6,62	6,11	8	9,67	7,89	6,73
Imaging, Magnetic Resonance	5,36	6,76	6,5	6,55	6,33	6,64	6,8	7,89	6,81	7,64	6,73

Continúa en la página siguiente. . .

Término MeSH	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Promedio
Obesity NOS	6,38	5,6	5,5	8	5,5	5	4,45	5	14,67	6,67	6,68
Stress, Oxidative	6,17	6	4,83	5,45	6,5	6,82	7,64	7,15	6,85	9,3	6,67
Ferritins	5	7,14	5,33	5,5	6,4	6,6	8,75	7,67	6,5	7,78	6,67
Livers	8,71	6,5	4,8	5	6,08	6,92	8,78	6,11	7,12	6,44	6,65
deficiency aspects	6,29	5,89	6,7	6,31	6,43	7,14	6,79	7,1	6,89	6,92	6,65
CADASIL Syndrome	6,7	7,78	4,33	5,89	6,53	6	6,35	6,84	7,32	8,69	6,64
Dystrophy, Muscular	4,25	4,87	4,72	7	7,25	5,75	5,15	11	8,55	7,71	6,63
epidemiologic	3	6,67	8,67	3,5	8,33	7,33	7,67	5,1	7,2	8,71	6,62
alpha-Galactosidases	8	3,88	6	4,33	7,67	9	5,17	4,75	10,71	6,5	6,60
Cytokines	5,5	5,4	6,38	6,89	3	8	5,83	6,2	9,44	9,33	6,60
mediterranean fever familial	5,22	7,91	7	6,92	6,92	7,4	6,06	6	6,19	6,22	6,58
Muscles	5	7,8	6,67	5,14	6,3	6,33	5,36	9,67	6,5	6,86	6,56
Lungs	6,18	3,67	6	6,4	7,33	8,55	5,75	6,13	7,39	8,18	6,56
Spectrometry, Mass	5	5,67	6	6,67	7,17	7,43	5	6,75	9,17	6,67	6,55
Epilepsies	6,33	5,5	7	8	5,5	6,38	5,33	8	6	7,14	6,52
Iron	6,73	6,61	5,36	5,5	6,72	6,47	7,61	6,35	7,7	6,06	6,51
DNA	5,43	7	5,2	5,43	8	7,5	5	7,22	7	7,25	6,50
Dystrophin	5,33	5,95	5,12	6,41	6,39	7,36	7,2	8,67	6,71	5,88	6,50
kearn sayre syndrome	6,6	3,88	5,71	5,22	8	7,25	8	5	9	5,67	6,43
Reaction, Polymerase Chain	6	5,25	6,25	6,25	6	8	5,5	7,86	6,88	6,33	6,43
microbicide	7	6,33	5,71	7	9,36	7	6	4,17	6,4	5,22	6,42
Complication Aspects	4,43	6,5	4,8	9	5,25	8,4	5	5,6	8,6	6,5	6,41
Dystrophy, Duchenne Muscular	4,82	5,77	4,87	6,33	6,37	6,89	7,35	7,77	6,05	7,79	6,40
haemochromatosis	6,08	7	6,19	6,52	6,06	6,74	6,36	6,04	7,13	5,84	6,40
factors transcription	4,33	5,33	6,88	5	8	7,12	7,67	7,9	5,5	6	6,37
Immunohistochemistry	5,67	5,33	6,33	6,57	4	8,67	6,9	7,83	4,6	7,71	6,36
Skeletal muscle structure	5,38	5,71	4,22	5,25	5,62	5,33	6,33	7,92	10,5	7	6,33

Continúa en la página siguiente...

Término MeSH	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Promedio
Diabetes mellitus NOS	6,57	6,92	4,4	5	5,83	5	7,33	10,33	6,83	5	6,32
Chain, Respiratory	4,67	5,44	5,57	4,33	6,2	7,7	6,4	6	8,88	7,86	6,31
Genotypes	5,86	7,9	5	4,4	5,12	7,82	5,83	6,64	9,14	5,14	6,29
Willi-Prader Syndrome	6,21	6,12	5,9	6,29	6	5,47	5,48	6	7,69	7,33	6,25
fibrose	3,83	7,08	5,25	6,33	7,5	5,43	5,71	8,15	6,56	6,56	6,24
Aspects of disease screening	5	5,67	6,4	5,33	5,17	8,07	6,3	4	8,44	7,86	6,22
Neurofibromatosis 1	6,06	4,86	6,47	5,95	7,24	6,11	6,74	6,17	6,78	5,82	6,22
diseases of the lung	6,5		3,5	8	4,2	6,83	7,12	3,83	7,5	8,5	6,22
Diagnosis, Prenatal	5,83	6	5,92	7,14	5,25	8,12	6,22	5,2	7,17	5,29	6,21
Inflammations	5,44	5,08	6,29	6,47	5,07	7,31	6,29	5,08	8,15	6,93	6,21
Diseases	5,42	6,09	5,68	6,15	5,71	7,86	6,04	7,55	5,54	6	6,20
transplantation aspects	6,6	5,78	6,8	6,33	4,33	7,5	5,83	3	8,18	7,56	6,19
Neuromas, Acoustic	5,57	6,12	4,4	5	4,75	7	8,5	6,88	4,75	8,78	6,18
Cystic Fibrosis Transmembrane Conductance Regulator	5,33	7,33	6	4,36	6,31	5,19	8	6,36	6,53	6,29	6,17
Syndromes	5,75	5,6	6,75	7,18	6,78	4,93	6,25	6,77	5,19	6,45	6,17
Degeneration, Hepatolenticular	5,67	7	7	4,83	4,4	6,4	7,43	6,33	6,92	5,67	6,17
gene therapies	6,22	5,57	5,12	6,82	5,79	8,89	5,25	7	5,25	5,5	6,14
Cells, Epithelial	5,79	6	5,5	5,07	6	6,65	7	6,07	6,09	6,88	6,11
disease lung obstructive	5,89	5,9	5,55	5,74	5,8	6,43	4,82	6,44	6,95	7,42	6,09
development aspects	3,86	7,33	6,5	5,36	7	5,43	4,67	5,82	6	8,82	6,08
Density, Bone	5,75	5,8	6	6,33	6,67	7,33	7,14	5,54	5,29	4,9	6,08
therapeutic aspects	7,4	4,6	5,33	5,36	6,88	2,5	7,22	4,4	9,54	7,38	6,06
diagnosis aspect	6,82	6,14	5,83	6,64	6,13	5,1	7,38	4,91	5,31	6,24	6,05
Composition, Body	4,33	6,33	6	5,12	6,6	7	6,43	6,25	5,62	6,75	6,04
Myopathies	4,6	3,86	5,57	8	6,57	7	4,43	9	7,71	3,5	6,02
Congenital MeSH qualifier	4,92	5,4	8,91	5,4	6,1	6,91	6,38	4,25	5,5	6,45	6,02
mdx Mice, Inbred	4,8	6,43	4,93	3,67	6	6	6,62	7,62	6,17	7,67	5,99

Continúa en la página siguiente...

Término MeSH	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Promedio
Steroid 21 Monooxygenase	5,57	5,42	5,07	5,5	7,27	5,73	7,42	5,73	6,17	6	5,99
charcot marie tooth disease	7,07	6,5	5,6	5,5	6,56		6,43	4	6,75	5,4	5,98
Transplantation, Lung	4	7	6	1,5	6,14	6,33	5	8,17	7,9	7,74	5,98
Cerebrovascular Accidents	6,5	5,5	4,38	5,25	5,17	6,4	6,33	6	8,75	5,14	5,94
Infections	4,75	5,67	5	5	5,73	8	6	5,75	6,08	7,42	5,94
Pseudomonas aeruginosa (bacteria) Disease, Fabry	5,13	7	6	5,41	5	5,83	4,86	6	6,55	7,59	5,94
Burkholderia cepacia	6	3	4,73	5,5	6,31	6,88	6,79	6,38	7,76	6	5,94
Retinitis Pigmentosa	5,13	6,36	5,92	5,42	6,43	5,8	3,25	7,6	8,45	4,7	5,91
Test, Pulmonary Function	7	6,57	6	6,14	4,5	3	6,88	5,14	6,67	7,14	5,90
Fibrosis, Cystic	4,62	6,08	7,62	5,88	5,5	6,09	5,8	5	5,75	6,08	5,84
bronchiectasis nos	5,39	5,81	5,54	4,97	5,84	6,08	5,64	5,99	6,59	6,53	5,84
Pseudomonas	9,33	6	5,17	3	7,5	5,12	4,44	6,14	6	5,45	5,82
Disease, Mitochondrial	6	7,25	4,86	6,25	5,5	5	6,78	5,2	5,83	4,92	5,76
alpha(1)-Antichymotrypsin	4,75	3,29	4,25	5,29	6,43	8	6,89	6,56	5,33	6,54	5,73
Channels, Chloride	5,17	3,14	6,3	7	5,57	4,75	6,55	4,67	6,45	7,71	5,73
Respiratory failure NOS	6,42	5,7	4,2	4,5	5,89	4,29	8	6,36	5,14	6,8	5,73
Brains	5,33	6,5	8	3,71	3,8	6	5,5	6,67	6,43	5	5,69
Infants (Newborn)	6,6	6,4	4	7,57	5	5,6	6,5	5,5	6,33	3	5,65
Infants	4,78	6,75	5,4	6,86	6	6,56	4,71	3	6,4	5,6	5,61
Life Quality	5,2	5	5,86	4,33	7	3,44	5,4	7	7,3	5,5	5,60
Neutrophile	4,25	3	4,67	4,33	5,5	6,75	5,38	6,67	7,38	7,42	5,54
Neurilemmomas	6,33	5	3,2	6,71	5	5,33	3,25	5,17	7,12	8	5,51
Prevalence aspects	5,4		3,67	7	6,75	7,33	4,33	4,57	4,88	5,57	5,50
Oxide, Nitric	4,71	4,5	6,5	5,62	6,17	6	4	6	6,22	5,2	5,49
Pathological accumulation of air in tissues	5,64	5	5,17	5,45	2,75	5,8	6,8	6,75	5,22	5,83	5,44
Children	4,5	5,2	4	3,67	6,25	6	5,75	5,11	6	7,5	5,40
	5,26	5,28	5,11	4,98	5,79	5,45	4,68	6,05	5,83	5,53	5,40

Continúa en la página siguiente...

Término MeSH	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Promedio
Chronic Obstructive Airways Disease	6,8	6,14	4,38	5,2	3,6	4,33	4,17	4,11	7,36	7,83	5,39
Growth & development aspects	4	6,2	4,29	6	6,75	5	6,33	4,5	5,4	5,39	5,39
Neurofibromatosis NOS	6,15	3,75	5,8	5,67	5,83	5	5,15	5,67	5,22	5,41	5,37
dornase alpha	4,5	4,9	5,2	4,2	4,33	4	4,6	4,4	6,5	11	5,36
alpha-1-antitrypsin	6,21	4,57	5,33	6,62	5,14	5,55	5,47	4	5,31	4,78	5,30
Deficiency, alpha 1-Antitrypsin	6,92	4,75	5,38	5,5	4,67	5	4,22	3,67	5,56	7,2	5,29
Hyperplasias, Congenital Adrenal	5,46	5,91	4,17	5,9	4,36	4,38	6,07	4,67	6,93	4,53	5,24
Asthmas	4,88	4,5	6,67	5,5	3,8	5,86	4,31	3,25	5,89	5,4	5,01
Therapeutic Procedures	1,67	5,5	5,44	3,44	4,5	3,5	5,62	6,44	6,33	6,14	4,86
Syndrome, Williams	5,27	4,8	4,62	4,91	3,82	5,5	4,17	4,75	4,69	5,29	4,78
Adults	3,83	2	4,67	5,33	4	3,4	3	4,2	4,89	5,33	4,07

Tabla B.4: *Media de coautoría por términos MeSH y año (SCI)*

Término MeSH	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Promedio
Alleles	6,66	7,75	8,24	7,62	6,88	7,98	8,07	28,71	8,4	10,58	10,09
Gene Frequency	7,8	7,36	6,33	7,33	7	7,86	7,9	17,83	9	12	9,04
Dna Mutational Analysis	8,34	7,14	7,84	7,62	8,12	9,04	8,63	10,42	11,62	10	8,88
Adaptor Proteins, Signal Transducing	10	6	11	10,75	4,9	13	9,4	8,12	9,1	6,5	8,88
Dna Primers	7,59	10,9	6,5	6,5	11,1	8,08	6,5	6,4	15	9,29	8,79
Adenoviridae	8,22	5	5,85	6,71	7	5,75	7,25	24,8	7,5	9	8,71
Abnormalities, Multiple	7,8	7,24	12,96	6,36	7,86	8,31	7,33	12	8,38	8,74	8,70
Base Sequence	7,89	9,38	6,8	6,58	9,77	7,66	6,87	10,04	12,15	9,75	8,69
Cohort Studies	9,22	7,31	5,6	5,27	12,14	8,35	8,45	10	7,03	11,2	8,46
Atrophy	4,75	4,67	9,14	5,14	6,62	4,86	7,82	18,44	6,4	16	8,38
Brain	7,74	8,61	5,93	7,82	7,69	8,05	9,31	8,66	10,5	9,16	8,35
Codon, Nonsense	6,75	9,1	5,5	7,75	6,25	7,6	12,67	10,11	9,22	8,25	8,32
Chromosome Mapping	6,22	6,12	6,54	15,36	7,36	8,14	6,89	8,5	9,17	7,33	8,16
Dna, Mitochondrial	9,27	5,56	6,5	8,05	7,41	9,25	13,5	7,94	5,67	8,09	8,12
Dna	6,44	7,33	6,5	6,82	6,73	5,9	5,77	15,5	11,93	8	8,09
Amino Acid Sequence	7,4	7,58	7,8	6,8	8,32	7,88	7,59	6,72	10,02	9,91	8,00
Adrenoleukodystrophy	3,71	10,64	12,25	8,1	7,17	7	7,38	7,57	7,5	8,67	8,00
Alternative Splicing	9,24	10,75	7,09	9,33	6,38	6,6	5,17	8,56	6,71	8,8	7,86
Familial Mediterranean Fever	7,67	9,92	3,5	6,5	8,5	5,71	2,67	6,2	9	18,83	7,85
Adrenal Gland Neoplasms	6,39	6,83	7,56	8,74	5,46	10,48	7,61	7,19	7,85	10,35	7,85
Exons	6,58	6,23	7,9	7,5	8,82	11,56	7,07	7,79	9,67	5,31	7,84
Genotype	6,52	6,53	5,6	6,38	8,07	6,3	7,67	13,5	7,69	10,13	7,84
Blotting, Western	7,86	10,57	5,89	6	6,75	7,1	7,87	8,67	8,29	9,21	7,82
Chromatography, High Pressure Liquid	6,83	5,4	7,43	5,89	8,64	9	6,86	8	5	12,75	7,58
Adenosine Triphosphatases	6,33	5,88	9,06	5,33	6,67	4,85	8,27	11,75	6,07	9,41	7,36
Axons	6,75	6	7,14	8,42	6	7,25	11,5	7,11	5,86	7,33	7,34
Animals, Newborn	6,71	7,57	6,91	6,75	8	6,56	7,57	7,93	7,38	7,33	7,27

Continúa en la página siguiente...

Término MeSH	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Promedio
Amino Acid Metabolism, Inborn Errors	10,15	5,17	5,8	9,11	6,14	6,56	5,88	4,22	10	9,67	7,27
Adenosine Triphosphate	6	7,88	8,22	4,5	9,12	8,17	8,23	6,92	7,58	5,92	7,25
Cell Line	8,84	4,91	5,6	5,5	8	7,1	6,24	11,04	7,88	7,25	7,24
Analysis Of Variance	7,1	8,55	5,08	7	7,36	6,16	8,33	7,89	7,56	7,27	7,23
Blotting, Southern	7,85	6,17	5,78	7,45	5,33	10,62	5,5	7	5	11,17	7,19
Charcotmarieetooth Disease	5,93	6,55	7,7	7,65	9,33	6,95	6,88	4,55	7,85	8,24	7,16
Cells, Cultured	6,67	6,21	5,76	6,42	7	7,36	8,05	6,94	7,41	9,32	7,11
Cos Cells	5,57	5,3	5,64	4,75	8,8	7,25	5,67	8,2	9,4	10,17	7,08
Casecontrol Studies	5,71	6,65	5,35	7	7,81	6,27	5,96	7,13	8,12	10,3	7,03
Hirschsprung Disease	6,9	6,56	6,14	7,88	6	6,33	4,67	11,75	10	2,5	6,87
Biopsy	7,57	6,62	6,27	7,11	7,2	7,33	6,78	5,56	5,89	7,35	6,77
Amino Acids	5,1	5,33	11,75	6,88	4,43	5,57	5,75	6	5,6	11	6,74
Carrier Proteins	6,48	7,56	7,88	7,71	3,78	5,73	5,86	7,5	4,71	9,86	6,71
Dystrophin	4,25	5,36	5,31	6,71	7,5	5,71	5,64	9,71	5,83	11	6,70
Arpinding Cassette Transporters	5,53	6,15	6,3	7,05	7,37	6,97	5,64	6,83	7,64	7,22	6,67
Amino Acid Substitution	7,06	4,5	7,89	6,28	7,44	6,77	5,7	5,64	7,09	8,11	6,65
Bacterial Proteins	4	7	6,67	8,2	8,67	6,67	4,8	5,67	8,55	5,92	6,62
Apoptosis	6,24	6,25	2,6	6	6,08	6,38	5,73	7,3	10,27	8,89	6,57
Crosssectional Studies	5,2	5,73	5,67	5,27	6,27	5,69	6,17	7,42	7	10,59	6,50
Antioxidants	4,25	4,8	5,83	6,44	7,5	6	7,88	6,75	5,53	9,86	6,48
Ataxia Telangiectasia	5,85	6,29	5,89	5,73	4,45	5,27	6,5	10,18	6,5	8,11	6,48
Absorptiometry, Photon	5,33	7,5	5,2	6,25	6,42	7	5,29	6,5	7,6	7	6,41
Administration, Oral	9,25	5,7	7,27	4,53	4,38	4,6	5,9	6,86	7,25	7,44	6,32
Dnabinding Proteins	6,67	7,33	7,35	5,94	5,73	4,14	3,25	10	6,22	6,4	6,30
Body Height	6,5	5,43	3,38	6,1	6,5	9,5	2,5	7,6	4,6	10,18	6,23
Angelman Syndrome	4,86	6,33	4,54	4,55	7,53	5,35	9	7,18	5,06	7,5	6,19
Biological Markers	5,52	3,5	7	6,14	6,58	6,68	5,14	6,57	6,4	7,65	6,12

Continúa en la página siguiente...

Término MeSH	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Promedio
Carcinoma, Hepatocelular	6,09	5,67	5,62	6,7	5	5,1	7,75	8	6	5,2	6,11
Followup Studies	6,43	5,33	4,4	5,43	6,56	5,69	6,54	7,36	6,75	6,4	6,09
Ferritins	3,5	8	6,08	7,67	5,5	4,88	4,86	4,4	9,6	6	6,05
Blood Glucose	2	8	4,5	5,75	5	5,67	7,5	7,36	6,57	7,78	6,01
Calcium	8,1	7,18	6,5	4,6	4,5	4,33	5,17	6	5,86	7,83	6,01
Cell Differentiation	5,75	4,9	3,75	4,86	9,83	4,6	6,88	7,22	6	6,13	5,99
Adrenal Hyperplasia, Congenital	5,8	3,69	5,71	6,36	5,3	5,5	6,06	7,79	6,47	7,22	5,99
Mutation	6,75	6,38	7	6	4,2	6	5,44	7	3,33	7,56	5,97
Disease Progression	6,2	6,38	5,69	5,13	4,67	6,58	4,33	7,82	6,53	5,86	5,92
Body Mass Index	5,67	5,3	5,33	6,5	4,88	5,67	6	5,57	6,65	7,44	5,90
Binding Sites	8	5,79	4,8	5,45	5,18	5,75	4,5	6,17	6,67	6,44	5,88
Diagnosis, Differential	4,23	4,45	7,9	4,58	5,94	5,65	6,33	6,58	5,5	7,5	5,87
Amino Acid Transport Systems, Basic	4,5	5			7	8,34	4,33	4,43	7	6	5,83
Aging	3,6	7	5,18	6,25	7,25	4,61	6,07	4,94	7,42	5,87	5,82
Cell Membrane	4,14	3,33	6,33	6,56	5,93	6,25	5,57	7,09	4,17	8,57	5,79
Fabry Disease	3,6	5,64	3,73	5,9	8,44	7	6,35	3,81	6,8	5,24	5,65
Cystic Fibrosis Transmembrane Conductance Regulator	5,43	5,05	4	4,82	6,37	7	5,33	6,2	6,94	5,29	5,64
Burkholderia Infections	4	4,56	5,67	4,5	7,3	7,5	4,56	7	4,33	6,2	5,56
Antimicrobial Cationic Peptides	1	7	4	7,12	4,94	7,17	3,5	6,12	7,73	6,67	5,53
Fragile X Mental Retardation Protein	4,14	3,6	5,67	6,38	4,6	6,86	6	7	4,67	6,3	5,52
Acute Disease	3,77	6,06	5,5	5	5,29	5,43	6,62	5,9	5,62	5,5	5,47
Cystic Fibrosis	5,23	5,27	4,74	3,93	5,39	5,22	5,85	5,34	6,5	6,88	5,44
Genetic Testing	3,43	7,8	4,33	5,78	5,45	4,83	7	5,12	6,83	3,67	5,42
Algorithms	6,12	4	3,33	3,25	5,78	5,12	4,33	6,43	8,86	6,67	5,39
Gaucher Disease	5,25	4,9	6	5,6	3	6,56	5,82	5,4	5,83	5,14	5,35
Cadasil				1	6,67	3,9	4,64	5,71	6,36	7,71	5,14
Biological Transport	5,62	4,14	4,36	4,25	4,1	4,33	6,67	4,12	6,29	7,5	5,14

Continúa en la página siguiente...

Término MeSH	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Promedio
Hemochromatosis	5	5	5	4,24	6,18	5,29	4,7	4,76	5,56	4,92	5,07
Achondroplasia	5	2,62	3,33	3,4	6,5	6,17	10,17	4,5	3,56	5	5,03
Chronic Disease	3,7	5	4	4,9	6	4,75	6,1	3,91	6	5,62	5,00
Administration, Inhalation	4,8	5,94	3,88	3,63	3,75	4,67	5,36	5,47	6	5,2	4,87
Antibacterial Agents	3,96	4,26	4,36	4,2	4,12	4,89	4,72	5,67	5,35	5,87	4,74
Asthma	2,67	5,12	5,69	3,6	7,2	2,67	4,38	4,75	5,62	4,56	4,63
Disease Models, Animal	4,2	4,42	3,3	2,75	5,4	4,4	5,17	4,5	5,43	5,74	4,53
Adrenal Cortex Hormones	5	3,5	3	2,33	3	3,8	6	4,4	4,8	4,82	4,07
Clinical Trials As Topic	1,33	2,5	2,12	6	1	2,38	2,4	9,31	4,67	5,62	3,73

Tabla B.5: *Media de coautoría por términos MeSH y año (Pubmed)*

B.4. Dispersión

Revistas	Trabajos	Acum. Trabajos	Acum. Revistas	Porcentaje trabajos	Porcentaje revistas	Dist. Teórica	Acum. Dist. Teórica	Porcentaje Dist. Teórica
1	359	359	1	1,85	0,05	0	0	0
1	321	680	2	3,50	0,10	0	0	0
1	302	982	3	5,06	0,14	0	0	0
1	218	1200	4	6,18	0,19	0,01	0,01	0,01
1	214	1414	5	7,28	0,24	0,01	0,02	0,02
1	210	1624	6	8,36	0,29	0,01	0,03	0,03
1	205	1829	7	9,42	0,34	0,01	0,04	0,04
1	199	2028	8	10,45	0,38	0,01	0,05	0,05
1	195	2223	9	11,45	0,43	0,01	0,06	0,06
1	194	2417	10	12,45	0,48	0,01	0,07	0,07
1	193	2610	11	13,44	0,53	0,01	0,08	0,08
1	188	2798	12	14,41	0,58	0,01	0,09	0,09
1	177	2975	13	15,32	0,62	0,01	0,10	0,10
1	172	3147	14	16,21	0,67	0,01	0,11	0,11
1	170	3317	15	17,08	0,72	0,01	0,12	0,12
1	151	3468	16	17,86	0,77	0,02	0,14	0,14
1	148	3616	17	18,62	0,82	0,02	0,16	0,16
1	144	3760	18	19,37	0,86	0,02	0,18	0,18
1	137	3897	19	20,07	0,91	0,02	0,20	0,20
1	132	4029	20	20,75	0,96	0,02	0,22	0,22
1	125	4154	21	21,40	1,01	0,02	0,24	0,24
1	122	4276	22	22,02	1,06	0,02	0,26	0,26
1	121	4397	23	22,65	1,10	0,02	0,28	0,28
1	114	4511	24	23,23	1,15	0,03	0,31	0,31
2	108	4727	26	24,35	1,25	0,03	0,34	0,34
2	99	4925	28	25,37	1,34	0,04	0,38	0,38
1	94	5019	29	25,85	1,39	0,04	0,42	0,42
1	91	5110	30	26,32	1,44	0,04	0,46	0,46
1	89	5199	31	26,78	1,49	0,05	0,51	0,51
2	88	5375	33	27,68	1,58	0,05	0,56	0,56
1	87	5462	34	28,13	1,63	0,05	0,61	0,61
2	85	5632	36	29,01	1,73	0,05	0,66	0,66
2	84	5800	38	29,87	1,82	0,05	0,71	0,71
1	83	5883	39	30,30	1,87	0,05	0,76	0,76
2	82	6047	41	31,15	1,97	0,05	0,81	0,81
1	79	6126	42	31,55	2,02	0,06	0,87	0,87
1	78	6204	43	31,95	2,06	0,06	0,93	0,93
1	77	6281	44	32,35	2,11	0,06	0,99	0,99
1	75	6356	45	32,74	2,16	0,06	1,05	1,05
1	74	6430	46	33,12	2,21	0,07	1,12	1,12
1	70	6500	47	33,48	2,26	0,07	1,19	1,19

Continúa en la página siguiente. ...

Revistas	Trabajos	Acum. Trabajos	Acum. Revistas	Porcentaje trabajos	Porcentaje revistas	Dist. Teórica	Acum. Dist. Teórica	Porcentaje Dist. Teórica
1	69	6569	48	33,83	2,30	0,08	1,27	0,22
1	67	6636	49	34,18	2,35	0,08	1,35	0,23
1	64	6700	50	34,51	2,40	0,09	1,44	0,25
1	63	6763	51	34,83	2,45	0,09	1,53	0,26
4	62	7011	55	36,11	2,64	0,09	1,62	0,28
1	61	7072	56	36,43	2,69	0,10	1,72	0,29
2	59	7190	58	37,03	2,78	0,10	1,82	0,31
1	58	7248	59	37,33	2,83	0,11	1,93	0,33
3	57	7419	62	38,21	2,98	0,11	2,04	0,35
2	56	7531	64	38,79	3,07	0,11	2,15	0,37
2	55	7641	66	39,36	3,17	0,12	2,27	0,39
1	54	7695	67	39,63	3,22	0,12	2,39	0,41
4	53	7907	71	40,73	3,41	0,13	2,52	0,43
1	50	7957	72	40,98	3,46	0,14	2,66	0,45
4	49	8153	76	41,99	3,65	0,15	2,81	0,48
3	48	8297	79	42,73	3,79	0,16	2,97	0,51
9	47	8720	88	44,91	4,22	0,16	3,13	0,53
1	45	8765	89	45,15	4,27	0,18	3,31	0,57
1	44	8809	90	45,37	4,32	0,19	3,50	0,60
3	42	8935	93	46,02	4,46	0,20	3,70	0,63
4	41	9099	97	46,87	4,66	0,21	3,91	0,67
4	40	9259	101	47,69	4,85	0,22	4,13	0,71
4	38	9411	105	48,47	5,04	0,25	4,38	0,75
2	37	9485	107	48,85	5,14	0,26	4,64	0,79
3	36	9593	110	49,41	5,28	0,28	4,92	0,84
5	35	9768	115	50,31	5,52	0,29	5,21	0,89
2	34	9836	117	50,66	5,62	0,31	5,52	0,94
4	33	9968	121	51,34	5,81	0,33	5,85	1
6	32	10160	127	52,33	6,10	0,35	6,20	1,06
8	31	10408	135	53,61	6,48	0,37	6,57	1,12
8	30	10648	143	54,84	6,87	0,40	6,97	1,19
2	29	10706	145	55,14	6,96	0,43	7,40	1,26
6	28	10874	151	56,01	7,25	0,46	7,86	1,34
5	27	11009	156	56,70	7,49	0,49	8,35	1,43
4	26	11113	160	57,24	7,68	0,53	8,88	1,52
2	25	11163	162	57,50	7,78	0,57	9,45	1,61
5	24	11283	167	58,11	8,02	0,62	10,07	1,72
12	23	11559	179	59,54	8,59	0,68	10,75	1,84
11	22	11801	190	60,78	9,12	0,74	11,49	1,96
11	21	12032	201	61,97	9,65	0,81	12,30	2,10
8	20	12192	209	62,80	10,03	0,90	13,20	2,26

Continúa en la página siguiente...

Revistas	Trabajos	Acum. Trabajos	Acum. Revistas	Porcentaje trabajos	Porcentaje revistas	Dist. Teórica	Acum. Dist. Teórica	Porcentaje Dist. Teórica
6	19	12306	215	63,38	10,32	0,99	14,19	2,42
12	18	12522	227	64,50	10,90	1,11	15,30	2,61
13	17	12743	240	65,63	11,52	1,24	16,54	2,83
14	16	12967	254	66,79	12,19	1,40	17,94	3,06
16	15	13207	270	68,02	12,96	1,60	19,54	3,34
19	14	13473	289	69,39	13,87	1,83	21,37	3,65
21	13	13746	310	70,80	14,88	2,12	23,49	4,01
29	12	14094	339	72,59	16,27	2,49	25,98	4,44
26	11	14380	365	74,07	17,52	2,97	28,95	4,95
40	10	14780	405	76,13	19,44	3,59	32,54	5,56
44	9	15176	449	78,17	21,56	4,43	36,97	6,32
51	8	15584	500	80,27	24	5,61	42,58	7,27
81	7	16151	581	83,19	27,89	7,33	49,91	8,53
75	6	16601	656	85,51	31,49	9,97	59,88	10,23
79	5	16996	735	87,54	35,29	14,36	74,24	12,68
128	4	17508	863	90,18	41,43	22,44	96,68	16,52
183	3	18057	1046	93,01	50,22	39,89	136,57	23,33
321	2	18699	1367	96,31	65,63	89,75	226,32	38,67
716	1	19415	2083	100	100	359	585,32	100

Tabla B.6: *Distribución de Bradford en SCI*

Revistas	Trabajos	Acum. Trabajos	Acum. Revistas	Porcentaje trabajos	Porcentaje revistas	Dist. Teórica	Acum. Dist. Teórica	Porcentaje Dist. Teórica
1	424	424	1	2,06	0,04	0	0	0
1	305	729	2	3,54	0,09	0	0	0
1	286	1015	3	4,93	0,13	0,01	0,01	0,01
1	282	1297	4	6,30	0,17	0,01	0,01	0,02
1	254	1551	5	7,54	0,22	0,01	0,01	0,03
1	248	1799	6	8,74	0,26	0,01	0,01	0,04
1	247	2046	7	9,95	0,30	0,01	0,01	0,05
1	211	2257	8	10,97	0,34	0,01	0,01	0,06
1	196	2453	9	11,92	0,39	0,01	0,01	0,07
1	195	2648	10	12,87	0,43	0,01	0,01	0,08
1	188	2836	11	13,79	0,47	0,01	0,01	0,09
1	158	2994	12	14,55	0,52	0,02	0,02	0,11
1	157	3151	13	15,32	0,56	0,02	0,02	0,13
1	155	3306	14	16,07	0,60	0,02	0,02	0,15
1	145	3451	15	16,78	0,65	0,02	0,02	0,17
1	144	3595	16	17,48	0,69	0,02	0,02	0,19
1	143	3738	17	18,17	0,73	0,02	0,02	0,21
1	131	3869	18	18,81	0,77	0,02	0,02	0,23
1	127	3996	19	19,42	0,82	0,03	0,03	0,26
1	124	4120	20	20,03	0,86	0,03	0,03	0,29
1	114	4234	21	20,58	0,90	0,03	0,03	0,32
1	112	4346	22	21,13	0,95	0,03	0,03	0,35
1	110	4456	23	21,66	0,99	0,04	0,04	0,39
1	107	4563	24	22,18	1,03	0,04	0,04	0,43
1	103	4666	25	22,68	1,08	0,04	0,04	0,47
1	98	4764	26	23,16	1,12	0,04	0,04	0,51
1	97	4861	27	23,63	1,16	0,05	0,05	0,56

Continúa en la página siguiente...

Revistas	Trabajos	Acum. Trabajos	Acum. Revistas	Porcentaje trabajos	Porcentaje revistas	Dist. Teórica	Acum. Dist. Teórica	Porcentaje Dist. Teórica
1	96	4957	28	24,10	1,20	0,05	0,61	0,09
4	93	5329	32	25,90	1,38	0,05	0,66	0,10
2	91	5511	34	26,79	1,46	0,05	0,71	0,10
1	88	5599	35	27,22	1,51	0,05	0,76	0,11
2	87	5773	37	28,06	1,59	0,06	0,82	0,12
1	86	5859	38	28,48	1,63	0,06	0,88	0,13
1	83	5942	39	28,88	1,68	0,06	0,94	0,14
2	79	6100	41	29,65	1,76	0,07	1,01	0,15
1	77	6177	42	30,03	1,81	0,07	1,08	0,16
2	76	6329	44	30,77	1,89	0,07	1,15	0,17
1	72	6401	45	31,12	1,94	0,08	1,23	0,18
1	70	6471	46	31,46	1,98	0,09	1,32	0,19
1	69	6540	47	31,79	2,02	0,09	1,41	0,20
2	68	6676	49	32,45	2,11	0,09	1,50	0,22
2	67	6810	51	33,10	2,19	0,09	1,59	0,23
4	66	7074	55	34,39	2,37	0,10	1,69	0,24
1	65	7139	56	34,70	2,41	0,10	1,79	0,26
1	64	7203	57	35,01	2,45	0,10	1,89	0,27
1	63	7266	58	35,32	2,49	0,11	2	0,29
2	61	7388	60	35,91	2,58	0,11	2,11	0,30
1	60	7448	61	36,20	2,62	0,12	2,23	0,32
2	59	7566	63	36,78	2,71	0,12	2,35	0,34
1	58	7624	64	37,06	2,75	0,13	2,48	0,36
4	54	7840	68	38,11	2,92	0,15	2,63	0,38
4	53	8052	72	39,14	3,10	0,15	2,78	0,40
2	52	8156	74	39,65	3,18	0,16	2,94	0,42
2	51	8258	76	40,14	3,27	0,16	3,10	0,45

Continúa en la página siguiente...

Revistas	Trabajos	Acum. Trabajos	Acum. Revistas	Porcentaje trabajos	Porcentaje revistas	Dist. Teórica	Acum. Dist. Teórica	Porcentaje Dist. Teórica
1	50	8308	77	40,38	3,31	0,17	3,27	0,47
1	49	8357	78	40,62	3,35	0,18	3,45	0,50
1	48	8405	79	40,86	3,40	0,18	3,63	0,52
3	47	8546	82	41,54	3,53	0,19	3,82	0,55
4	46	8730	86	42,44	3,70	0,20	4,02	0,58
3	44	8862	89	43,08	3,83	0,22	4,24	0,61
3	43	8991	92	43,71	3,96	0,23	4,47	0,65
2	42	9075	94	44,11	4,04	0,24	4,71	0,68
3	41	9198	97	44,71	4,17	0,25	4,96	0,72
5	40	9398	102	45,68	4,39	0,26	5,22	0,75
2	39	9476	104	46,06	4,47	0,28	5,50	0,79
1	38	9514	105	46,25	4,52	0,29	5,79	0,84
3	37	9625	108	46,79	4,65	0,31	6,10	0,88
2	36	9697	110	47,14	4,73	0,33	6,43	0,93
1	35	9732	111	47,31	4,77	0,35	6,78	0,98
4	34	9868	115	47,97	4,95	0,37	7,15	1,03
3	33	9967	118	48,45	5,08	0,39	7,54	1,09
5	32	10127	123	49,23	5,29	0,41	7,95	1,15
5	31	10282	128	49,98	5,51	0,44	8,39	1,21
7	30	10492	135	51	5,81	0,47	8,86	1,28
6	29	10666	141	51,85	6,06	0,50	9,36	1,35
1	28	10694	142	51,98	6,11	0,54	9,90	1,43
4	27	10802	146	52,51	6,28	0,58	10,48	1,51
10	26	11062	156	53,77	6,71	0,63	11,11	1,61
7	25	11237	163	54,62	7,01	0,68	11,79	1,70
8	24	11429	171	55,56	7,35	0,74	12,53	1,81
10	23	11659	181	56,67	7,78	0,80	13,33	1,93

Continúa en la página siguiente...

Revistas	Trabajos	Acum. Trabajos	Acum. Revistas	Porcentaje trabajos	Porcentaje revistas	Dist. Teórica	Acum. Dist. Teórica	Porcentaje Dist. Teórica
14	22	11967	195	58,17	8,39	0,88	14,21	2,05
11	21	12198	206	59,29	8,86	0,96	15,17	2,19
11	20	12418	217	60,36	9,33	1,06	16,23	2,35
9	19	12589	226	61,19	9,72	1,17	17,40	2,51
13	18	12823	239	62,33	10,28	1,31	18,71	2,70
12	17	13027	251	63,32	10,80	1,47	20,18	2,92
15	16	13267	266	64,49	11,44	1,66	21,84	3,16
17	15	13522	283	65,73	12,17	1,88	23,72	3,43
20	14	13802	303	67,09	13,03	2,16	25,88	3,74
18	13	14036	321	68,23	13,81	2,51	28,39	4,10
36	12	14468	357	70,33	15,35	2,94	31,33	4,53
31	11	14809	388	71,99	16,69	3,50	34,83	5,03
44	10	15249	432	74,13	18,58	4,24	39,07	5,65
50	9	15699	482	76,31	20,73	5,23	44,30	6,40
47	8	16075	529	78,14	22,75	6,62	50,92	7,36
83	7	16656	612	80,96	26,32	8,65	59,57	8,61
108	6	17304	720	84,11	30,97	11,78	71,35	10,31
110	5	17854	830	86,79	35,70	16,96	88,31	12,76
134	4	18390	964	89,39	41,46	26,50	114,81	16,59
238	3	19104	1202	92,86	51,70	47,11	161,92	23,40
345	2	19794	1547	96,22	66,54	106	267,92	38,72
778	1	20572	2325	100	100	424	691,92	100

Tabla B.7: Distribución de Bradford en Pubmed

Revistas	Trabajos	Acum. Trabajos	Acum. Revistas	Porcentaje trabajos	Porcentaje revistas	Dist. Teórica	Acum. Dist. Teórica	Porcentaje Dist. Teórica
1	23013	23013	1	3,40	0	0	0	0
1	16396	39409	2	5,83	0,01	0	0	0
1	15209	54618	3	8,08	0,01	0	0	0
1	13541	68159	4	10,08	0,02	0	0	0
1	12029	80188	5	11,86	0,02	0	0	0
1	10276	90464	6	13,38	0,02	0	0	0
1	9187	99651	7	14,74	0,03	0	0	0
1	9139	108790	8	16,09	0,03	0	0	0
1	8503	117293	9	17,35	0,04	0	0	0
1	7857	125150	10	18,51	0,04	0	0	0
1	7788	132938	11	19,66	0,04	0	0	0
1	7568	140506	12	20,78	0,05	0	0	0
1	7154	147660	13	21,84	0,05	0	0	0
1	6461	154121	14	22,79	0,06	0	0	0
1	6347	160468	15	23,73	0,06	0	0	0
1	6199	166667	16	24,65	0,06	0	0	0
1	5674	172341	17	25,49	0,07	0	0	0
1	5193	177534	18	26,26	0,07	0	0	0
1	5157	182691	19	27,02	0,08	0	0	0
1	5052	187743	20	27,77	0,08	0	0	0
1	5022	192765	21	28,51	0,09	0	0	0
1	4657	197422	22	29,20	0,09	0	0	0
1	4381	201803	23	29,85	0,09	0	0	0
1	4343	206146	24	30,49	0,10	0	0	0
1	4259	210405	25	31,12	0,10	0	0	0
1	4109	214514	26	31,73	0,11	0	0	0
1	4025	218539	27	32,32	0,11	0	0	0

Continúa en la página siguiente...

Revistas	Trabajos	Acum. Trabajos	Acum. Revistas	Porcentaje trabajos	Porcentaje revistas	Dist. Teórica	Acum. Dist. Teórica	Porcentaje Dist. Teórica
1	3777	222316	28	32,88	0,11	0	0	0
1	3680	225996	29	33,42	0,12	0	0	0
1	3503	229499	30	33,94	0,12	0	0	0
1	3483	232982	31	34,46	0,13	0	0	0
1	3373	236355	32	34,96	0,13	0	0	0
1	3370	239725	33	35,45	0,13	0	0	0
1	3227	242952	34	35,93	0,14	0	0	0
1	3143	246095	35	36,40	0,14	0	0	0
1	3078	249173	36	36,85	0,15	0	0	0
1	3037	252210	37	37,30	0,15	0	0	0
1	2873	255083	38	37,73	0,15	0	0	0
1	2867	257950	39	38,15	0,16	0	0	0
1	2839	260789	40	38,57	0,16	0	0	0
1	2819	263608	41	38,99	0,17	0	0	0
1	2780	266388	42	39,40	0,17	0	0	0
1	2709	269097	43	39,80	0,17	0	0	0
1	2694	271791	44	40,20	0,18	0	0	0
1	2680	274471	45	40,59	0,18	0	0	0
1	2661	277132	46	40,99	0,19	0	0	0
1	2579	279711	47	41,37	0,19	0	0	0
1	2543	282254	48	41,74	0,19	0	0	0
1	2500	284754	49	42,11	0,20	0	0	0
1	2460	287214	50	42,48	0,20	0	0	0
1	2402	289616	51	42,83	0,21	0	0	0
1	2327	291943	52	43,18	0,21	0	0	0
1	2275	294218	53	43,51	0,21	0	0	0
1	2208	296426	54	43,84	0,22	0	0	0

Continúa en la página siguiente. . .

Revistas	Trabajos	Acum. Trabajos	Acum. Revistas	Porcentaje trabajos	Porcentaje revistas	Dist. Teórica	Acum. Dist. Teórica	Porcentaje Dist. Teórica
1	2204	298630	55	44,17	0,22	0	0	0
1	2051	300681	56	44,47	0,23	0,01	0,01	0,01
1	2047	302728	57	44,77	0,23	0,01	0,02	0,02
1	1951	304679	58	45,06	0,23	0,01	0,03	0,03
1	1891	306570	59	45,34	0,24	0,01	0,04	0,04
1	1887	308457	60	45,62	0,24	0,01	0,05	0,05
1	1875	310332	61	45,90	0,25	0,01	0,06	0,06
1	1870	312202	62	46,17	0,25	0,01	0,07	0,07
1	1831	314033	63	46,44	0,26	0,01	0,08	0,08
1	1822	315855	64	46,71	0,26	0,01	0,09	0,09
1	1818	317673	65	46,98	0,26	0,01	0,10	0,10
1	1795	319468	66	47,25	0,27	0,01	0,11	0,11
1	1790	321258	67	47,51	0,27	0,01	0,12	0,12
1	1787	323045	68	47,78	0,28	0,01	0,13	0,13
1	1780	324825	69	48,04	0,28	0,01	0,14	0,14
1	1772	326597	70	48,30	0,28	0,01	0,15	0,15
1	1765	328362	71	48,56	0,29	0,01	0,16	0,16
1	1763	330125	72	48,82	0,29	0,01	0,17	0,17
1	1759	331884	73	49,08	0,30	0,01	0,18	0,18
1	1720	333604	74	49,34	0,30	0,01	0,19	0,19
1	1717	335321	75	49,59	0,30	0,01	0,20	0,20
1	1655	336976	76	49,84	0,31	0,01	0,21	0,21
1	1626	338602	77	50,08	0,31	0,01	0,22	0,22
1	1616	340218	78	50,32	0,32	0,01	0,23	0,23
1	1615	341833	79	50,56	0,32	0,01	0,24	0,24
1	1603	343436	80	50,79	0,32	0,01	0,25	0,25
1	1580	345016	81	51,03	0,33	0,01	0,26	0,26

Continúa en la página siguiente...

Revistas	Trabajos	Acum. Trabajos	Acum. Revistas	Porcentaje trabajos	Porcentaje revistas	Dist. Teórica	Acum. Dist. Teórica	Porcentaje Dist. Teórica
1	1549	346565	82	51,26	0,33	0,01	0,27	0
1	1546	348111	83	51,48	0,34	0,01	0,28	0
1	1510	349621	84	51,71	0,34	0,01	0,29	0
1	1508	351129	85	51,93	0,34	0,01	0,30	0
1	1497	352626	86	52,15	0,35	0,01	0,31	0
1	1490	354116	87	52,37	0,35	0,01	0,32	0
1	1383	355499	88	52,58	0,36	0,01	0,33	0
1	1378	356877	89	52,78	0,36	0,01	0,34	0
1	1358	358235	90	52,98	0,36	0,01	0,35	0
1	1340	359575	91	53,18	0,37	0,01	0,36	0
1	1328	360903	92	53,38	0,37	0,01	0,37	0
1	1306	362209	93	53,57	0,38	0,01	0,38	0
1	1282	363491	94	53,76	0,38	0,01	0,39	0
1	1212	364703	95	53,94	0,38	0,02	0,41	0
1	1211	365914	96	54,12	0,39	0,02	0,43	0
1	1183	367097	97	54,29	0,39	0,02	0,45	0
2	1180	369457	99	54,64	0,40	0,02	0,47	0
1	1167	370624	100	54,81	0,40	0,02	0,49	0
1	1125	371749	101	54,98	0,41	0,02	0,51	0
1	1096	372845	102	55,14	0,41	0,02	0,53	0
1	1071	373916	103	55,30	0,42	0,02	0,55	0
1	1066	374982	104	55,46	0,42	0,02	0,57	0
1	1064	376046	105	55,62	0,43	0,02	0,59	0
1	1049	377095	106	55,77	0,43	0,02	0,61	0
1	1041	378136	107	55,93	0,43	0,02	0,63	0
1	1039	379175	108	56,08	0,44	0,02	0,65	0
1	1033	380208	109	56,23	0,44	0,02	0,67	0

Continúa en la página siguiente. . .

Revistas	Trabajos	Acum. Trabajos	Acum. Revistas	Porcentaje trabajos	Porcentaje revistas	Dist. Teórica	Acum. Dist. Teórica	Porcentaje Dist. Teórica
2	1028	382264	111	56,54	0,45	0,02	0,69	0
1	1023	383287	112	56,69	0,45	0,02	0,71	0
1	1020	384307	113	56,84	0,46	0,02	0,73	0
1	1018	385325	114	56,99	0,46	0,02	0,75	0
2	999	387323	116	57,28	0,47	0,02	0,77	0
1	963	388286	117	57,43	0,47	0,02	0,79	0
1	962	389248	118	57,57	0,48	0,02	0,81	0
1	961	390209	119	57,71	0,48	0,02	0,83	0
1	948	391157	120	57,85	0,49	0,03	0,86	0
1	944	392101	121	57,99	0,49	0,03	0,89	0
1	939	393040	122	58,13	0,49	0,03	0,92	0
1	936	393976	123	58,27	0,50	0,03	0,95	0
1	917	394893	124	58,40	0,50	0,03	0,98	0
1	909	395802	125	58,54	0,51	0,03	1,01	0
1	903	396705	126	58,67	0,51	0,03	1,04	0
1	901	397606	127	58,80	0,51	0,03	1,07	0
1	886	398492	128	58,94	0,52	0,03	1,10	0
1	875	399367	129	59,07	0,52	0,03	1,13	0
1	871	400238	130	59,19	0,53	0,03	1,16	0
1	869	401107	131	59,32	0,53	0,03	1,19	0
1	853	401960	132	59,45	0,53	0,03	1,22	0
1	848	402808	133	59,57	0,54	0,03	1,25	0
1	844	403652	134	59,70	0,54	0,03	1,28	0
1	826	404478	135	59,82	0,55	0,03	1,31	0
1	824	405302	136	59,94	0,55	0,03	1,34	0
1	823	406125	137	60,06	0,55	0,03	1,37	0
1	822	406947	138	60,19	0,56	0,03	1,40	0

Continúa en la página siguiente...

Revistas	Trabajos	Acum. Trabajos	Acum. Revistas	Porcentaje trabajos	Porcentaje revistas	Dist. Teórica	Acum. Dist. Teórica	Porcentaje Dist. Teórica
1	814	407761	139	60,31	0,56	0,03	1,43	0
1	794	408555	140	60,42	0,57	0,04	1,47	0
1	792	409347	141	60,54	0,57	0,04	1,51	0
1	777	410124	142	60,66	0,58	0,04	1,55	0
1	754	410878	143	60,77	0,58	0,04	1,59	0
1	749	411627	144	60,88	0,58	0,04	1,63	0
1	748	412375	145	60,99	0,59	0,04	1,67	0
1	743	413118	146	61,10	0,59	0,04	1,71	0
1	740	413858	147	61,21	0,60	0,04	1,75	0
1	735	414593	148	61,32	0,60	0,04	1,79	0
1	725	415318	149	61,42	0,60	0,04	1,83	0
1	721	416039	150	61,53	0,61	0,04	1,87	0
2	701	417441	152	61,74	0,62	0,05	1,92	0,01
1	698	418139	153	61,84	0,62	0,05	1,97	0,01
1	697	418836	154	61,94	0,62	0,05	2,02	0,01
1	694	419530	155	62,05	0,63	0,05	2,07	0,01
2	692	420914	157	62,25	0,64	0,05	2,12	0,01
1	691	421605	158	62,35	0,64	0,05	2,17	0,01
1	687	422292	159	62,46	0,64	0,05	2,22	0,01
1	686	422978	160	62,56	0,65	0,05	2,27	0,01
1	682	423660	161	62,66	0,65	0,05	2,32	0,01
1	678	424338	162	62,76	0,66	0,05	2,37	0,01
1	676	425014	163	62,86	0,66	0,05	2,42	0,01
1	675	425689	164	62,96	0,66	0,05	2,47	0,01
1	666	426355	165	63,06	0,67	0,05	2,52	0,01
1	665	427020	166	63,15	0,67	0,05	2,57	0,01
1	664	427684	167	63,25	0,68	0,05	2,62	0,01

Continúa en la página siguiente...

Revistas	Trabajos	Acum. Trabajos	Acum. Revistas	Porcentaje trabajos	Porcentaje revistas	Dist. Teórica	Acum. Dist. Teórica	Porcentaje Dist. Teórica
1	653	428337	168	63,35	0,68	0,05	2,67	0,01
1	626	428963	169	63,44	0,68	0,06	2,73	0,01
1	614	429577	170	63,53	0,69	0,06	2,79	0,01
1	612	430189	171	63,62	0,69	0,06	2,85	0,01
1	606	430795	172	63,71	0,70	0,06	2,91	0,01
2	604	432003	174	63,89	0,70	0,06	2,97	0,01
1	599	432602	175	63,98	0,71	0,06	3,03	0,01
1	583	433185	176	64,07	0,71	0,07	3,10	0,01
1	582	433767	177	64,15	0,72	0,07	3,17	0,01
2	580	434927	179	64,32	0,72	0,07	3,24	0,01
1	573	435500	180	64,41	0,73	0,07	3,31	0,01
1	570	436070	181	64,49	0,73	0,07	3,38	0,01
2	569	437208	183	64,66	0,74	0,07	3,45	0,01
1	561	437769	184	64,74	0,75	0,07	3,52	0,01
1	557	438326	185	64,83	0,75	0,07	3,59	0,01
3	556	439994	188	65,07	0,76	0,07	3,66	0,01
1	544	440538	189	65,15	0,77	0,08	3,74	0,01
1	537	441075	190	65,23	0,77	0,08	3,82	0,01
2	534	442143	192	65,39	0,78	0,08	3,90	0,01
3	523	443712	195	65,62	0,79	0,08	3,98	0,01
1	522	444234	196	65,70	0,79	0,08	4,06	0,01
1	519	444753	197	65,78	0,80	0,09	4,15	0,01
1	518	445271	198	65,85	0,80	0,09	4,24	0,01
1	517	445788	199	65,93	0,81	0,09	4,33	0,01
1	513	446301	200	66,01	0,81	0,09	4,42	0,01
1	508	446809	201	66,08	0,81	0,09	4,51	0,01
1	505	447314	202	66,16	0,82	0,09	4,60	0,01

Continúa en la página siguiente...

Revistas	Trabajos	Acum. Trabajos	Acum. Revistas	Porcentaje trabajos	Porcentaje revistas	Dist. Teórica	Acum. Dist. Teórica	Porcentaje Dist. Teórica
1	503	447817	203	66,23	0,82	0,09	4,69	0,01
1	502	448319	204	66,30	0,83	0,09	4,78	0,01
1	501	448820	205	66,38	0,83	0,09	4,87	0,01
1	497	449317	206	66,45	0,83	0,09	4,96	0,01
1	490	449807	207	66,53	0,84	0,10	5,06	0,01
1	488	450295	208	66,60	0,84	0,10	5,16	0,01
1	485	450780	209	66,67	0,85	0,10	5,26	0,01
1	484	451264	210	66,74	0,85	0,10	5,36	0,01
2	481	452226	212	66,88	0,86	0,10	5,46	0,01
1	480	452706	213	66,95	0,86	0,10	5,56	0,01
1	479	453185	214	67,02	0,87	0,10	5,66	0,01
1	470	453655	215	67,09	0,87	0,10	5,76	0,02
1	465	454120	216	67,16	0,87	0,11	5,87	0,02
1	462	454582	217	67,23	0,88	0,11	5,98	0,02
2	460	455502	219	67,37	0,89	0,11	6,09	0,02
1	452	455954	220	67,43	0,89	0,11	6,20	0,02
1	451	456405	221	67,50	0,89	0,11	6,31	0,02
1	450	456855	222	67,57	0,90	0,11	6,42	0,02
1	448	457303	223	67,63	0,90	0,11	6,53	0,02
1	442	457745	224	67,70	0,91	0,12	6,65	0,02
2	441	458627	226	67,83	0,92	0,12	6,77	0,02
1	439	459066	227	67,89	0,92	0,12	6,89	0,02
1	438	459504	228	67,96	0,92	0,12	7,01	0,02
1	437	459941	229	68,02	0,93	0,12	7,13	0,02
1	436	460377	230	68,09	0,93	0,12	7,25	0,02
1	434	460811	231	68,15	0,94	0,12	7,37	0,02
1	430	461241	232	68,22	0,94	0,12	7,49	0,02

Continúa en la página siguiente. . .

Revistas	Trabajos	Acum. Trabajos	Acum. Revistas	Porcentaje trabajos	Porcentaje revistas	Dist. Teórica	Acum. Dist. Teórica	Porcentaje Dist. Teórica
1	424	461665	233	68,28	0,94	0,13	7,62	0,02
2	422	462509	235	68,40	0,95	0,13	7,75	0,02
1	419	462928	236	68,47	0,96	0,13	7,88	0,02
1	418	463346	237	68,53	0,96	0,13	8,01	0,02
1	416	463762	238	68,59	0,96	0,13	8,14	0,02
1	415	464177	239	68,65	0,97	0,13	8,27	0,02
1	414	464591	240	68,71	0,97	0,13	8,40	0,02
2	412	465415	242	68,83	0,98	0,14	8,54	0,02
1	411	465826	243	68,89	0,98	0,14	8,68	0,02
1	410	466236	244	68,95	0,99	0,14	8,82	0,02
1	409	466645	245	69,02	0,99	0,14	8,96	0,02
1	403	467048	246	69,07	1	0,14	9,10	0,02
1	400	467448	247	69,13	1	0,14	9,24	0,02
2	398	468244	249	69,25	1,01	0,15	9,39	0,02
1	397	468641	250	69,31	1,01	0,15	9,54	0,03
2	396	469433	252	69,43	1,02	0,15	9,69	0,03
1	394	469827	253	69,49	1,02	0,15	9,84	0,03
1	393	470220	254	69,54	1,03	0,15	9,99	0,03
2	391	471002	256	69,66	1,04	0,15	10,14	0,03
1	388	471390	257	69,72	1,04	0,15	10,29	0,03
1	387	471777	258	69,77	1,04	0,15	10,44	0,03
1	386	472163	259	69,83	1,05	0,15	10,59	0,03
1	384	472547	260	69,89	1,05	0,16	10,75	0,03
1	383	472930	261	69,94	1,06	0,16	10,91	0,03
1	382	473312	262	70	1,06	0,16	11,07	0,03
2	377	474066	264	70,11	1,07	0,16	11,23	0,03
1	376	474442	265	70,17	1,07	0,16	11,39	0,03

Continúa en la página siguiente...

Revistas	Trabajos	Acum. Trabajos	Acum. Revistas	Porcentaje trabajos	Porcentaje revistas	Dist. Teórica	Acum. Dist. Teórica	Porcentaje Dist. Teórica
1	375	474817	266	70,22	1,08	0,16	11,55	0,03
1	373	475190	267	70,28	1,08	0,17	11,72	0,03
1	366	475556	268	70,33	1,09	0,17	11,89	0,03
1	365	475921	269	70,39	1,09	0,17	12,06	0,03
1	360	476281	270	70,44	1,09	0,18	12,24	0,03
1	358	476639	271	70,49	1,10	0,18	12,42	0,03
1	355	476994	272	70,55	1,10	0,18	12,60	0,03
1	353	477347	273	70,60	1,11	0,18	12,78	0,03
2	352	478051	275	70,70	1,11	0,19	12,97	0,03
1	351	478402	276	70,75	1,12	0,19	13,16	0,03
1	349	478751	277	70,81	1,12	0,19	13,35	0,04
1	347	479098	278	70,86	1,13	0,19	13,54	0,04
1	343	479441	279	70,91	1,13	0,20	13,74	0,04
2	342	480125	281	71,01	1,14	0,20	13,94	0,04
1	341	480466	282	71,06	1,14	0,20	14,14	0,04
3	339	481483	285	71,21	1,15	0,20	14,34	0,04
1	338	481821	286	71,26	1,16	0,20	14,54	0,04
1	332	482153	287	71,31	1,16	0,21	14,75	0,04
1	330	482483	288	71,36	1,17	0,21	14,96	0,04
1	329	482812	289	71,41	1,17	0,21	15,17	0,04
1	324	483136	290	71,45	1,17	0,22	15,39	0,04
4	323	484428	294	71,65	1,19	0,22	15,61	0,04
1	322	484750	295	71,69	1,19	0,22	15,83	0,04
1	319	485069	296	71,74	1,20	0,23	16,06	0,04
2	317	485703	298	71,83	1,21	0,23	16,29	0,04
3	316	486651	301	71,97	1,22	0,23	16,52	0,04
1	315	486966	302	72,02	1,22	0,23	16,75	0,04

Continúa en la página siguiente...

Revistas	Trabajos	Acum. Trabajos	Acum. Revistas	Porcentaje trabajos	Porcentaje revistas	Dist. Teórica	Acum. Dist. Teórica	Porcentaje Dist. Teórica
1	314	487280	303	72,07	1,23	0,23	16,98	0,04
3	312	488216	306	72,21	1,24	0,24	17,22	0,05
2	311	488838	308	72,30	1,25	0,24	17,46	0,05
1	310	489148	309	72,34	1,25	0,24	17,70	0,05
1	308	489456	310	72,39	1,26	0,24	17,94	0,05
1	307	489763	311	72,43	1,26	0,24	18,18	0,05
1	305	490068	312	72,48	1,26	0,25	18,43	0,05
1	302	490370	313	72,52	1,27	0,25	18,68	0,05
1	297	490667	314	72,57	1,27	0,26	18,94	0,05
1	294	490961	315	72,61	1,28	0,27	19,21	0,05
1	293	491254	316	72,65	1,28	0,27	19,48	0,05
1	292	491546	317	72,70	1,28	0,27	19,75	0,05
1	291	491837	318	72,74	1,29	0,27	20,02	0,05
2	290	492417	320	72,83	1,30	0,27	20,29	0,05
1	287	492704	321	72,87	1,30	0,28	20,57	0,05
1	286	492990	322	72,91	1,30	0,28	20,85	0,06
1	281	493271	323	72,95	1,31	0,29	21,14	0,06
2	280	493831	325	73,04	1,32	0,29	21,43	0,06
1	278	494109	326	73,08	1,32	0,30	21,73	0,06
2	277	494663	328	73,16	1,33	0,30	22,03	0,06
1	276	494939	329	73,20	1,33	0,30	22,33	0,06
1	274	495213	330	73,24	1,34	0,31	22,64	0,06
1	273	495486	331	73,28	1,34	0,31	22,95	0,06
1	272	495758	332	73,32	1,34	0,31	23,26	0,06
1	271	496029	333	73,36	1,35	0,31	23,57	0,06
3	270	496839	336	73,48	1,36	0,32	23,89	0,06
1	269	497108	337	73,52	1,36	0,32	24,21	0,06

Continúa en la página siguiente...

Revistas	Trabajos	Acum. Trabajos	Acum. Revistas	Porcentaje trabajos	Porcentaje revistas	Dist. Teórica	Acum. Dist. Teórica	Porcentaje Dist. Teórica
1	268	497376	338	73,56	1,37	0,32	24,53	0,06
4	267	498444	342	73,72	1,39	0,32	24,85	0,07
1	266	498710	343	73,76	1,39	0,33	25,18	0,07
1	265	498975	344	73,80	1,39	0,33	25,51	0,07
4	264	500031	348	73,95	1,41	0,33	25,84	0,07
1	263	500294	349	73,99	1,41	0,33	26,17	0,07
1	261	500555	350	74,03	1,42	0,34	26,51	0,07
3	259	501332	353	74,15	1,43	0,34	26,85	0,07
1	258	501590	354	74,18	1,43	0,35	27,20	0,07
1	257	501847	355	74,22	1,44	0,35	27,55	0,07
1	255	502102	356	74,26	1,44	0,35	27,90	0,07
1	254	502356	357	74,30	1,45	0,36	28,26	0,07
1	253	502609	358	74,33	1,45	0,36	28,62	0,08
2	252	503113	360	74,41	1,46	0,36	28,98	0,08
1	250	503363	361	74,45	1,46	0,37	29,35	0,08
1	249	503612	362	74,48	1,47	0,37	29,72	0,08
1	248	503860	363	74,52	1,47	0,37	30,09	0,08
1	247	504107	364	74,56	1,47	0,38	30,47	0,08
2	246	504599	366	74,63	1,48	0,38	30,85	0,08
3	245	505334	369	74,74	1,49	0,38	31,23	0,08
1	244	505578	370	74,77	1,50	0,39	31,62	0,08
4	243	506550	374	74,92	1,51	0,39	32,01	0,08
1	242	506792	375	74,95	1,52	0,39	32,40	0,09
2	240	507272	377	75,02	1,53	0,40	32,80	0,09
1	239	507511	378	75,06	1,53	0,40	33,20	0,09
1	238	507749	379	75,09	1,53	0,41	33,61	0,09
2	236	508221	381	75,16	1,54	0,41	34,02	0,09

Continúa en la página siguiente. . .

Revistas	Trabajos	Acum. Trabajos	Acum. Revistas	Porcentaje trabajos	Porcentaje revistas	Dist. Teórica	Acum. Dist. Teórica	Porcentaje Dist. Teórica
2	235	508691	383	75,23	1,55	0,42	34,44	0,09
1	233	508924	384	75,27	1,56	0,42	34,86	0,09
1	232	509156	385	75,30	1,56	0,43	35,29	0,09
1	231	509387	386	75,34	1,56	0,43	35,72	0,09
1	230	509617	387	75,37	1,57	0,44	36,16	0,10
3	229	510304	390	75,47	1,58	0,44	36,60	0,10
2	228	510760	392	75,54	1,59	0,44	37,04	0,10
1	227	510987	393	75,57	1,59	0,45	37,49	0,10
2	225	511437	395	75,64	1,60	0,45	37,94	0,10
1	223	511660	396	75,67	1,60	0,46	38,40	0,10
2	221	512102	398	75,74	1,61	0,47	38,87	0,10
3	220	512762	401	75,84	1,62	0,48	39,35	0,10
1	219	512981	402	75,87	1,63	0,48	39,83	0,11
5	218	514071	407	76,03	1,65	0,48	40,31	0,11
3	217	514722	410	76,13	1,66	0,49	40,80	0,11
2	215	515152	412	76,19	1,67	0,50	41,30	0,11
2	213	515578	414	76,25	1,68	0,51	41,81	0,11
5	212	516638	419	76,41	1,70	0,51	42,32	0,11
1	210	516848	420	76,44	1,70	0,52	42,84	0,11
2	209	517266	422	76,50	1,71	0,53	43,37	0,11
3	208	517890	425	76,59	1,72	0,53	43,90	0,12
1	207	518097	426	76,62	1,73	0,54	44,44	0,12
1	206	518303	427	76,66	1,73	0,54	44,98	0,12
3	205	518918	430	76,75	1,74	0,55	45,53	0,12
5	204	519938	435	76,90	1,76	0,55	46,08	0,12
1	202	520140	436	76,93	1,77	0,56	46,64	0,12
1	201	520341	437	76,96	1,77	0,57	47,21	0,12

Continúa en la página siguiente...

Revistas	Trabajos	Acum. Trabajos	Acum. Revistas	Porcentaje trabajos	Porcentaje revistas	Dist. Teórica	Acum. Dist. Teórica	Porcentaje Dist. Teórica
1	200	520541	438	76,99	1,77	0,58	47,79	0,13
1	199	520740	439	77,02	1,78	0,58	48,37	0,13
1	198	520938	440	77,05	1,78	0,59	48,96	0,13
1	197	521135	441	77,07	1,79	0,59	49,55	0,13
3	196	521723	444	77,16	1,80	0,60	50,15	0,13
2	195	522113	446	77,22	1,81	0,61	50,76	0,13
1	193	522306	447	77,25	1,81	0,62	51,38	0,14
1	192	522498	448	77,28	1,81	0,62	52	0,14
2	191	522880	450	77,33	1,82	0,63	52,63	0,14
2	190	523260	452	77,39	1,83	0,64	53,27	0,14
1	189	523449	453	77,42	1,83	0,64	53,91	0,14
1	188	523637	454	77,44	1,84	0,65	54,56	0,14
2	187	524011	456	77,50	1,85	0,66	55,22	0,15
1	186	524197	457	77,53	1,85	0,67	55,89	0,15
1	184	524381	458	77,55	1,85	0,68	56,57	0,15
3	183	524930	461	77,64	1,87	0,69	57,26	0,15
2	182	525294	463	77,69	1,88	0,69	57,95	0,15
1	181	525475	464	77,72	1,88	0,70	58,65	0,16
2	180	525835	466	77,77	1,89	0,71	59,36	0,16
4	179	526551	470	77,88	1,90	0,72	60,08	0,16
4	178	527263	474	77,98	1,92	0,73	60,81	0,16
2	177	527617	476	78,03	1,93	0,73	61,54	0,16
1	176	527793	477	78,06	1,93	0,74	62,28	0,16
1	174	527967	478	78,08	1,94	0,76	63,04	0,17
1	173	528140	479	78,11	1,94	0,77	63,81	0,17
2	172	528484	481	78,16	1,95	0,78	64,59	0,17
2	171	528826	483	78,21	1,96	0,79	65,38	0,17

Continúa en la página siguiente...

Revistas	Trabajos	Acum. Trabajos	Acum. Revistas	Porcentaje trabajos	Porcentaje revistas	Dist. Teórica	Acum. Dist. Teórica	Porcentaje Dist. Teórica
3	170	529336	486	78,29	1,97	0,80	66,18	0,18
4	169	530012	490	78,39	1,98	0,81	66,99	0,18
2	168	530348	492	78,44	1,99	0,82	67,81	0,18
1	167	530515	493	78,46	2	0,83	68,64	0,18
3	165	531010	496	78,53	2,01	0,85	69,49	0,18
3	164	531502	499	78,61	2,02	0,86	70,35	0,19
5	163	532317	504	78,73	2,04	0,87	71,22	0,19
4	162	532965	508	78,82	2,06	0,88	72,10	0,19
1	161	533126	509	78,85	2,06	0,89	72,99	0,19
2	160	533446	511	78,89	2,07	0,90	73,89	0,20
2	158	533762	513	78,94	2,08	0,92	74,81	0,20
3	157	534233	516	79,01	2,09	0,93	75,74	0,20
3	156	534701	519	79,08	2,10	0,95	76,69	0,20
4	155	535321	523	79,17	2,12	0,96	77,65	0,21
3	154	535783	526	79,24	2,13	0,97	78,62	0,21
3	153	536242	529	79,31	2,14	0,98	79,60	0,21
2	152	536546	531	79,35	2,15	1	80,60	0,21
6	151	537452	537	79,49	2,17	1,01	81,61	0,22
5	150	538202	542	79,60	2,19	1,02	82,63	0,22
3	149	538649	545	79,66	2,21	1,04	83,67	0,22
4	148	539241	549	79,75	2,22	1,05	84,72	0,22
3	147	539682	552	79,82	2,24	1,06	85,78	0,23
1	146	539828	553	79,84	2,24	1,08	86,86	0,23
3	145	540263	556	79,90	2,25	1,09	87,95	0,23
5	144	540983	561	80,01	2,27	1,11	89,06	0,24
4	143	541555	565	80,09	2,29	1,13	90,19	0,24
6	142	542407	571	80,22	2,31	1,14	91,33	0,24

Continúa en la página siguiente...

Revistas	Trabajos	Acum. Trabajos	Acum. Revistas	Porcentaje trabajos	Porcentaje revistas	Dist. Teórica	Acum. Dist. Teórica	Porcentaje Dist. Teórica
6	141	543253	577	80,35	2,34	1,16	92,49	0,24
3	139	543670	580	80,41	2,35	1,19	93,68	0,25
2	138	543946	582	80,45	2,36	1,21	94,89	0,25
5	137	544631	587	80,55	2,38	1,23	96,12	0,25
3	136	545039	590	80,61	2,39	1,24	97,36	0,26
6	135	545849	596	80,73	2,41	1,26	98,62	0,26
3	134	546251	599	80,79	2,43	1,28	99,90	0,26
1	133	546384	600	80,81	2,43	1,30	101,20	0,27
3	132	546780	603	80,87	2,44	1,32	102,52	0,27
4	131	547304	607	80,94	2,46	1,34	103,86	0,27
6	130	548084	613	81,06	2,48	1,36	105,22	0,28
1	129	548213	614	81,08	2,49	1,38	106,60	0,28
10	128	549493	624	81,27	2,53	1,40	108	0,29
3	127	549874	627	81,32	2,54	1,43	109,43	0,29
4	126	550378	631	81,40	2,56	1,45	110,88	0,29
7	125	551253	638	81,53	2,58	1,47	112,35	0,30
2	124	551501	640	81,57	2,59	1,50	113,85	0,30
5	123	552116	645	81,66	2,61	1,52	115,37	0,31
4	122	552604	649	81,73	2,63	1,55	116,92	0,31
7	121	553451	656	81,85	2,66	1,57	118,49	0,31
5	120	554051	661	81,94	2,68	1,60	120,09	0,32
5	119	554646	666	82,03	2,70	1,63	121,72	0,32
2	118	554882	668	82,07	2,71	1,65	123,37	0,33
4	117	555350	672	82,13	2,72	1,68	125,05	0,33
8	116	556278	680	82,27	2,75	1,71	126,76	0,34
5	115	556853	685	82,36	2,77	1,74	128,50	0,34
5	114	557423	690	82,44	2,79	1,77	130,27	0,34

Continúa en la página siguiente...

Revistas	Trabajos	Acum. Trabajos	Acum. Revistas	Porcentaje trabajos	Porcentaje revistas	Dist. Teórica	Acum. Dist. Teórica	Porcentaje Dist. Teórica
4	113	557875	694	82,51	2,81	1,80	132,07	0,35
3	112	558211	697	82,56	2,82	1,83	133,90	0,35
6	111	558877	703	82,66	2,85	1,87	135,77	0,36
6	110	559537	709	82,75	2,87	1,90	137,67	0,36
3	109	559864	712	82,80	2,88	1,94	139,61	0,37
6	108	560512	718	82,90	2,91	1,97	141,58	0,37
6	107	561154	724	82,99	2,93	2,01	143,59	0,38
3	106	561472	727	83,04	2,94	2,05	145,64	0,39
5	105	561997	732	83,12	2,96	2,09	147,73	0,39
6	104	562621	738	83,21	2,99	2,13	149,86	0,40
3	103	562930	741	83,26	3	2,17	152,03	0,40
7	102	563644	748	83,36	3,03	2,21	154,24	0,41
4	101	564048	752	83,42	3,05	2,26	156,50	0,41
7	100	564748	759	83,52	3,07	2,30	158,80	0,42
5	99	565243	764	83,60	3,09	2,35	161,15	0,43
6	98	565831	770	83,68	3,12	2,40	163,55	0,43
3	97	566122	773	83,73	3,13	2,45	166	0,44
2	96	566314	775	83,76	3,14	2,50	168,50	0,45
7	95	566979	782	83,85	3,17	2,55	171,05	0,45
5	94	567449	787	83,92	3,19	2,60	173,65	0,46
8	93	568193	795	84,03	3,22	2,66	176,31	0,47
4	92	568561	799	84,09	3,24	2,72	179,03	0,47
5	91	569016	804	84,16	3,26	2,78	181,81	0,48
5	90	569466	809	84,22	3,28	2,84	184,65	0,49
8	89	570178	817	84,33	3,31	2,91	187,56	0,50
11	88	571146	828	84,47	3,35	2,97	190,53	0,50
2	87	571320	830	84,50	3,36	3,04	193,57	0,51

Continúa en la página siguiente...

Revistas	Trabajos	Acum. Trabajos	Acum. Revistas	Porcentaje trabajos	Porcentaje revistas	Dist. Teórica	Acum. Dist. Teórica	Porcentaje Dist. Teórica
10	86	572180	840	84,62	3,40	3,11	196,68	0,52
3	85	572435	843	84,66	3,41	3,19	199,87	0,53
9	84	573191	852	84,77	3,45	3,26	203,13	0,54
9	83	573938	861	84,88	3,49	3,34	206,47	0,55
3	82	574184	864	84,92	3,50	3,42	209,89	0,56
7	81	574751	871	85	3,53	3,51	213,40	0,56
12	80	575711	883	85,15	3,58	3,60	217	0,57
7	79	576264	890	85,23	3,60	3,69	220,69	0,58
11	78	577122	901	85,35	3,65	3,78	224,47	0,59
15	77	578277	916	85,53	3,71	3,88	228,35	0,60
17	76	579569	933	85,72	3,78	3,98	232,33	0,61
10	75	580319	943	85,83	3,82	4,09	236,42	0,63
4	74	580615	947	85,87	3,84	4,20	240,62	0,64
5	73	580980	952	85,93	3,86	4,32	244,94	0,65
10	72	581700	962	86,03	3,90	4,44	249,38	0,66
7	71	582197	969	86,11	3,92	4,57	253,95	0,67
12	70	583037	981	86,23	3,97	4,70	258,65	0,68
12	69	583865	993	86,35	4,02	4,83	263,48	0,70
8	68	584409	1001	86,43	4,05	4,98	268,46	0,71
7	67	584878	1008	86,50	4,08	5,13	273,59	0,72
9	66	585472	1017	86,59	4,12	5,28	278,87	0,74
9	65	586057	1026	86,68	4,16	5,45	284,32	0,75
7	64	586505	1033	86,74	4,18	5,62	289,94	0,77
9	63	587072	1042	86,83	4,22	5,80	295,74	0,78
6	62	587444	1048	86,88	4,24	5,99	301,73	0,80
11	61	588115	1059	86,98	4,29	6,18	307,91	0,81
11	60	588775	1070	87,08	4,33	6,39	314,30	0,83

Continúa en la página siguiente...

Revistas	Trabajos	Acum. Trabajos	Acum. Revistas	Porcentaje trabajos	Porcentaje revistas	Dist. Teórica	Acum. Dist. Teórica	Porcentaje Dist. Teórica
16	59	589719	1086	87,22	4,40	6,61	320,91	0,85
12	58	590415	1098	87,32	4,45	6,84	327,75	0,87
11	57	591042	1109	87,41	4,49	7,08	334,83	0,89
8	56	591490	1117	87,48	4,52	7,34	342,17	0,91
16	55	592370	1133	87,61	4,59	7,61	349,78	0,93
11	54	592964	1144	87,70	4,63	7,89	357,67	0,95
9	53	593441	1153	87,77	4,67	8,19	365,86	0,97
14	52	594169	1167	87,88	4,73	8,51	374,37	0,99
19	51	595138	1186	88,02	4,80	8,85	383,22	1,01
15	50	595888	1201	88,13	4,86	9,21	392,43	1,04
13	49	596525	1214	88,22	4,92	9,58	402,01	1,06
16	48	597293	1230	88,34	4,98	9,99	412	1,09
18	47	598139	1248	88,46	5,05	10,42	422,42	1,12
16	46	598875	1264	88,57	5,12	10,88	433,30	1,15
20	45	599775	1284	88,70	5,20	11,36	444,66	1,18
17	44	600523	1301	88,82	5,27	11,89	456,55	1,21
15	43	601168	1316	88,91	5,33	12,45	469	1,24
19	42	601966	1335	89,03	5,41	13,05	482,05	1,28
21	41	602827	1356	89,16	5,49	13,69	495,74	1,31
21	40	603667	1377	89,28	5,58	14,38	510,12	1,35
27	39	604720	1404	89,44	5,69	15,13	525,25	1,39
17	38	605366	1421	89,53	5,75	15,94	541,19	1,43
14	37	605884	1435	89,61	5,81	16,81	558	1,48
24	36	606748	1459	89,74	5,91	17,76	575,76	1,52
23	35	607553	1482	89,86	6	18,79	594,55	1,57
30	34	608573	1512	90,01	6,12	19,91	614,46	1,63
20	33	609233	1532	90,10	6,20	21,13	635,59	1,68

Continúa en la página siguiente...

Revistas	Trabajos	Acum. Trabajos	Acum. Revistas	Porcentaje trabajos	Porcentaje revistas	Dist. Teórica	Acum. Dist. Teórica	Porcentaje Dist. Teórica
27	32	610097	1559	90,23	6,31	22,47	658,06	1,74
36	31	611213	1595	90,40	6,46	23,95	682,01	1,81
33	30	612203	1628	90,54	6,59	25,57	707,58	1,87
30	29	613073	1658	90,67	6,71	27,36	734,94	1,95
39	28	614165	1697	90,83	6,87	29,35	764,29	2,02
42	27	615299	1739	91	7,04	31,57	795,86	2,11
41	26	616365	1780	91,16	7,21	34,04	829,90	2,20
41	25	617390	1821	91,31	7,37	36,82	866,72	2,29
43	24	618422	1864	91,46	7,55	39,95	906,67	2,40
37	23	619273	1901	91,59	7,70	43,50	950,17	2,51
45	22	620263	1946	91,73	7,88	47,55	997,72	2,64
42	21	621145	1988	91,87	8,05	52,18	1049,90	2,78
56	20	622265	2044	92,03	8,28	57,53	1107,43	2,93
69	19	623576	2113	92,22	8,56	63,75	1171,18	3,10
75	18	624926	2188	92,42	8,86	71,03	1242,21	3,29
64	17	626014	2252	92,59	9,12	79,63	1321,84	3,50
83	16	627342	2335	92,78	9,46	89,89	1411,73	3,74
81	15	628557	2416	92,96	9,78	102,28	1514,01	4,01
132	14	630405	2548	93,23	10,32	117,41	1631,42	4,32
99	13	631692	2647	93,43	10,72	136,17	1767,59	4,68
116	12	633084	2763	93,63	11,19	159,81	1927,40	5,10
144	11	634668	2907	93,87	11,77	190,19	2117,59	5,60
160	10	636268	3067	94,10	12,42	230,13	2347,72	6,21
195	9	638023	3262	94,36	13,21	284,11	2631,83	6,97
242	8	639959	3504	94,65	14,19	359,58	2991,41	7,92
316	7	642171	3820	94,98	15,47	469,65	3461,06	9,16
380	6	644451	4200	95,31	17,01	639,25	4100,31	10,85

Continúa en la página siguiente. . .

Revistas	Trabajos	Acum. Trabajos	Acum. Revistas	Porcentaje trabajos	Porcentaje revistas	Dist. Teórica	Acum. Dist. Teórica	Porcentaje Dist. Teórica
574	5	647321	4774	95,74	19,33	920,52	5020,83	13,29
864	4	650777	5638	96,25	22,83	1438,31	6459,14	17,10
1409	3	655004	7047	96,87	28,54	2557	9016,14	23,86
3497	2	661998	10544	97,91	42,70	5753,25	14769,39	39,09
14149	1	676147	24693	100	100	23013	37782,39	100

Tabla B.8: *Distribución de Bradford en las referencias bibliográficas*

B.5. Productividad de los autores

Autores	Trabajos	Acum. Trabajos	Acum. Autores	Porcentaje trabajos	Porcentaje autores	Dist. Teórica	Acum. Dist. Teórica	Porcentaje Dist. Teórica
1	130	130	1	4,21	0	0,01	0,01	0
1	99	229	2	7,42	0	0,01	0,02	0,01
1	88	317	3	10,27	0,01	0,02	0,04	0,02
2	86	403	5	13,06	0,01	0,02	0,06	0,03
1	80	483	6	15,65	0,01	0,02	0,08	0,04
3	79	562	9	18,21	0,02	0,02	0,10	0,05
1	74	636	10	20,61	0,02	0,02	0,12	0,06
1	73	709	11	22,97	0,02	0,02	0,14	0,07
1	71	780	12	25,28	0,02	0,03	0,17	0,08
1	70	850	13	27,54	0,02	0,03	0,20	0,09
1	69	919	14	29,78	0,03	0,03	0,23	0,11
1	67	986	15	31,95	0,03	0,03	0,26	0,12
1	66	1052	16	34,09	0,03	0,03	0,29	0,14
2	65	1117	18	36,20	0,03	0,03	0,32	0,15
1	64	1181	19	38,27	0,03	0,03	0,35	0,17
1	63	1244	20	40,31	0,04	0,03	0,38	0,18
1	62	1306	21	42,32	0,04	0,03	0,41	0,19
1	61	1367	22	44,30	0,04	0,03	0,44	0,21
2	60	1427	24	46,24	0,04	0,04	0,48	0,23
1	58	1485	25	48,12	0,05	0,04	0,52	0,25
1	57	1542	26	49,97	0,05	0,04	0,56	0,26
1	56	1598	27	51,78	0,05	0,04	0,60	0,28
2	55	1653	29	53,56	0,05	0,04	0,64	0,30
1	54	1707	30	55,31	0,05	0,04	0,68	0,32
2	53	1760	32	57,03	0,06	0,05	0,73	0,34
4	51	1811	36	58,68	0,06	0,05	0,78	0,37
2	50	1861	38	60,30	0,07	0,05	0,83	0,39

Continúa en la página siguiente. ...

Autores	Trabajos	Acum. Trabajos	Acum. Autores	Porcentaje trabajos	Porcentaje autores	Dist. Teórica	Acum. Dist. Teórica	Porcentaje Dist. Teórica
3	49	1910	41	61,89	0,07	0,05	0,88	0,42
3	48	1958	44	63,45	0,08	0,06	0,94	0,44
2	47	2005	46	64,97	0,08	0,06	1	0,47
3	46	2051	49	66,46	0,09	0,06	1,06	0,50
4	45	2096	53	67,92	0,10	0,06	1,12	0,53
2	44	2140	55	69,35	0,10	0,07	1,19	0,56
5	43	2183	60	70,74	0,11	0,07	1,26	0,59
8	42	2225	68	72,10	0,12	0,07	1,33	0,63
4	41	2266	72	73,43	0,13	0,08	1,41	0,67
6	40	2306	78	74,72	0,14	0,08	1,49	0,70
6	39	2345	84	75,99	0,15	0,09	1,58	0,75
10	38	2383	94	77,22	0,17	0,09	1,67	0,79
6	37	2420	100	78,42	0,18	0,09	1,76	0,83
8	36	2456	108	79,59	0,19	0,10	1,86	0,88
9	35	2491	117	80,72	0,21	0,11	1,97	0,93
8	34	2525	125	81,82	0,23	0,11	2,08	0,98
8	33	2558	133	82,89	0,24	0,12	2,20	1,04
10	32	2590	143	83,93	0,26	0,13	2,33	1,10
13	31	2621	156	84,93	0,28	0,14	2,47	1,16
9	30	2651	165	85,90	0,30	0,14	2,61	1,23
16	29	2680	181	86,84	0,33	0,15	2,76	1,30
17	28	2708	198	87,75	0,36	0,17	2,93	1,38
17	27	2735	215	88,63	0,39	0,18	3,11	1,47
23	26	2761	238	89,47	0,43	0,19	3,30	1,56
14	25	2786	252	90,28	0,45	0,21	3,51	1,66
27	24	2810	279	91,06	0,50	0,23	3,74	1,76
20	23	2833	299	91,80	0,54	0,25	3,99	1,88

Continúa en la página siguiente...

Autores	Trabajos	Acum. Trabajos	Acum. Autores	Porcentaje trabajos	Porcentaje autores	Dist. Teórica	Acum. Dist. Teórica	Porcentaje Dist. Teórica
38	22	2855	337	92,51	0,61	0,27	4,26	2,01
34	21	2876	371	93,20	0,67	0,29	4,55	2,15
47	20	2896	418	93,84	0,75	0,32	4,87	2,30
49	19	2915	467	94,46	0,84	0,36	5,23	2,47
44	18	2933	511	95,04	0,92	0,40	5,63	2,66
73	17	2950	584	95,59	1,05	0,45	6,08	2,87
86	16	2966	670	96,11	1,21	0,51	6,59	3,11
81	15	2981	751	96,60	1,35	0,58	7,17	3,38
95	14	2995	846	97,05	1,53	0,66	7,83	3,69
116	13	3008	962	97,47	1,74	0,77	8,60	4,06
159	12	3020	1121	97,86	2,02	0,90	9,50	4,48
192	11	3031	1313	98,22	2,37	1,07	10,57	4,99
207	10	3041	1520	98,54	2,74	1,30	11,87	5,60
312	9	3050	1832	98,83	3,30	1,60	13,47	6,35
374	8	3058	2206	99,09	3,98	2,03	15,50	7,31
580	7	3065	2786	99,32	5,03	2,65	18,15	8,56
808	6	3071	3594	99,51	6,48	3,61	21,76	10,26
1197	5	3076	4791	99,68	8,64	5,20	26,96	12,72
1965	4	3080	6756	99,81	12,19	8,12	35,08	16,55
3553	3	3083	10309	99,90	18,60	14,44	49,52	23,36
8586	2	3085	18895	99,97	34,09	32,50	82,02	38,69
36538	1	3086	55433	100	100	130	212,02	100

Tabla B.9: Distribución de Lotka en Pubmed

Autores	Trabajos	Acum. Trabajos	Acum. Autores	Porcentaje trabajos	Porcentaje autores	Dist. Teórica	Acum. Dist. Teórica	Porcentaje Dist. Teórica
1	151	151	1	6,44	0	0,01	0,01	0
1	89	240	2	10,24	0	0,02	0,03	0,01
1	82	322	3	13,74	0	0,02	0,05	0,02
1	79	401	4	17,11	0,01	0,02	0,07	0,03
1	75	476	5	20,31	0,01	0,03	0,10	0,04
1	74	550	6	23,46	0,01	0,03	0,13	0,05
2	73	623	8	26,58	0,01	0,03	0,16	0,07
1	71	694	9	29,61	0,01	0,03	0,19	0,08
1	69	763	10	32,55	0,02	0,03	0,22	0,09
1	63	826	11	35,24	0,02	0,04	0,26	0,11
1	61	887	12	37,84	0,02	0,04	0,30	0,12
1	60	947	13	40,40	0,02	0,04	0,34	0,14
1	58	1005	14	42,88	0,02	0,04	0,38	0,15
3	56	1061	17	45,26	0,03	0,05	0,43	0,17
3	53	1114	20	47,53	0,03	0,05	0,48	0,20
1	51	1165	21	49,70	0,03	0,06	0,54	0,22
2	49	1214	23	51,79	0,03	0,06	0,60	0,24
1	48	1262	24	53,84	0,04	0,07	0,67	0,27
3	47	1309	27	55,84	0,04	0,07	0,74	0,30
1	45	1354	28	57,76	0,04	0,07	0,81	0,33
5	44	1398	33	59,64	0,05	0,08	0,89	0,36
3	43	1441	36	61,48	0,05	0,08	0,97	0,39
4	42	1483	40	63,27	0,06	0,09	1,06	0,43
2	41	1524	42	65,02	0,06	0,09	1,15	0,47
2	40	1564	44	66,72	0,07	0,09	1,24	0,50
4	39	1603	48	68,39	0,07	0,10	1,34	0,55
4	38	1641	52	70,01	0,08	0,10	1,44	0,59

Continúa en la página siguiente...

Autores	Trabajos	Acum. Trabajos	Acum. Autores	Porcentaje trabajos	Porcentaje autores	Dist. Teórica	Acum. Dist. Teórica	Porcentaje Dist. Teórica
6	37	1678	58	71,59	0,09	0,11	1,55	0,63
5	36	1714	63	73,12	0,10	0,12	1,67	0,68
5	35	1749	68	74,62	0,10	0,12	1,79	0,73
7	34	1783	75	76,07	0,11	0,13	1,92	0,78
11	33	1816	86	77,47	0,13	0,14	2,06	0,84
7	32	1848	93	78,84	0,14	0,15	2,21	0,90
12	31	1879	105	80,16	0,16	0,16	2,37	0,96
9	30	1909	114	81,44	0,17	0,17	2,54	1,03
6	29	1938	120	82,68	0,18	0,18	2,72	1,11
10	28	1966	130	83,87	0,20	0,19	2,91	1,18
12	27	1993	142	85,03	0,22	0,21	3,12	1,27
14	26	2019	156	86,13	0,24	0,22	3,34	1,36
20	25	2044	176	87,20	0,27	0,24	3,58	1,46
18	24	2068	194	88,23	0,29	0,26	3,84	1,56
29	23	2091	223	89,21	0,34	0,29	4,13	1,68
33	22	2113	256	90,15	0,39	0,31	4,44	1,81
21	21	2134	277	91,04	0,42	0,34	4,78	1,94
41	20	2154	318	91,89	0,48	0,38	5,16	2,10
46	19	2173	364	92,70	0,55	0,42	5,58	2,27
52	18	2191	416	93,47	0,63	0,47	6,05	2,46
50	17	2208	466	94,20	0,71	0,52	6,57	2,67
65	16	2224	531	94,88	0,80	0,59	7,16	2,91
73	15	2239	604	95,52	0,92	0,67	7,83	3,19
95	14	2253	699	96,12	1,06	0,77	8,60	3,50
112	13	2266	811	96,67	1,23	0,89	9,49	3,86
143	12	2278	954	97,18	1,45	1,05	10,54	4,29
151	11	2289	1105	97,65	1,67	1,25	11,79	4,80

Continúa en la página siguiente. ...

Autores	Trabajos	Acum. Trabajos	Acum. Autores	Porcentaje trabajos	Porcentaje autores	Dist. Teórica	Acum. Dist. Teórica	Porcentaje Dist. Teórica
248	10	2299	1353	98,08	2,05	1,51	13,30	5,41
289	9	2308	1642	98,46	2,49	1,86	15,16	6,17
356	8	2316	1998	98,81	3,03	2,36	17,52	7,13
560	7	2323	2558	99,10	3,88	3,08	20,60	8,38
787	6	2329	3345	99,36	5,07	4,19	24,79	10,09
1174	5	2334	4519	99,57	6,85	6,04	30,83	12,54
2071	4	2338	6590	99,74	9,99	9,44	40,27	16,38
3946	3	2341	10536	99,87	15,97	16,78	57,05	23,21
9709	2	2343	20245	99,96	30,69	37,75	94,80	38,57
45731	1	2344	65976	100	100	151	245,80	100

Tabla B.10: *Distribución de Lotka en SCI*

C.1. Lista de abreviaturas y sus nombres desarrollados

CSIC	Consejo Superior de Investigaciones Científicas
CPAN	Comprehensive Perl Archive Network
DBI	Data Base Interface
EMEA	European Medicines Agency
ER	Enfermedades raras
ICD	International Classification of Diseases
INERGEN	Instituto de Investigación de Enfermedades Raras de Base Genética
IME	Índice Médico Español
ISI	Institute of Scientific Information
JCR	Journal Citations Report
MeSH	Medical Subject Headings
MIM	Mendelian Inheritance in Man
NORD	National Organization of Rare Diseases
OMIM	Online Mendelian Inheritance in Man
PERL	Practical Extraction and Report Language
RETICS	Redes Temáticas de Investigación Cooperativa Sanitaria
SCI	Science Citation Index
SQL	Structured Query Language
SPSS	Statistical Package for the Social Sciences
SSCI	Social Sciences Citation Index
UE	Unión Europea
UE-15	Países pertenecientes a la “Europa de los quince”
WHO	World Health Organization
WHOSIS	World Health Organization - Statistical Information System