

## La búsqueda de información y su asociación con la producción científica: el ejemplo de la sífilis y la gonorrea

Information search and its association to scientific production: the example of syphilis and gonorrhoea

María Sanz-Lorente<sup>1\*</sup>

Javier Sanz-Valero<sup>1</sup>

Carmina Wanden-Berghe<sup>2</sup>

<sup>1</sup>Universidad Miguel Hernández. Campus Sant Joan d'Alacant, España.

<sup>2</sup>Instituto de Investigación Sanitaria y Biomédica de Alicante (ISABIAL-FISABIO). Alicante, España.

\*Autor para la correspondencia: [msanzlor@gmail.com](mailto:msanzlor@gmail.com)

### RESUMEN

El objetivo de este trabajo fue analizar la asociación entre la búsqueda de información sobre sífilis y gonorrea a través de Google, y la producción científica sobre estas dos infecciones de transmisión sexual. Se realizó estudio descriptivo transversal. Los datos se obtuvieron de la consulta directa, *online*, *Google Trends* y MEDLINE (vía PubMed), con el uso de los términos “*Syphilis*” y “*Gonorrhoea*”. Las variables estudiadas fueron: volumen de búsqueda relativo (VBR), VBR mensual medio (VBRm), referencias (REF) y REF mensuales medias (REFm). La VBRm para el término *Syphilis* presentó ligera progresión creciente y ajuste exponencial ( $R^2= 0,05$ ;  $p= 0,42$ ); Para *Gonorrhoea* la evolución VBRm fue creciente con ajuste lineal ( $R^2= 0,67$ ;  $p< 0,01$ ). La REFm para la producción científica sobre la *Syphilis* mostró un ajuste lineal creciente ( $R^2= 0,42$ ;  $p= 0,01$ ) y para la REFm sobre la *Gonorrhoea* fue un ajuste exponencial creciente ( $R^2= 0,47$ ;  $p= 0,01$ ). La relación del VBRm para el término *Syphilis* en relación con su REFm dio una asociación escasa ( $R= 0,11$ ;  $p= 0,69$ ). Esta relación para la *Gonorrhoea* mostró una significativa correlación positiva ( $R= 0,67$ ;  $p= 0,01$ ). Los resultados obtenidos no permitieron obtener una conclusión firme que contestara el objetivo de este trabajo. A partir de ahora, con un mayor seguimiento, se podrán obtener resultados que confirmen, o no, la asociación entre la búsqueda de información y la producción científica sobre los temas relacionados con la salud.

**Palabras clave:** Acceso a la información; publicaciones; gestión de la información; asociación; sífilis; gonorrea.

## ABSTRACT

The purpose of the study was to analyze the association between information search about syphilis and gonorrhea in Google, and the scientific production about these two sexually transmitted infections. A descriptive cross-sectional study was conducted. Data were obtained by direct online consultation of Google Trends and MEDLINE (via PubMed) using the search terms "syphilis" and "gonorrhea". The variables analyzed were relative search volume (RSV), mean monthly RSV (RSVm), references (REF) and mean monthly REF (REFm). The RSVm for the term *syphilis* showed slight increasing progression and exponential adjustment ( $R^2= 0.05$ ;  $p= 0.42$ ); while for *gonorrhea* the RSVm evolution was increasing with linear adjustment ( $R^2= 0.67$ ;  $p< 0.01$ ). The REFm for the scientific production about *syphilis* displayed increasing linear adjustment ( $R^2= 0.42$ ;  $p= 0.01$ ), whereas the REFm for *gonorrhea* showed increasing exponential adjustment ( $R^2= 0.47$ ;  $p= 0.01$ ). The association of the RSVm for the term *syphilis* with its REFm was scant ( $R= 0.11$ ;  $p= 0.69$ ), while the term *gonorrhea* exhibited a significant positive correlation ( $R= 0.67$ ;  $p= 0.01$ ). The results obtained were not sufficient to reach a definitive conclusion in response to the aim of the study. From this moment onwards, with greater follow-up, further results may be obtained which will either confirm or disprove the association between information search and scientific production about health topics.

**Key words:** Access to information; publications; information management; association; syphilis; gonorrhea.

Recibido: 26/04/2019

Aceptado: 25/09/2019

## INTRODUCCIÓN

La misión de todo investigador, además de investigar, es también la de hacer llegar al tejido productivo los resultados de dichos estudios. La investigación contribuye sin duda a la generación del conocimiento y al desarrollo o crecimiento del entorno social, económico y productivo.<sup>(1)</sup> En consecuencia, el análisis y la evaluación de la información y el conocimiento resultante de la actividad científica es un elemento imprescindible para todos los programas de investigación pública, tecnología y desarrollo que se implementan en una sociedad, y es allí donde la Ciencia de la Información brinda una ayuda inestimable, al desarrollar técnicas e instrumentos para medir la producción de conocimiento y su transformación en bienes.<sup>(2)</sup>

Por otro lado, la Web 2.0 ofrece oportunidades sin precedentes para los pacientes y el público en general a la hora de buscar información sobre salud. De hecho, hace ya tiempo que buscan allí esta información, incluso antes que consultar con los profesionales,<sup>(3)</sup> Por ejemplo, Wikipedia es actualmente el quinto sitio más visitado en Internet<sup>(4)</sup> y una de las aplicaciones más globalizadas.<sup>(5)</sup>

*Eysenbach*,<sup>(6)</sup> en el año 2009, acuñó el término de «infodemiología» o «infoepidemiología» como un conjunto emergente de métodos de información de salud pública para analizar el comportamiento de la búsqueda, la comunicación y la publicación en Internet. Es decir, la «infodemiología» observa y analiza el comportamiento basado en la Web para conocer la conducta humana real a fin de predecir, evaluar, e incluso prevenir problemas relacionados con la salud que surgen constantemente en la vida cotidiana.<sup>(7)</sup>

Al inicio del siglo XXI, las enfermedades de transmisión sexual estaban entre las causas más comunes de enfermedad en el mundo. Diariamente, casi un millón de personas contraían una enfermedad de transmisión sexual (ETS), y aun excluyendo la infección por el virus de inmunodeficiencia humana (VIH), las ETS eran, y son, un problema de salud pública importante, de alta prevalencia, tanto por la carga de la enfermedad que generan, como por las complicaciones y secuelas que producen si no se diagnostican y tratan precozmente.<sup>(8)</sup>

Por tanto, las infecciones de transmisión sexual (ITS) figuran entre las cinco categorías principales de enfermedad por las que se busca información y ayuda. La hipótesis de que las poblaciones proporcionen datos sobre sus gustos e incluso sobre su enfermedad, a través del comportamiento de búsqueda de información en la Web, ya ha sido demostrada y existen estudios que relacionaron los datos obtenidos de las búsquedas de información con los casos de enfermedad; por ejemplo en la gripe, la hepatitis o el VIH/SIDA.<sup>(9)</sup> *Johnson y Mehta*,<sup>(10)</sup> en el año 2014, ya demostraron, estudiando las tendencias de los motores de búsqueda, que eran herramientas válidas para integrarlas en la vigilancia en tiempo real de las ITS. Además de este estudio sobre ITS, otros autores utilizaron las tendencias de búsqueda en Google para comprobar la relación con los datos de enfermedad, por ejemplo en el campo de la influenza,<sup>(11)</sup> el cuidado domiciliario,<sup>(9)</sup> o el virus de la inmunodeficiencia humana.<sup>(12)</sup>

Así pues, está más que demostrado que Google es un motor de búsqueda que facilita información -incluso sobre salud- a cualquier persona, que a través de los resultados obtenidos puede acceder fácilmente a los documentos científicos existentes en la Red. Asimismo, dado el repunte que se viene observando, en el siglo XXI, de los datos de incidencia y prevalencia de las ITS curables, y sobre todo de la sífilis y la gonorrea, se decidió el estudio de estas enfermedades.<sup>(8)</sup>

En consecuencia, el objetivo de este trabajo fue analizar la asociación entre la búsqueda de información sobre sífilis y gonorrea a través de Google, y la producción científica sobre estas dos infecciones de transmisión sexual.

## MÉTODOS

Se realizó un estudio descriptivo transversal. Los datos sobre búsqueda de información se obtuvieron de la consulta directa, mediante acceso *online*, a Google Trends (<https://trends.google.es/>), y los de la producción científica a partir de MEDLINE, vía PubMed (<https://www.ncbi.nlm.nih.gov/pubmed/>).

## Términos y búsqueda de los datos

Se utilizó las palabras “*Syphilis*” y “*Gonorrhea*” en Google Trends como «término de búsqueda» y para intentar asemejarse a las búsquedas realizadas por personas no expertas no se utilizó ningún tipo de filtro (búsqueda libre). En MEDLINE se realizó la búsqueda utilizando los *Medical Subject Headings* (MeSH): “*Syphilis*”[MeSH] y “*Gonorrhea*”[MeSH]. El periodo de estudio fue desde el año 2004 (primer año en el que Google Trends ofrece los datos) hasta el año 2018. La fecha de consulta fue el 4 de abril de 2019.

## Obtención y almacenamiento de los datos

Los resultados obtenidos fueron descargados, desde ambas plataformas, en formato normalizado *comma-separated values* (CSV) que permitió su almacenamiento posterior en un archivo Excel. El control de la calidad de esta información se efectuó mediante dobles tablas, con la corrección de las posibles inconsistencias mediante la consulta con la tabla original descargada. Para el análisis estadístico se utilizó el programa *Statistical Package for the Social Sciences* (SPSS) para Windows versión 22.0.

## Variables del estudio

1. *Volumen de búsqueda relativo (VBR)*: Resultado ofrecido por Google Trends cuyos valores se normalizan en una escala de 0 (un volumen de búsqueda relativo inferior al 1 % del volumen máximo) a 100 (el volumen de búsqueda relativo alcanza su máximo). Por ejemplo, un VBR = 25, representa el 25 % de la proporción de búsqueda observada más alta durante el período de estudio.
2. *VBR mensual medio (VBRm)*: Sumatorio de los VBR mensuales de un año determinado dividido por 12.
3. *Referencias (REF)*: Número de referencias bibliográficas anuales obtenidas de la búsqueda realizada en MEDLINE mediante el correspondiente MeSH:
  - Referencias mensuales medias (REFm): Número de referencias (REF) para un año determinado dividido por 12.

## Análisis de los datos

Para las variables cuantitativas (VBR, VBRm, REF y REFm) se calculó la media y su desviación estándar, la mediana, la amplitud intercuartílica (AIQ), el máximo y el mínimo. La evolución temporal de las tendencias de búsqueda se examinó mediante el análisis de regresión, donde se calculó el coeficiente de determinación. Para obtener la relación entre variables cuantitativas se usó el coeficiente de correlación de Pearson. El nivel de significación utilizado en todos los contrastes de hipótesis fue  $\alpha \leq 0,05$ .

## RESULTADOS

Al efectuar las búsquedas en Google Trends (Fig. 1) y en MEDLINE (Fig. 2) se pudieron obtener los resultados tanto para el VBRm desde Google Trends, como el REFm desde MEDLINE (vía PubMed), como muestra la tabla 1.

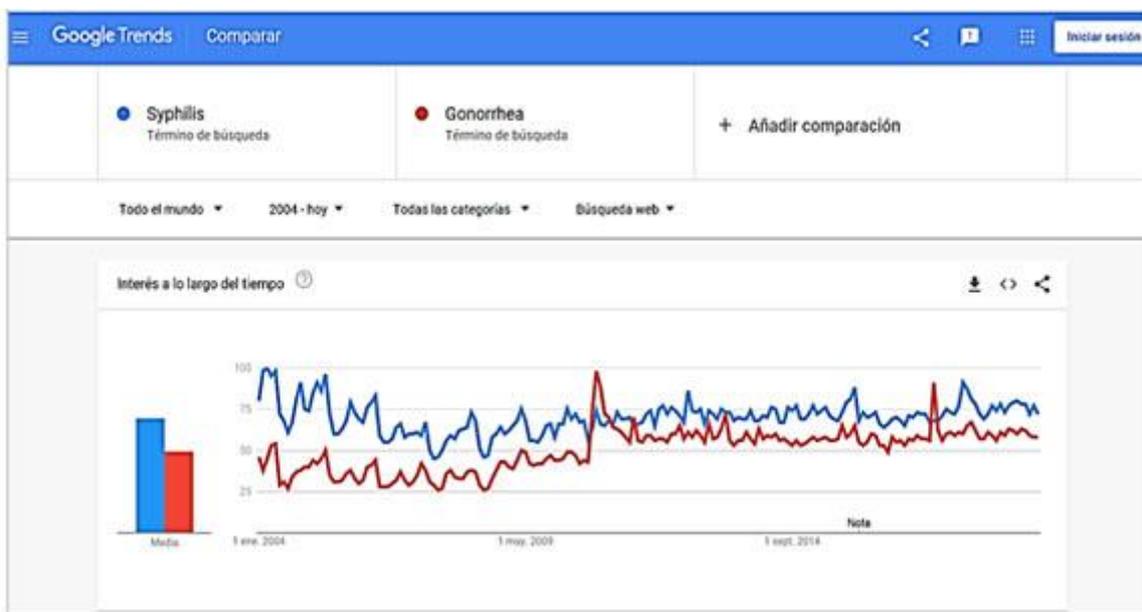


Fig. 1 – Búsqueda efectuada en Google Trends.

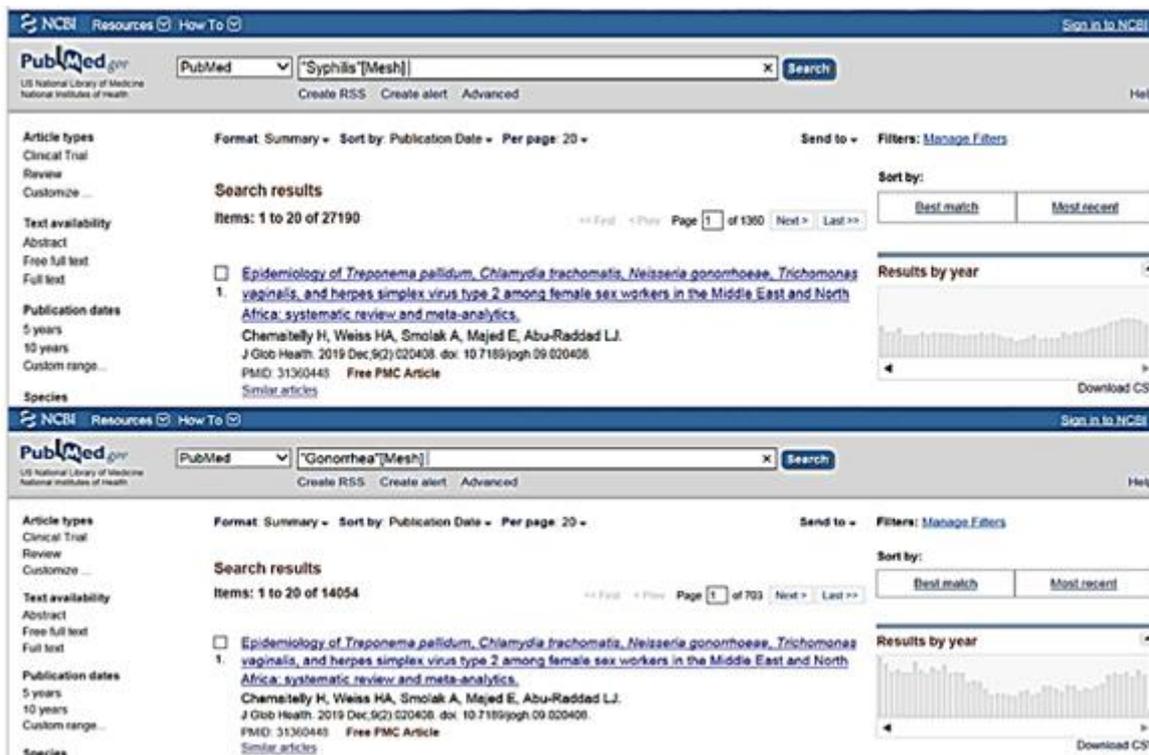


Fig. 2 – Búsquedas efectuadas en MEDLINE (vía PubMed).

**Tabla 1** - Volúmenes de búsqueda relativos mensuales (VBRm), obtenidos de Google Trends, y referencias mensuales medias (REFm), observadas en MEDLINE, para los términos *Syphilis* y *Gonorrhea*

Año	VBRm <sup>1</sup> Syphilis	VBRm <sup>1</sup> Gonorrhea	REFm <sup>2</sup> Syphilis	REFm <sup>2</sup> Gonorrhea
2004	76,92	37,33	24,83	13,08
2005	66,33	34,50	23,25	16,75
2006	58,67	29,58	26,67	16,08
2007	50,67	30,50	24,83	13,92
2008	53,92	31,08	27,33	15,08
2009	57,42	39,25	29,50	14,58
2010	60,83	52,83	30,50	16,08
2011	62,42	55,00	32,67	17,67
2012	66,67	53,33	35,67	21,92
2013	64,67	53,17	37,33	23,00
2014	64,75	51,00	39,08	22,25
2015	67,83	52,75	39,42	23,42
2016	62,08	50,08	38,92	21,42
2017	64,33	55,08	34,58	23,50
2018	69,67	54,42	24,25	15,25

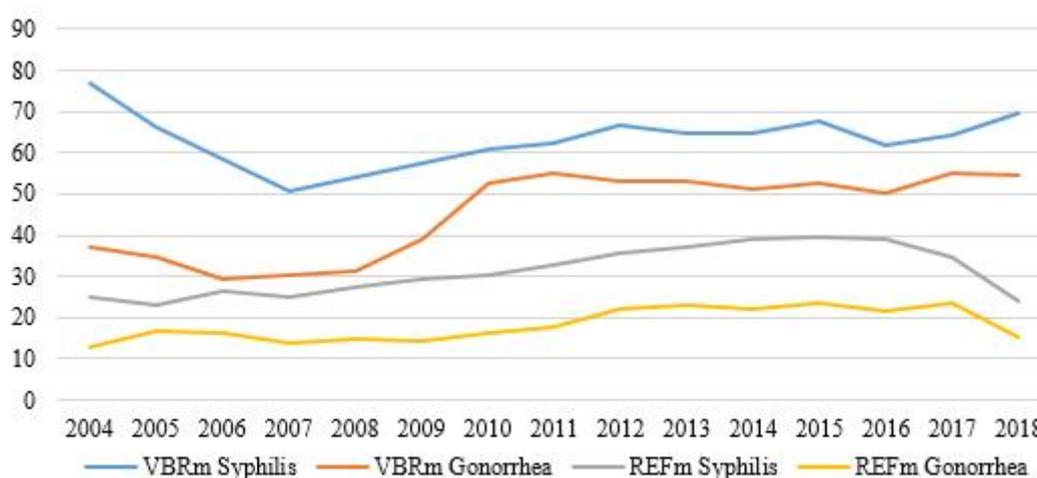
Los estadísticos, para todo el periodo analizado, tanto del VBR para *Syphilis* y *Gonorrhea*, como para las REF, obtenidas de las búsquedas realizadas pueden consultarse en la tabla 2.

**Tabla 2** - Estadísticos, para todo el periodo analizado, del volumen de búsqueda relativo mensual medio (VBRm) y del número de referencias mensuales medias (REFm), obtenidas de las búsquedas realizadas en Google Trends y MEDLINE para *Syphilis* y *Gonorrhea*

Datos estadísticos	VBRm <i>Syphilis</i>	VBRm <i>Gonorrhea</i>	REFm <i>Syphilis</i>	REFm <i>Gonorrhea</i>
Media	63,15 ± 1,67	45,33 ± 2,63	31,26 ± 1,53	18,27 ± 0,99
Mediana	64,33	51,00	30,50	16,75
AIQ <sup>1</sup>	8,00	18,83	12,50	7,17
Máximo	76,92	55,08	39,42	23,50
Mínimo	50,67	29,58	23,25	13,08

### Evolución temporal

De los resultados anuales, tanto del VBRm como las REFm, obtenidos para los dos términos en estudio, se pudo conocer su progreso a lo largo del periodo estudiado (Fig. 3).



Fuente: Datos obtenidos de Google Trends y MEDLINE.

**Fig. 3** - Evolución temporal del volumen de búsqueda relativo mensual medio (VBRm) y del número de referencias mensuales medias (REFm) para los dos términos del estudio: *Syphilis* y *Gonorrhea*.

Las tendencias de búsqueda obtenidas en Google Trends para el término *Syphilis* presentó una ligera progresión creciente y ajuste a un modelo exponencial ( $R^2= 0,05$ ;  $p= 0,42$ ). Para *Gonorrhea* la evolución VBRm fue claramente creciente con ajuste a un modelo lineal ( $R^2= 0,67$ ;  $p< 0,01$ ). El análisis de regresión de la variable REFm para la producción científica sobre *Syphilis* mostró un ajuste lineal creciente ( $R^2= 0,42$ ;  $p= 0,01$ ) y para la variable REFm de los documentos sobre *Gonorrhea* fue un ajuste exponencial creciente ( $R^2= 0,47$ ;  $p= 0,01$ ).

## Grado de relación

La relación del VBRm obtenido para el término de búsqueda *Syphilis* en relación con su producción científica REFm dio en el conjunto del periodo analizado una asociación escasa ( $R = 0,11$ ;  $p = 0,69$ ), mientras para la asociación entre VBRm y REFm para la *Gonorrhea* se observó una significativa correlación positiva ( $R = 0,67$ ;  $p = 0,01$ ).

## DISCUSIÓN

En este estudio se quiso comprobar si las necesidades de información, sobre las dos ITS curables (Sífilis y Gonorrea), se asociaban con la producción científica sobre estas. La primera observación que se pudo apreciar es que tanto la producción científica como la necesidad de información de estas dos enfermedades presentaban, desde el año 2004, un crecimiento constante. Este resultado era previsible, ya que la producción científica presenta *per se* crecimientos en torno al 8 % anual<sup>(13)</sup> y las consultadas impulsadas por la búsqueda clásica en Internet continuaron subiendo, según datos del cuarto trimestre de 2018.<sup>(14)</sup> También hay que tener en cuenta que según las teorías cuantitativas anunciadas por Price, el crecimiento de la producción científica se observaría en periodos de entre 15 a 30 años de evolución, situación que apenas se alcanza en el momento de este trabajo.<sup>(15)</sup>

Los estadísticos obtenidos indicaron que tanto VBRm como el REFm presentaron valores medios. No se destacó ningún hito (acontecimiento puntual y notorio en el VBR) que mostrara un especial interés en las búsquedas de información o un claro repunte en la producción científica. La localización de hitos (picos) es un hecho importante para la vigilancia epidemiológica, ya que se ha demostrado relación con repuntes de enfermedad<sup>(16,17)</sup> y también como réplica a campañas de información que provocaban un mayor interés en esta información.<sup>(18)</sup>

El grado de asociación observado presentó datos antagónicos. Por un lado, existió una adecuada relación entre la producción científica y las búsquedas realizadas en relación con la *Gonorrhea*. Pero, por el otro, la asociación obtenida al relacionar el número de artículos científicos con las búsquedas sobre *Syphilis* apenas dio una pobre correlación. Estos resultados no permiten obtener una conclusión firme que conteste el objetivo de este trabajo.

Si bien es cierto que anteriores estudios obtuvieron una adecuada correlación entre los resultados de la búsqueda de información y los de enfermedad (por ejemplo la influenza,<sup>(11)</sup> el cuidado domiciliario<sup>(9)</sup> o el virus de la inmunodeficiencia humana),<sup>(12)</sup> no se encontró, en la literatura científica, ningún documento que estudiara si se había investigado el posible crecimiento de las publicaciones científicas como consecuencia del interés de la población por esta información o como secuela del repunte de la enfermedad.

En cuanto a las posibles limitaciones del estudio, de acuerdo con Johnson & Mehta,<sup>(10)</sup> hay que tener presente que se trataba de un análisis de datos ecológicos y los hallazgos pudieron no ser representativos

a nivel individual; por ejemplo, las tendencias son poblacionales y no pueden concluir que solo los individuos infectados con ITS son, de hecho, aquellos que generan todo el volumen de búsqueda relacionado con estas enfermedades. Además, existe incertidumbre acerca de la causa de las tendencias en las búsquedas (aumento real en la infección, noticias, curiosidad, etc.) y cuándo ocurren (antes del diagnóstico de la ITS o después).<sup>(19)</sup> Por otro lado, los motivos que llevan a la publicación científica no siempre están motivados por la existencia de una enfermedad (por ejemplo: necesidades curriculares, satisfacción personal, etcétera).<sup>(20)</sup>

## Conclusiones

El seguimiento de las consultas *online*, a través de Google, puede revelar la preocupación de las personas y evaluar los cambios de comportamiento en relación con la información sobre salud, e incluso la necesidad de la generación de conocimiento en línea con la necesidad social de este.

Los resultados obtenidos no permitieron obtener una conclusión firme que contestara el objetivo de este trabajo, la asociación entre la búsqueda de información y la producción científica sobre las ciencias de la salud. A partir de ahora, con un mayor seguimiento, se podrán obtener resultados que confirmen, o no, la asociación entre la búsqueda de información y la producción científica sobre los temas relacionados con la salud.

## REFERENCIAS BIBLIOGRÁFICAS

1. Fundación Conocimiento y Desarrollo (Fundación CYD). Informe CYD; 2017 [monografía en Internet]. Barcelona, España: Fundación CYD; 2018 [acceso: 07/04/2019]. Disponible en: <https://bit.ly/2CCZQI8>
2. Arencibia R, de Moya Anegón F. La evaluación de la investigación científica: una aproximación teórica desde la cienciometría. ACIMED. 2008 [acceso: 07/04/2019];17(4):[aprox 79 pantallas]. Disponible en: <http://bit.ly/2XA9AvU>
3. Sanz-Valero J, López-Marcos P, Perdiguero-Gil E. Adequacy of the information on active pharmaceutical ingredients obtained in a Google search. Eur J Clin Pharm. 2015;17(1):8-14.
4. Amazon Company. Alexa, Top Sites (The top 500 sites on the Web); 2019 [acceso: 07/04/2019]. Disponible en: <http://www.alexa.com/topsites>
5. Pimienta D, Prado D. Medición de la presencia de la lengua española en la Internet: métodos y resultados. Rev Esp Doc Cient. 2016;39(3):e141.
6. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res. 2009;11(1):e11.

7. Mavragani A, Ochoa G, Tsagarakis KP. Assessing the methods, tools, and statistical approaches in Google Trends research: systematic review. *J Med Internet Res.* 2018;20(11):e270.
8. World Health Organization. Global incidence and prevalence of selected curable sexually transmitted infections - 2008 [monografía en Internet]. Geneva, Switzerland: WHO; 2012 [acceso: 12/05/2019]. Disponible en: <http://bit.ly/2W5V1mJ>
9. Sanz-Lorente M, Wanden-Berghe C. Tendencias temporales de los patrones de búsqueda de información sobre cuidado domiciliario “Home Care” u hospitalario “Hospital Care” a través de Google. *Hosp Domic.* 2018;2(3):93-9.
10. Johnson AK, Mehta SD. A comparison of Internet search trends and sexually transmitted infection rates using Google trends. *Sex Transm Dis.* 2014;41(1):61-3.
11. Orellano PW, Reynoso JI, Antman J, Argibay O. Using Google Trends to estimate the incidence of influenza-like illness in Argentina. *Cad Saude Publ.* 2015;31(4):691-700.
12. Chiu APY, Lin Q, He D. News trends and web search query of HIV/AIDS in Hong Kong. *PLoS One.* 2017;12(9):e0185004.
13. Bornmann L, Mutz R. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *J Assoc Inf Sci Technol.* 2015;66(11):2215-22.
14. Merkle Inc. Digital Marketing Report Q4 2018 [monografía en Internet]. Columbia, EE.UU.: Merkel; 2019 [acceso: 23/04/2019]. Disponible en: <http://bit.ly/2Vn2Io6>
15. Price DJS. Little science, big science. New York, EE.UU.: Columbia University Press; 1963.
16. Chan EH, Sahai V, Conrad C, Brownstein JS. Using web search query data to monitor dengue epidemics: a new model for neglected tropical disease surveillance. *PLoS Negl Trop Dis.* 2011;5(5):e1206.
17. Culquichicón-Sánchez C, Ramos-Cedano E, Chumbes-Aguirre D, Araujo-Chumacero M, Díaz Vélez C, Rodríguez-Morales AJ. Tecnologías de la Información y la Comunicación (TICs) en la vigilancia, prevención y control del dengue. *Rev Chilena Infectol.* 2015;32(3):363-4.
18. Ortiz-Martínez Y, Rios-González CM. Global impact of the World Hepatitis Day 2016: An evaluation using Google Trends. *J Infect Public Health.* 2017;10(5):690-1.
19. Sanz-Lorente M, Sanz-Valero J, Wanden-Berghe C. Tendencias temporales de los patrones de búsqueda de información sobre VIH/SIDA en España. *Rev Esp Comun Salud.* 2019. En prensa.
20. Castiel LD, Sanz-Valero J. Entre fetichismo e sobrevivência: o artigo científico é uma mercadoria acadêmica? *Cad Saude Publica.* 2007;23(12):3041-50.

### Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

### **Declaración de la contribución de los autores**

*Sanz-Valero* diseñó el estudio y preparó la propuesta metodológica; *Sanz-Lorente* y *Wanden-Berghe* analizaron los datos y redactaron la primera versión del manuscrito. *Sanz-Lorente* se encargó de la recogida y proceso de los datos. *Sanz-Valero* y *Wanden-Berghe* realizaron el análisis general y el análisis estadístico. Los tres autores revisaron la redacción del manuscrito y aprobaron la versión finalmente remitida.

### **Financiación**

La presente investigación no ha recibido ayudas específicas provenientes de agencias del sector público, sector comercial o entidades sin ánimo de lucro.