

Artículo original

Caracterización bibliométrica de la producción bibliográfica de los grupos de investigación cardio-cerebrovascular, España 1996-2004

Raúl I. Méndez-Vásquez^{a,*}, Eduard Suñén-Pinyol^a, Rosa Cervelló^b y Jordi Camí^c

^a *Bibliometría, Fundació Parc de Recerca Biomèdica de Barcelona (FPRBB), Barcelona, España*

^b *Institut Municipal d'Investigació Mèdica (IMIM), Barcelona, España*

^c *Departament de Ciències Experimentals i de la Salut (CEXS), Universitat Pompeu Fabra (UPF), Barcelona, España*

Historia del artículo:

Recibido el 4 de agosto de 2011

Aceptado el 13 de febrero de 2012

On-line el 23 de mayo de 2012

Palabras clave:

Análisis de coautoría

Normalización de autores

Grupos de investigación

Bibliometría

Cardio-cerebrovascular

RESUMEN

Introducción y objetivos: La abundancia de estudios macro sobre la producción en biomedicina en España pone de manifiesto la escasez de estudios micro que informan sobre la actividad de grupos de investigación, unidad básica del sistema de ciencia-tecnología. Entre las dificultades que explicarían dicha escasez, cuentan la ambigüedad del concepto de «grupo de investigación» y la existencia de firmas bibliográficas sinónimas y homónimas que falsean la correspondencia entre firmas bibliográficas y autores. El objetivo del presente estudio es caracterizar la producción bibliográfica en el campo cardio-cerebrovascular e identificar los grupos de investigación en esta área de estudio.

Métodos: Empleando la base de datos *National Citation Report for Spain* de Thomson-Reuters y el tesauro *Medical Subject Headings* de la *National Library of Medicine*, se definió el área cardio-cerebrovascular, y posteriormente se identificó los grupos de investigación mediante el análisis de coautorías y el concurso de un experto. Los grupos se caracterizaron bibliométricamente empleando indicadores de actividad y visibilidad.

Resultados: Se identificó un total de 93 grupos, formados por 772 autores distintos, a partir de un subconjunto inicial de 6.540 publicaciones en tema cardio-cerebrovascular. Los grupos detectados procedieron principalmente del sector sanitario y universitario, y se concentraron en las comunidades de Cataluña y Madrid. La producción científica atribuible a los grupos presentó indicadores de visibilidad superiores a la media del ámbito de biomedicina.

Conclusiones: La colaboración entre los sectores sanitario y universitario predominó en el área cardio-cerebrovascular; por el contrario, la colaboración internacional es una asignatura pendiente, con índices bajos en comparación con la media de biomedicina.

© 2012 Sociedad Española de Cardiología. Publicado por Elsevier España, S.L. Todos los derechos reservados.

Identification and Bibliometric Characterization of Research Groups in the Cardio-Cerebrovascular Field, Spain 1996-2004

ABSTRACT

Introduction and objectives: The abundance of macro-level studies on scientific production in the field of biomedicine in Spain only serves to highlight the scarcity of micro-level studies reporting on the activity of research groups—the basic units of the science and technology system. This lack of information may well be explained by the ambiguity inherent in the “research group” concept and by the existence of synonymous and homonymous bibliographic signatures that confuse the correspondence between these and the real authors. The aim of this study is to describe bibliographic production in cardio-cerebrovascular research and identify research groups active in the field.

Methods: Using Thomson-Reuters' *National Citation Report for Spain* database and the *National Library of Medicine Medical Subject Headings* thesaurus, we defined the field of cardio-cerebrovascular research and identified research groups through coauthorship analysis supported by the opinions of an expert. Groups were described in terms of bibliometric indicators of activity and visibility.

Results: Ninety-three groups made up of 772 different authors were identified from an initial subset of 6540 publications on cardio-cerebrovascular research. The groups we identified came mainly from the healthcare sector and the universities and were mostly located in the autonomous regions of Catalonia and the Community of Madrid. The scientific production attributable to the groups presented indicators of visibility above the mean for biomedicine.

Conclusions: Collaboration between the healthcare sector and the universities dominated cardio-cerebrovascular research, although international collaboration rates were poor, standing at levels below the mean for biomedicine.

Full English text available from: www.revespcardiol.org

© 2012 Sociedad Española de Cardiología. Published by Elsevier España, S.L. All rights reserved.

Keywords:

Coauthorship analysis

Author-name disambiguation

Research groups

Bibliometrics

Cardio-cerebrovascular

* Autor para correspondencia: Bibliometría, Fundació Parc de Recerca Biomèdica de Barcelona, Dr. Aiguader 88, 4.ª planta, 08003 Barcelona, España. Correo electrónico: rmendez@prbb.org (R.I. Méndez-Vásquez).

Abreviaturas

CSIC: Consejo Superior de Investigación Científica
 JCR: *Journal Citation Reports*
 MeSH: *Medical Subject Headings*

INTRODUCCIÓN

Si bien actualmente se acepta en general que la investigación es una actividad predominantemente colectiva¹, apenas se dispone de información sobre la producción científica de grupos de investigación, muy al contrario de lo que ocurre respecto a recuentos bibliométricos sobre comunidades autónomas, sectores institucionales y centros en el ámbito de biomedicina y ciencias de la salud². Las aproximaciones de dos estudios previos ejemplifican la diferencia entre el concepto administrativo y el funcional de grupo de investigación. Bordons y Zulueta³ analizaron diferentes facetas de la interdisciplinariedad en grupos españoles en cardiología y obtuvieron datos como el tamaño y la composición de los grupos a través de encuestas a investigadores en esta temática. En otras palabras, se analizaron datos reportados. En el segundo estudio, Valderrama-Zurián et al⁴ analizaron las redes de coautoría en las publicaciones de REVISTA ESPAÑOLA DE CARDIOLOGÍA, y llegaron a identificar 25 agrupaciones o, lo que es lo mismo conceptualmente, se partió de los resultados para obtener información sobre los investigadores responsables. Sin embargo, la detección de los grupos y su evaluación bibliométrica no fueron el objetivo de estos estudios, por lo que la información disponible a este respecto continúa siendo escasa. Además de la mencionada diferencia entre la concepción administrativa que enfatiza los recursos que se comparten y la afiliación institucional, y la funcional, que enfatiza la frecuencia y la regularidad con que se producen resultados^{5,6}, la caracterización bibliométrica de grupos exige atribuir las publicaciones a sus respectivos autores, proceso llamado normalización de autores. Este proceso representa un reto no sólo por los elevados volúmenes de publicaciones que es necesario tratar para conseguir muestras representativas, sino también por el hecho de que en las bases de datos no existe correspondencia directa entre autores (investigadores) y firmas bibliográficas. Dicha falta de conexión está causada por la presencia de firmas idénticas (homónimos) que reúnen falsamente la producción de varios autores distintos, así como en que hay variantes de firmas bibliográficas que corresponden a un único autor (sinónimos). El análisis de coautoría cuantifica la frecuencia con que los autores (firmas bibliográficas) concurren en las publicaciones, características que lo convierten en la principal herramienta de detección de grupos a partir de sus resultados. Los algoritmos informáticos de análisis de coautoría permiten tratar elevados volúmenes de publicaciones, si bien la distorsión que homónimos y sinónimos inducen con frecuencia invalida sus resultados. Se han ensayado diversas aproximaciones automáticas para minimizar esta distorsión: métodos probabilísticos⁷⁻⁹, grafos de estado finito¹⁰, algoritmos recursivos¹¹ o combinando firmas e instituciones¹²; sin embargo, los resultados no han sido plenamente satisfactorios. Los resultados del análisis de coautoría se analizan representando los autores como nodos, entre los que se dibujan aristas que representan la concurrencia en una publicación determinada¹³. Girvan y Newman¹⁴ estudiaron las redes de colaboración entre investigadores combinando esta metodología y las publicaciones en revistas científicas, y demostraron que es posible detectar grupos muy cohesionados entre los cuales hay enlaces laxos, propiedad que dieron por llamar *community structure*. Con la finalidad de ir cubriendo el déficit de información bibliométrica, se presenta en esta publicación el mapa de grupos de

investigación especializados en cardio-cerebrovascular en España. Su confección se fundamenta en el concepto funcional de grupo de investigación, es decir, definido por sus resultados —en este caso, sus publicaciones en revistas científicas—, en virtud de lo cual se considera grupo estable a todo núcleo de autores que firma regularmente de forma conjunta trabajos de investigación científica en un tema determinado⁶.

MÉTODOS

Nuestra aproximación consiste en un método semiautomático y cíclico que combina el análisis de coautoría y la normalización manual de firmas bibliográficas, de forma que los autores y su producción se identifican incrementalmente en cada ciclo. Dicho método se resume en siete etapas (figura) que se repitieron hasta obtener la aprobación del experto.

Detección de grupos de investigación

Etapas 1, fuente de datos y definición de la colección de estudio

La base de datos *National Citation Report* para España fue la fuente de datos del estudio. El *National Citation Report* es un producto de Thomson-Reuters que incluye los documentos publicados desde España en todos los campos de la ciencia durante el periodo 1981-2004. La colección de documentos en cardio-cerebrovascular se seleccionó mediante un filtro compuesto por 3.106 términos del tesoro de la *National Library of Medicine* de Estados Unidos (*Medical Subject Headings*¹⁵ [MeSH]). Mediante este filtro se seleccionaron 6.540 publicaciones entre 1996 y 2004 que presentaban en el título y/o las palabras clave (*keywords*) al menos uno de los términos del filtro (colección cardio-cerebrovascular).

Etapas 2, preselección de publicaciones y firmas bibliográficas o autores

La detección de grupos se basó en dos subconjuntos de publicaciones: a) las firmadas por autores con al menos el 50% de su producción en la colección cardio-cerebrovascular (filtro 50%), y b) las firmadas por autores con 10 o más documentos en dicha colección. Este segundo punto de corte se eligió con base

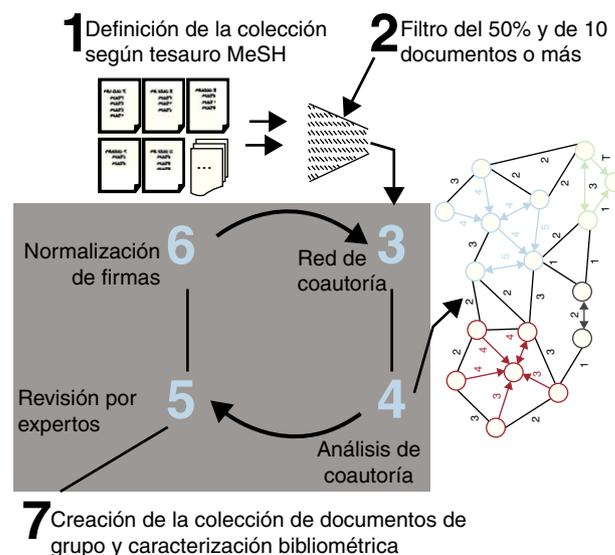


Figura. Desglose de las etapas de la detección de grupos de investigación. MeSH: *Medical Subject Headings*.

en la distribución del número de documentos por autor en la colección cardio-cerebrovascular.

Etapa 3, red de coautoría

Se construyó a partir de los documentos seleccionados en la etapa 2 representando las firmas bibliográficas como vértices y su relación de coautoría (la concurrencia de dos firmas distintas en una publicación) como aristas. Para cada vértice se obtuvo el número de veces que apareció en la colección (número de documentos), y para las aristas el número de veces que las dos firmas implicadas concurren en la colección (número de coautorías). Las relaciones de coautoría de la red se consideraron simétricas, es decir, que no se tuvo en cuenta la direccionalidad de las coautorías.

Etapa 4, análisis de coautoría

Se realizó sobre la red construida en la etapa 3 usando un algoritmo de elaboración propia que compara todos los nodos con sus respectivos vecinos agrupándolos en torno al vecino próximo más importante. Este se determinó aplicando cuatro criterios secuencialmente en caso de empate, lo que aseguró que los autores quedarán adscritos a un único grupo: *a)* número de publicaciones en coautoría; *b)* número de publicaciones del vecino; *c)* número de coautores del vecino, y *d)* la densidad local del vecino. El algoritmo se alimentó con los documentos seleccionados en la etapa 2, que se filtraron según el número mínimo de cuatro parámetros: *a)* número de autores por grupo; *b)* número de documentos por autor para formar grupo; *c)* número de documentos en coautoría para formar grupo, y *d)* número de coautores por autor para formar grupo. Por la estabilidad y la plausibilidad de sus resultados, se aplicaron respectivamente los siguientes valores: 4, 4, 3, 3.

Etapa 5, revisión por expertos

En el presente estudio participó un experto, procedente del entorno sanitario, con trayectoria en investigación científica en cardiología. Esta etapa consistió en la evaluación cualitativa de los grupos detectados en cada ciclo, refiriendo para cada caso su apreciación sobre la plausibilidad del grupo presentado. La evaluación del experto guio las sucesivas etapas de normalización de firmas.

Etapa 6, normalización manual de firmas bibliográficas

Por una parte, se unificaron las diferentes variantes de firma de un autor (unificación de sinónimos) y, por otra, se separaron las publicaciones de distintos autores que aparecían falsamente agrupados bajo una única firma bibliográfica en la fuente de datos (separación de homónimos). Como resultado de la etapa 4, las firmas sinónimas quedaron reunidas en un grupo junto a sus coautores más frecuentes. Así, su normalización sólo requirió crear una entidad «autor» en la base de datos y asociarla a las correspondientes firmas sinónimas. Estas entidades «autor» se crearon utilizando el nombre completo y el centro de filiación de los autores referidos en las publicaciones científicas y/o los documentos institucionales más recientes encontrados mediante búsquedas de pares de firmas (coautores) usando *Google Scholar* y *Google*. Los homónimos se separaron manualmente seleccionando las publicaciones en las que aparecía la firma en cuestión y un coautor determinado (generalmente el más frecuente), de forma similar a lo propuesto por Wooding et al¹¹. Las publicaciones así separadas se asociaron a su correspondiente entidad «autor» en la base de datos del estudio.

Etapa 7, documentos de grupos

Tras obtener el visto bueno del experto, se creó la colección de «documentos de grupo» asignando a cada grupo los documentos firmados por al menos uno de sus integrantes.

Clasificación temática

Los documentos de la colección cardio-cerebrovascular se ordenaron según una clasificación *ad-hoc*, desarrollada con la supervisión del experto, que agrupó la colección de 3.106 términos MeSH en 23 categorías. La relación de los términos por categoría están disponibles en el informe *web* del presente estudio. Asimismo se utilizó la clasificación del *Journal Citation Reports* (JCR) 2004 tal como Thomson-Reuters la proveyó al *National Citation Report* para España.

Normalización de las direcciones de centros de investigación y clasificación sectorial

Las variantes de los nombres de los centros referidas en las publicaciones se unificaron a una única denominación. Este proceso permitió atribuir las publicaciones analizadas a todas las organizaciones referidas de acuerdo con el método de asignación total. En virtud de dicho método, los documentos se asocian a todos los centros firmantes, lo que permite analizar la colaboración entre centros, territorios y sectores institucionales. Sin embargo, la asociación de un único documento a más de un centro origina que el recuento directo de los documentos por centro, territorios o sectores sume más que el total real, ya que los documentos asignados a más de un centro se cuentan más de una vez. Implícitamente, los documentos firmados por más de un centro se consideran producto de la colaboración entre centros, y análogamente para territorios y sectores institucionales.

La denominación de los centros procedió del Catálogo Nacional de Hospitales 2006 y el Registro Nacional de Universidades, Centros y Enseñanzas. Para las demás organizaciones se empleó el nombre referido en sus páginas *web* oficiales o en directorios acreditados. Asimismo, se obtuvo la dirección postal completa de cada organización, lo que permitió derivar su localización geográfica. Los centros detectados durante la normalización se clasificaron en cinco sectores institucionales según su forma jurídica o la naturaleza de su actividad. Así, en el sector universitario se incluyeron las universidades y los centros de su esfera, como escuelas universitarias e institutos propios. El sanitario incluyó los hospitales públicos y privados, los centros de investigación estrechamente relacionados con la investigación clínica y otro tipo de centros, como bancos de tejidos e instalaciones de diagnóstico por imagen, así como centros de atención primaria. El de los organismos públicos de investigación (OPI) incluyó los centros del Consejo Superior de Investigación Científica (CSIC), del Instituto de Salud Carlos III (ISCIII) y centros públicos de investigación de las comunidades autónomas. El sector de la Administración, las ONG y otros incluyó centros de la Administración, tanto central como autonómica, así como ONG y asociaciones científicas. El sector empresarial quedó formado principalmente empresas farmacéuticas.

Caracterización bibliométrica

El análisis bibliométrico se circunscribió a los documentos citables (artículos, revisiones y *proceeding papers*) de las colecciones cardio-cerebrovascular y de documentos de grupos, e incluyó

indicadores bibliométricos de: a) actividad: número de documentos (Docs); b) visibilidad: número de citas, la media de citas por documento (CD), el porcentaje de documentos no citados en el periodo de estudio (%NC) y la relación con la media de citas ponderada en España para el área cardio-cerebrovascular (MCE) (este último indicador se calculó dividiendo el CD de un documento entre el CD promedio de la colección cardio-cerebrovascular del mismo año de publicación; valores de MCE > 1 indican que se recibieron más citas que la media del área temática en España en el periodo de estudio); con carácter experimental, se calculó un índice de Hirsch¹⁶ de grupo (índice H) limitado al periodo temporal de estudio¹⁷, y c) indicadores de colaboración: el porcentaje de publicaciones en colaboración internacional (%Int), que incluye todo documento con al menos un centro firmante localizado en el extranjero, y el porcentaje de colaboración interregional o entre comunidades autónomas (%Reg).

Desglose del análisis

Los indicadores se calcularon por: a) comunidades autónomas; b) sectores institucionales; c) centros de investigación; d) grupos de investigación, y e) subáreas científicas. Este desglose requirió adscribir los grupos, con carácter orientativo, a la comunidad autónoma, el centro y la subárea temática más frecuente en sus documentos (adscripción principal) y a la que se repitió al menos el 80% de veces que la principal (adscripción secundaria). Se consideró que un agente (comunidades autónomas, centros, grupos, etc.) recibió mayor visibilidad cuando presentó de manera simultánea valores de CD, MCE y %Int superiores a la media de referencia, así como valores de %NC inferiores a esta.

RESULTADOS

Se puede ampliar y complementar los resultados del presente estudio acudiendo a la dirección web: <http://193.145.216.56/GruposCardioCerebrovascular>.

Caracterización de la colección cardio-cerebrovascular España, 1996-2004

La metodología aplicada recuperó 6.540 documentos (colección cardio-cerebrovascular España), el 63% firmados por centros con sede en Madrid y Cataluña. Este subconjunto de publicaciones recibió mayor visibilidad que la media de la colección (tabla 1). El 12,6% de la colección se publicó en colaboración entre comunidades autónomas, mientras que en el 61,3% había colaboración entre sectores. El sector que más colaboró fue el sanitario y la interacción más frecuente se produjo con centros universitarios, seguida de la colaboración entre los sectores universitario y OPI (tabla 2). En contraste, las publicaciones en colaboración internacional firmadas por centros del sector sanitario representaron un 17%, valor inferior al de los demás sectores institucionales, que en conjunto alcanzaron el 23% (tabla 3).

Cardiología clínica fue el área *ad-hoc* basada en términos MeSH más activa. Sin embargo, las áreas *ad-hoc* que recibieron mayor visibilidad fueron: Coagulación, plaquetas y trombosis, Farmacología cardiovascular y Síncope (tabla 4). Según el JCR, la colección cardio-cerebrovascular se distribuyó en 122 disciplinas distintas, de las cuales Sistema cardiovascular aglutinó mayor número de publicaciones (un tercio del total). REVISTA ESPAÑOLA DE CARDIOLOGÍA es la más utilizada de las 1.020 revistas analizadas. La distribución de la producción por disciplina JCR, así como por revista, se ha omitido por economía y está disponibles en el informe web.

Tabla 1
Producción en cardio-cerebrovascular por comunidades autónomas, España 1996-2004

Comunidad autónoma	Docs ^a	%Docs ^b	%Acum ^c	Citas ^d	CD ^e	MCE ^f	%NC ^g	%Int ^h
Cataluña	2.133	32,6	32,6	22.660	10,6	1,100	25,1	27,0
Comunidad de Madrid	2.003	30,6	63,2	21.398	10,7	1,080	29,7	24,9
Andalucía	720	11,0	74,3	4.328	6,0	0,660	34,0	15,4
Comunidad Valenciana	711	10,9	85,1	4.126	5,8	0,730	34,9	21,7
Galicia	387	5,9	91,0	2.607	6,7	0,830	28,4	16,8
Comunidad Foral de Navarra	254	3,9	94,9	1.788	7,0	0,950	26,8	13,8
Castilla y León	247	3,8	98,7	1.909	7,7	0,660	33,6	11,7
Región de Murcia	217	3,3	102,0	1.246	5,7	0,550	33,2	12,9
País Vasco	199	3,0	105,1	1.672	8,4	0,870	31,7	15,1
Aragón	172	2,6	107,7	1.023	6,0	0,590	32,6	24,4
Canarias	132	2,0	109,7	743	5,6	0,680	33,3	21,2
Cantabria	127	1,9	111,7	884	7,0	0,720	27,6	15,0
Principado de Asturias	109	1,7	113,3	950	8,7	0,750	29,4	14,7
Castilla-La Mancha	98	1,5	114,8	406	4,1	0,580	35,7	10,2
Extremadura	91	1,4	116,2	1.587	17,4	2,090	44,0	34,1
Islas Baleares	78	1,2	117,4	453	5,8	0,630	38,5	23,1
La Rioja	13	0,2	117,6	61	4,7	0,710	30,8	15,4
Total	6.540			55.519	8,5	0,910	30,3	22,7

^a Número de documentos.

^b Porcentaje de documentos respecto al total del área temática.

^c Porcentaje de documentos acumulado.

^d Número de citas recibidas en el periodo 1966-2004.

^e Media de citas por documento.

^f Relación con la media de citas ponderada en España para el área cardio-cerebrovascular.

^g Porcentaje de documentos no citados en el periodo de estudio.

^h Porcentaje de documentos publicados en colaboración internacional.

Tabla 2

Porcentajes de documentos en colaboración entre sectores en cardio-cerebrovascular, España 1996-2004

	Universitario	OPI	ADM, ONG	Empresas	Total
Sanitario	89,58	6,29	1,87	2,26	89,31
Universitario	0,00	73,16	10,17	16,67	10,20
OPI	0,00	0,00	50,00	50,00	0,46
ADM, ONG y otros	0,00	0,00	0,00	100,00	0,03
Empresas					0,00

ADM: Administración; OPI: organismo público de investigación.

Detección y caracterización bibliométrica de los grupos de investigación

El proceso de detección concluyó tras 12 ciclos y produjo 93 grupos formados por 772 autores distintos (8,3 investigadores de media), de los que el 28,0% eran mujeres. Esta observación contrasta con el porcentaje observado en la temática de psiquiatría (42,1%) (véase «Mapa bibliométrico de grupos en psiquiatría»,

disponible en: <http://bac.fundacionrecerca.cat/psiquiatria>). Once grupos de investigación mostraron porcentajes de documentos muy bajos en comparación con sus respectivos totales en el *National Citation Report*, por lo que sus indicadores podrían no ser representativos de su verdadera actividad.

Los 93 grupos detectados explicaron el 51,5% de los documentos y el 57,8% de las citas de la colección cardio-cerebrovascular y presentaron en conjunto valores de CD, MCE y %Reg superiores a los de la colección cardiocerebrovascular. Por el contrario, los valores de %NC y %Int fueron estadísticamente menores. La comparación con los indicadores del ámbito de biomedicina mostró resultados similares (tabla 5).

Relación del tamaño de los grupos con los indicadores bibliométricos

El número de integrantes por grupo, o tamaño de grupo, mostró una relación directa con el volumen de documentos de los grupos ($\rho = 0,770$; $p < 0,01$), con el número de citas recibidas ($\rho = 0,500$;

Tabla 3

Producción en cardio-cerebrovascular por sectores institucionales, España 1996-2004

Sector institucional	Docs ^a	%Docs ^b	%Acum ^c	Citas ^d	C/D ^e	MCE ^f	%NC ^g	%Int ^h
Sanitario	5.175	79,1	79,1	42.745	8,3	0,870	31,7	17,3
Universitario	3.859	59,0	138,1	27.327	7,1	0,860	30,4	21,1
OPI	487	7,4	145,6	7.022	14,4	1,350	18,5	38,2
Empresas	128	2,0	147,5	1.457	11,4	1,040	18,8	30,5
ADM, ONG y otros	101	1,5	149,1	1.171	11,6	1,160	18,8	23,8
Total	6.540			55.519	8,5	0,910	30,3	22,7

ADM: Administración; OPI: organismo público de investigación.

^a Número de documentos.

^b Porcentaje de documentos respecto al total del área temática.

^c Porcentaje de documentos acumulado.

^d Número de citas recibidas en el periodo 1966-2004.

^e Media de citas por documento.

^f Relación con la media de citas ponderada en España para el área cardio-cerebrovascular.

^g Porcentaje de documentos no citados en el periodo de estudio.

^h Porcentaje de documentos publicados en colaboración internacional.

Tabla 4

Producción en cardio-cerebrovascular por áreas *Medical Subject Headings*, España 1996-2004^a

Área temática	Docs ^b	%Docs ^c	Citas ^d	CD ^e	MCE ^f	%NC ^g	%Int ^h
Cardiología clínica	2.335	35,7	21.302	9,1	0,990	30,0	22,4
Coagulación, plaquetas y trombosis	1.740	26,6	19.465	11,2	1,030	25,1	29,1
Cardiopatía isquémica	1.339	20,5	14.737	11,0	1,090	28,5	22,4
Técnicas diagnósticas	1.304	19,9	9.719	7,5	0,900	32,1	20,1
Hipertensión arterial	1.174	18,0	12.524	10,7	1,020	29,1	19,2
Enfermedad cerebrovascular	1.077	16,5	10.783	10,0	0,990	35,2	16,7
Investigación vascular	823	12,6	6.730	8,2	0,960	30,7	23,2
Arritmia	631	9,6	4.896	7,8	0,970	32,6	22,2
Cirugía cardiovascular	589	9,0	5.303	9,0	0,940	29,2	21,2
Farmacología cardiovascular	552	8,4	6.305	11,4	1,090	29,5	23,2
Cirugía vascular	470	7,2	6.534	13,9	1,240	31,1	29,6
Valvulopatías	393	6,0	1.379	3,5	0,740	33,8	11,2
Síncope	103	1,6	1.284	12,5	1,350	27,2	24,3
Total	6.540	194,2	55.519	8,5	0,910	30,3	22,7

^a Subáreas temáticas con 100 o más documentos; la relación completa está disponible en el informe web: <http://193.145.216.56/GruposCardioCerebrovascular>.

^b Número de documentos.

^c Porcentaje de documentos respecto al total del área temática.

^d Número de citas recibidas en el periodo 1966-2004.

^e Media de citas por documento.

^f Relación con la media de citas ponderada en España para el área cardio-cerebrovascular.

^g Porcentaje de documentos no citados en el periodo de estudio.

^h Porcentaje de documentos publicados en colaboración internacional.

Tabla 5

Comparación de indicadores bibliométricos, cardio-cerebrovascular en España 1996-2004

	Docs ^a	Citas ^b	CD ^c	MCE ^d	%NC ^e	%Reg ^f	%Int ^g
Grupos de investigación	3.365	32.086	9,54 ^h	1,123 ^h	26,2	15,0 ^h	18,4
Colección cardio-cerebrovascular	6.540	55.519	8,49	1,037	30,3 ^h	12,6	22,7 ^h
Biomedicina 1996-2004	84.122	719.127	8,55 ^h	1,020 ^h	27,2 ^h	12,6 ^h	27,1 ^h

^a Número de documentos.^b Número de citas recibidas en el periodo de estudio.^c Media de citas por documento.^d Relación con la media de citas ponderada en España para el área cardio-cerebrovascular y según disciplinas del *Journal Citation Reports* para el caso del ámbito de biomedicina.^e Porcentaje de documentos no citados en el periodo de estudio.^f Porcentaje de documentos publicados en colaboración entre comunidades autónomas.^g Porcentaje de documentos publicados en colaboración internacional.^h Diferencias estadísticamente significativas ($p < 0,05$).

$p < 0,01$) y el índice H ($\rho = 0,480$; $p < 0,01$). No se observó relación directa entre el tamaño de grupo y CD, %Int o MCE.

Análisis territorial

Cataluña y Madrid aglutinaron en conjunto más del 50,0% de los grupos, los cuales firmaron el 74,0% de los documentos de grupos. Un grupo de investigación se adscribió a dos centros localizados en Madrid y Castilla-La Mancha, por lo que se obtienen 94 del recuento por comunidades autónomas. La Región de Murcia presentó la media de integrantes por grupo más elevada, mientras que Galicia y la Comunidad Foral de Navarra, las más bajas. Los documentos atribuidos a los grupos en Cataluña, Madrid y Extremadura (1 grupo) recibieron mayor visibilidad que la media del conjunto de grupos. El 8,6% de los documentos de grupo se publicó en colaboración. No se detectaron grupos de investigación en el País Vasco, novena comunidad según volumen de producción en este tema (tabla 6).

Análisis sectorial

El 74,2% de los grupos y el 80,0% de los documentos de grupo procedieron del sector sanitario, mientras que el 28,0% de los grupos

y el 30,4% de los documentos procedieron del universitario. Los grupos del sector sanitario presentaron la media de integrantes más elevada (9,1). Los demás grupos se distribuyeron como sigue: 3 en centros del sector OPI (2 en el Centro de Investigaciones Biológicas, CSIC, de Madrid y 1 en el Instituto de Investigaciones Biomédicas de Barcelona, CSIC), 1 en el sector empresarial (J. Uriach & Cía.), y otro en el sector de la Administración (Institut d'Estudis de la Salut en Barcelona). El 10,8% de los grupos se adscribieron simultáneamente a un hospital y a una universidad, los cuales firmaron el 15,1% del total de documentos en colaboración entre estos dos sectores. Los documentos firmados por centros del sector OPI presentaron los valores de CD y MCE más elevados del estudio.

Análisis temático según áreas Medical Subject Headings y Journal Citation Reports

Cardiología clínica, disciplina *ad-hoc* basada en términos MeSH, aglutinó más de la mitad de los grupos, los documentos y las citas. Las medias de integrantes por grupo más elevadas se observaron en las áreas MeSH de Cirugía cardiovascular y Coagulación, plaquetas y trombosis, 10,3 y 10,1 respectivamente. Excluyendo las áreas MeSH con menos de 3 grupos, las publicaciones de los 8

Tabla 6

Grupos en cardio-cerebrovascular por comunidades autónomas, España 1996-2004

Ordinal	Comunidad autónoma	Grupos, n (%)	MInt ^a	Docs ^b	%Docs ^c	%Acum ^d	Citas ^e	CD ^f	MCE ^g	%NC ^h	%Int ⁱ
1	Cataluña	34 (36,6)	8,5	1.460	43,4	43,4	16.700	11,4	1,150	22,5	23,2
2	Comunidad de Madrid	25 (26,9)	8,8	1.030	30,6	74,0	11.817	11,5	1,150	26,2	18,4
4	Comunidad Valenciana	11 (11,8)	9,7	373	11,1	85,1	1.918	5,1	0,640	33,5	9,9
3	Andalucía	7 (7,5)	6,3	174	5,2	90,3	882	5,1	0,590	31,6	7,5
5	Galicia	4 (4,3)	7,8	175	5,2	95,5	819	4,7	0,730	25,1	10,9
6	Comunidad Foral de Navarra	4 (4,3)	7,8	138	4,1	99,6	1.269	9,2	1,040	22,5	12,3
7	Castilla y León	2 (2,2)	9,5	64	1,9	101,5	352	5,5	0,580	26,6	14,1
8	Región de Murcia	2 (2,2)	13,5	133	4,0	105,4	706	5,3	0,610	32,3	4,5
13	Principado de Asturias	2 (2,2)	5,0	40	1,2	106,6	243	6,1	0,710	27,5	10,0
10	Aragón	1 (1,1)	2,0	23	0,7	107,3	97	4,2	0,460	39,1	73,9
12	Cantabria	1 (1,1)	4,0	21	0,6	107,9	133	6,3	0,630	33,3	4,8
15	Extremadura	1 (1,1)	3,0	22	0,7	108,6	333	15,1	1,620	18,2	72,7
	Total	93	8,3	3.365		108,6	32.086	9,5	1,123	26,2	18,4

^a Media de integrantes por grupo.^b Número de documentos atribuibles a los grupos de investigación.^c Porcentaje de documentos respecto al total de documentos de grupo.^d Porcentaje acumulado de documentos de grupo.^e Número de citas a los documentos de grupo entre 1996 y 2004.^f Media de citas por documento de grupo.^g Relación con la media de citas ponderada en España para el área cardio-cerebrovascular de los documentos de grupo.^h Porcentaje de documentos de grupo no citados en el periodo de estudio.ⁱ Porcentaje de documentos de grupo en colaboración internacional.

Tabla 7Grupos en cardio-cerebrovascular por áreas *Medical Subject Headings*, España 1996-2004

Área temática	Grupos, n (%)	MInt ^a	Docs ^b	%Docs ^c	Citas ^d	CD ^e	MCE ^f	%NC ^g	%Int ^h
Cardiología clínica	49 (52,7)	9,2	1.899	56,4	18.540	9,8	1,040	28,9	17,9
Técnicas diagnósticas	27 (29,0)	8,2	990	29,4	5.914	6,0	0,800	30,7	16,1
Cardiopatía isquémica	23 (24,7)	10	1.151	34,2	11.944	10,4	1,030	27,6	18,6
Coagulación, plaquetas y trombosis	17 (18,3)	10,1	788	23,4	6.660	8,5	0,760	22,8	20,2
Hipertensión arterial	11 (11,8)	7,4	508	15,1	4.956	9,8	1,190	23,8	15,7
Enfermedad cerebrovascular	9 (9,7)	7,6	375	11,1	3.699	9,9	1,080	22,7	8,5
Arritmia	9 (9,7)	6,6	346	10,3	3.111	9,0	1,090	30,9	22,3
Investigación vascular	8 (8,6)	8,5	207	6,2	2.926	14,1	1,260	15,0	20,3
Cirugía cardiovascular	7 (7,5)	10,3	235	7,0	1.486	6,3	0,660	36,6	12,3
Cirugía vascular	7 (7,5)	7,4	257	7,6	3.577	13,9	1,090	30,7	21,0
Farmacología cardiovascular	2 (2,2)	9	87	2,6	463	5,3	0,640	35,6	11,5
Valvulopatías	2 (2,2)	8	55	1,6	194	3,5	0,490	45,5	0,0
Síncope	1 (1,1)	4	13	0,4	45	3,5	0,340	53,8	7,7
Biología molecular	1 (1,1)	6	13	0,4	274	21,1	1,420	23,1	0,0
Aterosclerosis, aterogénesis y lípidos	1 (1,1)	6	13	0,4	274	21,1	1,420	23,1	0,0
Total	93 (187,1)	8,3	3.365		32.086	9,5	1	26,2	18,4

^a Media de integrantes por grupo.^b Número de documentos atribuibles a los grupos de investigación.^c Porcentaje respecto al total de documentos de grupo.^d Número de citas recibidas por los documentos de grupo entre 1996 y 2004.^e Media de citas por documento de grupo.^f Relación con la media de citas ponderada en España para el área cardio-cerebrovascular de los documentos de grupo.^g Porcentaje de documentos de grupo no citados en el periodo de estudio.^h Porcentaje de documentos de grupo en colaboración internacional.

grupos en vascular recibieron mayor visibilidad que la media del conjunto de grupos (tabla 7). La disciplina JCR Sistema cardiovascular aglutinó más de la mitad de los grupos detectados, así como del total de documentos y citas. En el subconjunto de disciplinas JCR con más de un grupo, Hematología presentó la media de integrantes por grupo más elevada y Enfermedad vascular periférica recibió mayor visibilidad.

Análisis de grupos

Tres grupos, 2 con sede en el *Hospital Clínic i Provincial* de Barcelona y otro en el *Hospital Vall d'Hebron*, se situaron de forma

Tabla 8

Valores de referencia de los indicadores bibliométricos de los grupos

Indicador	Mediana	Q1 (p75)	RIQ
Docs ^a	38	59	38
Citas ^b	242	470	360
CD ^c	6,23	10,45	6,38
MCE ^d	0,717	1,122	0,607
%NC ^e	24,53	18,18 ^f	15,67
% Int ^g	11,54	22,22	17,09
Índice H ^h	15	23	13

La relación completa de cuartiles de los indicadores bibliométricos se encuentra disponible en la dirección *web* del informe. Q1: primer cuartil; RIQ: amplitud intercuartílica o diferencia entre los percentiles 75 y 25.

^a Número de documentos.^b Número de citas recibidas.^c Media de citas por documento.^d Relación con la media de citas ponderada en España para el área cardio-cerebrovascular.^e Porcentaje de documentos no citados en el periodo de estudio.^f Debido a que el %NC tiene una significación más positiva cuanto menor es su valor, se muestra el valor correspondiente al percentil 25 o tercer cuartil.^g Porcentaje de documentos publicados en colaboración internacional.^h Índice H del grupo para el periodo 1996-2004.

simultánea en las primeras 3 posiciones según indicadores de actividad y visibilidad, se trata de grupos clínicos en cardiopatía isquémica, arritmia y cirugía cardiovascular. En relación con el MCE, 3 grupos presentaron los valores más elevados (MCE > 2,0): uno en la Universidad Autónoma de Madrid y otro en el Centro de Investigaciones Biológicas del CSIC, y el tercero en el *Hospital Clínic i Provincial de Barcelona*. Tres grupos con sede, respectivamente, en la Universidad de Extremadura, la Universidad de Zaragoza y el *Institut de Recerca Oncològica* en Barcelona publicaron más de la mitad de su producción en colaboración internacional. Los 3 grupos que presentaron los valores de índice H más elevados (> 35) se localizaron en Barcelona, en el *Hospital Clínic i Provincial* y en el *Institut d'Investigacions Biomèdiques August Pi i Sunyer*. Los valores de referencia de los indicadores bibliométricos empleados se presentan en la tabla 8.

DISCUSIÓN

El presente estudio caracterizó desde las perspectivas sectorial, temática y territorial la producción bibliográfica atribuible a los grupos de investigación en cardio-cerebrovascular.

El empleo de términos MeSH en la delimitación del área temática permitió clasificar más documentos en cardio-cerebrovascular que mediante disciplinas JCR, si bien derivó en un elevado grado de solapamiento, ya que las publicaciones se clasificaron según contuvieran o no los términos MeSH seleccionados para el estudio. Sistema cardiovascular, la disciplina JCR que más documentos aglutinó, solamente agrupó un tercio de los documentos de la colección de estudio. Las publicaciones de la colección cardio-cerebrovascular se distribuyeron en 122 disciplinas JCR distintas, lo que pone de manifiesto su limitación en la definición de disciplinas como la cardio-cerebrovascular.

La producción bibliográfica del área, así como la atribuible a los grupos, presentó distribuciones territorial y sectorial muy asimétricas. Una observación típica en estudios bibliométricos es que

pocos agentes son origen de la mayor parte de la producción y por lo general reciben mayor visibilidad. Cataluña y Madrid, así como el sector sanitario y el sector universitario, acumularon la mayor parte de la producción y de los grupos identificados en esta área, y además recibieron mayor visibilidad que los demás actores. Esta observación está en línea con un estudio previo de Bordons y Zulueta³ basado en encuestas a investigadores del área cardiovascular. La prominencia de la capacidad de producción bibliográfica en cardio-cerebrovascular de Madrid y Cataluña se mantiene aun normalizando por el número de habitantes en 2004 registrado por el Instituto Nacional de Estadística. En estos términos, únicamente la Comunidad Foral de Navarra las supera, si bien en valor absoluto su producción en esta área ocupa el sexto lugar.

En cardio-cerebrovascular, la colaboración entre sectores institucionales fue más importante que la colaboración entre comunidades autónomas y que la internacional (el 60 frente al 12 y el 23% respectivamente). Este elevado flujo de conocimiento se explica por la concurrencia de centros sanitarios y universitarios en las publicaciones de esta área que aparecieron en más del 80% de los casos. Este dato se debe analizar a la luz del fenómeno de doble filiación, frecuente entre investigadores clínicos ubicados en hospitales universitarios. Este fenómeno se ha observado en estudios previos⁴, y muy probablemente sería útil un análisis detallado de esta interacción tan relevante. Nuevas aproximaciones deberán explorar e instrumentalizar diferentes definiciones de «colaboración intersectorial» con el objeto de incrementar la precisión en su cuantificación.

Los grupos detectados son representativos de la comunidad científica española en cardio-cerebrovascular. En el presente análisis se detectaron 93 grupos resultado del análisis de 6.540 documentos publicados desde España durante 9 años en más de 1.000 revistas, que se clasificaron en 122 disciplinas JCR distintas. Los grupos procedieron de 5 sectores, con el sanitario y el universitario como principales fuentes, habiéndose identificado 3 grupos en centros del CSIC, y 2 más en una empresa farmacéutica y un centro de la Administración. El número de grupos detectados y su distribución sectorial están en línea con lo observado por Bordons y Zulueta³ en 2002. En la medida en que se cubra el déficit de información sobre grupos de investigación, será posible comparar el tamaño y los indicadores bibliométricos de esta comunidad con los de otros países del entorno, de forma que aporte elementos útiles en la gestión estratégica del potencial investigador en esta área científica. Si se considera el volumen de publicaciones en cardiología, definida según el JCR, área menor que la analizada en el presente estudio, España se sitúa en sexta posición en Europa y novena en el mundo¹⁸, posiciones que seguramente mejorarán en la medida que se estimule a los grupos de investigación con mayor potencial.

La metodología de detección de grupos de investigación es novedosa y robusta. A diferencia de otras metodologías de detección de grupos que es aplican una única vez¹¹, la presente propuesta consiste en un procedimiento iterativo que combina la normalización de firmas con el análisis de coautoría en la selección de autores pertinentes y la posterior detección de los grupos de investigación. Debido a que la normalización de firmas bibliográficas y la detección de grupos de investigación se basan en la frecuencia de coautoría, los resultados de la metodología presentada dependen únicamente del grado de depuración de la información bibliográfica que se alcance. Dicho grado de depuración es una función directa del número de veces que se repite el ciclo de depuración de firmas y de detección de grupos.

La normalización de firmas bibliográficas según coautoría supera las principales limitaciones de la normalización según centro de afiliación o área temática. La normalización por centro de afiliación o disciplina no permite distinguir a los autores

homónimos que trabajan en el mismo centro ni en la misma disciplina^{6,8,10}. Por el contrario, la depuración según coautoría permite diferenciarlos gracias a que, por lo general en ambos casos, los homónimos publican con distintos subconjuntos de coautores, principio que fundamenta el método propuesto por Wooding et al¹¹. Destaca asimismo la baja sensibilidad de esta metodología a la movilidad de los autores. Con frecuencia un autor se traslada a centros en los que previamente ha colaborado con otro autor, de modo que este sirve de «rastros» para identificar la producción del primero en su nuevo centro de filiación. En los casos en que no se dispone de este rastro, el hecho de que las publicaciones reflejen el tema y la filiación de los autores en los 3-5 años previos¹⁹, la información en internet identifica los cambios de filiación, pues generalmente está más actualizada. En estos casos es igualmente útil disponer de la información del currículum vitae del autor en cuestión. Entre las limitaciones de la metodología, destaca la detección de documentos publicados en solitario. Asimismo el método resulta ineficaz cuando se utilizan coautores homónimos como criterio de selección de publicaciones. Este sería el caso de firmas como: Rodríguez, A; Martínez, A, o Sánchez, A, que entre 2006 y 2008 registraron más de 200 documentos cada una en las bases de datos de Thomson-Reuters²⁰.

Las publicaciones atribuibles a los grupos de investigación recibieron mayor visibilidad que las de la colección cardio-cerebrovascular y que el conjunto de biomedicina. Datos en la bibliografía indican que este no es un fenómeno aislado; actualmente los grupos de investigación dominan la producción bibliográfica de mayor visibilidad en la mayor parte de las disciplinas científicas^{1,21,22}. Con todo, los grupos sólo explicaron la mitad de la producción del área, lo que indica que el método de detección aplicado fue restrictivo en la selección de los integrantes de grupo. Según esta hipótesis, los grupos detectados representarían a los núcleos de investigadores más cohesionados por sus prácticas de publicación, que además serían responsables del grueso de la producción en el área, conjunto al que se sumarían las aportaciones de disciplinas científicas cercanas a la cardio-cerebrovascular.

La producción en colaboración internacional atribuible a los grupos representó algo menos de una quinta parte del total, mientras que en el ámbito de la biomedicina durante el mismo periodo fue una tercera parte. Este bajo nivel de colaboración está en línea con lo observado en el subámbito de medicina clínica, que recoge las principales área de investigación clínica. Esta observación podría reflejar en parte el relativamente bajo interés en la colaboración internacional descrita por Bordons y Zulueta³.

Es deseable que futuros estudios dediquen esfuerzos a mejorar la resolución de los métodos de clasificación temática basados en términos MeSH e incluyan análisis diacrónicos que permitan observar cambios en el número y composición, así como en los temas de estudio de los grupos.

CONCLUSIONES

La investigación cardio-cerebrovascular del periodo analizado fue principalmente de carácter clínico, procedente del sector sanitario y concentrada en Cataluña y Madrid. Protagonizaron la colaboración regional los centros sanitarios y universitarios, probablemente como resultado de la frecuente doble filiación de los investigadores en esta área, si bien incrementar la colaboración internacional es una asignatura pendiente. La actividad atribuible a los grupos recibió mayor visibilidad tanto comparada con la producción del conjunto del área como con la del ámbito de biomedicina. Destacan por su alta productividad y visibilidad los grupos de primer nivel localizados en Cataluña y Madrid.

AGRADECIMIENTOS

Los autores agradecen la colaboración del Dr. Ginés Sanz Romero como experto externo en la elaboración del presente estudio. El Dr. Sanz es cardiólogo del *Hospital Clínic i Provincial de Barcelona* y director del Departamento de Investigación Cardiovascular Traslacional de Nuevas Tecnologías y Terapias del Centro Nacional de Investigaciones Cardiovasculares (CNIC).

FINANCIACIÓN

Estudio realizado por iniciativa y con la financiación del CNIC-ISCIH.

CONFLICTO DE INTERESES

Ninguno.

BIBLIOGRAFÍA

1. Wuchty S, Jones FJ, Uzzi B. The increasing dominance of teams in production of knowledge. *Science*. 2007;316:1036-9.
2. Méndez-Vásquez RI, Suñén-Pinyol E, Cervelló R, Camí J. Mapa bibliométrico de España 1996-2004: biomedicina y ciencias de la Salud. *Med Clin (Barc)*. 2008;130:246-53. Informe previo publicado [citado 20 Oct 2008] y disponible en: <http://bibliometria.prbb.org/MapaBiomedico2004>
3. Bordons M, Zulueta MA. La interdisciplinariedad en los grupos españoles de investigación en el área cardiovascular. *Rev Esp Cardiol*. 2002;55:900-12.
4. Valderrama-Zurián JC, González-Alcaide G, Valderrama-Zurián FJ, Aleixandre-Benavent R, Miguel-Dasitc A. Redes de coautorías y colaboración institucional en REVISTA ESPAÑOLA DE CARDIOLOGÍA. *Rev Esp Cardiol*. 2007;60:117-30.
5. Seglen PO, Aksnes DG. Scientific productivity and group size: A bibliometric analysis of norwegian microbiological research. *Scientometrics*. 2000;49:125-43.
6. Cohen JE. Size, age and productivity of scientific and technical research groups. *Scientometrics*. 1991;20:395-416.
7. Torvik VI, Weeber M, Swanson DR, Smalheiser NR. A probabilistic similarity metric for medline records: a model for author name disambiguation. *J Am Soc Inf Sci Technol*. 2005;56:140-58.
8. Costas R, Bordons M. Algoritmos para solventar la falta de normalización de nombres de autor en los estudios bibliométricos. *Investigación bibliotecológica*. 2007;21:13-32.
9. Soler JM. Separating the articles of authors with the same name [citado 20 Oct 2008]. Disponible en: <http://arxiv.org/abs/cs/0608004v1>
10. Galvez C, Moya-Anegón F. Approximate personal name-matching through finite-state graphs. *J Am Soc Inf Sci Technol*. 2007;58:1960-76.
11. Wooding S, Wilcox-Jay K, Lewison G, Grant J. Co-author inclusion: A novel recursive algorithmic method for dealing with homonyms in bibliometric analysis. *Scientometrics*. 2006;66:11-21.
12. Calero C, Buter R, Cabello Valdés C, Noyons E. How to identify research groups using publication analysis: an example in the field of nanotechnology. *Scientometrics*. 2006;66:365-76.
13. Wasserman S, Faust K. *Social network analysis*. Oxford: Cambridge University Press; 1994.
14. Girvan M, Newman MEJ. Community structure in social and biological networks. *Proc Natl Acad Sci U S A*. 2002;99:7821-6.
15. Medical Subject Heading Terms [citado 20 Oct 2008]. Disponible en: <http://www.ncbi.nlm.nih.gov/sites/entrez?db=mesh>
16. Hirsch JE. An index to quantify an individual's scientific research output. *Proc Natl Acad Sci U S A*. 2005;102:16569-72.
17. Van Raan AF. Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgement for 147 chemistry research groups. *Scientometrics*. 2006;67:491-502.
18. Aleixandre-Benavent R, Alonso-Arroyo A, Chorro-Gascó FJ, Alfonso-Manterola F, González-Alcaide G, Salvador-Taboada MJ, et al. La producción científica cardiovascular en España y en el contexto europeo y mundial (2003-2007). *Rev Esp Cardiol*. 2009;62:1404-17.
19. Grant J, Lewison G. Government funding of research and development. *Science*. 1997;278:878-9.
20. Méndez-Vásquez RI. Estar o no estar en el asunto: la evaluación individual del rendimiento científico. *Aten Primaria*. 2009;41:63-6.
21. Whitfield J. Group theory. *Nature*. 2008;455:720-3.
22. Teamwork. *Nat Genet*. 2009;41:1.