

Original article

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

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

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

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

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

Article history:

Received 4 August 2011

Accepted 13 February 2012

Available online 23 May 2012

Keywords:

Coauthorship analysis

Author-name disambiguation

Research groups

Bibliometrics

Cardio-cerebrovascular

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.

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

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

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 tesoro *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.

Palabras clave:

Análisis de coautoría

Normalización de autores

Grupos de investigación

Bibliometría

Cardio-cerebrovascular

* Corresponding author: Bibliometría, Fundació Parc de Recerca Biomèdica de Barcelona, Dr. Aiguader 88, 4.ª planta, 08003 Barcelona, Spain.

E-mail address: rmendez@prbb.org (R.I. Méndez-Vásquez).

Abbreviations

CSIC: Consejo Superior de Investigación Científica

JCR: Journal Citation Reports

MeSH: Medical Subject Headings

INTRODUCTION

While today we generally accept that research is primarily a collective activity,¹ little has been reported on the scientific production of research groups. This contrasts sharply with the numerous bibliometric accounts of Spain's autonomous regions, institutional sectors, and centers in the fields of biomedicine and the health sciences.² Two previous studies demonstrate the difference between the administrative and functional concepts of the research group. Bordons and Zulueta³ analyzed interdisciplinary aspects of Spanish cardiology groups gathering data on group size and composition via surveys to researchers on this topic. Ie, they analyzed data reported to them. In the second study, Valderrama-Zurián et al.⁴ analyzed the coauthorship networks revealed by publications in *Revista Española de Cardiología*, identifying 25 distinct groupings; ie, they used research results to gather information on the researchers who had produced them. However, neither the isolation of groups nor their bibliometric evaluation were the objective of these studies, hence the information available on the topic remains limited. In addition to the aforementioned difference between the administrative concept emphasizing shared resources and institutional affiliation and the functional concept emphasizing the frequency and regularity with which groups produce results,^{5,6} the bibliometric characterization of groups forces us to attribute publications to their respective authors through a process termed “author-name disambiguation.” This represents a challenge because we need to handle a substantial volume of publications to obtain representative samples and because no exact correspondence exists in databases between the real authors (researchers) and bibliographic signatures. This gap is caused by the presence of identical names (homonyms) that erroneously bring together the production of different authors, and also because individual authors may appear under different bibliographic signatures (synonyms). Coauthorship analysis quantifies the frequency with which authors (bibliographic signatures) coincide in publications, making it the principle means of isolating groups on the basis of their results. Specific computer algorithms enable us to handle large volumes of publications, although homonyms and synonyms do frequently distort and invalidate results. Several automated approaches aimed at minimizing distortion have been trialled: probabilistic methods,^{7–9} finite state graphs,¹⁰ recursive algorithms,¹¹ or combining names and institutions¹²; however, the results have not been wholly satisfactory. The results of coauthorship analysis are analyzed by presenting authors as nodes and drawing connections between them to show how they coincide in a specific publication.¹³ Girvan and Newman.¹⁴ studied collaboration networks of researchers, combining this with their scientific journal publications. They showed they could isolate highly cohesive groups with loose interconnections—a property they called “community structure”.

To overcome the lack of bibliometric information, the present study presents a map of specialist, cardio-cerebrovascular research groups in Spain that is based on the functional concept of the research group, ie, the group as defined by its results—in this case, scientific journal publications. Accordingly, we consider any nucleus of researchers who regularly coauthor scientific studies on a given topic to be a stable group.⁶

METHODS

Our semi-automated, cyclical approach combines coauthorship analysis with author-name disambiguation of bibliographic signatures to gradually identify more and more authors and their production in each cycle. We summarize this in 7 stages (Figure) that were repeated until they received the approval of our expert.

Detecting Research Groups

Stage 1, Source of Data and the Definition of the Study Collection

Our data source was the National Citation Report for Spain database. This is a Thomson-Reuters product that includes documents published from Spain in all fields of science for 1981–2004. The cardio-cerebrovascular collection was selected by using a 3106-term filter, drawn from the US National Library of Medicine (Medical Subject Headings [MeSH]) thesaurus.¹⁵ The filter identified 6540 publications (the cardio-cerebrovascular collection) with at least one filter term in the title and/or keywords and dated 1996–2004.

Stage 2, Pre-selection of Publications and Bibliographic Names or Authors

Groups were isolated on the basis of 2 publication subsets: a) those authored by researchers with $\geq 50\%$ of their production in the cardio-cerebrovascular collection (50% filter), and b) those authored by researchers with ≥ 10 documents in the collection. The second cutoff point was chosen because of the distribution of the number of documents per author in the collection.

Stage 3, Coauthorship Network

Based on the stage 2 selection, we constructed networks in which bibliographic signatures were represented as vertices and their coauthorship relation (the co-occurrence of two different bibliographic name in one publication) as edges. We calculated how often each vertex appeared in the collection (number of

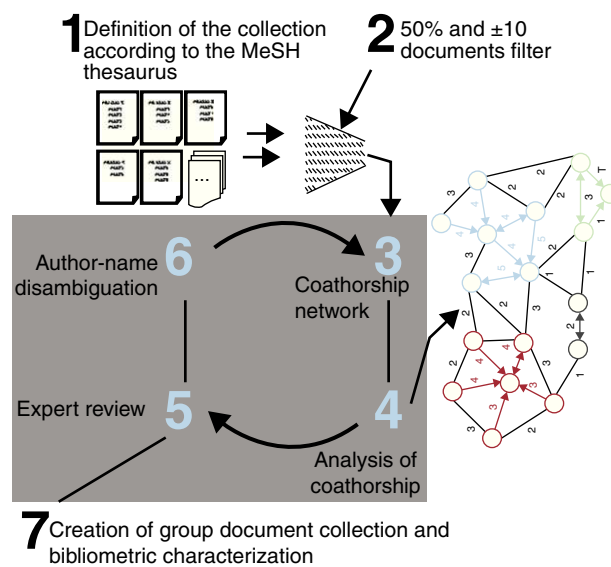


Figure. Breakdown of stages in detecting research groups. MeSH, Medical Subject Headings.

documents) and how often the two names involved coincided (number of coauthors). Coauthorship relationships in the network were considered symmetrical; ie, we took no account of directionality.

Stage 4, Coauthorship Analysis

We used our own algorithm to compare all nodes in the stage 3 network with their respective neighbors, forming groups around the nearest, most important neighbor. If necessary, to ensure authors were ascribed to one group only, we applied 4 criteria in sequence: *a*) number of coauthored publications; *b*) number of neighbor's publications; *c*) number of neighbor's coauthors, and *d*) local density around the neighbor. The algorithm used stage 2 documents, filtered by a minimum of 4 parameters: *a*) number of authors per group; *b*) number of documents per author to form a group, *c*) number of coauthored documents to form a group, and *d*) number of coauthors per author to form a group. To ensure the stability and plausibility of our results, we set these values at 4, 4, 3, and 3, respectively.

Stage 5, Expert Review

A healthcare expert with a background in scientific research in cardiology participated in the present study. At this stage, the expert made a qualitative evaluation of the groups detected in each cycle and reported an assessment of the plausibility of each one. This guided the subsequent author-name disambiguation stage.

Stage 6, Manual Author-name Disambiguation of Bibliographic Names

We first combined author-name variants (unifying synonyms) and then separated out the publications of different authors who had erroneously been grouped together under a single bibliographic name in the data source (separating homonyms). Following stage 4, synonymous bibliographic signatures had been gathered together in 1 group with their most frequent coauthors. Hence, their disambiguation required only 1 "author" field in the database and the corresponding synonymous bibliographic signatures were associated with this. These "author" entries were created using the author's full name and affiliation as given in scientific publications and/or the most recent institutional documents located by searching for pairs of bibliographic signatures (coauthors) using Google Scholar and Google. Homonyms were separated manually by selecting the publications in which a given bibliographic name appeared with a specific coauthor (generally the most frequent one), in a manner similar to that proposed by Wooding et al.¹¹ The publications separated out by this means were associated with the corresponding "author" entry in the study database.

Stage 7, Group Documents

Once the expert had given his approval, we created a "group documents" collection assigning to each group those documents for which at least 1 group member was named as author.

Topic-based Classification

The cardio-cerebrovascular documents were classified on an *ad hoc* basis—developed under the supervision of the expert—that grouped the 3106 MeSH terms into 23 categories. The list of terms by category is available in the online report of the present study.

We also used the *Journal Citation Reports* (JCR) 2004 classification as provided by Thomson-Reuters for the National Citation Report for Spain.

Disambiguation of Research Center Addresses and Sector-based Classification

Center-name variants appearing in publications were unified into a single name. This enabled us to attribute the publications analyzed to all the organizations in line with the total assignment method. Hence, documents were associated with all authors' centers of affiliation, facilitating analysis of collaboration between centers, regions, and institutional sectors. However, associating a given document to more than 1 center meant the document count per center, region or sector was greater than the actual total because documents assigned to more than 1 center were counted more than once. Implicitly, documents attributed to more than 1 center were considered a product of collaboration between centers and, by analogy, between regions and institutional sectors.

Center names were taken from the 2006 Spanish National Catalog of Hospitals (*Catálogo Nacional de Hospitales 2006*) and the National Registry of Universities, Centers and Teaching (Registro Nacional de Universidades, Centros y Enseñanzas). For other organizations, we used the name that appeared on official web pages or in accredited directories. Similarly, we obtained the full postal address of each organization, enabling us to ascertain its geographic location. The centers identified during the disambiguation process were classified in 5 institutional sectors on the basis of their legal status or the nature of their activity. The university sector included universities and centers in their orbit, such as university schools and institutes. The healthcare sector included public and private hospitals, research centers closely involved in clinical research, and other centers such as tissue banks and diagnostic imaging facilities, as well as primary care centers. The public research institution sector (PRI) included centers belonging to the Spanish Higher Science Research Council (*Consejo Superior de Investigación Científica* [CSIC]), the *Instituto de Salud Carlos III* (ISCIII), and public research centers in the autonomous regions. The Administration, NGO and others sector included state and regional centers, NGOs, and scientific associations. The business sector consisted principally of pharmaceutical companies.

Bibliometric Characterization

Bibliometric analysis was limited to the citable documents (articles, reviews and proceedings papers) in the cardio-cerebrovascular and research group document collections, and included bibliometric indicators of: *a*) activity: number of documents (Docs); *b*) visibility: number of citations, mean citations per document (CD), the percentage of documents not cited during the study period (%NC) and its relationship with the weighted mean of cardio-cerebrovascular citations in Spain (MCE), calculated by dividing document CD by the average cardio-cerebrovascular collection CD for the year of publication; MCE values >1 indicated they received more citations than the mean for the topic area in Spain during the study period (as an experiment, we calculated the Hirsch-index¹⁶ of the group [H-index] during the study period¹⁷), and *c*) indicators of collaboration: the percentage of publications involving international collaboration (%Int), including all documents with at least 1 author affiliated with a center outside of Spain, and the percentage of collaboration between autonomous regions (%Reg).

Breakdown

We calculated indicators for: a) autonomous regions; b) institutional sectors; c) research centers; d) research groups, and e) scientific subfields. This breakdown meant we had to tentatively ascribe groups to the autonomous region, center, and topic subfield that most frequently appeared in their documents (principle ascription) and that appeared with at least 80% of that frequency (secondary ascription). We considered that an agent (autonomous region, center, group, etc.) had greater visibility when it simultaneously presented CD, MCE and %Int values above, and %NC values below, the reference mean.

RESULTS

More detailed, broader-reaching results of the present study are available at <http://193.145.216.56/GruposCardioCerebrovascular>.

Description of the Cardio-Cerebrovascular Collection, Spain, 1996-2004

Our method retrieved 6540 documents (the cardio-cerebrovascular collection, Spain), 63% were authored by researchers affiliated with centers in Madrid and Catalonia. The visibility of this subgroup of publications was above the mean for the collection (Table 1). Collaboration between autonomous regions was present in 12.6% of the collection, whereas 61.3% involved collaboration between sectors. The sector most involved in collaborative publications was healthcare and interaction was most frequent between university centers, followed by collaboration between the university and PRI sectors (Table 2). In contrast,

healthcare sector publications involving international collaboration amounted to 17%, a figure lower than that of other institutional sectors which, as a group, reached 23% (Table 3).

The most active ad hoc MeSH term area was Clinical cardiology. However, the most visible ad hoc areas were Coagulation, platelets and thrombosis, Cardiovascular pharmacology, and Syncope (Table 4). The cardio-cerebrovascular collection was drawn from 122 different JCR disciplines; Cardiovascular system brought together the greatest number of publications (one third of the total). *Revista Española de Cardiología* was the most frequent of the 1020 journals analyzed. For reasons of space we have omitted production distribution by JCR discipline and by journal. This information is available online.

Detection and Bibliometric Description of Research Groups

The process of isolating groups concluded after 12 cycles and produced 93 groups made up of 772 different authors (mean 8.3 researchers); 28.0% were women. This last finding contrasts with the percentage (42.1%) observed in the field of psychiatry (see the bibliometric map of groups in psychiatry, *Mapa bibliométrico de grupos en psiquiatría*, at: <http://bac.fundaciorecerca.cat/psiquiatria>). Eleven research groups showed percentages of documents that were very low in comparison with their respective National Citation Report totals so their indicators may well not represent their true level of activity.

The 93 groups identified accounted for 51.5% of the documents and 57.8% of the citations of the cardio-cerebrovascular collection; as a whole they presented CD, MCE, and %Reg values above those of the collection. In contrast, the %NC and %Int values were statistically lower. A comparison with indicators for biomedicine showed similar results (Table 5).

Table 1
Production in Cardio-cerebrovascular Research by Autonomous Region, Spain, 1996-2004

| Autonomous region | Docs ^a | %Docs ^b | %Acum ^c | Citations ^d | CD ^e | MCE ^f | %NC ^g | %Int ^h |
|--------------------------------|-------------------|--------------------|--------------------|------------------------|-----------------|------------------|------------------|-------------------|
| Catalonia | 2133 | 32.6 | 32.6 | 22 660 | 10.6 | 1.100 | 25.1 | 27.0 |
| Community of Madrid | 2003 | 30.6 | 63.2 | 21 398 | 10.7 | 1.080 | 29.7 | 24.9 |
| Andalusia | 720 | 11.0 | 74.3 | 4328 | 6.0 | 0.660 | 34.0 | 15.4 |
| Valencian Community | 711 | 10.9 | 85.1 | 4126 | 5.8 | 0.730 | 34.9 | 21.7 |
| Galicia | 387 | 5.9 | 91.0 | 2607 | 6.7 | 0.830 | 28.4 | 16.8 |
| Chartered Community of Navarre | 254 | 3.9 | 94.9 | 1788 | 7.0 | 0.950 | 26.8 | 13.8 |
| Castile and León | 247 | 3.8 | 98.7 | 1909 | 7.7 | 0.660 | 33.6 | 11.7 |
| Region of Murcia | 217 | 3.3 | 102.0 | 1246 | 5.7 | 0.550 | 33.2 | 12.9 |
| Basque Country | 199 | 3.0 | 105.1 | 1672 | 8.4 | 0.870 | 31.7 | 15.1 |
| Aragon | 172 | 2.6 | 107.7 | 1023 | 6.0 | 0.590 | 32.6 | 24.4 |
| Canary Islands | 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 |
| Principality of Asturias | 109 | 1.7 | 113.3 | 950 | 8.7 | 0.750 | 29.4 | 14.7 |
| Castile-La Mancha | 98 | 1.5 | 114.8 | 406 | 4.1 | 0.580 | 35.7 | 10.2 |
| Extremadura | 91 | 1.4 | 116.2 | 1587 | 17.4 | 2.090 | 44.0 | 34.1 |
| Balearic Islands | 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 | 6540 | | | 55 519 | 8.5 | 0.910 | 30.3 | 22.7 |

^a Number of documents.

^b Percentage of documents with respect to the total for the topic area.

^c Cumulative percentage of documents.

^d Number of citations received in the period 1966-2004.

^e Mean number of citations per document.

^f Relationship with the weighted mean in Spain for the cardio-cerebrovascular field.

^g Percentage of documents not cited in the study period.

^h Percentage of documents published in international collaboration.

Table 2

Percentages of Documents Published in Collaboration Between Sectors in the Cardio-Cerebrovascular Field, Spain, 1996-2004

| | University | PRI | ADM, NGO | Business | Total |
|---------------------|------------|-------|----------|----------|-------|
| Healthcare | 89.58 | 6.29 | 1.87 | 2.26 | 89.31 |
| University | 0.00 | 73.16 | 10.17 | 16.67 | 10.20 |
| PRI | 0.00 | 0.00 | 50.00 | 50.00 | 0.46 |
| ADM, NGO and others | 0.00 | 0.00 | 0.00 | 100.00 | 0.03 |
| Business | | | | | 0.00 |

ADM, Administration; PRI, public research institutions.

Relation Between Group Size and Bibliometric Indicators

The number of members per group-group size-was directly related to the volume of group documents ($\rho=0.770$; $P<.01$), the number of citations received ($\rho=0.500$; $P<.01$), and the H-index ($\rho=0.480$; $P<.01$). We found no direct relationship between group size and CD, %Int, or MCE.

Table 3

Production in the Cardio-Cerebrovascular Field by Institutional Sectors, Spain, 1996-2004

| Institutional sector | Docs ^a | %Docs ^b | %Acum ^c | Citations ^d | CD ^e | MCE ^f | %NC ^g | %Int ^h |
|----------------------|-------------------|--------------------|--------------------|------------------------|-----------------|------------------|------------------|-------------------|
| Healthcare | 5175 | 79.1 | 79.1 | 42 745 | 8.3 | 0.870 | 31.7 | 17.3 |
| University | 3859 | 59.0 | 138.1 | 27 327 | 7.1 | 0.860 | 30.4 | 21.1 |
| PRI | 487 | 7.4 | 145.6 | 7022 | 14.4 | 1.350 | 18.5 | 38.2 |
| Business | 128 | 2.0 | 147.5 | 1457 | 11.4 | 1.040 | 18.8 | 30.5 |
| ADM, NGO and others | 101 | 1.5 | 149.1 | 1171 | 11.6 | 1.160 | 18.8 | 23.8 |
| Total | 6540 | | | 55 519 | 8.5 | 0.910 | 30.3 | 22.7 |

ADM, Administration; PRI, public research institutions.

^a Number of documents.

^b Percentage of documents with respect to the total for the topic area.

^c Cumulative percentage of documents.

^d Number of citations received in the period 1966-2004.

^e Mean number of citations per document.

^f Relationship with the weighted mean in Spain for the cardio-cerebrovascular field.

^g Percentage of documents not cited in the study period.

^h Percentage of documents published in international collaboration.

Table 4

Production in the Cardio-Cerebrovascular Field by Medical Subject Headings Areas, Spain, 1996-2004^a

| Topic area | Docs ^b | %Docs ^c | Citations ^d | CD ^e | MCE ^f | %NC ^g | %Int ^h |
|---------------------------------------|-------------------|--------------------|------------------------|-----------------|------------------|------------------|-------------------|
| Clinical cardiology | 2335 | 35.7 | 21 302 | 9.1 | 0.990 | 30.0 | 22.4 |
| Coagulation, platelets and thrombosis | 1740 | 26.6 | 19 465 | 11.2 | 1.030 | 25.1 | 29.1 |
| Ischemic heart disease | 1339 | 20.5 | 14 737 | 11.0 | 1.090 | 28.5 | 22.4 |
| Diagnostic techniques | 1304 | 19.9 | 9719 | 7.5 | 0.900 | 32.1 | 20.1 |
| High blood pressure | 1174 | 18.0 | 12 524 | 10.7 | 1.020 | 29.1 | 19.2 |
| Cerebrovascular disease | 1077 | 16.5 | 10 783 | 10.0 | 0.990 | 35.2 | 16.7 |
| Vascular research | 823 | 12.6 | 6730 | 8.2 | 0.960 | 30.7 | 23.2 |
| Arrhythmia | 631 | 9.6 | 4896 | 7.8 | 0.970 | 32.6 | 22.2 |
| Cardiovascular surgery | 589 | 9.0 | 5303 | 9.0 | 0.940 | 29.2 | 21.2 |
| Cardiovascular pharmacology | 552 | 8.4 | 6305 | 11.4 | 1.090 | 29.5 | 23.2 |
| Vascular surgery | 470 | 7.2 | 6534 | 13.9 | 1.240 | 31.1 | 29.6 |
| Valvular heart disease. | 393 | 6.0 | 1379 | 3.5 | 0.740 | 33.8 | 11.2 |
| Syncope | 103 | 1.6 | 1284 | 12.5 | 1.350 | 27.2 | 24.3 |
| Total | 6540 | 194.2 | 55 519 | 8.5 | 0.910 | 30.3 | 22.7 |

^a Topic subareas with ≥ 100 documents; a full list is available in the online report at: <http://193.145.216.56/GruposCardioCerebrovascular>.

^b Number of documents.

^c Percentage of documents with respect to the total for the topic area.

^d Number of citations received in the period 1966-2004.

^e Mean number of citations per document.

^f Relationship with the weighted mean in Spain for the cardio-cerebrovascular field.

^g Percentage of documents not cited in the study period.

^h Percentage of documents published in international collaboration.

Analysis by Location

Together, Catalonia and Madrid accounted for $>50.0\%$ of the groups, and these authored 74.0% of group documents. One research group was ascribed to 2 centers located in Madrid and the Castile-La Mancha autonomous region, adding 94 to the count for the autonomous regions. The Region of Murcia presented the highest mean per group; Galicia and the Chartered Community of Navarre had the lowest. Documents attributed to groups in Catalonia, Madrid, and Extremadura (1 group) had visibility above the mean of the groups as a whole. Collaborative publications accounted for 8.6% of group documents. No research groups were identified in the Basque Country, ranked ninth among the autonomous regions for volume of production in the field (Table 6).

Sector-Based Analysis

The healthcare sector accounted for 74.2% of groups and 80.0% of group documents; 28.0% of groups and 30.4% of documents were

Table 5
Comparison of Bibliometric Indicators, Cardio-Cerebrovascular Research in Spain, 1996–2004

| | Docs ^a | Citations ^b | CD ^c | MCE ^d | %NC ^e | %Reg ^f | %Int ^g |
|-----------------------------------|-------------------|------------------------|-------------------|--------------------|-------------------|-------------------|-------------------|
| Research groups | 3365 | 32 086 | 9.54 ^h | 1.123 ^h | 26.2 | 15.0 ^h | 18.4 |
| Cardio-cerebrovascular collection | 6540 | 55 519 | 8.49 | 1.037 | 30.3 ^h | 12.6 | 22.7 ^h |
| Biomedicine (1996–2004) | 84 122 | 719 127 | 8.55 ^h | 1.020 ^h | 27.2 ^h | 12.6 ^h | 27.1 ^h |

^a Number of documents.^b Number of citations received in the study period.^c Mean number of citations per document.^d Relationship with the weighted mean in Spain for the cardio-cerebrovascular field and according to the *Journal Citation Reports* disciplines in the field of biomedicine.^e Percentage of documents not cited in the study period.^f Percentage of documents published in collaboration between autonomous regions.^g Percentage of documents published in international collaboration.^h Statistically significant differences ($P < .05$).

from the university sector. Healthcare also presented the highest mean number of group members (9.1). The other groups were distributed as follows: 3 in PRI sector centers (2 in the *Centro de Investigaciones Biológicas*, CSIC, Madrid, and 1 in the *Instituto de Investigaciones Biomédicas de Barcelona*, CSIC, Barcelona); 1 in the business sector (J. Uriach & Cía); and 1 in the Administration sector (*Institut d'Estudis de la Salut*, Barcelona). Groups simultaneously ascribed to a hospital and a university represented 10.8% of the total; they authored 15.1% of all documents published in collaboration between these two sectors. The PRI sector documents presented the highest CD and MCE values in the study.

Medical Subject Headings Area and Journal Citation Reports Topic-based Analysis

Clinical cardiology, an ad hoc MeSH term-based discipline, accounted for more than half of the groups, documents, and citations. The highest mean group member figures were 10.3 and 10.1, respectively, in the MeSH areas of Cardiovascular surgery and Coagulation, platelets and thrombosis. Excluding those MeSH areas with <3 groups, the visibility of publications of the 8 vascular groups was above the mean for all the groups together (Table 7). The JCR

discipline of Cardiovascular system brought together more than half of the groups identified and of the total number of documents and citations. In the subgroup of JCR disciplines with >1 group, Hematology presented the highest mean number of group members and Peripheral vascular disease had the greatest visibility.

Analysis of Groups

Three groups, 2 at the *Hospital Clínic i Provincial de Barcelona* and 1 at the *Hospital Vall d'Hebron*, were simultaneously ranked in the first 3 positions for the activity and visibility indicators; these were clinical research groups in ischemic heart disease, arrhythmia and cardiovascular surgery. Three groups presented the highest MCE values (>2.0): 1 at the *Universidad Autónoma de Madrid*, 1 at the *CSIC Centro de Investigaciones Biológicas* in Barcelona, and 1 at the *Hospital Clínic i Provincial de Barcelona*. Three groups—at the *Universidad de Extremadura*, *Universidad de Zaragoza*, and the *Institut de Recerca Oncològica*, Barcelona—published more than half of their production in international collaboration. The 3 groups with the highest H-index values (>35) came from the *Hospital Clínic i Provincial* and the *Institut d'Investigacions Biomèdiques*

Table 6
Research Groups in the Cardio-cerebrovascular Field by Autonomous Regions, Spain, 1996–2004

| Ordinal | Autonomous region | Groups, no. (%) | MInt ^a | Docs ^b | %Docs ^c | %Acum ^d | Citations ^e | CD ^f | MCE ^g | %NC ^h | %Int ⁱ |
|---------|--------------------------------|-----------------|-------------------|-------------------|--------------------|--------------------|------------------------|-----------------|------------------|------------------|-------------------|
| 1 | Catalonia | 34 (36.6) | 8.5 | 1460 | 43.4 | 43.4 | 16 700 | 11.4 | 1.150 | 22.5 | 23.2 |
| 2 | Community of Madrid | 25 (26.9) | 8.8 | 1030 | 30.6 | 74.0 | 11 817 | 11.5 | 1.150 | 26.2 | 18.4 |
| 4 | Valencian Community | 11 (11.8) | 9.7 | 373 | 11.1 | 85.1 | 1918 | 5.1 | 0.640 | 33.5 | 9.9 |
| 3 | Andalusia | 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 | Chartered Community of Navarre | 4 (4.3) | 7.8 | 138 | 4.1 | 99.6 | 1269 | 9.2 | 1.040 | 22.5 | 12.3 |
| 7 | Castile and León | 2 (2.2) | 9.5 | 64 | 1.9 | 101.5 | 352 | 5.5 | 0.580 | 26.6 | 14.1 |
| 8 | Region of Murcia | 2 (2.2) | 13.5 | 133 | 4.0 | 105.4 | 706 | 5.3 | 0.610 | 32.3 | 4.5 |
| 13 | Principality of Asturias | 2 (2.2) | 5.0 | 40 | 1.2 | 106.6 | 243 | 6.1 | 0.710 | 27.5 | 10.0 |
| 10 | Aragon | 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 | 3365 | | 108.6 | 32 086 | 9.5 | 1.1230 | 26.2 | 18.4 |

^a Mean number of members per group.^b Number of documents attributable to the research groups.^c Percentage of documents with respect to the total number of group documents.^d Cumulative percentage of group documents.^e Number of citations of group documents between 1996 and 2004.^f Mean number of citations per group document.^g Relationship with the weighted mean of citations in Spain for the cardio-cerebrovascular field of the group documents.^h Percentage of documents not cited in the study period.ⁱ Percentage of group documents published in international collaboration.

Table 7
Research Groups in the Cardio-Cerebrovascular Field by Medical Subject Headings Areas, Spain, 1996–2004

| Topic area | Groups, no. (%) | MInt ^a | Docs ^b | %Docs ^c | Citations ^d | CD ^e | MCE ^f | %NC ^g | %Int ^h |
|---|-----------------|-------------------|-------------------|--------------------|------------------------|-----------------|------------------|------------------|-------------------|
| Clinical cardiology | 49 (52.7) | 9.2 | 1899 | 56.4 | 18 540 | 9.8 | 1.040 | 28.9 | 17.9 |
| Diagnostic techniques | 27 (29.0) | 8.2 | 990 | 29.4 | 5914 | 6.0 | 0.800 | 30.7 | 16.1 |
| Ischemic heart disease | 23 (24.7) | 10 | 1151 | 34.2 | 11 944 | 10.4 | 1.030 | 27.6 | 18.6 |
| Coagulation, platelets and thrombosis | 17 (18.3) | 10.1 | 788 | 23.4 | 6660 | 8.5 | 0.760 | 22.8 | 20.2 |
| High blood pressure | 11 (11.8) | 7.4 | 508 | 15.1 | 4956 | 9.8 | 1.190 | 23.8 | 15.7 |
| Cerebrovascular disease | 9 (9.7) | 7.6 | 375 | 11.1 | 3699 | 9.9 | 1.080 | 22.7 | 8.5 |
| Arrhythmia | 9 (9.7) | 6.6 | 346 | 10.3 | 3111 | 9.0 | 1.090 | 30.9 | 22.3 |
| Vascular research | 8 (8.6) | 8.5 | 207 | 6.2 | 2926 | 14.1 | 1.260 | 15.0 | 20.3 |
| Cardiovascular surgery | 7 (7.5) | 10.3 | 235 | 7.0 | 1486 | 6.3 | 0.660 | 36.6 | 12.3 |
| Vascular surgery | 7 (7.5) | 7.4 | 257 | 7.6 | 3577 | 13.9 | 1.090 | 30.7 | 21.0 |
| Cardiovascular pharmacology | 2 (2.2) | 9 | 87 | 2.6 | 463 | 5.3 | 0.640 | 35.6 | 11.5 |
| Valvular heart disease. | 2 (2.2) | 8 | 55 | 1.6 | 194 | 3.5 | 0.490 | 45.5 | 0.0 |
| Syncope | 1 (1.1) | 4 | 13 | 0.4 | 45 | 3.5 | 0.340 | 53.8 | 7.7 |
| Molecular biology | 1 (1.1) | 6 | 13 | 0.4 | 274 | 21.1 | 1.420 | 23.1 | 0.0 |
| Atherosclerosis, atherogenesis and lipids | 1 (1.1) | 6 | 13 | 0.4 | 274 | 21.1 | 1.420 | 23.1 | 0.0 |
| Total | 93 (187.1) | 8.3 | 3365 | | 32 086 | 9.5 | 1 | 26.2 | 18.4 |

^a Mean number of members per group.

^b Number of documents attributable to the research groups.

^c Percentage with respect to the total number of group documents.

^d Number of citations of group documents between 1996 and 2004.

^e Mean number of citations per group document.

^f Relationship with the weighted mean of citations in Spain for the cardio-cerebrovascular field of the group documents.

^g Percentage of documents not cited in the study period.

^h Percentage of group documents published in international collaboration.

August Pi i Sunyer, both in Barcelona. Bibliometric indicator reference values are in Table 8.

DISCUSSION

The present study describes the bibliographic production attributable to groups working in the field of cardio-cerebrovascular research by sector, topic, and geographical location.

Using MeSH terms to delimit the topic area meant we could classify more cardio-cerebrovascular documents than if we had

used the JCR disciplines. However, this approach did lead to a high degree of overlap because publications were classified according to whether or not they included the MeSH terms selected for the study. The Cardiovascular system—the JCR discipline bringing together the most documents—only retrieved one third of those in the study collection. Cardio-cerebrovascular collection documents were distributed over 122 JCR disciplines, showing the JCR's limitations when defining disciplines such as cardio-cerebrovascular.

Bibliographic production in the field, as well as that attributable to the groups, presented highly asymmetric location- and sector-based distributions. One finding typical of bibliometric studies is that only a few agents generate the greater part of the production and, generally, present greater visibility. Catalonia and Madrid, and the healthcare and university sectors, amass most of the production and groups detected in this area and, moreover, present greater visibility than the other actors. This coincides with results reported by Bordons and Zulueta,³ based on surveys of researchers in the cardiovascular field. The dominant position of Madrid and Catalonia in bibliographic production in cardio-cerebrovascular research is unchanged even after adjustment for the number of inhabitants in 2004, as recorded by Spanish National Statistics Institute. Following this adjustment, only the Chartered Community of Navarre surpasses them, although its production in absolute terms leaves it in sixth place.

In cardio-cerebrovascular research, collaboration between institutional sectors was more important than that between autonomous regions and international collaboration (60% vs 12% and 23%, respectively). This high level of knowledge flow is explained by the co-occurrence of healthcare centers and universities in >80% of publications in the field. These data should be analyzed in the light of the double affiliation phenomenon, frequent among clinical researchers based at university hospitals and which has been observed elsewhere.⁴ A detailed analysis of this highly relevant interaction would probably be useful. New

Table 8
Reference Values of the Groups' Bibliometric Indicators

| Indicator | Median | Q1 (P75) | IQR |
|------------------------|--------|--------------------|-------|
| Docs ^a | 38 | 59 | 38 |
| Citations ^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 |
| H-index ^h | 15 | 23 | 13 |

A full list of the bibliometric indicator quartiles is available in the online report. IQR, interquartile range or the difference between the 75th and 25th percentiles; Q1, first quartile.

^a Number of documents.

^b Number of citations.

^c Mean number of citations per document.

^d Relationship with the weighted mean in Spain for the cardio-cerebrovascular field.

^e Percentage of documents not cited in the study period.

^f The lower the %NC value, the more positive its significance, hence we give the value corresponding to the 25th percentile or third quartile.

^g Percentage of documents published in international collaboration.

^h H-index of the group for the period 1996–2004.

approaches should explore and design instruments to better define “inter-sector collaboration” and quantify it more precisely.

The groups detected are representative of the Spanish scientific community in cardio-cerebrovascular research. In the present study, we detected 93 groups by analyzing 6540 documents published from Spain over 9 years in more than 1000 journals and classified in 122 JCR disciplines. The groups came from 5 sectors: the healthcare and university sectors were the principle sources; 3 groups were identified at a CSIC center; 2 in a pharmaceutical company; and 1 Administration-managed center. The number of groups detected and their distribution by sector coincide with findings reported by Bordons and Zulueta.³ in 2002. As we gradually gather much-needed information on research groups, we will be better able to compare the size and bibliometric indicators of this community with its equivalent elsewhere in Europe, thus contributing to the strategic management of research potential in this field. If we consider that the volume of publications in cardiology—as defined by the JCR—covers an area smaller than that analyzed in the present study, Spain is ranked sixth in Europe and ninth in the world.¹⁸ These positions will surely improve as those research groups with the greatest potential are encouraged.

Our method of detecting research groups is new and robust. Unlike other methods, which are applied once only,¹¹ the present proposal consists of an iterative procedure combining author-name disambiguation with coauthorship analysis in selecting relevant authors and subsequently isolating research groups. Because bibliographic-name disambiguation and research-group detection are based on coauthorship frequency, the results of the method presented here depend solely on a thorough analysis of the bibliographic information collected. This is a direct function of the number of times the cycle of bibliographic-name disambiguation and group isolation is repeated.

Bibliographic-name disambiguation through coauthorship analysis overcomes the principle limitations of disambiguation by center, affiliation, or topic, which do not permit us to distinguish between homonymous authors working in the same center or discipline.^{6,8,10} In contrast, thorough, repeated coauthorship analysis enables us to differentiate between homonyms because, generally, homonymous authors publish with different coauthor subgroups, a principle underlying Wooding et al.'s proposed method.¹¹ Furthermore, our method's low sensitivity to author mobility is significant. Authors frequently move to centers where they have previously collaborated with other authors. This leaves a “trail” that helps us identify the author's production in their new center of affiliation. When no such trail exists, the fact that publications reflect authors' topics and affiliations from the previous 3–5 years¹⁹ means internet searches that can access more up-to-date sources can identify changes of affiliation. In these cases, access to the curriculum vitae of the author in question is equally useful. The lack of detecting solo-author publications is one of the outstanding limitations of this method. Similarly, it is inefficient when homonymous coauthors are used as the criterion for selecting publications. This would be the case for bibliographic signatures like: Rodríguez, A; Martínez, A, or Sánchez, A, who between 2006 and 2008 registered more than 200 documents each in the Thomson-Reuters databases.²⁰

Publications attributable to research groups presented greater visibility than those in the cardio-cerebrovascular collection and in biomedicine as a whole. Data in the literature indicate this is not an isolated phenomenon: research groups currently dominate the bibliographic production with greatest visibility in most scientific disciplines.^{1,21,22} In all, groups only account for half of the production in the field, which indicates that the method of detection applied was restrictive when selecting group members. According to this hypothesis, the groups detected would represent more cohesive nuclei of researchers linked by their publishing

practice who, moreover, would be responsible for the bulk of production in the field. The contributions of scientific disciplines closely related to the cardio-cerebrovascular field would be added to this.

Production involving international collaboration attributable to the groups represents less than one fifth of the total; in biomedicine, over the same period, it stood at one third. This low level of international collaboration coincides with that observed in the subfield of clinical medicine, which covers the principle areas of clinical research. This could partly reflect the relatively low interest in international collaboration described by Bordons and Zulueta.³

It is to be hoped that future studies will try to resolve the issues raised by MeSH term topic-based classification and include diachronic analyses that will enable us to observe changes in research group numbers, membership, and study topics.

CONCLUSIONS

Cardio-cerebrovascular research in Spain during the period analyzed was principally clinical, produced by the healthcare sector, and centered on Catalonia and Madrid. Regional collaboration was led by healthcare centers and universities, probably because of the commonly-found double affiliation of researchers in this field. International collaboration, however, remains little developed. Activity attributable to groups presented greater visibility when compared with production in the area as a whole and with the field of biomedicine. Top-class groups located in Catalonia and Madrid stand out for their high levels of productivity and visibility.

ACKNOWLEDGEMENTS

The authors wish to express their thanks to Dr Ginés Sanz Romero who acted as external expert in the preparation of the present study. Dr Sanz Romero is a cardiologist at the *Hospital Clínic i Provincial de Barcelona* and director of the Department of Translational Research in New Technologies and Therapies (*Departamento de Investigación Cardiovascular Traslacional de Nuevas Tecnologías y Terapias*) of the *Centro Nacional de Investigaciones Cardiovasculares* (CNIC).

FUNDING

This study was proposed and financed by the CNIC-ISCIH.

CONFLICTS OF INTEREST

None declared.

REFERENCES

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. Early published report [cited 20 Oct 2008] available from: <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-Dasit A. Redes de coautorías y colaboración institucional en REVISTA ESPAÑOLA DE CARDIOLOGÍA. *Rev Esp Cardiol*. 2007;60:117–30.
5. Seglen Per O, Aksnes Dag W. 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;9: 13 PM-32.
9. Soler JM. Separating the articles of authors with the same name [cited 20 Oct 2008]. Available from: <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 [cited 20 Oct 2008]. Available from: <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.