

IDENTIFICACIÓN Y ESTUDIO DE GRUPOS DE INVESTIGACIÓN A TRAVÉS DE INDICADORES BIBLIOMÉTRICOS

M. Ángeles Zulueta*, Aurello Cabrero**, Marfa Bordons**

Resumen: Se presenta un método para la identificación de los grupos de investigación activos en una determinada área o centro de investigación, a través de un análisis de co-autoría en las publicaciones científicas del área o centro objeto de estudio. La metodología expuesta se basa en programas de elaboración propia aplicados a una descarga de documentos del Science Citation Index. El proceso incluye tres etapas básicas: recuperación, descarga y normalización de los datos bibliográficos a estudiar; construcción del fichero de productividad de los autores; y delimitación de grupos a través de las frecuencias de co-autoría. La última etapa del proceso consiste en la caracterización de los grupos mediante indicadores bibliométricos: tamaño de grupo, producción, productividad, tasas de colaboración y especialización temática. Se ilustra el funcionamiento de la metodología mediante su aplicación a la producción española de una disciplina y de un centro de investigación en el período 1990-93.

Palabras clave: indicadores bibliométricos, grupos de investigación, publicaciones científicas, análisis de co-autoría.

Abstract: A methodology based on coauthorship analysis is shown for the identification of research teams in a given discipline or research centre. The method includes using in-house programs to analyse a collection of documents downloaded from the Science Citation Index database. Three basic stages can be distinguished in the process: retrieval, downloading and normalisation of bibliographic data, construction of the author productivity file, and delimitation of research teams through coauthorship analysis. As a final step, teams are characterised by means of bibliometric indicators: team size, production, productivity, collaboration rate and subject specialisation of groups. The Spanish production in both a discipline and a research centre, over the years 1990-1993, is analysed through the methodology to test its performance and main results.

Key-words: bibliometric indicators, research teams, scientific publications, co-authorship analysis.

1 Introducción

Los indicadores bibliométricos, basados en el análisis estadístico de datos cuantitativos procedentes de la literatura científica, constituyen en la actualidad una herramienta esencial para el estudio de la actividad investigadora (1, 2, 3). En los últimos años, se ha extendido el uso de los indicadores bibliométricos como complemento de

* Facultad de Documentación, Universidad de Alcalá.

** Centro de Información y Documentación Científica (CINDOC), CSIC. Madrid.

Correo-e: mbordons@cindoc.csic.es.

Recibido: 21-4-99.

otros indicadores científicos para analizar la situación de la investigación de un país, su evolución en el tiempo y su posición en el contexto internacional. La mayor parte de estos estudios se centran en el análisis de la producción de un país, de una disciplina científica o de una institución o centro de investigación, mientras que los estudios sobre grupos de investigación son menos comunes. Esto se debe en parte a la difusa definición del término «grupo», que no siempre se corresponde con la estructura formal de los departamentos, unidades o centros, y también a los problemas técnicos derivados de la aplicación de métodos estadísticos a pequeñas unidades como son los grupos.

Un grupo de investigación puede definirse como un colectivo de científicos que colaboran en el planteamiento y desarrollo de una investigación, compartiendo recursos materiales y económicos. La creciente especialización y complejidad de la investigación científica ha incrementado el papel de la colaboración entre autores y ha favorecido la creación de grupos de investigación, en los que se desarrolla el trabajo de forma colectiva para obtener resultados de forma más eficaz y eficiente. La relevancia adquirida por los grupos es tal, que han pasado a considerarse la mínima unidad del sistema investigador en muchas disciplinas. Los grupos son hoy una pieza clave en el desarrollo de la investigación en gran parte de las disciplinas científicas y tecnológicas, mientras que el trabajo individual prevalece en muchas áreas de las ciencias sociales y en las humanidades.

El interés de los estudios a nivel «micro» es diverso. Por un lado, el análisis de la actividad de los grupos a través de indicadores bibliométricos permite profundizar en el estudio de la estructura y dinámica del proceso investigador, analizado en lo que es su mínima unidad de funcionamiento y complementando a otros enfoques, como los procedentes de la sociología de la ciencia. Pero además, estos estudios pueden aportar datos de interés para la política científica de un país, como por ejemplo la identificación de los grupos de excelencia en una determinada área científica.

Existen en la literatura distintos estudios publicados que utilizan los indicadores bibliométricos para analizar la actividad de grupos de investigación (4, 5, 6). Sin embargo, hay que señalar que se aprecia una gran variabilidad en los criterios que se siguen para considerar un grupo de investigación. Los criterios más utilizados se basan en considerar como miembros de un grupo a los investigadores pertenecientes a un mismo departamento (7), los colaboradores en proyectos de investigación (8,9) o los co-autores en publicaciones científicas, pertenezcan o no al mismo centro (10). Entre las ventajas de esta última aproximación podemos apuntar el hecho de que no es necesario un conocimiento previo del conjunto a estudiar como ocurre en los otros métodos, en los que se precisa conocer la adscripción de los autores a un determinado departamento o a un proyecto de investigación para poder realizar los análisis.

En este trabajo se presenta una metodología para la identificación de grupos de investigación a través del análisis de co-autoría en publicaciones científicas extraídas de la base de datos Science Citation Index. Entre los trabajos más interesantes relacionados con este tema, podemos mencionar el trabajo de Stokes *et al.* (11) en el que identifican grupos de investigación en el campo de la DNApolimerasa, y el estudio de Peters *et al.* (12) en el que utilizan técnicas de análisis de conglomerados para identificar grupos de investigación de una universidad holandesa. El método seguido por Stokes no queda claramente definido en su trabajo, pero su base teórica es similar a la presentada en este artículo. La comparación entre el análisis de conglomerados y la metodología actual en una fase inicial de su desarrollo se presentó en un trabajo previo

(13), donde se evidenciaron las limitaciones del primer método y se puso de manifiesto las ventajas de perfeccionar la metodología propia.

La metodología expuesta en este artículo para la identificación y posterior caracterización bibliométrica de grupos de investigación se enmarca dentro de los estudios conducentes a analizar el sistema español de investigación. Se presenta un programa interactivo, capaz de identificar grupos de investigación atendiendo a la frecuencia de co-autoría en las publicaciones científicas, y que se aplica en dos niveles de análisis: grupos pertenecientes a una disciplina científica y grupos pertenecientes a un centro de investigación. El punto de partida es un conjunto de registros que corresponde a la producción científica de la disciplina o del centro objeto de estudio, y la red de autores se construye a partir de la frecuencia de co-autoría conjunta. Los procedimientos técnicos seguidos se explican en este trabajo, en el que se indican los principales problemas encontrados así como las soluciones propuestas.

2 Metodología

La comprensión de la metodología propuesta requiere asumir una serie de premisas básicas. Los grupos de investigación se definen en términos de co-autoría, es decir, los grupos están formados por aquellos autores que firman de forma conjunta un importante porcentaje de su producción, pero no se corresponden necesariamente con una determinada estructura administrativa o institucional. Se ha definido como «investigador principal» (IP) al autor más productivo dentro de un grupo. Aunque el «investigador principal» así definido no siempre coincide con el líder real del grupo, la coincidencia entre autor más productivo y líder grupal es muy frecuente (14, 15). De hecho, la persona que asume el liderazgo de un grupo con frecuencia participa en casi todos los resultados científicos obtenidos por el mismo.

La metodología que se muestra se basa en una serie de programas realizados «ad hoc» en DbaseIV. Estos programas interactúan entre sí y con los ficheros de datos bibliográficos que contienen las publicaciones del centro o disciplina que se desea analizar. La identificación de los grupos se realiza de forma interactiva, de forma que es posible especificar en cada ocasión, y variar de una vez a otra, las condiciones requeridas para la delimitación de los grupos. Una vez identificados los equipos, se crea un fichero final que contiene la composición de cada grupo (IP y miembros) y la producción del grupo y de cada uno de sus autores.

2.1 Etapas en la identificación de grupos

Los datos bibliográficos de partida requieren un tratamiento en parte manual y en parte automático atendiendo a los siguientes pasos.

1.º Recuperación y descarga de los datos bibliográficos objeto de estudio de la base de datos Science Citation Index (SCI), versión CD-ROM, y posterior conversión a una base de datos gestionada por DbaseIV. La base de datos resultante incluye los datos bibliográficos de cada documento, así como información adicional que se añade

por procedimientos semiautomáticos. Entre estos últimos datos se incluyen el factor de impacto de las revistas en el año de publicación de los documentos y el nivel de investigación (básico-aplicado) de la revista (16). Se añaden, además, los códigos de identificación de los centros firmantes de los documentos, atendiendo a un sistema de codificación semiautomática desarrollado en el CINDOC (17), a través del cual se asigna a cada centro un código que incluye tres niveles de información: provincia, sector institucional y centro de investigación. Se soluciona así el problema de la falta de normalización de las direcciones que aparecen en los registros del SCI, que dificultan el procesamiento automático de la información.

2.º Construcción del fichero de productividad de los autores, que incluye para cada autor su dirección y su frecuencia de publicación en el período estudiado.

Para identificar la dirección de cada autor se ha seguido el siguiente proceso:

- a) Para registros con una única dirección, se asigna la dirección correspondiente a cada uno de los autores del documento.
- b) Para registros con dos direcciones, si se han identificado todos los autores menos uno, por medio del paso anterior, como pertenecientes a uno de los centros firmantes, al autor restante se le asigna la segunda dirección.

Todos estos autores identificados van a formar parte de un fichero que adjudica a cada autor el código del centro al que pertenece.

En la construcción de este fichero se han detectado una serie de problemas que se detallan a continuación.

- a) Falta de normalización en los nombres de los autores. Un único autor puede aparecer registrado con diferentes nombres en la base de datos utilizada. Para solucionar este problema se ha desarrollado el siguiente procedimiento:
 - Control de nombres de autores en orden alfabético para detectar cambios producidos en las iniciales del nombre. Ej: Contreras L = Contreras LA.
 - Búsqueda de posibles variantes para un mismo nombre compuesto. Ej: Ruizcabello A = Cabello AR

La reunión de todos los autores pertenecientes a un mismo grupo facilita la identificación de todas las variantes que corresponden a un mismo nombre, que aparecen agrupadas en torno al IP del grupo.

Cerca de un 2% de los nombres de los autores puede corregirse por este procedimiento.

- b) Diferentes autores pueden tener el mismo nombre. Este es el caso de algunos apellidos muy comunes tales como García, Gómez, Pérez, etc. Para obviar en lo posible este problema, cada autor se identifica a través de la cadena «nombre autor-centro de trabajo», con lo que se reduce considerablemente la posibilidad de que se confundan autores distintos con un mismo nombre. Sin embargo, el hecho de considerar cada autor ligado a su centro de trabajo ocasiona algunos problemas ligados a la posible movilidad de los investigadores, ya que su producción estará dividida según hayan firmado como pertenecientes a uno u otro centro. No obstante, este sistema aporta más ventajas que inconvenientes a la hora de identificar a los autores, y además un cambio de dirección con frecuencia suele ir acompañado de un cambio de grupo.

3.º Delimitación de grupos. Se realiza a través de un programa que ordena todos los autores de mayor a menor producción. El punto de partida es el autor más productivo del fichero, y a partir de éste, el programa identifica y selecciona a los restantes componentes del grupo, atendiendo a su alta frecuencia de co-autoría con el investigador principal. Todos los autores que publican un alto porcentaje de su producción científica en colaboración con un determinado investigador principal son considerados miembros de su grupo. Todos los autores, salvo los IP, pueden ser adscritos a más de un grupo.

En cada caso, es necesario establecer «a priori» las condiciones en las que se va a basar el programa para identificar y delimitar los distintos grupos. Se exponen a continuación los parámetros básicos para la definición de los grupos y el valor que se ha adoptado como óptimo para cada uno de ellos. La determinación de los valores óptimos se ha realizado atendiendo a diversos estudios previos de identificación de grupos, en los que se han variado las condiciones de delimitación y se ha verificado la composición resultante con expertos de las áreas analizadas.

- Número mínimo de publicaciones de un autor para ser incluido en el estudio. Los autores ocasionales, con sólo una publicación en el período analizado, no se asignan a grupos. La razón para eliminar estos autores es que se considera inadecuado determinar su asignación a un determinado grupo en base a una sola publicación. De hecho, en la categoría de autores ocasionales pueden encontrarse tanto autores recién llegados al grupo, que no han tenido todavía tiempo de reunir un número importante de publicaciones, como autores de otras disciplinas o centros, que ocasionalmente colaboran con el grupo objeto de estudio.
- Porcentaje mínimo de publicaciones que un autor tiene que firmar junto con un investigador principal para ser incluido en su grupo. Se considera que un autor tiene que firmar al menos el 60% de su producción con un IP para ser asignado a su grupo. Se observó que el establecimiento de este umbral por encima del 60% conducía a una excesiva fragmentación de los grupos, mientras que cifras inferiores al 60% producían macro-grupos que englobaban varios grupos atendiendo al criterio de los expertos. El umbral del 60% incluye los autores que teniendo 3 documentos firman dos en colaboración con un IP (66%). Por el contrario, aquellos autores de dos documentos que colaboran en uno de ellos con el IP (50%), quedan excluidos.
- Número mínimo de autores/grupo. Se considera que un grupo debe tener al menos tres investigadores.
- Número mínimo de publicaciones de un autor para poder ser considerado IP. Se considera que un investigador principal debe tener al menos una publicación anual.
- Dependencia institucional única o múltiple de los miembros de un equipo. Los grupos identificados no responden a criterios institucionales o administrativos, sino a criterios de co-autoría. Por esta razón, se considera que los miembros de un equipo pueden pertenecer a distintos centros, siempre que se superen los umbrales de colaboración requeridos para definir un grupo. Es importante resaltar que el método desarrollado permite, para cada grupo, estudiar la composición y la adscripción institucional de cada uno de los componentes.

Una vez realizado el análisis de co-autoría, se obtiene un fichero que, para cada grupo, proporciona información sobre: número de identificación de grupo, nombre y dirección del investigador principal y de cada miembro de ese grupo, número de publicaciones de cada autor, número de publicaciones co-autoradas por cada autor con el IP del grupo y porcentaje que representa esta colaboración.

2.2 Estudio y caracterización de los grupos

Los grupos se pueden caracterizar en función de su composición y de sus hábitos de publicación. Para cada grupo se obtiene la siguiente información: tamaño de grupo (número de investigadores), producción (número de documentos en el período analizado), productividad (producción del grupo en función del número de investigadores o lo que es lo mismo, número de documentos por autor del grupo), tasas de colaboración nacional e internacional (porcentaje de documentos en colaboración nacional e internacional), carácter básico o aplicado de la investigación (atendiendo a las revistas de publicación), especialización temática (a través del principal subcampo de publicación) y factor de impacto esperado (factor de impacto medio de la producción del grupo en el principal subcampo de publicación). Para calcular el nivel de investigación medio de los grupos se utiliza la clasificación de Noma *et al.* (16) que asigna a cada revista del SCI un valor que oscila entre 1 (nivel aplicado-clínico) hasta 4 (nivel básico), con dos niveles de carácter intermedio.

El proceso general de identificación y caracterización de los grupos se puede aplicar a dos niveles de análisis diferentes: a) en una disciplina; y b) en un centro de investigación.

- a) Nivel disciplina, enfocado a identificar los grupos de investigación en una determinada disciplina. En este caso, para cada autor se calculan dos frecuencias de producción: frecuencia inferior o número de documentos del autor en la disciplina analizada, y frecuencia superior o número total de documentos del autor en el total de la base de datos SCI. Los grupos se identifican atendiendo a la frecuencia inferior, pero cuanto más cercanos estén ambos valores, mayor será la especialización del grupo en la disciplina analizada.
- b) Nivel centro, orientado a la identificación de grupos de investigación en un determinado centro.

Para ilustrar el funcionamiento de la metodología, se muestran en este trabajo los resultados de su aplicación sobre la producción española de una determinada disciplina, primero, y de un centro de investigación, a continuación. En ambos casos, la fuente de datos utilizada fue el Science Citation Index, versión CD-ROM, durante los años 1990-93. Como ejemplo de la aplicación de la metodología a una disciplina, se exponen en este trabajo los resultados del estudio de los grupos de investigación españoles en Sistema Cardiovascular, delimitada el área en función de la clasificación de revistas del SCI, es decir, que todos los documentos publicados en las revistas del subcampo «Sistema Cardiovascular» se consideraron de interés para el área. En relación con el estudio de un centro, se exponen los resultados del análisis de la producción de un centro de investigación del CSIC.

3 Resultados

El estudio de grupos se ha realizado en el área de Sistema Cardiovascular, décima disciplina más productiva en el conjunto de la Biomedicina española (18), y en el Centro de Biología Molecular (CBM) que es el tercer centro más productivo de España en el área biomédica (18). A continuación se muestran los resultados del estudio

3.1 Estudio de una disciplina

La producción científica española publicada en las revistas clasificadas bajo el epígrafe de Sistema Cardiovascular durante el periodo 1990-1993 ascendió a 744 documentos. El número medio de autores por documento fue de 5,78. Hubo un total de 4.303 ocurrencias de autores que correspondieron a 2.230 autores únicos (autor+dirección). De estos autores, el 68% fue responsable de un sólo documento en los cuatro años estudiados, y el 32% restante publicó dos o más documentos en dicho período (tabla I).

Los análisis de co-autoría permitieron identificar un total de 78 grupos. Estos grupos reunieron 566 autores (el 79% de los autores con más de un documento en el período estudiado) y 559 documentos (el 75% de los documentos del área). Se observó cierto solapamiento entre grupos: 46 autores (8% de los autores) estaban adscritos a más de un grupo, y 126 documentos (17%) se realizaron en colaboración entre varios equipos. La tabla I muestra algunos datos básicos de la identificación de los grupos.

Tabla I
Identificación de grupos de investigación en la disciplina de Sistema Cardiovascular

Número de documentos	744
Número de autores	2.230
— con más de 1 documento	718 (32,2%)
— con 1 documento	1.512 (67,8%)
Número de grupos	78
Número documentos adscritos a grupos	559 (75%)
Número autores adscritos a grupos	566 (78,8% de los autores con más de 1 documento)

La caracterización de los grupos se ha realizado sólo para el subconjunto de grupos con mayor producción, es decir, para 36 grupos que publicaron 2 o más documentos/año en el período analizado (tabla II). Los grupos estaban concentrados en Cataluña (39%) y Madrid (33%). La mayor parte de los equipos pertenecía al sector hospitalario (33 grupos, 92%), con escasa representación de la Universidad (3 grupos, 8%). En la tabla II se muestra el tamaño de grupo, su producción, productividad, nivel de investigación y factor de impacto medios para el conjunto de los 36 grupos. Hay que señalar que estas variables mostraron una distribución asimétrica, por lo que se muestra la media, moda y mediana para describir cada una de las variables.

Tabla II
Caracterización de los grupos españoles con mayor producción en el área Cardiovascular

<i>Característica</i>	<i>Media±desviación estándar (N=36)</i>
Tamaño del grupo	10,92±6,23 Me=9; Mo=7; Rango=3-27
Nº documentos/grupo	13,92±8,38 Me=11; Mo=8; Rango=8-43
Productividad	1,38±0,58 Me=1,27; Mo=0,8; Rango=0,75-4
Nivel de investigación	2,37±0,48 Me=2,12; Mo=3; Rango=1,73-3
Factor de impacto esperado	3,28±1,03 Me=3,38; Mo=4,66; Rango=0,72-5,81
Colaboración nacional (% doc. con más de una dirección nacional)	0,36±0,27 Me=0,35; Mo=0; Rango=0-1
Colaboración internacional (% doc. con participación de algún centro extranjero)	0,19±0,23 Me=0,11; Mo=0; Rango=0-1
Sin colaboración (% de doc. con una sola dirección)	0,45±0,24 Me=0,44; Mo=0,44; Rango=0-0,91

Me= mediana; Mo=moda. Datos de colaboración expresados en tanto por uno.

La figura 1 muestra la distribución del tamaño de grupo. Aunque en promedio los equipos estaban compuestos por 11 investigadores, los grupos más abundantes fueron los de 7 investigadores, y sólo un 17% de los equipos tenían más de 11 investigadores. El mayor tamaño de grupo se asoció a mayor producción, como muestra la correlación positiva encontrada entre ambas variables (figura 2).

Tabla III
Distribución de grupos por disciplina principal de publicación

<i>Disciplina principal</i>	<i>Número de grupos</i>
Cardiovascular	20
Hematología	10
Dermatología	2
Investigación biomédica	1
Medicina interna	1
Farmacología	1
Nefrología	1
Radiología	1

Figura 1
Tamaño de los grupos con mayor producción del área cardiovascular

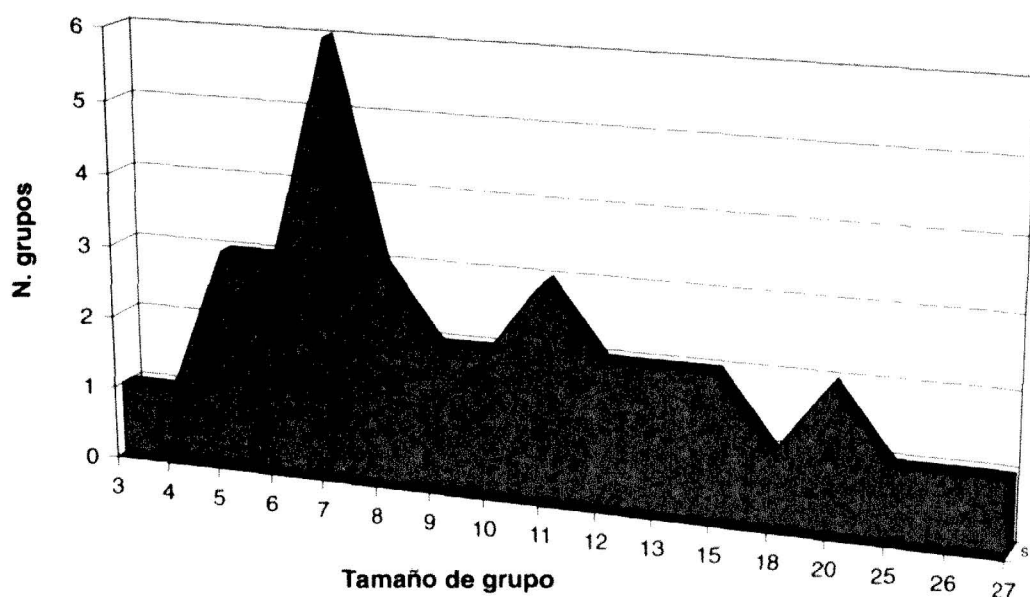
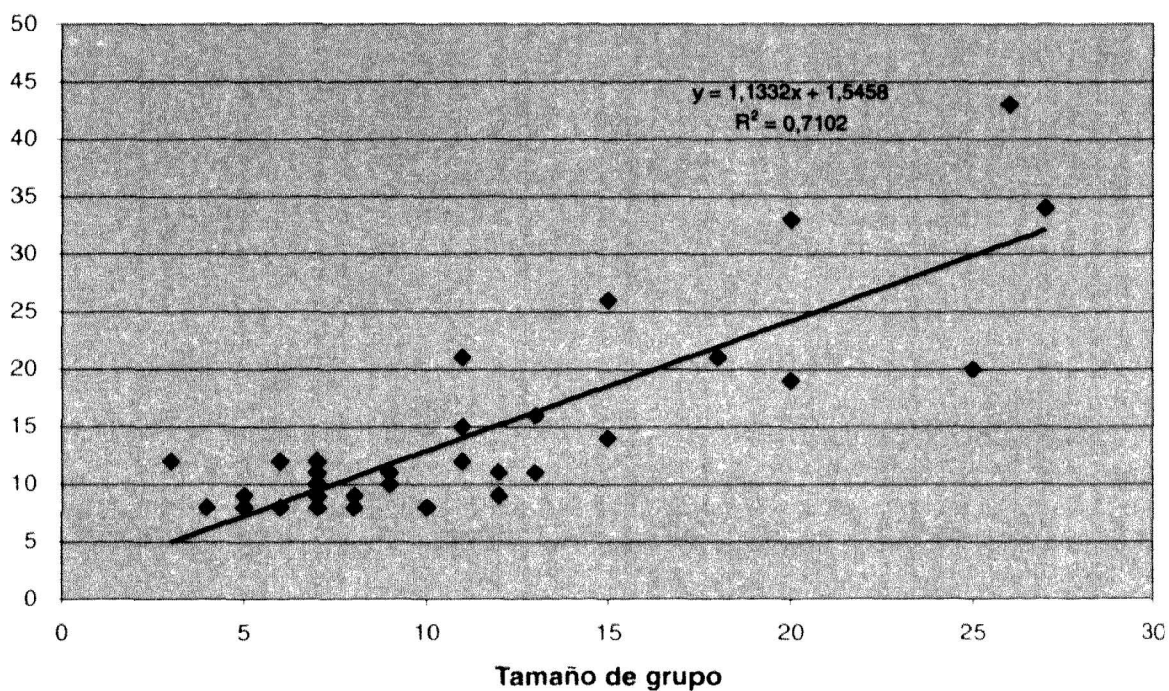


Figura 2
Relación entre la producción y el tamaño de grupo. Área cardiovascular



La principal disciplina en la que publicaron los grupos puede utilizarse para definir su área de especialización. En la tabla III se muestra dicha especialización para los distintos equipos.

La mitad de los grupos estaba especializada en Sistema Cardiovascular (20 grupos, 54%) y una cuarta parte en Hematología (10 grupos, 27%). Hubo un grupo que tenía dos disciplinas como principales especialidades, lo que explica la cifra final superior al número de grupos.

2. Estudio de un centro

Uno de los centros españoles con mayor incidencia en la investigación básica es el Centro de Biología Molecular (CBM), centro mixto dependiente del CSIC y de la Universidad Autónoma de Madrid. Este centro, durante el periodo estudiado, tuvo una producción científica de 599 documentos. El número medio de autores por documento ascendió a 4,6. La tabla IV muestra el número de autores diferentes identificados, el 61% de los cuales publicó sólo un documento en el período frente a un 39% con dos o más documentos.

A través del análisis de co-autoría se identificaron 37 grupos de investigación, que reunieron el 79% de los autores con más de un documento, y el 94% de los documentos publicados por el centro (tabla IV). Se observó un escaso solapamiento entre grupos, de forma que ningún autor pertenecía a más de un grupo.

Tabla IV
Identificación de grupos de investigación del CBM

Número de documentos	599
Número de autores	1.061
— con más de 1 documento	410 (39%)
— con 1 documento	651 (61%)
Número de grupos	37
Número documentos adscritos a grupos	561 (94%)
Número autores adscritos a grupos	321 (78,3% de los autores con más de 1 documento)

Se identificó un total de 31 grupos con más de 2 documentos/año, cuya caracterización se muestra en la tabla V.

La distribución de los grupos del CBM según su principal área de publicación se muestra en la tabla VI, donde se incluye para cada disciplina el número de documentos publicados por el centro y el número de grupos especializados en la misma. Se observa una alta concentración en el área de Bioquímica y Biología Molecular (52% de los equipos), seguida por Virología (20% de los grupos).

Los grupos del CBM mostraron una mayor producción y productividad que los grupos cardiovasculares. Subyace la actividad más clínica de éstos últimos grupos, sobre todo localizados en hospitales y con una actividad asistencial concomitante que en par-

Tabla V
Caracterización de los grupos con mayor producción del CBM

<i>Característica</i>	<i>Media±desviación estándar (N=31)</i>
Tamaño del grupo	9,44±6,17 Me=7; Mo=5; Rango=3-27
Nº documentos/grupo	19,4±15,65 Me=14; Mo=8; Rango=8-77
Productividad	1,98±0,68 Me=2,85; Mo=1,6; Rango=0,8-3,5
Nivel de investigación	3,79±0,26 Me=3,9; Mo=4; Rango=2,9-4
Factor de impacto relativo	1,631±0,82 Me=1,432; Rango=0,775-4,893
Colaboración nacional (% doc. con más de una dirección nacional)	0,26±0,26 Me=0,17; Mo=0; Rango=0-1
Colaboración internacional (% doc. con participación de algún centro extranjero)	0,22±0,21 Me=0,17; Mo=0; Rango=0-0,82
Sin colaboración (% de doc. con una sola dirección)	0,52±0,24 Me=0,53; Mo=0,80; Rango=0-0,98

Me= mediana; Mo=moda

Tabla VI
Distribución de los grupos por disciplina principal de publicación

<i>Disciplina</i>	<i>N.º de documentos CBM</i>	<i>N.º de grupos</i>
Bioquímica	216	16
Virología	103	6
Inmunología	80	2
Genética	45	2
Microbiología	43	2
Neurociencias	42	2
Biofísica	41	0
Citología	28	1
Embriología	19	2

te puede plasmarse en publicaciones nacionales, no consideradas en este estudio, y que además limita su dedicación a la investigación. Por el contrario, el CBM es un centro dedicado a la investigación, cuyos grupos desarrollan una actividad de carácter muy

básico. Atendiendo al principal subcampo de publicación, el 84% de los grupos del CBM y el 67% de los grupos del área cardiovascular publicaron sus trabajos en revistas de FI superior al promedio del país.

4 Discusión

La metodología expuesta permite profundizar en el estudio de la actividad investigadora, descendiendo a analizar la composición y la actividad de los grupos activos en un centro o área científica. La composición de los grupos se determina atendiendo a las frecuencias de co-autoría, y su actividad se caracteriza posteriormente mediante distintos indicadores bibliométricos. La aplicación del método presenta algunas limitaciones que se comentan a continuación.

- Dependencia de bases de datos con buena cobertura de autores y direcciones. La mayor parte de las bases de datos bibliográficas recogen sólo un número limitado de los autores firmantes de los trabajos y sólo la primera dirección, lo que las invalida como fuente de datos en la que aplicar la metodología propuesta. El método que se presenta está en principio limitado a las bases de datos que recogen todos los autores y direcciones de los documentos. Es el caso de la base de datos SCI, que además reúne ciertos criterios de normalización de los datos bibliográficos que facilitan su posterior tratamiento y codificación. Sin embargo, también podría aplicarse a otras bases de datos, como ICYT (producción científica española en ciencia y tecnología) o IME (producción científica española en medicina), que incluyen también completa información de los autores y sus lugares de trabajo.

- Validez variable del método según el área temática a analizar. La identificación de grupos se ha realizado con éxito en varias áreas biomédicas (19,20,21), en las que predomina el trabajo en equipo. No obstante, se cuestiona la utilidad del método en áreas de ciencias sociales y humanas, donde el trabajo individual prevalece sobre la colaboración y el trabajo en equipo. Un estudio bibliométrico reciente sobre el área de Documentación mostraba la escasa colaboración existente en dicha área, que presentó un índice de co-autoría medio de 1,3 autores por documento (22).

- Problemas técnicos en la aplicación de la metodología. El análisis de co-autoría se realiza automáticamente pero es necesaria cierta supervisión manual de los datos para evitar errores como los causados por la falta de normalización de los nombres de los autores, identificación incorrecta de autores con nombres muy comunes, o división de la producción de un autor que aparece con distintas direcciones. Estos errores pueden detectarse y evitarse mediante una cuidadosa supervisión de los grupos resultantes.

- Necesidad de asesoramiento externo. Resulta conveniente contrastar los resultados con algún experto del área o del centro objeto de análisis. La composición de los grupos identificados por análisis de co-autoría no siempre coincide con la de los grupos «reales» conocidos por los expertos. Resulta interesante contrastar las diferencias.

Por último, hay que señalar algunos aspectos diferenciales entre la identificación de grupos en una disciplina o en un centro de investigación. El estudio de un centro tiene la ventaja de que la producción objeto de análisis está muy bien delimitada. El

análisis de co-autoría se realiza sobre toda la producción de unos autores adscritos al centro que se analiza, y se obtiene una composición final de grupo más próxima a la composición real defendida por los expertos. Así, en el caso del estudio del CBM, el 93% de los documentos del centro se incluía en alguno de los grupos identificados, y sólo un 7% quedaba sin incluir, probablemente por ser aportaciones de autores individuales o de grupos que no alcanzan los criterios de producción definidos en el estudio.

En el caso del estudio de un área temática, la delimitación de ésta constituye la mayor dificultad. Hay que tener en cuenta que cuanto más correcta y exhaustiva sea la delimitación del área, más correctos serán también los resultados relativos a grupos. La delimitación basada en revistas de publicación, tiene la ventaja de ser fácilmente reproducible, pero es un método poco exhaustivo. Los trabajos de interés para una determinada especialidad se publican sobre todo en las revistas de dicha especialidad, pero también en revistas multidisciplinares o de especialidades próximas, y estas últimas no se consideran en la delimitación por revistas, por lo que se obtiene una imagen incompleta del área. En el caso del análisis de co-autoría, al incluir sólo parcialmente la producción de algunos autores, se identifican los grupos sobre una fracción de su producción, por lo que sólo se identifican aquellos grupos cuya producción está muy concentrada en revistas de su especialidad. De hecho, un 25% de los documentos del área cardiovascular quedaron sin asignar a grupos, documentos procedentes de autores que sólo ocasionalmente publican en revistas del área. No obstante, estudios previos indican que, aunque el enfoque por revistas no aporta una visión exhaustiva sobre los autores y grupos activos en un área, sí que permite obtener el núcleo básico de los mismos (21). Por otro lado, es posible observar la mayor o menor especialización de los grupos en el área en función de la diferencia entre la frecuencia de publicación inferior y superior de cada autor. En aquellos casos en los que se observa una diferencia importante entre ambas frecuencias, la composición del grupo es sólo orientativa y debe verificarse.

El estudio de un área temática tiene la ventaja de la mayor homogeneidad de la población a estudiar. La pertenencia de los grupos a una misma área nos permite asumir patrones similares de comportamiento, con lo que es posible realizar comparaciones entre grupos en función de su producción, productividad o factor de impacto de las revistas de publicación. Por el contrario, el estudio de un centro con actividad en varias áreas tiene el inconveniente de la mayor heterogeneidad de la actividad, con la dificultad consiguiente en el estudio comparativo de los grupos. Este problema aparece sobre todo al estudiar un hospital u otro tipo de centro multidisciplinar, pero es mínimo en estudios de centros especializados en una disciplina o área, como es el caso del CBM. De hecho, en este trabajo se observa que el principal subcampo de publicación (cardiovascular en el estudio por área y biología molecular en el caso del centro) permitía definir la actividad de más de la mitad de los grupos en ambos casos.

Los estudios realizados indican dos líneas principales de aplicación de este tipo de análisis. En primer lugar, esta metodología tiene indudable interés para profundizar en el conocimiento de un área e identificar a la población de investigadores y grupos más activos, así como caracterizar su actividad. No sólo es posible identificar los equipos más activos, sino también establecer quiénes presentan mayor productividad, publican en revistas de mayor factor de impacto o tienen más contactos internacionales. Pero además, es posible establecer patrones generales de comportamiento de los grupos y

profundizar en cómo estos patrones varían atendiendo al tipo de área o de investigación que realizan los autores.

El proceso descrito para la identificación de grupos de investigación se ha seguido hasta la fecha en distintas áreas biomédicas y ha sido objeto de constante revisión hasta llegar a la etapa actual de desarrollo. En cada caso, el análisis crítico de los resultados junto con los expertos de las áreas analizadas ha sido una importante fuente de experiencia, orientando sobre posibles modificaciones a introducir en el proceso. En la etapa actual de trabajo se pretende profundizar en el estudio de los grupos, aquí sólo esbozado en líneas generales, y valorar la posibilidad de introducir nuevos indicadores complementarios, tanto bibliométricos (número de citas) como de otro tipo (proyectos financiados).

Agradecimientos

Este trabajo ha sido realizado gracias a los proyectos SEC95-0082 y SEC97-1375, del Plan Nacional de I+D. Agradecemos la colaboración prestada por el Dr. Santos Barrión en el desarrollo del método y en la discusión de los resultados.

Bibliografía

1. López Piñero, J. M.; Terrada, M. L. Los indicadores bibliométricos y la evaluación de la actividad médico-científica. (IV) La aplicación de los indicadores. *Medicina Clínica (Barcelona)* 1992, vol. 98, n.º 10, p. 384-388.
2. Sancho, R. Indicadores bibliométricos utilizados en la evaluación de la ciencia y la tecnología. *Revista Española de Documentación Científica*, 1990, vol. 13, n.º 3-4, p. 842-865.
3. Van Raan, A. F. J. Advanced bibliometric methods to assess research performance and scientific development: basic principles and recent practical applications. *Research Evaluation*, 1993, vol. 3, n.º 3, p. 151-166.
4. Kretschmer, H. Cooperation structure, group size and productivity in research groups. *Scientometrics*, 1985, vol. 7, n.º 1-2, p. 39-53.
5. Nederhof, A. J.; Van Raan, A. F. J. A bibliometric analysis of six economics research groups: a comparison with peer review. *Research Policy*, 1993, vol. 22, p. 353-368.
6. Van Raan, A.F.J. Evaluation of research groups. En *Ciba Foundation Conference. The evaluation of scientific research*. Wiley eds., 1989, p. 169-187.
7. Moed, H. F.; Van Raan, A. F. J. Indicators of research performance. Applications in university research policy. En A. F. J. van Raan (ed.), *Handbook of quantitative studies of science and technology*. Amsterdam: Elsevier, 1988.
8. Moed, H. F.; Burger, W. J. M.; Frankfort, J. G.; Van Raan, A. F. J. The use of bibliometric data for the measurement of university research performance. *Research Policy*, 1985, vol. 14, p. 131-149.
9. Luukkonen, T. Citation indicators as measures of performance: case studies of cardiovascular research groups in Norway and Finland. En A. F. J. van Raan, A. J. Nederhof y H. F. Moed (eds.), *Science and technology indicators: their use in Science Policy and their role in science studies*. Leiden: DSWO Press, 1989.
10. Cohen, J. E. Size, age and productivity of scientific and technical research groups. *Scientometrics*, 1991, vol. 20, n.º 3, p. 395-416.
11. Stokes, T.D.; Hartley, J.A. Coauthorship, social structure and influence within specialties. *Social Studies of Science*, 1989, vol. 19, p. 101-125.

12. Peters, H. P. F.; Van Raan, A. F. J. Structuring scientific activities by coauthor analysis. An exercise on a university faculty level. *Scientometrics*, 1991, vol. 20, n.º 1, p. 235-255.
13. Bordons, M.; Zulueta, M. A.; Cabrero, A.; Barrigón, S. Identifying research teams with bibliometric tools. En M. E. Koenig, A. Brookstein (eds.), *Proceedings of the Fifth Biennial Conference of the International Society for Scientometrics and Informetrics*. Learned Information, Inc. Medford, 1995.
14. Vinkler, P. *Bibliometric analysis of publication activity of a scientific research institute. Informetrics 89/90*. Elsevier Science Publishers B.V. 1990, p. 309-334.
15. Lee Pao, M. Global and local collaborators: a study of scientific collaboration. *Information Processing & Management*, 1992, vol. 28, n.º 1, p. 99-109.
16. Noma, E. Subject classification and influence weights for 3000 journal. *Research report under CHI and NIH contracts*. Computer Horizons Inc. Research, New Jersey, 1986.
17. Fernández, M. T.; Cabrero, A.; Zulueta, M. A.; Gómez, I. Constructing a relational database for bibliometric analysis. *Research Evaluation*, 1993, vol. 3, n.º 1, p. 55-62.
18. Camf, J.; Zulueta, M.A.; Fernández, M.T.; Bordons, M.; Gómez, I. Producción científica española en biomedicina y ciencias de la salud durante el período 1990-1993 (Science Citation Index y Social Science Citation Index) y comparación con el período 1986-1989. *Medicina Clínica (Barcelona)*, 1997, vol. 109, n.º 13, p. 481-496.
19. Bordons, M.; Zulueta, M.A. Comparison of research team activity in two biomedical fields. *Scientometrics*, 1997, vol. 40, n.º 3, p. 423-436.
20. Bordons, M.; Zulueta, M. A.; Barrigón, S. Actividad científica de los grupos españoles más productivos en farmacología y farmacia durante el período 1986-1993 a través del Science Citation Index (SCI). *Medicina Clínica (Barcelona)*, 1998, vol.111, p. 489-495.
21. Zulueta, M.A.; Bordons, M. A global approach to the study of teams in multidisciplinary research areas through bibliometric indicators. *Research Evaluation*, 1999, vol. 8, n.º 2 (en prensa).
22. Jiménez, E.; Moya, F. Análisis de la autoría en revistas españolas de Biblioteconomía y Documentación, 1975-1995. *Revista Española de Documentación Científica*, 1997, vol. 20, n.º 3, p. 252-266.