

Uso de la herramienta Google Trends para estimar la incidencia de enfermedades tipo influenza en Argentina

Using Google Trends to estimate the incidence of influenza-like illness in Argentina

Uso da ferramenta Google Trends para estimar a incidência de síndrome gripal na Argentina

Pablo Wenceslao Orellano ¹

Julieta Itatí Reynoso ²

Julián Antman ³

Oswaldo Argibay ³

Abstract

The aim of this study was to find a model to estimate the incidence of influenza-like illness (ILI) from the Google Trends (GT) related to influenza. ILI surveillance data from 2012 through 2013 were obtained from the National Health Surveillance System, Argentina. Internet search data were downloaded from the GT search engine database using 6 influenza-related queries: flu, fever, cough, sore throat, paracetamol, and ibuprofen. A Poisson regression model was developed to compare surveillance data and internet search trends for the year 2012. The model's results were validated using surveillance data for the year 2013 and results of the Google Flu Trends (GFT) tool. ILI incidence from the surveillance system showed strong correlations with ILI estimates from the GT model ($r = 0.927$) and from the GFT tool ($r = 0.943$). However, the GFT tool overestimates (by nearly twofold) the highest ILI incidence, while the GT model underestimates the highest incidence by a factor of 0.7. These results demonstrate the utility of GT to complement influenza surveillance.

Human Influenza; Epidemiologic Models; Incidence

Resumen

El objetivo del presente estudio fue hallar un modelo para estimar la incidencia de enfermedades tipo influenza (ETI), a partir de los términos de búsqueda relacionados recolectados por el Google Trends (GT). Los datos de vigilancia de ETI para los años 2012 y 2013 se obtuvieron del Sistema Nacional de Vigilancia de la Salud de Argentina. Las búsquedas de Internet se obtuvieron de la base de datos del GT, usando 6 términos: gripe, fiebre, tos, dolor de garganta, paracetamol e ibuprofeno. Se desarrolló un modelo de regresión de Poisson a partir de datos del año 2012, y se validó con datos del 2013 y resultados de la herramienta Google Flu Trends (GFT). La incidencia de ETI del sistema de vigilancia presentó fuertes correlaciones con las estimaciones de ETI del GT ($r = 0,927$) y del GFT ($r = 0,943$). Sin embargo, el GFT sobreestimó el pico de incidencia por casi el doble, mientras que el modelo basado en el GT subestimó el pico de incidencia por un factor de 0,7. Estos resultados demuestran la utilidad del GT como un complemento para la vigilancia de la influenza.

Influenza Humana; Modelos Epidemiológicos; Incidencia

¹ Instituto Nacional de Tecnología Industrial, Rosario, Argentina.

² Hospital Interzonal General de Agudos "San Felipe", Buenos Aires, Argentina.

³ Dirección de Epidemiología, Ministerio de Salud de la Nación, Buenos Aires, Argentina.

Correspondencia

P. W. Orellano
Instituto Nacional de Tecnología Industrial,
Esmeralda y Ocampo,
Rosario, Santa Fé 2000,
Argentina.
porellano@gmail.com

Introducción

De acuerdo a la Organización Mundial de la Salud (OMS), las enfermedades respiratorias constituyen importantes causas de morbilidad y mortalidad a nivel mundial ¹, con 300 a 500 mil muertes anuales asociadas a la influenza ². En Argentina, la pandemia de influenza H1N1 2009 provocó una carga importante de enfermedad a lo largo del país, con 1,3 millones de casos de enfermedades tipo influenza, 10 mil hospitalizaciones y 990 fallecimientos ³, sobre un total de 40 millones de habitantes ⁴. En general, los patrones globales de influenza están caracterizados por picos de actividad focalizados en la estación invernal, con mayor duración de epidemias en los trópicos ⁵. La vigilancia epidemiológica de la dinámica temporal de la influenza juega un rol de importancia en el monitoreo y control de la enfermedad. Esta puede estar dirigida a la presentación clínica, a través de las enfermedades tipo influenza (ETI), o a la vigilancia de casos confirmados por laboratorio virológico. Debido a que la vigilancia epidemiológica se lleva adelante en forma pasiva, y depende de los registros médicos de todos los centros de salud, es posible de demoras en la notificación, así como de posibles subestimaciones ⁶. Por esta causa, el sistema de vigilancia debe ser continuamente evaluado, mientras que paralelamente se ensayan nuevas herramientas complementarias para estimar la incidencia de la influenza de forma indirecta, y mejorar la sensibilidad del registro. Un ejemplo es la utilización de sistemas de monitoreo basados en Internet, con diseños generados para participación voluntaria de la población ⁷. Una forma más general para el monitoreo indirecto de la incidencia de ETI es a través del monitoreo de las búsquedas de términos relacionados a la enfermedad a través de Internet ⁸. Esta experiencia se ha realizado con buscadores como Yahoo ⁹. Una herramienta genérica de amplio uso para el monitoreo de términos más buscados en Internet es Google Trends (GT), disponible en el sitio <http://www.google.com.ar/trends/>. En este sitio, los usuarios pueden consultar de forma gratuita el volumen de búsqueda de cualquier término (*gripe*, por ejemplo), obteniendo un resultado normalizado de la frecuencia relativa de búsqueda de ese término en el período considerado ¹⁰. Esta herramienta está disponible para búsquedas desde el año 2004, a nivel de país o región, y con una base semanal o mensual. La forma de utilizar estos datos es realizar regresiones entre los datos de vigilancia y los términos buscados, y utilizar los coeficientes calculados para estimar los casos de ETI, a partir de los términos de búsqueda. A partir de la herramienta GT, la compañía Google desarrolló en

el año 2008 un nuevo instrumento para estimar los casos de ETI a nivel de país y región ¹¹. Esta herramienta, denominada Google Flu Trends (GFT) (<http://www.google.org/flutrends/>) está basada en algoritmos propios que utilizan 45 términos de búsqueda, y una correlación con el porcentaje de consultas médicas por influenza, basada en años previos ¹². En años recientes se han realizado diversos estudios que validan el uso del GFT como complemento de la vigilancia epidemiológica, con resultados que demuestran su utilidad ^{13,14,15}. Sin embargo, también se han detectado errores substanciales en las estimaciones, sobre todo durante la pandemia de influenza H1N1 2009 y en el brote 2012/2013 de Estados Unidos, con subestimación y sobreestimación de casos respectivamente ^{11,16}. Estos errores dieron lugar a sucesivos ajustes en la parametrización del modelo ¹⁷, incluyendo una revisión y ajuste posterior a la pandemia del año 2009 ¹¹. Debido a estos problemas, y tomando en consideración que no es posible acceder a los términos de búsqueda y algoritmos utilizados en el GFT, los cuales no han sido divulgados por la compañía ^{10,11,18}, en el presente trabajo se prefirió utilizar los resultados de búsqueda de términos del GT para desarrollar un modelo local, sin recurrir a los resultados del GFT. Otras dos razones motivaron la elección del uso del GT para desarrollar un modelo propio: (i) la necesidad de adaptar de forma constante los modelos ¹⁶ y (ii) la conveniencia de obtener parámetros que relacionen las búsquedas de términos a nivel de país con los casos de ETI ¹⁹. Siguiendo esta línea de acción, el objetivo del presente trabajo fue obtener un modelo para estimar los casos de ETI, a partir de los términos de búsquedas en Internet provistos por el GT, y validar los resultados comparando el presente modelo y los resultados del GFT con los casos de ETI informados por el Sistema Nacional de Vigilancia de la Salud de Argentina (SNVS).

Métodos

Datos de vigilancia de la salud

Los datos para el cálculo de parámetros del modelo y para su validación se obtuvieron de registros de casos de ETI notificados al SNVS. Este sistema releva de forma semanal los casos de ETI que son registrados por centros de salud públicos y privados de toda Argentina, mediante un sistema de carga *on-line* de datos. Las ETI son entidades clínicas, sin confirmación de laboratorio, definidas por la presencia de fiebre (temperatura > 38°C) acompañada de tos u odinofagia, en ausencia de otros diagnósticos. Se

consideró la incidencia de ETI por cada 100 mil habitantes.

Datos de búsqueda de términos en Internet

La frecuencia relativa de búsquedas de términos relacionados a las ETI en Internet se obtuvo a través de la herramienta GT. Por medio de la misma puede consultarse cuántas búsquedas se han realizado en la página de Google (<https://www.google.com.ar/>) sobre un término específico, según el país o la región. En el presente trabajo se utilizó el total de búsquedas de la Argentina. La escala temporal de las búsquedas fue semanal, y coincidente con el comienzo y finalización de las semanas epidemiológicas consideradas por el sistema de vigilancia de la salud de Argentina. El resultado de las búsquedas para cada término es sometido a un ajuste por parte del grupo desarrollador del GT¹⁰. Los datos son escalados utilizando como denominador el total de búsquedas durante el período considerado para la misma región geográfica. En este trabajo, la búsqueda fue restringida a un año, para cada período (2012 y 2013). Por lo tanto, para cada año el valor semanal de búsquedas por término estuvo comprendido en un rango de 0-100. La elección de los términos se basó en las búsquedas más comunes realizadas vinculadas a la enfermedad, los síntomas y el tratamiento sintomático de las ETI, tomando en cuenta la relación observada entre las ventas de medicamentos y la incidencia de ETI²⁰. Los términos utilizados fueron *gripe*, *fiebre*, *tos*, *dolor de garganta*, *paracetamol* e *ibuprofeno*.

Datos del GFT

Las estimaciones de incidencia de ETI cada 100 mil habitantes, provenientes del GFT, se obtuvieron desde la página del buscador (<http://www.google.org/flutrends/>), seleccionando el total de Argentina para el año 2013.

Análisis estadístico

Se realizaron correlaciones simples de Pearson entre los datos de incidencia de ETI del SNVS y las búsquedas de cada término relacionado con las ETI para el año 2012, considerando tres posibilidades: (i) sin considerar retraso en la notificación; (ii) considerando una semana de retraso y (iii) considerando dos semanas de retraso. Luego, estas variables de predicción se utilizaron para estimar los parámetros según un modelo de regresión de Poisson. Este modelo es útil cuando la variable de resultado es numérica y discreta²¹. Debido a que es probable que el número de casos semanales de ETI se presente como una varia-

ble autocorrelacionada¹⁷, es decir, la incidencia de casos de una determinada semana puede ser dependiente de la incidencia de la semana previa, se utilizó un modelo de ecuaciones de estimación generalizadas (EEG)²², considerando el agrupamiento de semanas en cada mes. De esta manera se obtuvieron 12 grupos (uno por cada mes del año 2012), con 4 a 5 observaciones (semanas) cada uno. La estructura de la matriz de correlación del modelo de EEG se asumió como autoregresiva, por tratarse de medidas repetidas en el tiempo. Se analizó un modelo de regresión completo y se comparó con modelos reducidos anidados. En el modelo completo se intentó manejar la mayor cantidad de información posible para controlar por potenciales sesgos, y por lo tanto los 6 términos de búsqueda fueron mantenidos como variables independientes. Para hallar los modelos reducidos, se utilizó un procedimiento paso a paso hacia atrás, eliminando una a la vez las variables de predicción. El mejor conjunto de variables de predicción se determinó mediante el criterio de cuasi-verosimilitud, bajo el modelo de independencia (QIC, por sus siglas en inglés).

Los coeficientes de regresión, calculados para el año 2012, se utilizaron para estimar la incidencia de ETI de los años 2012 y 2013, a partir de los resultados de los términos de búsqueda del mismo año. La incidencia de casos de ETI, estimada por los modelos basados en el GT y por el GFT, se comparó por medio de una correlación de Pearson con la incidencia notificada al SNVS del año 2013. Se calcularon los coeficientes de correlación para las tres comparaciones, con el objetivo de realizar la validación del modelo.

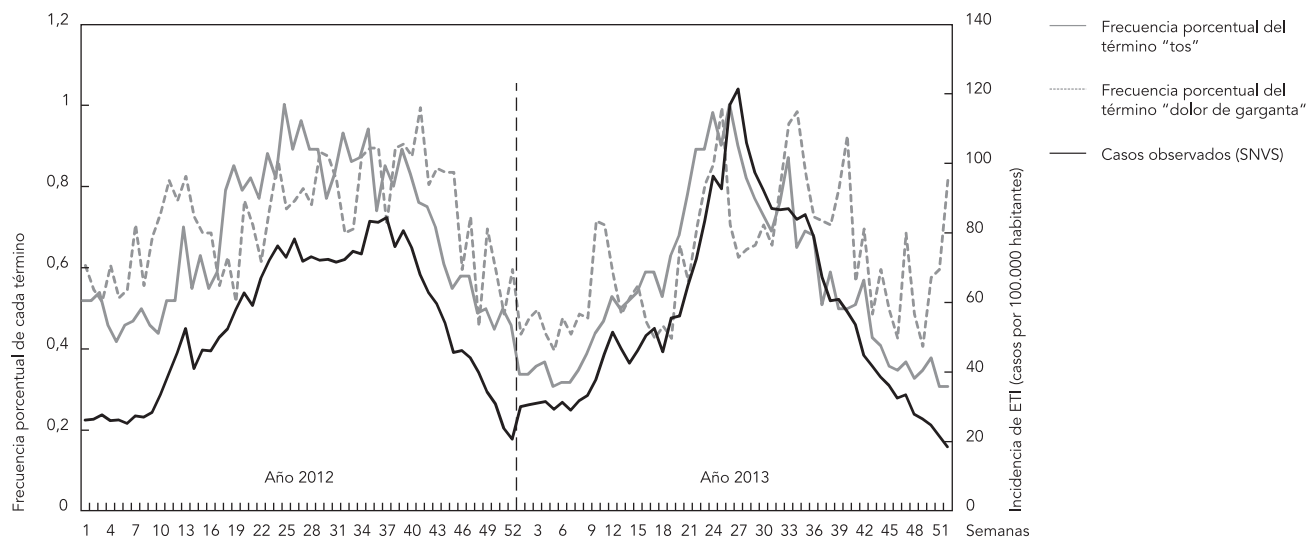
Los cálculos se realizaron mediante el programa R (The R Foundation for Statistical Computing, Viena, Austria; <http://www.r-project.org>), una plataforma de código abierto para análisis estadístico. Las ecuaciones de estimación generalizada se confeccionaron con el paquete "geepack"²³, y el QIC se estimó con el paquete "MuMIn", de la misma plataforma.

Resultados

La incidencia de casos de ETI registrados por el SNVS fue 2.795 y 2.913 por cada 100 mil habitantes para los años 2012 y 2013, respectivamente. En la Figura 1 se presenta la incidencia observada de ETI por semana, en conjunto con la frecuencia porcentual de búsquedas de los dos términos más relevantes para el modelo estimado: *tos* y *dolor de garganta*. Durante el año 2012 estos términos de búsqueda presentaron un máximo en las semanas 25 y 41 respectivamen-

Figura 1

Incidencia de enfermedades tipo influenza (ETI) y frecuencia porcentual de términos "tos" y "dolor de garganta".



SNVS: Sistema Nacional de Vigilancia de la Salud.

Fuentes: Dirección de Epidemiología, Ministerio de Salud de la Nación (Argentina), y Google Trends (<http://www.google.com.ar/trends/>).

te, mientras que la incidencia de ETI observada tiene su máximo en la semana 37. Este desplazamiento ocurre en un año en que no se observa un pico estacional marcado en la incidencia de las ETI. Sin embargo, para el año 2013 los máximos de búsqueda de estos términos coinciden temporalmente con la máxima incidencia de ETI observada, con una diferencia de sólo una semana. Debe tomarse en cuenta que el año 2013 presenta una estacionalidad más marcada. En la Tabla 1 se pueden observar los coeficientes de correlación de Pearson entre cada variable de predicción y los casos de ETI observados, para los diferentes retrasos temporales. Las mejores correlaciones se observan cuando no se considera un retraso entre la semana de búsqueda del término y la semana de notificación de los casos, y las peores ocurren cuando se considera un retraso de dos semanas.

Los coeficientes, obtenidos por el modelo completo de regresión de Poisson, se presentan en la Tabla 2 con errores estándar y valores de p . Los coeficientes correspondientes a las búsquedas de *tos* y *dolor de garganta* fueron significativos, mientras que los otros términos no presentaron significación estadística considerando un $\alpha = 0,05$. Ninguno de los modelos reducidos

fue superior al modelo completo de acuerdo a la comparación de valores del QIC. Sin embargo, se utilizó un modelo reducido manteniendo las variables de predicción más significativas, *tos* y *dolor de garganta*, para realizar la comparación de resultados entre modelos. En el modelo reducido, la variable *tos* presentó un coeficiente de regresión de 100,9 ($p < 0,01$) y la variable *dolor de garganta* presentó un coeficiente de 23,4 ($p = 0,17$).

La Figura 2 presenta los resultados de la correlación para la validación de los modelos completo y reducido basados en el GT y del modelo GFT. En el primero, se observa un valor de $r = 0,927$, en el segundo un valor de $r = 0,86$, mientras que para la validación del GFT se observa un valor de $r = 0,943$. La Figura 3 representa la evolución temporal de los valores observados y estimados de incidencia de ETI, para el año 2012 (etapa de ajuste) y el año 2013 (etapa de validación). En el año 2013, la incidencia máxima estimada por el modelo basado en el GT subestima a la incidencia máxima real por un factor de 0,7, aunque acompaña a la dinámica temporal. Los resultados del GFT presentan una marcada sobreestimación a lo largo de todo el año, con casi el doble de incidencia máxima estimada respecto

Tabla 1

Coefficientes de correlación de Pearson entre los casos de enfermedades tipo influenza (ETI) notificados al Sistema Nacional de Vigilancia de la Salud de Argentina (SNVS) y la frecuencia de búsqueda de cada término, considerando diferentes retrasos.

Término	Sin retraso	Valor de la correlación (IC95%)	
		1 semana de retraso	2 semanas de retraso
Gripe	0,50 (0,26; 0,68)	0,49 (0,25; 0,67)	0,51 (0,27; 0,69)
Fiebre	0,18 (-0,09; 0,43)	0,17 (-0,11; 0,43)	0,10 (-0,18; 0,37)
Tos	0,92 (0,87; 0,96)	0,92 (0,86; 0,95)	0,90 (0,82; 0,94)
Dolor de garganta	0,69 (0,51; 0,81)	0,65 (0,45; 0,78)	0,57 (0,35; 0,73)
Paracetamol	0,67 (0,49; 0,80)	0,69 (0,51; 0,81)	0,63 (0,42; 0,77)
Ibuprofeno	0,58 (0,36; 0,73)	0,58 (0,36; 0,74)	0,58 (0,37; 0,74)

IC95%: intervalo de confianza del 95%.

Tabla 2

Coefficientes de regresión con sus errores estándar y valor de p para el modelo de regresión de Poisson estimado.

Variable	Coefficiente de regresión (β)	Error estándar (S_{β})	Valor de p
Intersección	-9.510,5	3.999,4	0,017
Gripe	17,5	27,8	0,530
Fiebre	-8,7	32,2	0,787
Tos	321,6	27,2	< 0,001
Dolor de garganta	90,2	19,4	< 0,001
Paracetamol	27,4	20,2	0,175
Ibuprofeno	20,4	38,6	0,596

a la real, aunque también acompañan la dinámica temporal. El modelo reducido basado en el GT se presenta como poco sensible para acompañar la incidencia de casos notificados al SNVS, con ligeras variaciones en la incidencia estimada a lo largo de los dos años.

Discusión

En el presente trabajo se ha observado que los términos de búsqueda en Internet más relacionados con la incidencia de ETI fueron los relativos a los síntomas más comunes: *tos* y *dolor de garganta*. Esto coincide con las fuertes correlaciones halladas en estudios realizados en Corea de Sur, China y España para términos relacionados con los síntomas como *fiebre*, *tos* y *dolor de garganta*, en sus respectivos idiomas^{19,24,25}. Los restantes términos, vinculados a la definición de la enfermedad y al tratamiento, no fueron significativos en el modelo de regresión, aunque fueron conservados en el modelo completo para un

mejor control de potenciales sesgos²⁶. De hecho, aunque en este estudio no presentaron asociación significativa, los términos referidos al tratamiento presentaron buenas correlaciones en otros trabajos, por ejemplo el *Tamiflu*¹⁹. En este estudio no se utilizaron términos tan específicos como *influenza* o *Tamiflu*, ya que éstos son más utilizados por los médicos y no por la población general. De todas maneras, para representar en forma más exacta el pensamiento del público, respecto a los términos de búsqueda usados a nivel de país, es imprescindible realizar una encuesta poblacional o una encuesta a pacientes¹⁹.

Tanto los modelos basados en los términos del GT, como el GFT, presentaron fuertes correlaciones con los datos de ETI del SNVS. Debe observarse que estas correlaciones corresponden a resultados de la validación externa de los modelos, o sea, del desempeño frente a datos de un año diferente a la obtención de los parámetros. En otros estudios de validación del GFT se han observado también altas correlaciones, en diversas regiones de Estados Unidos^{12,13,27}. Sin

Figura 2

Correlaciones entre la incidencia de enfermedades tipo influenza (ETI), la incidencia estimada por los modelos basados en el Google Trends (GT) y la estimada por el Google Flu Trends (GFT) para el año 2013.

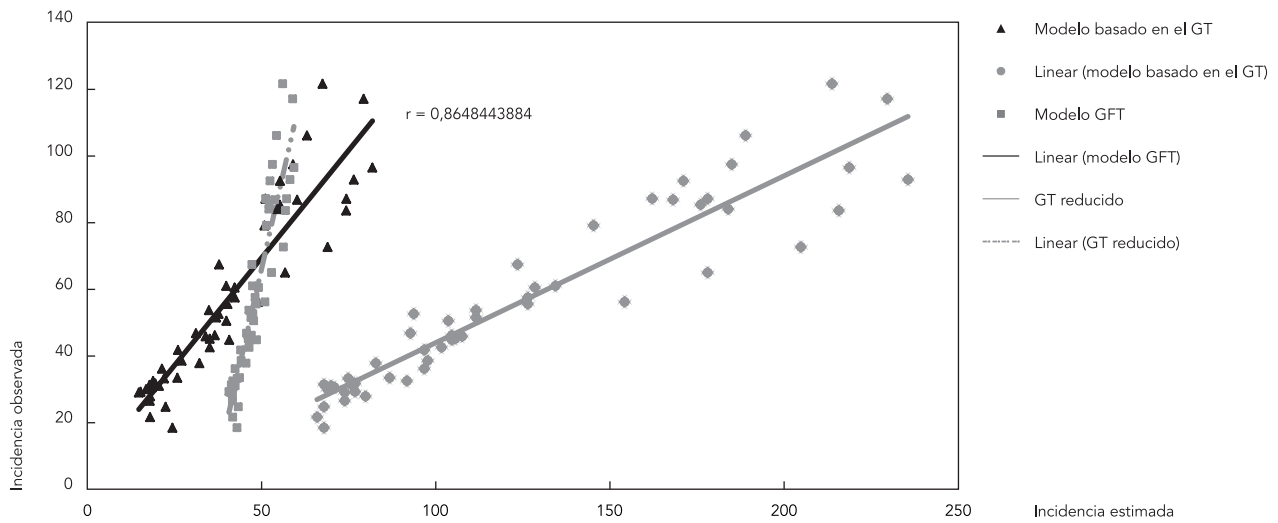
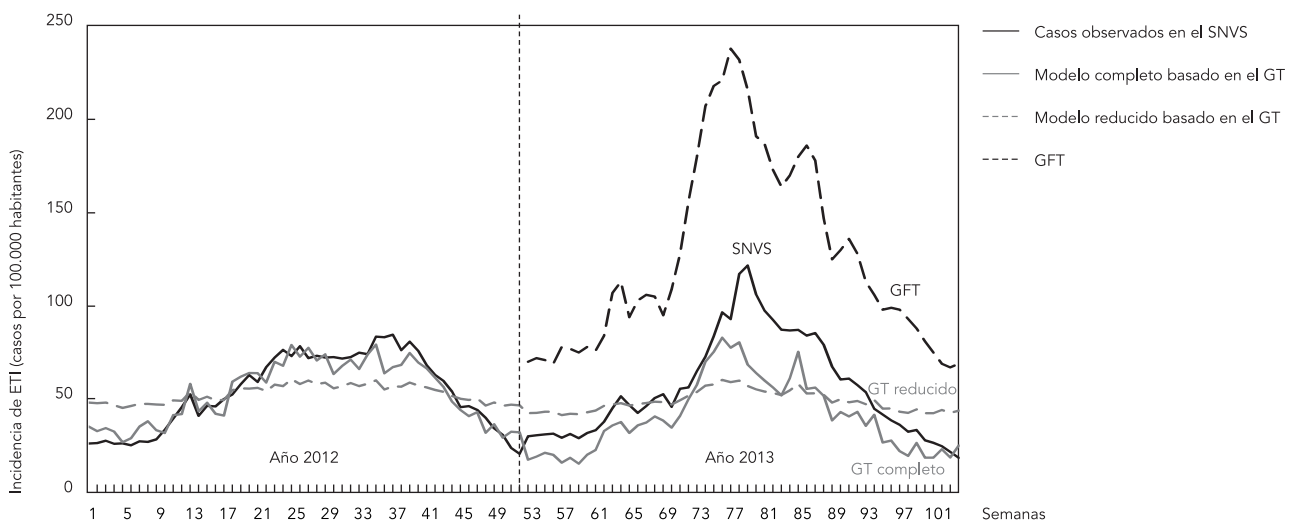


Figura 3

Incidencia de enfermedades tipo influenza (ETI) por semana notificados al Sistema Nacional de Vigilancia de la Salud (SNVS) de Argentina, y estimación de la incidencia según los modelos basados en el Google Trends (GT) y el Google Flu Trends (GFT).



embargo, aunque las correlaciones fueron fuertes, éstas describen la dinámica de unos datos respecto de otros, en lugar de comparar las magnitudes absolutas de las incidencias estimadas y las observadas. Se debe tomar en consideración que las correlaciones simples pueden ser inadecuadas para caracterizar el desempeño de estos modelos, y deben utilizarse en forma complementaria medidas de la intensidad de los brotes y el acople temporal entre los modelos y los datos de vigilancia¹¹. En este sentido, el GFT ha presentado inexactitudes respecto a las incidencias observadas, tanto sobreestimando como subestimando la incidencia real de la influenza, la ocurrencia temporal de los picos de incidencia y la aparición de patrones propios del GFT no reflejados en los datos observados^{11,16,17,28}. En el presente estudio, el GFT presenta una gran sobreestimación respecto a los casos observados notificados al SNVS, con casi el doble estimado respecto al observado durante el pico de incidencia. Por otro lado, el modelo completo basado en el GT presenta valores más cercanos a la incidencia real. El modelo reducido basado en el GT resulta poco sensible, sobreestimando y subestimando la incidencia observada durante todo el período considerado. Esto refuerza la decisión de incluir variables de predicción relevantes, aun las que no presentan asociación significativa individualmente. Son varios los factores que pueden contribuir al mejor desempeño de un modelo confeccionado a partir de datos de búsqueda local de términos en Internet. Por ejemplo, el impacto diferencial de los medios de comunicación en las diferentes sociedades, los períodos de difusión de alertas y reportes, la diferencia en las incidencias relativas entre países, diferentes comportamientos de búsqueda de los pacientes, etc. Además, los parámetros del modelo GFT se actualizan en forma permanente y se redefinen los algoritmos, en pos de un mejor ajuste¹⁷. En este sentido, el desarrollo de un modelo local permite una actualización más dinámica, y un mecanismo más transparente en el manejo de los modelos y los parámetros que relacionan los términos de búsqueda y las incidencias observadas. Justamente para mejorar esta dinámica, en el presente modelo se utilizó solamente un año previo para la estimación de los parámetros, en lugar de varios años. En el futuro se deberá estudiar la conveniencia de utilizar uno o más años para la obtención de los parámetros del modelo. Por otro lado, en el presente modelo no se consideró un retraso temporal entre las búsquedas de los términos en Internet y las incidencias observadas. En otros estudios se han utilizado con distinto éxito retrasos de una o dos semanas. Tanto en algunas regiones de Estados Unidos,

como en Canadá y en España, se han observado mejores correlaciones cuando se toman en cuenta retrasos^{14,25,29}, mientras que en otras regiones de Corea del Sur, China, Estados Unidos y Nueva Zelanda la ausencia de retraso produjo las mejores correlaciones^{19,24,27,28}. En el presente modelo no se consideró el retraso temporal, debido a que la enfermedad tiene un desarrollo corto y abrupto, por lo cual se asumió que las búsquedas en Internet no deberían preceder por muchos días a la consulta médica. Esta suposición fue reforzada por el análisis de las correlaciones simples para distintos retrasos temporales, ya que las mejores correlaciones se observaron cuando no se consideró el retraso temporal. Aun así, esta suposición debería ser puesta a prueba nuevamente en futuros estudios. Otras mejoras que pueden ser introducidas para optimizar el uso de modelos basados en términos de búsqueda son la aplicación de modelos de regresión multinivel, en lugar de ecuaciones de estimación generalizada, regresión con modelos de series temporales considerando estacionalidad¹⁷, la incorporación de variables meteorológicas de ajuste³⁰, modelos de simulación³¹, el análisis de outliers¹⁸, etc.

El presente estudio demuestra el uso de la herramienta GT como complemento a la vigilancia de casos de ETI. Esta herramienta, disponible a través de Google, es de acceso gratuito. Además, el desarrollo del modelo no requiere grandes recursos, y no precisa una capacitación avanzada en metodología estadística y epidemiología. Por consiguiente, los costos asociados al uso de la herramienta son bajos. En contraste, su utilidad puede ser considerable. Un aspecto que no se ha explorado en este estudio sobre los modelos relacionados al GT es la realización de pronósticos futuros de incidencia. Debido a la menor precisión de estas herramientas frente a los métodos tradicionales de vigilancia, se estima que los mejores pronósticos serían las proyecciones basadas en los datos observados por estos últimos, aun tomando en cuenta un retraso de hasta 3 semanas¹⁷. Por lo tanto, se prefirió utilizar el GT para estimar incidencias en tiempo presente, sin ensayar pronósticos futuros.

Son varias las limitaciones del presente estudio. Por un lado, la falta de precisión en cuanto a la definición de las ETI, la cual incluye a la influenza y a otras enfermedades febriles agudas. Esta falta de precisión es propia de la utilización de estas herramientas, ya que las búsquedas en Internet están dirigidas naturalmente a términos inespecíficos, como la tos. A su vez, la elección de estos términos complica aún más la precisión del método. Los síntomas, por ejemplo, son comunes a muchas otras enfermedades, incluso algunas sin ninguna relación con la influenza. Lo

mismo ocurre con algunos medicamentos para aliviar los síntomas, como el ibuprofeno. En este trabajo no se llevó adelante ningún estudio sobre la asociación entre los términos elegidos y las búsquedas de la población respecto a la influenza. Se deberían realizar estudios cualitativos que permitan describir cuales son los términos que los pacientes buscan para cada enfermedad, en particular para la influenza, de acuerdo a las realidades locales. En este estudio se tomó en cuenta para la regresión la autocorrelación entre las semanas. Es posible que los datos de incidencia de los meses hayan presentado algún nivel de autocorrelación también, ya que la incidencia de un mes podría influir en la incidencia del mes siguiente. Si bien la relación entre semanas contiguas debería ser mucho mayor, no debería descartarse una relación también entre meses subsiguientes. Por último, el SNVS tiene una mayor representación del sistema de salud público (70% de establecimientos cubiertos) respecto del privado (6% de establecimientos cubiertos) ³². En este sentido, los casos sobreestimados por el GFT podrían ser en realidad casos de ETI que consultaron en el sistema privado y no fueron notificados al SNVS. En contraste, el modelo basado en el

GT utiliza al mismo SNVS para ser parametrizado, y por lo tanto es consistente con los resultados de vigilancia observados, los cuales podrían presentar fuertes subregistros. Si esta limitación es tomada en cuenta, el GFT podría ser un mejor indicador de la incidencia real, mientras que el modelo basado en el GT sería un mejor indicador de las notificaciones registradas en el sistema de vigilancia.

En conclusión, el presente es el primer estudio en analizar el uso de herramientas de búsqueda de términos en Internet a través de Google en Argentina. Los resultados muestran la utilidad de esta herramienta para estimar la incidencia de las ETI sobre una base semanal, presentando altos valores de correlación entre los casos estimados y los observados por el sistema de vigilancia, y una moderada relación en las incidencias absolutas. En el futuro, es esperable que la aplicación de métodos de modelado superiores, y el estudio de la relación entre la influenza y los términos de búsqueda en Internet por parte de la población, puedan mejorar el uso de este potente instrumento para su utilización como apoyo a la vigilancia epidemiológica de esta enfermedad.

Resumo

O objetivo deste estudo foi encontrar um modelo para estimar a incidência da síndrome gripal com base nos termos de busca relacionados recolhidos pelo Google Trends (GT). Os dados de monitoramento de síndrome gripal entre 2012 e 2013 foram obtidos no Sistema Nacional de Vigilância de Saúde da Argentina. As pesquisas na Internet foram feitas baseando-se no banco de dados do GT usando 6 termos: gripe, febre, tosse, dor de garganta, paracetamol e ibuprofeno. Um modelo de regressão de Poisson foi desenvolvido valendo-se de dados de 2012. O modelo foi ajustado e validado com dados

de 2013 e comparado com os resultados da ferramenta Google Flu Trends (GFT). A incidência de síndrome gripal mostrou uma forte correlação com estimativas do GT ($r = 0,927$) e GFT ($r = 0,943$). No entanto, o GFT superestimou o pico de incidência por quase o dobro, e o modelo baseado no GT subestimou o pico de incidência por um fator de 0,7. Esses resultados demonstram a utilidade do GT como um suplemento para a vigilância da influenza.

Gripe Humana; Modelos Epidemiológicos; Incidência

Colaboradores

P. W. Orellano participó de la idea y diseño del estudio, procesamiento y análisis de los datos, redacción del trabajo original. J. I. Reynoso colaboró en el análisis de los datos y redacción del trabajo original. J. Antman y O. Argibay colaboraron en el procesamiento de los datos y revisión final del trabajo original.

Agradecimientos

A los revisores anónimos por las sugerencias aportadas para la mejora del manuscrito desde su versión original.

Referencias

1. Souza MF, Widdowson MA, Alencar AP, Gawryszewski VP, Aziz-Baumgartner E, Palekar R, et al. Trends in mortality from respiratory disease in Latin America since 1998 and the impact of the 2009 influenza pandemic. *Bull World Health Organ* 2013; 91:525-32.
2. Bonvehí PE, Istúriz RE, Labarca JA, Rüttimann RW, Vidal EI, Vilar-Compte D. Influenza among adults in Latin America, current status, and future directions: a consensus statement. *Rev Panam Salud Pública* 2012; 31:506-12.
3. Azziz-Baumgartner E, Cabrera AM, Chang L, Calli R, Kuszniarz G, Baez C, et al. Mortality, severe acute respiratory infection, and influenza-like illness associated with influenza A(H1N1)pdm09 in Argentina, 2009. *PLoS One* 2012; 7:e47540.
4. Instituto Nacional de Estadísticas y Censos. Estimaciones y proyecciones de población: total del país, 1950-2015: población por sexo y grupos quinquenales de edad. (Serie Análisis Demográfico, 30). http://estadisticas.tierradelfuego.gov.ar/wp-content/uploads/2013/11/Estimaciones_Proyecciones_País_1950_2015.pdf (accedido el 11/Abr/2014).
5. Bloom-Feshbach K, Alonso WJ, Charu V, Tamerius J, Simonsen L, Miller MA, et al. Latitudinal variations in seasonal activity of influenza and respiratory syncytial virus (RSV): a global comparative review. *PLoS One* 2013; 8:e54445.
6. Lindblade KA, Johnson AJ, Arvelo W, Zhang X, Jordan HT, Reyes L, et al. Low usage of government healthcare facilities for acute respiratory infections in Guatemala: implications for influenza surveillance. *BMC Public Health* 2011; 11:885.
7. Vandendijck Y, Faes C, Hens N. Eight years of the Great Influenza Survey to monitor influenza-like illness in Flanders. *PLoS One* 2013; 8:e64156.
8. Eysenbach G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. *AMIA Annu Symp Proc* 2006; 2006:244-8.
9. Polgreen PM, Chen Y, Pennock DM, Nelson FD. Using internet searches for influenza surveillance. *Clin Infect Dis* 2008; 47:1443-8.
10. Carneiro HA, Mylonakis E. Google trends: a web-based tool for real-time surveillance of disease outbreaks. *Clin Infect Dis* 2009; 49:1557-64.

11. Olson DR, Konty KJ, Paladini M, Viboud C, Simonsen L. Reassessing Google Flu Trends data for detection of seasonal and pandemic influenza: a comparative epidemiological study at three geographic scales. *PLoS Comput Biol* 2013; 9:e1003256.
12. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature* 2009; 457:1012-4.
13. Cook S, Conrad C, Fowlkes AL, Mohebbi MH. Assessing Google Flu Trends performance in the United States during the 2009 influenza virus A (H1N1) pandemic. *PLoS One* 2011; 6:e23610.
14. Malik MT, Gumel A, Thompson LH, Strome T, Mahmud SM. "Google Flu Trends" and emergency department triage data predicted the 2009 pandemic H1N1 waves in Manitoba. *Can J Public Health* 2011; 102:294-7.
15. Valdivia A, Lopez-Alcalde J, Vicente M, Pichule M, Ruiz M, Ordoñas M. Monitoring influenza activity in Europe with Google Flu Trends: comparison with the findings of sentinel physician networks – results for 2009-10. *Euro Surveill* 2010; 15:pii:19621.
16. Butler D. When Google got flu wrong. *Nature* 2013; 494:155-6.
17. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. *Science* 2014; 343:1203-5.
18. Ortiz JR, Zhou H, Shay DK, Neuzil KM, Fowlkes AL, Goss CH. Monitoring influenza activity in the United States: a comparison of traditional surveillance systems with Google Flu Trends. *PLoS One* 2011; 6:e18687.
19. Cho S, Sohn CH, Jo MW, Shin SY, Lee JH, Ryoo SM, et al. Correlation between national influenza surveillance data and google trends in South Korea. *PLoS One* 2013; 8:e81422.
20. Patwardhan A, Bilkovski R. Comparison: flu prescription sales data from a retail pharmacy in the US with Google Flu trends and US ILINet (CDC) data as flu activity indicator. *PLoS One* 2012; 7:e43611.
21. Myers RH, Montgomery DC, Geoffrey Vining G. Generalized linear models: with applications in engineering and the sciences. 2nd Ed. New Jersey: Wiley and Sons; 2010.
22. Hardin JW, Hilbe JM. Generalized estimating equations. New York: Chapman & Hall/CRC; 2003.
23. Højsgaard S, Halekoh U, Yan J. The R package gee-pack for generalized estimating equations. *J Stat Soft* 2006; 1:1-11.
24. Kang M, Zhong H, He J, Rutherford S, Yang F. Using Google Trends for influenza surveillance in South China. *PLoS One* 2013; 8:e55205.
25. Valdivia A, Monge-Corella S. Diseases tracked by using Google Trends, Spain. *Emerg Infect Dis* 2010; 16:168.
26. Kleinbaum DG, Klein M. Logistic regression: a self learning text. 3rd Ed. New York: Springer; 2010.
27. Dugas AF, Hsieh YH, Levin SR, Pines JM, Mareiniss DP, Mohareb A, et al. Google Flu Trends: correlation with emergency department influenza rates and crowding metrics. *Clin Infect Dis* 2012; 54:463-9.
28. Wilson N, Mason K, Tobias M, Peacey M, Huang QS, Baker M. Interpreting Google Flu Trends data for pandemic H1N1 influenza: the New Zealand experience. *Euro Surveill* 2009; 14:pii:19386.
29. Ortiz JR, Zhou H, Shay DK, Neuzil KM, Fowlkes AL, Goss CH. Monitoring influenza activity in the United States: a comparison of traditional surveillance systems with Google Flu Trends. *PLoS One* 2011; 6:e18687.
30. Dugas AF, Jalalpour M, Gel Y, Levin S, Torcaso F, Igusa T, et al. Influenza forecasting with Google Flu Trends. *PLoS One* 2013; 8:e56176.
31. Nsoesie E, Marathe M, Brownstein J. Forecasting peaks of seasonal influenza epidemics. *PLoS Currents* 2013; 5:ecurrents.outbreaks.bb1e-879a23137022ea79a8c508b030bc.
32. Giovanella L, Feo O, Faria M, Tobar S. Sistemas de Salud en Suramérica: desafíos para la universalidad, la integralidad y la equidad. Río de Janeiro: Instituto Suramericano de Gobierno en Salud; 2012.

Recibido el 08/May/2014

Versión final presentada el 03/Nov/2014

Aprobado el 08/Dic/2014