Chapter 2 / Capítulo 2

Applied bibliometrics. From data to publication (Spanish Edition)
ISBN: 978-9915-9680-6-3
DOI: 10.62486/978-9915-9680-6-3.ch02

©2025 The authors. This is an open access article distributed under the terms of the Creative Commons Attribution (CC BY) 4.0 License.

BRIEF HISTORICAL OVERVIEW / BREVE RECUENTO HISTÓRICO

En 1955, un joven químico llamado Eugene Garfield publicó un artículo titulado *"Citation Indexes for Science",* donde proponía un sistema para rastrear el impacto de las investigaciones. Nadie imaginaba que esta idea revolucionaría la ciencia, dando origen a la bibliometría moderna.[1]

2.1. La Era Analógica (1960-1990): Los Cimientos

Science Citation Index (SCI) – 1960

En 1960, el *Institute for Scientific Information* (ISI), fundado por Eugene Garfield, revolucionó la evaluación científica con el lanzamiento del *Science Citation Index* (SCI), el primer índice de citas comercializado a gran escala. Publicado originalmente en formato impreso, este sistema permitió, por primera vez, rastrear de manera sistemática cómo los artículos científicos se vinculaban entre sí a través de sus referencias bibliográficas.[2]

El SCI se basaba en un principio sencillo pero poderoso: ***"las citas son conexiones de conocimiento"***. Su metodología incluía rastreo de referencias cruzadas entre 600 revistas científicas seleccionadas (principalmente de EE.UU. y Europa), indexación inversa (en lugar de solo listar autores o temas, el SCI permitía buscar qué artículos citaban a un trabajo determinado, revelando su influencia posterior), cobertura multidisciplinaria (aunque enfocado en ciencias naturales, sentó las bases para futuros índices en ciencias sociales y humanidades)

El SCI transformó la forma de medir el impacto científico, antes del SCI, la productividad se medía por número de publicaciones. El índice introdujo la idea de que un artículo podía ser influyente incluso si su autor publicaba poco (ej: el paper de Watson y Crick sobre el ADN tuvo pocas publicaciones previas, pero cambió la biología). Identificó trabajos altamente citados que definieron campos enteros (ej: el artículo de 1953 de Miller sobre el origen de la vida). A pesar de sus limitaciones, como el sesgo anglófono, el enfoque occidental y el acceso restringido a instituciones privilegiadas, su influencia perdura a través de plataformas contemporáneas como Web of Science, mientras Google Scholar y Scopus han adoptado su lógica fundamental de citas cruzadas.[3]

El SCI no solo midió la ciencia: la hizo más transparente y conectada. Su historia nos recuerda que hasta las herramientas más disruptivas deben usarse con conciencia crítica.

Dato curioso:

> *El SCI original ocupaba 5 metros lineales de estanterías y solo instituciones élite como Harvard podían costearlo.*[4]

Leyes Fundamentales (Años 70-80). Los pilares matemáticos de la bibliometría

Durante las décadas de 1970 y 1980, la bibliometría consolidó su rigor científico gracias a la validación y aplicación masiva de dos leyes empíricas formuladas décadas antes: la Ley de Lotka (1926) y la Ley de Bradford (1934). Estas leyes, aunque desarrolladas inicialmente para describir patrones en la literatura científica, se convirtieron en herramientas esenciales para entender la productividad de los investigadores y la distribución del conocimiento en las revistas académicas.

Ley de Lotka: la desigualdad en la productividad científica

La Ley de Lotka, formulada por Alfred J. Lotka en 1926 pero masificada durante el auge cienciométrico de los años setenta, constituye uno de los hallazgos fundamentales sobre la distribución desigual de la productividad científica. Este principio establece que aproximadamente

el 10 % de los investigadores produce el 50 % de las publicaciones académicas, revelando un patrón de concentración del output científico que trasciende disciplinas y períodos históricos. La ley representa la aplicación específica del principio de Pareto al dominio científico, confirmando que la producción de conocimiento sigue distribuciones asimétricas donde pocos actores generan la mayor parte de los resultados investigativos.[5]

La evidencia empírica contemporánea valida persistentemente esta distribución desigual a través de múltiples campos del conocimiento. Esta distribución potencia el «efecto Mateo» en ciencia, donde investigadores establecidos con acceso a redes colaborativas, recursos institucionales y capital simbólico acumulado tienden a publicar con mayor frecuencia, reforzando así sus ventajas competitivas iniciales.

Las aplicaciones prácticas de la Ley de Lotka en evaluación académica son múltiples y significativas. Permite identificar investigadores altamente productivos para decisiones de contratación, promoción o asignación de financiamiento, proporcionando un punto de referencia cuantitativo para evaluar desempeños excepcionales. Simultáneamente, sirve como herramienta diagnóstica para detectar desigualdades estructurales en sistemas científicos específicos, informando políticas destinadas a democratizar las oportunidades de publicación y visibilidad académica.

Sin embargo, la ley presenta limitaciones metodológicas importantes en contextos contemporáneos. Su aplicabilidad resulta más robusta en disciplinas STEM que en humanidades, donde predominan la autoría única y formatos de publicación diversos que escapan a las métricas tradicionales. Tampoco anticipó la explosión de coautorías masivas característica de la ciencia de grandes equipos, donde publicaciones en genómica o física experimental pueden incluir miles de autores, redistribuyendo las dinámicas de productividad individual. Estas limitaciones subrayan la necesidad de contextualizar la ley dentro de los cambios paradigmáticos en las prácticas de autoría y colaboración científica del siglo XXI.

Ley de Bradford: el núcleo de las revistas clave

La Ley de Bradford (1943) representa un principio fundamental en bibliometría que describe la distribución desigual de literatura científica relevante entre las revistas académicas. Establece que un reducido núcleo de publicaciones periódicas concentra la mayoría de los artículos significativos en cualquier campo del conocimiento, mientras una periferia más extensa alberga contribuciones dispersas. Este patrón de concentración refleja cómo se estructura la comunicación especializada en la ciencia contemporánea, donde ciertas revistas funcionan como canales privilegiados para la diseminación de investigación de alto impacto.[6] La evidencia empírica confirma consistentemente este fenómeno de concentración entre múltiples disciplinas.

Las aplicaciones prácticas de esta ley son particularmente valiosas para la gestión de recursos de información. Las bibliotecas académicas utilizan análisis de zonas de Bradford para optimizar suscripciones, focalizando recursos limitados en el núcleo de revistas que maximiza el acceso a literatura relevante. Simultáneamente, la identificación de “zonas desérticas”, campos donde el conocimiento está ampliamente disperso entre numerosas fuentes, alerta sobre la necesidad de estrategias de búsqueda más comprehensivas, especialmente en áreas interdisciplinarias o estudios culturales específicos.

Sin embargo, la Ley de Bradford enfrenta limitaciones significativas en el ecosistema científico contemporáneo. Presenta un marcado sesgo anglófono, la mayoría de los artículos

publicados en las grandes bases de datos son en inglés, marginando contribuciones valiosas en otros idiomas. Además, la emergencia de la ciencia abierta y los repositorios de *preprints* está transformando sustancialmente estos patrones de concentración. En inteligencia artificial, por ejemplo, arXiv ha afectado el modelo tradicional al crear un canal alternativo de diseminación que compite con las revistas establecidas, demostrando la creciente fluidez en los patrones de comunicación científica.

Esta evolución hacia modelos más distribuidos sugiere que, aunque el principio de concentración sigue siendo relevante, sus manifestaciones concretas se están transformando rápidamente. La bibliometría contemporánea debe desarrollar herramientas capaces de capturar estas nuevas dinámicas mientras mantiene la utilidad analítica del concepto original de Bradford para comprender la estructura de la comunicación científica.

Limitaciones generales de estas leyes

Aunque revolucionarias, ambas leyes tienen problemas en la era digital:

1. Cobertura sesgada: Se basaron en datos de EE.UU. y Europa, ignorando la producción científica de Asia, África y América Latina. Ejemplo: La Ley de Bradford no predice bien el núcleo de revistas en agronomía africana, donde la investigación se publica en revistas locales.
2. Procesamiento manual: En los 70-80, los conteos se hacían a mano, lo que generaba errores (ej: duplicación de autores con nombres similares).
3. Contexto actual: Hoy, algoritmos como los de Scopus o Dimensions permiten análisis más dinámicos, pero las leyes siguen siendo la base teórica.

¿Por qué siguen importando en el siglo XXI?

Estas leyes bibliométricas históricas mantienen una sorprendente relevancia en el ecosistema científico digital del siglo XXI, demostrando que los patrones fundamentales de la comunicación académica trascienden los cambios tecnológicos. La Ley de Lotka continúa informando el desarrollo de modernos sistemas de evaluación, donde el Índice h y sus variantes incorporan su comprensión sobre la distribución desigual de la productividad científica. Esta perspectiva contextualiza métricas individuales dentro de patrones sistémicos más amplios, previniendo interpretaciones simplistas del desempeño investigador y reconociendo la naturaleza inherentemente asimétrica de la producción de conocimiento.

En el ámbito de las políticas científicas, la Ley de Bradford proporciona un marco analítico crucial para navegar la expansión exponencial de la comunicación académica. Su principio de concentración ayuda a identificar revistas depredadoras que operan fuera de los núcleos legítimos de cada disciplina, ofreciendo un criterio objetivo para distinguir canales de comunicación confiables. Esta aplicación resulta particularmente valiosa en contextos de acceso abierto, donde la proliferación de editoriales cuestionables requiere mecanismos robustos para garantizar la integridad de la comunicación científica.

La influencia más profunda de estas leyes se manifiesta en los fundamentos algorítmicos que sustentan las herramientas digitales contemporáneas. El algoritmo PageRank de Google Scholar, por ejemplo, implementa computacionalmente el principio de Bradford al asignar mayor peso a las citas provenientes de fuentes consideradas “núcleos” de autoridad académica. Simultáneamente, los sistemas de recomendación y descubrimiento de literatura científica incorporan entendimiento derivado de Lotka para priorizar contenido de investigadores altamente productivos e influyentes.

La vigencia de estos principios demuestra que, aunque las tecnologías de comunicación científica han experimentado transformaciones radicales, los patrones subyacentes de producción y diseminación del conocimiento mantienen regularidades estructurales observables. Esta continuidad convierte a las leyes bibliométricas clásicas en herramientas esenciales para comprender tanto las permanencias como las transformaciones en la dinámica científica contemporánea, proporcionando un marco interpretativo para navegar la complejidad del actual ecosistema de investigación.

2.2. La Revolución Digital (1990-2010): expansión y críticas

La Era en que la Bibliometría se Hizo Global (y Polémica)

La década de 1990 marcó el inicio de la bibliometría digital, transformando radicalmente cómo se mide y gestiona la ciencia. Con la llegada de Internet, los índices de citas dejaron atrás los formatos impresos y adoptaron plataformas en línea, ampliando su alcance, pero también generando nuevos desafíos éticos y metodológicos.

Web of Science (WoS) – 1997: El SCI se vuelve digital

Fue la versión digital del *Science Citation Index (SCI)*, lanzada por el *Institute for Scientific Information* (ISI). Permitió por primera vez búsquedas en tiempo real y análisis avanzados de citaciones. Pasó de 600 revistas (en papel) a 8 000 revistas indexadas, incluyendo ciencias sociales (Social Sciences Citation Index) y artes (Arts & Humanities Citation Index). Introdujo funciones como *Citation Reports* (para calcular el impacto de autores) y *Journal Impact Factor (JIF) en línea* (antes solo disponible en el *Journal Citation Reports* impreso).[7]

El JIF, creado en 1975 para evaluar revistas, comenzó a aplicarse erróneamente a investigadores individuales. Universidades exigían publicar en "revistas Q1" para contrataciones, ignorando la calidad real de los artículos.

Scopus (2004): el competidor que desafió la hegemonía

Desarrollado por Elsevier como una alternativa directa al Web of Science (WoS), Scopus amplió el horizonte de la indexación científica al incorporar una mayor cantidad de revistas no anglófonas, con un fuerte enfoque en publicaciones europeas y asiáticas. Entre sus aportes más relevantes, introdujo los perfiles de autor, sistemas pioneros en la desambiguación de nombres que anticiparon herramientas como ORCID. También incorporó métricas alternativas como el *CiteScore*, concebido como un indicador complementario al *Journal Impact Factor* (JIF).

Sin embargo, Scopus no estuvo exento de críticas. Se le ha señalado por presentar un sesgo comercial, favoreciendo en sus rankings a publicaciones asociadas a Elsevier. Además, su negativa a indexar *preprints* lo situó en desventaja frente a sistemas más abiertos como Google Scholar, al excluir de su base de datos una parte significativa de la literatura científica informal o "gris".

Google Scholar (2004): la disrupción democrática

El lanzamiento de Google Scholar representó una transformación profunda en el acceso al conocimiento científico. Su sistema de indexación automática permitió incorporar no solo artículos revisados por pares, sino también *preprint*s, como los de arXiv, libros, tesis y documentos técnicos, muchos de los cuales habían permanecido fuera del radar de las bases tradicionales. A diferencia de WoS y Scopus, Google Scholar se ofreció de forma gratuita, lo que democratizó el acceso a la literatura científica global. Su capacidad de capturar contenidos en idiomas y formatos tradicionalmente marginados, como estudios en español de América Latina, le otorgó una cobertura sin precedentes. Además, con el desarrollo de *Google Scholar Citations*,

los investigadores pudieron crear perfiles públicos para visibilizar su producción académica y sus métricas personales.

No obstante, esta apertura también generó controversias. Uno de los principales cuestionamientos ha sido la opacidad de sus algoritmos, ya que no se informa cómo se jerarquizan los resultados, lo que ha dado lugar a estrategias para manipular la visibilidad de ciertos trabajos mediante el uso intencional de palabras clave. Por otro lado, la calidad de los documentos indexados es motivo de debate, dado que Google Scholar incluye revistas depredadoras y materiales sin revisión por pares.

Impacto y críticas de esta revolución

La irrupción de plataformas como Scopus y Google Scholar marcó un antes y un después en la forma en que se accede, evalúa y difunde el conocimiento científico. Entre sus principales legados se encuentra la globalización de la ciencia, al otorgar visibilidad a investigaciones provenientes de regiones históricamente marginadas, como América Latina, Asia o África, más allá de los centros tradicionales de producción académica en Estados Unidos y Europa. Además, la digitalización masiva de citas permitió el desarrollo de nuevas métricas, siendo el índice h, propuesto por Hirsch en 2005, una de las más influyentes al medir simultáneamente la productividad y el impacto de un investigador.

No obstante, persisten problemas estructurales que estas plataformas no han logrado resolver. A pesar de las críticas acumuladas, el factor de impacto (JIF) continúa siendo utilizado como criterio decisivo en procesos de evaluación y financiamiento, como sucede en agencias nacionales como la ANID en Chile.[8] Esta persistencia refuerza una lógica cuantitativa que muchas veces ignora la calidad o el contexto del conocimiento producido. Por otro lado, la brecha digital sigue siendo un obstáculo importante: muchas instituciones en países de ingresos bajos o medianos no pueden costear el acceso a bases de datos comerciales como WoS o Scopus, lo que las obliga a depender de alternativas gratuitas como Google Scholar, con todas las limitaciones y riesgos que esto implica en términos de confiabilidad. Finalmente, el auge de las herramientas digitales ha dado lugar a nuevas formas de manipulación de métricas, como el citation stacking, una práctica que consiste en crear círculos de citas fraudulentas para inflar artificialmente el impacto de ciertas publicaciones, distorsionando así los procesos de evaluación académica.

2.3 La Era Moderna (2010-Presente): complejidad y democratización

Altmetría(2010)

A partir de 2010, la bibliometría comenzó a expandirse hacia formas más amplias de evaluación del impacto científico, dando origen a la altmetría. Este enfoque busca medir el impacto más allá del conteo de citas académicas, incorporando indicadores como menciones en redes sociales, medios de comunicación, blogs, Wikipedia y documentos de política pública. El lanzamiento de Altmetric.com en 2011 marcó un punto de inflexión, al ofrecer herramientas capaces de rastrear en tiempo real la circulación digital de artículos científicos.

Bibliometría Open Source(2017)

En 2017, la democratización de la bibliometría dio un salto significativo con la aparición de herramientas de código abierto, que facilitaron el acceso a análisis avanzados sin los altos costos asociados a plataformas comerciales. Entre las más destacadas se encuentran Bibliometrix, un paquete desarrollado en R que permite realizar análisis bibliométricos sofisticados, y VOSviewer, una herramienta ampliamente utilizada para la visualización de redes de coautoría, co-citación o términos clave. Estas plataformas permitieron que investigadores de instituciones con recursos limitados pudieran acceder a metodologías de análisis robustas, contribuyendo a

una descentralización del conocimiento bibliométrico.

Inteligencia Artificial y Minería de Textos(2020)

En la década de 2020, la integración de la inteligencia artificial revolucionó la manera en que se procesa y analiza la literatura científica. Modelos como GPT-4 abrieron la posibilidad de resumir tendencias a partir de millones de *abstracts* de forma automatizada y eficiente, permitiendo un análisis sintético de campos de investigación completos en tiempo récord. Paralelamente, plataformas como Dimensions.ai comenzaron a integrar bases de datos que conectan publicaciones científicas con patentes, subvenciones y resultados de investigación, ofreciendo una perspectiva mucho más completa del ciclo de producción y transferencia del conocimiento. Esta convergencia entre IA, minería de textos y ciencia abierta ha ampliado el alcance y la profundidad de la bibliometría contemporánea, posicionándola como una herramienta clave en la toma de decisiones científicas y de política pública.

Reto actual:

> *Cómo evitar que la IA genere "artículos zombis" (textos bien citados, pero sin sustento real).*

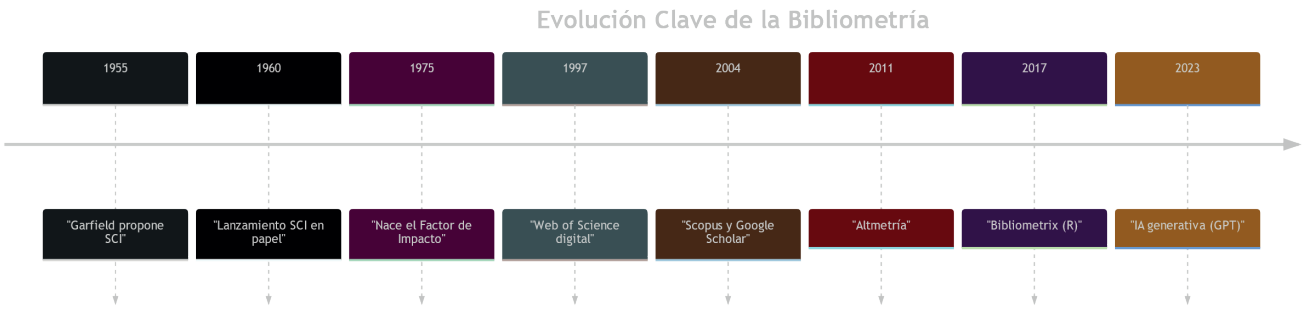

Figura 2.1. Línea de tiempo visual

2.4 Lecciones Históricas

1. De lo analógico a lo digital: Los saltos tecnológicos han democratizado el acceso, pero introducido nuevos sesgos (ej: sobrecarga informativa).
2. De lo cuantitativo a lo cualitativo: Las críticas al FI llevaron a métricas responsables (DORA, Leiden).
3. De lo académico a lo social: La altmetría amplió la noción de impacto científico.

"La bibliometría ya no es solo contar citas, sino entender cómo el conocimiento fluye —y a veces se estanca— en la sociedad."

La Sección II explorará las herramientas más usadas para la construcción de bibliometrías así cómo obtener los datos para crear los estudios bibliométricos, de una manera práctica.

Recapitulando

- La bibliometría tiene sus orígenes en la primera mitad del siglo XX, cuando surgieron los primeros intentos de cuantificar la producción científica.
- El término "bibliometría" fue acuñado por Paul Otlet (1934) y consolidado por Alan Pritchard (1969).

- Su desarrollo estuvo influenciado por los avances en documentación, estadística e informática.
- En los años 1920–1940 se publicaron los primeros estudios sobre productividad de autores y distribución de artículos.
- Lotka (1926) formuló la Ley de Productividad Científica, que describe la frecuencia desigual con que los investigadores publican.
- Bradford (1934) propuso su Ley de Dispersión, que explica cómo los artículos relevantes se concentran en un número reducido de revistas núcleo.
- Zipf (1949) introdujo la Ley de Frecuencia de Palabras, base para el análisis de términos y coocurrencias.
- En la década de 1950–1960, Eugene Garfield fundó el Institute for Scientific Information (ISI) y desarrolló el Science Citation Index (SCI).
- El SCI revolucionó la medición de la ciencia al permitir rastrear redes de citación entre publicaciones.
- En los años 70–80, el campo se institucionalizó con la aparición de la revista Scientometrics y los primeros congresos internacionales.
- En esta etapa, la bibliometría se amplió hacia la cienciometría, enfocada en la dinámica y política de la ciencia.
- A partir de los años 90, el uso de bases de datos electrónicas y software especializado permitió análisis más amplios y precisos.
- El surgimiento de Internet y los motores de búsqueda transformó radicalmente el acceso y la recopilación de información científica.
- Con la llegada de Google Scholar (2004) y Scopus (2004), se diversificaron las fuentes bibliométricas y los indicadores de impacto.
- Desde 2010, emergen la altmetría y la webmetría, orientadas al impacto social y digital de la ciencia.
- En esta etapa moderna, la bibliometría se integra con la inteligencia artificial, el big data y la ciencia abierta.
- Los métodos actuales permiten mapear redes de colaboración, tendencias temáticas y estructuras cognitivas.
- El desarrollo histórico refleja una transición desde un enfoque cuantitativo simple a una visión multidimensional y ética de la evaluación científica.
- La bibliometría contemporánea combina la tradición de Garfield, Lotka y Bradford con herramientas digitales avanzadas.
- En síntesis, su evolución ha convertido a la bibliometría en un pilar esencial de la investigación, la gestión del conocimiento y la política científica.

Preguntas de autoevaluación

1. ¿Quién acuñó el término “bibliometría” y en qué año se consolidó su uso?
2. ¿Qué describe la Ley de Lotka en relación con la productividad científica?
3. ¿Cuál es el principio central de la Ley de Bradford?
4. ¿Qué aportó Zipf al desarrollo de la bibliometría?
5. ¿Qué importancia tuvo Eugene Garfield en la historia de la disciplina?
6. ¿Qué papel desempeñó la creación del Science Citation Index en los años 60?
7. ¿Por qué la década de los 70–80 marcó la institucionalización de la bibliometría?
8. ¿Cómo influyeron las bases de datos electrónicas e Internet en la evolución del campo?
9. ¿Qué aportes introdujo la altmetría a partir de 2010?
10. ¿Qué caracteriza a la bibliometría contemporánea respecto a sus primeras etapas?

BIBLIOGRAFÍA

1. Sugimoto CR, Larivière V. Measuring research: What everyone needs to know. Oxford: Oxford University Press; 2018. https://global.oup.com/academic/product/measuring-research-9780190640125

2. Glänzel W. Bibliometrics as a research field: A course on theory and application of bibliometric indicators. Leuven: KU Leuven; 2003. https://www.kuleuven.be/metaforum/docs/pdf/lecture_glanzel_bibliometrics.pdf

3. Waltman L. A review of the literature on citation impact indicators. J Informetrics. 2016;10(2):365–91. https://doi.org/10.1016/j.joi.2016.02.007

4. Cronin B, Sugimoto CR, editors. Beyond bibliometrics: Harnessing multidimensional indicators of scholarly impact. Cambridge (MA): MIT Press; 2014. https://doi.org/10.7551/mitpress/9780262026792.001.0001

5. Hood WW, Wilson CS. The literature of bibliometrics, scientometrics, and informetrics. Scientometrics. 2001;52(2):291–314. https://doi.org/10.1023/A:1017919924342

REFERENCIAS BIBLIOGRÁFICAS

1. Garfield E. Citation indexes for science. Science. 1955;122(3159):108–11.

2. Garfield E. "Science Citation Index"—A new dimension in indexing. Science. 1964;144(3619):649–54.

3. Martín-Martín A, Thelwall M, Orduna-Malea E, Delgado López-Cózar E. Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: a multidisciplinary comparison of coverage via citations. Scientometrics. 2021;126(1):871–906. https://doi.org/10.1007/s11192-020-03690-4

4. Garfield E. The evolution of the science citation index. International Microbiology. 2007;10(1):65–9.

5. Lotka AJ. The frequency distribution of scientific productivity. Journal of the Washington Academy of Sciences. 1926;16(12):317–23.

6. Bradford SC. Classic paper: Sources of information on specific subjects. Collection Management. 1976;1(3–4):95–103. https://doi.org/10.1300/J105v01n03_06

7. Patsopoulos NA, Analatos AA, Ioannidis JPA. Relative citation impact of various study designs in the health sciences. JAMA. 2005;293(19):2362–6. https://doi.org/10.1001/jama.293.19.2362

8. Agencia Nacional de Investigación y Desarrollo (ANID). Concurso de Proyectos Fondecyt Regular 2024. 2024. https://anid.cl/concursos/concurso-de-proyectos-fondecyt-regular-2024/