



*Estudio de los flujos de citación científica
y su relación con los indicadores de impacto*

TESIS DOCTORAL

Doctoranda

Bárbara S. Lancho Barrantes

Director

Vicente P. Guerrero Bote

DEPARTAMENTO DE INFORMACIÓN Y COMUNICACIÓN

FACULTAD DE BIBLIOTECONOMÍA Y DOCUMENTACIÓN

2012

DEPARTAMENTO DE INFORMACIÓN Y COMUNICACIÓN
FACULTAD DE BIBLIOTECONOMÍA Y DOCUMENTACIÓN

Universidad de Extremadura

Badajoz, junio 2012

A mi madre

ÍNDICE

Prefacio	IX
Agradecimientos.....	XI
PARTE I: LÍNEAS GENERALES DEL TRABAJO	
1 Resumen.....	15
2 Introducción	23
2.1 Delimitación del estudio y antecedentes	23
2.1.1 Estructura del trabajo.....	25
2.1.2 Evaluación de la Investigación Científica	26
2.1.3 La herramienta para la evaluación: la Bibliometría.....	31
2.1.4 Los indicadores bibliométricos.....	37
2.1.5 Los diferentes hábitos de citación en la ciencia.....	41
2.2 Fuentes utilizadas	47
2.2.1 Scopus.....	47
2.2.2 SCImago Journal & Country Rank.....	49
2.3 Aplicación de Indicadores.....	57
2.3.1 Indicadores aplicados a las distribuciones de impacto	57
2.3.2 Indicadores de tasas de exportación e importación.....	58
2.3.3 Indicadores para medir la influencia de los perfiles de referencia y los hábitos de citación.....	59
2.3.4 Indicadores de colaboración científica	61

3 Justificación y Objetivos	65
3.1 Evaluar la investigación	65
3.2 Limitaciones de los indicadores bibliométricos	67
3.3 La falta de cobertura de las bases de datos y los hábitos de citación	68
3.4 Objetivos específicos de esta investigación	73
4 Discusión de resultados	77
5 Conclusiones	113
6 Perspectivas de investigación futuras.....	123
7 Referencias	125
PARTE II: ARTÍCULOS CIENTÍFICOS	
8 Papers.....	149
A1: The Iceberg Hypothesis revisited	151
A2: What lies behind the averages and significance of citation indicators in different disciplines?	189
A3: Citation flows in the zones of influence of scientific collaborations	223
A4: Citation increment between collaborating countries	253
9 Resúmenes de los artículos.....	293

Prefacio

Esta tesis ha sido posible por la financiación de la Consejería de Empleo, Empresa e Innovación de la Junta de Extremadura y el Fondo Social Europeo, gracias a sus ayudas para la concesión de becas y contratos en prácticas para la formación predoctoral de personal investigador (III PRI+D+I). Para formar parte del proyecto de investigación “Sistema Automático de Extracción de Frentes de Investigación” PRI06A200 cuyo investigador principal fue el Dr. Vicente P. Guerrero Bote.

Los resultados y la mayor parte de las conclusiones de este trabajo de tesis han sido presentados, en detalle, en cuatro trabajos realizados por la doctoranda en colaboración con su director de tesis, así como con miembros de su grupo de investigación, SCImago que pertenece al Consejo Superior de Investigaciones Científicas (CSIC) en Madrid.

Agradecimientos

Durante los años de elaboración de esta Tesis Doctoral son muchas las personas que han contribuido en mi progreso intelectual y personal.

En primer lugar agradecerle enormemente al Dr. Vicente P. Guerrero Bote, por creer en mí desde mis inicios, darme la oportunidad de realizar este trabajo y ayudarme a crecer profesionalmente. Gracias por haberme dirigido por el camino del conocimiento con la cautela, la rigurosidad y la perseverancia necesarias en esta profesión.

Al Dr. Félix de Moya Anegón y todo el Grupo de Investigación SCImago por el soporte y apoyo brindado en estos años de estudio y trabajo.

A mis padres y familia. Haciendo especial mención a mi madre por su aguante infinito y por haber estado a mi lado en todo momento animándome e impulsándome a continuar.

A mis mejores amigos por su comprensión y respeto sobre todo en estos últimos meses. Gracias por darme los mejores consejos, escucharme largas conversaciones y compartir conmigo los fines de semana de descanso y desconexión.

*Estudio de los flujos de citación científica
y su relación con los indicadores de impacto*



PARTE I: LÍNEAS GENERALES DEL TRABAJO

1 Resumen

La constante demanda de incremento en la financiación del proceso investigador cuyos resultados tienen gran influencia en el desarrollo económico y social de los países ha originado la necesidad creciente de cuantificar y evaluar la actividad científica, sus resultados y rendimientos. Las administraciones públicas encargadas de la distribución de los, siempre escasos, recursos económicos destinados al sistema de I+D+i, requieren, cada vez más, instrumentos de análisis que les permitan comprobar la eficacia de sus políticas y que les ayuden a racionalizar y adecuar sus inversiones de acuerdo con la planificación que han diseñado.

Por tanto, la evaluación resulta un elemento imprescindible, al menos para dos colectivos: para los propios científicos, a quienes les interesa saber si su investigación es de calidad, y para la administración pública y las empresas privadas, que son las encargadas de financiar las actividades científicas. A éstas les interesa comprobar si sus inversiones han sido productivas. Para ambos colectivos, la evaluación es un factor necesario puesto que dará a conocer a unos la importancia y el alcance de su investigación y a otros si la inversión realizada se ha canalizado de forma adecuada.

Tradicionalmente la investigación se ha evaluado a través de la opinión de expertos que, además de cuantificar los resultados, proporciona información sobre los aspectos de calidad de la misma. Este indicador, substancial en la valoración de grupos pequeños, de autores, proyectos y comunidades científicas muy reducidas, es de difícil aplicación a grandes y medianos colectivos con abundante producción, como centros de investigación, regiones o países, debido a la cantidad de recursos que consume. Por otro lado, al estar basado en percepciones, no está exento de inconvenientes, debido a que los científicos que hacen la valoración pueden incurrir en parcialidad, desconocimiento o desconfianza en las líneas emergentes.

El crecimiento de la producción científica en las últimas décadas así como su recopilación en bases de datos bibliográficas automatizadas han potenciado el uso de la bibliometría y la generación de indicadores para evaluar y medir los resultados de la actividad científica y tecnológica.

En la búsqueda de indicadores que aporten objetividad al proceso de evaluación surgen los indicadores bibliométricos, basados en el análisis de las publicaciones científicas. El uso de este tipo de indicadores con fines de evaluación es hoy una práctica común. Actualmente el factor de impacto, esta admitido como la herramienta principal para medir la visibilidad o difusión de una revista. Aunque no es el único indicador, es el más extendido, su fortaleza reside en su facilidad de cálculo, en su estabilidad y en su accesibilidad. Pero debido al uso arbitrario, no ponderado, que se hace de él, es controvertido, criticado y ha perdido su credibilidad (Moya-Anegón et al, 2005).

Glänzel & Moed, (2002) hacen una amplia revisión sobre sus ventajas y sus limitaciones así como del contexto de aplicación y las funciones empleadas para su cálculo.

Por ello aunque el impacto de los trabajos resulta importante, porque a través del recuento de citas recibidas por los mismos se puede conocer la repercusión de este trabajo en la comunidad científica. No es del todo una medida directa de calidad. No habría ningún problema si los índices bibliométricos se usaran realmente como lo que son, es decir, indicadores del nivel de difusión entre la comunidad científica. El problema aparece cuando estos factores son utilizados como índices de calidad, y en función de ello se considera que un artículo tiene calidad en función del prestigio que tiene la revista en la que fue publicado. Pinski y Narin ya en (1976) subrayaron que parecía más razonable dar un valor más alto a las citas provenientes de una revista prestigiosa que a las citas de una revista periférica.

Por este motivo, en la actualidad surgen nuevos indicadores. Por ejemplo los que se basan en el algoritmo de PageRank de Google. Tal es el caso del SJR (Grupo SCImago, 2007), utilizado en SCImago Journal & Country Rank, calculado también sobre revistas, pero considerando 3 años de citas y basado en los datos de Scopus. Igualmente el Eigenfactor y el Article Influence (Bergstrom, 2007), basados también en PageRank han sido incluidos recientemente en el Journal Citation Report del ISI junto con el Factor de Impacto de 5 años más apropiado para sectores de la ciencia donde la vida útil de los artículos es más dilatada, como en Matemáticas o Ciencias Sociales.

Pero es importante señalar que los indicadores bibliométricos difícilmente se pueden emplear para hacer comparaciones entre disciplinas científicas. Las principales causas que impiden estas comparaciones son: la falta de cobertura de las bases de datos en ciertas áreas (Braun et al., 2000, Grupo SCImago, 2006, Moya-Anegón et al., 2007) y los propios hábitos de citación (Pinski & Narin, 1976). Del mismo modo, el número medio de referencias por artículo que tienen las áreas temáticas, el tamaño de la comunidad científica (Gómez Caridad & Bordons Cangas, 1996), las diferentes proporciones de citación a los diferentes tipos de documentos citables (artículos, revistas y cartas) (Lundberg, 2007), o el periodo de tiempo necesario para llegar a la máxima citación (Moed, 2005) también pueden influir en que no se puedan hacer balances entre áreas científicas. Por tanto a veces en procesos de evaluación es necesaria una normalización de los mismos para cada disciplina, de modo que resulten valores comparables entre unas y otras (Moed et al., 1995; Van Leeuwen & Moed, 2002; Van Leeuwen et al., 2003; Podlubny, 2005; Sombatsompop & Markpin, 2005; Moya-Anegón et al., 2005; Van-Raan, 2006; Lundberg, 2007; Guerrero et al., 2007).

En ese sentido es difícil determinar la estructura de la ciencia en disciplinas, aunque son muchos los estudios que admiten como tal el Subject Category Listing del JCR del ISI. Dichas categorías, como indican Pudovkin & Garfield, (2002) fueron desarrolladas hace más de 40 años por métodos manuales para clasificar las revistas científicas, aunque a medida que fueron creciendo se establecieron subdivisiones.

Para la asignación de las revistas a categorías se sigue un proceso heurístico que tiene en cuenta las citas y utiliza entre otros el algoritmo de Hayne-Coulson.

A pesar de que hay autores que no lo comparten, se considera que las mencionadas categorías ISI son unidades informativas lo suficientemente explícitas como para poder ser utilizadas en la representación de las distintas disciplinas que componen la ciencia en general (Moya-Anegón et al, 2006).

Estas categorías son utilizadas en la actualidad, por ejemplo para la realización de mapas de la ciencia y de este modo observar las relaciones que pueden existir entre las diferentes categorías (Moya-Anegón et al, 2004) aplicando el análisis de citas y técnicas de poda como el algoritmo de Pathfinder (Schvaneveldt, Dearholt & Durso, 1988; Guerrero-Bote et al, 2006) o para la evaluación de la ciencia (Podlubny, 2005; Van Leeuwen & Moed, 2002; Van Leeuwen et al, 2003; Sombatsompop & Markpin, 2005; Moya-Anegón et al, 2005; Moya-Anegón et al, 2008).

Debido a todo esto, y como ya se ha dicho los indicadores basados en citas no son válidos para hacer comparaciones ya que los diferentes tipos de productos, canales de difusión y sobre todo los hábitos de publicación y citación de las distintas disciplinas, campos y países son diferentes y esto dificulta el análisis bibliométrico.

Por tanto el motivo principal de realizar esta tesis será estudiar los hábitos de citación y los perfiles de referenciación en la ciencia. Para ello nos hacemos una serie de preguntas previas.

Si se comparan las distribuciones de impacto de las disciplinas de las bases de datos ¿en qué medida depende que salgan las distribuciones diferentes? ¿Qué disciplinas científicas son las más exportadoras o importadoras de conocimiento? ¿Depende de la base de datos utilizada, de la propia estructura de disciplinas, y de la forma de adscribir las revistas? O ¿es una cuestión de la ventana de citación utilizada y de una fecha más reciente?

En el caso que haya disciplinas que en las distribuciones tengan unas tasas de exportación muy altas o muy bajas ¿tendrán alguna influencia en estas tasas los propios hábitos de citación de las disciplinas? ¿Tendrá influencia en esto los perfiles de referencia anuales de las diferentes disciplinas?

Así mismo ¿qué influencia tendrá el perfil de referenciación y los hábitos de citación de las disciplinas científicas en los indicadores bibliométricos que se basan en factores de impactos? ¿Y en los basados en el PageRank que ponderan las citasiones?

A lo largo de la literatura científica se ha venido diciendo que grandes niveles de colaboración conducen a mayor citación, grandes niveles de impacto, mayor calidad de los trabajos y mayor productividad de los autores dependiendo de los campos científicos (Lewinson, 1991; Narin et al., 1991; Glänzel, 2001; Glänzel, 2002; Leimu & Koricheva, 2005; Katz & Hicks, 1997; Persson et al., 2004; Hsu & Huang, 2010; Aksnes, 2003; Moya-Anegón, et al. 2008; Chinchilla, et. al., 2010) y que los potenciales beneficios de la colaboración científica pueden depender de la disciplina, el efecto de la colaboración aparece más positivamente

en las ciencias “duras”, como Física y Astronomía que en las “blandas”, como es el caso de la Sociología o las Ciencias Sociales (Stack, 2002; Bandyopadhyay, 2001; Moed et al. 1991; Bridgstock, 1991).

En este trabajo aparte de confirmar la idea de que la colaboración internacional conduce a grandes niveles de impacto se va a averiguar si hay una serie de países que se benefician más de sus colaboraciones que otros debido a los diferentes hábitos de publicación, citación, y colaboración. También se indagará en la procedencia de la citación obtenida por los diferentes países que colaboran en un trabajo y por los diferentes tipos de trabajos. Igualmente se va a investigar sobre el destino de la citación realizada por los países. Teniendo en cuenta también la producción de un país por un lado cuanto mayor sea la producción, lógicamente mayor citación podrá obtener, pero también mayor es la citación nacional puesta en juego.

2 Introducción

2.1 Delimitación del estudio y antecedentes

En el año 2007 se publicó un artículo llamado “The Iceberg Hypothesis: Inport-Export of Knowledge between Scientific Subject Categories” (Guerrero et al., 2007) con el fin de averiguar con los datos correspondientes al JCR de 1997, accesible a través del ISI Web of Knowledge, cuáles eran las categorías científicas mas exportadoras de conocimiento para ello se estudiaron las distribuciones de impacto con una ventana de citación de 2 años, concretamente de 1995 y 1996 y en el año de citación de 1997.

Con los resultados se arrojaron una serie de conclusiones, la primera conclusión es que las distribuciones rango/impacto en cada una de las disciplinas se adaptaban bien a una curva exponencial negativa como era de prever. Esto se debe al Efecto Mateo, es decir a que cuando un trabajo o una revista es citado aumenta la probabilidad de ser citado de nuevo.

Dentro de que todas las disciplinas vienen a adaptarse a la curva exponencial negativa, hay algunas que se adaptan menos, y esto es porque tienen un descenso más pronunciado que la aproximación exponencial y una cola también más prolongada. Dicho desajuste se puede localizar principalmente mediante el error cuadrático (relativizado respecto al Factor de Impacto medio) en relación con la aproximación.

De modo que hay categorías científicas que son más independientes, importando y exportando poco conocimiento y hay otras con mayores flujos de conocimiento. Las categorías científicas podrían ser vistas como Icebergs, cuya parte emergida se divisa en ocasiones desde otra categoría (Iceberg) en la distancia, mientras que otras veces la parte emergida apenas existe y el Iceberg (la categoría) solo es visible estando dentro de él. Esta metáfora explicaría como unas disciplinas científicas son exportadoras de ideas, porque el conocimiento que se genera dentro de ellas es visible desde otras que lo importan.

Mientras que otras disciplinas son generadoras de conocimiento para el autoconsumo aunque esta no es incompatible con la posibilidad de importar conocimientos de otras disciplinas eventualmente.

En todo caso, continuando con esta metáfora, cuanto más emerge el Iceberg más posibilidades existen que se vea desde él las partes emergidas de otros icebergs. Lo que es más visible desde el exterior permite divisar mejor el entorno.

Los resultados que emergieron del estudio dieron origen a una serie de interrogantes acerca de cómo era la citación en las diferentes disciplinas científicas. Este artículo sirvió de inspiración para elaborar la siguiente tesis doctoral.

2.1.1 Estructura del trabajo

Este trabajo de tesis consta de dos partes diferenciadas:

Parte I:

La parte I incluye los siguientes apartados: resumen, introducción, justificación y objetivos, discusión y conclusiones de los resultados obtenidos, perspectivas de investigación futuras y por último la bibliografía utilizada.

Parte II:

La parte II está constituida por los cuatro artículos donde han sido presentados los resultados de la investigación. Esta parte incluye dos apartados: los artículos en inglés y los resúmenes de los artículos en español.

Publicaciones:

- **Artículo 1:** The Iceberg Hypothesis revisited
- **Artículo 2:** What lies behind the averages and significance of citation indicators in different disciplines?
- **Artículo 3:** Citation flows in the zones of influence of scientific collaborations
- **Artículo 4:** Citation increment between collaborating countries

2.1.2 Evaluación de la Investigación Científica

En los años cincuenta se comienza a fermentar en los países desarrollados una relación entre ciencia, tecnología y producción, hasta ese momento desconocida. Es obvio que este proceso va a implantar cambios considerables en la actividad científica y su relación con la sociedad. Aparece también la Big Science, creando de este modo un cambio en toda la ciencia, denominada de este modo por el físico nuclear americano y premio Nobel Alvin Martin Weinberg para referirse a planes científicos “multi”: multidisciplinarios, multinacionales y multianuales. Todo parecía aumentar de tamaño, grandes presupuestos, mayor consumo, grandes complejos, grandes expectativas, mayores costes, concurrencia de diversas disciplinas, estructuras organizativas muy complicadas, pérdida de autonomía, y colaboración entre países para desarrollar proyectos conjuntos.

Desde mediados de los años sesenta hasta los setenta se marca el "fin de la inocencia" respecto al consumo en I+D que se tenía hasta entonces. Por ello se hará incuestionable la necesidad de controlar más la inversión en I+D y evaluar mejor los resultados de investigación. De hecho a finales de los setenta en Estados Unidos existía una cierta inquietud por las formas de evaluación que tenían las Agencias Públicas de Fomento para conceder fondos de investigación (Nuñez-Jover, J, 2012).

Ya en el trance de los años ochenta a los noventa nacen las llamadas “Políticas para la Innovación”. Poco a poco en este tiempo se observa una gran inquietud por el impacto que pudieran ejercer esas políticas de ciencia e innovación en el bienestar y la calidad de vida de los ciudadanos. Esto favoreció la evaluación del impacto social de la investigación y la tecnología en aspectos propios del ser humano como bienestar, calidad de vida, riesgos, salud, etc. Aparte de todo esto se demandó una mayor responsabilidad social por parte de las organizaciones, empresas o administraciones públicas en todos sus procesos. Los métodos de evaluación elegidos en aquella época eran cualitativos y de corte relativo, del estilo al método peer-review, cuyas técnicas quedaban dentro de un cuadro de evaluación interna de la propia comunidad científica.

Pero estos métodos resultaban insuficientes ya que muchas veces no se garantizaba la fiabilidad y validez de los juicios críticos. Debido a esto se fueron implantando progresivamente nuevos métodos cuantitativos más objetivos procedentes de las ciencias métricas.

Hoy en día la cultura de la evaluación se ha configurado como uno de los instrumentos más eficaces para el desarrollo y consolidación de una buena política científica.

Al hilo de esto Nederhof (1988) aseguró que en el universo de la investigación científica existen muchas disyuntivas por las cuales es preciso tomar decisiones oportunas y adecuadas, ya que esto repercute en la productividad y en el avance de la sociedad.

La evaluación también es considerada para algunos un instrumento político esencial gracias al cual se pueden asignar recursos o incentivos para suplir las necesidades de un país. Es valiosa para cambiar estructuras organizativas, procedimientos, técnicas, actitudes o incluso para incidir en el desarrollo de las diversas disciplinas científicas (De las heras & Polo, 2009). De ahí que la necesidad de evaluar sea un elemento imprescindible.

Pero sobre todo para dos segmentos de la población es necesaria: los científicos, a quienes les concierne conocer si su investigación es de interés y calidad, y la administración pública y las empresas privadas que son las encargadas de aportar los fondos que sustentan las actividades científicas. Para ambos colectivos, la evaluación es ineludible porque permite conocer a unos, la importancia y trascendencia de sus trabajos y a otros, si la inversión efectuada se ha canalizado de forma eficiente (Milanés, et. al. 2008)

Aunque evaluar la investigación implique "evaluar la calidad de ésta", es necesario también que se distingan otros conceptos, además del de calidad, como pueden ser: la importancia o el impacto (Martin & Irvine, 1983). Los tres (calidad, importancia e impacto) son arduos de calcular y sólo pueden medirse desde la valoración de los demás científicos (peer-review) o deducirse por la aplicación de la Bibliometría.

El concepto de calidad, en sí mismo, refleja la excelencia, es indicativo del valor que puede tener un trabajo científico, sea por su contribución al conocimiento, por su cuidada metodología, o por la singularidad o agudeza con la que se haya diseñado el proyecto de investigación. Pero al no ser la calidad un concepto cuantitativo, medible de forma numérica, requiere ser examinada por homólogos, es decir, mediante juicios subjetivos que pueden, algunas veces, incurrir en sesgos debido a los propios intereses o desconocimientos de los evaluadores en la materia oportuna. Por otro lado la importancia o relevancia científica es una cualidad que se refiere a la influencia potencial que aquella investigación tiene o puede tener para el avance del conocimiento científico.

El inconveniente es que el reconocimiento de la relevancia se enfrenta por un lado a los problemas propios dentro de cada disciplina y por otro a la lentitud con que, un supuesto avance científico se introduce en el mundo de la ciencia y se tiene en cuenta.

En último lugar, el impacto, que revelaría la repercusión inmediata de un trabajo en la comunidad científica (Camí, 1997)

Pero debido a que para evaluar algunos de los conceptos anteriores los métodos cualitativos no resultan ecuanímenes se empieza a usar técnicas de carácter cuantitativo originarias de la bibliometría con la intención de restar subjetividad a los procesos de evaluación (Nuñez-Jover, 2012)

A partir de estas técnicas bibliométricas y otras técnicas de utilidad como las redes sociales, se ha desplegado otro abanico de útiles instrumentos de estudio apoyadas en mapas de la ciencia, que faciliten representar el estado presente de la producción científica de diferentes dominios institucionales y geográficos (Leydesdorff, 2004; Moya Anegón & et al., 2006).

Por otro lado, como consecuencia del progreso tecnológico, se examinan nuevas fuentes de datos para su explotación, como es el caso del Currículum Vitae (CV) para lograr analizar el cometido científico del investigador y se establecen indicadores de evaluación más específicos adaptados a las realidades de países con escasos recursos (Cañibano, Otamend & Andujar, 2008).

Existe una gran diversidad tipológica de organizaciones internacionales y nacionales cuyas funciones son evaluar y acreditar los resultados de investigación. Los siguientes tipos de organismos pueden ser: entidades gubernamentales, organismos autónomos pero dependientes del gobierno, empresas privadas con reconocimiento gubernamental y empresas privadas.

Situándonos en concreto en el caso español destacan los siguientes organismos: ANEP (Agencia Nacional de Evaluación y Prospectiva), la auténtica introductora en España de la evaluación por pares, CNEAI (Comisión Nacional de Evaluación de la Actividad Investigadora) creada en 1989 y la ANECA (Agencia Nacional de Evaluación de la Calidad y Acreditación) creada en 2002. Además de estas agencias estatales cada comunidad autónoma ha desarrollado agencias propias.

2.1.3 La herramienta elegida para la evaluación: la Bibliometría

Desde hace años, la Bibliometría, viene manifestando ser una herramienta de gran utilidad, fundamentalmente, para la evaluación de los resultados de investigación y para el análisis de la ciencia y la tecnología (Price, 1970; Garfield, 1979; Martin and Irvine, 1983; Braun, Glänzel and Schubert, 1988; Van Raan, 2004).

No existe acuerdo sobre cuáles fueron los comienzos de la Bibliometría, sin embargo, los primeros trabajos sobre este tema datan de principios del siglo XIX. En esos momentos se produce un proceso de “matematización” de las Ciencias Sociales (Psicología, Sociología, Economía, Educación, etc). Esto empieza porque los bibliógrafos de entonces empezaron a emplear los métodos estadísticos a las descripciones bibliográficas considerándolas procedimientos de análisis y observación. Cada uno de esos elementos pueden ser objetos de cuantificación. Estos trabajos bibliométricos donde más se desplegaron fueron en Estados Unidos y Gran Bretaña.

En 1923 F. W. Hulme en una conferencia en la Universidad de Cambridge utiliza el término “Statistical Bibliography” para detallar el método utilizado en su estudio sobre autores y revistas referenciadas en el International Catalogue of Scientific Literature (López-López y Tortosa, 2002). Hulme (1923) la definía de la siguiente manera “la reunión e interpretación de estadísticas relativas a libros y publicaciones periódicas; puede emplearse para una variada gama de situaciones y una casi ilimitada cantidad de medidas, entre ellas, demostrar movimientos

históricos, determinar el uso nacional o universal de libros y revistas en la investigación, aclarar en situaciones locales el uso de libros y revistas” y, también, “presentar para cada período el equivalente bibliográfico correspondiente al crecimiento y desarrollo de las actividades intelectuales de la humanidad”

Pero en 1969 se decide cambiar este término Bibliografía Estadística puesto que era un término confuso, no descriptivo y que se podía confundir con una bibliografía de estadística, es en ese año cuando Alan Pritchard acuña a estos métodos con el nombre de Bibliometrics en su artículo *Statistical bibliography or bibliometrics*, ya que las Ciencias de la Información estaban experimentando un rápido desarrollo y había una gran necesidad de medir datos de bibliografías.

Por ende Pritchard (1969) la definió como la aplicación de métodos estadísticos y matemáticos a los libros y otros medios de comunicación mediante el recuento y análisis de las distintas facetas de dicha comunicación. En otras palabras, es la aplicación de tratamientos cuantitativos a las propiedades del discurso escrito y los comportamientos típicos de éste.

La gran mayoría de los autores están de acuerdo que fue el inventor de este término, sin embargo, para algunos, el primero que empleó el término bibliometría fue Paul Otlet en 1934 para referirse a la dimensión cuantitativa de la bibliografía (Urbizagástegui Alvarado, 2007).

A lo largo del tiempo han surgido una variedad de designaciones y denominaciones para nombrar a las disciplinas métricas de la información, que ha dado lugar a que haya desacuerdos entre los especialistas en cuanto al concepto, denominación, contenido y límites. Jiménez Contreras (2005) plantea como posibles causas el origen disperso de la disciplina y la rápida evolución de las tecnologías de la información, mientras que Canales et al. (2002) lo relaciona al desarrollo de la Ciencia, que lleva consigo el desarrollo de las disciplinas métricas como instrumento para el análisis cuantitativo de su objeto de estudio. Pero a pesar de esto hay una marcada preferencia por la palabra bibliometría, que desde sus orígenes hasta la actualidad sigue siendo la más utilizada y difundida. Actualmente se refiere al método científico y la aproximación cuantitativa que permite estudiar la naturaleza y trayectoria de una disciplina (mientras dé lugar a publicaciones) y además es el estudio inferencial de todas las formas de comunicación escritas adaptadas a la literatura científica (Sancho, 1990)

Mientras la Ciencia de la Ciencia se arraigaba académicamente con la obra de Price (1965) los países occidentales creaban la Organization of Economic Cooperation and Development (OECD). El gran logro de esta organización fue el Manual de Frascati que resultó el origen de una serie de manuales sobre medición de la ciencia.

Al mismo tiempo en la Unión Soviética emerge una nueva especialidad académica bajo la denominación de Naukometriya (Medición de la Ciencia) el factor más elemental de la Naukometriya es el surgimiento de un manual en 1969 titulado *Scientometrics. The study of Science as an Information Process*. En este

trabajo se afrontaron temas como los análisis de citas, la estructura interna de los frentes de investigación y se plantearon cuestiones para la instauración de un centro que estudiara la ciencia desde el punto de vista estadístico (Granovsky, 2001).

En la década de los setenta nadie se oponía a la consolidación de los análisis métricos como campo científico independiente. Como ejemplo de esto, en 1973 se publica el informe Science Indicators como petición del congreso de los Estados Unidos a la National Science Foundation. Este informe fue pionero y ofreció una serie de indicadores de carácter puramente bibliométrico. La fuente primordial de extracción de datos fue Science Citation Index que fue elegido por su carácter multidisciplinar, por la afiliación institucional de sus autores y por la amplia cobertura de revistas (Narin, et al., 2000)

En ese mismo año de 1973 cuando la NSF presento su informe salió a la luz un estudio sobre cocitaciones (Small, 1973) que permitió el mapeo de la ciencia, la caracterización de frentes de investigación y la identificación de colegios invisibles.

En España el análisis estadístico empieza a penetrar con la figura de López Piñero en 1973, con la traducción de la obra de Price. Es el introductor de la Bibliometría en España. De esta manera López Piñero y Terrada (1992) la definen de la siguiente forma: “El análisis bibliométrico es un método documental que ha alcanzado un importante desarrollo durante las décadas. Sus objetivos fundamentales son, por una parte, el estudio del tamaño, crecimiento y distribución de los documentos científicos y, por otra, la indagación de la

estructura y dinámica de los grupos que producen y consumen dichos documentos y la información que contienen”

En 1978 se unió a la OECD otra organización, la UNESCO, que publicó unas recomendaciones sobre la estandarización en la recopilación de datos. *Recommendations Concerning the International Standardization of Statistics on Science and Technology*. La década de los 70 se cerró con la puesta en escena de la primera revista científica especializada en este campo: *Scientometrics*, Fundada por Tibor Braun y publicada conjuntamente por la Akademiai Kiado de Hungría.

Los años ochenta suponen uno de los momentos más innovadores para la Bibliometría ya que se comienzan a utilizar nuevos procedimientos estadísticos para el análisis de los datos, además se produce su asentamiento académico en las universidades como pone de manifiesto la creación de instituciones y grupos de investigación especializados. El más destacado de todos ellos es Centre for Science and Technology Studies (CWTS) de Leiden University o la Information Science and Scientometrics Research Unit de la Hungarian Academician of Sciences. El CWTS asentó las bases de los llamados indicadores bibliométricos modernos.

En los noventa la Comisión Europea a través de de la Oficina de Estadísticas de I+D e Innovación de la Dirección General XIII (EUROSTAT) se incorpora a la elaboración de estadísticas. Esta tendencia la siguen la UNESCO y organismos creados ad hoc como es el caso de Francia (OST- Observatoire de Science and Technology), Holanda (NOWT- Dutch Observatory of Science and Technology) o España (FECYT- Fundación Española para la Ciencia y la Tecnología)

Durante la década de los 90 y gracias al nacimiento de Internet la bibliometría cambia radicalmente ya que se pueden consultar las bases de datos de manera remota como es el caso de Medline, Science Citation Index, etc.

El siglo XXI ha traído consigo el surgimiento de la base de datos multidisciplinar, Scopus, que eliminó el monopolio que tenía la base de datos ISI de Thomsom Scientific sobre los índices de citas (Codina, 2005). En este siglo también se han creado nuevos buscadores como Google Scholar, Scirus, CiteSeer (Jacso, 2005). Todos estos cambios se verán reflejados en los indicadores bibliométricos.

2.1.4 Los indicadores bibliométricos

Un indicador es una medida utilizada para evaluar cualquier actividad y proporciona información sobre los resultados de un trabajo realizado. Los indicadores bibliométricos, basados en publicaciones científicas, surgen con el fin de aportar objetividad numérica al proceso de evaluación de las actividades que se desarrollan en el ámbito de la ciencia y la tecnología, aportando información sobre los resultados del proceso investigador, valorando la actividad científica, y la influencia tanto del trabajo como de las fuentes. Desde sus inicios el objetivo primordial de este tipo de medidas ha sido analizar los outputs o resultados científicos. La utilización de este tipo de indicadores con fines de evaluación es hoy una práctica habitual.

Rosa Sancho (1990) recoge en su publicación todos los aspectos que se pueden examinar con los indicadores bibliométricos. De la extensa enumeración de estos aspectos podemos hacer mención a los más importantes:

- Crecimiento de cualquier campo de la ciencia, según la variación cronológica del número de trabajos publicados en él.
- Envejecimiento de los campos científicos, según la “vida media” de las referencias de sus publicaciones.
- Evolución cronológica de la producción científica, según el año de publicación de los documentos.

- Productividad de los autores o instituciones, medida por el número de sus trabajos.
- Colaboración entre los científicos e instituciones, medida por el número de autores por trabajo o centros de investigación que colaboran.
- Impacto o visibilidad de las publicaciones dentro de la comunidad científica internacional, medido por el número de citas que reciben éstas por parte de trabajos posteriores.
- Análisis y evolución de las fuentes difusoras de los trabajos, por medio de indicadores de impacto de las fuentes.
- Dispersión de las publicaciones científicas entre las diversas fuentes, etc.

Moed (1988) afirma que los indicadores bibliométricos son una herramienta provechosa, esencialmente porque proporcionan doble información cuantitativa por un lado sobre la producción (output) y por otro sobre el impacto de esa producción. Asimismo, y a pesar de potenciales restricciones, los indicadores bibliométricos poseen un papel notable en la toma de decisiones en política científica y en evaluaciones del rendimiento de la investigación. Por ejemplo, en investigación básica o pura los indicadores bibliométricos pueden jugar un papel vital como herramienta de evaluación de grupos de investigación grandes o de científicos individuales (Zachos, 1991).

El Factor de Impacto, es el protagonista indiscutible de los indicadores bibliométricos de visibilidad e impacto que se utilizó originariamente como parámetro de prestigio de las revistas y consecutivamente se ha utilizado para la evaluación de resultados científicos. Desde que el investigador norteamericano Eugene Garfield lo propusiera en los años setenta, los índices de impacto se han convertido en uno de los temas núcleo de la Bibliometría. Es el indicador bibliométrico para la evaluación más extendido pero ha sido criticado en algunas ocasiones debido al uso arbitrario, no ponderado que se hace de él (Glänzel & Moed, 2002; Moya-Anegón et al, 2005)

A lo largo del tiempo y para eliminar sesgos, se han ido sucediendo una serie de modificaciones en el indicador con el fin de minimizar, entre otras cosas, las desigualdades entre disciplinas científicas, relativas a los diferentes hábitos de publicación y citación (Braun, 1985; Sen, 1992; Ramírez, 2002). Por otro lado se han revelado en muchos estudios sus limitaciones que interfieren en distintos aspectos del mismo, entre ellas: el reparto sesgado de las citas, las diferencias propias entre disciplinas, los defectos intrínsecos de las bases de datos de donde se extrae la información, etc. y normalmente aparecen las críticas a su uso generalizado y más o menos incorrecto (Nature 2005)

Para salvar las limitaciones del factor de impacto, surgen nuevos indicadores, que como subrayaron Pinski y Narin (1976) parecía más razonable dar un valor más alto a las citas provenientes de una revista prestigiosa que a las citas de una revista periférica. Estos nuevos indicadores se basan en el algoritmo de PageRank de Google. Tal es el caso del SJR (Grupo SCImago, 2007), utilizado en SCImago Journal & Country Rank. Igualmente el Eigenfactor y el Article Influence (Bergstrom, 2007).

Pero es importante señalar que estos indicadores difícilmente se pueden emplear para hacer comparaciones entre disciplinas científicas debido a los diferentes hábitos de citación (Pinski & Narin, 1976), o según las áreas temáticas (diferente media de referencias por artículo), o debido al tamaño de la comunidad científica, o el carácter más analítico o descriptivo de la literatura (Gómez Caridad & Bordons Cangas, 1996), también puede influir las diferentes proporciones de citación a los diferentes tipos de documentos citables (artículos, revistas y cartas) (Lundberg, 2007), o al diferente periodo de tiempo necesario para llegar a la máxima citación (Moed, 2005). Por tanto a veces en procesos de evaluación es necesaria una normalización de los mismos para cada disciplina, de modo que resulten valores comparables entre unas y otras (Moed, 1995; Van Leeuwen & Moed, 2002; Van Leeuwen et al, 2003; Podlubny, 2005; Sombatsompop & Markpin, 2005; Moya-Anegón et al, 2005; Van-Raan, 2006; Lundberg, 2007; Guerrero et al., 2007)

2.1.5 Los diferentes hábitos de citación en la ciencia

Los encargados de evaluar las actividades de I+D+i usan los indicadores bibliométricos como instrumento para valorar sus actividades, pero los diferentes productos, medios de difusión y modos de publicación y/o citación de las distintas áreas y naciones, dificulta este análisis. Como se ha apuntado anteriormente los indicadores difícilmente se pueden emplear para hacer comparaciones entre disciplinas científicas. Uno de los principales factores que lo dificultan son los diferentes hábitos de citación y los perfiles de referenciación que tienen las áreas científicas (Pinski & Narin, 1976).

Como diría Jordi Camí (1997) “desde una perspectiva meramente comparativa de los análisis de citas entre áreas temáticas, publicar un trabajo en una revista de bioquímica cuyo Factor de Impacto es de seis no supone necesariamente la probabilidad de obtener un impacto relativo superior a lo que significa publicar un trabajo de salud pública en una revista de Factor de Impacto igual a cuatro, que es el mayor Factor de Impacto que alcanzan las revistas de esta disciplina. Éstas deberían ser razones suficientes para demostrar a nuestros apreciados compañeros impactólatras que la mezcla y/o la comparación de trabajos de distintas disciplinas basadas únicamente en el FI de la revista es una solemne barbaridad. Precisamente para soslayar esta importante limitación de los análisis de citas, en bibliometría se está recurriendo a distintas fórmulas de relativización y contextualización, cuya aplicación depende del tipo de comparaciones que se desee realizar”.

Es difícil determinar la estructura de la ciencia en disciplinas, aunque son muchos los estudios que admiten como tal el Subject Category Listing del JCR del ISI. Dichas categorías, como indican Pudovkin & Garfield, (2002) fueron desarrolladas hace más de 40 años por métodos manuales para clasificar las revistas científicas, aunque a medida que fueron creciendo se establecieron subdivisiones. Para la asignación de las revistas a categorías se sigue un proceso heurístico que tiene en cuenta las citas y utiliza entre otros el algoritmo de Hayne-Coulson. A pesar de que hay autores que no lo comparten, se considera que las mencionadas categorías ISI son unidades informativas lo suficientemente explícitas como para poder ser utilizadas en la representación de las distintas disciplinas que componen la ciencia en general (Moya-Anegón et al, 2006).

Estas categorías son utilizadas en la actualidad, por ejemplo para la realización de mapas de la ciencia y de este modo observar las relaciones que pueden existir entre las diferentes categorías (Moya-Anegón et al, 2004) aplicando el análisis de citas y técnicas de poda como el algoritmo de Pathfinder (Schvaneveldt, Dearholt & Durso, 1988; Guerrero-Bote et al, 2006) o para la evaluación de la ciencia (Podlubny, 2005; Van Leeuwen & Moed, 2002; Van Leeuwen et al, 2003; Sombatsompop & Markpin, 2005; Moya-Anegón et al, 2005; Moya-Anegón et al, 2008)

La bibliometría no está libre de juicios externos. Aunque fue diseñada para evaluar disciplinas pertenecientes a las Ciencias Básicas o Puras como, las Ciencias Naturales, Ciencias Exactas y de la Vida, hoy en día también es utilizada, en menor medida y con mayores críticas, en la evaluación de las Ciencias Sociales (Lemoine et al., 1997). Una de sus primordiales limitaciones es que no da una perspectiva cierta de lo que ocurre en las ciencias sociales porque los científicos sociales tienen hábitos de publicación y de citación desiguales. Efectivamente, la bibliometría se basa fundamentalmente en revistas "científicas" de amplia visibilidad, y tiende a dejar de lado aspectos y productos característicos de la producción en ciencias sociales. Además, los hábitos de citación de los investigadores que pertenecen a las ciencias sociales son muy particulares. Los libros, manuales, documentos de trabajo, memorias de seminarios y páginas web tienden a ser más citados que las revistas (Ordoñez Matamoros et al., 2009)

Se pueden estudiar muchos aspectos de los hábitos de citación de los países y las diferentes disciplinas científicas pero uno de ellos es su relación con la colaboración científica y por tanto las tendencias de las disciplinas a la hora de citar y referenciar, si suelen referenciar más los trabajos de sus colaboradores o por el contrario se referencian más su propia producción nacional.

Algunos investigadores han analizado el grado de colaboración con la co-autoría de las publicaciones (Carpintero & Peiró, 1983; Alcaín & Gálvez, 1998). En efecto, existe una cierta correlación entre el índice de co-autoría y el impacto de los trabajos, por lo que las comunidades científicas en el aumento de la visibilidad de sus redes crecer y reunir a un mayor número de asociados (Moya et al., 2008).

En general, altos niveles de colaboración conducen a altos niveles de impacto, una mayor calidad de los trabajos publicados, y una mayor productividad de los autores en sus campos de la ciencia en particular (Lewinson & Cunningham, 1991; Narin, Stevens & Whitlow, 1991; Glänzel, 2001; Glänzel, 2002; Leimu & Koricheva, 2005; Katz & Hicks, 1997; Persson, Glanzel & Danell, 2004; Hsu & Huang, 2010; Aksnes, 2003; Moya-Anegón et al. Hassan, González & Moya, 2010)

Los beneficios potenciales de la colaboración científica pueden depender de la disciplina. El efecto de la colaboración en impacto científico parece ser más positivo en las ciencias "duras" como la física y la astronomía, que en las ciencias "suaves" tales como la sociología o las ciencias sociales (Stack, 2002; Bandyopadhyay, 2001; Moed, Bruin, Nederhof & Tijssen, 1991; Bridgstock, 1991), con el comportamiento de la citación a veces difieren considerablemente de un campo a otro (Lancho, Guerrero & Moya, 2010 a,b).

También los beneficios que se obtienen de la colaboración científica dependerán de los diferentes tipos de colaboración (Leimu & Koricheva, 2005; Katz & Hicks, 1997): (a) doméstica (todos los autores de la misma institución); (b) colaboración institucional interna (todos los autores del mismo país, sino de más de una institución) y (c) la colaboración internacional (autores de más de un país) (Leimu & Koricheva, 2005).

Pero este incremento de citación obtenido puede variar de unas naciones a otras (Inzelt, Schubert & Schubert, 2009; Zhao & Guan, 2011). Por diversas razones puede haber diferencias de colaboración entre países, por el tamaño del país (Luukkonen et al., 1993; Ding, Foo & Chowdhury, 1999), por las propias tendencias de los países (hábitos de citación de los países, patrones, costumbres) (Katz & Hicks, 1997; Glänzel, 2001; Singh, 2005; Sooryamoorthy, 2009; Gazni & Didegah, 2011), por su productividad e importancia (países con mayor o menor impacto) (Price & Beaver, 1966; Zuckerman, 1968; Pao, 1981; Glänzel, Schubert & Czerwon, 1999; King, 2004; Zhao & Guan, 2011), o por la proximidad geográfica (países que colaboran sólo con países que están próximos a él, la vecindad geográfica) (Kraut & Egido, 1988; Moed et al., 1991; Katz, 1994; Glänzel & Schubert, 2001; Leta & Chaimovich, 2002; Goldfinch et al, 2003; Leimu & Koricheva, 2005; Ma & Guan, 2005; Figg & Al., 2006)

Por otro lado de unas disciplinas científicas a otras puede haber también diferencias, ya que hay disciplinas en las que el porcentaje de colaboración es mayor que en otras. (Garfield, 1979; Bridgstock, 1991; Bandyopadhyay, 2001; Molteni & Zulueta, 2002; Bandyopadhyay, 2004; Sudhier, 2007; Abt, 2007; Ardanuy et al., 2009; Franceschet & Costantini, 2010; Ardanuy, 2011; Gazni & Didegah, 2011)

2.2 Fuentes utilizadas

Los análisis de citas dependen en gran medida de la base de datos utilizada, por lo que tanto sus potencialidades como sus limitaciones deben ser minuciosamente examinadas (Moed, 2005). De ahí que en nuestro trabajo Scopus y el portal SJR hayan sido utilizados no sólo como fuentes de información sino a la vez como objeto mismo de estudio desde una perspectiva bibliométrica.

2.2.1 Scopus

La fuente original de los datos para elaborar los indicadores y para la realización de este trabajo ha sido Scopus. Se eligió esta base de datos ya que es una novedosa herramienta de navegación que engloba la mayor colección multidisciplinar a nivel mundial de resúmenes, referencias e índices de literatura científica, técnica y médica.

Es una base de datos europea que nace en 2006 y que brinda una gran cantidad de resúmenes y citas. Contiene más de 18.000 títulos de todas las áreas de la ciencia y también incluye más de 5.000 editores internacionales.

Scopus es la mayor base de datos bibliográfica científica a nivel mundial. Ofrece información completa sobre citas, relacionando inequívocamente al documento citante con el citado (si la referencia a este último está incluida en la base de datos), además de otras funcionalidades relacionadas con este asunto. Debido a ello, se pueden realizar análisis e informes de la producción científica y su impacto que difícilmente se pueden llevar a cabo utilizando otra base de datos.

Scopus incluye dos métricas de impacto como son SCImago Journal Rank (SJR) y también Source Normalized Impact per Paper (SNIP) de la Universidad de Leyden. Ambos ofrecidos en Scopus Journal Analyzer.

Para extraer los datos para este estudio las consultas a Scopus se hicieron en SQL para obtener la información en la base de datos relacional de ORACLE. La información bibliográfica fuente, descargada de Scopus con fines de investigación, se ha almacenado de forma distribuida en un sistema de bases de datos con toda la información integrada y de forma relacionada.

Para la realización de gráficos, tablas y la aplicación de indicadores, la información necesaria se volcó a hojas de cálculo donde se procedió al cálculo de los indicadores.

2.2.2 SCImago Journal & Country Rank

De las publicaciones vaciadas de Scopus, el estudio se ha realizado con aquellas que aparecen en el Web accesible a través de SCImago Journal & Country Rank (SJR). El SJR es un portal creado por el grupo SCImago que incluye indicadores científicos de revistas y países desarrollados a partir de los datos suministrados por Scopus a principios de 2007 (SJR, 2007). En la actualidad, este es el único portal de acceso libre que proporciona indicadores científicos de todas las revistas incluidas en Scopus, consintiendo de esta manera comparar el impacto de las revistas de la Web Of Science con Scopus.

Se han utilizado los datos de este portal para el cálculo de las distribuciones de impacto y las tasas de exportación e importación, siempre refiriéndonos a citas y referencias realizadas en trabajos publicados en 2006 de trabajos publicados en 2003, 2004 y 2005.

Para descubrir la influencia de los perfiles de referencia anuales y los hábitos de citación en las distribuciones bibliométricas de las diferentes categorías y áreas científicas se han utilizado los documentos publicados en 2007 presentes en la base de datos de Scopus (el número de registros utilizados correspondientes a los documentos publicados en el 2007 incluidos en Scopus es de 1.645.036). Estos documentos están clasificados por áreas y así mismo se recuperan todas sus referencias, tanto las que se refieren a documentos presentes en la base de datos en años anteriores (el número de referencias dentro del universo de la base de datos es 5.842.139) como las referencias que se dirigen fuera del universo de la

base de datos (un total de 41.838.425 referencias contabilizadas)

En la base de datos Scopus están categorizados los Ítems, sin embargo SCImago lo que categoriza son las revistas metiendo cada una de ellas en las categorías donde tienen Ítems.

Por ello en el presente trabajo se han agrupado de dos formas los datos, según la categorización de las Revistas donde cada registro corresponde a una revista en cada categoría y según la categorización de los Ítems, que resulta ventajoso ya que se evita el solapamiento.

Hay que decir que a lo largo del tiempo se han ido sucediendo varias metodologías para la clasificación de revistas científicas: la más usual es la que agrupa las revistas científicas en las llamadas categorías temáticas. Actualmente existen varias clasificaciones de este tipo, por grupos científicos (Centre for Science and Technology Studies, CWTS), por áreas (Agencia Nacional de Evaluación y Prospectiva, ANEP), o por categorías (Journal Citation Report, JCR y SCImago Journal and Country Rank, SJR). Cada revista en el SJR y JCR es asignada a una o más categorías. En el SJR las 15.000 revistas indexadas en Scopus son clasificadas en 27 áreas y en 295 categorías y en el JCR las 8.700 revistas incluidas en el WoS en 127 en su Science Edition y 55 en su Social Sciences Edition. Esta clasificación se basa principalmente en una supervisión de los títulos y de sus áreas de competencia y sólo parcialmente en relaciones de citación entre las revistas. Pinski y Narin (1976) exploraron el agrupamiento de las revistas en categorías basándose en las relaciones de citación entre las revistas.

Otro método alternativo para la clasificación de las revistas científicas es el análisis de co-citación de revistas, desarrollado por Moya-Anegón et. al. (2006), basado en el número de veces que un par de revistas es citada por la misma fuente.

Por ello aunque en Scopus existen 295 categorías científicas, en el presente estudio se suprimen aquellas categorías que tengan menos de 6 revistas. Por tanto el total de categorías que quedaría para analizar sería 261. Las categorías científicas que se van a estudiar son: Accounting, Acoustics and Ultrasonics, Advanced and Specialized Nursing, Aerospace Engineering, Aging, Agricultural and Biological Sciences (miscellaneous), Agronomy and Crop Science, Algebra and Number Theory, Analysis, Analytical Chemistry, Anatomy, Anesthesiology and Pain Medicine, Animal Science and Zoology, Anthropology, Applied Mathematics, Applied Microbiology and Biotechnology, Applied Psychology, Aquatic Science, Archeology, Architecture, Artificial Intelligence, Arts and Humanities (miscellaneous), Astronomy and Astrophysics, Atmospheric Science, Atomic and Molecular Physics, and Optics, Automotive Engineering, Behavioral Neuroscience, Biochemistry, Biochemistry, Genetics and Molecular Biology (miscellaneous), Bioengineering, Biological Psychiatry, Biomaterials, Biomedical Engineering, Biophysics, Biotechnology, Building and Construction, Business and International Management, Business, Management and Accounting (miscellaneous), Cancer Research, Cardiology and Cardiovascular Medicine, Catalysis, Cell Biology, Cellular and Molecular Neuroscience, Ceramics and Composites, Chemical Engineering (miscellaneous), Chemical Health and Safety,

Chemistry (miscellaneous), Civil and Structural Engineering, Clinical Biochemistry, Clinical Psychology, Cognitive Neuroscience, Colloid and Surface Chemistry, Communication, Community and Home Care, Complementary and Alternative Medicine, Computational Mathematics, Computational Mechanics, Computational Theory and Mathematics, Computer Graphics and Computer-Aided Design, Computer Networks and Communications, Computer Science (miscellaneous), Computer Science Applications, Computer Vision and Pattern Recognition, Computers in Earth Sciences, Condensed Matter Physics, Control and Optimization, Control and Systems Engineering, Critical Care and Intensive Care Medicine, Critical Care Nursing, Cultural Studies, Decision Sciences (miscellaneous), Demography, Dentistry (miscellaneous), Dermatology, Development, Developmental and Educational Psychology, Developmental Biology, Developmental Neuroscience, Discrete Mathematics and Combinatorics, Drug Discovery, Earth and Planetary Sciences (miscellaneous), Earth-Surface Processes, Ecological Modeling, Ecology, Ecology, Evolution, Behavior and Systematics, Economic Geology, Economics and Econometrics, Economics, Econometrics and Finance (miscellaneous), Education, Electrical and Electronic Engineering, Electrochemistry, Electronic, Optical and Magnetic Materials, Embryology, Emergency Medicine, Endocrinology, Endocrinology, Diabetes and Metabolism, Energy (miscellaneous), Energy Engineering and Power Technology, Engineering (miscellaneous), Environmental Chemistry, Environmental Engineering, Environmental Science (miscellaneous), Epidemiology, Experimental and Cognitive Psychology, Filtration and Separation, Finance, Fluid Flow and Transfer Processes, Food Science, Forestry, Fuel Technology,

Gastroenterology, Gender Studies, Genetics, Genetics (clinical), Geochemistry and Petrology, Geography, Planning and Development, Geology, Geometry and Topology, Geophysics, Geotechnical Engineering and Engineering Geology, Geriatrics and Gerontology, Global and Planetary Change, Hardware and Architecture, Health (social science), Health Informatics, Health Information Management, Health Policy, Health Professions (miscellaneous), Health, Toxicology and Mutagenesis, Hematology, Hepatology, Histology, History, History and Philosophy of Science, Horticulture, Human Factors and Ergonomics, Human-Computer Interaction, Immunology, Immunology and Allergy, Immunology and Microbiology (miscellaneous), Industrial and Manufacturing Engineering, Infectious Diseases, Information Systems, Information Systems and Management, Inorganic Chemistry, Insect Science, Instrumentation, Internal Medicine, Language and Linguistics, Law, Leadership and Management, Library and Information Sciences, Linguistics and Language, Logic, Management Information Systems, Management of Technology and Innovation, Management Science and Operations Research, Management, Monitoring, Policy and Law, Marketing, Materials Chemistry, Materials Science (miscellaneous), Mathematical Physics, Mathematics (miscellaneous), Mechanical Engineering, Mechanics of Materials, Media Technology, Medical Laboratory Technology, Medicine (miscellaneous), Metals and Alloys, Microbiology, Microbiology (medical), Modeling and Simulation, Molecular Biology, Molecular Medicine, Multidisciplinary, Nature and Landscape Conservation, Nephrology, Neurology, Neurology (clinical), Neuropsychology and Physiological Psychology, Neuroscience (miscellaneous), Nuclear and High Energy Physics, Nuclear Energy

and Engineering, Numerical Analysis, Nursing (miscellaneous), Nutrition and Dietetics, Obstetrics and Gynecology, Ocean Engineering, Oceanography, Oncology, Ophthalmology, Organic Chemistry, Organizational Behavior and Human Resource Management, Orthopedics and Sports Medicine, Otorhinolaryngology, Paleontology, Parasitology, Pathology and Forensic Medicine, Pediatrics, Perinatology and Child Health, Pharmaceutical Science, Pharmacology, Pharmacology (medical), Pharmacology, Toxicology and Pharmaceutics (miscellaneous), Pharmacy, Philosophy, Physical and Theoretical Chemistry, Physical Therapy, Sports Therapy and Rehabilitation, Physics and Astronomy (miscellaneous), Physiology, Physiology (medical), Plant Science, Political Science and International Relations, Pollution, Polymers and Plastics, Process Chemistry and Technology, Psychiatric Mental Health, Psychiatry and Mental Health, Psychology (miscellaneous), Public Administration, Public Health, Environmental and Occupational Health, Pulmonary and Respiratory Medicine, Radiation, Radiological and Ultrasound Technology, Radiology, Nuclear Medicine and Imaging, Rehabilitation, Religious Studies, Renewable Energy, Sustainability and the Environment, Reproductive Medicine, Rheumatology, Safety Research, Safety, Risk, Reliability and Quality, Sensory Systems, Signal Processing, Social Psychology, Social Sciences (miscellaneous), Sociology and Political Science, Software, Soil Science, Space and Planetary Science, Spectroscopy, Statistical and Nonlinear Physics, Statistics and Probability, Statistics, Probability and Uncertainty, Strategy and Management, Stratigraphy, Structural Biology, Surfaces and Interfaces, Surfaces, Coatings and

Films, Surgery, Theoretical Computer Science, Tourism, Leisure and Hospitality Management, Toxicology, Transplantation, Transportation, Urban Studies, Urology, Veterinary (miscellaneous), Virology, Waste Management and Disposal, Water Science and Technology.

Éstas categorías se han reclasificado en 27 áreas o aglomerados de orden superior, adoptando para ello la división del conocimiento propuesta por Elsevier, teniendo en cuenta que una única categoría temática puede pertenecer a diferentes áreas (Moya et. al, 2004)

El sistema de clasificación de Scopus contiene 27 áreas, lo que refleja su amplia cobertura temática. Las grandes áreas clasificadas son las siguientes:

Life Sciences: Agricultural and Biological Sciences, Biochemistry, Genetics and Molecular Biology, Immunology and Microbiology, Neuroscience, Pharmacology, Toxicology and Pharmaceutics.

Physical Sciences: Chemical Engineering, Chemistry, Computer Science, Earth and Planetary Science, Energy, Engineering, Environmental Science, Materials Science, Mathematics, Physics and Astronomy.

Social Sciences: Arts and Humanities, Business, Management and Accounting, Decision Sciences, Economics, Econometrics and Finance, Psychology, Social Sciences.

Health Sciences: Medicine, Nursing, Veterinary, Dentistry, Health Professions.

Para la parte de la colaboración científica internacional por un lado se han elegido los 20 países con mayor producción en 2004, evitando así problemas de significación estadística en los países de menor producción.

Los países elegidos han sido: Estados Unidos, China, Reino Unido, Japón, Alemania, Francia, Canadá, Italia, España, India, Australia, Corea, Holanda, Rusia, Brasil, Taiwán, Suiza, Polonia y Turquía.

Más tarde para analizar la citación y la referenciación entre los países colaboradores se seleccionaron, para acotar el estudio y por razones de espacio, los 9 países más productivos de 2004, Estados Unidos, China, Reino Unido, Japón, Alemania, Francia, Canadá, Italia, España. Estos países compondrán el grupo A y el grupo B. En el grupo denominado A, de los que se estudia el destino de sus referencias en el caso de colaborar o no, y el grupo denominado B, de los que se estudia la citación obtenida y procedente de los países A en el caso de colaborar o no con ellos.

2.3 Aplicación de Indicadores

2.3.1 Indicadores aplicados a las distribuciones de impacto

- Impacto Total: se calcula con base a un periodo de 3 años.
 - A = Número de veces en que los artículos publicados en el periodo 2003-2004-2005 han sido citados por las publicaciones a las que se les da seguimiento a lo largo del año 2006.
 - B = Número de artículos publicados en el periodo 2003-2004-2005.
 - Impacto Total 2006 = A/B
- Impacto Interno: de manera similar al Factor de Impacto, pero, teniendo en cuenta solamente las citas procedentes de la propia Categoría.
- Impacto Externo: de manera similar al Factor de Impacto, pero, teniendo en cuenta solamente las citas no procedentes de la propia Categoría.

Se han hecho representaciones de los impactos frente a los rangos de ordenación decreciente, para cada categoría. Para cada una de ellas se han calculado líneas de tendencia exponenciales. Con respecto a estas se han utilizado como indicadores de la distribución los coeficientes a (pendiente) y b (desplazamiento), y de la adaptación a la tendencia exponencial el coeficiente R^2 y el error cuadrático. Este último indicador ha sido de gran utilidad para detectar las categorías cuya distribución tiene un pico más pronunciado que la función exponencial.

Finalmente se han correlacionado los distintos indicadores para ver la influencia de cada indicador en la adaptación de cada distribución a la tendencia logarítmica.

2.3.2 Indicadores de tasas de exportación e importación

- **Tasa de Exportación de Conocimiento (%Export):** Se puede definir como el porcentaje de citas recibidas desde documentos de otras categorías con respecto al total de citas recibidas.

$$\% \text{ Export} = \frac{100 \cdot (TotCitation - SubCitation)}{TotCitation}$$

- **Tasa de Importación de Conocimiento (%Import):** Se puede definir como el porcentaje de referencias a documentos de otras categorías con respecto al total de referencias.

$$\% \text{ Import} = \frac{100 \cdot (References - SubCitation)}{References}$$

2.3.3 Indicadores para medir la influencia de los perfiles de referencia y los hábitos de citación

- **Porcentaje de referencias en Scopus:** Porcentaje de referencias de las diferentes categorías a documentos que están en la base de datos Scopus.
- **Porcentaje de referencias en la ventana de citación de 3 años:** Porcentaje de referencias a trabajos publicados en la ventana de citas de 3 años (2004, 2005, 2006)
- **Porcentaje de referencias en la ventana de citación de 2 años:** Porcentaje de referencias a trabajos publicados en la ventana de citas de 2 años (2005, 2006)
- **Porcentaje en la ventana de citación de 3 años en Scopus:** Porcentaje de referencias a trabajos publicados en la ventana de citación de 3 años de citas incluidos en Scopus (2004, 2005, 2006)
- **Porcentaje en la ventana de citación de 2 años en Scopus:** Porcentaje de referencias a trabajos publicados en la ventana de citación de 2 años incluidos en Scopus (2005, 2006)
- **Referencias por artículo:** Promedio de número de referencias por artículo que tienen las categorías y áreas.

- **Referencias por artículo en la ventana de citación de 3 años:** Promedio de número de referencias a trabajos publicados en la ventana de 3 años de citas (2004, 2005, 2006) por artículo.

- **Referencias por artículo en la ventana de citación de 2 años:** Promedio de número de referencias a trabajos publicados en la ventana de 2 años de citas (2005, 2006) por artículo.

- **Referencias a trabajos de Scopus por artículo en la ventana de citación de 3 años:** Promedio de número de referencias a trabajos publicados en la ventana de 3 años incluidos en Scopus (2004, 2005, 2006) por artículo.

- **Referencias a Scopus por artículo en la ventana de citación de 2 años:** Promedio de número de referencias a trabajos publicados en los 2 años incluidos en la ventana de citas de Scopus (2005, 2006) por artículo.

- **Referencias por artículos en Scopus:** Promedio de número de referencias a documentos en Scopus por artículo.

- **SJR (SCImago Journal Rank):** Dicho indicador expresa el número medio ponderado de las citas recibidas en un año por los documentos publicados en la revista seleccionados en los tres años anteriores (SCImago Journal & Country Rank, 2007-2009)

- **Factor de Impacto 4 años:** Puede ser visto como el número medio de citas en 2007 a los documentos de una revista que se publicaron durante los cuatro años anteriores.
- **Factor de Impacto 3 años:** Puede ser visto como el número medio de citas en 2007 a los documentos de una revista que se publicaron durante los tres años anteriores.
- **Factor de Impacto 2 años:** Puede ser visto como el número medio de citas en 2007 a los documentos de una revista que se publicaron durante los dos años anteriores.

2.3.4 Indicadores de colaboración científica

Para estudiar cómo son los hábitos de citación de los 20 países más productivos de 2004 cuando colaboran científicamente con otros países se calculan estos indicadores:

- **Citation per papers:** Promedio de citación obtenida por los documentos de 2004 de los documentos de 2005-2007.
- **References per papers:** Promedio de referencias hechas en documentos de 2005-2007 hacia documentos del 2004.
- **Influence on paper per paper:** Influencia que ejercen los documentos de 2004 en los documentos de 2005 a 2007.

A diferencia de los anteriores indicadores que son meras ratios, este indicador tiene en cuenta tanto el tamaño del conjunto de target papers (lógicamente a mayor tamaño mayor posibilidad de captar citas) y el tamaño del conjunto de source papers (lógicamente a mayor tamaño mayor posibilidad de captar citas).

De este modo se tiene en cuenta no solo el promedio de citas por trabajo sino también el máximo que podría llegar a tener. Normalmente no se ha tenido en cuenta el tamaño del conjunto de source papers, puesto que este suele ser el conjunto de la producción mundial. Sin embargo, cuando se hacen agrupaciones teniendo en cuenta la procedencia de la citación esto puede tener una influencia capital si generan conjuntos fuentes de diferentes tamaños. Este tipo de indicadores fueron utilizados por primera vez por Pichappan (1995) para el análisis de la citación de las revistas de un campo temático.

En todos los casos, tanto los targets como los source papers se desglosan por procedencia (domestica (de la misma nación) o no domestica) y por tipos de colaboración (con o sin colaboración internacional) (solamente se distingue la colaboración con instituciones de distintos países). Incluso se hacen desgloses específicos para la citación hacia o procedente de los países colaboradores.

Para analizar cómo es la citación y la referenciación entre los 9 países colaboradores más productivos de 2004, se utilizan una serie de indicadores:

– **Citation Rate Increment of Colaborator (CRIC)**

$$CRIC_{A \rightarrow B} = \frac{\frac{Citation_{A \rightarrow AB}}{Citation_{AB}}}{\frac{Citation_{A \rightarrow B}}{Citation_B}}$$

Citation B: Citas obtenidas por los documentos de B.

Citation AB: Citas obtenidas por los documentos de A en colaboración con B.

CitationA->B: Citas procedentes de A obtenidas por los documentos de B.

CitationA->AB: Citas procedentes de A obtenidas por los documentos de A en colaboración con B.

Como se puede observar, el mismo conjunto de documentos en colaboración da lugar a dos indicadores el Citation Rate Increment of Colaborator A cuando colabora con B, y el Citation Rate Increment of Colaborator B cuando colabora con A. A la mis vez, permite hallar dos promedios para cada país, su Citation Rate Increment Average when Colaborate (CRIAC), y el Citation Rate Increment obtained from its collaborators (CRIOC).

– **Domestic Citation Rate Comparison when Collaborate (DCRCC):**

$$DCRCC_{A \rightarrow B} = \frac{\frac{Citation_{A \rightarrow AB}}{Citation_{AB}}}{\frac{Citation_{A \rightarrow A}}{Citation_A}}$$

Como se puede observar, el mismo conjunto de documentos en colaboración da lugar a dos indicadores el Domestic Citation Rate Comparison of A when Collaborate with B, y el Domestic Citation Rate Comparison of B when Collaborate with A. A la misma vez, permite hallar dos promedios para cada país, su Domestic Citation Rate Comparison Average (DCRCA) y Domestic Citation Rate of Collaborators Comparison Average (DCRCCA).

– **Domestic Impact Rate Increment when Collaborate (DIRIC)**

$$DIRCC_{A \rightarrow B} = \frac{\frac{Citation_{A \rightarrow AB}}{Ndoc_{AB}}}{\frac{Citation_{A \rightarrow A}}{Ndoc_A}}$$

Como se puede observar, el mismo conjunto de documentos en colaboración da lugar a dos indicadores el Domestic Impact Rate Increment of A when Collaborate with B, y el Domestic Impact Rate Increment of B when Collaborate with A. A la mis vez, permite hallar dos promedios para cada país, su Domestic Impact Rate Increment Average (DIRIA) y Domestic Impact Rate of Collaborators Increment Average (DIRCIA).

3 Justificación y Objetivos

3.1 Evaluar de la Investigación

Que la investigación es crucial para el desarrollo y la evaluación ineludible para la investigación es algo hoy día indiscutible. Por ende, el análisis y la evaluación de la información y el conocimiento es un componente indispensable para todos los programas públicos de investigación, tecnología y desarrollo que se efectúan en nuestra sociedad. Es allí donde las Ciencias de la Información ofrecen un apoyo incalculable, al desarrollar técnicas y herramientas que permitan medir la producción de conocimiento y su transformación en bienes (Arencibia & Moya, 2008)

Gracias a las diferentes métricas (bibliometría, cienciometría e informetría, etc) se han desarrollado indicadores para medir la información. Al margen de las ventajas e inconvenientes que tienen estos indicadores ampliamente debatidos (Geisler, 2005; Nederhof, 2005) constituyen la herramienta estrella en la gestión de la política científica y tecnológica y en procesos de toma de decisiones importantes.

El desarrollo, expansión y consolidación de los sistemas de Ciencia y Tecnología se ha conllevado a que aparezcan nuevas necesidades emergidas de la propia sociedad y de las distintas políticas científicas. De este modo la evaluación se convierte en una herramienta clave para la asignación o repartición de recursos materiales o financieros. Proporciona a su vez la definición de nuevos alicientes y la validación de los resultados en ciertas áreas científicas en relación con las necesidades nacionales (Sanz Menéndez, 2004)

En el siglo XXI la evaluación de la investigación implica una concepción totalizadora y multidimensional, donde la revisión por peer-reviews constituye una unidad más, en conjunto con cuestionarios especializados, sistemas econométricos, estudios con perspectivas futuras y análisis bibliométricos. Esta imagen que tenemos de la evaluación como herramienta tecnológica para la precisión en la investigación, en sus resultados, sus instituciones y sus autores, ayudan a la efectividad y eficiencia de los sistemas de Investigación + Desarrollo + Innovación (I+D+I), sustentándolos de una mayor coherencia y visión estratégica, que posibilita su integración a los sistemas de gestión y dirección de la ciencia y a los procesos de toma de decisiones.

De ahí, la importancia y necesidad de estudiar, explorar y desarrollar nuevas técnicas, herramientas y metodologías para conseguir una óptima evaluación de la investigación; en este trabajo este es nuestro objetivo principal.

3.2 Limitaciones de los indicadores bibliométricos

Los indicadores bibliométricos que se utilizan para la evaluación no solamente tienen en cuenta aspectos cuantitativos sino también aspectos cualitativos basándose en las referencias de los trabajos, incluidas también en las principales bases de datos (ISI o Scopus, principalmente), y en la idea de que, a pesar de sus distintas motivaciones (Brooks, 1995), una cita es un reconocimiento a un trabajo anterior. Sin embargo, en muchos momentos se ha puesto en duda la aplicación de los indicadores bibliométricos y las propias bases de datos internacionales a ciertas disciplinas científicas. Para algunos, los indicadores bibliométricos, contruidos a partir de estas bases de datos son efectivos en contextos en los que la investigación se difunde esencialmente a través de revistas científicas que se ubican normalmente dentro de áreas de las Ciencias Básicas (Filippo & Fernández, 2002)

Ya Pinski & Narin (1976) apuntaban una serie de limitaciones a la hora de aplicar equitativamente los indicadores a las diferentes áreas. La primera de las limitaciones fue que los reviews tenían mayor factor de impacto en comparación con otros medios de publicación de trabajos. La segunda limitación que encontraron fue que las citas no estaban ponderadas. Y la tercera es que no había normalización de las diferentes características que tienen a la hora de referenciar los diferentes campos científicos.

En Guerrero et al (2007) concluyeron que debido a que algunas de las disciplinas científicas son más exportadoras de ideas porque el conocimiento que se genera dentro de ellas es visible desde otras que lo importan. Tendrá que ver en esto, ¿que esas disciplinas estén en la base de datos del ISI o en Scopus que tiene una cobertura mayor? ¿Y si se cambia la ventana de citación cambiaran las distribuciones? En este estudio se analizará la coherencia entre las tasas de exportación e importación de ambas bases de datos. Se llegara a comprobar si existe correlación entre los indicadores bibliométricos (SJR y Factor de Impacto) con el tamaño de referencias que contienen las disciplinas científicas para ver si influye dicho tamaño en las distribuciones.

3.3 La falta de cobertura de las bases de datos y los hábitos de citación

Autores han apuntado una serie de causas que pueden influir en las distintas distribuciones y promedios o la distinta significación estadísticas de los indicadores bibliométricos en las distintas disciplinas: la falta de cobertura de las bases de datos en ciertas áreas (Braun et al., 2000, Grupo SCImago, 2006, Moya-Anegón et al., 2007) tanto por las revistas, como principalmente por los tipos documentales y los hábitos de citación de las diferentes áreas científicas (Broadus, 1971; Clemens et al., 1995; Cronin et al., 1997; Hargens, 2000; Hicks & Potter, 1991; Kyvik, 2003; Lewison, 2001; Lindholm-Romantschuk & Warner, 1996; Nederhof et al., 1989; Nock, 2001; Price, 1970; Small & Crane, 1979; Thompson, 2002).

Respecto a las coberturas, existen varios estudios que viene a concluir que no existen grandes sesgos en cuanto a la cobertura. Braun et al. (2000) compararon la cobertura de ISI-WoS con relación al Ulrich's International Periodicals Directory, universalmente aceptado como la base de datos que contiene la casi totalidad de las revistas editadas en el mundo. Más tarde en un trabajo de Moya-Anegón et al., (2007) se analizó la cobertura de Scopus en relación con Ulrich's Core tomando como modelo la metodología de Braun et al., (2000). Dichos trabajos concluyeron que la cobertura en ambas bases de datos era equilibrada, estadísticamente hablando, en términos de temas, países, idiomas y editores con excepción de las Ciencias Sociales, Artes y las Humanidades.

Parece que otro de los motivos por lo que los indicadores son más bajos en las áreas de Ciencias Sociales y Humanidades son los diferentes hábitos de citación, que es un punto clave en investigación científica ya que los trabajos científicos deben poseer referencias. El criterio primordial es que toda investigación se respalde en trabajos anteriores. Un apartado digno de destacar son los frentes de investigación en campos nuevos, en los cuales existe una innovación de todo tipo y resulta difícil, sino imposible, citar fuente alguna. (Dimitri, 2004)

Pero cada área científica tiene sus propios hábitos de citación, por ejemplo en el área de Ciencias Sociales y Humanidades hay una menor utilización de las revistas como vehículo de publicación en beneficio de las monografías que, a su vez, también son más citadas (Broadus, 1971; Clemens et al., 1995; Cronin et al.,

1997; Hargens, 2000; Hicks & Potter, 1991; Kyvik, 2003; Lewison, 2001; Lindholm-Romantschuk & Warner, 1996; Nederhof et al., 1989; Nock, 2001; Price, 1970; Small & Crane, 1979; Thompson, 2002). Las bases de datos de ISI y Scopus únicamente incluyen contenidos de revistas dejando de lado materiales como las monografías, o los informes, que son de gran relevancia en el proceso de comunicación de los resultados de la investigación en algunas disciplinas, como las Ciencias Sociales, Humanidades o las Ingenierías, que tienen tendencia a publicar en congresos.

Nederhof (2006) busca soluciones para esta situación, y cree que la realización de estudios bibliométricos en Ciencias Sociales y Humanidades debe considerar, la inclusión de documentos fuente, ajenos a los indexados por el ISI o Scopus.

Hay que tener en cuenta también, que en las Ciencias Sociales hay una mayor propensión a la investigación individual, frente a la investigación en equipo habitual en Ciencias Básicas. Cabe agregar que existe una tendencia a que aumente el impacto de las publicaciones según aumenta el número de autores y por tanto, la visibilidad que se obtiene en colaboración es mayor (Moya-Anegón et. al., 2006)

Como, para los distintos indicadores bibliométricos lo que se utilizan son las referencias, y no todas las referencias, sino solamente las referencias con destino a trabajos incluidos en la propia base de datos, nuestra hipótesis es que el perfil anual de referencias a documentos incluidos en el propio universo de la base de

datos tendrá una gran influencia en la significación estadística de dichos indicadores en las distintas áreas científicas. Por este motivo, nos proponemos estudiar dicho perfil, para los documentos publicados en un año natural, pudiendo determinar la proporción de referencias que tienen como destino documentos de la propia base de datos, así como su distribución anual. De este modo también podremos determinar la influencia de la ventana de citación en los distintos indicadores.

¿Qué influencia tendrá el perfil de referenciación y los hábitos de citación de las disciplinas científicas en los indicadores bibliométricos en los basados en los factores de impactos? ¿Y en los basados en el PageRank que ponderan las citas? ¿Será solamente una cuestión de cobertura de las bases de datos? ¿Se podrá obtener alguna idea a aplicar en las grandes bases de datos referenciales para mejorar la evaluación global de la ciencia?

Por otro lado y con el fin de estudiar los hábitos de citación, Gómez Caridad et al. (1999) consideran que la colaboración internacional aumenta la visibilidad de los trabajos de investigación, al publicarse en revistas de mayor impacto, que los trabajos en colaboración nacional. Narin & Whitlow (1990) han encontrado certeza de que los trabajos, con múltiple autoría internacional, duplican la frecuencia de citas de aquellos que no presentan colaboración internacional. Schmoch & Shubert (2008) sugieren que los documentos con colaboración internacional son altamente citados porque su comunidad es potencialmente más

grande. Puesto en otras palabras: documentos con colaboración internacional serán más citados simplemente porque más gente los conoce.

Esto y el solapamiento de las comunidades podrían explicar la diferencia de impacto entre las colaboraciones en los diferentes ámbitos, ya que en ámbitos cercanos el solapamiento de comunidades es mayor que en ámbitos más alejados. A ello se suma un posible sesgo a lo nacional en los distintos países que podría favorecer a los trabajos con colaboración internacional.

Tras estas aproximaciones donde se viene a concluir que si se colabora, especialmente con instituciones de diferentes países, se obtendrá un mayor impacto, mayor tasa de citación que en caso de no colaborar, ¿es cierto que los trabajos en colaboración internacional obtienen un mayor impacto? ¿De dónde viene ese impacto? ¿Viene en mayor medida del resto de naciones con las que se colabora? ¿Podrá deberse este impacto a los socios colaboradores internacionales? ¿Obtienen a la vez los trabajos en colaboración más impacto nacional? ¿Tienen los países un sesgo de citas a su propio país? ¿Tendrán más impacto los trabajos en colaboración por ser nacionales de más países?

3.4 Objetivos específicos de esta investigación

1. Comprobar si las distribuciones de impacto del estudio de Guerrero et al (2007) son generalizables.
 - Diseñar un experimento similar al de 2007 con otra base de datos y otras características temporales.
 - Averiguar si eligiendo otra ventana de citación u optando por una fecha más reciente son análogas las distribuciones de impacto.
 - Descubrir si es la propia estructura de disciplinas en las bases de datos o la forma de clasificar las revistas lo que hace que los datos puedan resultar desiguales.
 - Analizar si que una disciplina se adapte bien a una ley depende de la base de datos seleccionada.

2. Probar si las distribuciones de las tasas de exportación e importación del experimento de ISI-1997 son universales.
 - Analizar si son parecidas esas tasas en la base de datos de Scopus que tiene mayor cobertura.
 - Estudiar si influyen la forma de clasificar las revistas que tiene Scopus en dichas tasas.
 - Observar si al cambiar la ventana de citación o elegir una fecha más reciente cambian las distribuciones.

3. Estudiar los hábitos de citación en las distintas áreas científicas.
 - Analizar los perfiles de referenciación de las distintas disciplinas.
 - Revelar si intervienen los propios perfiles de referencia anuales en las distribuciones.

4. Comprobar si influyen los hábitos de citación o la propia cobertura de la base de datos en los indicadores bibliométricos tanto el Factor de Impacto como los que ponderan las citas basadas en el Page Rank como es el caso del SJR.
 - Comprobar si influyen en los indicadores bibliométricos de una disciplina, su tamaño o el número de referencias promedio.
 - Estudiar la influencia de los perfiles de referencia y los hábitos de citación en los indicadores bibliométricos.

5. Examinar si la colaboración científica, especialmente la colaboración internacional conduce a altos índices de impacto.
 - Detectar si existe correspondencia entre el tráfico de citas a/desde los países colaboradores y la producción de dichos países.
 - Analizar la procedencia y destino de la citación de los diferentes países y los diferentes tipos de trabajos, si procede o destina en mayor medida al resto de naciones con las que se colabora o puede deberse a los socios colaboradores internacionales.

- Estudiar si obtienen a la vez los trabajos en colaboración más impacto nacional.
 - Averiguar si existe un sesgo de citas a su propio país mayor que a otros países.
6. Estudiar cómo varía el incremento de referenciación a los países con los que se colabora y la citación recibida de los países colaboradores.
- Analizar qué países proporcionan mayor incremento de referencias hacia los países colaboradores y que países obtienen mayor incremento de citas de los países colaboradores.
 - Estudiar cómo es la referenciación doméstica a los trabajos en colaboración en comparación a los trabajos sin colaboración.
 - Investigar si existe incremento en el impacto nacional, que los países proporcionan a su producción en colaboración y el que obtienen de sus colaboradores.
 - Indagar como varían estos incrementos de unas disciplinas a otras.

4 Discusión de resultados

En esta sección presentamos la discusión conjunta de los resultados obtenidos en los distintos trabajos de acuerdo con el tema de investigación tratado en el conjunto global de la tesis.

El tema principal de investigación ha sido ya introducido junto con los objetivos planteados en el apartado anterior; por lo tanto exponemos los resultados derivados de nuestro estudio.

En este apartado incluimos únicamente la discusión de los resultados dando respuesta a las preguntas de la investigación más relevantes ya que éstas han quedado expuestas y desarrolladas con detalle en cada publicación junto con las fuentes, metodologías utilizadas, limitaciones encontradas y líneas futuras de investigación propuestas.

En el estudio de Guerrero-Bote (2007) se incluyó una verificación exhaustiva de todas las categorías del Journal Citation Index en 1997. Concluyendo, entre otras cosas, que hay disciplinas científicas más exportadoras o importadoras de conocimiento que otras.

Para el presente estudio, de las publicaciones vaciadas en Scopus, el cálculo se ha hecho con aquellas que aparecen en el Web accesible a través de SCImago Journal & Country Rank (SJR) para el cálculo de las distribuciones de impacto y las tasas de exportación e importación, siempre ciñéndose a citas y referencias realizadas en trabajos publicados en 2006 de trabajos publicados en 2003, 2004 y 2005.

Se cuenta también con la anterior hipótesis del artículo ‘The Iceberg Hypothesis: Inport-Export of Knowledge between scientific subject categories’ donde se observó la alta correlación existente entre la tasa de exportación de conocimiento y la tasa de importación de conocimiento, es decir entre el porcentaje de citas recibidas de otras categorías (con respecto a todas las citas recibidas) y el porcentaje de referencias realizadas a otras categorías (con respecto al total de referencias realizadas).

Tabla 1. Coeficiente R2 y error cuadrático para las categorías con distribución de impactos en Scopus 2006.

CATEGORÍA	R ²	E ²	CATEGORÍA	R ²	E ²
Education	0.995	0.004	Ocean Engineering	0.976	0.020
Algebra and Number Theory	0.982	0.004	Health (social science)	0.977	0.020
History and Philosophy of Science	0.982	0.005	Physical Therapy, Sports Therapy and Rehabilitation	0.977	0.020
Religious Studies	0.974	0.005	Orthopedics and Sports Medicine	0.979	0.021
Architecture	0.942	0.005	Nuclear Energy and Engineering	0.963	0.021
Control and Optimization	0.965	0.006	Hematology	0.987	0.021
Environmental Chemistry	0.992	0.007	Safety, Risk, Reliability and Quality	0.981	0.021
Medical Laboratory Technology	0.980	0.007	Agronomy and Crop Science	0.971	0.021
Development	0.991	0.007	Biomaterials	0.968	0.022
Environmental Science (miscellaneous)	0.993	0.007	Civil and Structural Engineering	0.960	0.022
Social Sciences (miscellaneous)	0.989	0.007	Water Science and Technology	0.963	0.022
Health Policy	0.991	0.007	Economics and Econometrics	0.980	0.022
Law	0.989	0.008	Inorganic Chemistry	0.972	0.023
Paleontology	0.987	0.008	Emergency Medicine	0.960	0.023
Public Health, Environmental and Occupational Health	0.993	0.008	Ceramics and Composites	0.970	0.023
Sociology and Political Science	0.989	0.008	Mathematical Physics	0.926	0.024
Finance	0.990	0.009	Otorhinolaryngology	0.964	0.024
Economics, Econometrics and Finance (miscellaneous)	0.979	0.009	Social Psychology	0.961	0.025
Horticulture	0.987	0.009	Human Factors and Ergonomics	0.959	0.025
History	0.979	0.009	Mechanical Engineering	0.973	0.025
Industrial and Manufacturing Engineering	0.987	0.009	Demography	0.961	0.025
Geology	0.990	0.009	Nature and Landscape Conservation	0.976	0.026
Language and Linguistics	0.987	0.010	Radiological and Ultrasound Technology	0.977	0.026
Archeology	0.964	0.010	Public Administration	0.935	0.026
Political Science and International Relations	0.983	0.010	Community and Home Care	0.957	0.026
Developmental and Educational Psychology	0.988	0.011	Control and Systems Engineering	0.982	0.027
Astronomy and Astrophysics	0.991	0.011	Organic Chemistry	0.982	0.027
Tourism, Leisure and Hospitality Management	0.970	0.011	Radiation	0.984	0.027
Rehabilitation	0.984	0.012	Human-Computer Interaction	0.965	0.027
Geometry and Topology	0.973	0.012	Management, Monitoring, Policy and Law	0.969	0.027
Linguistics and Language	0.985	0.013	Philosophy	0.940	0.027
Computational Mechanics	0.963	0.013	Statistical and Nonlinear Physics	0.923	0.028
Signal Processing	0.980	0.014	Modeling and Simulation	0.964	0.028
Earth and Planetary Sciences (miscellaneous)	0.980	0.014	Computer Vision and Pattern Recognition	0.985	0.029
Health Informatics	0.981	0.014	Geriatrics and Gerontology	0.979	0.029
Analysis	0.939	0.015	Health Information Management	0.943	0.029
Forestry	0.979	0.015	Management Science and Operations Research	0.930	0.030
Urban Studies	0.965	0.015	Automotive Engineering	0.937	0.030
Nutrition and Dietetics	0.985	0.015	Logic	0.879	0.031
Statistics and Probability	0.980	0.015	Pharmacology (medical)	0.980	0.031
Clinical Psychology	0.986	0.015	Filtration and Separation	0.961	0.031
Pediatrics, Perinatology and Child Health	0.986	0.015	Economic Geology	0.968	0.032
Accounting	0.983	0.016	Acoustics and Ultrasonics	0.967	0.032
Applied Psychology	0.968	0.016	Condensed Matter Physics	0.974	0.032
Cultural Studies	0.958	0.016	Nuclear and High Energy Physics	0.966	0.032
Fluid Flow and Transfer Processes	0.979	0.016	Marketing	0.975	0.034
Pathology and Forensic Medicine	0.988	0.017	Parasitology	0.967	0.034
Psychology (miscellaneous)	0.985	0.017	Artificial Intelligence	0.976	0.034
Anthropology	0.971	0.017	Numerical Analysis	0.750	0.035
Chemical Engineering (miscellaneous)	0.979	0.017	Geochemistry and Petrology	0.952	0.035
Business and International Management	0.987	0.018	Epidemiology	0.977	0.035
Dermatology	0.987	0.018	Gender Studies	0.917	0.035
Organizational Behavior and Human Resource Management	0.975	0.018	Discrete Mathematics and Combinatorics	0.939	0.036
Health Professions (miscellaneous)	0.972	0.018	Communication	0.905	0.037
Geotechnical Engineering and Engineering Geology	0.970	0.018	Aquatic Science	0.944	0.037
Materials Chemistry	0.984	0.018	Pollution	0.960	0.038
Surfaces and Interfaces	0.988	0.018	Oceanography	0.963	0.038
Radiology, Nuclear Medicine and Imaging	0.987	0.018	Histology	0.935	0.038
Experimental and Cognitive Psychology	0.973	0.019	Earth-Surface Processes	0.960	0.039
Geography, Planning and Development	0.976	0.019	Information Systems and Management	0.904	0.039
Environmental Engineering	0.976	0.019	Computer Graphics and Computer-Aided Design	0.956	0.040
Mechanics of Materials	0.963	0.019	Decision Sciences (miscellaneous)	0.906	0.041
Insect Science	0.963	0.020	Pharmacy	0.969	0.041
Animal Science and Zoology	0.966	0.020	Atmospheric Science	0.956	0.042
Ecology	0.987	0.020	Food Science	0.970	0.042

Tabla 1. (cont.) Coeficiente R2 y error cuadrático para las categorías con distribución de impactos en Scopus 2006.

CATEGORÍA	R ²	E ²	CATEGORÍA	R ²	E ²
Computer Science (miscellaneous)	0.975	0.042	Reproductive Medicine	0.926	0.108
Dentistry (miscellaneous)	0.951	0.042	Statistics, Probability and Uncertainty	0.881	0.111
Renewable Energy, Sustainability and the Environment	0.927	0.042	Fuel Technology	0.930	0.115
Instrumentation	0.964	0.043	Management of Technology and Innovation	0.929	0.120
Obstetrics and Gynecology	0.971	0.045	Health, Toxicology and Mutagenesis	0.911	0.122
Analytical Chemistry	0.961	0.045	Gastroenterology	0.948	0.122
Aging	0.960	0.045	Electrochemistry	0.905	0.124
Computers in Earth Sciences	0.961	0.045	Neuroscience (miscellaneous)	0.957	0.129
Ecological Modeling	0.903	0.045	Transplantation	0.891	0.133
Microbiology (medical)	0.961	0.046	Leadership and Management	0.865	0.134
Computer Science Applications	0.940	0.047	Pharmacology	0.945	0.137
Computational Mathematics	0.888	0.047	Internal Medicine	0.960	0.138
Pharmacology, Toxicology and Pharmaceutics (miscellaneous)	0.967	0.048	Cardiology and Cardiovascular Medicine	0.940	0.142
Atomic and Molecular Physics, and Optics	0.959	0.050	Business, Management and Accounting (miscellaneous)	0.898	0.147
Waste Management and Disposal	0.916	0.051	Plant Science	0.914	0.166
Soil Science	0.938	0.052	Nursing (miscellaneous)	0.873	0.168
Hardware and Architecture	0.971	0.053	Information Systems	0.923	0.169
Biotechnology	0.969	0.053	Advanced and Specialized Nursing	0.832	0.171
Ecology, Evolution, Behavior and Systematics	0.953	0.054	Space and Planetary Science	0.924	0.182
Pulmonary and Respiratory Medicine	0.975	0.055	Geophysics	0.870	0.183
Clinical Biochemistry	0.956	0.057	Spectroscopy	0.883	0.187
Critical Care Nursing	0.874	0.059	Computer Networks and Communications	0.900	0.190
Arts and Humanities (miscellaneous)	0.852	0.059	Energy (miscellaneous)	0.906	0.206
Stratigraphy	0.890	0.059	Infectious Diseases	0.941	0.208
Complementary and Alternative Medicine	0.938	0.061	Biophysics	0.916	0.209
Anesthesiology and Pain Medicine	0.964	0.062	Embryology	0.789	0.220
Electronic, Optical and Magnetic Materials	0.969	0.063	Pharmaceutical Science	0.912	0.224
Bioengineering	0.963	0.064	Immunology and Allergy	0.945	0.234
Neurology (clinical)	0.956	0.064	Genetics	0.940	0.237
Cellular and Molecular Neuroscience	0.907	0.064	Biomedical Engineering	0.872	0.249
Psychiatry and Mental Health	0.972	0.064	Microbiology	0.921	0.252
Critical Care and Intensive Care Medicine	0.967	0.064	Computational Theory and Mathematics	0.861	0.281
Surfaces, Coatings and Films	0.957	0.065	Behavioral Neuroscience	0.821	0.304
Urology	0.960	0.065	Software	0.842	0.315
Applied Microbiology and Biotechnology	0.943	0.066	Library and Information Sciences	0.780	0.371
Anatomy	0.946	0.066	Molecular Biology	0.918	0.393
Aerospace Engineering	0.925	0.066	Sensory Systems	0.838	0.414
Neurology	0.942	0.066	Materials Science (miscellaneous)	0.848	0.429
Physiology (medical)	0.908	0.070	Theoretical Computer Science	0.800	0.437
Nephrology	0.963	0.071	Molecular Medicine	0.864	0.450
Colloid and Surface Chemistry	0.905	0.072	Medicine (miscellaneous)	0.867	0.552
Psychiatric Mental Health	0.626	0.073	Management Information Systems	0.819	0.585
Immunology and Microbiology (miscellaneous)	0.925	0.073	Endocrinology, Diabetes and Metabolism	0.887	0.602
Veterinary (miscellaneous)	0.921	0.074	Cancer Research	0.857	0.653
Biological Psychiatry	0.924	0.075	Media Technology	0.722	0.667
Building and Construction	0.877	0.077	Engineering (miscellaneous)	0.766	0.680
Safety Research	0.913	0.077	Biochemistry, Genetics and Molecular Biology (miscellaneous)	0.821	0.691
Transportation	0.875	0.077	Agricultural and Biological Sciences (miscellaneous)	0.811	0.700
Ophthalmology	0.955	0.078	Drug Discovery	0.764	0.711
Neuropsychology and Physiological Psychology	0.946	0.078	Catalysis	0.822	0.793
Developmental Neuroscience	0.914	0.079	Endocrinology	0.738	0.823
Hepatology	0.977	0.080	Physical and Theoretical Chemistry	0.764	0.882
Metals and Alloys	0.931	0.081	Cell Biology	0.861	0.983
Applied Mathematics	0.876	0.086	Genetics (clinical)	0.819	0.992
Cognitive Neuroscience	0.943	0.089	Polymers and Plastics	0.660	1.100
Energy Engineering and Power Technology	0.930	0.089	Physiology	0.740	1.371
Mathematics (miscellaneous)	0.901	0.089	Electrical and Electronic Engineering	0.656	1.612
Toxicology	0.928	0.092	Process Chemistry and Technology	0.700	1.702
Virology	0.907	0.092	Chemistry (miscellaneous)	0.716	1.760
Chemical Health and Safety	0.894	0.096	Immunology	0.723	2.114
Strategy and Management	0.938	0.100	Oncology	0.690	2.460
Global and Planetary Change	0.934	0.101	Physics and Astronomy (miscellaneous)	0.586	2.697
Rheumatology	0.953	0.101	Biochemistry	0.607	2.803
Surgeny	0.938	0.102	Developmental Biology	0.624	3.799
Structural Biology	0.956	0.105	Multidisciplinary	0.622	4.325

Tras observar la tabla 1 hay que añadir que un 78% de las categorías caen dentro de un error cuadrático de menos de 0.12. Con este umbral que hemos elegido como ejemplo se demostraría que la mayoría de las categorías de Scopus en 2006 (a partir de ahora llamaremos Scopus-06 para abreviar) se acercan bastante a la aproximación logarítmica. Hay algunas excepciones como el caso de *Developmental Biology* ó *Biochemistry*, etc. Que tienen un error muy alto, esto se debe a la irregularidad de impactos que existen en sus revistas, algunas con un impacto muy alto y otras con un impacto muy bajo. Por ejemplo entre *Annual Review of Biochemistry* e *Indian Journal of Agricultural Biochemistry* que hay una gran diferencia en el total de citas que reciben. Coincide que ambas pertenecen al área de *Biochemistry*, un campo científico esencial para la formación de numerosos profesionales que trabajan en campos muy diversos, todos ellos relacionados con los sistemas biológicos a nivel molecular y es, a su vez, un área muy productiva para el desarrollo científico-técnico, con importantes aplicaciones en diferentes áreas de gran interés social. Parecido ocurría en el trabajo de Guerrero et al. (2007) (a partir de ahora llamaremos ISI-97 para abreviar), ya que un 80% de las categorías tenían un error cuadrático menor que 0.12 y era *Medicine, General & Internal*, la categoría con mayor producción científica de su área, quien obtenía uno de los errores cuadráticos más altos. Es interesante apuntar, gracias al trabajo de Olmeda-Gómez, C. et al. (2007), que la Medicina tiene una elevada actividad intelectual común con la Biología Molecular, Celular y Genética, en particular con la Bioquímica y la Inmunología. También se relacionan con las especialidades periféricas como la Ética Médica,

Rehabilitación, Andrología, Medicina Legal, Enfermería, etc. Pero tiene un bajo grado de interdisciplinariedad y es una disciplina poco colonizada por otras áreas.

Para asegurarnos que efectivamente lo que se observa es cierto, y que las distribuciones de impacto son generalizables, y no le influye que le cambiemos de base de datos ni de ventana de citación ni de parámetros temporales utilizados. Hayamos la correlación entre los indicadores utilizados tanto en el estudio de ISI-97 como en el presente estudio Scopus-06.

Tabla 2. Analogía entre los indicadores calculados.

	PEARSON
<i>Impacto Externo</i>	0.97
<i>Impacto Medio</i>	0.88
<i>E2</i>	0.99
<i>R2</i>	0.85
<i>a</i>	0.9

El coeficiente más alto, como era de esperar, recae en la comparación entre el Impacto Exterior de ISI-97 con el Impacto Exterior de Scopus-06 ($r=0.97$). También son altos los resultados de la confrontación entre el Impacto Medio de ISI-97 con el Impacto Medio de Scopus-06 ($r=0.88$), el coeficiente *a* de ISI-97 con el coeficiente *a* de Scopus-06 ($r=0.90$), el R^2 de ISI-97 con el R^2 de Scopus-06 y error cuadrático de ISI-97 con el error cuadrático de Scopus-06 ($r=0.85$).

Las distribuciones de impacto del estudio de Guerrero-Bote (2007) son generalizables. Coincide que son análogas en la base de datos de Scopus. Los resultados no son dependientes de la base de datos elegida, aunque el experimento se haya hecho con una de mayor cobertura.

Hay que decir que existen ligeras diferencias en el orden de las áreas que mejor se adaptan a la exponencial negativa en un experimento y en otro. Por ello, además del área General, con un valor inferior en Scopus, las áreas con unos valores más altos de error cuadrático son Bioquímica, Genética y Biología Molecular, Inmunología y Microbiología y Química. Y aquellas con los valores más bajos son Econometría, Economía y Finanzas, Arte y Humanidades, Profesiones de la Salud y Psicología. Del mismo modo, además de General, con un valor más alto en Scopus, las materias con los valores más bajos en el R^2 se Bioquímica, Genética y Biología Molecular, Enfermería, Química, e Inmunología y Microbiología. Y aquellos con los valores más altos se Econometría, Economía y Finanzas, Psicología, y Profesiones de la Salud. Finalmente, aparte de los generales, las áreas con mayor valor de coeficiente a son Bioquímica, Genética y Biología Molecular, Neurociencias, Inmunología y Microbiología y Química. Y aquellos con los valores más bajos son de Artes y Humanidades, Matemáticas, Ciencias Sociales, y Econometría, Economía y Finanzas.

Las distribuciones de las tasas de importación y exportación de las diferentes categorías son muy distintas, hay algunas categorías que tienen unas tasas de exportación muy altas. Como por ejemplo el caso las que pertenecen a las

Ciencias de la Vida, por el contrario había algunas que tenían las tasas de exportación e importación más bajas como es el caso de las que pertenecen a las Ciencias Sociales. Más tarde se pudo comprobar que a lo largo de la literatura se había dicho que las diferentes áreas científicas tienen distintos hábitos de citación y esto podría ser la razón de que las distribuciones salgan tan diferentes en las disciplinas.

Debido a esto vamos a analizar las diferentes áreas de Scopus, para ver sus tendencias a la hora de referenciar. En la siguiente tabla podemos ver las 27 áreas de Scopus y una serie de indicadores relativos a esta tendencia.

Tabla 3. Diferentes porcentajes de referencias por artículos, en total, en Scopus, en la ventana de citación de 3 años, y en la ventana de citación de 2 años.

Áreas	Referencias por artículo	Referencias ventana 3 años/artículo	Referencias ventana 2 años/artículo	Referencias por artículo en Scopus	Referencias ventana 3 años/artículo en Scopus	Referencias ventana 2 años/artículo en Scopus
General	21.9	6.1	4.1	13.0	5.1	3.5
Agricultural and Biological Sciences	33.6	6.3	3.9	15.3	4.8	3.1
Arts and Humanities	43.1	4.4	2.7	4.3	1.2	0.7
Biochemistry, Genetics and Molecular Biology	36.5	9.8	6.5	24.8	9.0	5.9
Business, Management and Accounting	21.1	3.6	2.2	5.9	1.5	0.9
Chemical Engineering	19.4	4.3	2.8	9.9	3.4	2.2
Chemistry	33.3	8.3	5.5	18.6	6.9	4.6
Computer Science	18.7	4.8	3.1	6.5	2.3	1.4
Decision Sciences	23.4	3.9	2.3	7.6	2.0	1.2
Earth and Planetary Sciences	32.7	6.4	4.1	13.8	4.5	2.9
Economics, Econometrics and Finance	26.4	4.8	3.0	7.8	2.0	1.2
Energy	13.8	3.1	2.0	4.9	1.7	1.1
Engineering	14.2	3.3	2.1	5.7	2.0	1.2
Environmental Science	32.9	6.6	4.1	14.7	4.6	2.9
Immunology and Microbiology	33.9	9.3	6.2	23.3	8.5	5.6
Materials Science	20.9	4.9	3.1	11.0	3.8	2.3
Mathematics	20.4	4.4	2.9	6.8	2.3	1.4
Medicine	23.1	5.9	3.8	13.9	4.9	3.2
Neuroscience	44.5	10.7	6.9	29.7	9.9	6.4
Nursing	17.5	4.7	3.1	9.3	3.3	2.1
Pharmacology, Toxicology and Pharmaceutics	31.4	8.4	5.6	20.3	7.2	4.8
Physics and Astronomy	23.9	6.0	4.1	11.1	3.7	2.3
Psychology	39.2	6.3	3.8	15.2	4.2	2.5
Social Sciences	35.1	6.4	4.1	6.0	1.8	1.1
Veterinary	23.2	4.2	2.6	10.6	3.2	1.9
Dentistry	23.8	4.6	2.8	13.3	4.0	2.4

La tabla 3 muestra valores absolutos promedio y se pueden apreciar a simple vista tres factores que explican la variabilidad de los datos: el primero de ellos sería el número medio de referencias de cada área (Referencias/Artículo mostrado en columna 1), el segundo factor sería el número medio de referencias recientes (Referencias en la Ventana de Citación 3 años/Artículo y Referencias en la Ventana de Citación 2 años/Artículo mostrado en las columnas 2 y 3) y el tercer factor es el relativo al número medio de referencias en Scopus (Referencias/Artículo en Scopus mostrado en la columna 4). Sin embargo también se muestran en las últimas columnas de la tabla 3 otros dos indicadores que aúnan los dos últimos factores.

Por ello, y teniendo en cuenta el primer factor cabe apuntar que son Neuroscience (44.5) y Arts and Humanities (43.1) las áreas que más referencias por artículo contienen. Seguida de éstas están Psychology, Biochemistry, Genetics and Molecular Biology, Social Sciences. Sin embargo las áreas que menos referencian en sus artículos son Energy, Engineering, Nursing, Computer Science, la mayoría de ellas relativas a la Scopus Category de Ciencias Físicas. La diferencia entre el mayor valor y el menor es cuatro veces más.

Cuando observamos en la tabla 3 el segundo factor, el relativo al número medio de referencia a documentos recientes (en la ventana de citación de 2 y 3 años). Las áreas que menor valor tienen son Energy, Engineering, Business, Management and Accounting, Decision Sciences, etc. Y en el lado opuesto están las áreas Pharmacology, Toxicology and Pharmaceutics, Immunology and Microbiology,

Biochemistry, Genetics and Molecular Biology, Neuroscience. La diferencia entre la de mayor valor y la de menor es tres veces más.

Sin embargo en el tercer factor, el número medio de referencias por artículos en la base de datos Scopus, el área con mayor referencias a documentos Scopus es, Neuroscience (29.7) casualmente esta área es también la que contienen mas referencias por artículo. Seguida de esta nos encontramos con Biochemistry, Genetics and Molecular Biology, Immunology and Microbiology, Pharmacology, Toxicology and Pharmaceutics, etc. Los valores más bajos de referencias por artículo en Scopus, efectivamente están en Arts and Humanities, Business, Management and Accounting, Social Sciences y algunas relacionadas con las Ciencias Físicas, Energy, Engineering, Computer Science, etc. La diferencia entre la de mayor valor y la de menor es seis veces más.

La primera columna se diferencia bastante de la segunda y la tercera con una correlación de 0.70 y de la cuarta con 0.60. Sin embargo la tercera y cuarta columna tienen una correlación bastante alta de 0.98. De hecho podemos ver como Arts and Humanities está en la primera columna en los primeros puestos, en la segunda y tercera columna pasa a una situación bastante posterior y en la cuarta columna se sitúa en la última posición. Con Social Science ocurre algo parecido.

Por último, cuando unimos los factores Referencias en la Ventana de Citación de 3 años/Artículo y Referencias en la Ventana de Citación de 2 años/Artículo y Referencias/Artículo en Scopus, teniendo en cuenta por tanto las referencias hacia artículos que están en la base de datos Scopus y que están dentro de la ventana de citación de 2 y 3 años. Cabe indicar que Arts and Humanities (1.2 y 0.7),

Business, Management and Accounting, Energy y Social Sciences, obtienen los datos más bajos. Por el contrario es Neuroscience (9.9 y 6.4), Biochemistry, Genetics and Molecular Biology, Immunology and Microbiology, y Pharmacology, Toxicology and Pharmaceutics relativas a la Scopus Category de Ciencias de la Vida quien tiene los valores más altos.

La diferencia entre la de mayor valor y la de menor es nueve veces más cuando se tiene en cuenta la ventana de citación de 3 años y seis veces más cuando se toma la ventana de citación de 2 años.

La diferencia es todavía mayor con los indicadores que aúnan los dos factores, lo cual es lógico porque aúnan las diferencias de uno y otro. Lo mismo ocurre con la diferencia entre el primero y el último.

Tabla 4. Porcentaje de referencias en Scopus (artículos publicados en 2007).

Áreas	Porcentaje de referencias en Scopus	Porcent.Ref. Ventana 3 años en Scopus	Porcent.Ref. Ventana 2 años en Scopus	Porcent.Ref. Ventana 3 años	Porcent.Ref. Ventana 2 años
Immunology and Microbiology	68.6	25.0	16.4	27.5	18.2
Biochemistry, Genetics and Molecular Biology	67.8	24.6	16.1	26.8	17.7
General	59.3	23.3	15.8	27.6	18.9
Pharmacology, Toxicology and Pharmaceutics	64.5	23.1	15.1	26.7	17.8
Neuroscience	66.7	22.3	14.4	23.9	15.6
Medicine	60.2	21.4	13.9	25.3	16.6
Chemistry	55.9	20.8	13.7	25.0	16.6
Health Professions	57.1	20.1	12.8	25.5	16.6
Nursing	53.3	18.7	11.9	26.9	17.6
Materials Science	52.4	18.0	11.1	23.6	15.0
Chemical Engineering	51.2	17.8	11.3	22.2	14.3
Dentistry	55.7	16.7	10.1	19.2	11.8
Physics and Astronomy	46.5	15.5	9.5	25.0	17.1
Agricultural and Biological Sciences	45.6	14.4	9.1	18.6	11.7
Environmental Science	44.5	14.1	8.8	20.0	12.6
Engineering	40.0	14.0	8.7	23.4	15.0
Earth and Planetary Sciences	42.1	13.7	8.8	19.4	12.5
Veterinary	45.6	13.6	8.3	18.0	11.2
Energy	35.3	12.6	7.9	22.8	14.6
Computer Science	34.6	12.2	7.6	25.4	16.4
Mathematics	33.2	11.1	7.0	21.3	14.1
Psychology	38.9	10.7	6.4	16.2	9.8
Decision Sciences	32.4	8.7	5.0	16.6	9.8
Economics, Econometrics and Finance	29.4	7.5	4.4	18.3	11.3
Business, Management and Accounting	28.0	7.2	4.3	17.3	10.5
Social Sciences	17.0	5.1	3.1	18.1	11.8
Arts and Humanities	10.0	2.9	1.7	10.2	6.3

Al observar la tabla 4, como muestran los datos relativizados con respecto al número de referencias, solamente podemos apreciar los dos últimos factores: la referenciación hacia documentos recogidos en la base de datos (Porcentaje referencias en Scopus mostrado en la primera columna) y la referenciación hacia documentos recientes (mostrado en las dos últimas columnas Porcentaje en Ventana de Citación de 3 años o Porcentaje en Ventana de Citación de 2 años).

No obstante, existen otros dos indicadores mostrados en la tabla 4, que añadan de nuevo estos dos factores.

Las áreas que menor porcentaje de referencias contienen hacia documentos incluidos en la base de datos Scopus son las relativas a las Arts and Humanities, Business, Management and Accounting, Economics, Econometrics and Finance, Decision Sciences, etc. todas ellas pertenecientes a la Scopus Category de Ciencias Sociales. En el lado opuesto y con un mayor porcentaje estarían las áreas relativas a las Scopus Category de Ciencias de la Vida y Ciencias de la Salud, como es el caso de Immunology and Microbiology, Biochemistry, Genetics and Molecular Biology, Pharmacology, Toxicology and Pharmaceutics, etc. Es importante apuntar que la diferencia entre la de mayor valor y la de menor es siete veces más.

Por otro lado, en cuanto a la referenciación de documentos actuales (en la ventana de citación de 2 y 3 años) el orden es bastante similar aunque no igual al mostrado en la primera columna (con unas correlaciones de 0.78 y 0.76 con Porcentaje en Ventana de Citación de 3 años y Porcentaje en Ventana de Citación de 2 años respectivamente).

Las áreas que menor porcentaje tienen son Arts and Humanities, Psychology, Decision Sciences, Business, Management and Accounting, etc. En el lado opuesto están las áreas relativas a las Ciencias de la Vida y la Salud como es el caso de Immunology and Microbiology, Pharmacology, Toxicology and Pharmaceutics, Biochemistry, Genetics and Molecular Biology, Nursing, etc. La diferencia entre la de mayor valor y la de menor es tres veces más.

En la parte central de la tabla 4, donde aparecen los indicadores que unifican los dos factores apuntados anteriormente, el de referenciar documentos Scopus y referenciar documentos actuales, Porcentaje en Ventana de Citación de 3 años en Scopus y Porcentaje en Ventana de Citación de 2 años en Scopus, el orden es exactamente el mismo de la primera columna, sin embargo como se aúnan los dos factores la diferencia del primero al último aumenta llegando a ser casi diez veces más el primero que el último.

Podemos apuntar acerca del porcentaje que sigue siendo Arts and Humanities con 2.9 cuando la ventana de citación es de 3 años, y 1.7 cuando la ventana de citación es de 2 años, la que obtiene los datos más bajos. Otras áreas a las que le sucede lo mismo son Social Sciences, Business, Management and Accounting, Economics, Econometrics and Finance, Decision Sciences, etc. Por el contrario quien obtiene un mayor porcentaje en este caso es Immunology and Microbiology con 25.0 cuando la ventana de citación es de 3 años y 16.4 cuando la ventana de citación es de 2 años.

En la figura 1 se muestra el porcentaje de referencias por año en artículos publicados en 2007 (hasta 1980). En la leyenda entre paréntesis se muestra el porcentaje total incluido en esta figura.

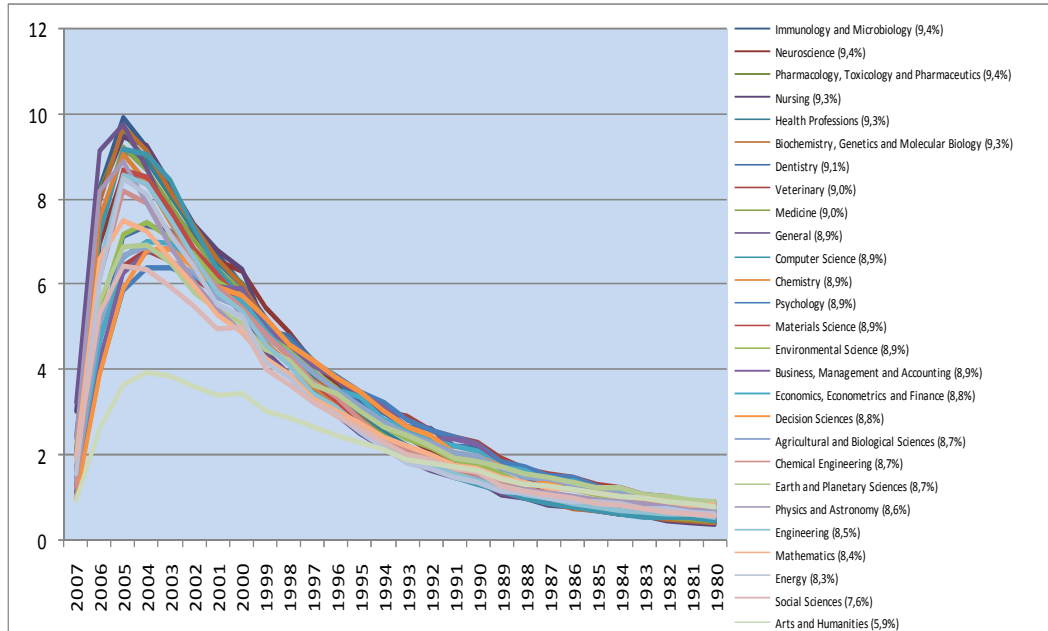


Figura 1. Porcentaje del total de referencias por años de artículos publicados en 2007 incluidos en Scopus.

Se puede observar en la figura 1 como el porcentaje de Immunology and Microbiology (94.3%) (El área que tiene mayor porcentaje) en comparación con Arts and Humanities (58.7%) (El área de menor porcentaje) es casi el doble. Arts and Humanities tiene una mayor tendencia a referenciar fuentes originales, manuales clásicos, normalmente muy antiguos, y estos documentos se salen notablemente del rango temporal que muestra la grafica. Es por eso que la línea de esta área se separa considerablemente del resto. A ella le siguen Social Sciences (76.1%), Energy (82.7%) o Mathematics (84.0%), así como algunas otras áreas enmarcadas en las Ciencias Físicas o en las Ciencias Sociales.

Sin embargo en el lado opuesto están las áreas de las Ciencias de la Vida y las Ciencias de la Salud, como es el caso de Immunology and Microbiology (94.3%), Neuroscience (93.7%), Pharmacology, Toxicology and Pharmaceutics (93.7%) se puede ver como estas áreas tiene las líneas más altas en los últimos años. Ya que su tendencia es citar artículos actuales porque la obsolescencia en estas áreas es muy alta.

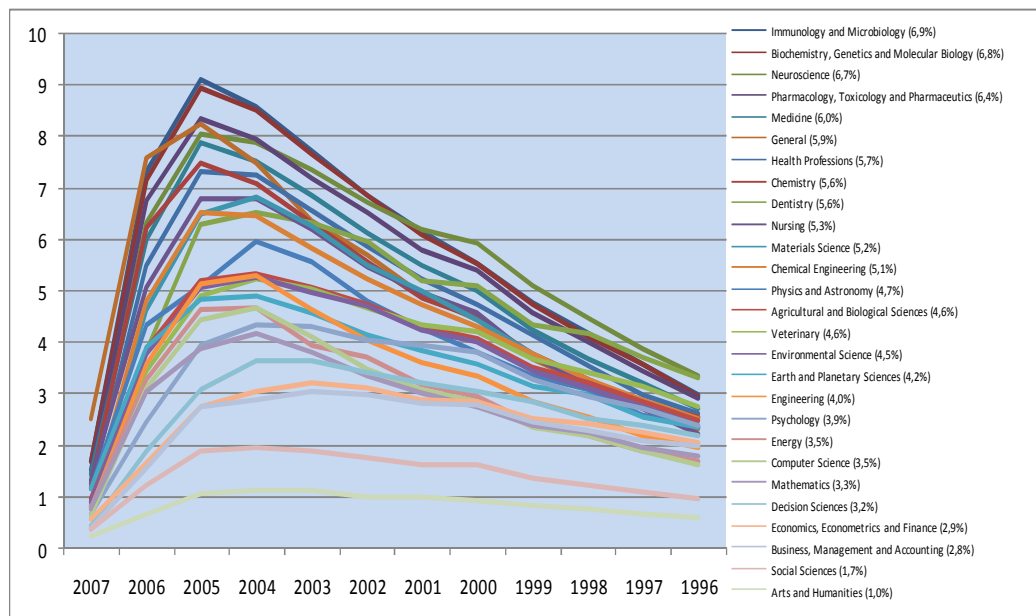


Figura 2. Porcentaje de referencias por años en Scopus de artículos publicados en 2007 incluidos en Scopus.

En la figura 2 se muestran, a diferencia de la anterior, los porcentajes de las áreas con referencias hacia documentos de la base de datos Scopus. El rango temporal es también distinto ya que se toman los años de los documentos que hay en Scopus solamente de 1996 a 2007. Podemos decir tras observar la figura que los

datos en general son más bajos en comparación con la anterior figura, esto se debe a que Scopus no cubre la totalidad de las referencias de algunas áreas y a una menor cobertura temporal. En la figura las áreas que menor porcentaje tienen son las de Arts and Humanities, Social Sciences, Business, Management and Accounting, Economics, Econometrics and Finance, etc. Los valores son tan bajos que podemos decir que casi todas las referencias de estas áreas salen fuera de la base de datos. Hay que indicar que algunas de las áreas incluidas en la Scopus Category de Ciencias Sociales en la figura 1 tenían un rank más alto que en la figura 2, éstas son Psychology, Business, Management and Accounting, Economics, Econometrics and Finance y Decision Sciences. En el lado contrario con valores altos están Immunology and Microbiology, Biochemistry, Genetics and Molecular Biology, Neuroscience, etc. Todas ellas encuadradas dentro de las Scopus Categories de Ciencias de la Vida y Ciencias de la Salud.

Cuando miramos la figura 2 nos damos cuenta que el porcentaje de referencias incluidas en Scopus en Artes y Humanidades es 6 veces menor que en Immunology and Microbiology. Y otra cosa a apreciar en esta figura es que la línea de Artes y Humanidades no se separa notablemente del resto como sucedía en la anterior figura.

Sin embargo, una cosa que nos parece oportuno apuntar es el caso de Psychology (88.8% porcentaje en la figura 1 y 38.9% porcentaje en la figura 2) que tiene un puesto alto en comparación con otras áreas pertenecientes a la Scopus Category de Ciencias Sociales. Esto puede deberse a su relación con la medicina.

Podemos decir que hay bastante coherencia entre la tabla 4 y las figuras 1 y 2, ya que en ambas se afirma que son las áreas pertenecientes a la Scopus Category de Ciencias Sociales quien obtiene datos más bajos y las áreas relativas a la Scopus Category de Ciencias de la Vida y la Salud las que tienen datos superiores.

Para el cálculo de los distintos indicadores bibliométricos lo que se utilizan son las referencias, y no todas las referencias, sino solamente las referencias con destino a trabajos incluidos en la propia base de datos. El perfil anual de referencias a documentos incluidos en el propio universo de la base de datos tendrá una gran influencia en la significación estadística de dichos indicadores en las distintas áreas científicas. Por este motivo, vamos a estudiar dicho perfil, para los documentos publicados en un año natural, pudiendo determinar la proporción de referencias que tienen como destino documentos de la propia base de datos, así como su distribución anual. De este modo también podremos determinar la influencia de la ventana de citación en los distintos indicadores.

Tabla 5. Correlación Pearson entre los indicadores bibliométricos y los diferentes porcentajes de referencias.

	SJR	FI 4AÑOS	FI 3AÑOS	FI 2AÑOS	MEDIA
Ref/art	0.436	0.448	0.455	0.458	0.449
Ref/art en Scopus	0.834	0.882	0.898	0.904	0.880
Ref. Vent. 3años/art en Scopus	0.870	0.899	0.916	0.929	0.903
Ref. Vent. 2años/art en Scopus	0.877	0.900	0.916	0.931	0.906
Ref. Vent. 3años/art	0.820	0.849	0.866	0.883	0.854
Ref. Vent. 2años/art	0.830	0.853	0.870	0.892	0.861
Total Referencias en Scopus	0.394	0.363	0.376	0.368	0.375
Total Referencias	0.277	0.263	0.274	0.261	0.269
Total Artículos	0.112	0.101	0.111	0.103	0.107
Porcentaje de Referencias en Scopus	0.703	0.724	0.739	0.756	0.731
Porcent. Vent. 3años en Scopus	0.722	0.719	0.734	0.757	0.733
Porcent. Vent. 2años en Scopus	0.737	0.727	0.740	0.765	0.742
Porcent. Vent. 3años	0.541	0.543	0.557	0.581	0.556
Porcent. Vent. 2años	0.549	0.544	0.557	0.585	0.559

En la tabla 5 se muestran las correlaciones entre los indicadores: SJR y Factores de Impacto según la ventana de citación de 2, 3 y 4 años y los elaborados con las referencias de las diferentes áreas, que se calculan teniendo en cuenta las referencias totales, las referencias a la base de datos de Scopus y las referencias recientes en la ventana de citación de 2 y 3 años.

Teniendo en cuenta que tanto el tamaño de las listas de referencia como el tamaño de las disciplinas científicas se han apuntado algunas veces como causa de las diferentes Citation Rates y por tanto de los promedios en los indicadores bibliométricos llama la atención la poca relación entre los valores de los indicadores y el número medio de referencias por artículo, el total de referencias

en un área, el total de artículos en esa área e incluso el total de referencias en Scopus. Por ejemplo, la correlación entre el número total de referencias y los indicadores es 0.269, la correlación entre el número medio de referencias por artículo con los indicadores es 0.449 y la correlación entre el total de artículos y los indicadores es 0.107. Por tanto, hay que decir que el tamaño del número de artículos y referencias no tiene influencia en los indicadores.

Igualmente llama la atención la alta correlación que presentan los promedios de referencias recientes en Scopus, lo cual después de verlo resulta lógico, puesto que para el cálculo de los indicadores se tiene en cuenta las referencias recientes en la base de datos utilizada. También tienen una alta correlación el resto de promedios de referencias, por detrás de esto se sitúan las rates de referencias recientes en Scopus.

El dato más alto estaría en la correlación entre Referencias en Ventana de Citación de 2 años en Scopus/Artículo y el Factor de Impacto de 2 años, evidentemente, como hemos apuntado anteriormente, si un área referencia documentos que están en Scopus y dentro de la ventana de citación de 2 años tendrán reciprocidad con el factor de impacto de 2 años.

Las correlaciones más importantes a resaltar serían las relativas a Referencias en Ventana de Citación de 3 años en Scopus/Artículo y Referencias en Ventana de Citación de 2 años en Scopus/Artículo con SJR y FI 2 años, FI 3 años y FI 4 años.

Ya que si un área tiende a referenciar documentos recientes y que estén en bases de datos internacionales como es Scopus tiende a ser mas citado, por lo tanto a tener mayor factor de impacto.

Pero la máxima correlación de los indicadores medios es la de Referencias en la Ventana de Citación de 2 años en Scopus/Artículo y FI 2 años, evidentemente, ya que si un área hace referencia a documentos que están en Scopus y dentro de la ventana de citación de 2 años tendrán correlación con el factor de impacto de 2 años.

Otras analogías a destacar serían las relativas al SJR con Referencias en la Ventana de Citación de 2 años en Scopus/Artículo o Referencias/Artículo en Scopus con FI 2 años o FI 3 años. Sin embargo hay que decir que la mínima correlación es la relativa a Referencias por Artículo con SJR, FI 4 años, FI 3 años y FI 2 años. Esto significa que el número medio de referencias por artículo que tienen las áreas no tienen mucha consecuencia en la media de los indicadores.

Finalmente podemos afirmar que son las áreas de Artes y Humanidades y Ciencias Sociales quien obtiene los indicadores más bajos y las Ciencias de la Vida y la Salud las que tienen datos superiores. Sin embargo cabe apuntar el caso de Psychology que tiene el puesto alto en comparación con otras áreas relacionadas con las Ciencias Sociales debido a que es una de las áreas más relacionadas con la medicina.

Desde otro punto de vista, a lo largo de la bibliografía científica se ha venido a concluir que si se colabora, especialmente con instituciones de diferentes disciplinas o de diferentes países, se obtendrá un mayor impacto, mayor tasa de citación que en caso de no colaborar. Pero claro no es lo mismo colaborar con una disciplina científica que con otra, también hay disciplinas que tienden más a colaborar entre ellas que con otras o colaborar con un país que con otro. Con lo cual los hábitos de citación tendrán influencia en la colaboración científica. Y la colaboración científica en todas las áreas y países no será la misma.

Tabla 6. Se muestran los 20 países ordenados por número total de documentos, porcentaje de documentos en colaboración, porcentaje de citas procedentes de los países colaboradores y porcentaje de referencias destinadas a los países colaboradores.

Países	Doc. 2004	Doc. 2005-2007	%Doc. Colab. 2004	% Doc. Colab. 2005-2007	% Cit. de Colaboradores	% Ref. a Colaboradores
United States	410.521	1.359.565	23.28	25.29	15.57	14.36
China	113.292	552.902	16.09	13.99	24.18	32.65
United Kingdom	107.143	362.408	37.46	38.99	30.86	30.24
Japan	110.622	347.570	19.79	20.94	35.30	34.89
Germany	98.949	326.678	39.65	41.47	30.65	29.49
France	69.591	232.091	41.64	43.33	32.51	31.07
Canada	55.929	196.960	39.27	40.78	39.06	40.04
Italy	54.112	183.950	34.81	36.64	33.39	33.67
Spain	39.776	144.564	32.88	34.89	33.29	32.92
India	33.169	128.505	18.22	18.28	33.61	33.08
Australia	35.886	127.554	38.03	38.85	36.18	35.26
Korea, Republic Of	31.332	120.173	25.00	25.45	38.68	42.53
Netherlands	30.168	104.253	45.14	45.90	35.05	35.36
Russian Federation	35.000	100.774	30.93	33.20	34.89	32.77
Brazil	21.658	88.335	27.75	25.35	36.16	35.50
Taiwan	21.071	81.275	17.49	18.53	38.25	42.70
Switzerland	22.049	75.615	55.05	56.95	35.58	35.59
Sweden	21.280	69.274	45.75	48.90	34.38	32.30
Poland	20.378	67.225	31.32	31.18	36.80	34.40
Turkey	18.170	65.901	16.75	15.39	32.61	36.60

Observando la tabla a modo de contexto, se puede ver que USA obtiene un reducido porcentaje de citas de sus colaboradores y proporciona un pequeño porcentaje de referencias a sus socios. Entendemos que esto se debe a su gran producción y por lo tanto al gran número de referencias nacionales puestas en juego. Es decir, como la producción de USA es tan grande, su aportación, la citación procedente del mismo país reduce el porcentaje del resto de países (e igualmente con respecto a las referencias). Sin embargo China, que pasa de tener el 27% de la producción de USA en 2004 al 40% en 2005-2007, a pesar de tener poco porcentaje de colaboración (que decrece en el segundo periodo), sí que destina gran porcentaje de sus referencias a países colaboradores. Turkey, que es el país que menos produce y que menos documentos en colaboración tiene, consigue un porcentaje alto de citas de colaboradores y referencia bastante a sus países colaboradores, mucho más que USA y China.

A pesar de que Switzerland tiene el mayor porcentaje de documentos en colaboración no es el que más porcentaje de citación recibe de sus países colaboradores.

Los países que mayor porcentaje de citas obtienen de sus países colaboradores y que mayor porcentaje de referencias dan a dichos países son, Canadá, Republic of Korea, Australia y Taiwan y no son países que se caractericen por su gran producción científica.

De manera general se puede decir que el promedio no nacional supera al nacional. El Citation per Collaboration Paper es bastante superior al Citation per Without Collaboration Paper y al Citation per General Paper, y esta diferencia procede principalmente de los documentos no nacionales.

Por otro lado en los indicadores relativos a los references ocurre lo mismo aunque con las diferencias más amortiguadas. El promedio de referencias no domésticas es muy superior al de referencias domésticas en todos los casos. El promedio de References per Collaboration Paper es bastante superior al de References per Without Collaboration Paper y al Reference per General Paper.

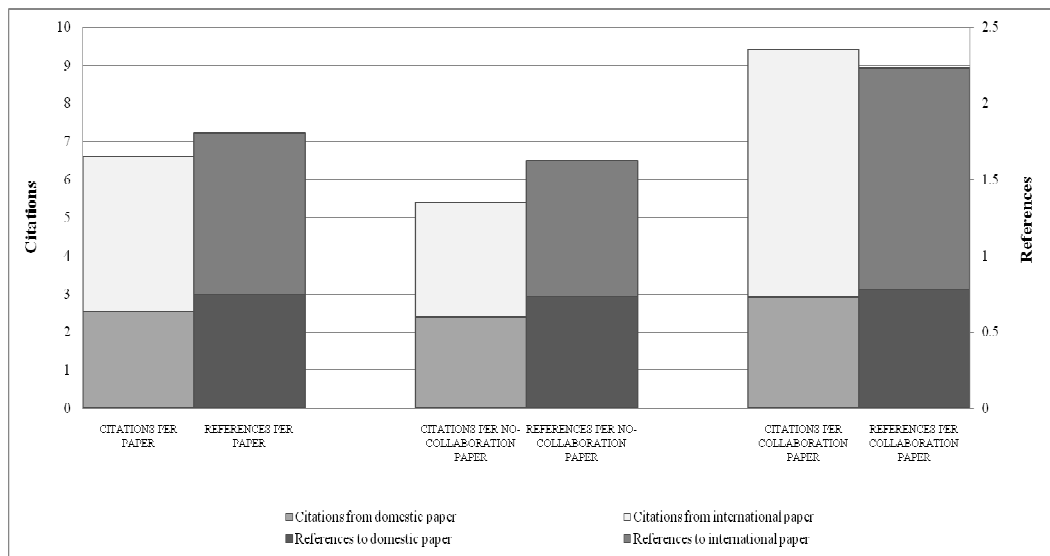


Figura 3. Comparación de promedio de citaciones (por documentos en general, por documentos sin colaboración, y por documentos con colaboración) y referencias (por documentos en general, por documentos sin colaboración, y por documentos con colaboración) de los 20 países con mayor producción en 2004, distinguiéndose en ambos casos documentos nacionales y documentos no nacionales.

El tráfico porcentual de citas con los países colaboradores disminuye a medida que aumenta la producción, lo cual es lógico simplemente en términos matemáticos, a menor producción menos que referenciar o menor el número de papers de donde captar citación por tanto mayor porcentaje supone el tráfico de citación exterior.

Estados Unidos y China que casi el 50% (China en menor medida) proviene de sus citas nacionales (recordemos que eran los países con mayor producción documental en 2004 pero con menor porcentaje de citación de sus colaboradores internacionales). Los países que menos citas nacionales reciben son Taiwan y Turkey, eran los que mayor porcentaje de citas recibían de sus socios colaboradores internacionales. Con respecto a las referencias, ocurre lo mismo.

Hay cierta correlación entre citación y referencias, pero en los países con menor citación se observa un estancamiento del número de referencias.

El promedio general que reciben los países de sus colaboradores es de 2.9 y el promedio que se recibe del resto de naciones es 3.9. Y una correlación Pearson entre ambos datos de 0.84.

Aunque hay países que se salen un poco de esto, como es el caso de EEUU y China, que reciben menos, en comparación con los demás países, de los colaboradores internacionales, y da menos a sus colaboradores. El caso de China puede ser ligeramente diferente, puesto que capta poca citación, mientras sí que tiene un mayor número de referencias no nacionales.

Por otro lado, el promedio de referencias que dan los países a los países colaboradores es de 0.71 y el promedio que se da al resto de naciones es 0.91. Y una correlación Pearson entre ambos datos de 0.71.

Sweden, Switzerland, Netherlands son los tres países de mayor Citation per Collaboration Paper, y son los que mayor influencia nacional tienen estos son que también son los de mayor influencia not domestic, probablemente por la alta proporción de trabajos con colaboraciones simultáneas nacionales e internacionales.

En el lado opuesto están los países con mayor producción China y EEUU, casualmente también son países que resultaron tener una gran citación nacional por documentos en colaboración. Es decir, esta gran citación nacional se debe a la gran producción nacional y al gran número de referencias que aporta esta producción, porque si tenemos en cuenta dicho volumen resultan ser los que menor sesgo hacia su producción tienen. Existe una correlación negativa entre la producción y la Influence on domestic paper per paper of the country de -0.6, mientras que la correlación entre las influencias domésticas y no domésticas es inferior a 0.38.

Desglosando estas influencias por documentos con o sin colaboración internacional, se puede decir que, el promedio de influencia nacional es el más alto, esto quiere decir que por norma general todos los países tienden a influir más en su producción. Los trabajos en colaboración también tienen mayor influencia a nivel general (tanto doméstica como no doméstica) y crece más la no doméstica porcentualmente, por lo que en ellos es menor la diferencia entre la influencia doméstica y no doméstica.

En general podemos concluir que la ciencia no tiene fronteras. Ciertamente hay una mayor influencia en el entorno cercano de los autores, que no tiene por qué coincidir con los entornos nacionales, que se diluye a medida que se amplía dicho entorno, mientras que, en un entorno pequeño se maximiza el sesgo producido por la autocitación. Los mayores sesgos hacia lo doméstico se producen en países pequeños y en desarrollo.

Existen evidencias de que los trabajos con colaboración obtienen una mayor citación (Lewinson & Cunningham, 1991; Narin, Stevens, & Whitlow, 1991; Glänzel, 2001; Glänzel, 2002; Leimu & Koricheva, 2005; Katz & Hicks, 1997; Persson, Glanzel, & Danell, 2004; Hsu & Huang, 2010; Aksnes, 2003; Moya-Anegón et al., 2008; Chinchilla, Vargas, Hassan, González, & Moya, 2010).

Este incremento en la citación puede depender del tipo de colaboración, por ejemplo de la colaboración científica internacional, entre instituciones de diferentes países, es de la que más incremento de citación se recibe (Narin, Stevens, & Whitlow, 1991, Katz & Hicks 1997, Goldfinch et al.2003). Aparte de esto, el incremento no solo proviene del país con el que colabora sino también del resto de países, porque gracias a esto se aumenta la red de difusión (Lancho-Barrantes, Guerrero-Bote, Chinchilla & Moya-Anegón, 2011).

Por otro lado de unas disciplinas científicas a otras puede haber también diferencias, ya que hay disciplinas en las que el porcentaje de colaboración es mayor que en otras. (Garfield, 1979; Bridgstock, 1991; Bandyopadhyay, 2001; Molteni & Zulueta, 2002; Bandyopadhyay, 2004; Sudhier, 2007; Abt, 2007; Ardanuy et al., 2009; Franceschet & Costantini, 2010; Ardanuy, 2011; Gazni & Didegah, 2011)

Tabla 7. Los 9 países con mayor producción en 2004 ordenados por número total de documentos, con el correspondiente porcentaje de documentos en colaboración, porcentaje de citas de los países colaboradores, porcentaje de referencias a los países colaboradores, Impacto total (JIF), Impacto nacional (NACJIF), Reference per Paper (RpP) y National Reference per paper (NacRpP).

Country	Papers 2004	% Collab. Papers 2004	% Cit. from Colaborators	% Ref. to colaboradores	JIF	NACJIF	RpP	NacRpP
United States	410521	23.28	15.57	14.36	8.37	4.49	2.08	1.36
China	113292	16.09	24.18	32.65	2.77	1.52	0.99	0.31
Japan	110622	19.79	35.30	34.89	7.70	2.11	1.95	0.62
United Kingdom	107143	37.46	30.86	30.24	5.23	1.80	1.44	0.57
Germany	98949	39.65	30.65	29.49	7.23	2.18	2.10	0.66
France	69591	41.64	32.51	31.07	6.75	1.74	1.96	0.52
Canada	55929	39.27	39.06	40.04	7.48	1.67	2.01	0.48
Italy	54112	34.81	33.39	33.67	6.72	1.77	2.06	0.52
Spain	39776	32.88	33.29	32.92	6.06	1.71	2.01	0.47

En la Tabla 7 se observa que USA es el país con mayor producción en 2004 pero obtiene sólo un pequeño porcentaje de citas de sus colaboradores y a su vez les proporciona sólo un pequeño porcentaje de referencias. Entendemos que esto es debido a su gran producción, y por lo tanto al gran número de referencias nacionales involucradas. Es decir, la producción de USA es tan grande por tanto sus citas nacionales reducen el porcentaje de otros países (y de manera análoga con respecto a las referencias) (Lancho et al., 2011).

Como podemos ver es el país que posee mayor Impacto Nacional (NacJIF) y el que posee mayor NacRpP.

Seguida de la producción de USA encontramos a China, pero es el país que menor porcentaje de documentos en colaboración tiene, también tiene un porcentaje pequeño de citas de sus colaboradores pero no es un país que destine poco porcentaje de referencias a sus colaboradores sin embargo es el país con menor Impacto y menor RpP.

En el último lugar del ranking encontramos a España que es el país que menos documentos tiene en 2004 sin embargo el porcentaje de documentos en colaboración es mayor que otros con mayor producción. Este país recibe un porcentaje elevado de citas de sus colaboraciones y responde de la misma manera con sus referencias a sus colaboradores. A pesar de esto es uno de los países con menor impacto total y con menor RpP.

Cuando se colabora se obtiene un Citation Rate Increment of the Colaborator. En todos los casos se recibe mayor incremento de citas de un país con el que se colabora que en el caso de no colaborar.

Pero algunos países colaboradores tienen mayor Citation Rate Increment Average when Colaborate (CRIAC) (hacia los países colaborares) que otros.

Los países China (3.96) y Japón (4.21) son los países colaboradores que menor porcentaje de documentos en colaboración tienen en 2004 al igual que de 2005 a 2007 pero sin embargo son dos de los países de mayor Citation Rate Increment Average when Colaborate (CRIAC) (su ratio de referenciación hacia un país aumenta en torno a 4 veces (de promedio) al colaborar).

Estados Unidos supera a China y Japón en porcentaje de documentos en colaboración. Sin embargo es el país con menor Citation Rate Increment Average when Colaborate (CRIAC) (su ratio de referenciación hacia un país aumenta en torno a 1 vez (de promedio) al colaborar). Por tanto es el país en el que la ratio de referenciación al colaborar y al no colaborar está más igualada. Recordemos que el porcentaje de referencias que destinaba a sus colaboradores era muy pequeño (14%). Un factor que puede influir en este menor CRIAC es su gran tamaño, a mayor producción el entorno está más diluido. Otro factor es que es el país con el impacto total más alto, aunque la mayor parte de este impacto viene de los demás países. Esto hace que la ratio que supone su referenciación sea pequeña y también su variación.

El impacto de los 9 países A, tiene una correlación negativa con su Citation Rate Increment Average when Colaborate (CRIAC) de -0,64, y con la desviación típica de estos incrementos de -0,71. Es decir a mayor impacto, menor incremento al colaborar y menos depende del país colaborador dicho incremento.

También hay países colaboradores que reciben mayor Citation Rate Increment obtained from its collaborators (CRIOC) que otros. Existe una correlación negativa de -0,51 entre el Citation Rate Increment obtained from its collaborators (CRIOC) y su impacto.

Casualmente Estados Unidos, China y Japón, los países con mayor producción en 2004, son los países que mayor incremento de citación obtienen de los países A, pero no de forma equitativa de todos los países con los que colaboran. Estados Unidos recibe mucho de aquellos con menor impacto (de hecho su CRIOC con el impacto tiene una correlación negativa de -0,71). Mientras que la cosa varía en el caso de Japón y de China que hasta hacerse dicha correlación ligeramente positiva en el caso de China. Estos reciben mayor incremento con aquellos en los que existe una mayor barrera lingüística, Italia, España y Francia.

Italia y España (con un menor impacto) tienen menor CRIOC. La mayoría de éste se debe a países como Canadá, Japón o China.

Sin embargo la Domestic Citation Rate es menor cuando se colabora.

Pero algunos países como Estados Unidos, Reino Unido, Alemania (países con alto impacto) se acercan más a la igualdad. De hecho, existe una correlación de 0,92 entre el Domestic Citation Rate Comparison Average (DCRCA) y el impacto de cada país.

Alemania tiene una Domestic Citation Rate superior cuando colaboran con ciertos países que de forma global. Concretamente cuando colabora con Rusia, Ukraine, Croatia, Bulgaria, casualmente países de bajo impacto total y de su círculo de países del Este.

En ese sentido existe una correlación negativa de -0,55 del impacto con su Domestic Citation Rate of Collaborators Comparison Average (DCRCCA). Es decir cuanto menor es su impacto más se aproxima la Domestic Citation Rate obtenida de sus colaboradores a la global de los mismos.

Pero sin embargo, el Domestic Impact Rate es mayor cuando se colabora. Por tanto si la DCR es menor cuando se colabora y el DIR es mayor, lo que ocurre es que se ganan más citas también del resto de naciones por eso la DCR disminuye al colaborar.

Pero, algunos países colaboradores tienen mayor aumento de DIR al colaborar que otros. Cuanto mayor impacto total tienen los países, tiende a ser menor el DIRIA (la diferencia entre el impacto doméstico y el dado a las colaboraciones) (correlación -0,31), y tiende a ser mayor el DIRCIA (la diferencia que los demás hacen de tus colaboraciones respecto a su producción) (correlación 0,50).

Los países que más se alejan de la igualdad, entre el impacto de lo doméstico y el dado a las colaboraciones, es decir, tienen mayor DIRIA, son Italia, España, China (que aun siendo el país de menor impacto no es el que mayor impacto comparativo da) y Francia.

Los países que mayor DIRCIA (la diferencia entre el impacto domestico y el que se recibe de las colaboraciones) tienen son España, Canadá, Japón y Alemania. China es el país que menor DIRCIA tiene y es el país que menor impacto posee.

Para comprobar si los países se comportan de manera similar en todas las disciplinas, o si por el contrario hay diferencias entre ellas, se eligen cuatro: Medicina, Ciencias Sociales, Ingenierías y Físicas.

Hay que decir que las áreas de Ciencias Sociales y de Ingenierías son las que menor impacto total tienen de todas 2.05 y 2.10 respectivamente, mientras que Medicina posee un impacto total de 5.43 y Físicas de 5.86.

En las cuatro disciplinas científicas escogidas hay que decir que en todas se obtiene un Citation Rate Increment of the Colaborator (CRIC), como sucedía en el caso general.

Sin embargo algunos países tienen un mayor Citation Rate Increment Average when Colaborate (CRIAC) (hacia los países colaboradores) en algunas disciplinas que en el caso general. En Ciencias Sociales, algunos países tienen un incremento del triple que en el caso general. En las áreas de Medicina y Físicas los países se comportan de manera similar al estudio general.

Por otro lado en el caso de Estados Unidos no existen diferencias de comportamiento entre el estudio general y las diferentes disciplinas ya que en ambos casos su ratio de referenciación hacia un país aumenta en torno a 2 veces (de promedio) al colaborar.

Como sucedía en el estudio general las correlaciones, entre Citation Rate Increment Average when Colaborate (CRIAC) frente al impacto, vuelven a salir negativas, o positivas pero poco significativas en las cuatro disciplinas.

También en las diferentes disciplinas científicas hay países colaboradores que reciben mayor Citation Rate Increment obtained from its collaborators (CRIOC) que otros.

En Ciencias Sociales reciben un incremento de citación de los colaboradores de casi el triple que en los estudios generales. Las áreas de Medicina y Físicas son las áreas con menos Citation Rate Increment Obtained from its Collaborators (CRIOC), incluso menos que en los estudios generales. Excepto China que en el campo de las Físicas supera en porcentaje al estudio general.

Existen correlaciones negativas entre Citation Rate Increment Obtained from its Collaborators (CRIOC) y su impacto en las áreas de Ingenierías (-0,88) y Físicas (-0,74), cosa que ocurría en el estudio general (-0,51), e insignificantes en Medicina y Ciencias Sociales.

Por otro lado, en las cuatro disciplinas científicas lo que ocurre es que la Domestic Citation Rate (DCR) es menor cuando se colabora.

Por ejemplo en las Ciencias Sociales es China quien supera a Estados Unidos aproximándose de este modo más a la igualdad, recordemos que China tenía un impacto bajo en todas las disciplinas científicas, especialmente en Ciencias Sociales e Ingenierías. Parece que los países que menos impacto tienen son los

países que menos se aproximan a la igualdad. Canadá es una excepción en Ciencias Sociales, donde es el país que posee el DCRCCA más bajo, incluso más bajo que el que poseía en el estudio general, un país con un impacto alto en esta disciplina.

En las cuatro disciplinas científicas hay que decir que el Domestic Impact es mayor cuando se colabora. Sin embargo hay algunas disciplinas científicas como Medicina y Físicas donde los países poseen los valores más altos de DIRIA exceptuando el caso de USA. El país que coincide en todas las disciplinas científicas por tener un mayor DIRCIA alejándose de este modo de la unidad y además coincide que es el que más se aleja de la unidad en el estudio general es España.

5 Conclusiones

Las principales conclusiones obtenidas en el estudio son:

1. Las distribuciones de impacto de Guerrero et al. (2007) son generalizables. Coincide que son análogas en la base de datos de Scopus. Gracias al E^2 de las distribuciones se puede afirmar que en Scopus la mayoría de las disciplinas se acercan bastante a una exponencial. Con lo cual más de 250 disciplinas son exportadoras de conocimiento. Este indicador ha sido de gran utilidad para detectar las categorías más exportadoras cuya distribución tiene un pico más pronunciado que la función exponencial.
 - Los resultados no son dependientes de la base de datos elegida, aunque el experimento se haya hecho con una de gran cobertura.
 - Una ventana de citación más amplia y una fecha más reciente no tienen influencia en las distribuciones de impacto ni de exportación/importación.
 - La propia estructura de categorías y la forma de adscribir las revistas que tienen las bases de datos no influyen en las distribuciones.
 - Hay que decir que existen ligeras diferencias en el orden de las áreas que mejor se adaptan a la exponencial negativa en las bases de datos. Por ello, además del área General, con un valor inferior en Scopus-06, las áreas con

unos valores más altos de error cuadrático son Bioquímica, Genética y Biología Molecular, Inmunología y Microbiología y Química. Y aquellas con los valores más bajos son Econometría, Economía y Finanzas, Arte y Humanidades, Profesiones de la Salud y Psicología. Del mismo modo, además de General, con un valor más alto en Scopus-06, las materias con los valores más bajos en el R2 se Bioquímica, Genética y Biología Molecular, Enfermería, Química, e Inmunología y Microbiología. Y aquellos con los valores más altos se Econometría, Economía y Finanzas, Psicología, y Profesiones de la Salud. Finalmente, aparte de los generales, las áreas con mayor valor de coeficiente a son Bioquímica, Genética y Biología Molecular, Neurociencias, Inmunología y Microbiología y Química. Y aquellos con los valores más bajos son de Artes y Humanidades, Matemáticas, Ciencias Sociales, y Econometría, Economía y Finanzas.

2. Las distribuciones de exportación e importación de ISI-1997 son universales. Ya que concuerdan las distribuciones de exportación de ISI con las distribuciones de la base de datos de Scopus.

- Al no depender estas distribuciones de la base de datos empleada, los resultados de las tasas de exportación e importación van a salir muy similares.
- Las tasas ordenadas por disciplinas en un experimento y en otro son similares. En ambas bases de datos, coincide que las categorías más exportadoras e importadoras pertenecen al área de Ciencias de la Vida.

- La ventana temporal utilizada no influye en las distribuciones de exportación.
 - La propia estructura de categorías y la forma de adscribir las revistas no influyen en las distribuciones de exportación e importación.
 - Los perfiles de citación de las distintas categorías afectan sobre todo a la significación estadística tanto de las distribuciones de importación/exportación como a las de impactos.
3. Se confirma la existencia de distintos hábitos de citación en las distintas disciplinas científicas.
- Son las disciplinas de Ciencias Sociales o Artes y Humanidades, las que tienen tendencia a citar documentos que no recogen las bases de datos internacionales.
 - Pese a esto, Psychology tiene un puesto alto en comparación con otras áreas pertenecientes a la gran área de las Ciencias Sociales. Esto puede deberse a su relación con la medicina. En el lado opuesto y con un mayor porcentaje estarían las áreas que pertenecen a la gran área de Ciencias de la Vida y la Salud. La diferencia entre el área de mayor valor y la de menor es siete veces más. Por otro lado, en cuanto a la referenciación de documentos actuales (en la ventana de citación de 2 y 3 años). Las áreas que menor porcentaje tienen son las relativas a la gran área de Ciencias Sociales. Por el contrario, las áreas con porcentajes más altos son las relativas a las Ciencias de la Vida y la Salud.

La diferencia entre la de mayor valor y la de menor es tres veces más. Cuando se añan los factores de referenciar documentos Scopus y referenciar documentos actuales la diferencia del primer valor al último aumenta, llegando a ser casi diez veces más el primero que el último.

- Las áreas que mas documentos recientes citan son las relativas a las Ciencias de la Vida y las Ciencias de la Salud, como es el caso de Immunology and Microbiology, Neuroscience, Pharmacology, Toxicology and Pharmaceutics ya que su tendencia es citar artículos actuales porque la obsolescencia en estas áreas es muy alta. En el lado opuesto están Arts and Humanities tiene una mayor tendencia a referenciar fuentes originales, manuales clásicos, normalmente muy antiguos, a ella le siguen, Social Sciences, Energy o Mathematics, así como algunas otras áreas enmarcadas en las Ciencias Físicas o en las Ciencias Sociales.
- Las áreas con menor porcentaje de referencias hacia documentos de la base de datos Scopus son Arts and Humanities, Social Sciences, Business, Management and Accounting, Economics, Econometrics and Finance. En el lado contrario con valores altos están Immunology and Microbiology, Biochemistry, Genetics and Molecular Biology, Neuroscience.

4. Los perfiles de citación de las distintas categorías afectan sobre todo a la significación estadística tanto de las distribuciones de importación/exportación como a las de impactos.
 - El perfil anual de referencias a documentos incluidos en el propio universo de la base de datos y los hábitos de citación tendrán una gran influencia en la significación estadística de los indicadores bibliométricos basados en el impacto en las distintas áreas científicas.
 - Como consecuencia de los hábitos de citación en algunas disciplinas los datos varían ya que si un área tiende a referenciar documentos recientes y que estén en bases de datos internacionales como es Scopus tiende a ser más citado y por tanto a tener un mayor JIF.
 - Del mismo modo, estos dos factores también influirán en los valores de los indicadores bibliométricos de las disciplinas, basados en el PageRank que ponderan las citaciones.
 - Existe poca correlación entre el número medio de referencias por artículo, el total de referencias en un área, el total de artículos de esa área e incluso el total de referencias en Scopus con los valores de los indicadores SJR y Factor de Impacto (2, 3 y 4 años). Pese a que fueron apuntadas como una de las causas de las diferentes Citation Rates y por tanto de los promedios en los indicadores bibliométricos. Que haya bastante correlación entre las Referencias a Scopus en la ventana de citación de 2 y 3 años con los

indicadores SJR y JIF, es lógico, ya que dichos indicadores se forman con los datos de Scopus de los últimos años. La correlación más significativa sería la de Referencias a Scopus en la ventana de citación de 2 y el JIF2, como hemos apuntado anteriormente, si un área referencia documentos que están en Scopus y dentro de la ventana de citación de 2 años tendrán reciprocidad con el JIF de 2 años.

- Tanto los valores promedios de los indicadores bibliométricos, SJR y Factor de Impacto, su distribución, como la significación estadística de los mismos va a depender, no sólo de la cobertura de la base de datos si no principalmente de los hábitos de citación, de la citación a documentos recientes y presentes en la base de datos.
 - Con el fin de incrementar la capacidad para la evaluación global de la ciencia de las bases bibliográficas sus estrategias de aumento de la cobertura deben de ir en la dirección de igualar el número medio de referencias recientes hacia las revistas fuentes de la base de datos para todas las disciplinas.
5. La colaboración científica, en especial la colaboración científica internacional conduce a grandes niveles de impacto.
- Existe una correlación negativa significativa entre la producción y el porcentaje de tráfico de citas a/desde los países colaboradores, un tanto esperable matemáticamente.

- De manera general, el promedio de impacto no doméstico supera al doméstico. Así como, el promedio de impacto de los documentos en colaboración supera al de los sin colaboración (en más de un 70%).
- De este impacto no doméstico de los trabajos en colaboración, en torno al 40% viene de los países colaboradores. Existiendo una gran correlación entre la citación de los colaboradores y la del resto de países. El impacto no doméstico promedio de los trabajos en colaboración es superior al de los trabajos sin colaboración, incluso sin contabilizar el procedente de los países colaboradores.
- La influencia en la producción doméstica es superior a la influencia en la producción no doméstica. Como era de esperar, los trabajos en colaboración tienen mayor influencia y hay menor diferencia entre la influencia doméstica y no doméstica.
- Algunos países que tienen gran influencia sobre los documentos no domésticos pero influyen poco en los domésticos, es el caso de Estados Unidos, Reino Unido, Alemania o Japón. En el lado opuesto podemos decir que hay países con una gran influencia doméstica y baja influencia no doméstica como puede ser el caso de Turquía y Polonia o también Brasil.

- En general podemos concluir que la ciencia no tiene fronteras. Ciertamente hay una mayor influencia en el entorno cercano de los autores, que no tiene por qué coincidir con los entornos nacionales, que se diluye a medida que se amplía dicho entorno, mientras que, en un entorno pequeño se maximiza el sesgo producido por la autocitación. Los mayores sesgos hacia lo doméstico se producen en países pequeños y en desarrollo.
6. Cuando se colabora con un país se obtiene más incremento de citación de ese país que en caso de no colaborar con él.
- Cuando se colabora se obtiene un Citation Rate Increment of the Colaborator (CRIC). En todos los casos se recibe mayor incremento de citas de un país con el que se colabora que en el caso de no colaborar. Pero algunos países colaboradores tienen mayor Citation Rate Increment Average when Colaborate (CRIAC) (referencias hacia los países colaboradores) que otros y también hay países colaboradores que reciben mayor Citation Rate Increment obtained from its collaborators (CRIOC) (citas obtenidas de los países colaboradores) que otros. Parece que hay una tendencia a que en los países con mayor impacto el incremento al colaborar sea menor.
 - Por otro lado la Domestic Citation Rate es menor cuando se colabora, aunque algunos países se acercan bastante a la igualdad en la Domestic Citation Rate Comparison Average (DCRCA) y la Domestic Citation Rate of Collaborators Comparison Average (DCRCCA). Normalmente países con alto impacto son

los que más se aproximan a la igualdad en DCRCA. Mientras que la tendencia de la DCRCCA es a ser mayor en los países con menor impacto.

- Por último, el Domestic Impact Rate es mayor cuando se colabora, aunque cuanto mayor es el impacto menor es la diferencia entre el Domestic Impact global y el dado a las colaboraciones (DIRIA) y mayor es la diferencia que hacen los demás (DIRCIA) con respecto al Domestic Impact global. Por tanto si la DCR es menor cuando se colabora y el DIR es mayor, lo que ocurre se gana mayor incremento de citas también del resto de naciones por eso la DCR disminuye al colaborar.
- En las cuatro disciplinas científicas escogidas hay que decir que en todas se obtiene un Citation Rate Increment of the Colaborator (CRIC). Sin embargo en Ciencias Sociales, algunos países tienen una Citation Rate Increment Average when Colaborate (CRIAC) del triple que en el caso general y en Ingenierías de casi el doble. También en Ciencias Sociales e Ingenierías son las disciplinas científicas donde los países colaboradores reciben mayor Citation Rate Increment obtained from its collaborators (CRIOC) y las áreas de Medicina y Físicas son las que menos CRIOC tienen, incluso menos que en los estudios generales. Parece que hay una tendencia en estos países a que sea menor el incremento al colaborar en las disciplinas de mayor impacto.

- Por otro lado, en las cuatro disciplinas científicas ocurre que la Domestic Citation Rate (DCR) es menor cuando se colabora. Sin embargo se puede decir que algunos países se acercan más a la igualdad en algunas disciplinas que en otras. Por ejemplo en las Ciencias Sociales es China quien supera a Estados Unidos aproximándose de este modo más a la igualdad, recordemos que China tenía un impacto bajo en todas las disciplinas científicas, especialmente en Ciencias Sociales e Ingenierías. Parece que los países que menos impacto tienen son los países que menos se aproximan a la igualdad. Canadá es una excepción en Ciencias Sociales, donde es el país que posee el DCRCCA más bajo, incluso más bajo que el que poseía en el estudio general, un país con un impacto alto en esta disciplina.
- Por último, en las cuatro disciplinas científicas hay que decir que el Domestic Impact es mayor cuando se colabora. Sin embargo hay algunas disciplinas científicas como Medicina y Físicas donde los países poseen los valores más altos de DIRIA exceptuando el caso de USA. El país que coincide en todas las disciplinas científicas por tener un mayor DIRCIA alejándose de este modo de la unidad y además coincide que es el que más se aleja de la unidad en el estudio general es España.

6 Perspectivas de investigación futuras

Finalmente, en esta sección, incluimos las perspectivas de investigación futuras más importantes con respecto a los temas de investigación planteados.

De nuestro último artículo surgieron una serie de resultados, que permitió afirmar que cuando los países colaboraban entre si recibían más incremento de citación que en caso de no colaborar. Este incremento variaba de unas naciones a otras. Por otro lado en el mismo trabajo se incluyó un análisis de cómo resultaba este incremento de citación de los países colaboradores en las diferentes disciplinas científicas, ya que a lo largo de la literatura siempre se había dicho que había disciplinas en las que el porcentaje de colaboración era mayor en unas que en otras. Las cuatro disciplinas elegidas fueron (Medicina, Ciencias Sociales, Ingenierías y Físicas) efectivamente los resultados de este artículo corroboraron que el incremento que los países colaboradores proporcionaban era diferente entre áreas.

Por todo ello surge la idea de realizar un estudio más completo y general donde puedan tomar participación todas las disciplinas de la Ciencia, de este modo se trataría de averiguar si hay países que tienen preferencia por unas determinadas áreas a la hora de colaborar entre ellos y si estas áreas guardan algún tipo relación entre sí.

Igualmente nos gustaría analizar si existe la tendencia de colaboración de los países en vías de desarrollo en unas áreas concretas con los países desarrollados. Y como es el incremento resultante de estas colaboraciones.

7 Referencias

Abt, H. (2007). The frequencies of multinational papers in various sciences. *Scientometrics*, 72 (1), 105–115.

Aksnes, D. (2003). Characteristics of highly cited papers. *Research Evaluation*, 12 (3), 159–170.

Alcaín, M., & Gálvez, M. (1998). Evolución de las revistas españolas de psicología. *Papeles del Psicólogo*, 70, 35–42.

Ardanuy, J. (2012). Scientific collaboration in Library and Information Science viewed through the Web of Knowledge: the Spanish case. *Scientometrics*, 90 (3), 877-890.

Ardanuy, J., Urbano, C., & Quintana, L. (2009). A citation analysis of Catalan literary studies (1974-2003): Towards a bibliometrics of humanistic studies in minority languages. *Scientometrics*, 81(2), 347–366.

Arencibia Jorge, R., & Moya Anegón, F. (2008). La evaluación de la investigación científica: una aproximación teórica desde la Cienciometría. In *ACIMED*. Editorial de Ciencias Médicas ECIMED.

Bandyopadhyay, A. K. (2004). Authorship collaboration in physics philosophy and political science, IASLIC National Seminar, 11th, Kolkata, 2004, 405-409.

Bandyopadhyay, A. K. (2001). Authorship pattern in different disciplines. *Annals of Library and Information Studies*, 48(4), 139-147.

Bergstrom, C. (2007). Eigenfactor: Measuring the value and prestige of academic journals. *C&RL New*, 68(5).

Borrego, Ángel & Urbano, Cristóbal (2006). La evaluación de revistas científicas en Ciencias Sociales y Humanidades. *Información, Cultura y Sociedad*, 14, p.11-27.

Braun, T; Glänzel, W & Schubert, A. (2000). "How balanced is the Science Citation Index's journal coverage? A preliminary overview of macrolevel statistical data". En: Cronin, B; Barsky Atkins, H (eds.). *The Web of knowledge, a festschrift in honor of Eugene Garfield*. Canada: American Society of Information Science. pp. 251-277.

Braun, T., Glänzel, W., & Schubert, A. (1985). *Scientometric indicators: a 32 country comparative evaluation of publishing performance and citation impact*: World Scientific.

Bridgstock, M. (1991). The quality of single and multiple authored papers— An unresolved problem. *Scientometrics*, 21, 37–48.

Broadus, R. N. (1971). The literature of the social sciences: a survey of citation studies. *International Social Sciences Journal*, 23: 236–243

Brooks, T.A. (1985). Private acts and public objects: an investigation of citer motivations. *Journal of the American Society for Information Science*, 36(4):223-229.

Camí, J. (1997). Impactolatría: diagnóstico y tratamiento. *Medicina Clínica*, 109 (13), 515- 524

Cañibano C., Otamend J., & Andujar, I. (2008). Measuring and assessing researcher mobility from CV analysis: the case of the Ramón y Cajal Programme in Spain. *Research Evaluation*. 17(1), 17-31.

Carpintero, H., & Peiró, J. (1983). The significance of the bibliometric methodology to the studies of the history of psychology. *Revista de Historia de la Psicología*, 4(1), 21–32.

Cawkell AE. (1978). Evaluating scientific journals with Journal Citation Reports — a case study in acoustics. *Journal of the American Society for Information Science*, 29, 41-46.

Chinchilla-Rodríguez, Z., Vargas-Quesada, B., Hassan-Montero, Y., González-Molina, A., & Moya Anegón, F. (2010). New approach to the visualization of international scientific collaboration. *Information Visualization*, 9(4), 277–287.

Clemens, E. S., Powell, W. W., Mcilwaine, K., Okamoto, D. (1995). Careers in print: Books, journals, and scholarly reputations. *American Journal of Sociology*, 101: 433–494.

Cronin, B., Snyder, H., Atkins, H. (1997). Comparative citation rankings of authors in monographic and journal literature: a study of sociology. *Journal of Documentation*, 53: 263–273.

Deis, L.F. & Goodman, D. (2005). Web of Science (2004 version) and Scopus. The Charleston Advisor, 6. Text available at: <http://www.charlestonco.com/comp.cfm?id=43>

Dimitri, Pedro Jorge (2004). Los hábitos de citación de los investigadores que publicaron para el Instituto Nacional de la Administración Pública.

Ding, Y., Foo, S., & Chowdhury, G. (1999). A bibliometric analysis of collaboration in the field of information retrieval. *The International Information & Library Review*, 30 (4), 367-376.

Egghe, L. (2009). Mathematical derivation of the impact factor distribution. *Journal of Informetrics*, 3(4), 290–295.

Figg, W. D., Dunn, L., Liewehr, D. J., Steinberg, S. M., Thurman, P. W., Barrett, J. C., & Birkinshaw, J. (2006). Scientific collaboration results in higher citation rates of published articles. *Pharmacotherapy*, 26 (6), 759–767.

Filippo, Daniela de & Fernández, María Teresa (2002). Bibliometría: importancia de los indicadores bibliométricos. En: *El estado de la ciencia*. p. 69-76. Red Iberoamericana de Indicadores de Ciencia y Tecnología (RICYT).

Franceschet, M., & Costantini, A. (2010). The effect of scholar collaboration on impact and quality of academic papers. *Journal of Informetrics*, 4 (4), 540–553.

García, Ángeles & Román, Adelaida (1998). Las publicaciones periódicas de Historia Antigua, Prehistoria y Arqueología: difusión internacional. En *Trabajos de Prehistoria*. Vol. 55, no. 1, 139-146.

Garfield, E. (1979). *Citation indexing. Its theory and application in science, technology and humanities*. New York: Wiley.

Garfield, E. (1979). Is citation analysis a legitimate evaluation tool?. *Scientometrics*, 1(4), 359-375.

Garfield E. (1972). Citation analysis as a tool in journal evaluation. *Science* 178:471-499

Gazni, A. & Didegah, F. (2011). Investigating different types of research collaboration and citation impact: a case study of Harvard University's publications. *Scientometrics*, 87 (2), 251-265.

Geisler E. (2005). The measurement of scientific activity: Research directions in linking philosophy of science and metrics of science and technology outputs. *Scientometrics*. 62(2):269-84.

Giménez, E., Román, A. & Sánchez, J. M. (1999). Aplicación de un modelo de evaluación a las revistas científicas españolas de Economía: una aproximación metodológica. En *Revista Española de Documentación Científica*. Vol. 22, no. 3, 309-324.

Glänzel, W. (2002). Coauthorship patterns and trends in the sciences: A bibliometric study with implications for database indexing and search strategies, 1980-1998. *Library Trends*, 50(3), 461-473.

Glänzel, W. (2001). Domestic characteristics in international scientific coauthorship relations. *Scientometrics*, 51(1), 69-115.

Glänzel, W. (2001). National characteristics in international and scientific co-authorship relations. *Scientometrics*, 51 (1), 69–115.

Glänzel, W. (1996). A bibliometric approach to the social sciences. National research performances in 6 selected social science areas, 1990-1992. *Scientometrics*, 35 : 291–307.

Glänzel, W., & Moed, H. (2002). Journal impact measures in bibliometric research. *Scientometrics*, 53(2), 171–193.

Glänzel, W., & Schubert, A. (2003). A new classification scheme of science fields and subfields designed for scientometric evaluation purposes. *Scientometrics*, 56(3), 257–367.

Glänzel, W., & Schubert, A. (2001). Double effort = Double impact? A critical view at international coauthorship in chemistry. *Scientometrics*, 50(2), 199–214

Glänzel, W., Schubert, A., & Czerwon, J. (1999). A bibliometric analysis of international scientific cooperation of the European Union (1985-1995). *Scientometrics*, 45 (2), 185–202.

Goldfinch S., Dale, T., & De Roue, K. (2003). Science from the periphery: Collaboration network and «Pheriphery effects» in the citation of New Zeland Crown Research Institutes articles, 1992-2000. *Scientometrics*, 57 (3), 321-337.

Gómez Caridad, I. & Bordons Cangas, M. (1996). Limitaciones en el uso de los Indicadores Bibliométricos para la Evaluación Científica. *Política Científica*. Vol. 46, 21-26.

Gómez, I., Fernández, M.T., & Sebastián, J. (1999). Analysis of the structure of international scientific cooperation networks through bibliometric indicators. *Scientometrics*, 44(3), 441–457.

Grupo SCImago (2006). Análisis de la cobertura de la base de datos Scopus. *El profesional de la información*, Vol.15, nº2: 144-145

Guerrero-Bote, V. P., Zapico-Alonso, F., Espinosa-Calvo, M. E., Gómez-Crisóstomo, R., & Moya-Anegón, F. (2007). The iceberg hypothesis: import-export of knowledge between scientific subject categories. *Scientometrics*, 71(3), 423–441.

Guerrero-Bote, V. P., Zapico-Alonso, F., Espinosa-Calvo, M. E., Gómez-Crisóstomo, R., & Moya-Anegón, F. (2006). Binary pathfinder: An improvement to the Pathfinder algorithm. *Information Processing and Management*, 42(6), 1484–1490.

Hargens, L. L. (2000). Using the literature: reference networks, reference contexts, and the social structure of scholarship. *American Sociological Review*, 65 : 846–865.

Hicks, D.(1999). The difficulty of achieving full coverage of international social science literature and the bibliometric consequences. *Scientometrics*, 44: 193–215.

Hjorland B, & Albrechtsen H. (1995). Toward a New Horizon in Information-Science - Domain-Analysis. *Journal of the American Society for Information Science*. 46(6):400-25.

Hsu, J.W., & Huang, D.W. (2010). Correlation between impact and collaboration. *Scientometrics*, 86(2), 317–324.

Hulme, F.W. (1923). *Statistical bibliography in relation to the growth of modern civilisation*. Grafton, London, p. 44.

Inzelt, A., Schubert, A., & Schubert, M. (2009). Incremental citation impact due to international co authorship in Hungarian higher education institutions. *Scientometrics*, 78 (1), 37-43.

Jiménez Contreras, Evaristo (2005). Bases para un concepto de las “-metrías” [en línea]. *Cognosfera*. Universidad de Granada de España. Disponible en: http://www.ugr.es/~rruizb/cognosfera/sala_de_estudio/bibliometria/las_metrias.DOC [consulta: 24 mayo 2012].

Jiménez Contreras, E., Ruiz Pérez, R., & Delgado-Lopez-Cozar, E. (2005). Evaluación del impacto de las revistas latino-americanas y españolas de Psiquiatría a través de las bases de datos [comunicación]. In 9º Congreso Mundial de Informação em Saúde e Bibliotecas, Salvador de Bahía (Brasil) ,20-23 septiembre de 2005

Katz, J. (1994). Geographical proximity and scientific collaboration. *Scientometrics*, 31 (1), 31-43.

Katz, J., & Hicks, D. (1997). How much is a collaboration worth? A calibrated bibliometric model. *Scientometrics*, 40(3), 541–554.

King, D. A. (2004). The scientific impact of nations. *Nature*, 430, 311–316.

Kraut R., Egado C., & Galeguer, J. (1988). Patterns of contact and communication in scientific research collaboration. 'CSCW '88: Proceedings of the 1988 ACM conference on Computer-supported cooperative work', ACM, New York, NY, USA, 1-12.

Kyvik, S.(2003). Changing trends in publishing behaviour among university faculty, 1980–2000. *Scientometrics*, 58: 35–48.

Lancho Barrantes, B., Guerrero Bote, V., & Moya Anegón, F. (2010a). The iceberg hypothesis revisited. *Scientometrics*, 85(2), 443–461.

Lancho Barrantes, B., Guerrero Bote, V., & Moya Anegón, F. (2010b). What lies behind the averages and significance of citation indicators in different disciplines? *Journal of Information Science*, 36(3), 371–382.

Lancho Barrantes, B., Guerrero Bote, V., Chinchilla Rodríguez, Z., & Moya Anegón, F. (2012). Citation flows in the zones of influence of scientific collaborations. *Journal of the American Society for Information Science and Technology*, 63(3), 481-489.

Leimu, R., & Koricheva, J. (2005). Does scientific collaboration increase the impact of ecological articles? *Bioscience*, 55 (5), 438–443.

Leta, J., & Chaimovich, H. (2002). Recognition and international collaboration: the Brazilian case. *Scientometrics*, 53 (3), 325–335.

Lewinson, G., & Cunningham, P. (1991). Bibliometric studies for the evaluation of trans-domestic research. *Scientometrics*, 21(2), 223–244.

Lewison, G. (2001). Evaluation of books as research outputs in history of medicine. *Research Evaluation*, 10: 89–95.

Leydesdorff, L. (2004). Clusters and maps of science journals based on biconnected graphs in Journal Citation Reports. *Journal of Documentation*. 60(4), 371-427.

Leydesdorff, L., & Bensman, S. (2006). Classification and powerlaws: The logarithmic transformation. *Journal of the American Society for Information Science and Technology*, 57(11), 1470–1486.

Lindholm-Romantschuk, Y., & Warner, J. (1996). The role of monographs in scholarly communication: an empirical study of philosophy, sociology and economics. *Journal of Documentation*, 54 : 389–404.

López Piñero, JM & Terrada, ML. (1992). Los indicadores bibliométricos y la evaluación de la actividad médico-científica (I). Usos y abusos de la bibliometría. *Medicina Clínica (Barcelona)*. 98:64-68.

López Piñero JM & Terrada ML. (1992). Los indicadores bibliométricos y la evaluación de la actividad médico-científica (II). La comunicación científica en las distintas áreas de las ciencias médicas. *Medicina Clínica (Barcelona)*. 98:101-106.

López Piñero JM, Terrada ML. (1992). Los indicadores bibliométricos y la evaluación de la actividad médico-científica (III). Los indicadores de producción, circulación y dispersión, consumo de la información y repercusión. *Medicina Clínica (Barcelona)*. 98:142-148.

López Piñero, JM & Terrada, ML. (1992). Los indicadores bibliométricos y la evaluación de la actividad médico-científica (IV). La aplicación de los indicadores. *Medicina Clínica (Barcelona)*. 98:384-388.

Lundberg, J. (2007). Lifting the crown—citation z-score. *Journal of Informetrics*, 1, 145–154.

Luukkonen, T., Tijseen, R. J. W., Persson, O., & Sivertsen, G. (1993). The measurement of international scientific collaboration. *Scientometrics*, 28(1), 15-36.

Ma, N., & Guan, J. C. (2005). An exploratory study on collaboration profiles of Chinese publications in Molecular Biology. *Scientometrics*, 65 (3), 343–355.

Martin BR & Irvine J. (1983). Assessing basic research: some partial indicators of scientific progress in radio astronomy. *Research Policy*, 12: 61-90.

Merton, R. K. (1968). The Matthew effect in science. *Science*, 159, 56–63.

Milanés-Guisado Y, Pérez-Rodríguez Y, Peralta-González M, Ruiz-Ramos, M. (2008). Science evaluation studies: Theoretical and metric approaches. *Acimed: revista cubana de los profesionales de la información y la comunicación en salud*, 18;6.

Moed, H. F. (2005). *Citation analysis in research evaluation* (p. 346). Dordrecht: Springer.

Moed, H. F., De Bruin, R. E., & Vanleeuwen, T. N. (1995). New Bibliometric Tools for the Assessment of National Research Performance – Database Description, Overview of Indicators and First Applications. *Scientometrics*, 33(3), 381-422.

Moed, H., De Bruin, R., Nederhof, A., & Tijssen, R. (1991). International scientific co-operation and awareness within the European Community: Problems and perspectives. *Scientometrics*, 21(3), 291–311.

Moed, H. F. (1988). The Use of on-line databases for bibliometric analysis. In: L. Egghe, R. Rousseau (Eds), *Informetrics 87/88*. Elsevier Science Publishers, Amsterdam, pp. 133–146

Molteni, V., & Zulueta, M. A. (2002). Análisis de la visibilidad internacional de la producción científica argentina en las Bases de datos Social Science Citation Index y Arts and Humanities Citation Index de 1990–2000: estudio bibliométrico. *Revista Española de Documentación Científica*, 25(4), 455–465.

Moya Anegón, F., Chinchilla Rodríguez, Z., Corera Álvarez, E., González Molina, A., Hassan Montero, Y., & Vargas Quesada, B. (2008). *Indicadores bibliométricos de la actividad científica española: 2002–2006*. Madrid: Fecyt.

Moya Anegón, F., Chinchilla Rodríguez, Z., Vargas Quesada, B., Corera Álvarez, E., Muñoz Fernández, F., González Molina, A., & Herrero Solana, V. (2007). Coverage analysis of Scopus: A journal metric approach. *Scientometrics*, 73(1), 53–78.

Moya-Anegón, F., Chinchilla-Rodríguez, Z., Vargas-Quesada, B., González-Molina, A., Muñoz-Fernández, F. J., Corera-Álvarez, E., & Herrero-Solana, V. (2006). Visualizing and analyzing the spanish science structure: ISI web of science 1990-2005. *El Profesional de la información*, 15, (4), 258-269

Moya-Anegón, F. d., Chinchilla-Rodríguez, Z., Vargas-Quesada, B., Muñoz-Fernández, F. J., Corera-Álvarez, E., & Herrero-Solana, V. (2005). Análisis de dominio institucional: la producción científica de la Universidad de Granada. *Revista española de documentación científica* , 170-195.

Moya Anegón, F., Vargas Quesada, B., Herrero Solana, V., Chinchilla Rodríguez, Z., Corera Álvarez, E., & Muñoz Fernández, F. J. (2004). A new technique for building maps of large scientific domains based on the cocitation of classes and categories. *Scientometrics*, 61, 129–145.

Narin, F., & Whitlow, E. (1990). Measurement of scientific cooperation and coauthorship in CEC-related areas of science. Luxembourg: European Community.

Narin, F., Stevens, K., & Whitlow, E. (1991). Scientific cooperation in Europe and the citation of multidomestically authored papers. *Scientometrics*, 21(3), 313–323.

Nature 2005 Editorial. Not-so-deep impact. *Nature*. 435: 1003-1004.

Nederhof, A. J. (2006). Bibliometric monitoring of research performance in the Social Sciences and the Humanities: a review. En: *Scientometrics*. Vol. 66, no. 1, 81-100.

Nederhof, A. J., Zwaan, R. A., De Bruin, R. E., & Dekker, P. J. (1989). Assessing the usefulness of bibliometric indicators for the humanities and the social sciences. *Scientometrics*, 15 : 423–435.

Nederhof, A. (1988). The validity and reliability of evaluation of scholarly performance. Van Raan AFJ. *Handbook of quantitative studies of science and technology* , 193-228.

Nock, D. A. (2001). Careers in print: Canadian sociological books and their wider impact, 1975–1992. *Canadian Journal of Sociology/Cahiers canadiens de sociologie*, 26: 469–485.

Núñez Jover, J. (2012). La ciencia y la tecnología como procesos sociales. Lo que la educación científica no debería olvidar. Disponible en: <http://www.oei.es/salactsi/nunez03.htm>

Pao, M. L. (1981). Coauthorship as communication measure. *Library Research*, 2, 327-338.

Persson, O., Glänzel, W., & Danell, R. (2004). Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. *Scientometrics*, 60(3), 421–432.

Pichappan, P. (1995). A dual refinement of journal self-citation measures. *Scientometrics*, 33(1), 13–21.

Pinski, G., & Narin, F. (1976). Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics. *Information Processing and Management*, 12(5), 297–312.

Podlubny, I. (2005). Comparison of scientific impact expressed by the number of citations in different fields of science. *Scientometrics*, 64(1), 95–99.

Price, D. J. S. (1970). Citation measures of hard science, soft science, technology, and non-science. In: C. E. Nelson, D. Pollack (Eds), *Communication Among Scientists and Engineers*. Lexington, Mass., Lexington books.

Price, D. J. S. (1965). Network of scientific papers. *Science*. 149:510-515.

Price, D. J. S., & Beaver, D. B. (1966). Collaboration in an invisible college. *American psychologist*, 21 (11), 1011-1018.

Pritchard, A. (1969). Statistical bibliography or bibliometrics?. *Journal of Documentation*. v. 25, n. 4, p. 348-349.

Pudovkin, A., & Garfield, E. (2002). Algorithmic procedure for finding semantically related journals. *Journal of the American Society for Information Science and Technology*, 53(13), 1113–1119.

Ramírez A.M.; García E.O. & Del Río J.A. (2000). Renormalized Impact Factor. *Scientometrics*, 47 (1), pp. 3-9

Román Román, A. & Alcaín, M.D. (2009) Cómo citan los investigadores de Ciencias Humanas y Sociales. Jornadas de Indicadores .FESABID- IX Jornadas.

Sanz Menéndez, L. (2004). Evaluación de la investigación y sistema de ciencia. Disponible en: <http://www.ipp.csic.es/RePec/ipp/wpaper/dt0407.pdf> [Consultado: 10 de mayo de 2012]

Schmoch, U., & Schubert, T. (2008). Are international co-publications an indicator for quality of scientific research?. *Scientometrics*, 74(3), 361–77.

Schvaneveldt, R. W., Dearholt, D. W., & Durso, F. T. (1988). Graph theoretic foundations of Pathfinder networks. *Computers and Mathematics with Applications*, 15(4), 337–345.

SCImago Journal & Country Rank. (2008). SCImago Research Group. Accessed 26 March, 2009 from <http://www.scimagojr.com/>

Sen, B K. (1992). Normalised impact factor. *Journal of Documentation*, 48, 3.

Singh, J. (2005). Collaborative networks as determinants of knowledge diffusion patterns. *Management Science*, 51 (5), 756–770.

Small, H. G., Crane, D. (1979). Specialties and disciplines in science and social science: an examination of their structure using citation indexes. *Scientometrics*, 1 : 445–461.

Sombatsompop, N., & Markpin, T. (2005). Making an equality of ISI impact factors for different subject fields. *Journal of the American Society for Information Science and Technology*, 56(7), 676–683.

Sooryamoorthy, R. (2009). Do types of collaboration change citation? Collaboration and citation patterns of South African science publications. *Scientometrics*, 81(1), 177–193.

Stack, S. (2002). Gender and scholarly productivity: The case of criminal justice. *Journal of Criminal Justice*, 30(3), 175–182.

Sudhier Pillai, K.G. (2007). Authorship patterns in physics literature: An informetric study on citations in doctoral theses of the Indian Institute of Science. *Annals of Library and Information Studies*, 54, 90-94.

Terrada, ML & Peris Bonet, R. (1982). Bibliometría de la literatura pediátrica española (1974-1981). *Anuario Español de Pediatría*. 17(14):105-114.

Thompson, J. W. (2002). The death of the scholarly monograph in the humanities? Citation patterns in literary scholarship. *Libri*, 52 (3) : 121–136.

Urbizagástegui Alvarado, R. (2007). A bibliometria: história, legitimação e estrutura En: Batista Brandão Toutain, L. M. *Para Entender a Ciência da Informação*. Salvador, Brasil: EDUFBA, 185-217.

Van Leeuwen, T. N., & Moed, H. F. (2002). Development and application of journal impact measures in the Dutch science system. *Scientometrics*, 53(2), 249–266.

Van Leeuwen, T. N., Visser, M. S., Moed, H. F., Nederhof, T. J., & Van Raan, A. F. J. (2003). The Holy Grail of science policy: Exploring and combining bibliometric tools in search of scientific excellence. *Scientometrics*, 57(2), 257–280.

Van Raan, AFJ. (2006). Statistical properties of bibliometric indicators: research group indicator distributions and correlations. *JASIST* 57 (3): 408-430.

Van Raan, AFJ. (2005). Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics*, 62(1):133-43.

Waltman, L., & Van Eck, N. J. (2009). Some comments on Egghe's derivation of the impact factor distribution. *Journal of Informetrics*, 3(4), 363–366.

Wormell, I. (1998). Informetric analysis of the international impact of scientific journals: How 'International' are the international journals? *Journal of Documentation*, 54(5), 584–605.

Zhao, Q., & Guan, J. (2011). International collaboration of three 'giants' with the G7 countries in emerging nanobiopharmaceuticals. *Scientometrics*, 87 (1), 159-170.

Zuckerman, H. A. (1968). Patterns of name ordering among authors of scientific papers: A study of social symbolism and its ambiguity. *American Journal of Sociology*, 74, 276-291.

*Estudio de los flujos de citación científica
y su relación con los indicadores de impacto*

PARTE II: ARTÍCULOS

8 Papers

A1: The Iceberg Hypothesis revisited

A2: What lies behind the averages and significance of citation indicators in different disciplines?

A3: Citation flows in the zones of influence of scientific collaborations

A4: Citation increment between collaborating countries

ARTÍCULO

A1: The Iceberg Hypothesis revisited

Bárbara S. Lancho Barrantes, Vicente P. Guerrero Bote and Félix de Moya

Anegón

In press in:

Scientometrics (2010) 85:443–461

The iceberg hypothesis revisited

Bárbara S. Lancho-Barrantes (1), Vicente P. Guerrero-Bote (1) and

Félix Moya-Anegón (2)

(1) Departamento de Información y Comunicación, Grupo SCImago, Universidad de Extremadura, 06071, Badajoz, Spain.

(2) Grupo SCImago, CSIC, CCHS, IPP, 28037, Madrid, Spain.

Abstract

A study is described of the rank/JIF (Journal Impact Factor) distributions in the high-coverage Scopus database, using recent data and a three-year citation window. It includes a comparison with an older study of the Journal Citation Report categories and indicators, and a determination of the factors most influencing the distributions. While all the specific subject areas fit a negative logarithmic law fairly well, those with a greater External JIF have distributions with a more sharply defined peak and a longer tail—something like an iceberg. No S-shaped distributions, such as predicted by Egghe, were found. A strong correlation was observed between the knowledge export and import ratios.

Finally, data from both Scopus and ISI were used to characterize the rank/JIF distributions by subject area.

Keywords

Knowledge export; Scientometrics; Journal impact factor; Citation analysis

Introduction

At present, the Journal Impact Factor (JIF) is the most extensively used tool to measure the visibility and diffusion of a scientific journal. Its strength lies in its stability, facility of calculation, and accessibility. But these factors have perhaps also been a source of weakness since they have led to its sometimes being used arbitrarily and uncritically, and hence to its being disputed and criticized. Indeed, its failure to normally take into account a series of agents which affect it directly but are not necessarily connected with any journal has been a source of loss of credibility. Its advantages and limitations, the context of its application, and the functions used in its calculations are reviewed in Glänzel and Moed (2002).

Nevertheless, it is difficult to use citation-based indicators for comparisons across disciplines because the various fields of science differ in their referencing patterns (Pinski and Narin 1976; Garfield 1979), their proportions of different types of publication, average citation rates, ages of the references, (Lundberg 2007), and time-lag to reach the citation peak (Moed 2005; Lancho-Barrantes et al., in press).

In this sense, neither is it clear how to structure science into different disciplines. Many studies use the Subject Category Listing of Thomson Scientific's JCR (Journal Citation Report). These categories were developed manually more than forty years ago to classify scientific journals, with subdivisions being added as the categories grew. To assign a journal to a category, this listing follows a heuristic that takes the journal's references into account using, among others, the Hayne–Coulson algorithm (Pudovkin and Garfield 2002).

Despite there being many authors who do not agree, and we also have classification schemes of science fields and subfields designed for scientometric evaluation purposes (Glänzel and Schubert 2003), it is generally accepted that the JCR categories are sufficiently explicit informative units with which to represent the disciplines comprising the whole of science. They are presently used, for example, to construct maps of science that show how the different categories are related (Moya-Anegón, et al. 2004) by applying such citation analysis and pruning techniques as the Pathfinder algorithm (Schvaneveldt et al. 1988; Guerrero-Bote et al. 2006), and to evaluate scientific production in different fields (Podlubny 2005; Van Leeuwen and Moed 2002; Van Leeuwen et al. 2003; Sombatsompop and Markpin, 2005; Moya-Anegón et al. 2008).

Scopus offers a similar classification, with the whole of science divided into 27 subject areas, further subdivided into 295 specific subject areas. Another difference is that in this case it is the documents themselves rather than the journals which are assigned to the different specific subject areas.

In spite of the widespread use of the JIF, little work has been done to study the rank/JIF distributions in the different categories. In one such study, Guerrero-Bote et al. (2007) concluded that, as was to be expected, the rank/JIF distributions in most Subject Categories fit a negative logarithmic law fairly well. The exceptions that fit less well have a sharper than exponential initial descent and a longer tail. The factor with the greatest influence on this deviation was the category's exterior visibility, which was concentrated in just a very few journals. In Subject Categories with a high exterior visibility, the effect of this concentration was that the initial descent of the rank/JIF distribution was steeper than exponential. Also, the slope of the logarithmic law was strongly correlated with the External JIF and its standard deviation. The study also showed how certain scientific disciplines are exporters of ideas in the sense that the knowledge generated within them is visible from other disciplines. This suggested an "Iceberg" metaphor in the sense that scientific categories can be seen as icebergs, some of whose tips above the surface are high enough to be sighted in the distance from some other iceberg (category), while in others the emerged part hardly exists at all, and the iceberg (the category) is only visible when one is actually on it (i.e., in the same category).

Recently, Egghe (2009) presented a mathematical analysis of the rank-order distribution of JIFs based on the central limit theorem, predicting that the distribution would be S-shaped. This is, however, not concordant with the distributions shown in Guerrero-Bote et al. (2007). In the same issue of the same journal, Waltman and Van Eck (2009) showed that its analysis is not in agreement with certain empirical data, and argued that the assumption of Egghe (2009) that the articles published in each journal can be considered as random samples of the population of all articles published in a field is unrealistic because they are neither totally random nor of the same size for the application of the central limit theorem.

Although the study of Guerrero-Bote (2007) included an exhaustive verification of all the Journal Citation Report subject categories, this was not done in either the Egghe or the Waltman and Van Eck studies. Hence a number of questions suggest themselves. To what measure do these conclusions depend on the database used? What is the influence of the subject category structure and the form of assigning the journals? Do the conclusions still hold? Does the width of the citation window have any effect?

Our purpose with the present study is to look for an answer to these questions. To this end, we shall study the Rank/JIF distributions in the Scopus database which has greater coverage, considering a more recent date and a wider citation window. But, rather than just checking the same indicators as in the earlier study in another database (with other parameters), we also propose to characterize the rank/JIF distributions by subject area using data from the experiment of Guerrero-Bote et al. (2007).

Data and method

We used Scopus as the data source for the computation of the indicators because it best represents the overall structure of world science at a global scale. Scopus is the world's largest scientific database. It covers all the journals included in the Thomson Scientific Web of Science (WoS) and more (Moya-Anegón et al. 2007), and its coverage is statistically balanced in terms of subjects, countries, languages, and publishers. The greater part of the data was retrieved from the SCImago Journal & Country Rank (<http://www.scimagojr.com/>) website, limiting the citations to references in papers published in 2006 to papers published in 2003, 2004, and 2005. The data was downloaded during September 2008, and was stored in a relational database.

Scopus also provides a subject classification consisting of 295 specific subject areas grouped into 26 subject areas. In addition, there is the General subject area containing multidisciplinary journals, such as Nature or Science. It is the papers which are classified into the different specific subject areas instead of

the journals as in JCR. The subject areas are grouped into four categories on the Scopus “Basic Search” page (see the Scopus website, www.scopus.com, visited on 7 August 2009).

The four Scopus categories are:

- Life Sciences (4,300 titles): Agricultural & Biological Sciences; Biochemistry, Genetics & Molecular Biology; Immunology & Microbiology; Neuroscience; Pharmacology, Toxicology & Pharmaceutics.

- Physical Sciences (7,200 titles): Chemical Engineering; Chemistry; Computer Science; Earth & Planetary Science; Energy; Engineering; Environmental Science; Materials Science; Mathematics; Physics & Astronomy.

- Social Sciences (5,300 titles): Arts & Humanities; Business, Management & Accounting; Decision Sciences; Economics, Econometrics and Finance; Psychology; Social Sciences.

- Health Sciences (6,800 titles, including 100% coverage of Medline titles): Medicine; Nursing; Veterinary; Dentistry; Health Professions.

In addition to this pre-classification, we assigned the journals to those specific subject areas in which they have papers classified. We limited the computations to the 260 specific subject areas with more than 6 journals in the SCImago Journal & Country Rank.

For each journal, we calculated the number of citable papers (Articles, Reviews, and Conference Papers published in 2003, 2004, or 2005), the total number of citations received (made in 2006 to papers published in 2003, 2004, or 2005), the number of subcitations (citations from journals of the same specific subject area) received (made in 2006 to papers published in 2003, 2004, or 2005), and the number of references (made in 2006 to papers published in 2003, 2004, or 2005). These data were used to calculate the following indices:

- Knowledge Export Ratio (%) (Wormell 1998) (%export):

$$\% \text{ Export} = \frac{100 \cdot (TotCitations - SubCitations)}{TotCitations}$$

- Knowledge Import Ratio (%) (%import):

$$\% \text{ Import} = \frac{100 \cdot (References - SubCitations)}{References}$$

- Internal Journal Impact Factor: similar to the JIF, but only taking into account the citations from that same Specific Subject Area (Subcitations).
- External Journal Impact Factor: similar to the JIF, but only taking into account the citations not from that same Specific Subject Area.

In the definitions of these indices, it has to be borne in mind that, since many journals are assigned to more than one specific subject area, the references from one such journal to any of these specific subject areas are Subcitations.

For each specific subject area, we plotted the impacts against the descending-order ranks, and calculated the best-fit exponential trend line. This line is computed as a least squares fit of the impact factor to the logarithms of the ranks. We used as indices of the distribution the slope a and intersect b of these trend lines, and, from the fits, the R^2 coefficient and the deviation divided by the Mean JIF squared. This last index was particularly useful in detecting the specific subject areas whose distribution has a sharper than exponential peak.

We then performed a correlation analysis of the different indices in order to study their influence on the fit of each distribution to its logarithmic approximation.

Finally, to compare the results of the two studies (henceforth we shall use ISI-97 to refer to the previous study, and Scopus-06 to refer to the present study) and to characterize the rank/JIF distributions by disciplines, we adopted the 27 subject areas of Scopus since the specific subject areas/categories of the two databases are not equivalent. To this end, besides the given association of each Scopus specific subject area with a Scopus subject area, we associated the ISI-97 categories with Scopus areas.

Results and discussion

Figure 1 shows the distribution of the Specific Subject Area Algebra and Number Theory. This is the one which most closely fits a logarithmic distribution. By way of illustration, the figure includes the External JIF, the Export Ratio, and an exponential trend line with respect to the JIF.

This trend line corresponds to the expression: $y = -0.40 \cdot \ln(x) + 1.71$

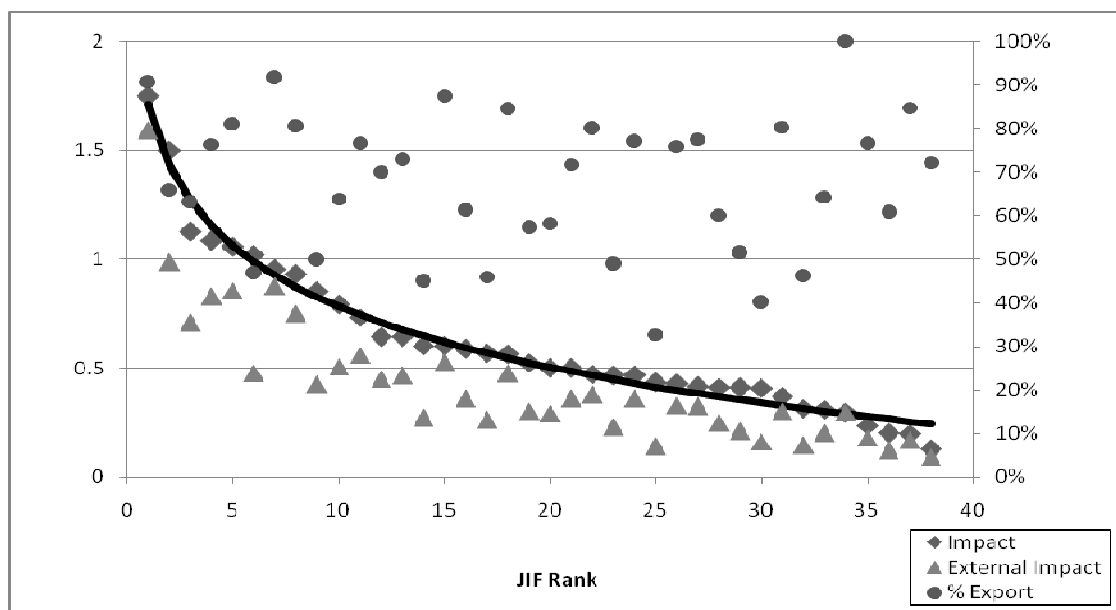


Figure 1: Distribution of the JIFs of the Subject Category Algebra and Number Theory. Added are the curve corresponding to the logarithmic trend of the distribution, and the External JIFs and export rates of each of the journals.

The coefficient $a = -0.40$ characterizes the sharpness of the trend, and the coefficient $b = 1.71$ the intersect, and hence the height of the peak. The goodness of the fit to this logarithmic approximation is characterized by the values of R^2 and the squared deviation:

$$R^2 = 0.9815$$
$$\left(\frac{Diff}{JIF} \right)^2 = 0.0015$$

This specific subject area was not in the JCR in 1997, and therefore its study is new. But comparing it with Applied Mathematics, for example, whose trend line was given by the expression: $y = -0.33 \cdot \ln(x) + 1.75$

with goodness-of-fit parameters:

$$R^2 = 0.9715$$
$$\left(\frac{Diff}{JIF} \right)^2 = 0.0050$$

one observes that the two distributions are similar.

The close correlation of the External JIF with the JIF (larger JIFs usually correspond to larger External JIFs) was also observed in ISI-97.

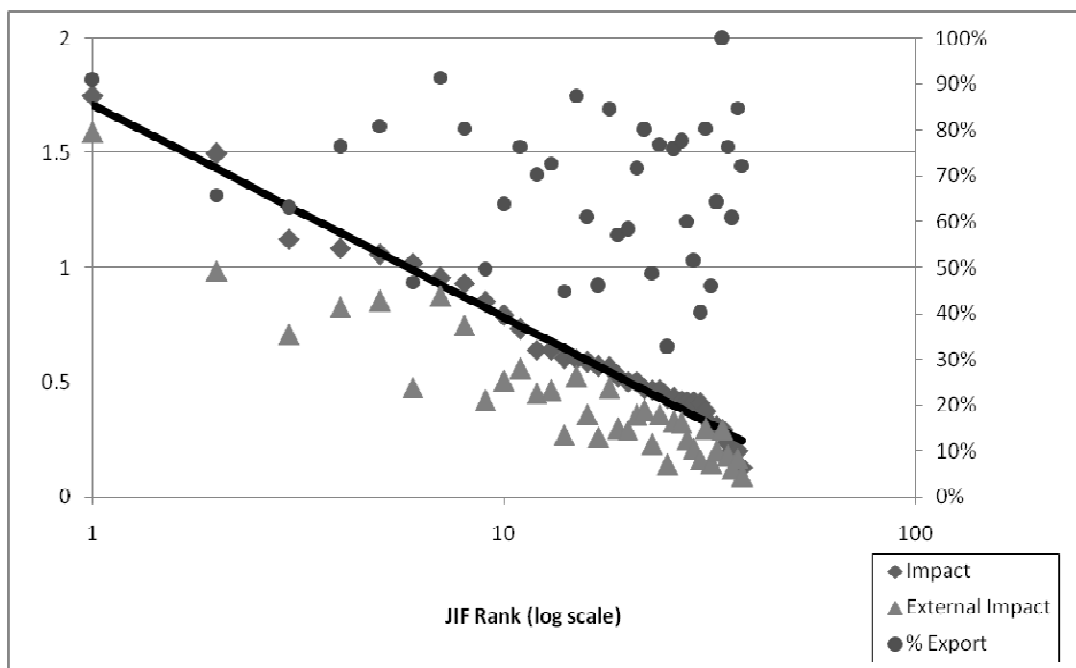


Figure 2: As Fig. 1, but with a logarithmic scale for the ranks.

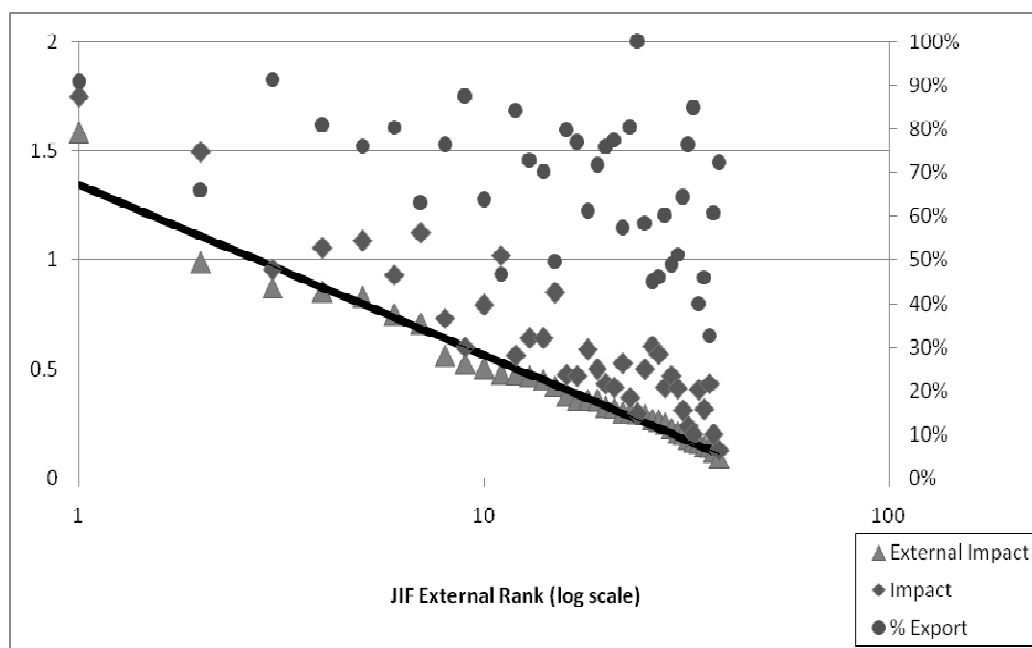


Figure 3: Distribution of the External JIFs of the Subject Category *Algebra and Number Theory* on a logarithmic scale. Added are the export rates and the straight line trends corresponding to the External JIFs and the JIFs.

Figure 2 shows the same data but with the ranks on a logarithmic scale to better appreciate the goodness of the fit of the distribution to a logarithmic function. (Of course, the slope and intercept are the same values as given above for Fig. 1: $a = -0.40$ and $b = 1.71$.) No relationship was observed between the export ratio and the JIF. This contradicts the idea that journals with a greater JIF also have a higher export ratio and vice versa.

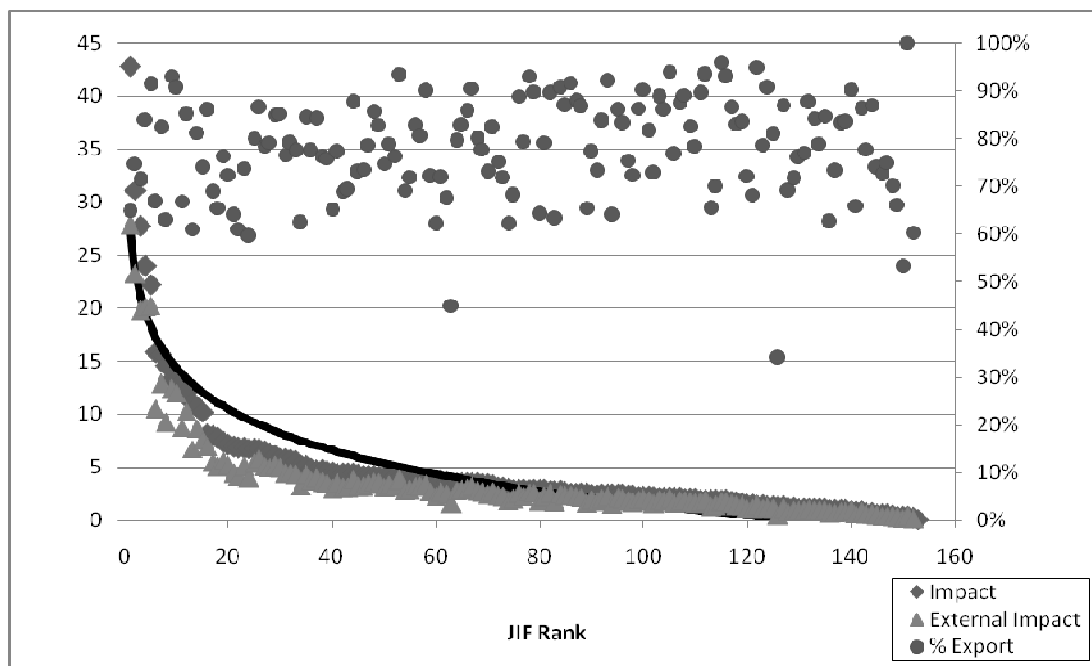


Figure 4: Distribution of the JIFs of the Subject Category Cell Biology. Added are the curve corresponding to the logarithmic trend of the distribution, and the External JIFs and export rates of each of the journals.

Figure 3 is a plot of the same data but ranked by External JIF. One sees that the Matthew Effect (Merton 1968), the mechanism of preferential attachment which is well known for generating negative logarithmic laws (Leydesdorff and Bensman 2006), is even more marked.

Indeed, the trend of the distribution is steeper than a negative logarithmic law, and the external component accounts for most of the slope of the JIFs.

Figure 4 shows the distribution of the Specific Subject Area Cell Biology. This is one which gives the poorest fit to the logarithmic approximation.

The trend line shown corresponds to the expression:

$$y = -5.58 \cdot \ln(x) + 27.21$$

The values of R^2 and the squared deviation that characterize the goodness of fit to this logarithmic approximation were:

$$R^2 = 0.8609$$
$$\left(\frac{Diff}{JIF} \right)^2 = 0.9834$$

The previous study (ISI-97), which used the WoS database with a 2-year citation window and references from the year 1997, gave slightly different parameters: the goodness of fit to the logarithmic approximation was characterized by the values:

$$R^2 = 0.8155$$
$$\left(\frac{Diff}{JIF} \right)^2 = 1.4078$$

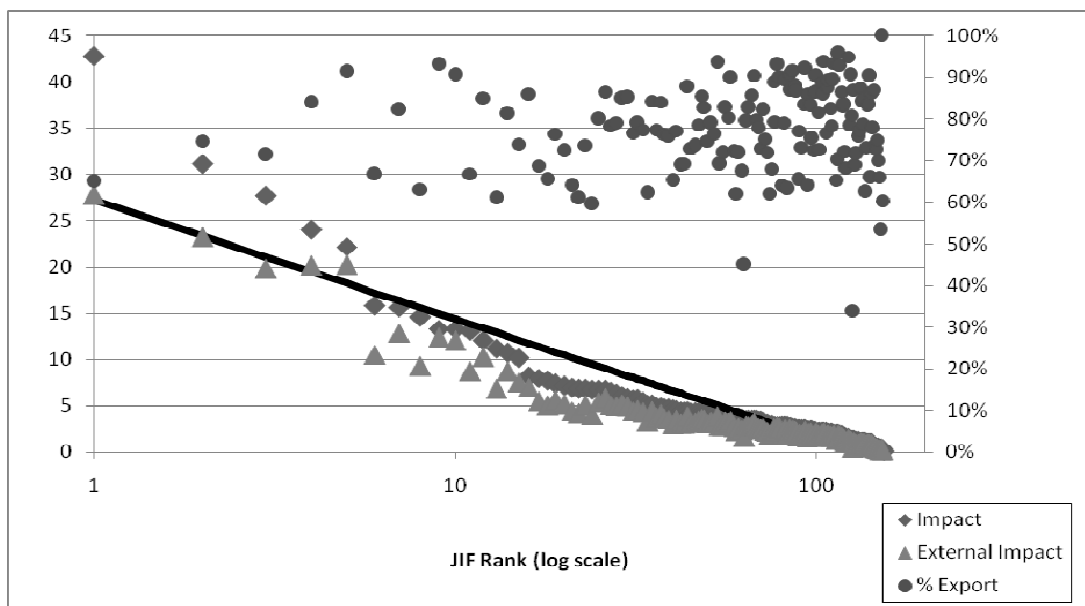


Figure 5: As Fig. 4, but with a logarithmic scale for the ranks.

Its logarithmic trend line corresponded to the expression:

$$y = -4.87 \cdot \ln(x) + 22.19$$

One observes that the distribution for this specific subject area has a sharper initial descent than the logarithmic line, and a longer tail. This can also be seen in Fig. 5, which is a plot of the same data but with a logarithmic scale for the ranks.

The features shown for this Scopus Specific Subject Area are typical of all the specific subject areas that do not fit the logarithmic approximation well, as was reflected more clearly in the value of the squared deviation than in the R2 coefficient.

Figure 6 shows the distribution of the external component of the JIFs. One observes in this case that it accounts for more than half the slope of the logarithmic approximation, and most of the deviation with respect to that approximation.

As was expected given what had been observed in the previous study, in all the specific subject areas the distributions of the JIFs were fairly close to an exponential, so that in the first JIFs there was a major jump from one to the next, with these jumps getting smaller as one advanced in the ranking, until reaching the tail where the variations were very small. The data presented in Table 1 show that 78% of the specific subject areas had a squared deviation lower than 0.12, which means that most of the Scopus-06 specific subject areas fit the logarithmic approximation fairly well. The exceptions such as Developmental Biology or Biochemistry with a high squared deviation were because they had very large differences in the impact factors of the top-ranked journals and the rest. For example, there was already a large gap between Biochemical Society Transactions and Annual Review of Biochemistry, the top two journals of the Biochemistry specific subject area, and a similarly large gap before the third journal. The case was similar in ISI-97 in which 80% of the categories had a squared deviation less than 0.12, and Medicine, General & Internal was the category with the highest production and also one with the highest squared deviations.

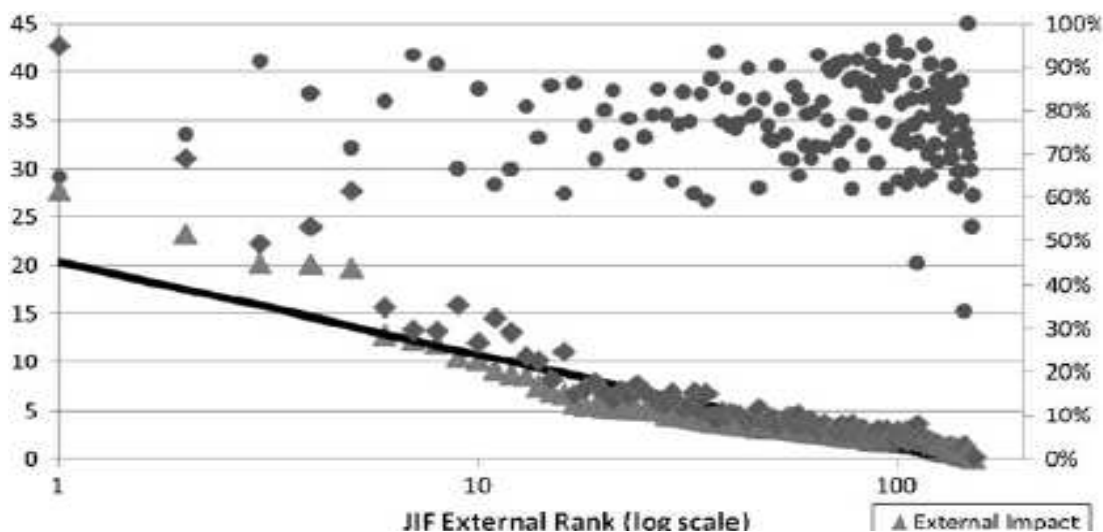


Figure 6: Distribution of the External JIFs of the Specific Subject Area Cell Biology on a logarithmic scale. Added are the export rates and the straight line trends corresponding to the External JIFs and the JIFs

The average R^2 of the Scopus-06 specific subject areas was 0.93, and in ISI-97 it was 0.92. The squared deviation in both studies was 0.20. Thus they both reflect an overall good fit of the specific subject areas to the logarithmic approximation.

As was mentioned above, to compare more precisely the results of the two studies, we used the 27 Scopus Subject Areas. To this end, the means of the exterior journal impact factor, average journal impact factor, a, R^2 , and squared deviation in both Scopus-06 and ISI-97 for all these Scopus Subject Areas were subjected to a correlation analysis. The results are given in Table 2.

One observes in the table that all these indicators were strongly correlated between the two databases, i.e., their behaviour was similar not only overall, but also when observed by subject area.

With respect to the squared deviations (Fig. 7), apart from the General subject area, the greatest difference occurred in Chemical Engineering. This was due to the Process Chemistry and Technology specific subject area whose top-ranked journal had major differences in journal impact factor, with the first ranked, *Catalysis Today*, differing greatly from the mean. Immunology & Microbiology had a higher value in Scopus-06 because of the journal *Annual Review of Immunology* which also had a high impact factor with respect to the rest. Two other areas with the same behaviour were Neuroscience and Pharmacology, Toxicology & Pharmaceuticals.

Figure 8 shows that there were also only slight differences in R^2 . For example, the specific subject areas belonging to the subject area of Economics, Econometrics & Finance, with a higher mean in Scopus-06 ($R^2 = 0.98$) than in ISI-97 ($R^2 = 0.84$), fit the logarithmic approximation better in Scopus-06. The same was the case with the subject area of Psychology with a higher mean value in Scopus-06 ($R^2 = 0.97$) than in ISI-97 ($R^2 = 0.90$), and in Pharmacology, Toxicology & Pharmaceuticals (Scopus-06 $R^2 = 0.90$, ISI-97 $R^2 = 0.81$). In Decision Sciences, however, the mean was slightly higher in ISI-97 ($R^2 = 0.98$) than in Scopus-06 ($R^2 = 0.90$).

Table 1 R² coefficient and squared deviation for the Specific Subject Areas with an impact distribution in 2006

SUBJECT CATEGORY	R ²	Diff ²	SUBJECT CATEGORY	R ²	Diff ²
Computer Science (miscellaneous)	0.975	0.042	Reproductive Medicine	0.926	0.108
Dentistry (miscellaneous)	0.951	0.042	Statistics, Probability and Uncertainty	0.881	0.111
Renewable Energy, Sustainability and the Environment	0.927	0.042	Fuel Technology	0.930	0.115
Instrumentation	0.964	0.043	Management of Technology and Innovation	0.929	0.120
Obstetrics and Gynecology	0.971	0.045	Health, Toxicology and Mutagenesis	0.911	0.122
Analytical Chemistry	0.961	0.045	Gastroenterology	0.948	0.122
Aging	0.960	0.045	Electrochemistry	0.905	0.124
Computers in Earth Sciences	0.961	0.045	Neuroscience (miscellaneous)	0.957	0.129
Ecological Modeling	0.903	0.045	Transplantation	0.891	0.133
Microbiology (medical)	0.961	0.046	Leadership and Management	0.865	0.134
Computer Science Applications	0.940	0.047	Pharmacology	0.945	0.137
Computational Mathematics	0.888	0.047	Internal Medicine	0.960	0.138
Pharmacology, Toxicology and Pharmaceutics (miscellaneous)	0.967	0.048	Cardiology and Cardiovascular Medicine	0.940	0.142
Atomic and Molecular Physics, and Optics	0.959	0.050	Business, Management and Accounting (miscellaneous)	0.898	0.147
Waste Management and Disposal	0.916	0.051	Plant Science	0.914	0.166
Soil Science	0.938	0.052	Nursing (miscellaneous)	0.873	0.168
Hardware and Architecture	0.971	0.053	Information Systems	0.923	0.169
Biotechnology	0.969	0.053	Advanced and Specialized Nursing	0.832	0.171
Ecology, Evolution, Behavior and Systematics	0.953	0.054	Space and Planetary Science	0.924	0.182
Pulmonary and Respiratory Medicine	0.975	0.055	Geophysics	0.870	0.183
Clinical Biochemistry	0.956	0.057	Spectroscopy	0.883	0.187
Critical Care Nursing	0.874	0.059	Computer Networks and Communications	0.900	0.190
Arts and Humanities (miscellaneous)	0.852	0.059	Energy (miscellaneous)	0.906	0.206
Stratigraphy	0.890	0.059	Infectious Diseases	0.941	0.208
Complementary and Alternative Medicine	0.938	0.061	Biophysics	0.916	0.209
Anesthesiology and Pain Medicine	0.964	0.062	Embryology	0.789	0.220
Electronic, Optical and Magnetic Materials	0.969	0.063	Pharmaceutical Science	0.912	0.224
Bioengineering	0.963	0.064	Immunology and Allergy	0.945	0.234
Neurology (clinical)	0.956	0.064	Genetics	0.940	0.237
Cellular and Molecular Neuroscience	0.907	0.064	Biomedical Engineering	0.872	0.249
Psychiatry and Mental Health	0.972	0.064	Microbiology	0.921	0.252
Critical Care and Intensive Care Medicine	0.967	0.064	Computational Theory and Mathematics	0.861	0.281
Surfaces, Coatings and Films	0.957	0.065	Behavioral Neuroscience	0.821	0.304
Urology	0.960	0.065	Software	0.842	0.315
Applied Microbiology and Biotechnology	0.943	0.066	Library and Information Sciences	0.780	0.371
Anatomy	0.946	0.066	Molecular Biology	0.918	0.393
Aerospace Engineering	0.925	0.066	Sensory Systems	0.838	0.414
Neurology	0.942	0.066	Materials Science (miscellaneous)	0.848	0.429
Physiology (medical)	0.908	0.070	Theoretical Computer Science	0.800	0.437
Nephrology	0.963	0.071	Molecular Medicine	0.864	0.450
Colloid and Surface Chemistry	0.905	0.072	Medicine (miscellaneous)	0.867	0.552
Psychiatric Mental Health	0.626	0.073	Management Information Systems	0.819	0.585
Immunology and Microbiology (miscellaneous)	0.925	0.073	Endocrinology, Diabetes and Metabolism	0.887	0.602
Veterinary (miscellaneous)	0.921	0.074	Cancer Research	0.857	0.653
Biological Psychiatry	0.924	0.075	Media Technology	0.722	0.667
Building and Construction	0.877	0.077	Engineering (miscellaneous)	0.766	0.680
Safety Research	0.913	0.077	Biochemistry, Genetics and Molecular Biology (miscellaneous)	0.821	0.691
Transportation	0.875	0.077	Agricultural and Biological Sciences (miscellaneous)	0.811	0.700
Ophthalmology	0.955	0.078	Drug Discovery	0.764	0.711
Neuropsychology and Physiological Psychology	0.946	0.078	Catalysis	0.822	0.793
Developmental Neuroscience	0.914	0.079	Endocrinology	0.738	0.823
Hepatology	0.977	0.080	Physical and Theoretical Chemistry	0.764	0.882
Metals and Alloys	0.931	0.081	Cell Biology	0.861	0.983
Applied Mathematics	0.876	0.086	Genetics (clinical)	0.819	0.992
Cognitive Neuroscience	0.943	0.089	Polymers and Plastics	0.660	1.100
Energy Engineering and Power Technology	0.930	0.089	Physiology	0.740	1.371
Mathematics (miscellaneous)	0.901	0.089	Electrical and Electronic Engineering	0.656	1.612
Toxicology	0.928	0.092	Process Chemistry and Technology	0.700	1.702
Virology	0.907	0.092	Chemistry (miscellaneous)	0.716	1.760
Chemical Health and Safety	0.894	0.096	Immunology	0.723	2.114
Strategy and Management	0.938	0.100	Oncology	0.690	2.460
Global and Planetary Change	0.934	0.101	Physics and Astronomy (miscellaneous)	0.586	2.697
Rheumatology	0.953	0.101	Biochemistry	0.607	2.803
Surgery	0.938	0.102	Developmental Biology	0.624	3.799
Structural Biology	0.956	0.105	Multidisciplinary	0.622	4.325

Table 1 continued

SUBJECT CATEGORY	R ²	Diff ²	SUBJECT CATEGORY	R ²	Diff ²
Education	0.995	0.004	Ocean Engineering	0.976	0.020
Algebra and Number Theory	0.982	0.004	Health (social science)	0.977	0.020
History and Philosophy of Science	0.982	0.005	Physical Therapy, Sports Therapy and Rehabilitation	0.977	0.020
Religious Studies	0.974	0.005	Orthopedics and Sports Medicine	0.979	0.021
Architecture	0.942	0.005	Nuclear Energy and Engineering	0.963	0.021
Control and Optimization	0.965	0.006	Hematology	0.987	0.021
Environmental Chemistry	0.992	0.007	Safety, Risk, Reliability and Quality	0.981	0.021
Medical Laboratory Technology	0.980	0.007	Agronomy and Crop Science	0.971	0.021
Development	0.991	0.007	Biomaterials	0.968	0.022
Environmental Science (miscellaneous)	0.993	0.007	Civil and Structural Engineering	0.960	0.022
Social Sciences (miscellaneous)	0.989	0.007	Water Science and Technology	0.963	0.022
Health Policy	0.991	0.007	Economics and Econometrics	0.980	0.022
Law	0.989	0.008	Inorganic Chemistry	0.972	0.023
Paleontology	0.987	0.008	Emergency Medicine	0.960	0.023
Public Health, Environmental and Occupational Health	0.993	0.008	Ceramics and Composites	0.970	0.023
Sociology and Political Science	0.989	0.008	Mathematical Physics	0.926	0.024
Finance	0.990	0.009	Otorhinolaryngology	0.964	0.024
Economics, Econometrics and Finance (miscellaneous)	0.979	0.009	Social Psychology	0.961	0.025
Horticulture	0.987	0.009	Human Factors and Ergonomics	0.959	0.025
History	0.979	0.009	Mechanical Engineering	0.973	0.025
Industrial and Manufacturing Engineering	0.987	0.009	Demography	0.961	0.025
Geology	0.990	0.009	Nature and Landscape Conservation	0.976	0.026
Language and Linguistics	0.987	0.010	Radiological and Ultrasound Technology	0.977	0.026
Archeology	0.964	0.010	Public Administration	0.935	0.026
Political Science and International Relations	0.983	0.010	Community and Home Care	0.957	0.026
Developmental and Educational Psychology	0.988	0.011	Control and Systems Engineering	0.982	0.027
Astronomy and Astrophysics	0.991	0.011	Organic Chemistry	0.982	0.027
Tourism, Leisure and Hospitality Management	0.970	0.011	Radiation	0.984	0.027
Rehabilitation	0.984	0.012	Human-Computer Interaction	0.965	0.027
Geometry and Topology	0.973	0.012	Management, Monitoring, Policy and Law	0.969	0.027
Linguistics and Language	0.985	0.013	Philosophy	0.940	0.027
Computational Mechanics	0.963	0.013	Statistical and Nonlinear Physics	0.923	0.028
Signal Processing	0.980	0.014	Modeling and Simulation	0.964	0.028
Earth and Planetary Sciences (miscellaneous)	0.980	0.014	Computer Vision and Pattern Recognition	0.985	0.029
Health Informatics	0.981	0.014	Geriatrics and Gerontology	0.979	0.029
Analysis	0.939	0.015	Health Information Management	0.943	0.029
Forestry	0.979	0.015	Management Science and Operations Research	0.930	0.030
Urban Studies	0.965	0.015	Automotive Engineering	0.937	0.030
Nutrition and Dietetics	0.985	0.015	Logic	0.879	0.031
Statistics and Probability	0.980	0.015	Pharmacology (medical)	0.980	0.031
Clinical Psychology	0.986	0.015	Filtration and Separation	0.961	0.031
Pediatrics, Perinatology and Child Health	0.986	0.015	Economic Geology	0.968	0.032
Accounting	0.983	0.016	Acoustics and Ultrasonics	0.967	0.032
Applied Psychology	0.968	0.016	Condensed Matter Physics	0.974	0.032
Cultural Studies	0.958	0.016	Nuclear and High Energy Physics	0.966	0.032
Fluid Flow and Transfer Processes	0.979	0.016	Marketing	0.975	0.034
Pathology and Forensic Medicine	0.988	0.017	Parasitology	0.967	0.034
Psychology (miscellaneous)	0.985	0.017	Artificial Intelligence	0.976	0.034
Anthropology	0.971	0.017	Numerical Analysis	0.750	0.035
Chemical Engineering (miscellaneous)	0.979	0.017	Geochemistry and Petrology	0.952	0.035
Business and International Management	0.987	0.018	Epidemiology	0.977	0.035
Dermatology	0.987	0.018	Gender Studies	0.917	0.035
Organizational Behavior and Human Resource Management	0.975	0.018	Discrete Mathematics and Combinatorics	0.939	0.036
Health Professions (miscellaneous)	0.972	0.018	Communication	0.905	0.037
Geotechnical Engineering and Engineering Geology	0.970	0.018	Aquatic Science	0.944	0.037
Materials Chemistry	0.984	0.018	Pollution	0.960	0.038
Surfaces and Interfaces	0.988	0.018	Oceanography	0.963	0.038
Radiology, Nuclear Medicine and Imaging	0.987	0.018	Histology	0.935	0.038
Experimental and Cognitive Psychology	0.973	0.019	Earth-Surface Processes	0.960	0.039
Geography, Planning and Development	0.976	0.019	Information Systems and Management	0.904	0.039
Environmental Engineering	0.976	0.019	Computer Graphics and Computer-Aided Design	0.956	0.040
Mechanics of Materials	0.963	0.019	Decision Sciences (miscellaneous)	0.906	0.041
Insect Science	0.963	0.020	Pharmacy	0.969	0.041
Animal Science and Zoology	0.966	0.020	Atmospheric Science	0.956	0.042
Ecology	0.987	0.020	Food Science	0.970	0.042

Figure 9 shows the mean a coefficient corresponding to the slope of the logarithmic law. It was almost always greater in absolute value in Scopus-06, meaning that the Scopus specific subject areas generally have sharper impact distributions. This could be a reflection of the Matthew effect (Merton 1968) and because Scopus-06 includes a greater number number of journals in its database (and hence deals with more citations) and a wider citation window. The subject areas of Arts & Humanities, Nursing, and Immunology & Microbiology had similar average values of a in Scopus-06 and in ISI-97. The most significant differences were in Chemical Engineering, Computer Science, Energy, and Business, Management & Accounting perhaps because in these subject areas there are greater increases in the numbers of journals in Scopus-06 with respect to ISI-97.

Table 2 Correlations between ISI-97 and Scopus-06 for various indicators

	PEARSON
External Impact	0.97
JIF Average	0.88
E2	0.99
R2	0.85
a	0.9

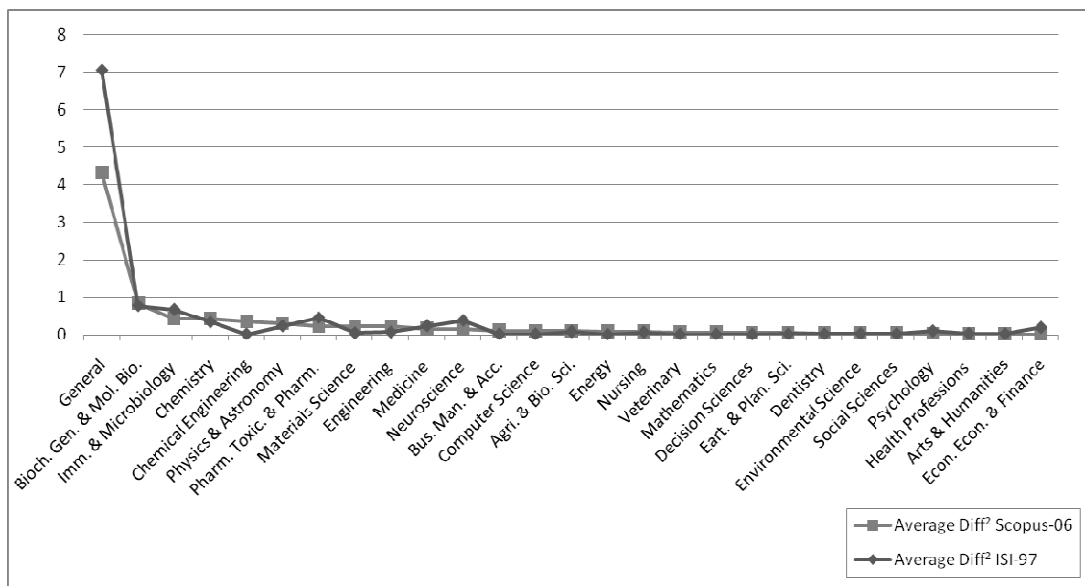


Figure 7. Average square deviation in Subject Areas in both ISI-97 and Scopus06.

Returning to the impact distributions in order to determine which factors influence how well a specific subject area fits the logarithmic approximation, we present in Table 3 the correlations between the different indicators studied, showing only those whose values were greater than 0.5. In this table, one observes that there was a strong relationship (correlation coefficient of 0.85) between the import and export of knowledge, i.e., that specific subject areas with a high import ratio usually also have a high export ratio. This was also observed in ISI-97.

In the next column, one sees that the Exterior JIF was strongly related with the Mean JIF, the JIF standard deviation, and the a and b coefficients. This indicates that, in a specific subject area, as the External JIF increases, so does the Mean JIF and its SD (as was to be expected) and the slope of the logarithmic approximation,

and the fit becomes poorer with the distributions having a sharper descent and longer tails (very similar to ISI-97).

The Internal JIF has some correlation (correlation coefficients 0.54 and 0.55, respectively) with the Export and Import Ratios (this was not noticeable in ISI-97). Relative to the External JIF, the correlation of the Internal JIF with both the Mean JIF and its SD was weaker, and it was also weaker (but still fairly strong) with the two coefficients of the logarithmic approximation (similar to ISI-97).

The absolute SD of the JIF was strongly correlated with many indices. However, its value relative to the Mean JIF was correlated only with the coefficients of the logarithmic approximation, with the squared deviation (indicating that the greater the deviation with respect to the mean, the more poorly the distribution approximates the logarithmic), and negatively with the R² coefficient (similar to ISI-97).

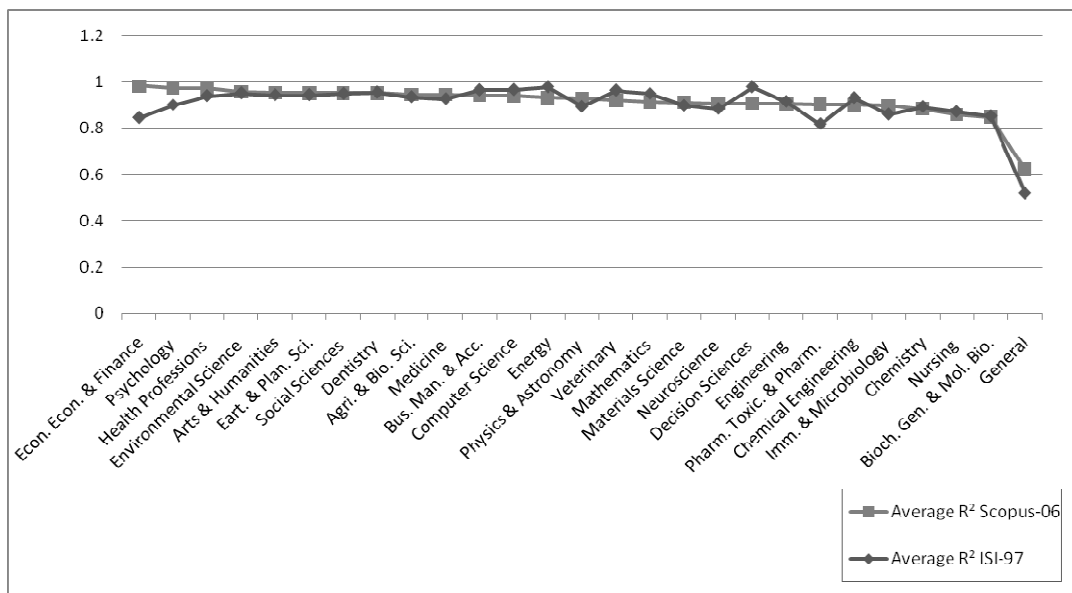


Figure 8. Average R^2 in Subject Areas in both ISI-97 and Scopus06.

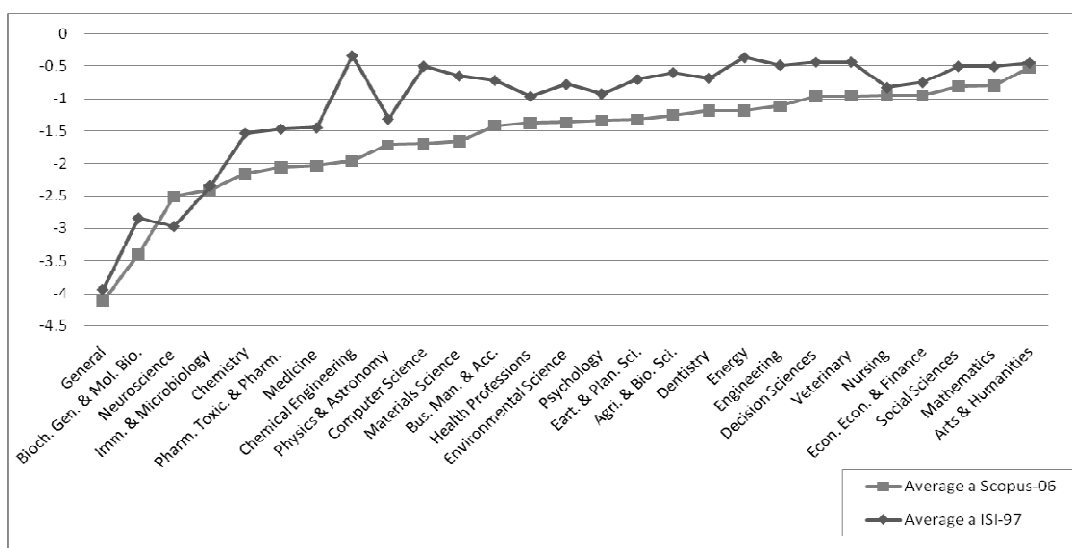


Figure 9. Average a coefficient in Subject Areas in both ISI-97 and Scopus06.

The coefficients of the logarithmic approximation were strongly correlated with most of the indices as well as with each other (a correlation coefficient of 0.92), the latter indicating that the height of the peak and the slope of the approximation usually increase together (similar to ISI-97).

The R^2 coefficient had strong correlations with the squared deviation and, curiously, with the SD, as we remarked above—in both cases negative (similar to ISI-97).

The squared deviation (normalized to the Mean JIF), which is an indicator of a failure to fit the logarithmic approximation, was strongly correlated with the SD of the JIF, as is logical since the greater the dispersion of the JIFs around the mean, the greater also will be the deviations from the logarithmic approximation.

It was also strongly correlated with the coefficients of the logarithmic approximation, indicating that the fits were poorer in distributions with steeper slopes. It had a negative correlation with R^2 (similar to ISI-97).

Table 3. Coefficients of correlation between the different indices calculated.

	%Export	%Import	External Impact	R ²	a	b	JIF Average	S.D. JIF	S.D. JIF/JIF Av.	Internal Impact	Diff ²
Total Cites						0.65		0.53			
Self-Cites						0.57				0.57	
Own-Cites						0.58				0.56	
Citables											
References						0.67		0.54			
%Export	1.00	0.85								-0.54	
%Import	0.85	1.00								-0.55	
External Impact			1.00		-0.77	0.68	0.78	0.72			
Internal Impact	-0.54	-0.55			-0.56	0.62	0.56	0.57		1.00	
a			-0.77		1.00	-0.92	-0.87	-0.97		-0.56	-0.58
b			0.68		-0.92	1.00	0.83	0.95	0.50	0.62	0.55
JIF Average			0.78		-0.87	0.83	1.00	0.81		0.56	
R ²				1.00				-0.58	-0.62		-0.83
R				0.99				-0.58	-0.62		-0.83
Journals											
S.D. JIF			0.72	-0.58	-0.97	0.95	0.81	1.00	0.59	0.56	0.69
S.D. JIF/JIF Av.				-0.62		0.50		0.59	1.00		0.72
Diff ²				-0.83	-0.57	0.55		0.69	0.72		1.00

Conclusions

The first conclusion is that, as in ISI-97, observing the squared deviation Diff2, one can state that the rank/JIF distributions in most of the specific subject areas fit a negative logarithmic law fairly well. Those that fit less well have a sharper descent than exponential and a longer tail, similar to ISI-97. No S-shaped distribution was observed in any of the specific subject areas.

This means that the JIFs distributions are highly skewed, that is in all categories there are a few journals that accumulate the majority of the category's JIF.

The values of R2 for most of the specific subject areas were close to unity. This was similar in ISI-97.

The a and b logarithmic approximation coefficients were greater in (absolute) value than those of ISI-97, meaning that these distributions fit the logarithmic approximation with logarithmic laws that are sharper than those of ISI-97. This could be due to the greater citation volume included in Scopus-06 both because the database is larger and because the citation window considered was three years instead of two.

There was little difference between ISI-97 and Scopus-06 in how close the categories approximated a logarithmic law. The subject areas belonging to Social Sciences presented the closest fits in both studies. Those with the poorest fits belonged to Health Sciences in ISI-97 and to Life Sciences (and some also to Physics) in Scopus-06.

Apart from General (the Scopus subject area that includes the multidisciplinary journals) which had a lower value in Scopus-06, the subject areas with the highest values of the squared deviation were Immunology & Microbiology, Chemistry, and Biochemistry, Genetics & Molecular Biology. Those with the lowest values were Economics, Econometrics & Finance, Arts & Humanities, Health Professions, and Psychology.

Similarly, apart from General which had a higher value in Scopus-06, the subject areas with the lowest values of R^2 were Nursing, Chemistry, Immunology & Microbiology, and Biochemistry, Genetics & Molecular Biology. Those with the highest values were Economics, Econometrics & Finance, Psychology, and Health Professions.

Apart from General, the subject areas with the greatest values of the slope were Neuroscience, Immunology & Microbiology, Chemistry, and Biochemistry, Genetics & Molecular Biology. Those with the lowest values were Arts & Humanities, Mathematics, Social Sciences, and Economics, Econometrics & Finance.

The comparison by subject area of the two experiments failed to reveal any significant differences. Perhaps the largest difference was the increased slope of the logarithmic approximations. The rest of the differences could be due to the distinct form of organizing the two databases.

The correlation analysis of the various indicators showed smaller correlations between them than in ISI-97. However, the dependencies were similar. The most important was the correlation of the external JIF with the slope of the logarithmic approximation and the squared deviation.

This means that there are specific subject areas whose distribution is sharper even than that foreseen by the logarithmic approximation giving a high square deviation and a small R^2 , which is usually accompanied by or even a consequence of the impact external to the specific subject area. Evidently for any subject area it is desirable to capture the attention of other subject areas. This also may be a consequence of the artificial division of the science.

For instance, extensive disciplines are usually subdivided to obtain subject areas of a similar size which imply high citation traffic between them.

In sum therefore, the results are not dependent on the database or the time parameters used, since the categories of ISI-97 that are visible from the outside also showed this property in Scopus-06 in spite of the differences in coverage, in dates, in citation window, in structure, and in that structure's form of classifying journals.

The results have thus reconfirmed the appropriateness of the Iceberg metaphor, according to which specific subject areas of science can be seen as icebergs, some of whose tips above the surface are high enough to be sighted in the distance from some other iceberg (specific subject area), while in others the emerged part hardly exists at all, and the iceberg (the specific subject area) is only visible when one is actually on it (i.e., from within that same specific subject area).

Acknowledgments This work was financed by the Junta de Extremadura—Consejería de Educación Ciencia & Tecnología and the Fondo Social Europeo as part of research project PRI06A200, and by the Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica 2008-2011 and the Fondo Europeo de Desarrollo Regional (FEDER) as part of research projects TIN2008-06514-C02-01 and TIN2008-06514-C02-02.

References

Egghe, L. (2009). Mathematical derivation of the impact factor distribution. *Journal of Informetrics*, 3(4), 290–295.

Garfield, E. (1979). *Citation indexing. Its theory and application in science, technology and humanities*. New York: Wiley.

Glänzel, W., & Moed, H. (2002). Journal impact measures in bibliometric research. *Scientometrics*, 53(2), 171–193.

Glänzel, W., & Schubert, A. (2003). A new classification scheme of science fields and subfields designed for scientometric evaluation purposes. *Scientometrics*, 56(3), 257–367.

Guerrero-Bote, V. P., Zapico-Alonso, F., Espinosa-Calvo, M. E., Gómez-Crisóstomo, R., & Moya-Anegón, F. (2006). Binary pathfinder: An improvement to the Pathfinder algorithm. *Information Processing and Management*, 42(6), 1484–1490.

Guerrero-Bote, V. P., Zapico-Alonso, F., Espinosa-Calvo, M. E., Gómez-Crisóstomo, R., & Moya-Anegón, F. (2007). The iceberg hypothesis: import-export of knowledge between scientific subject categories. *Scientometrics*, 71(3), 423–441.

Lancho-Barrantes, B. S., Guerrero-Bote, V. P., & Moya-Anegón, F. (in press). What lies behind the averages and significance of citation indicators in different disciplines? *Journal of Information Science*.

Leydesdorff, L., & Betsman, S. (2006). Classification and powerlaws: The logarithmic transformation. *Journal of the American Society for Information Science and Technology*, 57(11), 1470–1486.

Lundberg, J. (2007). Lifting the crown—citation z-score. *Journal of Informetrics*, 1, 145–154.

Merton, R. K. (1968). The Matthew effect in science. *Science*, 159, 56–63.

Moed, H. F. (2005). *Citation analysis in research evaluation* (p. 346). Dordrecht: Springer.

Moya-Anegón, F., Chinchilla-Rodríguez, Z., Corera-Álvarez, E., González-Molina, A., Hassan-Montero, Y., & Vargas-Quesada, B. (2008). Indicadores bibliométricos de la actividad científica española (2002–2006). Madrid: FECYT.

Moya-Anegón, F., Chinchilla-Rodríguez, Z., Vargas-Quesada, B., Corera-Álvarez, E., Muñoz-Fernández, F. J., González-Molina, A., et al. (2007). Coverage analysis of Scopus: a journal metric approach. *Scientometrics*, 73(1), 53–78.

Moya-Anegón, F., Vargas-Quesada, B., Herrero-Solana, V., Chinchilla-Rodríguez, Z., Corera-Álvarez, E., & Muñoz-Fernández, F. J. (2004). A new technique for building maps of large scientific domains based on the cocitation of classes and categories. *Scientometrics*, 61, 129–145.

Pinski, G., & Narin, F. (1976). Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics. *Information Processing and Management*, 12(5), 297–312.

Podlubny, I. (2005). Comparison of scientific impact expressed by the number of citations in different fields of science. *Scientometrics*, 64(1), 95–99.

Pudovkin, A., & Garfield, E. (2002). Algorithmic procedure for finding semantically related journals. *Journal of the American Society for Information Science and Technology*, 53(13), 1113–1119.

Sancho, R. (1990). Indicadores bibliométricos utilizados en la evaluación de la ciencia y la tecnología. Revisión bibliográfica. *Revista Española de Documentación Científica*.13:842-865.

Schvaneveldt, R. W., Dearholt, D. W., & Durso, F. T. (1988). Graph theoretic foundations of Pathfinder networks. *Computers and Mathematics with Applications*, 15(4), 337–345.

SCImago Journal & Country Rank. (2008). SCImago Research Group. Accessed 26 March, 2009 from <http://www.scimagojr.com/>.

Sombatsompop, N., & Markpin, T. (2005). Making an equality of ISI impact factors for different subject fields. *Journal of the American Society for Information Science and Technology*, 56(7), 676–683.

Van Leeuwen, T. N., & Moed, H. F. (2002). Development and application of journal impact measures in the Dutch science system. *Scientometrics*, 53(2), 249–266.

Van Leeuwen, T. N., Visser, M. S., Moed, H. F., Nederhof, T. J., & Van Raan, A. F. J. (2003). The Holy Grail of science policy: Exploring and combining bibliometric tools in search of scientific excellence. *Scientometrics*, 57(2), 257–280.

Waltman, L., & Van Eck, N. J. (2009). Some comments on Egghe's derivation of the impact factor distribution. *Journal of Informetrics*, 3(4), 363–366.

Wormell, I. (1998). Informetric analysis of the international impact of scientific journals: How 'International' are the international journals? *Journal of Documentation*, 54(5), 584–605.

ARTÍCULO

A2: What lies behind the averages and significance of citation indicators in different disciplines?

Bárbara S. Lancho Barrantes, Vicente P. Guerrero Bote and Félix de Moya Anegón

In press in:

Journal of Information Science June 2010 vol. 36 no. 3 371-382

What lies behind the averages and significance of citation indicators in different disciplines?

Bárbara S. Lancho-Barrantes and Vicente P. Guerrero-Bote

Universidad de Extremadura, Grupo SCImago, Spain

Félix Moya-Anegón

CSIC, CCHS, IPP, Grupo SCImago, Spain

Abstract

The limitations of citation-based indicators include a lack of coverage, no normalization with respect to the length of reference lists (with a potential bias in favour of reviews), and different citation habits. As a consequence, the distributions of the indicators are not comparable across different disciplines. Here we show that the most popular journal citation indicators used in quality assessment – the journal impact factors of Thomson Scientific and the scientific journal rankings of Scopus – are strongly correlated with the proportion of within-database references, and even more so with the number of within-database recent references per paper. No significant correlations were found with other bibliometric magnitudes. We anticipate that these results will be a starting point for more sophisticated indicator models that take this dependence

into account, and for the design of strategies aimed at extending such bibliometric databases as Thomson Scientific's Science Citation Index or Elsevier's Scopus to improve their capacity to evaluate all sciences.

Keywords: Bibliometrics; Citation Indicators; Scientific Indicators

1. Introduction

Scientific evaluations ideally ought to take into account all the results obtained in research and all the means chosen for their publication – journal articles, conference papers, reviews, patents, reports, monographs, etc. Evaluations of large volumes of research, however, still only consider the scholarly communications included in the major databases, whether multidisciplinary or specialized, national or international. The reason is that these databases permit the automated computerized handling of massive quantities of data.

The bibliometric indicators used to evaluate research take qualitative as well as quantitative aspects into account: they use the citations only of papers included in the main databases (principally Thomson Reuters' Web of Science and Elsevier's Scopus) and are based on the idea that, in spite of different motivations [1], citations constitute recognition of foregoing work [2]. However, doubts have often been expressed about the applicability to certain disciplines of these bibliometric

indicators and international databases. While the bibliometric indicators built from these databases are normally effective in basic science contexts in which research is disseminated mainly via scientific journals [3], other research fields have varying yearly average citation rates [4], and the values of the indicators are almost always lower in the areas of engineering, social sciences, and humanities [5].

Various causes have been posited for the lack of overall statistical significance and normalization observed for these citation-based indicators: the paucity of database coverage in certain areas, not only in journals but more importantly in other types of documents [6–8], and the distinct habits or patterns of citation in different areas of science [9–21].

Firstly, with respect to the coverage, Braun, Glänzel and Schubert [6] compared the Web of Science coverage with that of Ulrich's International Periodicals Directory, universally accepted as the database which contains most of the world's journals. Based on the methodological approach of that work, Moya-Anegón et al. [8] analyzed the Scopus coverage with respect to the Ulrich Core. Both studies concluded that the two databases indeed have a statistically balanced coverage in terms of subjects, countries, languages and publishers, except for the two areas of social sciences and humanities.

Secondly, with respect to the effect of differing citation habits, it appears that these contribute to the low values of the bibliometric indicators in social sciences and humanities. Citations are of course of fundamental importance in all scholarly communication, since they show how a scientific paper is founded on foregoing work. But each scientific area has its own citation habits and in these two areas in particular there is less use of scientific journals as the vehicle for publication than in other areas, but more reliance on monographs, which naturally are more cited [9–20, 22, 23]. The problem then is that the two major databases (Scopus and Web of Science) only include journal material, leaving aside monographs or reports despite their importance in the communication of research results in disciplines such as social sciences and humanities (monographs) and engineering and computer science (conference papers).

In the computation of citation-based bibliometric indicators, not all references are used, only those to papers included in the corresponding database. Our hypothesis is that the annual profile of references to documents included in the database will have a major effect on the statistical significance and the average value of those indicators in different areas of science.

Recently, Althouse et al. [24] used multiple regressions between sets of predictor variables and the Web of Science database to explore the reasons for impact factor differences across fields. They concluded that the fraction of citations to Journal Citation Report (JCR)-indexed literature is the greatest contributor to these differences, with secondary contributions from differences in the number of

citations per paper and in the fraction of recent references. They defined a weighted journal impact factor (JIF) as the total of a journal's citations divided by the total number of papers in the corresponding two-year citation window.

The present work concentrates on journal metric indicators, looking at the profile of references to indexed documents from the papers published in a given year to determine the influence on the citation-based journal metric indicators of the rate of citation to papers included in the database and/or to recent papers. The indicators used for this purpose were the JIF with different citation windows, the SJR (SCImago Journal Rank) [25], and Scopus. The results showed that the habits of citation concentration explain the differences between fields in the statistical significance and average values of these citation-based bibliometric indicators. This finding will also be of value for designers of database expansion strategies aimed at improving the capacity to evaluate all of science.

2. Data and method

The original source of the data used to calculate the indicators and perform the rest of the study was Scopus. This database, which is larger than the Web of Science [8, 26], was selected as representing the composition of world science on a large scale.

Most data were retrieved from the website SCImago Journal and Country Rank (www.scimagojr.com) maintained by the SCImago Research Group. The data were downloaded in April 2009 (limiting the citations to those included in works published in 2007). This data were stored in a relational database integrating all the information.

The main part of this study used only the 1,645,036 Scopus database documents published in 2007 (we used all the document types so as to avoid any bias, although the primary production – articles, reviews and conference papers – accounted for 90%). The data corresponding to all of their references were retrieved, both to documents present in the Scopus Source Set (SSS) database in previous years (5,842,139) and to those outside the SSS (41,838,425). For the purposes of the present work, we define the SSS as the set of the more than 21,000 source journals of Scopus with a temporal range from 1996 to 2007. When the cited reference is present in the SSS, Scopus identifies it unequivocally, so that both classes of data can be retrieved without error.

These documents are classified by category and area. There are 295 ‘specific subject areas’ organized into 26 ‘subject areas’, plus a ‘general subject area’ containing multidisciplinary journals such as *Nature* or *Science*. These subjects are grouped into four categories on the Scopus Basic Search page (see the Scopus website www.scopus.com, accessed 7 August 2009). The four Scopus categories are:

- Life sciences: agricultural and biological sciences; biochemistry, genetics and molecular biology; immunology and microbiology; neuroscience; pharmacology, toxicology and pharmaceuticals.
- Physical sciences: chemical engineering; chemistry; computer science; earth and planetary science; energy; engineering; environmental science; materials science; mathematics; physics and astronomy.
- Social sciences: arts and humanities; business, management and accounting; decision sciences; economics, econometrics and finance; psychology; social sciences.
- Health sciences: medicine; nursing; veterinary; dentistry; health professions.

Table 1 lists the number of documents in the SSS by area, the corresponding total number of references, and the percentages of papers and references (the two percentages sum to more than 100 due to the overlap of the subject areas).

By way of summary, Table 2 presents the percentages of papers and references of the four Scopus categories.

The citation quality indicators considered were:

- SJR (SCImago Journal Rank) indicator [25]: this expresses a weighted average number of citations received in a year by the documents published in the selected journal in the three previous years. The citations are weighted with the prestige or status of the citing journal using a technique similar to Google's PageRank.
- JIF4Y: this represents the average number of citations given in 2007 to those papers in a journal that were published during the four previous years.
- JIF3Y: this represents the average number of citations given in 2007 to those papers in a journal that were published during the three previous years.
- JIF2Y: this represents the average number of citations given in 2007 to those papers in a journal that were published during the two previous years..

Table 1

The areas of the Scopus Source Set with the corresponding number of papers published in 2007, their references, and their percentages

Subject areas	Papers	%Papers	References	%References
General	17,160	1.04	376,281	0.89
Agricultural and biological sciences	128,908	7.80	4,335,268	10.24
Arts and humanities	10,790	0.65	465,140	1.10
Biochemistry, genetics and molecular biology	193,592	11.72	7,074,116	16.71
Business, management and accounting	37,243	2.25	784,242	1.85
Chemical engineering	69,096	4.18	1,341,503	3.17
Chemistry	119,844	7.25	3,986,474	9.42
Computer science	95,537	5.78	1,786,991	4.22
Decision sciences	9240	0.56	216,091	0.51
Earth and planetary sciences	68,961	4.17	2,256,517	5.33
Economics, econometrics and finance	20,648	1.25	545,098	1.29
Energy	26,217	1.59	361,027	0.85
Engineering	223,074	13.50	3,165,209	7.48
Environmental science	66,135	4.00	2,176,535	5.14
Immunology and microbiology	55,202	3.34	1,873,562	4.43
Materials science	108,355	6.56	2,262,981	5.35
Mathematics	80,921	4.90	1,652,356	3.90
Medicine	485,158	29.37	11,207,833	26.48
Neuroscience	28,608	1.73	1,273,218	3.01
Nursing	25,825	1.56	452,156	1.07
Pharmacology, toxicology and pharmaceuticals	42,855	2.59	1,347,212	3.18
Physics and astronomy	134,915	8.17	3,219,210	7.61
Psychology	22,248	1.35	871,741	2.06
Social sciences	72,362	4.38	2,541,277	6.00
Veterinary	16,748	1.01	388,565	0.92
Dentistry	7724	0.47	183,777	0.43
Health professions	10,458	0.63	246,611	0.58

Table 2

Summary of percentages of papers and references per Scopus category

Scopus category	% Papers	% References
Health	31.34	28.47
Life	24.25	33.42
Physical	46.52	40.61
Social	9.12	11.24

We computed the Pearson correlation coefficients of the above citation quality indicators with the following bibliometric magnitudes:

- References/Article: average number of references per article.
- References Window 3Y/Article: average number of references to papers published in the three-year citation window (2004, 2005, 2006) per article.
- References Window 2Y/Article: average number of references to papers published in the two-year citation window (2005, 2006) per article.
- References/Article in SSS: average number of references to SSS-indexed papers per article.
- References Window 3Y/Article in SSS: average number of references to SSS-indexed papers published in the three-year citation window (2004, 2005, 2006) per article.

Table 3

Average references per article: in total and considering only the Scopus Source Set (SSS), and for two- (2Y) and three-year (3Y) citation windows

Subject areas	Ref/art	Ref Win 3Y/art	Ref Win 2Y/art	Ref/art in SSS	Ref Win 3Y/art in SSS	Ref Win 2Y/art in SSS
General	21.9	6.1	4.1	13.0	5.1	3.5
Agricultural and biological sciences	33.6	6.3	3.9	15.3	4.8	3.1
Arts and humanities	43.1	4.4	2.7	4.3	1.2	0.7
Biochemistry, genetics and molecular biology	36.5	9.8	6.5	24.8	9.0	5.9
Business, management and accounting	21.1	3.6	2.2	5.9	1.5	0.9
Chemical engineering	19.4	4.3	2.8	9.9	3.4	2.2
Chemistry	33.3	8.3	5.5	18.6	6.9	4.6
Computer science	18.7	4.8	3.1	6.5	2.3	1.4
Decision sciences	23.4	3.9	2.3	7.6	2.0	1.2
Earth and planetary sciences	32.7	6.4	4.1	13.8	4.5	2.9
Economics, econometrics and finance	26.4	4.8	3.0	7.8	2.0	1.2
Energy	13.8	3.1	2.0	4.9	1.7	1.1
Engineering	14.2	3.3	2.1	5.7	2.0	1.2
Environmental science	32.9	6.6	4.1	14.7	4.6	2.9
Immunology and microbiology	33.9	9.3	6.2	23.3	8.5	5.6
Materials science	20.9	4.9	3.1	11.0	3.8	2.3
Mathematics	20.4	4.4	2.9	6.8	2.3	1.4
Medicine	23.1	5.9	3.8	13.9	4.9	3.2
Neuroscience	44.5	10.7	6.9	29.7	9.9	6.4
Nursing	17.5	4.7	3.1	9.3	3.3	2.1
Pharmacology, toxicology and pharmaceutics	31.4	8.4	5.6	20.3	7.2	4.8
Physics and astronomy	23.9	6.0	4.1	11.1	3.7	2.3
Psychology	39.2	6.3	3.8	15.2	4.2	2.5
Social sciences	35.1	6.4	4.1	6.0	1.8	1.1
Veterinary	23.2	4.2	2.6	10.6	3.2	1.9
Dentistry	23.8	4.6	2.8	13.3	4.0	2.4
Health professions	23.6	6.0	3.9	13.5	4.7	3.0

- References Window 2Y/Article in SSS: average number of references to SSS-indexed papers published in the two year citation window (2005, 2006) per article. Rate in SSS: percentage of references to SSS-indexed papers.
- Rate in 3Y Window: percentage of references to papers published in the three-year citation window (2004, 2005, 2006).
- Rate in 2Y Window: percentage of references to papers published in the two-year citation window (2005, 2006).
- Rate in 3Y Window in SSS: percentage of references to SSS-indexed papers published in the three-year citation window (2004, 2005, 2006).
- Rate in 2Y Window in SSS: percentage of references to SSS-indexed papers published in the two-year citation window (2005, 2006).

3. Results

Table 3 lists the values of the bibliometric magnitudes corresponding to absolute average numbers of references. One observes by simple inspection that there is a marked variation in the numbers for three basic factors:

1. The average number of references per article in each area (References/Article in column 1);
2. The average number of recent references per article (References Window 3Y/Article and References Window 2Y/Article in columns 2 and 3, respectively);
3. The average number of SSS-indexed references (References/Article in SSS in column 4).

The two magnitudes corresponding to the last two columns take these last two factors into account conjointly.

For the first factor – the average number of references per article – neuroscience (44.5) and arts and humanities (43.1) are the areas with the highest values. Following them are psychology, biochemistry, genetics and molecular biology, and social sciences. The areas with the lowest values are energy, engineering, nursing, and computer science, most of which belong to the Scopus physical sciences category. The highest value is more than three times greater than the lowest.

For the second factor – the averages of recent references (two- and three-year windows) – the areas with the highest values are pharmacology, toxicology and pharmaceuticals, immunology and microbiology, biochemistry, genetics and molecular biology, and neuroscience. Those with the lowest values are energy, engineering, business, management and accounting, and decision sciences. The highest value is more than three times greater than the lowest.

For the third factor – the average number of SSS-indexed references per article – the area with the highest value is neuroscience (29.7) which was also the area with the highest average number of total references per article. Following this are biochemistry, genetics and molecular biology, immunology and microbiology, pharmacology, toxicology and pharmaceuticals. The areas with the lowest values are arts and humanities, business, management and accounting, social sciences, and some of the areas included in the Scopus physical sciences category (energy,

engineering, and computer science). The highest value is more than six times greater than the lowest.

The first column differs from the second and the third (correlation coefficient = 0.70) and also from the fourth (correlation coefficient = 0.60), while the third and fourth columns are strongly correlated (correlation coefficient = 0.98). This situation is exemplified by the subject area of arts and humanities which is ranked top in the first column, lower in the second and third columns, and bottom in the fourth column. The case is similar with social sciences.

Lastly, in the last two columns corresponding to bibliometric magnitudes which include the effects of both factors – references to articles which are both recent and SSS-indexed – one observes that neuroscience (9.9 and 6.4), biochemistry, genetics and molecular biology, immunology and microbiology, and pharmacology, toxicology and pharmaceuticals, all belonging to the Scopus life sciences category, have the highest values, while arts and humanities (1.2 and 0.7), business, management and accounting, energy, and social sciences have the lowest values. The highest value is approximately nine times greater than the lowest in both citation windows. This greater difference is logical since the two factors are considered together.

In Table 4, only the effects of the last two factors can be observed since the values are relative to the average number of references per article. The first column corresponds to the factor of citing SSS-indexed papers, the following two columns to the factor of recent referencing (two- and three-year windows), and again the

last two columns correspond to bibliometric magnitudes which take both these factors into account conjointly.

With respect to the percentages of references to SSS-indexed papers, the highest values are for areas belonging to the Scopus life sciences and health sciences categories: immunology and microbiology, biochemistry, genetics and molecular biology, and pharmacology, toxicology and pharmaceuticals. The lowest values are for areas belonging to the Scopus social sciences category: arts and humanities, business, management and accounting, economics, econometrics and finance, and decision sciences. The highest value is approximately seven times greater than the lowest.

With respect to the percentages of references to recent papers (two or three years in age), the rank order is similar to, although not the same as, the first column (correlation coefficients of 0.78 and 0.76 with the Rate in 3Y Window and the Rate in 2Y Window, respectively).

Table 4

Percentages of references that are in the Scopus Source Set (SSS), and of recent references (in articles published in 2007) ordered by rate in two- (2Y) and three-year window SSS

Subject areas	Rate in SSS	Rate in 3Y Window	Rate in 2Y Window	Rate in 3Y Window SSS	Rate in 2Y Window SSS
Immunology and microbiology	68.6	27.5	18.2	25.0	16.4
Biochemistry, genetics and molecular biology	67.8	26.8	17.7	24.6	16.1
General	59.3	27.6	18.9	23.3	15.8
Pharmacology, toxicology and pharmaceuticals	64.5	26.7	17.8	23.1	15.1
Neuroscience	66.7	23.9	15.6	22.3	14.4
Medicine	60.2	25.3	16.6	21.4	13.9
Chemistry	55.9	25.0	16.6	20.8	13.7
Health professions	57.1	25.5	16.6	20.1	12.8
Nursing	53.3	26.9	17.6	18.7	11.9
Materials science	52.4	23.6	15.0	18.0	11.1
Chemical engineering	51.2	22.2	14.3	17.8	11.3
Dentistry	55.7	19.2	11.8	16.7	10.1
Physics and astronomy	46.5	25.0	17.1	15.5	9.5
Agricultural and biological sciences	45.6	18.6	11.7	14.4	9.1
Environmental science	44.5	20.0	12.6	14.1	8.8
Engineering	40.0	23.4	15.0	14.0	8.7
Earth and planetary sciences	42.1	19.4	12.5	13.7	8.8
Veterinary	45.6	18.0	11.2	13.6	8.3
Energy	35.3	22.8	14.6	12.6	7.9
Computer science	34.6	25.4	16.4	12.2	7.6
Mathematics	33.2	21.3	14.1	11.1	7.0
Psychology	38.9	16.2	9.8	10.7	6.4
Decision sciences	32.4	16.6	9.8	8.7	5.0
Economics, econometrics and finance	29.4	18.3	11.3	7.5	4.4
Business, management and accounting	28.0	17.3	10.5	7.2	4.3
Social sciences	17.0	18.1	11.8	5.1	3.1
Arts and humanities	10.0	10.2	6.3	2.9	1.7

The highest percentages, apart from the special multidisciplinary area general, are for areas of the life sciences: immunology and microbiology, pharmacology, toxicology and pharmaceuticals, biochemistry, genetics and molecular biology, and nursing. The lowest are for arts and humanities, psychology, decision sciences,

and business, management and accounting. The highest value is three times greater than the lowest.

The last two columns of Table 4 correspond to the effect of including the two factors together – the Rate in 3Y Window in SSS and the Rate in 2Y Window in SSS. They show a similar rank order as the first column, but now, as the two factors are combined, the differences are enhanced, the highest value being almost 10 times greater than the lowest. In particular, the percentages for immunology and microbiology are 25.0 and 16.4, respectively, and for arts and humanities 2.9 and 1.7, respectively.

Figure 1 shows, by area and for papers published in 2007, the percentages of references to each year relative to the total number of references. The accumulated percentages of the 12 years included in the plots are given in parentheses in the legend.

The 12-year accumulated percentages of nursing (75.0%) and immunology and microbiology (74.8%) (the two highest) are more than twice that of arts and humanities (36.5% – the lowest). While the culture in arts and humanities is one of citing original sources and classical books which will normally be older than the temporal range of the figure, the difference relative to the rest of the areas is still surprisingly great. Above arts and humanities are social sciences (55.2%), psychology (59.5%), earth and planetary sciences (60.0%), and other areas belonging to the Scopus physical sciences and social sciences categories. The

highest percentages are of areas in the Scopus life sciences and health sciences categories: nursing (75.0%), immunology and microbiology (74.8%), biochemistry, genetics and molecular biology (74%), pharmacology, toxicology and pharmaceuticals (74%), health professions (72%). This reflects their tendency to cite recent articles as they are fields with rapid obsolescence.

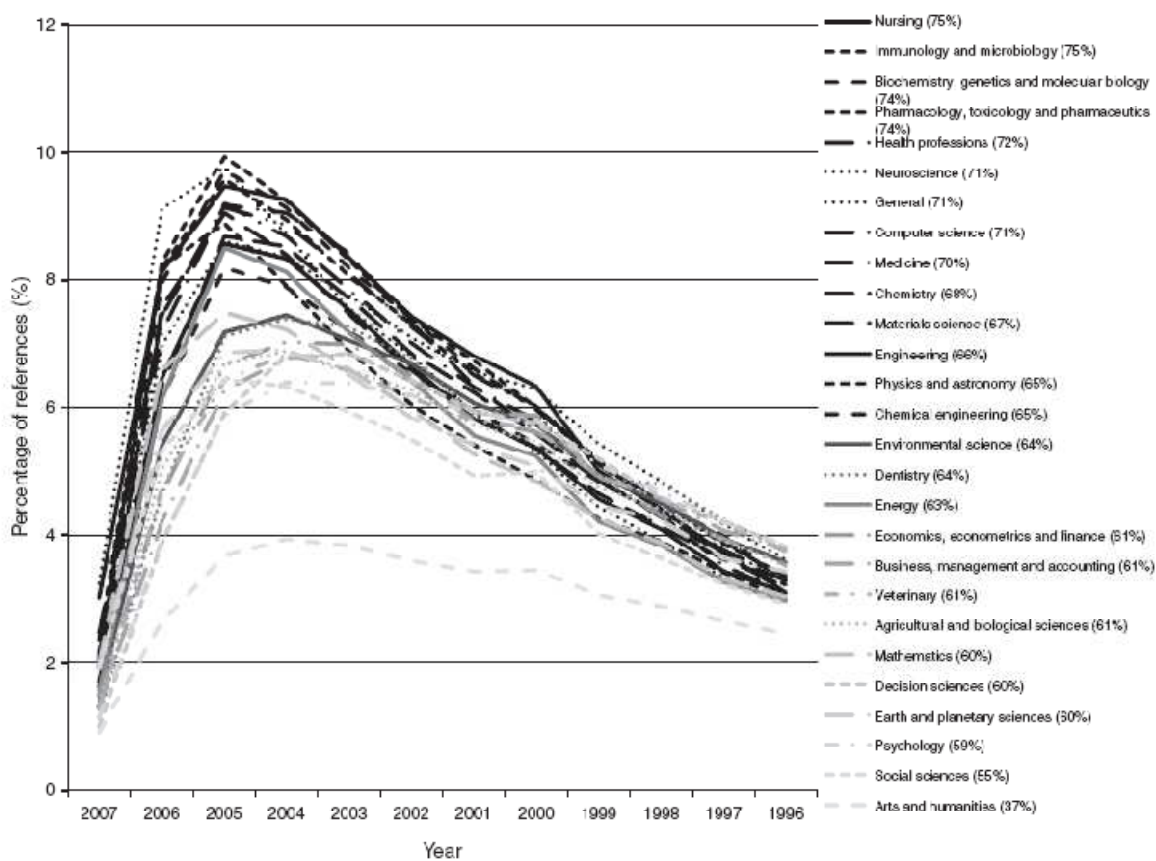


Fig. 1. Percentages of references to each year relative to the total number of references for 2007 papers. In parentheses in the legend, the sum of these percentages.

Figure 2 shows the percentages by area of references to SSS-indexed papers. The values are of course lower because SSS does not cover all references. The areas with the lowest values are arts and humanities, social sciences, business, management and accounting, and economics, econometrics and finance. Two areas belonging to the Scopus social sciences category – business, management and accounting, and economics, econometrics and finance – had a higher rank in Figure 1. The highest values correspond to immunology and microbiology, biochemistry, genetics and molecular biology, and neuroscience, all belonging to the Scopus life sciences and health sciences categories. While the immunology and microbiology total percentage in Figure 2 is nearly seven times greater than that of arts and humanities, one observes that the curve of the latter is notably less distant from other areas than in Figure 1, and that the different areas are more evenly spread out.

There is a clear coherence between the patterns that emerged in Table 4 and Figures 1 and 2, with the Scopus social sciences category areas having the lowest values, and the areas of life sciences and health sciences the highest. They are likewise coherent with the results of other studies performed on the Institute for Scientific Information (ISI) databases [23], including in the overlap zone which involves such fields as mathematics, technology-oriented science, and some social science areas.

The second percentage relative to the references of this period (1996–2007) that is included in the legend of Figure 2 is a reflection of how thoroughly Scopus covers the citations of the different areas.

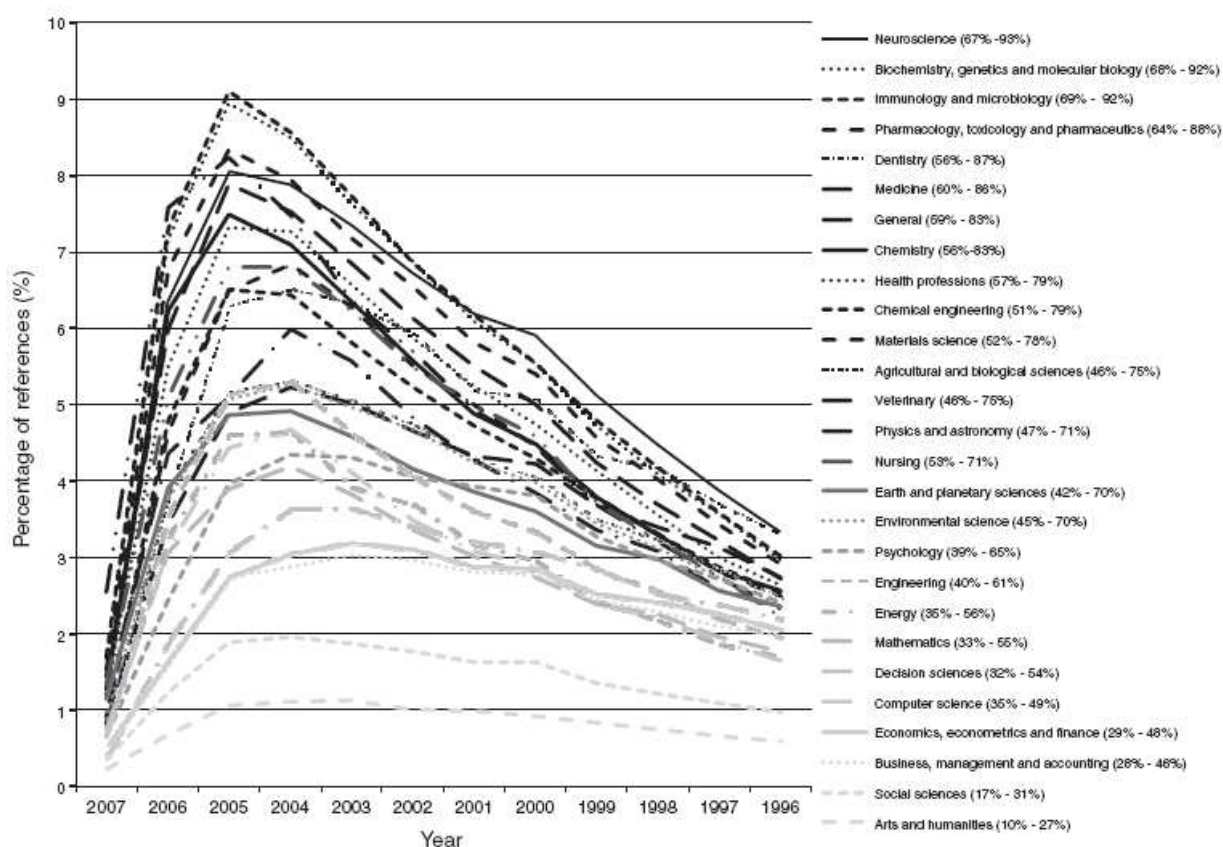


Fig. 2. As in Figure 1, but now percentage of references to articles in the Scopus Source Set for each of the last 12 years relative to the total references in articles published in 2007. In parentheses in the legend, the sum of these percentages, and the percentage this sum represents of all the references to this period.

The values of this percentage show that Scopus has excellent coverage of the references ($> 80\%$) in the main fields of biochemistry, genetics and molecular biology, the human biological sciences, and chemistry; very good coverage (60–80%) in chemical engineering, physics and astronomy, materials science, animal and plant biological sciences, geosciences, psychology, engineering, and computer science; good coverage (40–60%) in energy, mathematics, decision sciences, and economics; and moderate coverage ($<40\%$) in social sciences and arts and humanities. This is also fairly similar to the findings of Moed with respect to the ISI Web of Science [2].

Psychology is of note here (38.9% of all its references are in this period in SSS, and 65.4% of the references in this same period are in SSS) as it has a high rank order compared with the rest of the Scopus social sciences category. This could be because of its mixed health and social science character.

Table 5 presents the results of the analysis of correlations between the SJR and JIF (two-, three- and four-year citation windows) citation quality indicators averaged for each area and the basic bibliometric magnitudes that have been considered in the present study.

Given that both the different average size of the reference lists and the different sizes of the various scientific disciplines (as represented by the size of their total body of references) have been suggested as causes of the distinct citation rates,

and consequently of the different averages and statistical significances of the bibliometric indicators, it is perhaps surprising that no marked relationship was found between the average indicators and the mean number of references per article, the total number of references in an area, the total number of articles in an area, and even the total number of SSS-indexed references in an area. For example, the mean correlation between the average indicators and the total number of references was 0.269, that of the average number of references per article was 0.449, and that of the total number of articles was 0.107. It can thus be concluded that the size of neither the discipline itself nor its reference lists has any significant influence on the bibliometric indicators.

Table 5
Pearson correlation coefficients between standard citation quality indicators and basic bibliometric magnitudes

	SJR	JIF4Y	JIF3Y	JIF2Y	Average
Ref/art	0.436	0.448	0.455	0.458	0.449
Ref/art in SSS	0.834	0.882	0.898	0.904	0.880
Ref Win 3Y/art in SSS	0.870	0.899	0.916	0.929	0.903
Ref Win 2Y/art in SSS	0.877	0.900	0.916	0.931	0.906
Ref Win 3Y/art	0.820	0.849	0.866	0.883	0.854
Ref Win 2Y/art	0.830	0.853	0.870	0.892	0.861
Total References in SSS	0.394	0.363	0.376	0.368	0.375
Total References	0.277	0.263	0.274	0.261	0.269
Total Articles	0.112	0.101	0.111	0.103	0.107
Rate in SSS	0.703	0.724	0.739	0.756	0.731
Rate in 3Y Window SSS	0.722	0.719	0.734	0.757	0.733
Rate in 2Y Window SSS	0.737	0.727	0.740	0.765	0.742
Rate in 3Y Window	0.541	0.543	0.557	0.581	0.556
Rate in 2Y Window	0.549	0.544	0.557	0.585	0.559

Also worthy of note is the strong correlation between the average indicators and the recent SSS-indexed references. In hindsight, this seems logical since the indicators are computed by taking into account recent references to the database. The rest of the magnitudes representing average numbers of references also present a high correlation. They are followed by the percentage ('Rate') of recent references to SSS papers.

If one tries to distinguish which factor – citing recent papers or citing papers in SSS – has the greater effect, corresponding to referencing in SSS would seem to have more influence, with the difference being more appreciable in the percentage magnitudes (Rate in SSS versus Rate in 2Y Window, with a difference of 0.17) than in the averages of references per article (References/Article in SSS versus References Window 2Y/Article, with a difference of 0.02).

The strongest correlation found was between References Window 2Y/Article in SSS and JIF2Y. Again, this was to be expected since, if an area refers to papers that are in the two-year window and that are included in the database, it will have a high average two-year citation window JIF.

4. Conclusions

The first striking conclusion is the rejection of the hypothesis that is widely accepted in the scientific literature that there is a relationship between the size of the discipline and the statistical significances and averages of citation-based bibliometric indicators. Moreover, there was no significant influence of the size of the reference list.

A second conclusion is that the averages of the SJR and JIF indicators are strongly correlated with the average number of references to recent papers included in the database. This implies that the statistical significance and average values of those indicators depend on the citation habits of the different disciplines: the greater the traffic of citations to recent works present in the database, the greater will be the

average values and statistical significance of the indicators. This is quite consistent with the findings of Althouse et al. [24], and indeed was to be expected since those indicators are calculated using the recent citations of the database. More sophisticated indicator models are starting to take this dependence into account [27].

For example, arts and humanities is one of the areas with most references per article. But when the date of the references is limited to the period 1996–2007, it becomes the area with fewest references, far behind the rest. Indeed, it has often been observed in the scientific literature that the mean age of the references in this area is greater than in the rest. When only references to papers of the period (1996–2007) included in the database are considered, the different areas become more clearly separated (this is consistent with other published studies on the ISI databases [23]). At the bottom are other areas of the Scopus social sciences category. At the opposite extreme, with the highest value, is immunology and microbiology reflecting its habit of citing recent work due to the rapid obsolescence of the literature in this area.

The Scopus life sciences and health sciences categories are those with the highest reference rates to SSS-indexed papers, while the Scopus social sciences category has the lowest rate, although psychology (included in the social sciences category) is slightly higher, possibly due to its link with the health sciences.

The case is similar, although to a less pronounced extent, for the citation of recent papers. As was to be expected, the greatest differences were found when the two factors were considered conjointly.

In this regard, in order to increase their capacity to evaluate all the sciences, the strategies for the expansion of such bibliometric databases as the Web of Science or Scopus need to avoid imbalances in citations across disciplines in the database, improve temporal coverage, and include new publications which, in the case of humanities and social sciences, would be both journals and monographs.

With respect to the main fields, one can say that Scopus has excellent coverage of the references ($> 80\%$) in the main fields of biochemistry and molecular biology, the human biological sciences, and chemistry; very good coverage (60–80%) in chemical engineering, physics and astronomy, materials, the animal and plant biological sciences, geosciences, psychology, engineering, and computer science; good coverage (40–60%) in energy, mathematics, decision sciences, and economics; and moderate coverage ($<40\%$) in other social sciences and arts and humanities. This is not inconsistent with the findings of Braun et al. [6] or Moya-Anegón et al. [8] since those studies found that the database is balanced relative to Ulrich's in certain areas. Indeed, there is a major overlap between the areas that have a higher percentage of SSS-indexed references and those that are more balanced relative to their presence in the Ulrich web source. This suggests that, as mentioned above, for certain areas it is not only necessary to extend the temporal coverage but also the content coverage. This is also fairly similar to the findings

of Moed with respect to the ISI Web of Science [2]. In this connection the Scopus team recently announced their intention of increasing the coverage of journals in the arts and humanities to 3500 by adding all top global journals using the European Reference Index for the Humanities (ERIH) of the European Science Foundation. ISI has also extended its database with 700 regional journals.

Acknowledgments

This work was financed by the Junta de Extremadura – Consejería de Educación Ciencia and Tecnología and the Fondo Social Europeo as part of research project PRI06A200, and by the Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica 2008–2011 and the Fondo Europeo de Desarrollo Regional (FEDER) as part of research projects TIN2008–06514-C02–01 and TIN2008–06514-C02–02.

References

- [1] T.A. Brooks, Private acts and public objects: an investigation of citer motivations, *Journal of the American Society for Information Science* 36(4) (1985) 223–229.
- [2] H.F. Moed, *Citation Analysis in Research Evaluation* (Springer, Dordrecht, 2005).

- [3] D. Filippo and M.T. Fernández, Bibliometría: importancia de los indicadores bibliométricos. In: *El estado de la ciencia* (Red Iberoamericana de Indicadores de Ciencia y Tecnología (RICYT), 2002).
- [4] J. Lundberg, Lifting the crown – citation z-score, *Journal of Informetrics* 1 (2007) 145–154.
- [5] V.P. Guerrero-Bote, F. Zapico-Alonso, M.E. Espinosa-Calvo, M.R. Gómez-Crisóstomo and F. Moya-Anegón, The Iceberg Hypothesis: Import-export of knowledge between scientific subject categories, *Scientometrics* 71(3) (2007) 423–441.
- [6] T. Braun, W. Glänzel and A. Schubert, How balanced is the Science Citation Index's journal coverage? A preliminary overview of macrolevel statistical data. In: B. Cronin and H.B. Atkins (eds), *The Web of Knowledge, a Festschrift in Honor of Eugene Garfield* (American Society of Information Science, Canada, 2000).
- [7] Grupo SCImago, Análisis de la cobertura de la base de datos Scopus, *El profesional de la información* 15(2) (2006) 144–145.
- [8] F. Moya-Anegón, Z. Chinchilla-Rodríguez, B. Vargas-Quesada, E. Corera-Álvarez, F.J. Muñoz-Fernández and A. González-Molina, Coverage analysis of Scopus: a journal metric approach, *Scientometrics* 73(1) (2007) 53–78.

- [9] R.N. Broadus, The literature of the social sciences: a survey of citation studies, *International Social Sciences Journal* 23 (1971) 236–243.
- [10] E.S. Clemens, W.W. Powell, K. Mcilwaine and D. Okamoto, Careers in print: books, journals, and scholarly reputations, *American Journal of Sociology* 101 (1995) 433–494.
- [11] B. Cronin, H. Snyder and H. Atkins, Comparative citation rankings of authors in monographic and journal literature: a study of sociology, *Journal of Documentation* 53 (1997) 263–273.
- [12] L.L. Hargens, Using the literature: reference networks, reference contexts, and the social structure of scholarship, *American Sociological Review* 65 (2000) 846–865.
- [13] S. Kyvik, Changing trends in publishing behaviour among university faculty, 1980–2000, *Scientometrics* 58 (2003) 35–48.
- [14] G. Lewison, Evaluation of books as research outputs in history of medicine, *Research Evaluation* 10 (2001) 89–95.
- [15] Y. Lindholm-Romantschuk and J. Warner, The role of monographs in scholarly communication: an empirical study of philosophy, sociology and economics, *Journal of Documentation* 54 (1996) 389–404.

- [16] A.J. Nederhof, R.A. Zwaan, R.E. De Bruin and P.J. Dekker, Assessing the usefulness of bibliometric indicators for the humanities and the social sciences, *Scientometrics* 15 (1989) 423–435.
- [17] D.A. Nock, Careers in print: Canadian sociological books and their wider impact, 1975–1992, *Canadian Journal of Sociology/Cahiers canadiens de sociologie* 26 (2001) 469–485.
- [18] D.J. Price, Citation measures of hard science, soft science, technology, and non-science. In: C.E. Nelson and D. Pollack (eds), *Communication Among Scientists and Engineers* (Lexington Books, Lexington, 1970).
- [19] H.G. Small and D. Crane, Specialties and disciplines in science and social science: An examination of their structure using citation indexes, *Scientometrics* 1 (1979) 445–461.
- [20] J.W. Thompson, The death of the scholarly monograph in the humanities? Citation patterns in literary scholarship, *Libri* 52(3) (2002) 121–136.
- [21] E. Garfield, *Citation Indexing. Its Theory and Application in Science, Technology and Humanities* (Wiley, New York, 1979).
- [22] W. Glänzel, A bibliometric approach to the social sciences. National research performances in 6 selected social science areas, 1990–1992, *Scientometrics* 35 (1996) 291–307.

[23] W. Glänzel and U. Schoepflin, A bibliometric study of reference literature in the sciences and social sciences, *Information Processing and Management* 35 (1) (1999) 31–44.

[24] B.M. Althouse, J.D. West, C.T. Bergstrom and T. Bergstrom, Differences in impact factor across fields and over time, *Journal of the American Society for Information Science and Technology* 60(1) (2008) 27–34.

[25] B. González-Pereira, V.P. Guerrero-Bote and F. Moya-Anegón, The SJR indicator: A new indicator of journals' scientific prestige (unpublished manuscript, Madrid, 2009). Available at: <http://arxiv.org/ftp/arxiv/papers/0912/0912.4141.pdf> (accessed 6 January 2010).

[26] L. Leydesdorff, F. Moya-Anegón and V.P. Guerrero-Bote, Journal maps on the basis of Scopus data: a comparison with the Journal Citation Reports of the ISI, *Journal of the American Society for Information Science and Technology* (in press).

[27] H.F. Moed, Measuring contextual citation impact of scientific journals (unpublished manuscript, Amsterdam, 2009). Available at <http://arxiv.org/ftp/arxiv/papers/0911/0911.2632.pdf> (accessed 6 January 2010).

ARTÍCULO

A3: Citation Flows in the Zones of Influence of Scientific Collaborations

Bárbara S. Lancho Barrantes, Vicente P. Guerrero Bote,

Zaida Chinchilla Rodríguez and Félix de Moya Anegón

In press in:

Journal of the American Society for Information Science and Technology.

Volume 63, Issue 3, pages 481–489, March 2012

Citation Flows in the Zones of Influence of Scientific Collaborations

Bárbara S. Lancho Barrantes & Vicente P. Guerrero Bote

Departamento de Información y Comunicación, Grupo Scimago, Universidad de Extremadura, Plazuela Ibn Marwan, 06071 Badajoz, Spain.

Email: {bslancho, guerrero}@unex.es

Zaida Chinchilla Rodríguez & Félix de Moya Anegón

Grupo Scimago, CSIC, CCHS, IPP, C/Albasanz, 26-28. 28037 Madrid, Spain.

E-mail: {zaida.chinchilla, felix.demoya}@cchs.csic.es.

Abstract

Domestic citation to papers from the same country and the greater citation impact of documents involving international collaboration are two phenomena that have been extensively studied and contrasted. Here, however, we show that it is not so much a national bias, but that papers have a greater impact on their immediate environments, an impact that is diluted as that environment grows. For this reason, the greatest biases are observed in countries with a limited production. Papers that involve international collaboration have a greater impact in general, on the one hand, because they have multiple “immediate environments,” and on the other because of their greater quality or prestige. In short, one can say that science knows no frontiers.

Certainly there is a greater impact on the authors' immediate environment, but this does not necessarily have to coincide with their national environments, which fade in importance as the collaborative environment expands.

Introduction

Some workers have analyzed the degree of collaboration using the coauthorship of publications (Carpintero & Peiró, 1983; Alcaín & Gálvez, 1998). There is indeed some correlation between the coauthorship index and the impact of papers, so that scientific communities gain in visibility as their networks grow and bring together a greater number of associates (Moya et al., 2008). In general, high levels of collaboration lead to high levels of impact, greater quality of the papers published, and greater productivity of the authors in their particular scientific fields (Lewinson & Cunningham, 1991; Narin, Stevens, & Whitlow, 1991; Glänzel, 2001; Glänzel, 2002; Leimu & Koricheva, 2005; Katz & Hicks, 1997; Persson, Glanzel, & Danell, 2004; Hsu & Huang, 2010; Aksnes, 2003; Moya-Anegón et al., 2008; Chinchilla, Vargas, Hassan, González, & Moya, 2010)

The potential benefits of scientific collaboration may depend on the discipline. The effect of collaboration on scientific impact appears to be more positive in the “hard” sciences such as physics and astronomy, than in the “soft” sciences such as sociology or social sciences (Stack, 2002; Bandyopadhyay, 2001; Moed, Bruin, Nederhof, & Tijssen, 1991; Bridgstock, 1991), with citation behaviour sometimes

differing considerably from one field to another (Lancho, Guerrero, & Moya, 2010a, 2010b).

The benefits will also depend on the different types of collaboration (Leimu & Koricheva, 2005; Katz & Hicks, 1997): (a) domestic in-house collaboration (all authors from the same institution); (b) domestic institutional collaboration (all authors from the same country but from more than one institution); and (c) international collaboration (authors from more than one country; (Leimu&Koricheva). Although institutional collaboration is more important than domestic in-house collaboration, international collaboration is even more so in the sense that it increases the citation rates far above those of domestic national collaboration (Narin et al., 1991; Katz & Hicks; Goldfinch, Dale, & De Roue, 2003; Sooryamoorthy, 2009).

Gómez, Fernandez, and Sebastian (1999) consider that international collaboration increases the visibility of research papers since they are published in journals of greater impact than those of national collaborations. Narin and Whitlow (1990) find evidence that papers with a multiple international authorship have double the citation frequency of those without such collaboration. Schmoch and Schubert (2008) suggest that international papers are more highly cited because their potential community is larger. In other words, international papers may be more highly cited simply because more people know about them.

This and the overlap of the communities involved could explain the difference in impact between collaborations in different fields, because the overlap of communities in fields that are close to each other is greater than in fields that are farther apart. To this one must add a possible national bias in the various countries involved in a collaboration that could work to the benefit of papers with international collaboration.

Are these suggestions and theoretical expectations of a greater impact and citation rate of international collaborations actually borne out in reality, however? If so, where does that impact come from? Does it come mainly from the countries included in the collaboration? Could the impact be because of the international partners in the collaboration? Do these papers also have a greater national impact? Do countries have a citation bias towards papers from their own country? Is the greater impact of such international collaboration papers because the authors come from more countries?

The main hypothesis of the present study was that science knows no frontiers—that there is no national citation bias. If there is a certain bias towards what is nearby, this simply reflects the density of the authors' personal relationships and therefore of the information channels (both incoming and outgoing) that they use, a density that becomes diluted with increasing size of the area considered. This means that collaborative papers may gain in citations because of the increase in these informal channels of information dissemination, but above all, there exists a general gain in impact.

In the present work, we inquire into the provenance of the citations received by the different countries and the different types of papers, and into the targets of the citations made by the different countries. We shall take a country's production into account in two ways: first, that clearly the greater the production, the more citations the country has the potential to receive, but also the more domestic citations (or references) it will introduce into the arena.

Data and Methods

We used Scopus as the data source for the computation of the indicators because it best represents the overall structure of world science at a global scale. Scopus is the world's largest scientific database. It covers most of journals included in Thomson Reuters ScientificWeb of Science (WoS) and more (Moya-Anegón et al., 2007), and its coverage is statistically balanced in terms of subjects, countries, languages, and publishers. The greater part of the data was retrieved from the SCImago Journal & Country Rank (<http://www.scimagojr.com/>) database, limiting the search to citations to papers published in 2004 from papers published in 2005, 2006, and 2007. The data were downloaded in December 2009.

We chose the 20 countries with the greatest production in 2004, which are the countries with the greatest statistical significance, after which the following countries show a sharp drop in production. It is not the purpose of this study to make inferences from a random sample, but to study what happens in these 20 countries with the greatest production.

They account for nearly 90% of overall scientific production, and nearly 97% of the scientific production involving international collaboration.

The main indicators used are as follows:

- Citations per paper: Average citations received by the papers of 2004 from the papers of 2005–2007, calculated by dividing the number of citations/references by the number of target papers.
- References per paper: Average references in papers of 2005–2007 to papers of 2004, calculated by dividing the number of citations/references by the number of source papers.

- Impact on papers per paper: The impact (or influence) of the papers of 2004 on the papers of 2005–2007, calculated by dividing the number of citations/references by the number of target papers and by the number of source papers.

$$I_{pp} = \text{citations (references)} / \text{num. target papers} \cdot \text{num. source papers}$$

For example, to calculate the impact on nondomestic papers per collaboration paper for Germany, in the numerator, we put the citations from the nondomestic papers corresponding to the case of Germany to Germany's international collaboration papers. In this case, the number of target papers would be the number of Germany's collaboration papers of 2004, and the number of source papers would be the number of nondomestic papers for the case of Germany from 2005 to 2007. To calculate the impact on collaboration papers per nondomestic paper for Germany, in the numerator, we put the citations from Germany's international collaboration papers to the nondomestic papers corresponding to the case of Germany. In this case, the number of target papers would be the number of nondomestic papers for the case of Germany of 2004, and the number of source papers would be the number of Germany's collaboration papers from 2005 to 2007.

Unlike the previous two indicators, which are simple ratios, this last indicator takes into account the size of both the target set of papers (obviously, the larger this set, the greater the likelihood of attracting more citations) and the source set (again of course, the larger this set, the more likely that references will be made to the target set). This takes into account not only the average citations per paper, but also the maximum that this ratio could reach. The size of the set of source papers has normally not been taken into account because the sources are usually taken to be the set of the global production of papers. However, grouping citations in accordance with their provenance can have a major effect if source sets of different sizes are generated. This type of indicator was first used by Pichappan (1995) for the citation analysis of journals of a particular subject field.

In all cases, both the target and the source papers are classified by source (domestic, i.e., of the same country, or nondomestic), by collaboration type (with or without international collaboration, only distinguishing collaboration with institutions of other countries), and specific classifications for citations to or from the collaborating countries.

TABLE 1. The 20 countries ranked by total number of papers, with the corresponding percentage of papers in collaboration, percentage of citations from collaborating countries, and percentage of references to collaborating countries.

Country	Papers 2004	Papers 2005 to 2007	% Collab. Papers 2004	% Collab. Papers 2005 to 2007	% Cit. from Collaborators	% Ref. to Collaborators
United States	410521	1359565	23.28	25.29	15.57	14.36
China	113292	552902	16.09	13.99	24.18	32.65
Japan	110622	347570	19.79	20.94	35.30	34.89
United Kingdom	107143	362408	37.46	38.99	30.86	30.24
Germany	98949	326678	39.65	41.47	30.65	29.49
France	69591	232091	41.64	43.33	32.51	31.07
Canada	55929	196960	39.27	40.78	39.06	40.04
Italy	54112	183950	34.81	36.64	33.39	33.67
Spain	39776	144564	32.88	34.89	33.29	32.92
Australia	35886	127554	38.03	38.85	36.18	35.26
Russian Federation	35000	100774	30.93	33.20	34.89	32.77
India	33169	128505	18.22	18.28	33.61	33.08
Republic Of Korea	31332	120173	25.00	25.45	38.68	42.53
Netherlands	30168	104253	45.14	45.90	35.05	35.36
Switzerland	22049	75615	55.05	56.95	35.58	35.59
Brazil	21658	88335	27.75	25.35	36.16	35.50
Sweden	21280	69274	45.75	48.90	34.38	32.30
Taiwan	21071	81275	17.49	18.53	38.25	42.70
Poland	20378	67225	31.32	31.18	36.80	34.40
Turkey	18170	65901	16.75	15.39	32.61	36.60

Results and Discussion

By way of context, one observes in Table 1 that the United States gets only a small percentage of citations from its collaborators and in turn provides them with only a small percentage of references. We understand this to be because of its large production and hence the large number of national references involved, i.e.,

because U.S. production is so large, its domestic citations reduce the percentage from other countries (and analogously with regard to references).

But China, which went from 27% of U.S. production in 2004 to 40% in 2005–2007, despite having only a small percentage of collaboration (which even decreased in the second period), devotes a major proportion of its references to collaborating countries. Turkey, which is the country with the smallest production of the 20 countries considered and with the fewest collaboration papers, receives a high percentage of citations from collaborators, and references many of its collaborating countries, far more in relative terms than the United States or China. Although Switzerland has the highest percentage of collaboration papers, its percentage of citations received from its collaborating countries is not the highest.

The countries with the highest percentages of citations received from their collaborating countries and the highest percentages of references given to those countries are Canada, Republic of Korea, Australia, and Taiwan. These are not countries characterized by a great volume of scientific production.

The Pearson correlation coefficient between the two percentages (citations received from collaborating countries and references to those collaborating countries) is 0.88. This implies a major degree of reciprocity between references and citations to and from collaborators. The correlation coefficient between the percentage of citations received from collaborators and the papers published in 2005–2007 is negative, -0.9 . And that between the percentage of references to

collaborators and papers published in 2004 is also negative, -0.84 , i.e., the percentage traffic of citations with collaborating countries decreases as production increases. This makes sense simply in mathematical terms: the lower the production, the fewer domestic papers there will be to reference, and, vice versa, the fewer domestic papers there will be from which to receive citations, and hence the greater the respective percentages that will correspond to the traffic to and from nondomestic papers.

One can say in general that for the three citation indicators shown in Figure 1 (of the collaboration, no-collaboration, and overall papers), the nondomestic average exceeds the domestic average. The value of citations per collaboration paper is considerably greater than those of the citations per no-collaboration paper and citations per paper overall, with this difference originating mainly from nondomestic papers.

The case is similar for the indicators relating to references, although the differences are less marked. The average of nondomestic references is far greater than that of the domestic references in all cases, and the average references per collaboration paper is considerably greater than those of the references per no-collaboration paper and per paper overall.

As one can see in Figure 2, the greater part of these high values of citations and references per collaboration paper comes from nondomestic papers. The average domestic citation obtained in collaborative papers is 1.9, while the average citation per paper coming from domestic papers is 6.85, three times greater. With respect to references, the average of domestic references per collaborative paper targeted at collaborative papers is 0.52, while the average of references to nondomestic papers per collaboration paper is 1.62, i.e., triple.

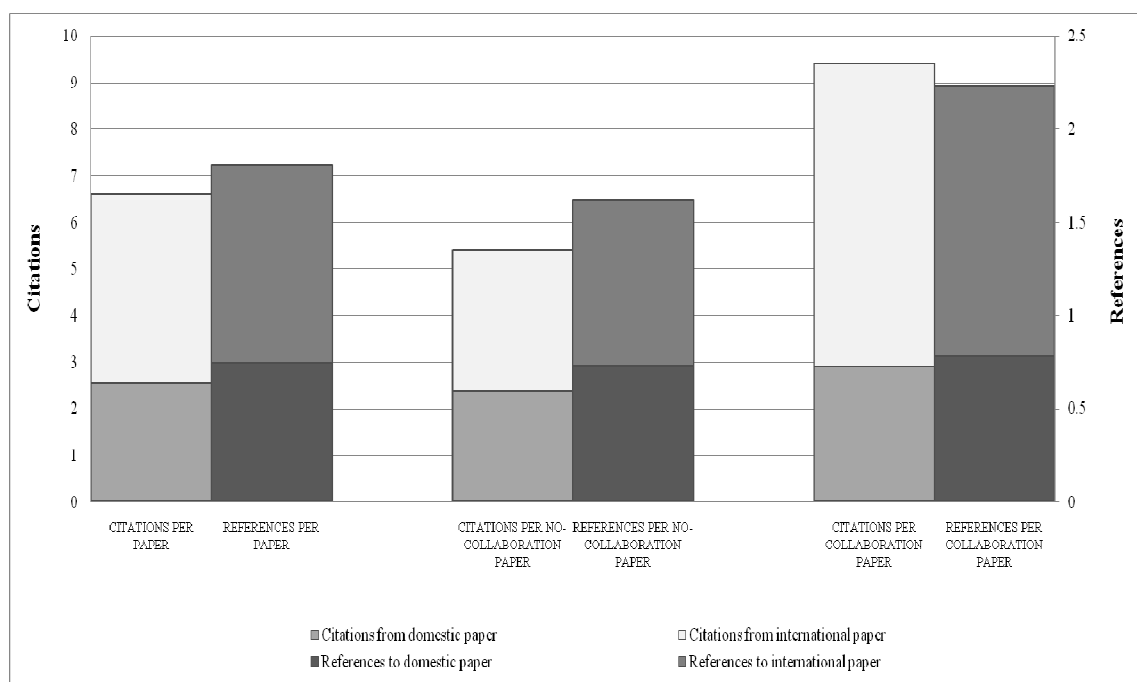


FIG. 1. Comparison of weighted average citations (per paper overall, per paper without collaboration, and per paper with collaboration) and references (per paper overall, per paper without collaboration, and per paper with collaboration) of the 20 countries with the highest production in 2004, distinguishing in both cases between domestic and nondomestic papers.

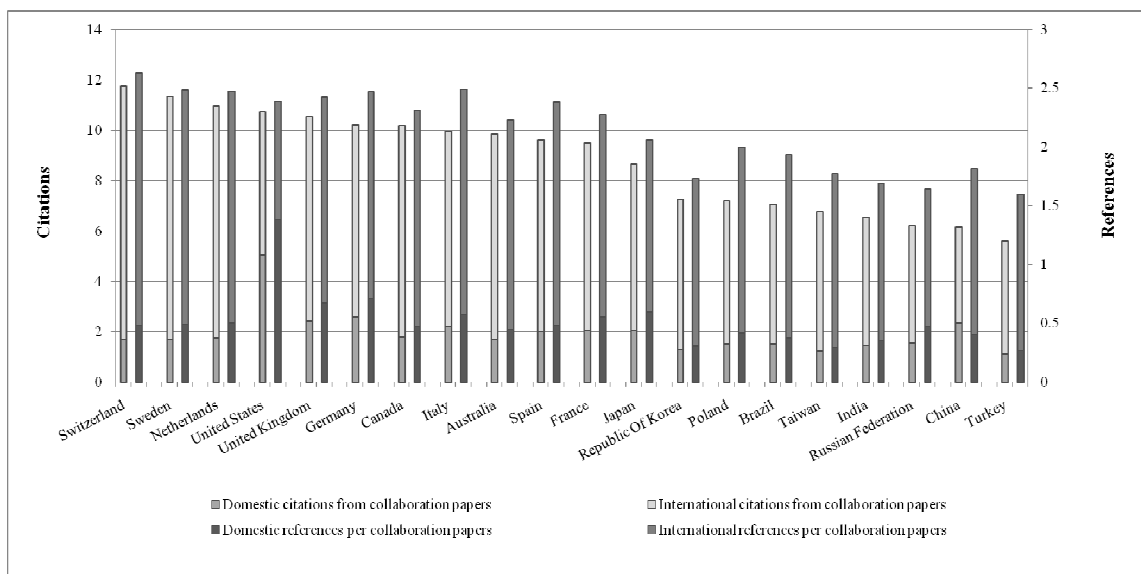


FIG. 2. Citations per collaboration paper (domestic and nondomestic) and references per collaboration paper (domestic and nondomestic), ordered from greater to lesser values of the former of these two parameters.

But it has to be said that there are some cases that deviate somewhat from these averages. For example, for the United States and China, almost 50% (to a lesser extent China) come from their domestic citations (recall that they were the countries with the greatest production of documents in 2004, but with the lowest percentage of citations from their international collaborators). The countries receiving the fewest domestic citations, Taiwan and Turkey, were those that received the greatest percentage of citations from their international collaborators. The same is the case with respect to the references.

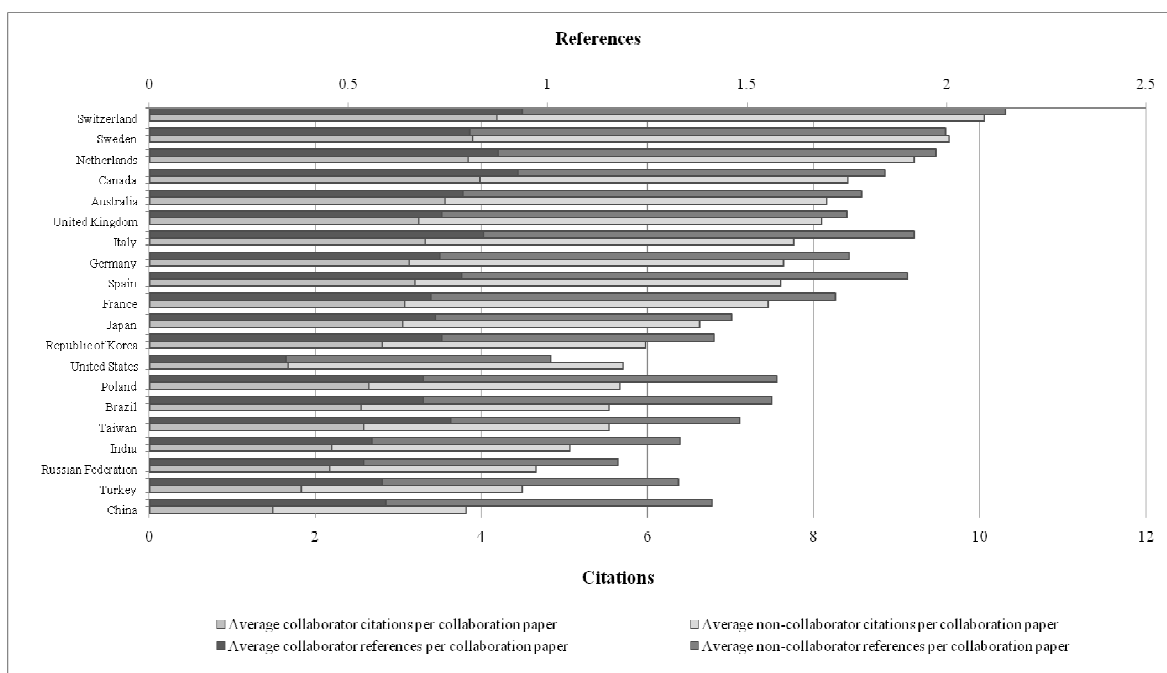


FIG. 3. Comparison between nondomestic citations per collaboration paper and references to nondomestic papers per collaboration paper, showing the average citations from/to collaborator and noncollaborator countries, ordered from greater to lesser values of nondomestic citations per collaboration paper.

Although at first glance there is some correlation between citations and references, in countries with the fewest citations there is a stagnation in the number of references.

As one observes in Figure 3, the greater part of the nondomestic citations per international collaboration paper comes from other countries. The overall average received from the collaborators is 2.9, and the average received from other countries is 3.9. The Pearson correlation coefficient between the two is 0.84.

Again, there are countries that deviate a little from this pattern, as is the case of the United States and China, which, compared with other countries, receive less from their international collaborators and in turn give them less. The case of China may be slightly different because although it captures relatively few citations, it does make a greater number of nondomestic references. The average of references given to collaborating countries is 0.71 and to other countries 0.91. The Pearson correlation coefficient between the two sets of data is 0.71. The Pearson correlation coefficient between the average collaborator citations and references is 0.81, and between the same two sets of data for the other countries is 0.90.

The value of the average nondomestic citations of the collaboration papers is greater than that of the no-collaboration papers, even excluding the citations of the collaborating countries (see Figure 4). The same is the case with the references.

There are countries for which the difference in the averages is quite noticeable, as is the case of the Russian Federation, China, and Poland, and other countries for which the indicators are quite similar, as is the case of Switzerland, the Netherlands, Canada, and the United States. Switzerland, Sweden, and the Netherlands are the countries that make most references to nondomestic papers in their no-collaboration papers, and are also the countries with the highest nondomestic citation per paper in their international collaborations, although their productions are quite small.

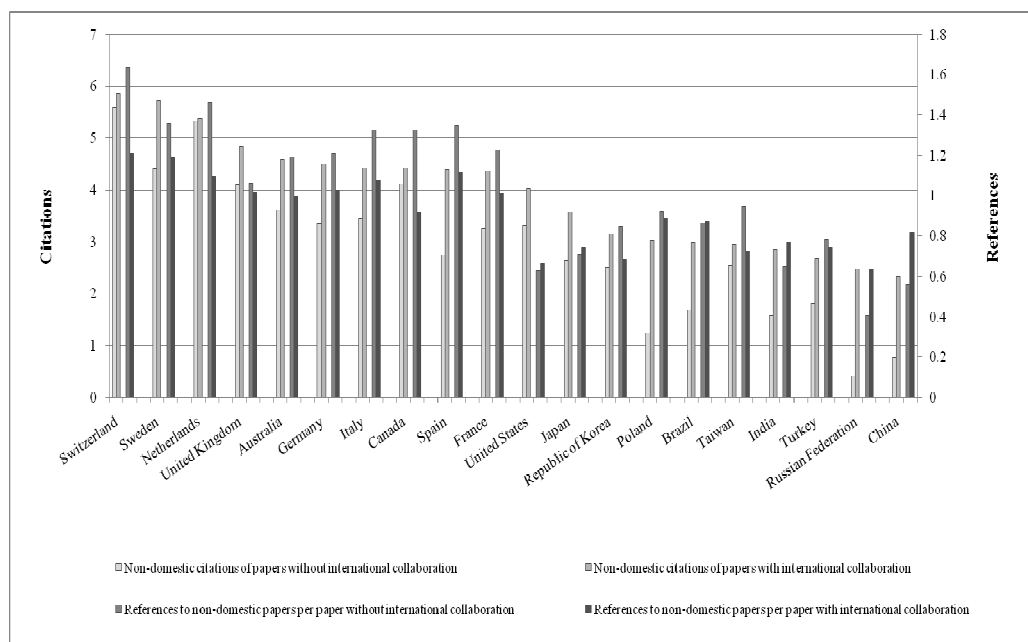


FIG. 4. Average nondomestic citations of papers without international collaboration, average nondomestic citations of papers with international collaboration (excluding citations from the collaborating countries), references to nondomestic papers per paper without international collaboration, and references to the nondomestic papers per collaboration paper (excluding those corresponding to the collaborators).

Naturally, the impact on domestic papers per paper of the country is greater than both the impact on nondomestic papers per paper of the country and the impact on papers of the country per nondomestic paper, as shown in Figure 5 in which two different scales have had to be used.

Coincidentally, the first three countries are the three with the greatest values of citations per collaboration paper, and the greatest national impact. These are Sweden, Switzerland, and the Netherlands, which are also those with the greatest nondomestic impact, probably because of their high proportion of papers with both national and international collaboration.

At the opposite end are the countries with the greatest production, China and the United States. Coincidentally, they are also countries with high values of domestic citations per collaboration paper. This high value of domestic citations is thus due to the large domestic production and the great number of references contributed by this production, because if this volume is taken into account, then these countries turn out to be the least biased towards their own production.

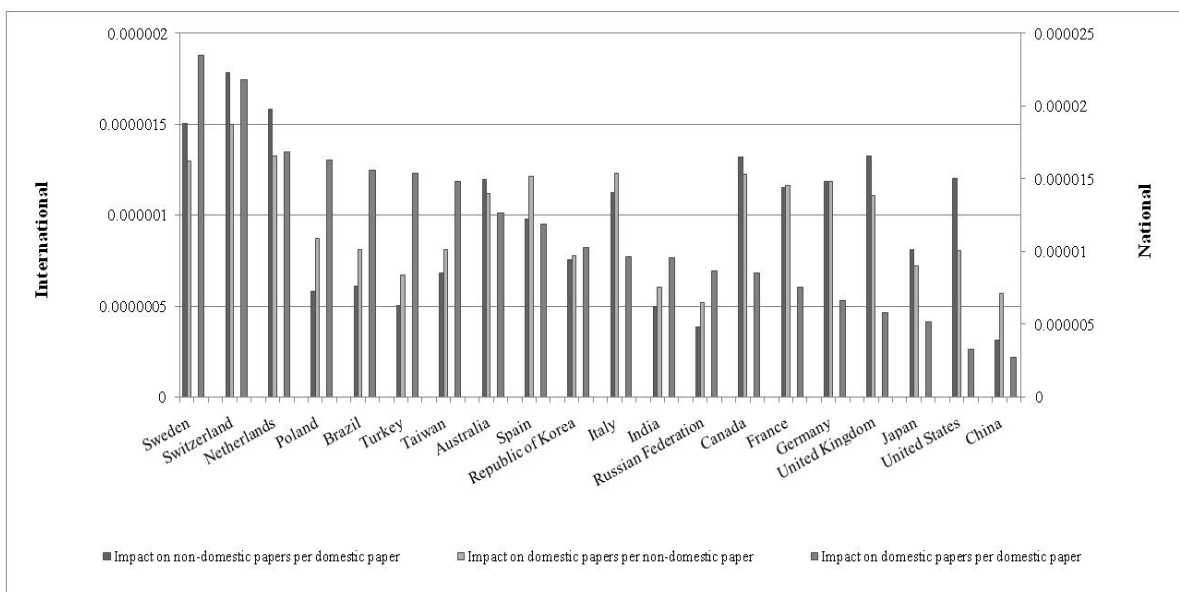


FIG. 5. Impact on nondomestic papers per domestic paper, impact on domestic papers per nondomestic paper, and impact on domestic papers per domestic paper, ordered from greater to lesser values of this last parameter.

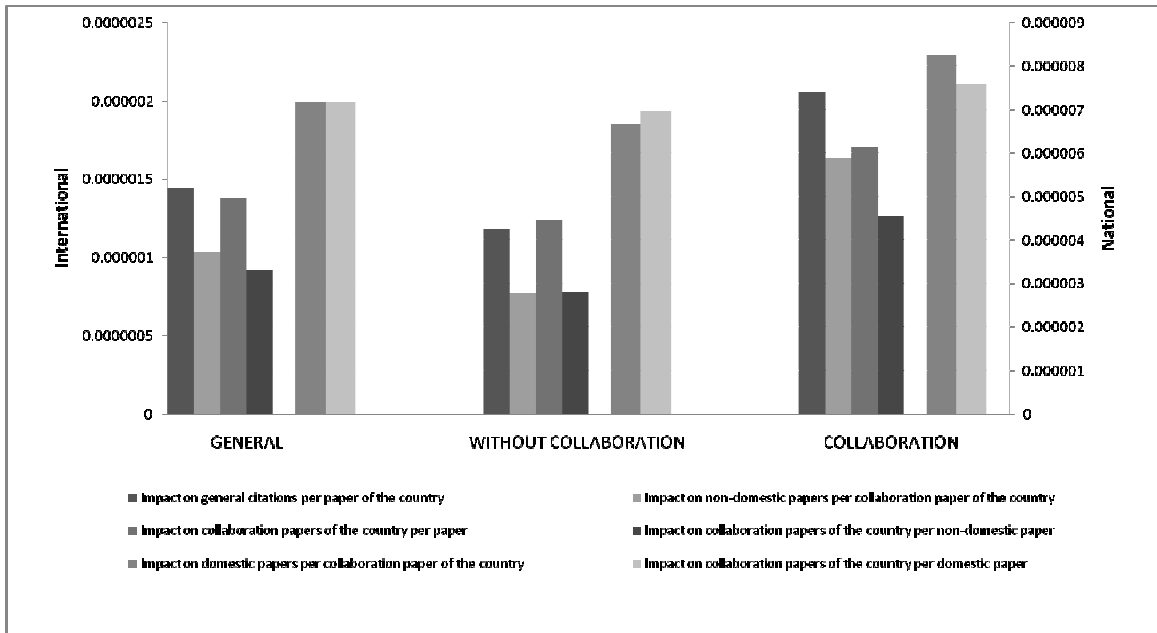


FIG. 6. Weighted averages for the production of 2004 of the impact on general papers (1st), on domestic papers (5th), and on nondomestic papers (2nd) per domestic paper, per domestic paper, and per collaboration paper, and impact on domestic papers, on papers without collaboration, and on collaboration papers per general paper (3rd), per domestic paper (6th), and per nondomestic paper (4th).

There is a negative correlation of -0.6 between production and the impact on domestic papers per paper of the country, while the correlation between domestic and nondomestic impacts is less than 0.38.

Disaggregating these influences by documents with or without international collaboration, one finds that the average national impact is the greatest. This means that, as a rule, all countries tend to have more impact on their own production (Figure 6). Collaborative papers also have a greater impact overall (both domestic and nondomestic), with the nondomestic growing more percentage-wise, so that in them there is less difference between the domestic and nondomestic impacts.

Disaggregating by country the impact of papers with international collaboration, one finds the same general features. As in Figure 5, one observes groups of countries with similar behaviour:

- Great production, with little domestic impact (United States, China, and Japan).
- Major domestic and nondomestic impact, but with a not very prominent production (Switzerland, Sweden, and the Netherlands).
- Little impact, whether domestic or nondomestic (India, Russia, Taiwan, and Korea).
- Major nondomestic impact, but little domestic impact (U.S., U.K., Germany, and Japan).

- Major domestic impact, but little nondomestic impact (Turkey, Poland, and Brazil – with Russia coming close to this group in Figure 7).

It has to be said that simple observation of the figures shows that for all the countries, the impact on the collaboration papers per domestic paper is fairly large, although the collaboration papers have a greater impact on the domestic papers than vice versa.

Interpretation

Although some countries with a large production, such as China and the United States, do not have a large percentage of international collaboration papers, there is no significant correlation between scientific production and percentage of collaboration. However, there is a significant negative correlation between production and the percentage traffic of citations to/from the collaborating countries, as was indeed to be expected mathematically.

In general, the average nondomestic citation per paper exceeds the domestic, and the average citation per paper of collaboration papers exceeds that of the no-collaboration papers (by more than 70%). Although there is an increase in the average domestic citation per paper of collaboration papers, there is comparatively a greater increase in the nondomestic citation per paper (by more than 110%).

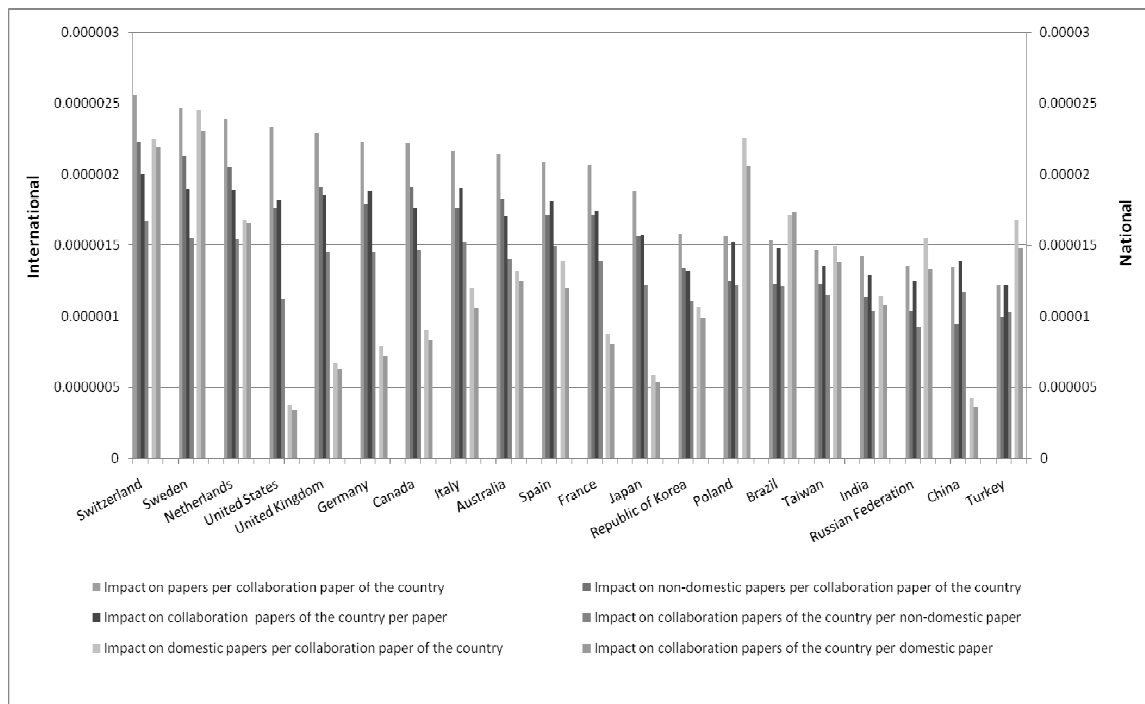


FIG 7. Impact on papers per collaboration paper of the country, impact on nondomestic papers per collaboration paper of the country, impact on collaboration papers of the country per paper, impact on collaboration papers of the country per nondomestic paper, impact on domestic papers per collaboration paper of the country, and impact on collaboration paper of the country per domestic paper, ordered from greater to lesser values of the first of these parameters.

Some 40% of this nondomestic citation per paper of collaboration papers comes from the collaborating countries, and there is a strong correlation between the collaborator citations and citations from the rest of the countries. The average nondomestic citation per paper of collaboration papers is greater than that of the no-collaboration papers, even when excluding that coming from the collaborating countries.

The impact on domestic production is greater than on nondomestic production. As expected, collaboration papers have a greater impact and there is less difference between the domestic and nondomestic impact.

There is a negative correlation between a country's production and its impact on domestic papers per paper. The countries with the largest productions tend to have the least impact on their own production, as is especially the case for the United States and China. On the contrary, the countries of small sizes in terms of scientific production may be favored in this sense, because either their international frontiers with other countries form more a part of their authors' immediate environments or, for some of them (e.g., Netherlands, Switzerland), they are also small in area with high population densities. Both these factors would naturally enhance the dissemination of ideas between countries. However, no significant correlation is observed between domestic and nondomestic impact.

Some countries (United States, United Kingdom, Germany, and Japan) have a great impact on nondomestic papers but little impact on domestic papers. At the opposite end of the spectrum, there are countries with a major domestic impact but little nondomestic impact, as is the case for Turkey and Poland, and Brazil. Emerging countries have, understandably, less impact outside their country as most of their researchers do not yet perform at the scientific frontier.

In general, one may conclude that science knows no national frontiers. Certainly there is a greater impact on authors' immediate environments, which do not necessarily coincide with national boundaries. This impact fades as the environment grows in size, while a small environment maximizes the bias produced by self-citation. Indeed, the greatest biases towards domestic production occur in small and developing countries.

Acknowledgments

This work was financed by the Junta de Extremadura e Consejería de Educación Ciencia & Tecnología and the Fondo Social Europeo as part of the pre-doctoral studentship PRE07052 and the research group grant GR10019, and by the Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica 2008e2011 and the Fondo Europeo de Desarrollo Regional (FEDER) as part of research projects TIN2008-06514-C02-01 and TIN2008-06514-C0202.

References

Aksnes, D. (2003). Characteristics of highly cited papers. *Research Evaluation*, 12(3), 159–170.

Alcaín, M., & Gálvez, M. (1998). Evolución de las revistas españolas de psicología. *Papeles del Psicólogo*, 70, 35–42.

Bandyopadhyay, A.K. (2001). Authorship pattern in different disciplines. *Annals of Library and Information Studies*, 48(4), 139–147.

Bridgstock, M. (1991). The quality of single and multiple authored papers– An unresolved problem. *Scientometrics*, 21, 37–48.

Carpintero, H., & Peiró, J. (1983). The significance of the bibliometric methodology to the studies of the history of psychology. *Revista de Historia de la Psicología*, 4(1), 21–32.

Chinchilla, Z., Vargas, B., Hassan, Y., González, A., & Moya, F. (2010). New approach to the visualization of international scientific collaboration. *Information Visualization*, 9(4), 277–287.

Glänzel, W. (2001). Domestic characteristics in international scientific coauthorship relations. *Scientometrics*, 51(1), 69–115.

Glänzel, W. (2002). Coauthorship patterns and trends in the sciences: A bibliometric study with implications for database indexing and search strategies, 1980-1998. *Library Trends*, 50(3), 461–473.

Goldfinch, S., Dale, T., & De Roue, K. (2003). Science from the periphery: Collaboration network and ‘periphery effects’ in the citation of New Zealand Crown Research Institutes articles, 1992–2000. *Scientometrics*, 57(3), 321–337.

Gómez, I., Fernández, M.T., & Sebastián, J. (1999). Analysis of the structure of international scientific cooperation networks through bibliometric indicators. *Scientometrics*, 44(3), 441–457.

Hsu, J.W., & Huang, D.W. (2010). Correlation between impact and collaboration. *Scientometrics*, 86(2), 317–324.

Katz, J., & Hicks, D. (1997). How much is a collaboration worth? A calibrated bibliometric model. *Scientometrics*, 40(3), 541–554.

Lancho, B., Guerrero, V., & Moya, F. (2010a). The iceberg hypothesis revisited. *Scientometrics*, 85(2), 443–461.

Lancho, B., Guerrero, V., & Moya F. (2010b). What lies behind the averages and significance of citation indicators in different disciplines? *Journal of Information Science*, 36(3), 371–382.

Leimu, R., & Koricheva, J. (2005). Does scientific collaboration increase the impact of ecological articles? *Bio Science*, 55, 438–443.

Lewinson, G., & Cunningham, P. (1991). Bibliometric studies for the evaluation of trans-domestic research. *Scientometrics*, 21(2), 223–244.

Moed, H., Bruin, R., Nederhof, A., & Tijssen, R. (1991). International scientific co-operation and awareness within the European Community: Problems and perspectives. *Scientometrics*, 21(3), 291–311.

Moya, F., Chinchilla, Z., Corera, E., González, A., Hassan, Y., & Vargas, B. (2008). *Indicadores bibliométricos de la actividad científica española: 2002–2006*. Madrid: Fecyt.

Moya, F., Chinchilla, Z., Vargas, B., Corera, E., Muñoz, F., González, A., & Herrero, V. (2007). Coverage analysis of Scopus: A journal metric approach. *Scientometrics*, 73(1), 53–78.

Narin, F., Stevens, K., & Whitlow, E. (1991). Scientific cooperation in Europe and the citation of multidomestically authored papers. *Scientometrics*, 21(3), 313–323.

Narin, F., & Whitlow, E. (1990). *Measurement of scientific cooperation and coauthorship in CEC-related areas of science*. Luxembourg: European Community.

Persson, O., Glanzel, W., & Danell, R. (2004). Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. *Scientometrics*, 60(3), 421–432.

Pichappan, P. (1995). A dual refinement of journal self-citation measures. *Scientometrics*, 33(1), 13–21.

Schmoch, U., & Schubert, T. (2008). Are international co-publications an indicator for quality of scientific research?. *Scientometrics*, 74(3), 361–77.

Sooryamoorthy, R. (2009). Do types of collaboration change citation? Collaboration and citation patterns of South African science publications. *Scientometrics*, 81(1), 177–193.

Stack, S. (2002). Gender and scholarly productivity: The case of criminal justice. *Journal of Criminal Justice*, 30(3), 175–182.

ARTÍCULO

A4: Citation increments between collaborating countries

Bárbara S. Lancho Barrantes, Vicente P. Guerrero Bote and Félix de Moya

Anegón

Aceptado en *Scientometrics*

DOI 10.1007/s11192-012-0797-3

Citation increments between collaborating countries

Bárbara S. Lancho Barrantes

Departamento de Información y Comunicación, Grupo SCImago, Universidad de Extremadura, Plazuela Ibn Marwan, 06071 Badajoz, Spain.

E-mail: bslancho@unex.es

Vicente P. Guerrero Bote

Departamento de Información y Comunicación, Grupo SCImago, Universidad de Extremadura, Plazuela Ibn Marwan, 06071 Badajoz, Spain.

E-mail: guerrero@unex.es

Félix de Moya Anegón

Grupo SCImago, CSIC, CCHS, IPP, C/Albasanz, 26-28. 28037 Madrid, Spain.

E-mail: felix.demoya@cchs.csic.es

Abstract

International collaboration enhances citation impact. Collaborating with a country increments the citations received from it. But some collaborating countries provide greater increments in this sense than others, and likewise some countries receive greater increments from their partner countries than others. We observed a certain tendency for these increments to be lower in countries with greater impacts. Also, all the countries studied had higher Domestic Impacts as a result of collaborating, although this increment was less than that obtained from other countries.

Finally, there were differences in the behaviour of the countries between the various scientific disciplines, with the effects being greatest in Social Sciences, followed by Engineering.

Keywords

Citation increment – Citation analysis – Scientific collaboration in subject areas – Scientometrics

Introduction

There is evidence that work performed in collaboration receives more citations (Lewinsonand Cunningham 1991; Narin et al. 1991; Katz and Hicks1997; Glänzel 2001; Glänzel 2002; Aksnes 2003; Persson et al. 2004; Leimu and Koricheva 2005; Moya-Anegón et al. 2008; Chinchilla-Rodríguez et al. 2010; Hsu and Huang 2010).

This increment in citation can depend on the type of scientific collaboration. For example, the greatest increment in citation comes from collaboration with institutions of different countries (Narin et al. 1991; Katz and Hicks 1997; Goldfinch et al.2003). Moreover, the increment comes not only from the collaborating country but also from other countries as a result of the expansion of the dissemination network (Lancho-Barrantes et al. 2012).

This increment in citations received can vary from one country to another (Inzelt et al. 2009; Zhao and Guan 2011). Various factors lead to there being differences between countries in their scientific work in collaboration – size of the country (Luukkonen et al. 1993; Ding et al. 1999), culture (citation habits and culturally acceptable behaviour) (Katz and Hicks 1997; Glänzel 2001; Singh 2005; Sooryamoorthy 2009; Gazni and Didegah2011), productivity and importance (countries with varying degrees of impact) (Price and Beaver 1966; Zuckerman 1968; Pao 1981; Glänzel et al. 1999; King 2004; Zhao and Guan 2011), and

geographical proximity (some countries essentially collaborate only with countries that are close geographically) (Kraut et al. 1988; Moed et al. 1991; Katz 1994; Glänzel and Schubert 2001; Leta and Chaimovich 2002; Goldfinch et al. 2003; Leimu and Koricheva 2005; Ma and Guan 2005; Figg et al. 2006).

There can also be differences between scientific disciplines, with the percentage of collaboration being much higher in some than in others (Garfield 1979; Bridgstock 1991; Bandyopadhyay 2001; Molteni and Zulueta 2002; Bandyopadhyay 2004; Sudhier 2007; Abt 2007; Ardanuy et al. 2009; Franceschet and Costantini 2010; Gazni and Didegah 2011; Ardanuy 2012).

The following questions thus arise. How does the increment in citations to and from collaborating countries vary from one country to another? Is the proportion of domestic citations to collaborative works greater or less than to works without collaboration? Is there an increment in the domestic impact of the collaborative production, and is there an increment in the impact obtained from collaborators? And finally, how do these increments vary between scientific disciplines?

The main idea examined in the present work is that when a country collaborates with another it receives an increment in citation relative to what would have been the case if there had been no such collaboration. In particular, we study the percentage increment in the number of citations to the collaborating countries, the citation rate to domestic collaborative documents, and the increment in impact

that is obtained. We then compare these quantities between four different disciplines.

Data and Methods

We used Scopus as the data source for the computation of the indicators because it best represents the overall structure of world science at a global scale. Scopus is the world's largest scientific database. It covers most of the journals included in the Thomson Scientific Web of Science (WoS) and more (Moya-Anegón et al. 2007), and its coverage is statistically balanced in terms of subjects, countries, languages, and publishers. The greater part of the data was retrieved from the SCImago Journal & Country Rank (<http://www.scimagojr.com/>) database, limiting the citations to papers published in 2004 from papers published in 2005, 2006, and 2007. The data was downloaded during December 2009.

To narrow the study, and for reasons of space, we selected the 9 most productive countries in 2004 to form Group L, and the 60 countries with a production of more than 1000 documents in 2004 (including of course those of Group L) to form Group W. We then studied the citations from the countries of Group L in the cases of collaborating or not with the countries of Group W.

To this end, we designed the following indicators:

- Citation Rate Increment from the Collaborator (CRIC)

$$CRIC_{A \rightarrow B} = \frac{\frac{\text{Citation}_{A \rightarrow AB}}{\text{Citation}_{AB}}}{\frac{\text{Citation}_{A \rightarrow B}}{\text{Citation}_B}}$$

A, B: Papers of countries A and B, respectively.

AB: Papers with collaboration between countries A and B.

Citations_{X→Y}: Citations from the set of papers A to the set of papers B.

Thus, the same set of documents in collaboration gives rise to two indicators: the Citation Rate Increment of Collaborator A when collaborating with B, and the Citation Rate Increment of Collaborator B when collaborating with A. One then determines two averages for each country – its Citation Rate Increment Average when Collaborating (CRIAC) and its Citation Rate Increment Obtained from Collaborators (CRIOC).

- Domestic Citation Rate Comparison when Collaborating (DCRCC)

$$DCRCC_{A \rightarrow B} = \frac{\frac{Citation_{A \rightarrow AB}}{Citation_{AB}}}{\frac{Citation_{A \rightarrow A}}{Citation_A}}$$

Again in this case, the same set of documents in collaboration gives rise to two further indicators: the Domestic Citation Rate Comparison of A when collaborating with B, and the Domestic Citation Rate Comparison of B when collaborating with A. Similarly, one determines two averages for each country – its Domestic Citation Rate Comparison Average (DCRCA) and its Domestic Citation Rate of Collaborators Comparison Average (DCRCCA).

- Domestic Impact Rate Increment when Collaborating (DIRIC)

$$DIRCC_{A \rightarrow B} = \frac{\frac{Citation_{A \rightarrow AB}}{Ndoc_{AB}}}{\frac{Citation_{A \rightarrow A}}{Ndoc_A}}$$

Once again, the same set of documents in collaboration gives rise to another two indicators: the Domestic Impact Rate Increment of A when collaborating with B, and the Domestic Impact Rate Increment of B when collaborating with A. The corresponding two averages for each country are – its Domestic Impact Rate Increment Average (DIRIA) and its Domestic Impact Rate of Collaborators Increment Average (DIRCIA).

To quantify in a rough and rapid form the relationships between the data sets, we used the Pearson correlation coefficient, although we are aware that for such small sets (9 items), the correlation coefficients are only statistically significant at a level of $p < 0.01$ from 0.798 upwards, and for $p < 0.05$ from 0.667 upwards.

Results and Discussion

Table 1 shows the USA to have been the country with the greatest production in 2004. Nevertheless, it obtained only a small percentage of citations from its collaborators, and in turn provided them with only a small percentage of its references. We understand this to be due to its high level of production, and the consequent large number of domestic references involved. I.e., the USA's production is so large that its domestic citations reduce the percentage from other countries (and similarly with respect to its references) (Lancho-Barrantes et al. 2012). It was the country with the greatest impact, and one of the countries with the greatest RpP. Following the USA in production was China, but this was the

country with the lowest percentage of documents in collaboration, and had a small percentage of citations from its collaborators. It was not one of the countries, however, which allocated only a small percentage of references to their collaborators. It had little total impact and a low value of RpP.

In last place in this 9-country ranking, Spain published the fewest documents in 2004, but its percentage of documents in collaboration was greater than that of the top three producing countries. The country received a high percentage of citations from its collaborators and responded similarly with its references to its collaborators. Nonetheless, it was one of the countries with low values of impact and RpP.

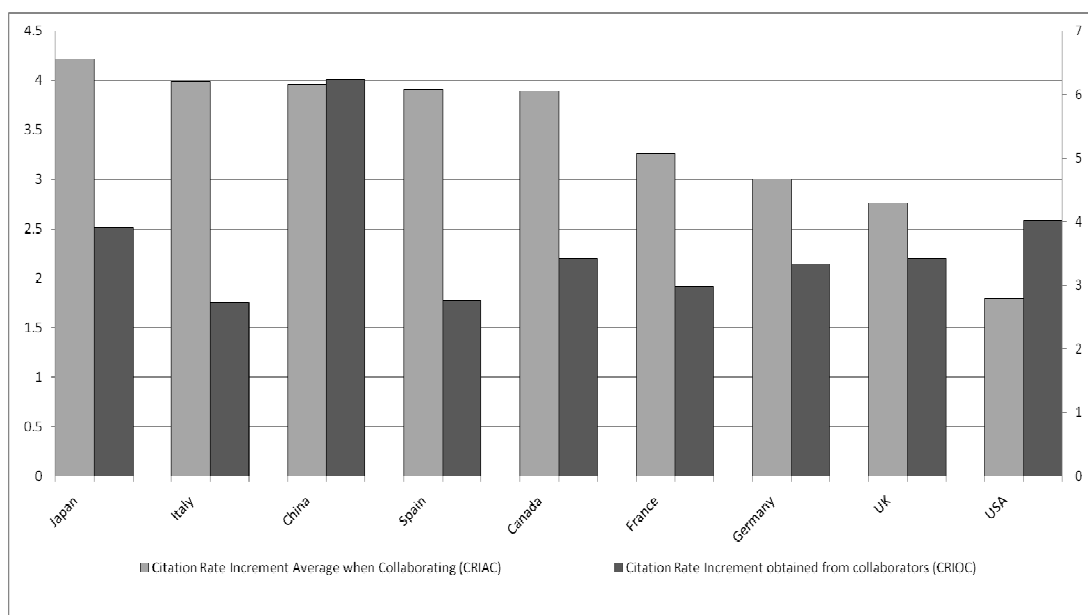
Table 1. The 9 countries with the greatest production in 2004 ordered by total number of documents, with their corresponding percentage of documents in collaboration, percentage of citations from collaborating countries, percentage of references to collaborating countries, total impact (JIF), domestic impact (DomJIF), references per paper (RpP), and domestic references per paper (DomRpP).

Country	Papers 2004	% Collab. Papers 2004	% Cit. fromCollaborators	% Ref. tocollaborators	JIF	NACJIF	RpP	NacRpP
UnitedStates	410521	23.28	15.57	14.36	8.37	4.49	2.08	1.36
China	113292	16.09	24.18	32.65	2.77	1.52	0.99	0.31
Japan	110622	19.79	35.30	34.89	7.70	2.11	1.95	0.62
UnitedKingdom	107143	37.46	30.86	30.24	5.23	1.80	1.44	0.57
Germany	98949	39.65	30.65	29.49	7.23	2.18	2.10	0.66
France	69591	41.64	32.51	31.07	6.75	1.74	1.96	0.52
Canada	55929	39.27	39.06	40.04	7.48	1.67	2.01	0.48
Italy	54112	34.81	33.39	33.67	6.72	1.77	2.06	0.52
Spain	39776	32.88	33.29	32.92	6.06	1.71	2.01	0.47

General Results

When a country collaborates, it obtains a Citation Rate Increment from the Collaborator (CRIC). In every case studied, a greater percentage of citations was received from a country with which it collaborated than in the case of non-collaboration.

Fig. 1. Comparison between the Citation Rate Increment Average when Collaborating (CRIAC) and the Citation Rate Increment Obtained from Collaborators (CRIOC) of the 9 countries with the greatest production in 2004. Ordered by CRIAC.



But some collaborating countries presented higher values of the Citation Rate Increment Average when Collaborating (CRIAC) (to their collaborator countries) than others.

China (3.96) and Japan (4.21) were the countries with the lowest percentages of documents in collaboration in 2004 (China 19%; Japan 20%), as also from 2005 to 2007 (China 6%; Japan 21%). Yet they were two of the countries with the highest values of Citation Rate Increment Average when Collaborating (CRIAC). In particular, their citation rate to a country increased by a factor of about 4 (on average) when there was collaboration. One reason may be the small number of references per paper (China has only about 50% of references per paper compared with the average, and Japan less than 75%). Added to this is that, after the USA, Japan was the country with the lowest percentage of non-domestic references. The number of citations in play was therefore small, so that any increment in citations to collaborative work would have had a large effect percentage-wise.

The situation of Italy (3.98), Spain (3.91), and Canada (3.89) was similar to that of China and Japan. However, they had greater percentages of documents in collaboration (Canada 39%, Italy 35%, and Spain 33%). They were also countries that gave a high percentage of their references to their collaborators, perhaps because they tend to convert their collaborations into their own production, and because, being smaller, their research environment and channels of dissemination are more clearly defined. Perhaps also these five countries, except for Canada, are those with the greatest language barriers, so that there would be a greater difference in the number of citations to other countries according to whether they were partners in a collaboration or not. These countries did not distribute the increment in the same way, but instead it was oriented to those countries which had a greater impact than theirs.

The USA surpassed China and Japan in the percentage of documents in collaboration. However it had the lowest value of Citation Rate Increment Average when Collaborating (CRIAC) (its citation rate to a given country on collaboration with it increased only by about 1 time on average). It was thus the country with most equality between the collaboration and non-collaboration citation rates. One recalls (Table 1) that the percentage of references it gave to its collaborators was very small (14%).

One factor that may have influenced the low value of CRIAC is its large size in terms of production – the greater the production, the more dilute is the research environment. Another factor is that it was the country with the greatest total impact, although most of this impact came from other countries. As a result, its relative citation rate from collaborating countries was small, and hence so would have been any variation in that rate. Unlike the countries discussed previously, the USA's distribution of this increment was very homogeneous – a standard deviation of 0.37 compared to the value of 2.0 for China.

The impact of the 9 countries of Group L was negatively correlated (correlation coefficient of -0.64) with the Citation Rate Increment Average when Collaborating (CRIAC) and with the standard deviation of those increments (correlation coefficient of -0.71). I.e., the greater the impact, the lower the increment on collaborating, and the less that increment depended on the collaborating country.

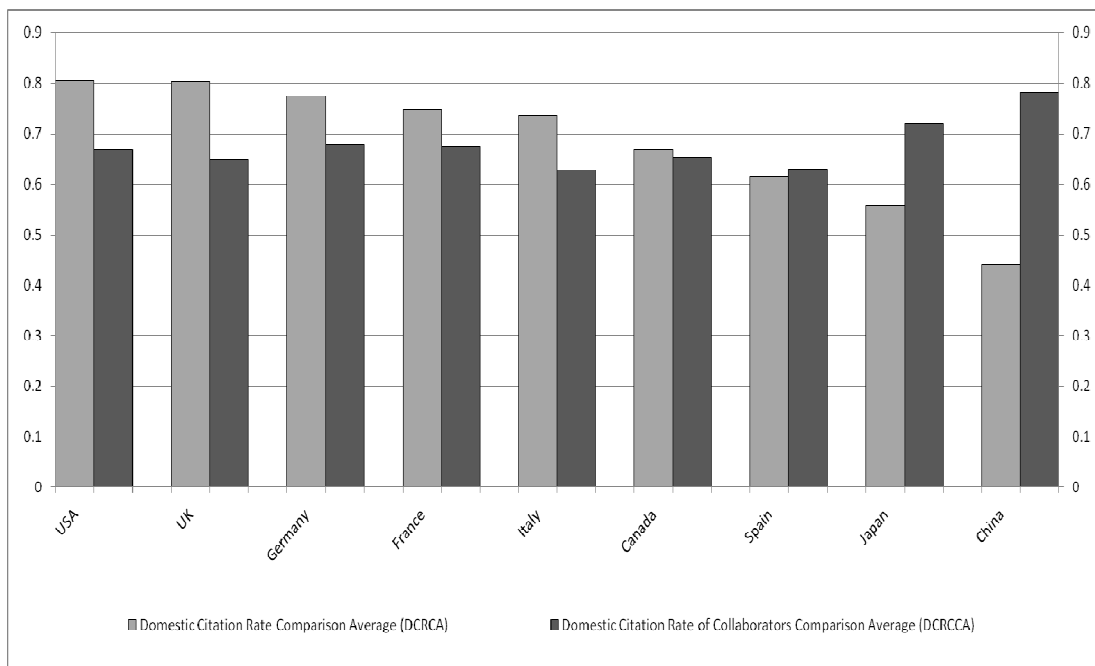
There were also countries which received a greater Citation Rate Increment Obtained from Collaborators (CRIOC) than others. There was a negative correlation of -0.51 between these values of CRIOC and impact. Coincidentally the USA, China, and Japan, the countries with the greatest production in 2004, were the countries with the greatest increment in the number of citations obtained from the Group L countries, but not equally from all the countries with which they collaborated. While the USA received much from countries with less impact (indeed, its CRIC was negatively correlated with impact, with a coefficient of -0.71), this was not the case for Japan and China, with the corresponding correlation being even slightly positive in the case of China.

These two countries received a greater increment from those countries in which there are greater language barriers – Italy, Spain, and France.

Italy and Spain (with less impact) had lower values of CRIOC, with most of the contribution coming from Canada, Japan, or China.

The Domestic Citation Rate (DCR) was found to be lower with collaboration.

Fig. 2. Comparison between Domestic Citation Rate Comparison Average (DCRCA) and Domestic Citation Rate of Collaborators Comparison Average (DCRCCA) of the 9 countries with the greatest production in 2004. Ordered by DCRCA.



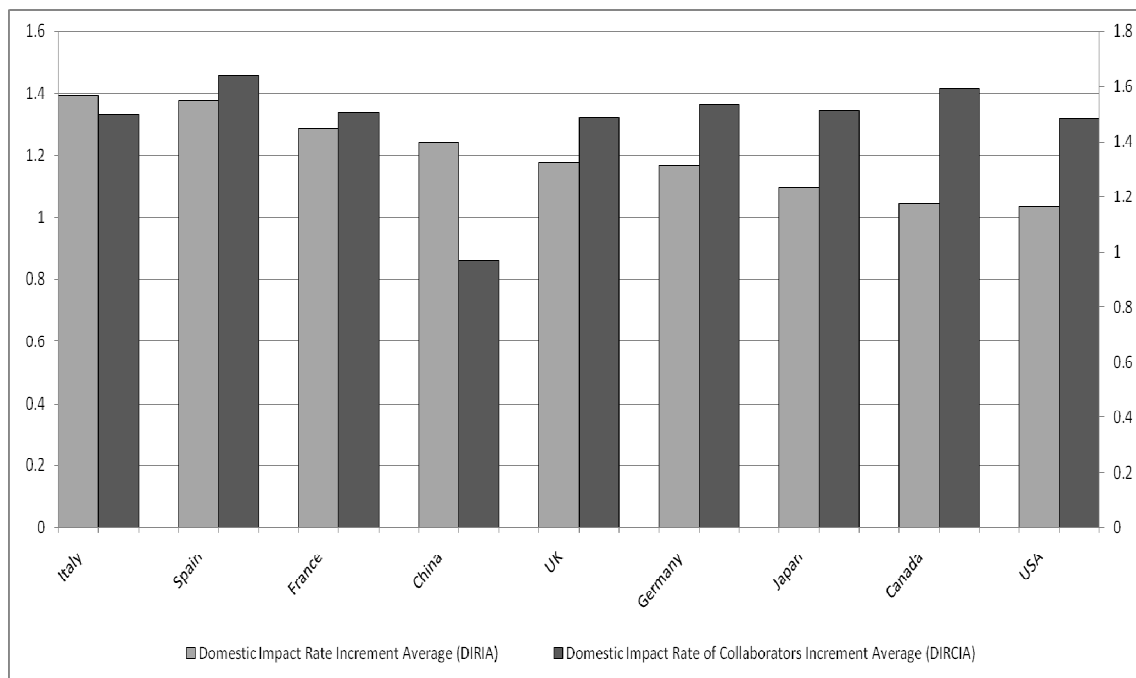
The UK, USA, and Germany –countries with high impact– came close to equality. Indeed, there was a correlation of 0.92 between the Domestic Citation Rate Comparison Average (DCRCA) and the impact of each country. With the opposite tendency were China, Japan, and Spain (countries with less impact) which were farther from equality. Canada did not fit the general pattern since, despite having a high impact, it was not very close to equality. It was actually the country with the lowest percentage of self-citation, although its percentage of citations to collaborations was even smaller.

Germany had a greater Domestic Citation Rate when collaborating with certain countries than overall, particularly when collaborating with Russia, Ukraine, Croatia, and Bulgaria, countries with low impact in its research environment of the countries of Eastern Europe.

China, Japan, and Germany were the countries receiving the greatest DCRCCA, i.e., a Domestic Citation Rate of its Collaborators closer to the value these had overall. In this sense, there was a negative correlation of -0.55 between their Domestic Citation Rate of Collaborators Comparison Average (DCRCCA) and impact. I.e., the lower the impact, the closer their Domestic Citation Rate obtained from their collaborators to the overall value of this index.

The Domestic Impact Rate (DIR), however, was found to be greater with collaboration. Therefore, if the DCR is lower when there is collaboration, and the DIR is higher, then the case is that more impact (citations) is also gained from other countries, so that the DCR decreases with collaboration.

Fig. 3. Comparison between Domestic Impact Rate Increment Average (DIRIA) and Domestic Impact Rate of Collaborators Increment Average (DIRCIA) of the 9 countries with the greatest production in 2004. Ordered by DIRIA.



Some collaborating countries received a greater increment in DIR from collaboration than others. The greater their total impact, the lower tended to be their DIRIA –the difference between the domestic impact and the impact provided to their collaborations (domestic impact when collaborating)– with a correlation coefficient of -0.31, and the greater tended to be their DIRCIA –the difference that the other countries make from the collaborations with respect to production– with a correlation coefficient of 0.50.

The countries which were farthest from equality between the domestic impact and the impact provided to their collaborations, i.e., those with the greatest values of DIRIA, were Italy, Spain, China (which, even though it was the country with the

lowest impact, it was still not that with the greatest comparative impact), and France.

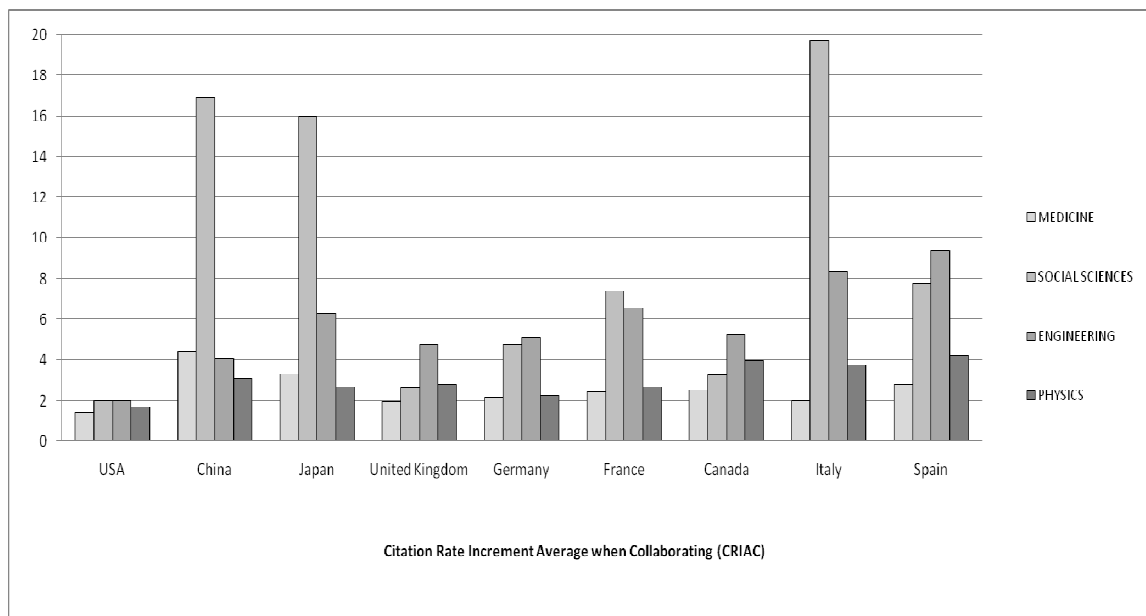
The countries with the greatest DIRCIA –the difference between domestic impact and that received from collaborations– were Spain, Canada, Japan, and Germany. China (the country with least impact) had the lowest DIRCIA.

Specific Results (Scientific Areas)

To check whether or not the countries behaved similarly in all disciplines, we selected four areas to study in depth: Medicine, Social Sciences, Engineering, and Physics. It must be said that the areas of Social Sciences and Engineering had the lowest total impact of all (2.05 and 2.10, respectively), while Medicine had a total impact of 5.43 and Physics 5.86.

With all four of these scientific disciplines, there was a positive Citation Rate Increment from the Collaborator (CRIC), as in the general case.

Fig. 4. Citation Rate Increment Average when Collaborating (CRIAC) of the 9 countries with the greatest production in 2004 in the Medicine, Social Sciences, Engineering, and Physics specific areas.



However, some countries had a greater Citation Rate Increment Average when Collaborating (CRIAC) towards their collaborating countries in some disciplines than in the general case. Indeed, in the Social Sciences, three countries more than tripled the increment of the general case – Italy (19.6), China (16.5), and Japan (15.9). The situation of Spain (7.7) was similar to that of the above three countries since, in this specific area, its CRIAC value was significantly higher than the general case. In Engineering, Spain (9.4), Italy (8.3), and France (6.5) almost doubled the increment of the general case, i.e., they had a high impact in this discipline compared with other countries. In the areas of Physics and Medicine, the countries deviated little from the general case – their citation rates to a country increased some fourfold when collaborating.

In the case of the USA, there were no differences between the general case and the different specific disciplines. In both cases, the citation rate to a country increased by about 2 times (on average) with collaboration. It was not only the country with the highest total impact in general, but also in all four specific disciplines studied except Social Sciences, in which it ranked second after the UK.

For some countries, the increment they provided to their collaborators in the area of Social Sciences was almost double that of Engineering, and five times higher in the cases of Physics or Medicine.

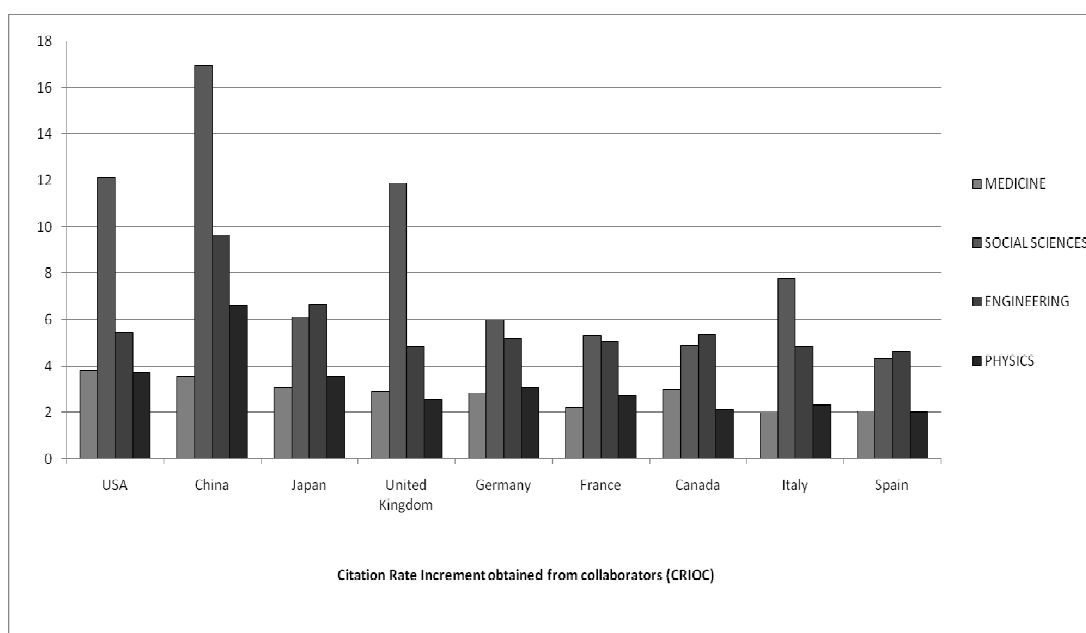
The countries which provided the greatest values of Citation Rate Increment Average when Collaborating (CRIAC) in Social Sciences and Engineering were the same four countries giving the greatest increment in the general studies – China, Japan, Italy, and Spain. While Italy and Spain were two of the countries with the greatest impact within the area of Engineering, the low average impact of this discipline compared with Medicine and Physics meant that the actual value of the impact of Italy and Spain in Engineering was less than in Medicine and Physics. Indeed, as was noted above, Physics and Medicine have greater impact than Social Sciences and Engineering, so that in general the impact would be lower in all countries in these last two disciplines.

As was the case in the general study, the correlations between the Citation Rate Increment Average when Collaborating (CRIAC) and impact were positive or negative but of little significance in the four disciplines.

There were negative correlations in the case of Medicine (-0.89) and Social Sciences (-0.55), and non-significant correlations in Physics (-0.09) and Engineering (0.08), while in the general study the correlation coefficient was -0.64 with a standard deviation of -0.71. The areas closest to the overall correlation were Social Sciences and Medicine.

In the different scientific disciplines, there were collaborating countries that received a greater Citation Rate Increment Obtained from Collaborators (CRIOC) than others.

Fig. 5. Citation Rate Increment Obtained from Collaborators (CRIOC) of the 9 countries with the greatest production in 2004 in the Medicine, Social Sciences, Engineering, and Physics specific areas.



In Social Sciences, China (16.9), the USA (12), and the UK (11.8) received an increment from their collaborators that almost tripled that found in the general study. In Engineering, the values for China (9.6) and Japan (6.6) almost doubled those of the general study, perhaps partially because they had a lower impact in this discipline than overall. Spain received less CRIOC from other countries in all four disciplines, although in the Social Sciences and Engineering its CRIOC exceeded that of the general study (as in the previous case, this may have been in part because these were the two disciplines for which it had the lowest impact).

The areas of Medicine and Physics were those with the lowest values of CRIOC, even lower than in the general study. The only exception was China in Physics, with a value that surpassed that of the general study. Indeed, for China the impact in Physics exceeded that of the general studies, while the opposite was the case for the other three scientific areas considered.

There were negative correlations between the values of CRIOC and impact in the areas of Engineering (-0.88) and Physics (-0.74), as had been the case in the general study (-0.51). There were no significant correlations in Medicine or Social Sciences.

There were strong correlations between the Citation Rate Increment Average when Collaborating (CRIAC) and the Citation Rate Increment Obtained from Collaborators (CRIOC) in all four disciplines, as had also been the case in the general study.

For all four disciplines, the Domestic Citation Rate (DCR) was lower with collaboration.

Fig. 6. Domestic Citation Rate Comparison Average (DCRCA) of the 9 countries with the greatest production in 2004 in the Medicine, Social Sciences, Engineering, and Physics specific areas.

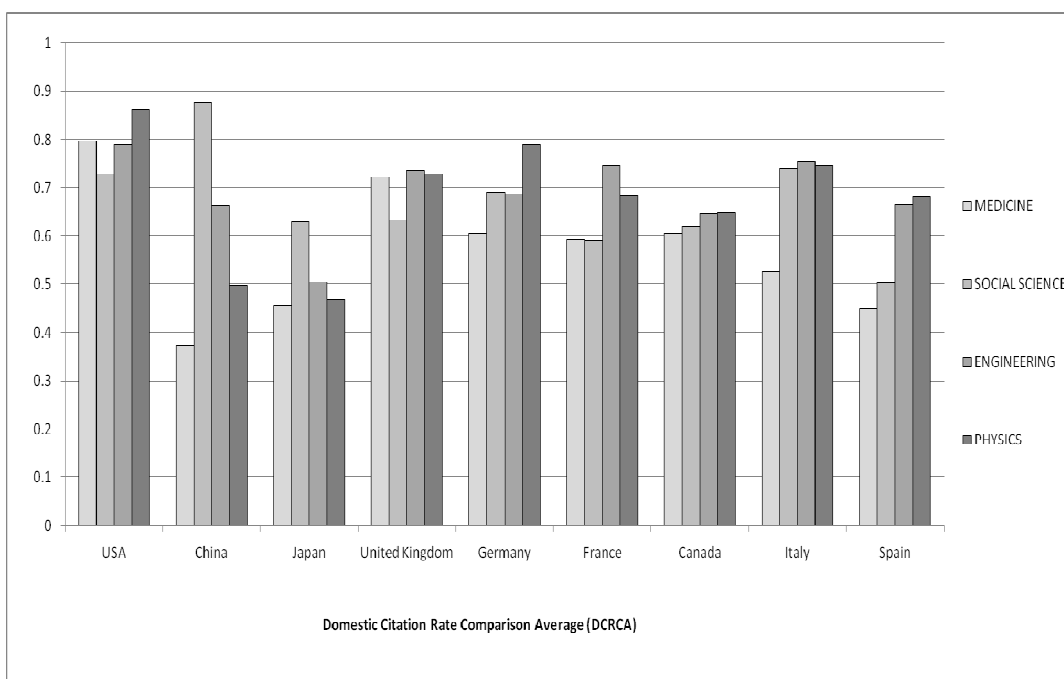


Figure 6 shows the Domestic Citation Rate Comparison Average (DCRCA) for the countries of Group L in the four disciplines. In all cases, the collaborators had a greater citation rate to their domestic production than to their documents in collaboration, as we had also found to be the case in the general study.

However, some countries came closer to equality in some disciplines than in others. In particular, the USA was the country that most closely approached equality in almost all the disciplines, as had also been the case in the general study. This may have been influenced by its being the country with the greatest total impact.

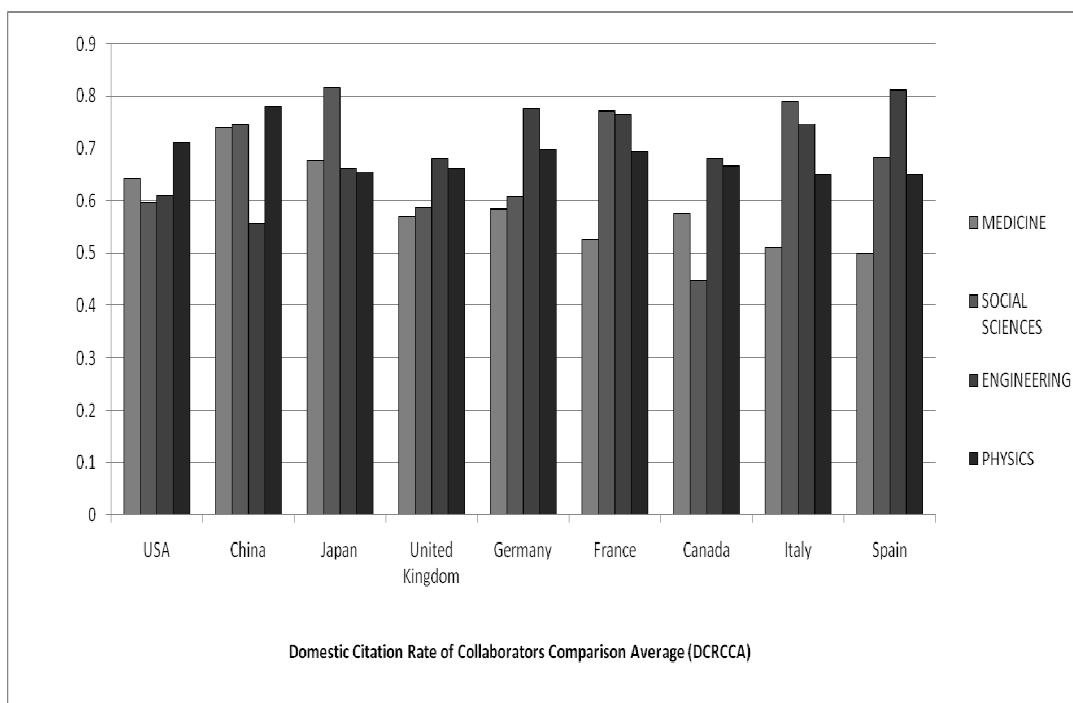
In Social Sciences, it was China rather than the USA which came closest to equality, while, in the field of Medicine, it had the lowest DCRCA of all. Indeed, unlike the other countries whose values of DCRCA were more homogeneous, China presented an erratic behaviour between scientific disciplines. Indeed, it had the highest value of all 9 countries in Social Sciences (0.87), and the lowest in Medicine (0.37). Japan was somewhat similar in this sense, standing out as being the only country to present low values in all four disciplines.

The UK and Canada presented quite homogeneous patterns of DCRCA across the four disciplines, although in both countries the values were lower than in the general studies.

The general study showed a fairly strong correlation between the total impact of collaborators and their Domestic Citation Rate Comparison Average (DCRCA) (0.92). The partial study also yielded fairly strong positive correlations between these indicators in the areas of Medicine (0.85) and Physics (0.80). In Engineering, the correlation was acceptable and positive (0.60). In the Social Sciences however, it was weak and negative (-0.27), and was thus clearly not coherent with the other scientific areas studied.

There was little difference between the disciplines in terms of the Domestic Citation Rate of Collaborators Comparison Average (DCRCCA) in any of the Group L countries

Fig. 7. Domestic Citation Rate of Collaborators Comparison Average (DCRCCA) of the 9 countries with the greatest production in 2004 in the Medicine, Social Sciences, Engineering, and Physics specific areas.



The USA and the UK were the countries with the most homogeneous DCRCCA values across the four disciplines. Japan and Italy came close to equality in the Social Sciences, and Spain and Germany in Engineering.

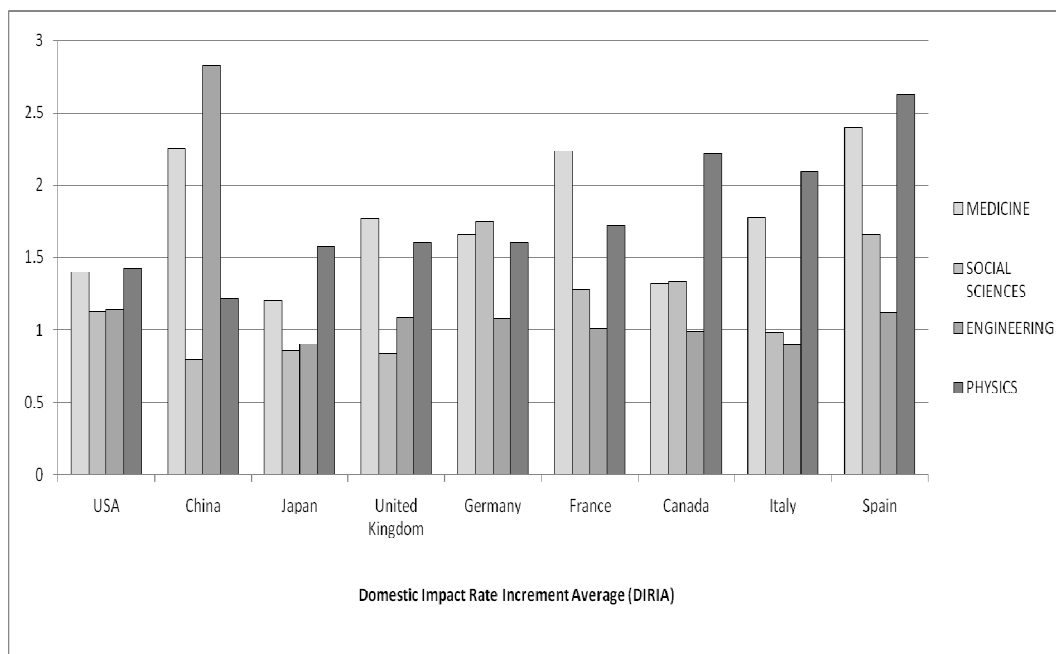
The country with the lowest DCRCCA in Social Sciences was Canada, with a value even lower than it had in the general study. In Engineering and Physics, its DCRCCA values were not at all close to equality, and were fairly similar to each other.

With respect to the correlations of this index (DCRCCA) with impact, three were negative – Medicine (-0.50), Social Sciences (-0.56), and Physics (-0.49). The other, Engineering, was positive (0.45), but not significant. I.e., as was found in the general study, the lower the impact, the closer the Domestic Citation Rate obtained from the collaborators approached the overall value.

There was a strong correlation between Domestic Citation Rate Comparison Average (DCRCA) and Domestic Citation Rate of Collaborators Comparison Average (DCRCCA).

In all four scientific disciplines, the Domestic Impact was greater when there was collaboration. Therefore, as was also found in the general study, if the DCR is less when there is collaboration and the DIR is greater then, in all the disciplines, a given country gains more from other countries even from those with which it does not collaborate.

Fig. 8. Domestic Impact Rate Increment Average (DIRIA) of the 9 countries with the greatest production in 2004 in the Medicine, Social Sciences, Engineering, and Physics specific areas.

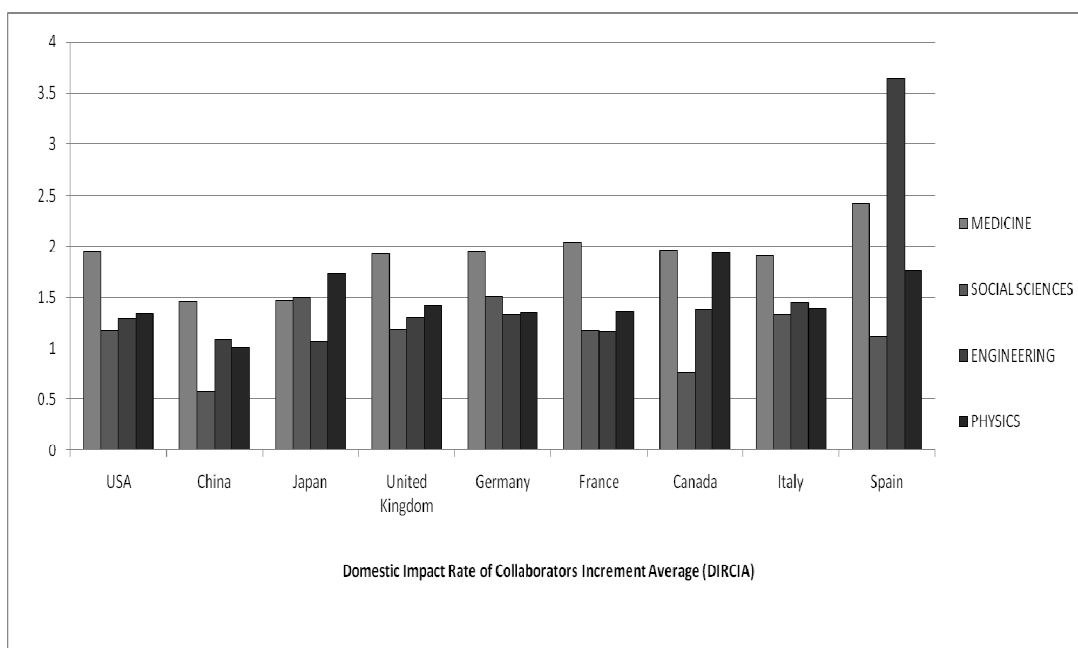


With respect to the Domestic Impact Rate Increment Average (DIRIA) for the different scientific disciplines, China had the highest value of all 9 countries in Engineering (2.82), and the lowest in Social Sciences (0.79). The USA showed a fairly uniform behaviour over all four disciplines, with low values not exceeding 1.5. The highest value in Physics (2.6) corresponded to Spain.

The areas of Medicine and Physics generally gave the highest values of DIRIA, the exceptions being China, Germany, and Canada.

One recalls that in the general study the correlation between impact and Domestic Impact Rate Increment Average (DIRIA) was negative (-0.31) but not statistically significant. The case was similar in the four scientific disciplines studied – the correlations were negative in the area of Medicine (-0.50), Social Sciences (-0.25), and Engineering (-0.73), and, although positive in Physics (0.39), this was not significant.

Fig. 9. Domestic Impact Rate of Collaborators Increment Average (DIRCIA) of the 9 countries with the greatest production in 2004 in the Medicine, Social Sciences, Engineering, and Physics specific areas.



In each of the four scientific disciplines examined, as well as in the general study, Spain was the country which was furthest from equality between the impact that

its collaborators gave its documents in collaboration and the impact given to its own production.

There were strong correlations in all the disciplines between the Domestic Impact Rate Increment Average (DIRIA) and the Domestic Impact Rate of Collaborators Increment Average (DIRCIA).

In the general study, it was observed that the greater the impact, the greater the difference that other countries tended to make of the collaborations relative to the production. The correlation coefficient was found to be 0.50. The partial studies were generally coherent with this trend, with the correlations being positive in Medicine (0.49), Engineering (0.27), and Physics (0.43), although for Social Sciences the correlation was weak and negative (- 0.30). This is thus further evidence that, except in the Social Sciences, the greater the impact the less tends to be the difference between the domestic impact and that given to collaborators, and the greater tends to be the difference that the other countries make of the collaborations with respect to their production.

Interpretation

When a country is involved in collaborations, it receives a positive Citation Rate Increment from the Collaborator (CRIC). In all cases, a greater citation increment was received from collaborating countries than from non-collaborating countries. But some countries had higher values of the Citation Rate Increment Average when Collaborating (CRIAC) (to their collaborating countries) than others, and there were also countries receiving a greater Citation Rate Increment Obtained

from Collaborators (CRIOC) than others. There seemed to be a tendency for the countries with greater impact to have a smaller increment from collaborating.

The Domestic Citation Rate was lower with collaboration, although some countries came very close to equality in the Domestic Citation Rate Comparison Average (DCRCA) and the Domestic Citation Rate of Collaborators Comparison Average (DCRCCA). Typically, it was the countries with high impact that most approached equality in the DCRCA, while the trend in the values of the DCRCCA was for countries with less impact to have greater values.

The Domestic Impact Rate was greater with collaboration, but the greater the impact the smaller the difference between the total Domestic Impact and that given to collaborators (DIRIA), and the greater the difference between the total Domestic Impact and that received from other collaborating countries (DIRCIA). Therefore, if the DCR is lower when there is collaboration (see below), the reason is that the higher DIR naturally leads to an increase also in non-domestic citations, thus diluting the proportion of domestic citations.

In the four scientific disciplines selected for in-depth study, in all there was a positive Citation Rate Increment from the Collaborator (CRIC). But in Social Sciences, some countries had a Citation Rate Increment Average when Collaborating (CRIAC) which was greater than that of the general case by a factor of three, and in Engineering by a factor of nearly two. Indeed, the areas of Social Sciences and Engineering were those in which collaborating countries received greater Citation Rate Increments Obtained from their Collaborators (CRIOC),

while the areas of Physics and Medicine received lower CRIOC increments, even less than in the general studies. There seemed to be a trend in these countries for the increment to be less for collaborations in the disciplines of greater impact.

In all four disciplines studied, the Domestic Citation Rate (DCR) was less when collaborating. There were differences between countries, however, in that some came closer to equality in some disciplines than in others. In the Social Sciences for example, China led the USA in coming closer to equality. Nevertheless, one must bear in mind that China had a low impact in all scientific disciplines, especially in Social Sciences and Engineering. It seems that the countries with less impact are those that least approach equality. Canada was an exception in Social Sciences, being a country with a high impact in this discipline, but having the lowest value of DCRCCA – lower even than its value in the general study.

Finally, in all four scientific disciplines the Domestic Impact was greater with collaboration. Except for the case of the USA which had a fairly uniform distribution of values, the greatest values of DIRIA generally corresponded to the disciplines of Physics and Medicine. The country which had values of DIRCIA most diverging from unity, and which was also the furthest from unity in the general study, was Spain.

Acknowledgments

This work was financed by the Junta de Extremadura e Consejería de EducaciónCiencia&Tecnología and the Fondo Social Europeo as part of the pre-doctoral studentship PRE07052 and the research group grant GR10019, and by the Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica 2008e2011 and the Fondo Europeo de Desarrollo Regional (FEDER) as part of research projects TIN2008-06514-C02-01 and TIN2008-06514-C0202.

References

Abt, H. (2007). The frequencies of multinational papers in various sciences. *Scientometrics*, 72 (1), 105–115.

Aksnes, D. (2003). Characteristics of highly cited papers. *Research Evaluation*, 12(3), 159–170.

Ardanuy, J. (2012). Scientific collaboration in Library and Information Science viewed through the Web of Knowledge: the Spanish case. *Scientometrics*, 90 (3), 877-890.

Ardanuy, J., Urbano, C., & Quintana, L. (2009). A citation analysis of Catalan literary studies (1974-2003): Towards a bibliometrics of humanistic studies in minority languages. *Scientometrics*, 81(2), 347–366.

Bandyopadhyay, A. K. (2004). Authorship collaboration in physics philosophy and political science, IASLIC National Seminar, 11th, Kolkata, 2004, 405-409.

Bandyopadhyay, A.K. (2001). Authorship pattern in different disciplines. *Annals of Library and Information Studies*, 48(4), 139-147.

Bridgstock, M. (1991). The quality of single and multiple authored papers- an unresolved problem. *Scientometrics*, 21 (1), 37-48.

Chinchilla-Rodríguez, Z., Vargas-Quesada, B., Hassan-Montero, Y., González-Molina, A., & MoyaAnegón, F. (2010). New approach to the visualization of international scientific collaboration. *Information Visualization*, 9(4), 277–287.

Ding, Y., Foo, S., & Chowdhury, G. (1999). A bibliometric analysis of collaboration in the field of information retrieval. *The International Information & Library Review*, 30 (4), 367-376.

Figg, W. D., Dunn, L., Liewehr, D. J., Steinberg, S. M., Thurman, P. W., Barrett, J. C., & Birkinshaw, J. (2006). Scientific collaboration results in higher citation rates of published articles. *Pharmacotherapy*, 26 (6), 759–767.

Franceschet, M., & Costantini, A. (2010). The effect of scholar collaboration on impact and quality of academic papers. *Journal of Informetrics*, 4 (4), 540–553.

Garfield, E. (1979). Is citation analysis a legitimate evaluation tool?. *Scientometrics*, 1(4), 359-375.

Gazni, A. & Didegah, F. (2011). Investigating different types of research collaboration and citation impact: a case study of Harvard University's publications. *Scientometrics*, 87 (2), 251-265.

Glänzel, W. (2002). Coauthorship patterns and trends in the sciences: A bibliometric study with implications for database indexing and search strategies, 1980-1998. *Library Trends*, 50(3), 461– 473.

Glänzel, W. (2001). National characteristics in international and scientific co-authorship relations. *Scientometrics*, 51 (1), 69–115.

Glänzel, W., & Schubert, A. (2001). Double effort = Double impact? A critical view at international coauthorship in chemistry. *Scientometrics*, 50(2), 199–214

Glänzel, W., Schubert, A., & Czerwon, J. (1999). A bibliometric analysis of international scientific cooperation of the European Union (1985-1995). *Scientometrics*, 45 (2), 185–202.

Goldfinch S., Dale, T., & De Roue, K. (2003). Science from the periphery: Collaboration network and «Pheriphery effects» in the citation of New Zeland Crown Research Institutes articles, 1992-2000. *Scientometrics*, 57 (3), 321-337.

Hsu, J.W., & Huang, D.W. (2010). Correlation between impact and collaboration. *Scientometrics*, 86(2), 317–324.

Inzelt, A., Schubert, A., & Schubert, M. (2009). Incremental citation impact due to international co authorship in Hungarian higher education institutions. *Scientometrics*, 78 (1), 37-43.

Katz, J. S. (1994). Geographical proximity and scientific collaboration. *Scientometrics*, 31 (1), 31-43.

Katz, J., & Hicks, D. (1997). How much is a collaboration worth? A calibrated bibliometric model. *Scientometrics*, 40(3), 541–554.

King, D. A. (2004). The scientific impact of nations. *Nature*, 430, 311–316.

Kraut R., Egidio C., & Galeguer, J. (1988). Patterns of contact and communication in scientific research collaboration. 'CSCW '88: Proceedings of the 1988 ACM conference on Computer-supported cooperative work', ACM, New York, NY, USA, 1-12.

Lancho-Barrantes, B. S., Guerrero-Bote, V. P., Chinchilla-Rodríguez, Z., & Moya-Anegón, F. (2012). Citation flows in the zones of influence of scientific collaborations. *Journal of the American Society for Information Science and Technology*, 63(3), 481-489.

Leimu, R., & Koricheva, J. (2005). Does scientific collaboration increase the impact of ecological articles? *Bioscience*, 55 (5), 438–443.

Leta, J., & Chaimovich, H. (2002). Recognition and international collaboration: the Brazilian case. *Scientometrics*, 53 (3), 325–335.

Lewinson, G., & Cunningham, P. (1991). Bibliometric studies for the evaluation of trans-domestic research. *Scientometrics*, 21(2), 223–244.

Luukkonen, T., Tijssen, R. J. W., Persson, O., & Sivertsen, G. (1993). The measurement of international scientific collaboration. *Scientometrics*, 28(1), 15-36.

Ma, N., & Guan, J. C. (2005). An exploratory study on collaboration profiles of Chinese publications in Molecular Biology. *Scientometrics*, 65 (3), 343–355.

Moed, H. F., De Bruin, R. E., Nederhof, A. J., & Tijssen, T. J. W. (1991). International scientific co operation and awareness within the European Community: Problems and perspectives. *Scientometrics*, 21(3), 291-311.

Molteni, V., & Zulueta, M. A. (2002). Análisis de la visibilidad internacional de la producción científica argentina en las Bases de datos Social Science Citation Index y Arts and Humanities Citation Index de 1990–2000: estudio bibliométrico. *Revista Española de Documentación Científica*, 25(4), 455–465.

Moya-Anegón, F., Chinchilla-Rodríguez, Z., Corera-Álvarez, E., González-Molina, A., Hassan-Montero, Y., & Vargas-Quesada, B. (2008). Indicadores bibliométricos de la actividad científica española: 2002–2006. Madrid: FECYT.

Moya-Anegón, F., Chinchilla-Rodríguez, Z., Vargas-Quesada, B., Corera-Álvarez, E., Muñoz-Fernández, F. J., González-Molina, A., & Herrero-Solana, V. (2007). Coverage analysis of Scopus: A journal metric approach. *Scientometrics*, 73(1), 53–78.

Narin, F., Stevens, K., & Whitlow, E. (1991). Scientific cooperation in Europe and the citation of multidomestically authored papers. *Scientometrics*, 21(3), 313–323.

Pao, M. L. (1981). Coauthorship as communication measure. *Library Research*, 2, 327–338.

Persson, O., Glänzel, W., & Danell, R. (2004). Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. *Scientometrics*, 60(3), 421–432.

Price, D. J.S., & Beaver, D.B. (1966). Collaboration in an invisible college. *American psychologist*, 21(11), 1011–1018.

SCImago Journal & Country Rank.(2008). SCImago Research Group. Accessed 26 March, 2009 from <http://www.scimagojr.com/>

Singh, J. (2005). Collaborative networks as determinants of knowledge diffusion patterns. *Management Science*, 51 (5), 756–770.

Sooryamoorthy, R. (2009). Do types of collaboration change citation? Collaboration and citation patterns of South African science publications. *Scientometrics*, 81(1), 177–193.

Sudhier Pillai, K.G. (2007). Authorship patterns in physics literature: An informetric study on citations in doctoral theses of the Indian Institute of Science. *Annals of Library and Information Studies*, 54, 90-94.

Zhao, Q., & Guan, J. (2011). International collaboration of three ‘giants’ with the G7 countries in emerging nanobiopharmaceuticals. *Scientometrics*, 87 (1), 159-170.

Zuckerman, H. A. (1968). Patterns of name ordering among authors of scientific papers: A study of social symbolism and its ambiguity. *American Journal of Sociology*, 74, 276-291.

9 Resúmenes en español de los Artículos

A1: The Iceberg Hypothesis revisited

Bárbara S. Lancho Barrantes, Vicente P. Guerrero Bote and Félix de Moya

Anegón

Publicado en *Scientometrics* (2010) 85:443–461

En el estudio se analizan las distribuciones Rango / Impacto de la base de datos Scopus que tiene una gran cobertura, utilizando datos recientes y una ventana de la citación de tres años. En este estudio también se incluye una comparación con un estudio anterior de las categorías del Journal Citation Report y la determinación de los factores que más influyen en las distribuciones. Si bien todas las áreas temáticas específicas se adaptan a una curva logarítmica negativa bastante bien, las que tienen un mayor Impacto Externo tienen distribuciones con un pico más definido y una cola más larga algo así como un Iceberg. Una fuerte correlación se observó entre los ratios de exportación e importación de conocimiento.

Por último, los datos de ambas bases de datos Scopus e ISI se utilizaron para caracterizar las distribuciones de rango / impacto por área temática.

A2: What lies behind the averages and significance of citation indicators in different disciplines?

Bárbara S. Lancho Barrantes, Vicente P. Guerrero Bote and Félix de Moya Anegón

Publicado en Journal of Information Science June 2010 vol. 36 no. 3 371-382

Las limitaciones de los indicadores basados en citas incluyen una falta de cobertura, una no normalización de relaciones con respecto a la longitud de las listas de referencia (con un posible sesgo en favor de las revisiones), y diferentes hábitos de citación. Como consecuencia, las distribuciones de los indicadores no son comparables entre diferentes disciplinas. Aquí se muestra que los indicadores de citas de revistas más populares utilizados en la evaluación de la calidad - los factores de impacto de revistas de Thomson Scientific y la clasificación científica de las revista de Scopus - están fuertemente correlacionadas con la proporción de referencias dentro de la base de datos, y más aún con el número de referencias dentro de la base de datos de los últimos por el papel. No se encontraron correlaciones significativas con otras magnitudes bibliométricas. Anticipamos que estos resultados serán un punto de partida para más modelos de indicadores más sofisticados que tengan esto en cuenta, y para el diseño de estrategias destinadas a extender esas bases de datos bibliométricas como Thomson Scientific's Science Citation Index o Scopus de Elsevier para mejorar su capacidad de evaluar toda la Ciencia.

A3: Citation Flows in the Zones of Influence of Scientific Collaborations**Bárbara S. Lancho Barrantes, Vicente P. Guerrero Bote,****Zaida Chinchilla Rodríguez y Félix de Moya Anegón****Publicado en Journal of the American Society for Information Science and Technology. Volume 63, Issue 3, pages 481–489, March 2012**

Con el experimento llevado a cabo se corrobora que las investigaciones que conllevan colaboración internacional tienen un mayor impacto que las que no lo conllevan. Este incremento de citación no solo proviene de los países colaboradores sino que en mayor medida de otros países. Existe una mayor influencia sobre un entorno cercano de las instituciones. Sin embargo no hay sesgo nacional, sino que en aquellos países pequeños, sus fronteras se adaptan más a ese entorno cercano, por lo que la influencia es mayor. En los países grandes las fronteras incluyen un entorno más amplio por lo que se obtienen menores valores de influencia. Los trabajos en colaboración también presentan una mayor influencia en los distintos entornos. Estas influencias están también moduladas por el tamaño científico del país. Curiosamente tienen más influencia los domestic papers en los documentos en colaboración que estos últimos en los primeros. Al contrario tienen más influencia los collaboration papers en los no domestic papers que los no domestic papers en los collaboration papers. En definitiva podemos concluir que la ciencia no tiene fronteras.

Ciertamente hay una mayor influencia en el entorno cercano de los autores, que no tiene por qué coincidir con los entornos nacionales, que se diluye a medida que se amplía dicho entorno. Mismamente en un entorno pequeño se maximiza el sesgo producido por la autocitación.

A4: Citation increments between collaborating countries**Bárbara S. Lancho Barrantes, Vicente P. Guerrero Bote and Félix de Moya****Anegón****Aceptado en *Scientometrics*.****En proceso de publicación.**

La colaboración internacional aumenta la citación (impacto). En todos los casos se recibe mayor incremento de citas de un país con el que se colabora que en el caso de no colaborar con él. Pero algunos países colaboradores tienen mayor incremento (hacia los países colaboradores) que otros y también hay países colaboradores que reciben mayor incremento que otros. Parece que hay una tendencia a que en los países con mayor impacto el incremento al colaborar sea menor.

Por otro lado parece ser que hay una propensión a que los países quieran incrementar más la citación a su producción nacional que su producción en colaboración. Pero aunque ocurre esto, parece que el impacto de los países es mayor cuando se colabora, lo que sucede es que se gana impacto también del resto de naciones.

Finalmente, en las diferentes disciplinas científicas los países se comportan de manera distinta en unas que en otras.