

Boletín GripeNet Temporada 2, 5º entrega:

Big data y Epidemiología: Google flu trends (3): ¿Cuáles son sus límites?

En las últimas semanas venimos analizando Google flu trends, la herramienta del gigante informático para estimar y predecir la influencia gripal a través del análisis de las entradas en su buscador. En el último boletín, hablábamos de los dos principales escollos que los investigadores tuvieron que solventar para proporcionar una herramienta fiable y robusta: la elección de los términos de búsqueda como indicadores fidedignos de la existencia de un cuadro gripal y el establecimiento de una escala adecuada que les permitiese, temporada tras temporada, re-escalar el número de búsquedas en Google hasta obtener una estimación de la incidencia gripal en cada momento y lugar.

Como veremos hoy, el éxito de Google flu trends es, cinco años después de su implementación, un éxito relativo. De los cinco picos epidémicos que han tenido lugar en el hemisferio norte desde su funcionamiento, éste ha emitido predicciones seriamente desviadas en algunas áreas en dos ocasiones: durante la atípica temporada gripal del 2009 y después de que Google reajustara los algoritmos tras el primer error el pasado invierno. En este boletín hablaremos de estos problemas y de sus posibles causas, que, como veremos, imponen una frustrante cortapisa conceptual a este tipo de métodos de observación epidemiológica indirecta.

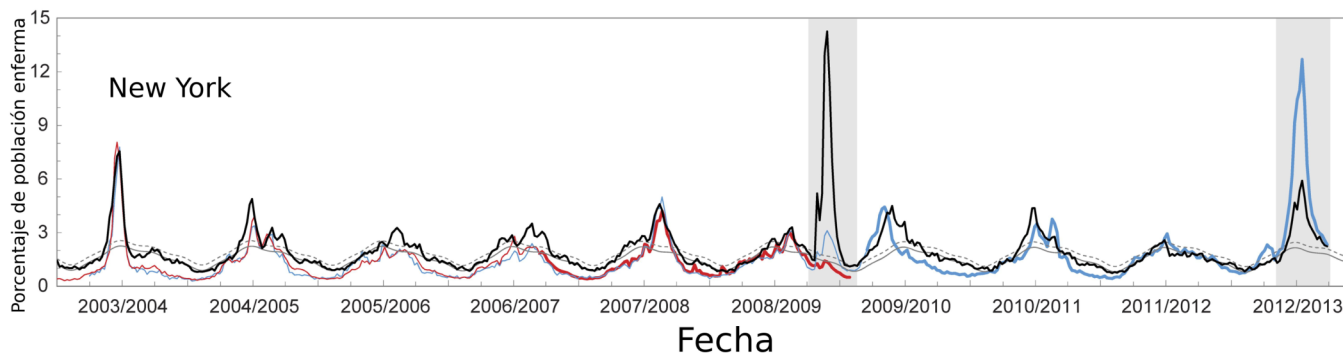
Los mil nombres de la gripe de 2009.

Como ya sabrá el lector asiduo a este Boletín, la aparición de una nueva variante de gripe –la consabida “gripe-A”- en la primavera de 2009 supuso un auténtico terremoto mediático no sólo dentro de la comunidad científica, médica y epidemiológica, sino también para la población en general.

Tal y como expusimos en la temporada pasada, la rapidez y contundencia con la que el nuevo virus se propagó por todos los rincones del planeta, sólo se vio superada por el estupor y la confusión generalizados entre la población, alimentados por una torre de Babel mediática de alcance global. **¿Qué era aquello? ¿Era gripe o era más grave que la gripe? ¿Había vacuna? ¿Era segura? Y por encima de todo: ¿Cuál era el nombre más adecuado para el nuevo virus? ¿Gripe mexicana, gripe porcina, gripe nueva o gripe A?** A día de hoy, una gran parte de la población mundial desconoce las respuestas a algunas de estas sencillas preguntas.

Fue precisamente la novedad del virus, unida a toda esta vaguedad terminológica difundida a través de los medios de comunicación lo que provocó un efecto inesperado en el comportamiento de los usuarios a la hora de realizar consultas en Google. La gente, preocupada por las noticias sobre el nuevo virus, aunque seguía buscando en Internet información sobre la gripe, lo hacía introduciendo nuevos términos en el buscador. El efecto de este cambio de comportamiento, obviamente, no pudo ser previsto por Google flu trends con antelación: en el 2008, cuando el algoritmo se implementó, nadie podía imaginar que sólo unos meses más tarde términos como “gripe mexicana” o “gripe A” iban a estar entre los más buscados en la red por personas preocupadas por la enfermedad.

Esta situación se tradujo en que el algoritmo de Google fue esencialmente incapaz de predecir el pico epidémico en la primavera de 2009 en localizaciones como la ciudad de Nueva York, tal como vemos en la siguiente imagen:

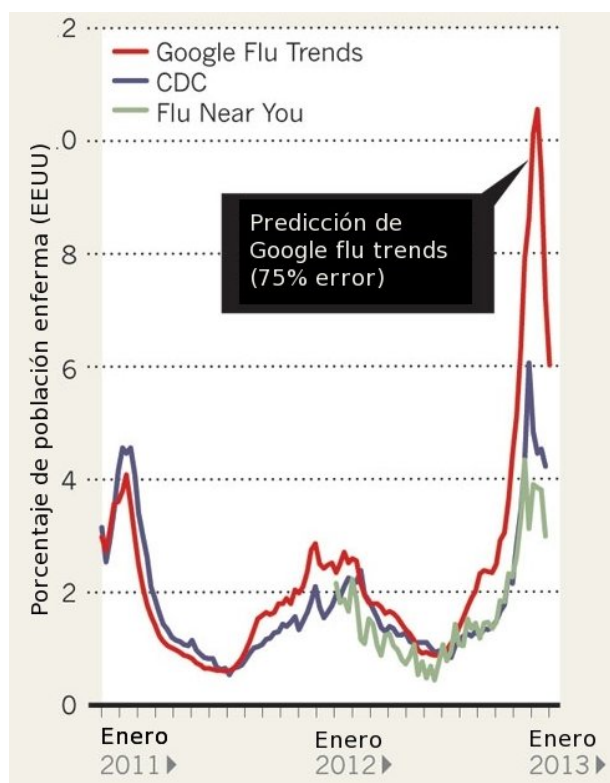


Estimaciones de la incidencia gripal para la ciudad de Nueva York. Serie negra: mediciones de la red de médicos centinela del CDC. Rojo: google flu trends (primera versión) Azul celeste: Google flu trends (recalibrado). Como puede observarse (franjas grises), Google flu trends no previó la alta incidencia de gripe en la primavera de 2009 en la ciudad de Nueva York; ciudad en la que el método también falló el pasado invierno. Imagen adaptada desde <http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1003256>

A raíz del problema, los responsables del método desarrollaron una nueva versión del mismo en la cual todos estos nuevos términos relacionados con la nueva variedad de gripe fueron reintroducidos (en azul en la imagen), recalibrando así el algoritmo.

Búsquedas en Google e infecciones: dos magnitudes (elásticamente) ligadas

El reajuste del método tras el fiasco del año 2009, volvió de nuevo a dar buenos resultados durante los inviernos de los años 2010, 2011 y 2012, tal y como veíamos en la anterior figura. Sin embargo, como se ve más abajo, para la población de Estados Unidos (no sólo para la de Nueva York), el nuevo método predijo el pasado invierno un pico epidémico el doble de severo que lo que se observó realmente, lo cual supone un fallo igual de grave que infravalorar la incidencia gripal.



¿Cuál fue la razón de este segundo fallo, si se supone que los términos de búsqueda no cambiaron con respecto al año anterior?

En este caso, si bien la situación no está tan clara como en el 2009, todo apunta a que la razón estriba en la potente campaña que las autoridades sanitarias estadounidenses lanzaron en los medios de comunicación en aras de fomentar medidas de prevención ante una gripe, que ese invierno, se esperaba más virulenta de lo habitual.

Estimaciones de la incidencia gripal para EEUU. El pasado invierno, Google Flu trends aportó una predicción de un 10.5% de prevalencia gripal entre la población durante el pico de la epidemia. Esta predicción resultó ser un 75% mayor que los datos reportados por la red de médicos del centro de control de enfermedades (CDC). Flu Near You es una plataforma de vigilancia participativa similar a GripeNet (<https://flunearyou.org/>). Imagen adaptada desde <http://www.nature.com/news/when-google-got-flu-wrong-1.12413>

Dicha campaña, lanzada a través de los medios de comunicación avivó de manera temporal el tema de la gripe entre la población, incentivando a los individuos a concienciarse del riesgo epidemiológico y a informarse por su cuenta, lo cual aumentó el tráfico de información relacionado con la gripe en el buscador.

Como consecuencia, el algoritmo de Google flu trends, incapaz de valorar estas circunstancias, interpretó que el aumento de búsquedas no podía sino deberse a un aumento de la incidencia, emitiendo una predicción exageradamente superior a lo que finalmente se registraría.

Conclusiones: los problemas intrínsecos de la vigilancia epidemiológica indirecta

Como hemos detallado en los últimos boletines, la vigilancia epidemiológica indirecta consiste en la estimación de los niveles de incidencia de una enfermedad a través de la medida de otra variable más accesible y estrechamente relacionada con la anterior. En el caso de Google flu trends, esta variable es la cantidad de búsquedas relacionadas con la enfermedad que la población realiza en el principal buscador de la red.

Los dos errores descritos en las predicciones de Google flu trends están relacionados con los principales aspectos del método. **Por un lado, el problema en el 2009 fue un cambio súbito e impredecible de los términos tradicionales usados por la población en el buscador para hacer consultas sobre la gripe, mientras que en 2012 lo que varió fue la escala cuantitativa entre cantidad de búsquedas y número de enfermos.**

Esta situación pone de manifiesto la limitación intrínseca a este tipo de métodos, ya que los fenómenos que produjeron los fallos en el 2009 y en el 2012 pueden repetirse en cualquier momento y originar nuevos errores en las predicciones; incluso aunque los algoritmos se recalibren cada año. Los términos de búsqueda relacionados con la incidencia gripal dependen de un gran número de factores sociológicos coyunturales que pueden cambiar sin previo aviso en cualquier momento.

Por todas estas razones, resulta difícil concebir un futuro sistema de vigilancia epidemiológica de la gripe netamente indirecto que no se complemente con las actuales redes de vigilancia epidemiológica formadas por médicos de atención primaria o con otros enfoques orientados en esta dirección como el modelo participativo de GripeNet.

En el próximo boletín analizaremos otro enfoque distinto al problema de la vigilancia epidemiológica: la vigilancia participativa. Repasaremos las iniciativas de vigilancia epidemiológica participativa que se están desarrollando fuera de Europa y con las que la red europea InfluenzaNet (de la que GripeNet.es forma parte) colabora activamente. Hablaremos de los distintos enfoques, las diferencias y los problemas a los que se enfrentan.

¡Feliz semana a todos!