CIENCIAS
UNIVERSIDAD DE LOS ANDES
MÉRIDA VENEZUELA

DEPARTAMENTO DE BIOLOGÍA

Trabajo especial de grado

"Redes sociales como herramienta de epidemiología digital para la salud pública"

Autor: Br. Ilvanna Salas León

Tutor: Ascanio Rojas A.

Tutor Académico: Wilfredo Quiñones

Mérida, noviembre de 2017

Resumen

La epidemiología es una herramienta esencial para la vigilancia de salud pública. Tradicionalmente la información epidemiológica proviene de organismos de salud, donde el personal recibe y procesa los datos para la posterior toma de decisiones, pero este proceso es lento, subjetivo y costoso. Recientemente los medios digitales como redes sociales se han implementado como nuevas fuentes de datos que aportan información amplia, en tiempo real, a un menor costo y puede incluir una población mucho más amplia. La cantidad de datos obtenidos de estas fuentes son abrumadores, y su análisis excede la capacidad humana. Es por esto que han surgido nuevas herramientas como la extracción de conocimiento de bases de datos que permiten su colección y análisis para la toma de decisiones basadas en evidencias, un factor muy importante en salud pública, ya que permite la prevención de eventos, estudios focalizados y un mejor uso de los recursos. En este estudio se plantea que los medios como redes sociales son útiles como herramientas para epidemiología digital en Venezuela. Para probarlo se usaron datos formales (boletines epidemiológicos de Venezuela 2004-2016), datos informales (tendencias de Google 2004- 2017), y se evaluó si esta información podía aplicarse en el diseño de un modelo epidemiológico predictivo para las enfermedades incluidas (dengue, diarrea, difteria, fiebre, hepatitis, leishmaniasis, rubéola sarampión y VIH). Los datos informales fueron corroborados con información de la Organización Mundial de la Salud, y se evidenció que el comportamiento de las tendencias de Google puede relacionarse con eventos epidemiológicos en la población; por lo tanto, se sugiere su aplicación como fuente de información para alertas tempranas y respuesta, para actualizar y complementar el sistema de salud venezolano.

Palabras clave: Salud pública, alerta temprana, vigilancia epidemiológica, redes sociales, tendencias.

“Los sistemas de transporte se mueven más rápido que un período de incubación de una enfermedad y los profesionales de la salud pública se limitan a sus propios esfuerzos humanos. A medida que evolucionan las fuentes de enfermedades, nuestros métodos de prevención también deberían. No estamos seguros de si nuestra acción salvará vidas, pero ciertamente estamos seguros de que la inacción mata.”

Dhesi Raja & Rainier Mallol

Agradecimientos

Comienzo por agradecera Venezuela, mi país, que me ha hecho ser lo que soy, agradezco a mi alma mater, la Universidad de Los Andes, donde me formé no solo como profesional, sino como un mejor ciudadano. A la facultad de ciencias, todos los profesores y estudiantes que hacen vida en ella, por cultivar en mí el pensamiento científico, la disciplina, pasión por la ciencia y un sinfín de aportes de los que estar agradecida. A Ascanio Rojas, mi tutor, por su orientación y paciencia, innumerables enseñanzas que me abrieron nuevas puertas en el mundo del conocimiento, por escucharme y aconsejarme sobre la vida per se.

Agradezco a la vida por haberme puesto en una familia maravillosa, pues no podría llegar a tan alto si no estuviera parada “sobre hombros de gigantes”. Le doy las gracias a Dayra León, mi mamá, por tanto, amor, esfuerzo y vocación, siempre con una hermosa sonrisa. A Tulio Salas, mi papá, por ser mi apoyo incondicional, por ser mi mejor amigo y haber creído en mí siempre. A Dayvanna Salas, mi hermana, por ser mi protectora, amiga y compañía durante toda mi vida. A mis abuelos, tíos y primos (quisiera mencionarlos a todos, pero no me alcanzaría el espacio) gracias por su amor, apoyo y compañía.

Le doy las gracias a Alejandro Rodríguez, mi pareja perfecta, por su amor, compañía y apoyo, por creer ciegamente en mí y mis sueños, acompañarme y ayudarme en la construcción de mi futuro.

A mis amigos, por escucharme y aconsejarme, por muchas tortas, tazas de té, cervezas y comida chatarra compartida, ¡sin eso no podría decir que pasé por la universidad!

Índice

Introducción

La epidemiología es una herramienta esencial para realizar cuatro funciones fundamentales: la vigilancia de salud pública en salud pública, la investigación de enfermedades, los estudios analíticos, y la evaluación de programas. Mediante la vigilancia de salud pública un servicio de salud recoge, analiza, interpreta y divulga los datos de salud, en forma continua y sistemática. Un servicio local de salud utiliza la vigilancia en salud pública para evaluar la salud de una comunidad. Cuando un departamento dispone de información sobre los patrones de ocurrencia de las enfermedades y el potencial de otros daños a la salud, puede investigar, prevenir y controlar efectivamente las enfermedades de la comunidad. [1]

Tradicionalmente, la epidemiología se ha basado en datos de los organismos de salud pública mediante el personal de salud, hospitales, consultorios médicos y en el campo. Sin embargo, en años recientes han surgido nuevas fuentes de datos donde estos son recogidos directamente de los individuos siguiendo los rastros que dejan como consecuencia de las nuevas formas de comunicación y un mayor uso de dispositivos electrónicos. [2]

Estas fuentes en línea proporcionan una imagen de la salud global que a menudo es diferente de la imagen creada por los sistemas de vigilancia tradicionales. De hecho, estos datos se han convertido en fuentes de datos invaluables para la generación de nuevos sistemas de vigilancia de la salud pública que operan en fronteras, llenan los vacíos en la infraestructura de salud pública y complementan los sistemas tradicionales de vigilancia existentes. Además, se obtiene información de las epidemias semanas, o incluso meses, antes que la proporcionada por métodos de epidemiología clásica, implicando un menor costo, menor subjetividad y una población de estudio más amplia. [2] [3] [4]

Desde el punto de vista de la minería de datos, una red social es un conjunto de datos heterogéneos y multirrelacionados representados por un gráfico. El gráfico es típicamente muy grande, con nodos que corresponden a objetos y "*edges*" que corresponden a enlaces que representan relaciones o interacciones entre los objetos. Tanto nodos como enlaces tienen atributos. Los objetos pueden tener etiquetas de clase. Los enlaces pueden ser unidireccionales y no se requiere que sean binarios. [5]

Las redes sociales no necesariamente son de contexto social. En el mundo real hay muchas redes, tecnológicas, de negocios, económicas y biológicas. Algunos ejemplos son gráficos de llamadas telefónicas, la propagación de virus computacionales y la World Wide Web (WWW). En biología los

ejemplos abarcan desde redes epidemiológicas, redes celulares y metabólicas, hasta la red neuronal de *Caenorhabditis elegans*(la única criatura cuya red neuronal ha sido completamente mapeada). Por otro lado, el intercambio de mensajes por email en corporaciones, salas de chat, sex webs, son ejemplos de sociología. [5]

Las redes sociales (del pequeño mundo) han recibido una atención considerable recientemente, y reflejan el concepto del "mundo pequeño" que originalmente se enfocó en redes entre individuos. La frase "qué pequeño es el mundo" captura la sorpresa inicial entre dos extraños cuando se dan cuenta de que están conectados indirectamente con el otro por algún conocido en común. [5]

En 1967 un sociólogo de Harvard, Stanley Milgram y sus colegas, condujeron un experimento en el cual a personas de Kansas y Nebraska se les pidió dirigir una carta a extraños en Boston, dándola a un amigo que creyeran que podía conocer a los extraños en Boston. La mitad de las cartas se entregaron exitosamente con no más de 5 intermediarios. Otros estudios realizados por Milgram y otros, conducidos en otras ciudades, mostraron que parece ser universal los "seis grados de separación" entre un individuo cualquiera y otro en el mundo. [5]

Las redes sociales (*small world*) han sido caracterizadas como poseedoras de un alto nivel de agrupación local de una fracción pequeña de nodos (nodos que están interconectados unos con otros) los cuales a su vez no están a más de un par de grados de separación de los nodos restantes. Se cree que muchas redes sociales, físicas y biológicas, exhiben esas características del "pequeño mundo". ¿Por qué todo este interés en las redes (*small world*) y redes sociales en general? ¿Cuál es el interés en caracterizar las redes y desenterrarlas para aprender acerca de su estructura? La razón es porque la estructura siempre afecta a la función. Por ejemplo, la topología de las redes sociales afecta la propagación de una enfermedad infecciosa a través de una población estructurada. [5]

El Boletín Epidemiológico Semanal fue creado por el Dr. Darío Curiel Sánchez, eminente sanitarista venezolano, cuando se fundó la División de Epidemiología y Estadística Vital, en 1938. Miembro correspondiente de esta Academia, diseñó y ejecutó el Plan de Campaña Nacional Preventiva de Vacunación Antivariólica, con el cual se erradicó la viruela del país. ¿Qué es el Boletín Epidemiológico Semanal? En esencia es el resumen de la situación de salud de nuestro país, mediante las cifras de casos y muertes de enfermedades de notificación obligatoria, comunicadas semanalmente por los médicos que ejercen en cargos públicos o privados. Son enfermedades transmisibles directamente, o a través de vectores, de un enfermo a un sano. ¿Por qué se requiere de esa información? El servicio de epidemiología de un país necesita conocer con la mayor precisión y rapidez los casos y muertes por enfermedades transmisibles que ocurran en su territorio, no sólo para tener información sobre ellas y planear campañas según los problemas locales o tomar decisiones inmediatas, sí son necesarias, sino también para informar prontamente a determinados organismos

internacionales sobre la situación reinante en el país en materia de enfermedades transmisibles, obligación contraída por Venezuela en tratados internacionales. [6] [7]

En Venezuela, no se ha llevado a cabo un estudio que involucre el uso de estas nuevas fuentes de datos, razón por la cual en este trabajo se pretende demostrar que el uso de medios digitales y redes sociales pueden ser útiles como herramientas para la vigilancia epidemiológica de salud pública en el país.

Para llevar a cabo la investigación se usarán los datos de dichos boletines, los cuales serán llamados datos formales, y datos de los medios digitales, que serán llamados datos informales, y luego de ser depurados se aplicarán técnicas de minería de datos, capaces de dar lugar a un modelo matemático que permita hacer la predicción epidemiológica para diferentes enfermedades en Venezuela. A partir de estos resultados se espera conocer la situación de salud actual de país, y poder continuar con el trabajo de estos datos a futuro para garantizar la información necesaria para mejorar la preparación del sector salud, además de motivar la formación de estudiantes en esta área necesarios para la actualización y mejoramiento del sistema de salud venezolano.

CAPÍTULO I - **Marco teórico**

I.1. Epidemiología

La palabra epidemiología viene del griego "*epi*" que significa sobre o encima, "*demos*" que significa población y "*logos*" que significa estudio de. [1]

Es el estudio de la distribución y de los determinantes de los fenómenos relacionados con la salud en poblaciones específicas, y la aplicación de este estudio al control de problemas sanitarios. [8]

Epidemiología y salud pública

Como disciplina de la salud pública, la epidemiología está fundamentada en la concepción de que la información epidemiológica debe ser utilizada para promover y proteger la salud pública. De hecho, la epidemiología supone tanto el quehacer de la ciencia como la práctica de la salud pública. [1]

Salud pública

Se refiere a los esfuerzos colectivos por mejorar la salud de una población. La epidemiología es una herramienta de la salud pública. [8]

Medición de la salud y la enfermedad

Cuantificar la salud y la enfermedad es fundamental en la epidemiología, y a pesar de que existen diversas medidas, esto no se realiza en diversos países, lo cual produce problemas por falta de información. [8]

Enfermedad

Se refiere a todos los cambios desfavorables de la salud, incluyendo lesiones, traumatismos y salud mental. [8]

Salud

En epidemiología, se refiere a la ausencia de enfermedad. [8]

Caso

Una definición de caso es una serie de criterios estandarizados para decidir si una persona tiene una enfermedad particular u otra condición relacionada con la salud. Al utilizar una definición de caso estandarizada, se asegura que cada caso es diagnosticado de la misma manera, independientemente de cuándo o dónde ocurrió o quien lo identificó. Se puede comparar el número de casos de la enfermedad ocurridos en un tiempo y lugar dados con el número de casos ocurridos en otro tiempo y otro lugar. Por ejemplo, con una definición de caso, se pueden comparar el número de casos de influenza H1N1 ocurridos en Caracas durante el 2009 con el número ocurrido durante 2008; además, se puede comparar el número de casos ocurridos en Mérida el mismo año. Con una definición de caso estándar, cuando se encuentra una diferencia en la ocurrencia de la enfermedad, se sabe que es más una diferencia real y no el resultado de posibles diferencias en las maneras en que se hizo el diagnóstico. [1]

Población expuesta al riesgo

Un aspecto importante para cuantificar la frecuencia de enfermedad es estimar correctamente el tamaño de la población que se considera. Es importante que se incluya sólo a las personas potencialmente susceptibles. Por lo tanto, la población expuesta al riesgo es la susceptible a contraer la enfermedad, y puede definirse según factores geográficos, demográficos o ambientales. [8]

Incidencia y Prevalencia

La incidencia es la velocidad en la que se producen nuevos casos, durante un tiempo determinado en una población específica. La prevalencia es la frecuencia de casos de una enfermedad en una población, en un momento dado. [8]

I.2. Redes Sociales

Red Social

Es una estructura social compuesta por un conjunto de actores (como organismos u organizaciones) que están relacionados de acuerdo a algún criterio (relación profesional, amistad, parentesco). Las redes sociales consisten en dos elementos: 1) Los individuos (nodos) y 2) el vínculo social entre ellos. [9]

Una vez que todos los nodos e individuos son conocidos, se puede dibujar una imagen de la red y discernir la ubicación de cada persona dentro de esta, ubicando a cada individuo en un espacio social análogo al espacio geográfico. [9]

Grados de separación

Dentro de una red se puede hablar de la distancia entre dos personas, en la que la más corta es entre una persona y otra. Por ejemplo, una persona está a un grado de separación de su amigo, dos grados de un amigo de su amigo, tres grados del amigo del amigo de su amigo, y así sucesivamente. [9]

Los vínculos sociales

Los vínculos sociales no están restringidos a amigos, ya que un individuo puede estar conectado a su pareja, hermano, compañero de trabajo, entre otros. Para efectos de discusión puede hacerse referencia a los individuos bajo estudio y a la gente a la que están conectados. [9]

Los vínculos sociales se pueden describir como **"*edges*":** relaciones no dirigidas entre nodos (como hermanos, o esposos), o **arcos**: relaciones dirigidas entre un nodo y otro (como dos amigos, en el que A identifica a B como su amigo, pero no es reciproco de B). Ambos pueden ser medidos en escala binaria, o pueden ser valorados. [8]

Tipos de redes sociales

Puede haber redes de: **Un modo** que incluyen un solo tipo de nodo (nodo = pacientes), o de **dos modos** con dos tipos de nodos (nodo = pacientes y doctores). [9]

Análisis de las redes

En su forma más simple, el análisis de redes se enfoca en conexiones entre nodos homogéneos, básicamente en la información sobre los nodos. [9]

Los individuos conectados en una red pueden tener características diferentes, (nivel académico, comportamientos de salud, posiciones políticas, entre otros.). Esto puede ser atribuido a tres procesos:

1. **Homofilia**

Los individuos escogen estar conectados, basados en los atributos o comportamientos compartidos. [9]

2. **Inducción**

Los atributos o comportamientos de una persona causan atributos o comportamientos análogos en otros. [9]

3. **Confusión**

Individuos conectados conjuntamente experimentan exhibiciones que los hacen compartir un atributo o comportamiento. [9]

I.1. Descubrimiento del conocimiento en bases de Datos (en inglés Knowledge Discovery in Databases -KDD)

La frase “Descubrimiento de conocimiento en bases de datos” fue acuñada por primera vez en el primer taller de KDD en 1989, para enfatizar que el conocimiento es el producto final de un descubrimiento basado en datos, y se ha popularizado en campos de inteligencia artificial y aprendizaje automático. [3]

El proceso de descubrimiento conocido como *Knowledge Discovery in Databases* (KDD) se refiere al proceso no trivial de descubrir conocimiento e información potencialmente útil dentro de los datos

contenidos en algún repositorio de información. No es un proceso automático, es un proceso iterativo que exhaustivamente explora grandes volúmenes de datos para determinar sus relaciones. Es un proceso que extrae información de calidad que puede usarse para dibujar conclusiones basadas en relaciones o modelos dentro de los datos. [3]

En un nivel abstracto, el campo del KDD está dirigido a transformar datos muy voluminosos y difíciles de entender, a formas más compactas, como una aproximación descriptiva, o un modelo generado por los datos. En el centro del proceso, está la aplicación de métodos específicos de minería de datos para el descubrimiento y la aplicación de patrones. [3]

¿Por qué es necesario el KDD?

Sea ciencia, mercadeo, finanzas, cuidado de la salud, o cualquier otro campo, el enfoque clásico de análisis de datos se basa fundamentalmente en uno o más analistas, convirtiéndose en interfaz entre los datos y el producto. Para estas y muchas más aplicaciones, esta forma manual de sondear los datos es lenta, cara y altamente subjetiva. El aumento dramático de los volúmenes de los datos hace que este tipo de análisis de datos manual sea impráctico en muchos aspectos. [3]

En vista de que la informática nos ha permitido recopilar más datos de los que podemos procesar, lo más natural es recurrir a técnicas computacionales para ayudarnos a desentrañar patrones de los volúmenes de datos masivos. Por lo tanto, KDD intenta dirigir un problema que en la era de la información digital es una realidad para todos: sobrecarga de datos. [3]

KDD en el mundo real

Tiene aplicaciones científicas (astronomía, genómica proteómica,ecología, cambio climático), marketing (sistemas de comercialización de bases de datos), detección de fraudes, manufactura (aerolíneas), y otras áreas. [3]

Diferencia entre KDD y Minería de datos (en inglés *Data Mining*)

El **KDD** se refiere al proceso global de descubrir conocimientos útiles a partir de datos, y la **minería de datos** se refiere a un paso particular en este proceso. La minería de datos es la aplicación de algoritmos específicos para la extracción de patrones de los datos. [3]

Los pasos adicionales en el KDD como: preparación de datos, limpieza de los mismos, la incorporación de conocimientos previos apropiados, la interpretación adecuada de los resultados, son esenciales para garantizar el conocimiento que se deriva de los datos. La aplicación ciega de minería

de datos puede ser una actividad riesgosa ya que fácilmente conduce al descubrimiento de patrones no válidos. [3]

El componente de minería de datos de KDD depende en gran medida de las técnicas conocidas, desde aprendizaje automático, reconocimiento de patrones, y estadística para encontrar patrones a partir de datos. Además, el KDD se centra en el proceso general del descubrimiento de conocimiento a partir de datos, cómo se almacenan y cómo se accede a ellos, cómo se puede escalar algoritmos para que corran eficientemente con datos masivos, cómo los resultados pueden ser analizados y visualizados, y cómo puede ser útil la interacción hombre máquina. [3]

CAPÍTULO II - **Descripción del proyecto**

II.1. Estado actual del problema

A partir de 2004, se publican lastendencias de búsqueda en Google, capaz de explorar las tendencias en historias en tiempo real por categoría y ubicación. Las tendencias en historias utilizan la tecnología del gráfico de conocimiento en Búsqueda de Google, Google noticias y Youtube para detectar cuando un tema está en tendencias en estas tres plataformas. [10]

En 2006, un grupo de investigadores epidemiólogos y elaboradores de software del Hospital infantil de Boston fundaron “Health Map” el líder global en utilizar fuentes informales en línea para el monitoreo de brotes de enfermedades y la vigilancia en tiempo real de las emergentes amenazas para la salud pública. [11]

Posteriormente, en 2008 surge “*Google Flu Trends*”, un servicio Web operado por Google, para proporcionar estimaciones de la actividad de influenza en más de 25 países. El modelo fue lanzado en 2008 y actualizado en 2009, 2013 y 2014 para ayudar a predecir los brotes de influenza. De igual forma surgió una herramienta que aplicaba los mismos modelos para hacer predicciones de dengue: “*Google Dengue Trends*”; sin actualización reciente de las tendencias. Han surgido nuevos proyectos, como el “*Flu Prediction Project*” del Instituto de Ciencia Cognitiva de Osnabrück e IBM Watson, que combinan datos de redes sociales con datos del “*Center for Disease Control and Prevention(CDC)*” y modelos estructurales que infieren la propagación espacial y temporal de la enfermedad. [12] [13]

De esta forma se han venido aplicando estas herramientas, como tendencias y redes sociales para la predicción de epidemias y vigilancia epidemiológica a tiempo real. Numerosos estudios acerca del tema se han realizado alrededor del mundo, por ejemplo, en 2010 Rumi Chunara y colaboradores establecieron patrones epidemiológicos al principio de la epidemia de cólera en Haití, y evidenciaron

que los patrones se correspondían bastante bien con la información oficial aportada por el CDC. Así mismo, Nicholas Christakis y James Fowler en 2010 publicaron un estudio realizado en 2009 en la universidad de Harvard, en el cual evidenciaron que las redes sociales sirven como un sensor temprano de la detección de enfermedades contagiosas. También en 2009, Cynthia Chew y Gunther Eysenbach llevaron a cabo un estudio en Canadá, usando información de Twitter para analizar la epidemia de H1N1, el cual ilustra el potencial del uso de redes sociales en "Infodemiología" para salud pública. [4] [9] [14]

Una vez más en 2012, Yusheng Xie y colaboradores, en Northwestern University, hicieron uso de estas herramientas para su publicación "Detección y seguimiento de brotes de enfermedades por minería de datos de redes sociales", y afirman que la detección de brotes epidémicos es una tremenda oportunidad para que la comunidad de investigadores cree impactos en el mundo real, y que la detección temprana de los brotes es crucial para mejorar las políticas de los profesionales, pacientes y funcionarios de la salud. Además, solicitan a otros investigadores la construcción de una plataforma con la arquitectura para el análisis de datos con algoritmos escalables, que constantemente podría proporcionar inteligencia colectiva y crear consciencia de los brotes epidémicos a tiempo real. [15]

En 2013, Rumi Chunara publica el estudio "Evaluación del entorno social en línea para la prevalencia de la obesidad", y con información de Facebook logro deducir que las personas con intereses en actividades sedentarias están asociadas a la población con mayor prevalencia de obesidad, y los patrones obtenidos coincidieron con la información del CDC. [16]

Estos son algunos ejemplos de aplicaciones que se han venido dando a estas herramientas en la vigilancia epidemiológica, en diferentes lugares y enfermedades de distinta naturaleza, y así como estos se pueden mencionar innumerables estudios realizados acerca del tema.

II.2. Planteamiento del problema

En Venezuela, en 1938 se comenzaron a publicar boletines epidemiológicos semanales, que incluyen 73 enfermedades endémicas y epidémicas de notificación obligatoria, que permiten conocer el comportamiento y llevar un control sanitario de la situación de salud del país. Los últimos años, la publicación de dichos boletines ha sido tardía y no abarca suficiente información, y los retrasos de esta publicación tiene grandes consecuencias para la salud pública, ya que aún en ausencia de información, las enfermedades siguen desarrollándose y afectando a la población. Dada esta situación, en este estudio se plantea una forma alternativa de hacer predicciones epidemiológicas que permitan conocer el estado actual de salud de Venezuela, y una fuente de información para el futuro,

haciendo el uso de las redes sociales como herramienta para la vigilancia epidemiológica de la salud pública. [5] [6]

II.3. Hipótesis

Los medios digitales y las redes sociales pueden servir de herramienta para la vigilancia epidemiológica en salud pública.

II.4. Objetivos de la investigación

Objetivo General

Desarrollar un modelo matemático que permita la predicción de brotes infecciosos en Venezuela

Objetivos específicos

-Seleccionar y descargar el conjunto de datos epidemiológicos y tendencias de interés por ciertas enfermedades entre 2004 y 2016

-Analizar de forma gráfica las propiedades de los datos para conocer de qué forma se comportan.

-Elaborar una base de datos que permita su posterior uso y modificación.

-Interpretar, evaluar y comparar los datos obtenidos de distintas fuentes.

-Generar información útil para la vigilancia epidemiológica en salud pública.

II.5. Justificación de la investigación

La vigilancia epidemiológica es una herramienta que favorece a la salud pública, y permite controlar las epidemias o disminuir su incidencia en la población; dado que el obtener información oficial de instituciones de salud públicas o gubernamentales puede ser un proceso lento y complejo, lo contrario ocurre con los datos generados por los usuarios de las redes sociales, son candidatos excelentes para ser analizados y construir patrones que permitan predecir comportamientos sociales o de una enfermedad a tiempo real, lo que permite dar un paso adelante en la prevención y la logística requerida para afrontar la problemática tratada.

Por otro lado, la sociedad es cada vez más tecnológica, y en el campo de la biología se produce una abrumadora cantidad de información almacenada en las bases de datos, datos con los que se

pueden realizar innumerables investigaciones, a partir de las cuales se han generado numerosas publicaciones recientemente.

Se dice que estamos viviendo en la era de la información, sin embargo, actualmente estamos viviendo en la era de los datos. Terabytes de datos se vierten en la "*World Wide Web*" (WWW) y en varios dispositivos de almacenamiento, todos los días, provenientes de negocios, sociedad, ciencia e ingeniería, medicina y muchos de los aspectos de la vida diaria. Este crecimiento explosivo del volumen de datos es un resultado de la computarización de nuestra sociedad y del desarrollo rápido de poderosas herramientas de almacenamiento de datos, todo esto verifica que estamos en la era de los datos. Son necesarias herramientas poderosas y versátiles para automatizar la transformación de tremendas cantidades de datos a conocimiento organizado. Esta necesidad ha llevado al nacimiento de la minería de datos, es un campo joven, dinámico y prometedor, que puede ser visto como un resultado natural de la evolución de la tecnología de la información. [17]

La abundancia de datos, en conjunto con la necesidad de herramientas de análisis, describe una situación rica en datos, pero pobre en información. El rápido crecimiento de ingentes cantidades de datos, colectados y almacenados y de grandes repositorios de estos ha excedido la habilidad humana para su comprensión y análisis, sin el uso de herramientas especializadas. Frecuentemente las decisiones importantes son tomadas por intuición, sin basarse en la información almacenada en dichos repositorios, simplemente porque las personas que toman estas decisiones no tienen las herramientas para extraer este valioso conocimiento enterrado en las vastas cantidades de datos. Se han hecho esfuerzos para desarrollar sistemas expertos basados en tecnología y conocimientos, que confían a usuarios expertos la introducción manual del conocimiento en bases de datos. Desafortunadamente estos sistemas están acompañados de errores y consumen altas cantidades de dinero y tiempo. Esta distancia creciente entre los datos y la información abre las puertas al desarrollo de herramientas de minería de datos que pueden convertir las "tumbas de datos" en "pepitas de oro" del conocimiento. [17]

II.6. Limitaciones de la investigación

- Difícil acceso a datos oficiales del sector de la salud.
- Las redes sociales tienen mayor incidencia en regiones con mayor acceso a internet.
- Los datos obtenidos de los medios informales pueden ser más representativos de ciertos grupos etarios.
- Los reportes de medios informales pueden incluir falsos positivos

CAPÍTULO III - **Diseño metodológico**

III.1.1. Metodología

Tipo de estudio o investigación

Es un estudio descriptivo, ya que se limita a una descripción de la frecuencia de una enfermedad en una población (se realizan mediciones sin ningún tipo de intervención)

Enfoque de la investigación

Se plantea un estudio cuantitativo, para explicar una realidad social desde un punto de vista externo y objetivo.

Área de estudio:

Epidemiología digital descriptiva

Población

El estudio se plantea para la población de Venezuela, comprendida entre los años 2004 y 2017, ya que las tendencias de Google comienzan a estar disponibles para 2004. Se ve representada principalmente la población con acceso a internet y su entorno. Según el reporte de tendencias digitales "Penetración y usos de internet en Venezuela, Reporte 2016". [18] La distribución de la población de Venezuela con acceso a internet se describe en la tabla 1, y en la tabla 2 se muestran los usos más frecuentes que se da al internet por la población.

Tabla 1. Distribución de la población de Venezuela con acceso a internet en 2016

Estrato social		Según edad		Según sexo	
Estrato	**(%)**	**Edad (años)**	**(%)**	**Sexo**	**(%)**
A/B	2	7- 12	16	Femenino	50
C	19	13 - 17	18	Masculino	50
D	41	18-24	23		
C	38	25-34	23		
		35-49	19		
		50-55	2		

Tabla 2. Principales Usos de Internet en Venezuela 2016	
Actividad	**Porcentaje (%)**
Enviar y recibir correos electrónicos	88
Realizar operaciones bancarias	83
Leer noticias	82
Mantener Redes sociales (Facebook, LinkedIn	78
Buscar información bancaria	67
Buscar información para el trabajo	67
Buscar trabajo	67
Ver videos	67
Chatear	61
Realizar trámites en un sitio web del gobierno	60
Comprar productos y/o servicios	59
Publicar fotos	59
Buscar información sobre productos para decidir	57
Buscar información para estudios	57
Actualizar mi estado en sitios como Twitter	57

Enfermedades Incluidas en el estudio

El criterio usado para elegir las enfermedades incluidas fue seleccionar de las enfermedades con mayor número de reportes semanales según los boletines epidemiológicos de Venezuela, aquellas con una sintomatología que pudiera ser de algún modo percibida e identificada. En las tablas mostradas a continuación se encuentra una breve descripción de cada una de ellas y la relevancia de su estudio, información obtenida de la Organización Mundial de la Salud.

Tabla 3. Enfermedades incluidas en el estudio y descripción breve

Enfermedad	Descripción
Dengue	Enfermedad viral con síntomas gripales y en ocasiones evoluciona hasta convertirse en un cuadro potencialmente mortal llamado dengue grave
Diarrea	Deposición, tres o más veces al día (o con una frecuencia mayor que la normal para la persona) de sueltas o líquidas
Difteria	Enfermedad bacteriana aguda que afecta sobre todo la mucosa de las vías respiratorias superiores (nariz, amígdalas, faringe, laringe), la piel o, en raras ocasiones, mucosas de otras zonas, como las conjuntivas, la vagina o el oído.
Fiebre	Es un síndrome (conjunto de síntomas y signos) cuyo signo principal es la hipertermia, aunque no es imprescindible
Hepatitis A / B / C	La hepatitis es la inflamación del hígado. La afección puede remitir espontáneamente o evolucionar hacia una fibrosis (cicatrización), una cirrosis o un cáncer de hígado. Los virus de la hepatitis son la causa más frecuente de las hepatitis, que también pueden deberse a otras infecciones, sustancias tóxicas (por ejemplo, el alcohol o determinadas drogas) o enfermedades autoinmunitarias.
Malaria	El paludismo es una enfermedad febril aguda. En un individuo no inmune, los síntomas suelen aparecer entre 10 y 15 días tras la picadura del mosquito infectivo. Puede resultar difícil reconocer el origen palúdico de los primeros síntomas (fiebre, dolor de cabeza y escalofríos), que pueden ser leves. Si no se trata en las primeras 24 horas, el paludismo por P.*falciparum* puede agravarse, llevando a menudo a la muerte.
Rubéola	La rubéola es una infección vírica aguda y contagiosa. Si bien por lo general la enfermedad es leve en los niños, tiene consecuencias graves en las embarazadas, porque puede causar muerte fetal o defectos congénitos en la forma del síndrome de rubéola congénita
Sarampión	El sarampión es una enfermedad muy contagiosa causada por un virus de la familia de los paramixovirus. El virus infecta el tracto respiratorio y se extiende al resto del organismo. Se trata de una enfermedad humana que no afecta a los animales.
VIH / SIDA	El virus de la inmunodeficiencia humana (VIH) infecta a las células del sistema inmunitario, alterando o anulando su función. La infección produce un deterioro progresivo del sistema inmunitario, con la consiguiente "inmunodeficiencia". Se considera que el sistema inmunitario es deficiente cuando deja de poder cumplir su función de lucha contra las infecciones y enfermedades. El síndrome de inmunodeficiencia adquirida (SIDA) es un término que se aplica a los estadios más avanzados de la infección por VIH y se define por la presencia de alguna de las más de 20 infecciones oportunistas o de cánceres relacionados con el VIH.

Tabla 4. Relevancia del estudio de las enfermedades incluidas en el estudio

Enfermedad	Relevancia
Dengue	Alrededor de la mitad de la población del mundo corre el riesgo de contraer esta enfermedad
Diarrea	Las enfermedades diarreicas son la segunda mayor causa de muerte de niños menores de cinco años
Difteria	Durante el año 2015, cinco países notificaron casos de difteria. En Venezuela, se notificaron 447 casos sospechosos de difteria (324 en 2016 y 123 en 2017)
Fiebre	Es un indicador de una amplia variedad de enfermedades
Hepatitis A / B / C	La hepatitis vírica causó 1,34 millones de muertes en 2015
Malaria	En 2015 hubo unos 212 millones de casos mundiales de paludismo y casi la mitad de la población mundial corría el riesgo de padecer el paludismo.
Rubéola	Se calcula cada año nacen en el mundo aproximadamente 100.000 niños con síndrome de rubéola congénita
Sarampión	En 2015 hubo 134200 muertes por sarampión en todo el mundo, es decir, cerca de 367 por día o 15 por hora
VIH / SIDA	El VIH, que continúa siendo uno de los mayores problemas para la salud pública mundial, se ha cobrado ya más de 35 millones de vidas. A finales de 2016 había aproximadamente 36,7 millones de personas infectadas por el VIH en el mundo, y en ese año se produjeron 1,8 millones de nuevas infecciones y un millón de personas fallecieron en el mundo por causas relacionadas con este virus.

III.2. Recolección de datos

Extracción de datos control

Inicialmente se llevó a cabo de forma manual la extracción de datos de los boletines epidemiológicos de Venezuela, publicados por el Ministerio del Poder Popular para la Salud, tomando en cuenta los reportes comprendidos entre 2004-2016. Con estos se realizó en Excel una base de datos control, organizados en tablas para cada enfermedad, y su vez para cada año. Generando un aproximado de 10 tablas para cada enfermedad, las cuales incluyen el número de casos confirmados o sospechosos de cada enfermedad para cada una de las 52/53 semanas correspondientes a cada año. En la tabla 4 se observa parte de una de las tablas realizadas para Malaria en 2016.

*En la tabla se ven representadas solo las primeras 22 semanas del año, pero están constituidos por 52 o 53 semanas.

Al terminar la extracción de estos datos control se obtuvo la información mostrada en la tabla 6.

Tabla 5. Datos control obtenidos en los boletines epidemiológicos de Venezuela para Malaria en 2016

Semana epidemiologica	Casos Malaria
1	4578
2	3606
3	3585
4	4034
5	3786
6	3453
7	3080
8	4270
9	4146
10	3222
11	4223
12	3069
13	4631
14	2950
15	3432
16	3023
17	3971
18	4188
19	4227
20	4410
21	2540
22	4958

Tabla 6. Tablas en la base de datos control y los años que representan, para cada enfermedad tomada en cuenta para la investigación, y el número de semanas incluidas

Enfermedad	Tablas en la base de datos	Semanas incluidas
Dengue	12 (2004-2016)	678
Diarrea	10 (2006-2016)	574
Difteria	10 (2006-2016)	574
Fiebre	10 (2006-2016)	574
Hepatitis A	10 (2006-2016)	574
Hepatitis B	10 (2006-2016)	574
Hepatitis C	10 (2006-2016)	574
Hepatitis no específica	10 (2006-2016)	574
Infecciones agudas respiratorias	10 (2006-2016)	574
Leishmaniasis	10 (2006-2016)	574
Malaria	12 (2004-2016)	678
Sarampión	10 (2006-2016)	574
VIH	10 (2006-2016)	574
Rubéola	10 (2006-2016)	574
TOTAL	144	8244

Uso del buscador de tendencias de Google

La página principal de tendencias de búsqueda en Google es capaz de explorar las tendencias en historias en tiempo real por categoría y ubicación.

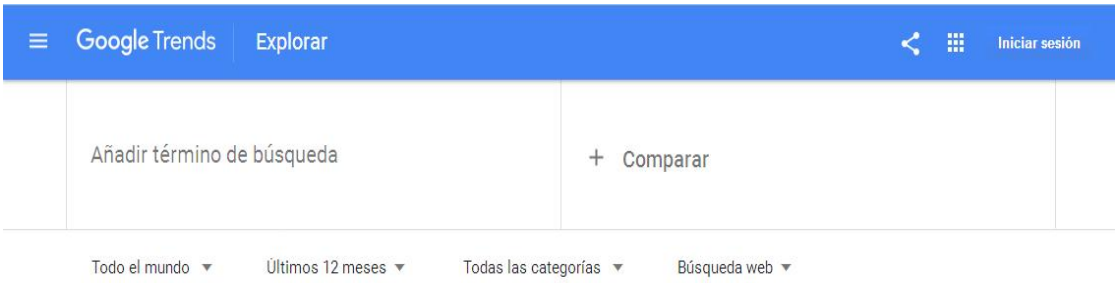

Figura 1. Visión inicial del buscador de tendencias de Google

A partir de esta página se obtuvieron las tendencias de cada una de las enfermedades en Venezuela, para cada año tomado en cuenta, información obtenida de esta forma para cuatro de las cinco plataformas de búsqueda: Búsquedas Web en Google, búsqueda imágenes en Google, noticias de Google y búsquedas en Youtube.

Esta herramienta además ofrece una opción arrojada por el buscador al momento de escribir la palabra cuya tendencia se desea obtener, bien sea el término o tema relacionado con la búsqueda. En este estudio ambas fueron tomadas en cuenta, para compararlas posteriormente.

Figura 2. Opciones de tendencias sugeridas por el buscador de tendencias de Google al escribir la palabra que se desea buscar

Para realizar la búsqueda, se seleccionó el país “Venezuela” y la selección “todas las categorías” se mantuvo constante. Se usaron las cuatro formas de búsqueda mencionadas; seleccionadas para cada año y cada una de las enfermedades, tanto para el término como para la búsqueda.

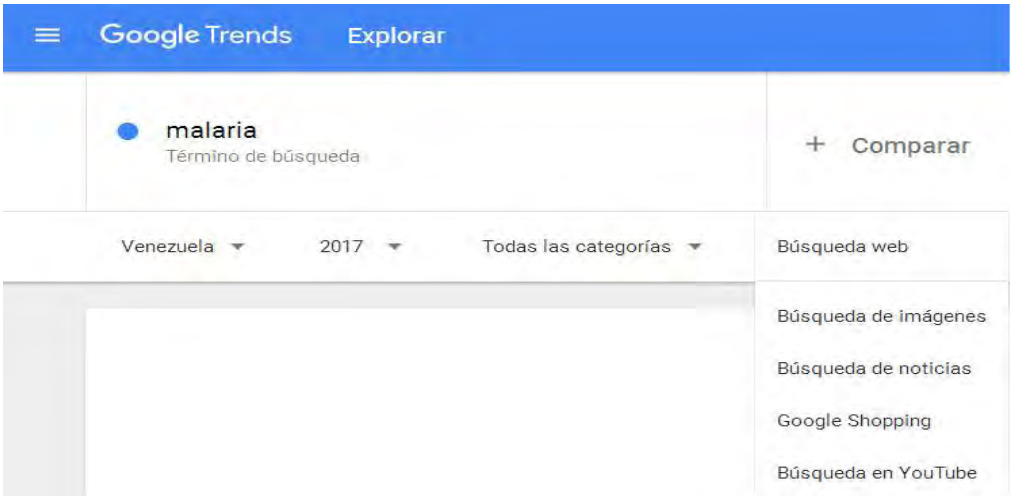

Figura 3. Configuración de búsqueda usada para buscar las tendencias obtenidas, y la ilustración de las plataformas disponibles para obtener tendencias en la página de tendencias de Google.

Las tendencias arrojadas por el buscador están representadas de forma gráfica, y los datos fueron descargados en un archivo “CSV” (Abreviatura del inglés de “Valores separados por comas”) e incluidos en las tablas mencionadas (tablas que incluyen el número de casos comprobados o sospechosos para cada enfermedad por año). Además, se obtuvo y descargo el interés por subregión para cada una de las tendencias.

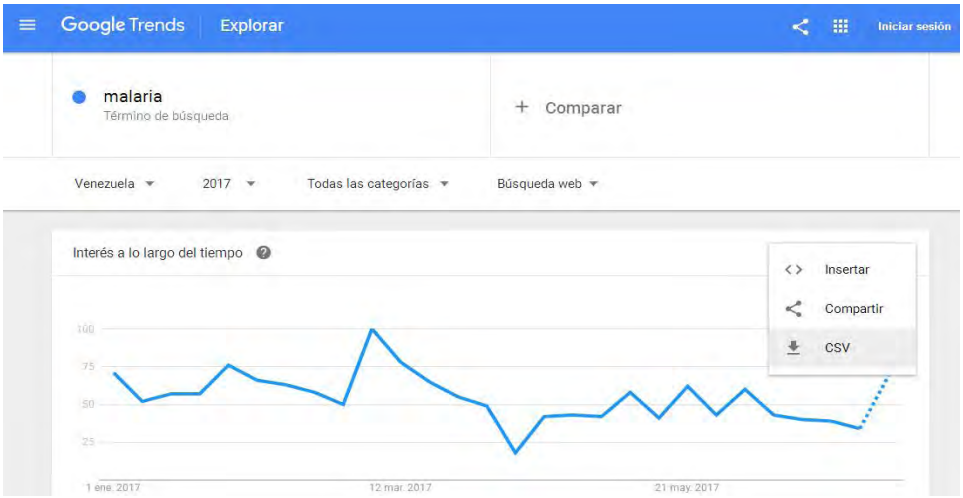

Figura 4. Ilustración de la selección de la opción “CSV” para descargar los datos graficados por la página de tendencia de Google

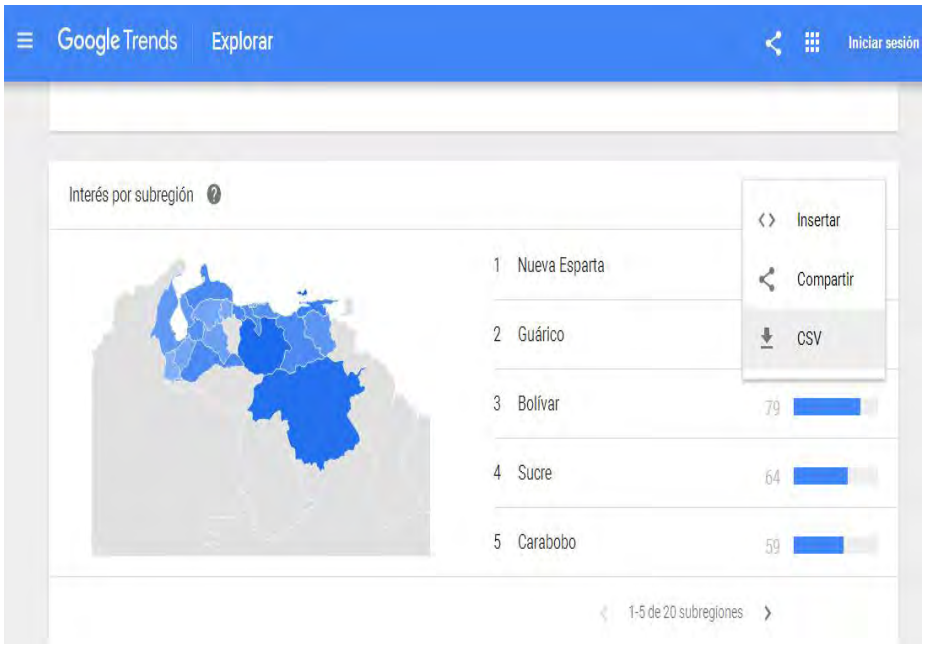

Figura 5. Ilustración del resultado obtenido en la página de tendencias de Google. Se muestra el interés de los usuarios de Google, por subregión, en tendencias de búsqueda en web de malaria en el año 2017

Una vez descargados los datos de la página de tendencias de Google, se realizó una nueva base de datos, de la misma forma de la base de datos control, una tabla por año para cada enfermedad, incluyendo: Datos control y datos de las tendencias de: búsqueda web en Google, búsqueda de imágenes en Google, noticias de Google y búsquedas en Youtube.

Tabla 7. Casos de malaria según los datos control, y el número de búsquedas en Web según la página de tendencias de Google, para malaria en el año 2012. Se incluyen los datos para la búsqueda del término, el tema y la palabra Plasmodium, como palabra clave asociada con el tema.

MALARIA 2012			Web	
Semana epidemiologica	Casos Malaria	Malaria (Término)	Malaria (Tema)	Plasmodium
1	829	3	12	2
2	1084	10	26	5
3	1067	8	36	1
4	1132	7	39	4
5	1065	10	32	1
6	1016	12	30	2
7	1037	8	23	4
8	872	7	21	0
9	959	7	31	0
10	824	8	25	4
11	915	15	31	5
12	752	11	27	0
13	650	7	20	1

Nota: La tabla muestra 7 solo 13 semanas representativas de 52.

Tabla 8. Casos de malaria según los datos control, y el número de búsquedas de imágenes según la página de tendencias de Google, para malaria en el año 2012. Se incluyen los datos para la búsqueda del término, el tema y la palabra Plasmodium, como palabra clave asociada con el tema.

MALARIA 2012			Imagenes	
Semana epidemiologica	Casos Malaria	Malaria (Término)	Malaria (Tema)	Plasmodium
1	829	0	0	0
2	1084	0	9	9
3	1067	8	23	0
4	1132	17	23	8
5	1065	0	8	0
6	1016	8	17	0
7	1037	0	17	0
8	872	15	17	10
9	959	0	9	0
10	824	9	17	9
11	915	14	34	0
12	752	11	30	13
13	650	9	9	17

De igual forma se procedió con todas las plataformas, para todas las enfermedades; usando en cada caso palabras clave asociadas a la enfermedad para comparar comportamientos en distintas formas de búsqueda, para obtener lo que se resume en la tabla 8.

III.3. Representación gráfica

Posteriormente, se realizó un gráfico para cada una de estas tablas, para comparar temporalmente los datos control extraídos de los boletines epidemiológicos de Venezuela y los datos proporcionados por la página de tendencias de Google, y así evaluar si hubo coincidencia en el comportamiento de dichos datos comparados. Se obtuvo un total de 540 gráficos.

Para obtener información de la epidemia, como su origen, duración, patrón de propagación, entre otras características, se planteó tratar dichos gráficos como epi-curvas. En su estudio sobre curvas epidémicas, Michell T., [31] define una epi-curva como una representación gráfica del número de casos epidémicos de acuerdo con la fecha de la aparición de la enfermedad, y plantea que permite obtener algunas características de la epidemia, descritas a continuación.

El patrón de propagación de la epidemia puede obtenerse de la forma global de la curva, la cual puede revelar el tipo de epidemia (origen común, origen puntual o propagado). Una epidemia de origen común es aquella en la cual la gente está expuesta intermitente o continuamente a una fuente dañina común. El período de exposición puede ser corto o largo. Una exposición intermitente en una

epidemia de origen común frecuentemente resulta en una epi-curva con picos irregulares que reflejan el tiempo y extensión de la exposición. La exposición continua hará, frecuentemente, que los casos aumenten gradualmente (y a lo mejor en "meseta" más que en pico). Una curva epidémica con una pendiente aguda hacia arriba y una pendiente gradual hacia abajo, típicamente describe una epidemia de origen puntual. Una epidemia de origen puntual es una epidemia de origen común, en la cual el periodo de exposición es relativamente corto y todos los casos ocurren dentro de un periodo de incubación. Una epidemia propagada es aquella que pasa de persona a persona, por lo cual este tipo de epidemias pueden durar más que las de origen común y pueden llevar a múltiples oleadas de infección si ocurren casos secundarios y terciarios. La clásica curva epidémica propagada tiene una serie de picos progresivamente más altos, siendo cada uno un periodo de incubación aparte, pero en la realidad la curva epidémica puede verse algo diferente. Se puede obtener información adicional acerca de la magnitud de la epidemia entre subpoblaciones estratificando la curva epidémica, separando la muestra en varias sub-muestras de acuerdo con criterios específicos, como género, edad, síntomas clínicos o ubicación geográfica.

Observando la curva se puede tener una idea de cuándo comenzó la epidemia y su duración, a lo que se llama tendencia en el tiempo de la epidemia. También se puede obtener información acerca de los casos aislados, los cuales pueden proveer información importante. Por ejemplo, un caso temprano puede no ser parte de la epidemia; más bien, puede representar el nivel basal de la enfermedad. Sin embargo, éste también puede representar la fuente de la epidemia, como un manipulador de comida enfermo, o puede ser un caso expuesto antes que los otros. Un caso tardío puede no ser parte de la epidemia; pero alternativamente, un caso tardío puede representar a un individuo que tuvo un periodo de incubación largo, que fue expuesto más tarde que los demás, o que fue un caso secundario (adquirió la enfermedad de un caso primario).

Otro factor importante es el periodo de exposición/periodo de incubación de la epidemia; si el tiempo de exposición es conocido, las curvas epidémicas pueden ser usadas para estimar el periodo de incubación de la enfermedad, y esto puede facilitar la identificación del agente causal. En epidemias de origen común, que involucran enfermedades con periodos de incubación conocidos, las curvas epidémicas pueden ayudar a determinar el periodo probable de exposición. Esto puede hacerse ubicando el periodo de incubación promedio para el organismo y desde el caso pico contar hacia atrás el tiempo promedio del periodo de incubación. Ahora las exposiciones potenciales durante este marco de tiempo pueden ser investigadas con la esperanza de encontrar la fuente de la epidemia.

III.4. "Sentiment analysis"

Se puede definir el análisis sentimental ("*Sentiment analysis*") como el tratamiento computacional de las opiniones, sentimientos y fenómenos subjetivos en los textos. [32] Este análisis se llevó a cabo

para ampliar la información de las tendencias, con la finalidad de incluir el análisis de las redes sociales Facebook y Twitter, y medir el interés de la población por las enfermedades tomadas en cuenta. Para este análisis de hizo uso de R, un entorno y lenguaje de programación con un enfoque al análisis estadístico.

R es una implementación de software libre del lenguaje de programación S, pero con soporte de alcance estadístico. Se trata de uno de los lenguajes más utilizados en investigación por la comunidad estadística, siendo además muy popular en el campo de la minería de datos, la investigación biomédica, la bioinformática y las matemáticas financieras. A esto contribuye la posibilidad de cargar diferentes bibliotecas o paquetes con funcionalidades de cálculo y gráficas. [33]

R se distribuye bajo la licencia GNU,GPL (La Licencia Pública General de GNU o más conocida por su nombre en inglés GNU General Public License es la licencia de derecho de autor más ampliamente usada en el mundo del software libre y código abierto). Está disponible para los sistemas operativosWindows, Macintosh, Unix y GNU/Linux. R proporciona un amplio abanico de herramientas estadísticas (modelos lineales y no lineales, pruebas estadísticas, análisis de series temporales, algoritmos de clasificación y agrupamiento, etc.) y permite hacer de forma sencilla una variedad de representaciones graficas de buena calidad. R puede integrarse con distintas bases de datos, además de disponer de extensas bibliotecas que facilitan su utilización desde lenguajes de programación interpretados como Perl y Python. [34]

El paquete “twitteR”, [34] contiene una función que permite obtener los comentarios de Twitter que incluyen una palabra clave, para una ubicación geográfica específica y un N determinado, así como otra serie de funciones. Para hacer uso de este paquete fue necesario obtener información como el API *key* y *access token* de Twitter, lo cual se consigue creando una aplicación en la página de desarrolladores de Twitter.

Así mismo R cuenta el paquete “RFacebook”, [35] que contiene funciones aplicadas a Facebook, y aunque este paquete incluye entre sus funciones la descarga de publicaciones acerca de una palabra clave, esta función no puede ser usada por cambios en las políticas de privacidad de Facebook a partir de 2015. Por lo tanto, no fue posible incluir datos de Facebook en este estudio.

III.5. Construcción de la base de datos

Una vez concluida la fase de recopilación de los datos, estos fueron organizados y transformados para construir una base de datos en MySQL.

MySQL es un sistema de gestión de bases de datosrelacional desarrollado bajo licencia dual GPL/Licencia comercial por Oracle Corporation y está considerada como la base datos *código abierto* más popular del mundo. [36]

MySQL es patrocinado por una empresa privada, que posee los derechos de la mayor parte del código. Esto es lo que posibilita el esquema de doble licenciamiento anteriormente mencionado. MySQL es usado por muchos sitios web grandes y populares, como Wikipedia, Google, Facebook, Twitter, Flickr y YouTube. [36]

Las bases de datos relacionales, como MySQL, organizan los datos en tablas (como una hoja de cálculo) y pueden vincular valores de tablas entre sí. En términos generales, son mejores en el manejo de grandes conjuntos de datos y son más eficientes en el almacenamiento y acceso a los datos que los CSV la compresión y la indexación. Los datos permanecen en la base de datos MySQL hasta que se accede a ellos a través de una consulta, que es diferente de cómo R aborda *data.frames* y CSV. Al acceder a los datos almacenados en un archivo data.frame o CSV en R, todos los datos deben caber en la memoria. Sin embargo, esto se convierte en un problema si se usa un conjunto de datos grande o si se le usa una computadora con <= 4 GB de RAM. En estos casos, cada vez que se carga el conjunto de datos o se realiza una operación que requiere mucha memoria, la computadora se ralentizará. La ventaja es que las bases de datos SQL es que solo cargan en la memoria RAM los datos con los que se está trabajando, cuando estos han sido seleccionados, dejando suficiente memoria para el análisis real. [37]

Otra razón por la que es preferible usar las bases de datos SQL (abreviatura del inglés*Strutured Query Lanaguage*) es la seguridad. Al almacenar los datos en un archivo.Rdata, se pone la confianza en el sistema operativo, o una alternativa de almacenamiento basada en la nube (por ejemplo, Dropbox o CrashPlan) para hacer una copia de seguridad de los datos. Otra opción sería almacenar los datos en una base de datos de Excel, y si este fallara, se podría correr el riesgo de perder años de colección de datos, riesgos que no se corren al usar una base de datos SQL, que además permite incluir usuarios con acceso limitado a la base de datos, lo que hace más fácil compartir datos con colaboradores. [37]

Para hacer uso de MySQL fue necesario implementar el servidor HTTP Apache,un servidor webHTTP de código abierto, para plataformas Unix (BSD, GNU/Linux, etc.), Microsoft Windows, Macintosh y otras, que implementa el protocolo HTTP/1.1 y la noción de sitio virtual. [38]

Para el manejo de la base de datos se incluyó phpMyAdmin, una herramienta escrita en PHP con la intención de manejar la administración de MySQL a través de páginas web, utilizando Internet que permite crear y eliminar Bases de Datos, tablas, y manipular los campos, ejecutar cualquier sentencia

SQL, administrar claves en campos, administrar privilegios, exportar datos en varios formatos y está disponible en 72 idiomas. Se encuentra disponible bajo la licencia GPL Versión 2. [39]

PHP es un lenguaje de programación de código abierto originalmente diseñado para el desarrollo web de contenido dinámico. Fue uno de los primeros lenguajes de programación que se podían incorporar directamente en el documento HTML en lugar de llamar a un archivo externo que procese los datos. El código es interpretado por un servidor web con un módulo de procesador de PHP que genera la página web resultante. PHP ha evolucionado por lo que ahora incluye también una interfaz de línea de comandos que puede ser usada en aplicaciones gráficas independientes. Puede ser usado en la mayoría de los servidores web al igual que en casi todos los sistemas operativos y plataformas sin ningún costo. Este lenguaje forma parte del software libre publicado bajo la licencia PHP, que es incompatible con la Licencia Pública General de GNU debido a las restricciones del uso del término PHP. [40]

Para emplear estas herramientas se usó un sistema de infraestructura de internet llamada WAMP, cuyas siglas corresponden a Windows como sistema operativo, Apache como servidor Web, MySQL como gestor de base de datos y PHP como lenguaje de programación, en Linux estas herramientas vienen por omisión en la instalación del sistema operativo. [41]

A continuación, se muestra el *script* usado en R para la descarga de datos de Twitter, su conversión y envío a la base de datos en MySQL. Como se observa en la figura 6, comienza con la instalación de los paquetes requeridos, seguido de la conexión con Twitter y MySQL, y a búsqueda en Twitter, para finalmente enviar los datos a la misma.

```
##Start_twitter_oauth

install.packages("RCurl")
install.packages("twitteR")
require (twitteR)
require (RCurl)
consumer_key <- 'XXX'
consumer_secret <- 'XXX'
access_token <- 'XXX'
access_secret <- 'XXX'
setup_twitter_oauth(consumer_key, consumer_secret, access_token, access_secret)

#Connecting to database
install.packages("RMySQL")
library(RMySQL)

##Connecting to MySQL: Once the RMySQL library is installed create a database connection object.
mydb = dbConnect(MySQL(), user='XXX', password='XXX', dbname='XXX', host='XXX')

##DB register backend

register_mysql_backend("tweets", "XXX", "XXX", "XXX")

##search_twees
palabra_clave_tweets<- searchTwitter('palabra_clave', geocode='10.48059,-66.90361,1000mi',n=500, lang="es")

#Set as data frame and view

palabra_clave_tweets_df = do.call("rbind", lapply(palabra_clave_tweets, as.data.frame))
View(palabra_clave_tweets_df)

##Write dataframe into db
dbWriteTable(mydb, name='palabra_clave_tweets', value=palabra_clave_tweets_df)
```

Figura 6. *Script* usado en R para la descarga de datos de Twitter y su envió a la base de datos MySQL

III.6. Análisis estadístico

Cuando se evalúa la relación entre dos variables es importante determinar cómo se comportan. Las relaciones lineales son muy comunes, pero las variables también pueden tener una relación no lineal, motona y es posible que no haya relación entre las variables. [42]

Como lo recomienda la bibliografía se comenzó por crear gráficos de dispersión de las variables para evaluar la relación e identificar el modelo de dispersión que podía representar los datos. Para su interpretación se tomaron en cuenta los siguientes criterios:

- Una relación lineal es una tendencia en los datos que se puede modelar mediante una línea recta. Se utiliza el coeficiente de correlación de Pearson para examinar la fuerza y la dirección de la relación lineal entre dos variables continuas.
- Si una relación entre dos variables no es lineal, la tasa de aumento o descenso puede cambiar a medida que una variable cambia, causando un "patrón de curva" en los datos. Esta tendencia en forma de curva se podría modelar mejor mediante una función no lineal, como una función cuadrática o cúbica, o se podría transformar para convertirla en lineal.
- En una relación monótona, las variables tienden a moverse en la misma dirección relativa, pero no necesariamente a un ritmo constante. En una relación lineal, las variables se mueven en la misma dirección a un ritmo constante. Se usa el coeficiente de correlación de Spearman para examinar la fuerza y la dirección de la relación monótona entre dos variables continuas u ordinales.

En vista de que los gráficos de dispersión de los datos no tienen una tendencia lineal, se escogió hacer el cálculo de correlación con el coeficiente de Spearman usado para relaciones monótonas, en las cuales se incluye la relación lineal. En una relación monótona, las variables tienden a moverse en la misma dirección relativa, pero no necesariamente a un ritmo constante. Para calcular la correlación de Spearman, es necesario jerarquizar los datos sin procesar y luego se calcula el coeficiente de correlación con los datos jerarquizados. [42]

Una vez calculado, el valor del coeficiente de correlación puede variar de −1 a +1. Mientras mayor sea el valor absoluto del coeficiente, más fuerte será la relación entre las variables. Para la

correlación de Spearman, un valor absoluto de 1 indica que los datos ordenados por rango son perfectamente lineales. Por ejemplo, una correlación de Spearman de −1 significa que el valor más alto de la Variable A está asociado con el valor más bajo de la Variable B, el segundo valor más alto de la Variable A está asociado con el segundo valor más bajo de la Variable B y así sucesivamente. Por otro lado, el signo del coeficiente indica la dirección de la relación. Si ambas variables tienden a aumentar o disminuir a la vez, el coeficiente es positivo y la línea que representa la correlación forma una pendiente hacia arriba. Si una variable tiende a incrementarse mientras la otra disminuye, el coeficiente es negativo y la línea que representa la correlación forma una pendiente hacia abajo. [43]

Para determinar si la correlación entre las variables era significativa, se comparó elvalor *p* con su nivel de significancia denotado como α o alfa, con un valor de 0.05, que indica que el riesgo de concluir que existe una correlación, cuando en realidad no es así, es 5%. El valor *p* indica si el coeficiente de correlación es significativamente diferente de 0. El criterio usado para considerar si había o no correlación fue:

- Valor p ≤ α: La correlación es estadísticamente significativa: Sí el valor *p* es menor que o igual al nivel de significancia, se puede concluir que la correlación es diferente de 0.
- Valor p > α: La correlación no es estadísticamente significativa: Sí el valor *p* es mayor que el nivel de significancia, no se puede concluir que la correlación es diferente de 0. [43]

Para llevar a cabo el análisis se hizo uso del Software estadístico Minitab versión 18

Los resultados más representativos fueron incluidos en la tabla 10.

III.7. Elaboración de un modelo epidemiológico

El propósito del modelado de epidemias es entender los procesos por los cuales ocurre la propagación de las mismas y así tener bases racionales para formular programas de prevención más efectivos y combatir los brotes. El proceso epidémico es esencialmente un modelo de crecimiento poblacional, donde la enfermedad está representada por individuos infectados, y los recursos restantes (limitantes) por los susceptibles a la infección. [44]

En teoría, para comprender completamente el proceso de propagación de una enfermedad, se deben considerar todos los factores biológicos relevantes, mientras que, en la práctica, un estudio tan complejo no es factible a escala humana. Por lo tanto, existen otras formas de abordar el problema con modelos simplificados de diseminación de la enfermedad. Estosmodelos no están enfocados en el fenómeno a nivel celular, sino más bien tratan de entender la propagación de la enfermedad a través de la red de comunidades e individuos a diferentes niveles. [44]

El modelado y las simulaciones automatizadas proporcionan una poderosa herramienta que permiten estudiar la evolución espacio-temporal de enfermedades. Se pueden usar para comprender los roles individuales y los factores que influyen en la propagación o para medir los efectos de diversas intervenciones como estrategias farmacéuticas o campañas de prevención dirigidas a las personas o al medio ambiente. [45]

Distintos autores proponen clasificar los modelos epidemiológicos según cuatro categorías de la siguiente manera: Modelos compartimentados, metapoblaciones, modelos de red y modelos basados en agentes.

Modelos compartimentados

Los modelos de compartimiento suponen que una población puede ser dividida en un conjunto de compartimentos (o estados) de acuerdo con el nivel del desarrollo de la enfermedad en los individuos. En tal modelo, la transmisión ocurre cuando los individuos están en contacto con personas infectadas. Estos modelos suponen que los individuos tienen una estructura relacional regular entre ellos y dentro de los compartimentos. De hecho, la mezcla dentro y entre los compartimentos se supone que es aleatoria o dirigida por reglas de transición que permiten especificar cómo cambian las personas a otro compartimento. Por lo tanto, la cantidad de nuevas infecciones es proporcional al producto del número de individuos infectados por el de individuos susceptibles. En otras palabras, las interacciones entre los individuos que conducen a la transmisión de la enfermedad son homogéneas, por lo tanto, los modelos de compartimentos son naturalmente deterministas para una población en un compartimiento dado. Posteriormente, las ecuaciones diferenciales se utilizan comúnmente para describir la evolución de la enfermedad en el sistema, al mezclarse los individuos en los diferentes compartimentos con ciertas probabilidades. Además, la estructura de las poblaciones puede estar integrada dentro de estos modelos agregando nuevos compartimentos correspondientes a varias características individuales de la población, como la edad, los comportamientos de riesgo, el estado social, etc. [46] Las siguientes sub secciones presentan los principales modelos compartimentales utilizados en epidemiología y la representación gráfica de cada uno de ellos en las figuras 7-10.

Modelo SI

Comenzamos con quizás el modelo más simple de contagio, llamado Modelo Susceptible-Infectado. En este tipo de modelo, la enfermedad dentro del hospedador solo se reduce a dos estados: susceptible (S) e infectado (I). Un individuo en el estado susceptible no ha contraído la enfermedad, pero podría, si entra en contacto con alguien que está infectado, mientras que un individuo en el estado infectado corresponde a alguien que ha contraído la enfermedad. Este modelo de dos estados asume que un individuo infectado retiene su estado para siempre. El enfoque aquí consiste en

modelar la propagación de la enfermedad por una cierta probabilidad de pasar al estado infectado. [46]

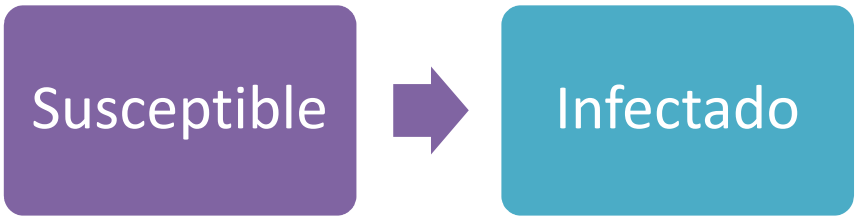

Figura 7. Representación gráfica de modelo SI

Modelo SIR

Basado en el mismo principio, el modelo SIR es una evolución del Modelo SI, que conserva los dos estados del modelo SI y agrega un nuevo estado: el estado recuperado. Este modelo se llama Modelo Susceptible-infectado-recuperado (SIR). Puede ser visto como un enfoque más realista en el caso de enfermedades donde las personas se recuperan de la infección después de un tiempo determinado porque su sistema inmune rechaza al patógeno responsable de la enfermedad. Este tipo de modelo supone que las personas conservan su inmunidad a la enfermedad después de tal recuperación para que no puedan contagiarse de nuevo. Este modelo también se puede usar para estudiar otra enfermedad donde las personas no se recuperan, pero mueren después de un intervalo de tiempo. De hecho, en términos de modelado, tanto la recuperación como la muerte pueden ser representadas por el estado R de este modelo. [46] A continuación se muestra su representación gráfica.

Figura 8. Representación gráfica del modelo SIR

Modelo SIS

El modelo SIS es otra evolución del modelo SI que permite la reinfección. En el modelo SIS, un individuo, una vez infectado, puede volver al estado S y puede infectarse una y otra vez. Para representar este comportamiento el modelo SI, debe permitir que un individuo infectado (es decir, en el estado I) regrese al estado S. El modelo SIS es interesante para enfermedades que pueden

infectar personas más de una vez. Esto puede ser debido a que el sistema inmune no confiere inmunidad a las víctimas después de la recuperación, o confiere una inmunidad limitada. Sin embargo, este tipo de modelo parece limitado si uno quiere estudiar el efecto de varias intervenciones en individuos, ya que no permite que un individuo se cure completamente. [46]

Figura 9. Representación gráfica del modelo SIS

Modelo SIRS

El modelo SIRS aprovecha los modelos SIR y SIS y los combina. De hecho, integra la recuperación y la posibilidad de reinfección. Este modelo toma en cuenta el hecho de que la inmunidad puede ser temporal. Cuando los individuos están en el estado R, ganan inmunidad, como en el Modelo SIR, pero esta inmunidad es temporal. Después de un cierto período de tiempo, pierden su inmunidad y se vuelven susceptibles. El modelo SIRS permite ir más allá de los límites establecidos en los modelos SIR y SIS. Proporciona un modelo con oscilaciones más realistas, ya que incluye la inmunidad temporal de los individuos recuperados, una característica de muchas enfermedades. [46]

Tomando en cuenta las enfermedades estudiadas, su clasificación de acuerdo a los modelos compartimentados se muestra en la tabla 9, además se indican los distintos tipos de transmisión, lo cual también es un factor importante al momento de estudiar la propagación de brotes infecciosos.

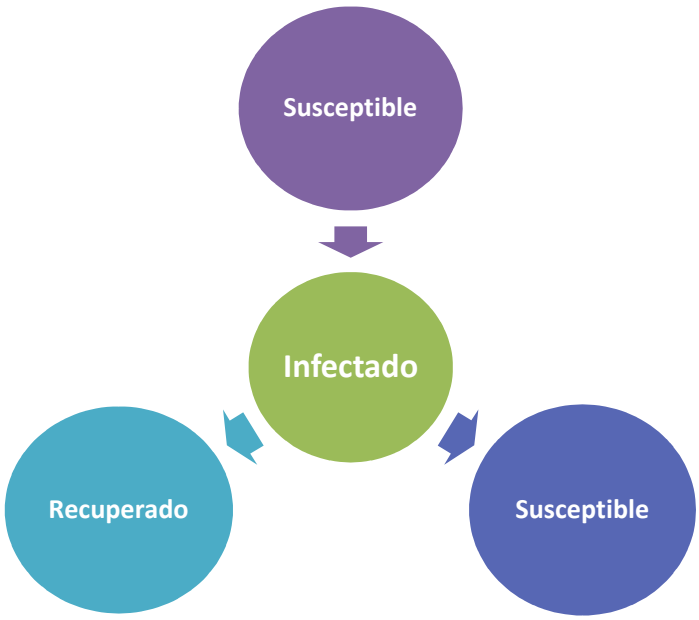

*Figura **10.** Representación gráfica del modelo SIRS*

Tabla 9. Asignación del modelo compartimentado para cada una de las enfermedades y su modo de transmisión.

Enfermedad	Transmisión	Modelo de compartimientos
Dengue	Transmitida por mosquitos infectados	SIRS
Diarrea	La infección se transmite por alimentos o agua de consumo contaminados, o bien de una persona a otra como resultado de una higiene deficiente	SIRS
Difteria	Por contacto con un paciente o un portador; con menor frecuencia, contacto con artículos contaminados por secreciones de lesiones de personas infectadas.	SIRS
Fiebre	Depende de la causa (viral, bacteriana, parasitaria, entre otras)	SIRS
Hepatitis A / B / C	La hepatitis A y la E son causadas generalmente por la ingestión de agua o alimentos contaminados. Las hepatitis B, C y D se producen por el contacto con humores corporales infectados. Son formas comunes de transmisión de estos últimos la transfusión de sangre o productos sanguíneos contaminados, los procedimientos médicos invasores en que se usa equipo contaminado y, en el caso de la hepatitis B, la transmisión de la madre a la criatura en el parto o de un miembro de la familia al niño, y también el contacto sexual.	SIR / SIR / SIS
Malaria	Causada por parásitos del género Plasmodium que se transmiten al ser humano por la picadura de mosquitos hembra infectados del género Anopheles	SIRS
Rubéola	El virus de la rubéola se transmite por gotículas en el aire, cuando las personas infectadas estornudan o tosen. Los humanos son el único huésped conocido.	SIR
Sarampión	El virus del sarampión se transmite a través del contacto directo y del aire.	SI

VIH / SIDA	El virus de la inmunodeficiencia humana (VIH) infecta a las células del sistema inmunitario, alterando o anulando su función. La infección produce un deterioro progresivo del sistema inmunitario, con la consiguiente "inmunodeficiencia". Se considera que el sistema inmunitario es deficiente cuando deja de poder cumplir su función de lucha contra las infecciones y enfermedades. El síndrome de inmunodeficiencia adquirida (SIDA) es un término que se aplica a los estadios más avanzados de la infección por VIH y se define por la presencia de alguna de las más de 20 infecciones oportunistas o de cánceres relacionados con el VIH. El VIH puede transmitirse por las relaciones sexuales vaginales, anales u orales con una persona infectada, la transfusión de sangre contaminada o el uso compartido de agujas, jeringuillas u otros instrumentos punzantes. Asimismo, puede transmitirse de la madre al hijo durante el embarazo, el parto y la lactancia.	SI

Modelos de metapoblación

Parte de la idea de que el papel de la estructura de la población no puede ser ignorada en el entendimiento de la propagación de enfermedades humanas. Mientras los modelos de compartimientos asumen una distribución aleatoria dentro de las poblaciones y se basan en atributos individuales, se ha demostradoque los modelos tienen que integrar diferentes tipos de interacciones sociales. Debido a que involucran la estructura de la población, pueden ser considerados como una relación entre modelos de compartimentos y redes. [46]

Los metamodelos suponen una distribución aleatoria entre subpoblaciones de la misma forma que los modelos de compartimentos e introducen nuevos tipos de interacciones basadas en la estructura espacial de las poblaciones. Los modelos epidémicos de metapoblaciones se basan en la estructura espacial del ambiente y el conocimiento detallado de las interacciones de las subpoblaciones. Individuos dentro de cada subpoblación se clasifican en estados como infectados, susceptibles, recuperados, y se aplican enfoques de modelos compartimentados como SIR, donde se considera que las personas en el mismo lugar pueden estar en contacto y cambiar su estado de acuerdo a la dinámica de la infección. La interacción entre subpoblaciones es el resultado del movimiento de individuos de una subpoblación a otra. Por lo tanto, es importante describir los patrones de desplazamiento de las personas con precisión. Aunque reducen la complejidad de las redes reales,

los metamodelos ayudan a obtener una idea del importante papel que juega la estructura de la población en la propagación de la enfermedad. [46]

Limitaciones del modelo: Aunque los metamodelos han demostrado ser útiles para comprender la posible progresión de brotes epidémicos, todos los autores subrayan la necesidad de tener datos más precisos sobre el comportamiento de la población, y en los boletines epidemiológicos de Venezuela la única información reflejada es el número de casos para cada enfermedad, para cada semana del año y en la mayoría de los casos un único valor referente al país completo. [46]

Modelo de red

En epidemiología, es fácil entender que las redes sociales han encontrado aplicaciones rápidamente, ya que la estructura y la naturaleza de la red de interacciones entre los individuos es un factor significativo del brote y la evolución de las enfermedades. De hecho, la transmisión de la enfermedad a menudo depende de la naturaleza de las interacciones que mantiene un individuocon su entorno. Por ejemplo, al considerar el contexto de las ITS, la probabilidad de que un individuo se contagie de una enfermedad depende de las características de su red. Por lo tanto, los modelos de red son particularmente útiles en epidemiología. [46]

La epidemiología considera los contactos entre los individuos desde dos enfoques:

-Contactos personales: Los contactos personales son el primer tipo de contactos que han sido explotado en epidemiología. Los contactos personales de un individuo se refieren a tipos de contactos con otras personas y son identificables por la observación, por ejemplo, parentesco, amistad, actividad de ocio, contactos íntimos, etc. Numerosos modelos de red han usado contactos personales para modelar la propagación de enfermedades, porque generalmente son rutas naturales e identificables, particularmente en el campo de enfermedades de transmisión sexual, parece ser más adecuado a este tipo de modelo. [46]

-Contactos Geográficos: En el caso de enfermedades como tuberculosis (TB) o SARS, que se transmiten no solo a través de contactos personales, sino que una enfermedad puede propagarse por (a) contacto directo entre individuos, (b) el entorno o (c) fómites. Diversos estudios muestran que la inclusión de contactos geográficos es una valiosa contribución. Ellos demuestran de hecho que la incorporación de contactos geográficos "proporciona una buena forma de encontrar posibles conexiones entre los pacientes y ver el papel que juegan esas ubicaciones geográficas en la enfermedad. [46]

Limitaciones del modelo: Finalmente, los modelos de red resultan ser un enfoque más realista que los modelos simples, como el compartimiento o modelos de metapoblación, ya que son más adecuados

para la complejidad de relaciones reales. Sin embargo, la minería de datos de estos modelos sigue siendo difícil por diversas razones. En primer lugar, los contactos en una red social son amplios, irregulares y dinámicos. Además, como el conjunto de datos reales no está disponible, los modelos de red a menudo se basan en conjuntos de datos generados, que posiblemente no representan el mundo real. Estudios experimentales sobre el impacto de intervenciones farmacéuticas, políticas o de conciencia requieren un gran número de ejecuciones, y para obtener modelos completos y realistas, los modelos tienen que considerar una gran cantidad de parámetros. De hecho, la diversidad entre los factores es crucial para entender la extensión espacial y temporal de una enfermedad. [46]

CAPÍTULO IV - **Resultados**

Una vez extraída la información de los boletines epidemiológicos, y finalizado el proceso de descarga de datos de las tendencias de Google (buscador, imágenes, noticias y YouTube) se obtuvo un total de 540 tablas distribuidas como se muestra en la tabla 10. Estas tablas incluyen información acerca de las 14 enfermedades señaladas, para 10 años (2006-2016), tanto de datos formales como informales.

Tabla 10. Tablas incluidas en la base de datos al incluir datos control y datos de las tendencias.

	Tablas	Tablas datos tendencias				Total de tablas
Enfermedad	Datos Control	Web	Imágenes	Noticias	Youtube	por enfermedad
Dengue	12	12	6	6	6	42
Diarrea	10	10	6	6	6	38
Difteria	10	10	6	6	6	38
Fiebre	10	10	6	6	6	38
Hepatitis A	10	10	6	6	6	38
Hepatitis B	10	10	6	6	6	38
Hepatitis C	10	10	6	6	6	38
Hepatitis no específica	10	10	6	6	6	38
Infecciones respiratorias agudas	10	10	6	6	6	38
Leishmaniasis	10	10	6	6	6	38
Malaria	12	12	6	6	6	42
Rubéola	10	10	6	6	6	38
Sarampión	10	10	6	6	6	38
VIH	10	10	6	6	6	38
Total de tablas por plataforma	144	144	84	84	84	540

Las tablas obtenidas se representaron de forma gráfica, con la intención de observar el comportamiento de los datos para cada enfermedad en cada año, representando el número de casos reportados por semana epidemiológica en el caso de los datos formales, y el número de búsquedas por semana epidemiológica en caso de los datos informales. Además, se buscaba determinar si el

comportamiento de los datos formales e informales era similar. En el apartado de anexos, se observa parte de los gráficos obtenidos, en los que se permite observar el comportamiento de los datos, y pareciera verse cierta relación entre los datos formales e informales, pero para poder verificar dicha información se prosiguió a realizar un análisis estadístico de los datos.

Análisis estadístico

Como se explicó en la metodología, el análisis estadístico comenzó por una prueba de calidad de los datos, realizando gráficos de dispersión de las variables para evaluar la relación e identificar el modelo de dispersión que podía representar los datos de cada enfermedad en cada año.

A continuación, en las figuras 11-16, se muestran algunos de los gráficos obtenidos al tratar de identificar el modelo de dispersión para los datos formales e informales, y en cada uno se observan varios modelos de dispersión, con los valores de prueba de bondad de ajuste, en el recuadro derecho superior, que indican cual sería el modelo de dispersión más adecuado para los datos.

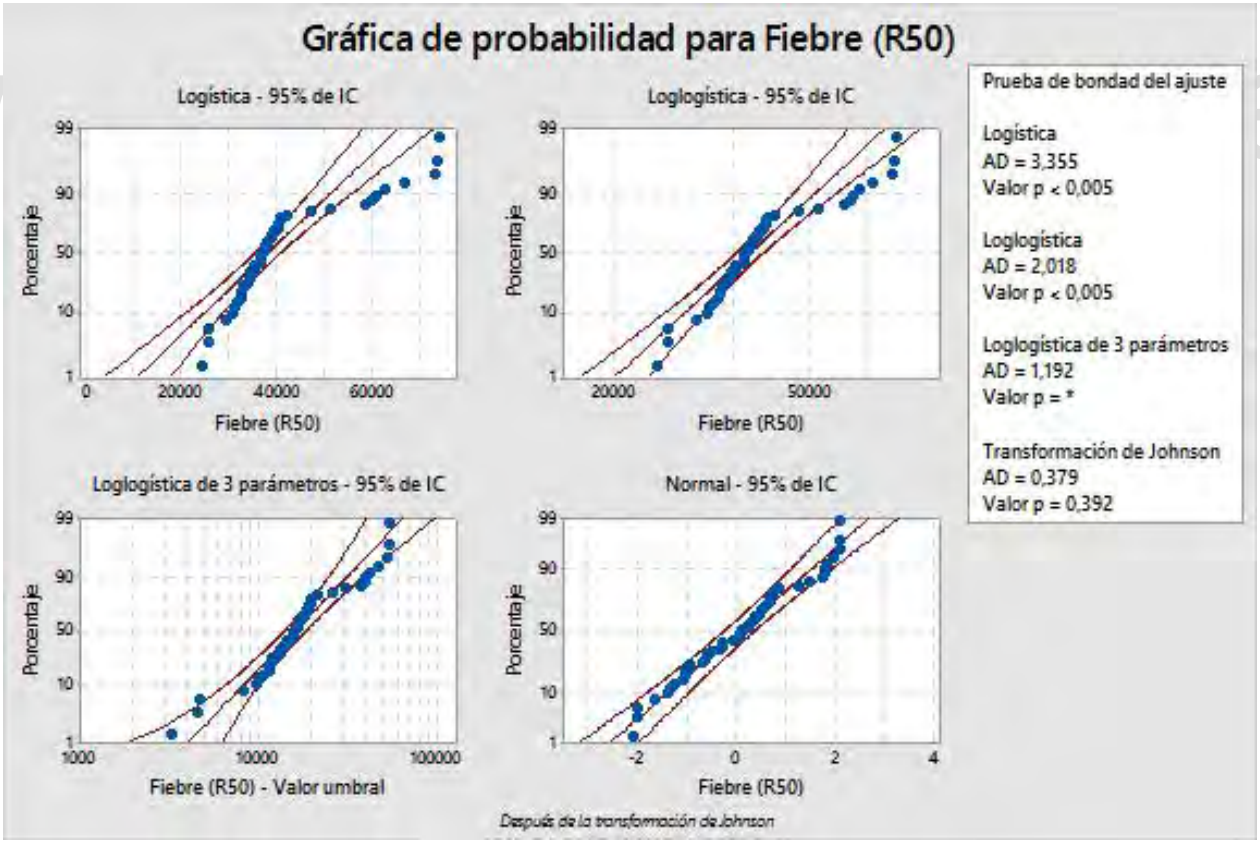

Figura 11. Identificación de gráfico de dispersión para datos oficiales de Fiebre 2016

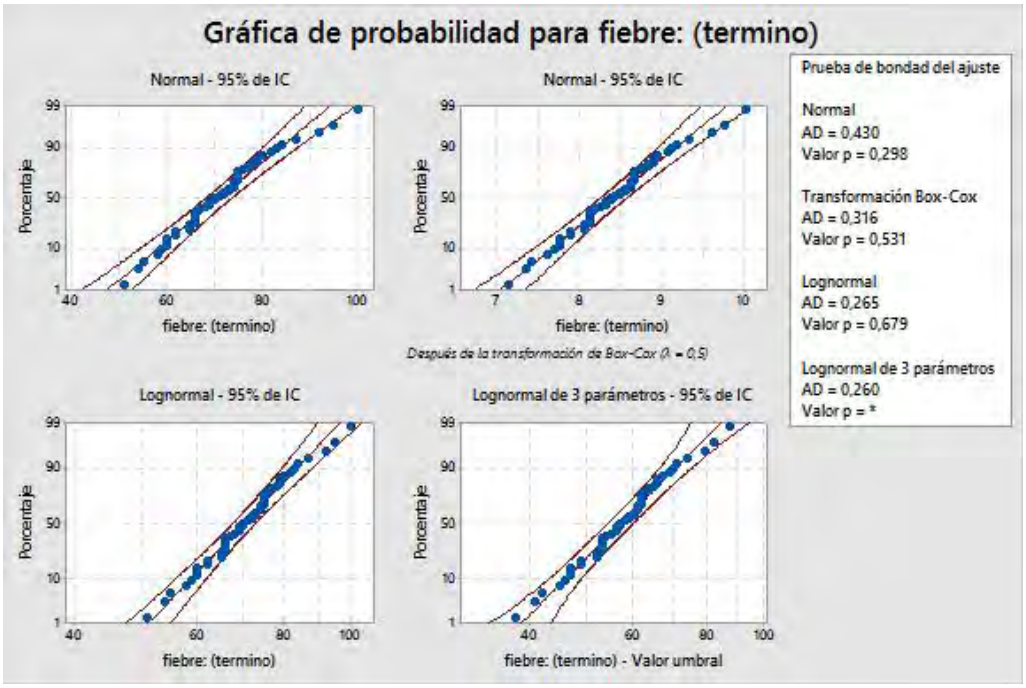

Figura 12. Identificación de gráfico de dispersión para datos informales de Fiebre 2016

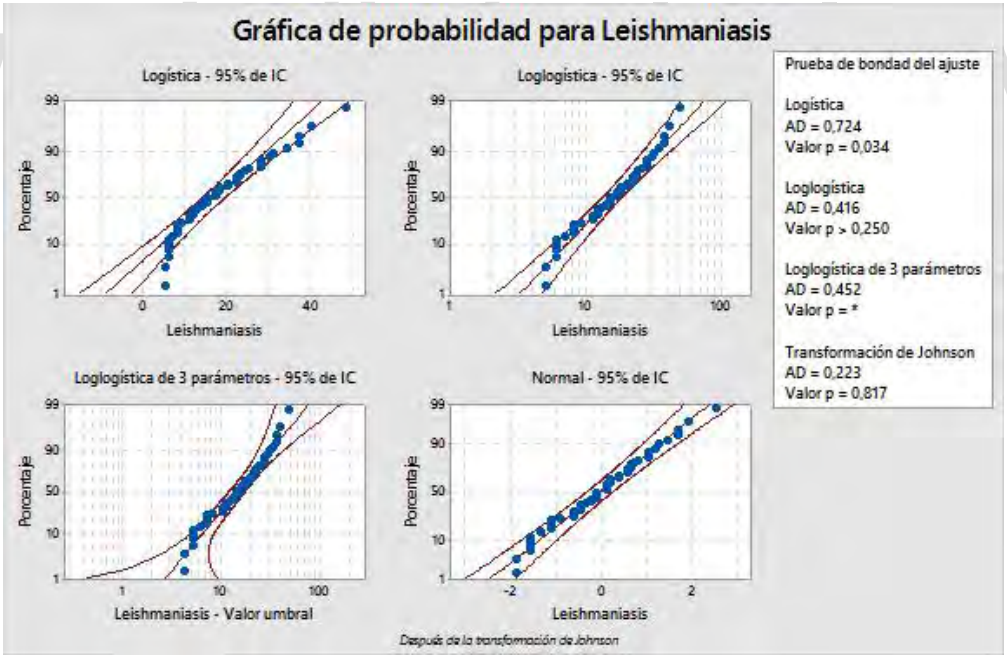

Figura 13. Identificación de gráfico de dispersión para datos oficiales de Leishmaniasis 2016

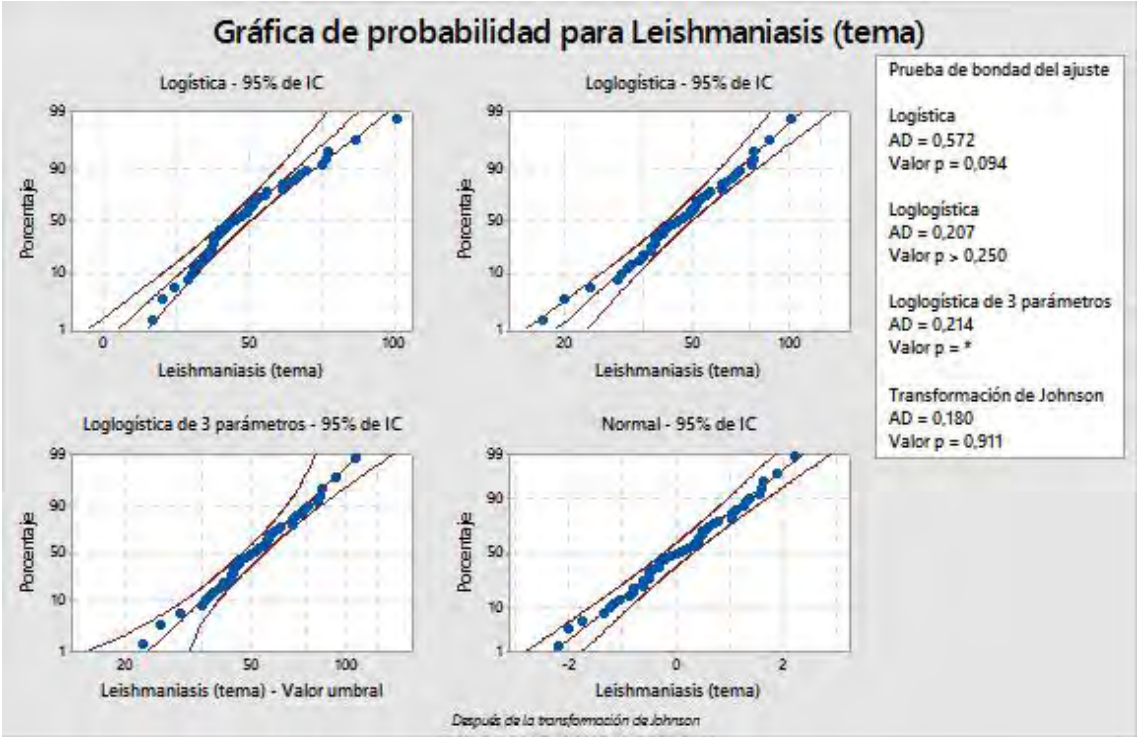

Figura 14. Identificación de gráfico de dispersión para datos informales de Leishmaniasis 2016

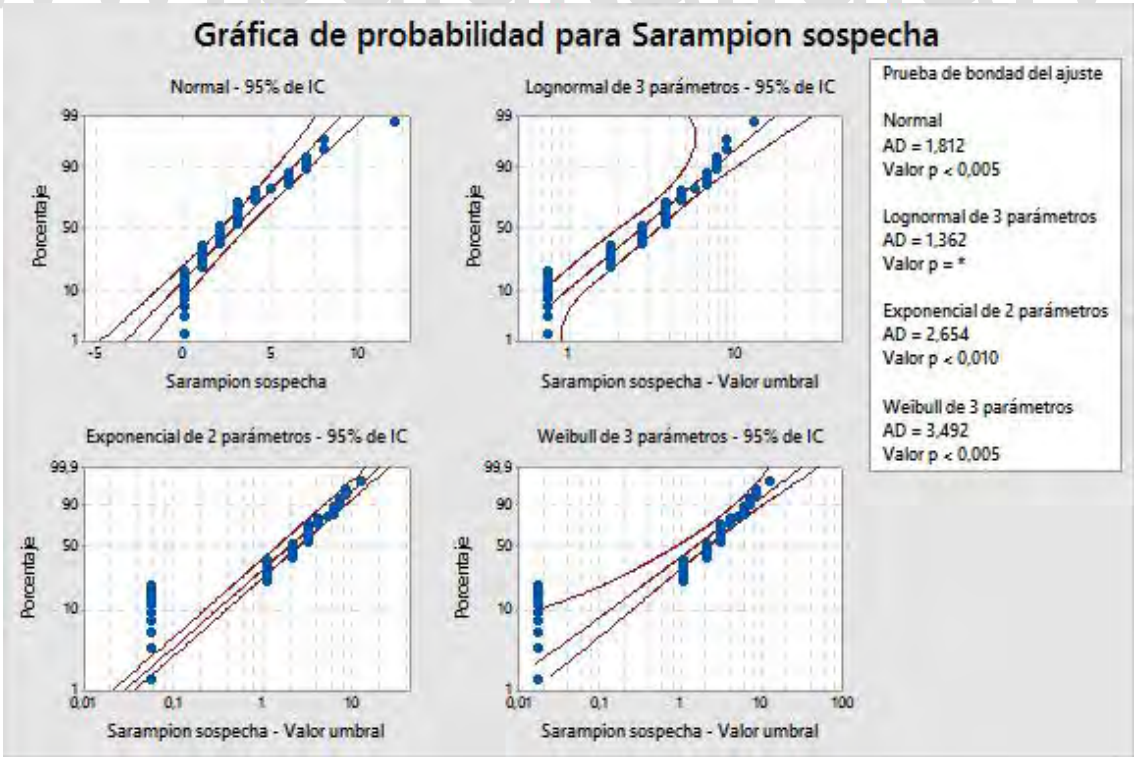

Figura 15. Identificación de gráfico de dispersión para datos oficiales de Sarampión 2016

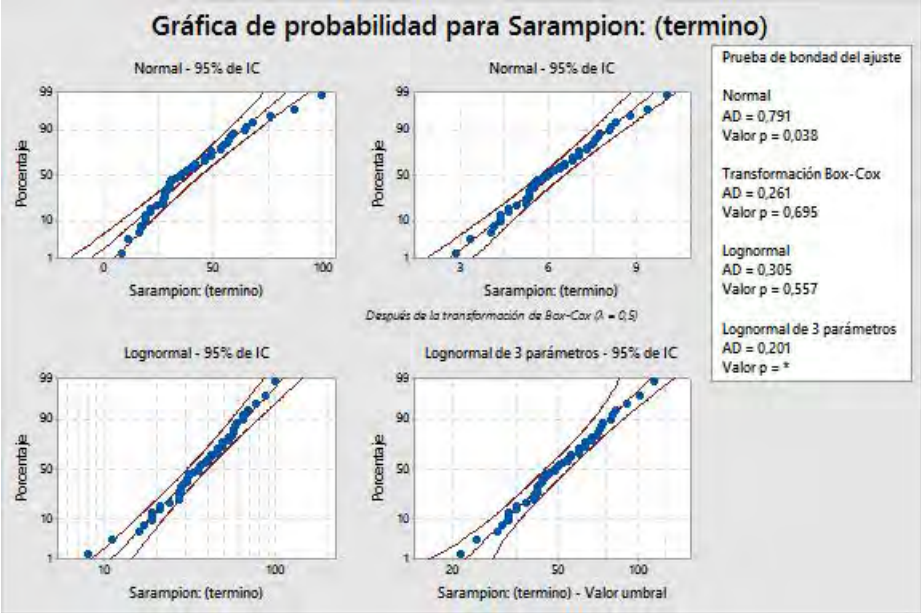

Figura 16. Identificación de gráfico de dispersión para datos informales de Sarampión 2016

Los modelos de dispersión indican que la relación entre las variables (datos formales e informales) no es lineal y más bien se acerca a una relación monótona, por ende se usó el coeficiente de correlación de Spearman.

A continuación, se muestran los valores de correlación entre las variables, para los años 2006, 2010 y 2016 de algunas de las enfermedades, para ilustrar el comportamiento de los datos en distintos periodos de tiempo, en los que el acceso de la población a internet es distinto, y también el sistema de salud oficial del país.

Tabla 11. Valores representativos para tres años de correlación de Spearman (rho) y su valor de p

Enfermedad	Año	rho - p (control/termino)	rho - p (control/tema)	rho - p (termino/tema)
Dengue	2006	**0,469 – 0**	0,434 - 0	0,972 - 0
	2010	0,353 - 0,01	0,332 - 0,016	0,995 - 0
	2016	- 0,472 - 0	- 0,472 - 0	0,998 - 0
Difteria	2006	*	*	0,352 - 0
	2010	*	*	0,615 - 0
	2016	*	*	0,999 - 0
Fiebre	2006	0,243 - 0,083	0,249 - 0,075	0,576 - 0
	2010	0,44 – 0	0,504 - 0	0,518 - 0
	2016	0,627 – 0	0,634 - 0	0,74 - 0
Hepatitis	2006	0,157 - 0,266	0,376 - 0,006	0,368 - 0,007
	2010	0,148 - 0,304	0,185 - 0,199	0,875 - 0

	2016	0,251 - 0,073	0,305 - 0,208	0,917 - 0
Leishmaniasis	2006	0,217 - 0,122	0,262 - 0,061	0,65 - 0
	2010	0,138 - 0,339	0,065 - 0,656	0,871 - 0
	2016	0,262 - 0,064	0,236 - 0,095	0,919 - 0
Malaria	2006	0,128 - 0,366	0,19 - 0,177	0,444 - 0,001
	2010	0,03 - 0,83	-0,131 - 0,354	0,786 - 0
	2016	0,169 - 0,232	0,214 - 0,128	0,824 - 0
Rubéola	2006	**0,651 – 0**	0,606 - 0	0,391 - 0,004
	2010	0,34 - 0,016	0,372 - 0,008	0,877 - 0
	2016	- 0,013 - 0,927	- 0,089 - 0,529	0,926 - 0
Sarampión	2006	**0,829 – 0**	0,809 - 0	0,99 - 0
	2010	0,055 - 0,705	0,062 - 0,671	0,903 - 0
	2016	- 0,001 - 0,995	0,075 - 0,595	0,711 - 0

(*) Los valores no pueden ser calculados ya que en todo el año el número de casos en el boletín epidemiológico fue 0

Tendencias 2017

Como no fue posible correlacionar los datos formales e informales como era esperado, se llevó a cabo la descarga de datos de las tendencias de Google 2017 para cada una de las enfermedades y se graficó el número de búsquedas para cada semana epidemiológica, de forma similar que una epi-curva, pero no puede llamarse así en vista de que las buscas no representan directamente el número de casos en un evento epidemiológico. En los siguientes gráficos se representa el comportamiento de las tendencias de Google para algunas de las enfermedades en 2017, con la intención de compararlos con la información publicada por la OMS en sus actualizaciones epidémicas para el mismo año.

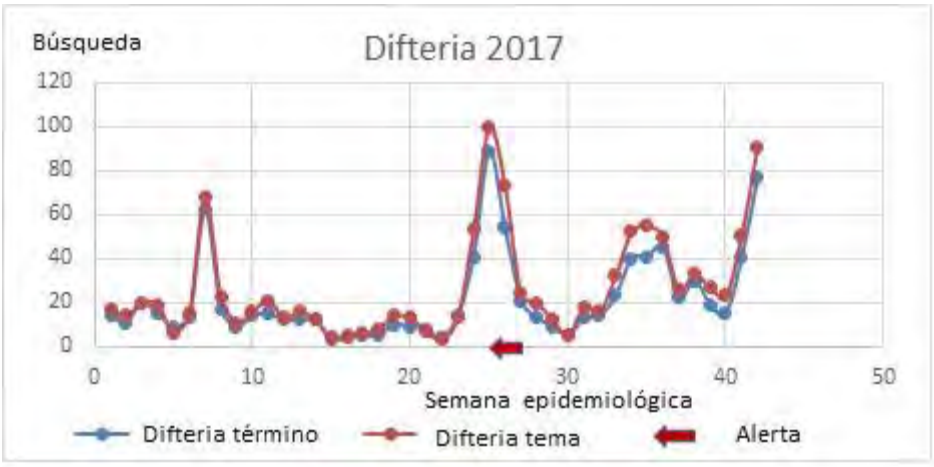

Figura 17. Gráfico epidemiológico de datos informales de difteria en Venezuela 2017

En la tabla 12 se observa el interés de la población de acuerdo con las tendencias de búsqueda, divididos por subregión, y se puede evidenciar cuales de las regiones muestran más interés en el tema. Esta información es comparada con los reportes de casos publicados por la OMS en su actualización epidemiológica de difteria 2017, información que publican con detalle únicamente en la actualización epidemiológica de difteria 2017.

Tabla 12. Interés por subregión en búsquedas de "Difteria" en 2017 y reportes de alerta de la OMS

Región	Término	Tema	Casos OMS
Amazonas	50	100	0
Trujillo	80	74	3
Monagas	71	80	26
Apure	34	61	19
Mérida	55	61	3
Zulia	56	60	0
Carabobo	55	60	1
Anzoátegui	52	57	37
Nueva Esparta	32	52	1
Distrito Capital	33	43	9
Táchira	38	43	0
Guárico	34	39	0
Bolívar	34	39	282
Cojedes	38	23	6
Aragua	27	33	0
Sucre	31	31	10
Yaracuy	31	28	4
Vargas	30	26	5
Portuguesa	24	30	2
Falcón	25	30	0
Miranda	29	29	29
Delta Amacuro	28	18	0
Lara	17	21	0
Barinas	13	16	2

Así mismo, en la figura 18 se ven representadas las tendencias de Google 2017 para sarampión, y en el mismo gráfico se indica con una flecha el evento reportado por la OMS en su actualización epidemiológica de sarampión 2017, y en la tabla que le sigue, el interés mostrado por la población para distintas regiones del país.

Figura 18. Gráfico epidemiológico de datos informales de Sarampión en Venezuela 2017

Tabla 13. Interés por subregión en "Sarampión" en 2017 según tendencias de Google

Región	Sarampión: (término)	Sarampión: (tema)
Bolívar	82	100
Sucre	44	59
Anzoátegui	49	54
Barinas	0	52
Aragua	44	47
Táchira	40	47
Mérida	26	47
Guárico	26	47
Carabobo	31	43
Miranda	26	42
Zulia	30	42
Distrito Capital	33	41
Lara	27	33
Monagas	30	30
Monagas	20	28

En las figuras 19-23, se observa la representación gráfica de las tendencias de Google para algunas de las enfermedades, lo que permite tener una idea del comportamiento de los datos para las mismas en el tiempo transcurrido de 2017, y las tablas 13-17, muestran cómo se distribuyó el interés de la población en cada una de las enfermedades mostradas.

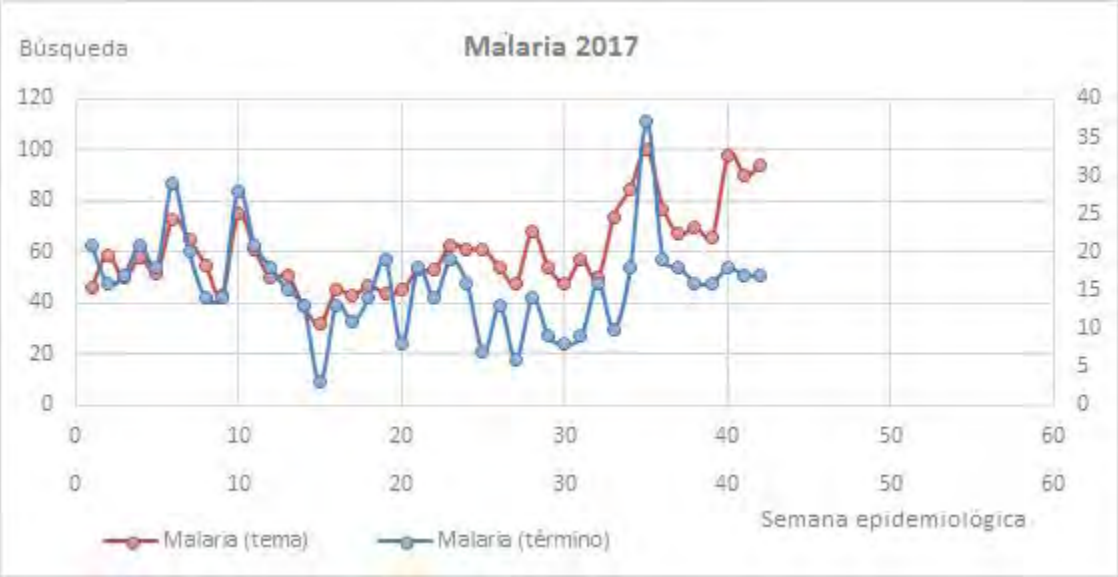

Figura 19. Gráfico epidemiológico de datos informales de Malaria en Venezuela 2017

Tabla 13. Interés por subregión en "Malaria" en 2017 según tendencias de Google

Región	Malaria (término)	Malaria: (tema)
Amazonas	0	100
Bolívar	12	87
Delta Amacuro	0	75
Sucre	13	65
Anzoátegui	10	58
Monagas	9	45
Guárico	14	35
Nueva Esparta	10	35
Distrito Capital	8	34
Aragua	12	27
Apure	0	27
Cojedes	0	26
Miranda	9	25
Zulia	6	24
Trujillo	10	23
Carabobo	8	21
Yaracuy	10	19
Barinas	7	19
Vargas	3	19
Mérida	5	16
Portuguesa	3	13
Falcón	5	12
Lara	4	10

Figura 20. Gráfico epidemiológico de datos informales de VIH en Venezuela 2017

Tabla 14. Interés por subregión en "VIH" en 2017 según tendencias de Google

Región	VIH (termino)	VIH: (tema)
Trujillo	61	65
Delta Amacuro	44	44
Portuguesa	65	69
Apure	65	73
Zulia	59	70
Guárico	41	50
Barinas	48	56
Bolívar	53	63
Nueva Esparta	62	68
Mérida	56	68
Aragua	49	56
Sucre	53	64
Yaracuy	38	46
Falcón	60	60
Cojedes	32	32
Vargas	36	50
Distrito Capital	46	58
Anzoátegui	52	56
Carabobo	48	56
Monagas	41	44
Táchira	47	54
Miranda	38	48
Lara	40	45

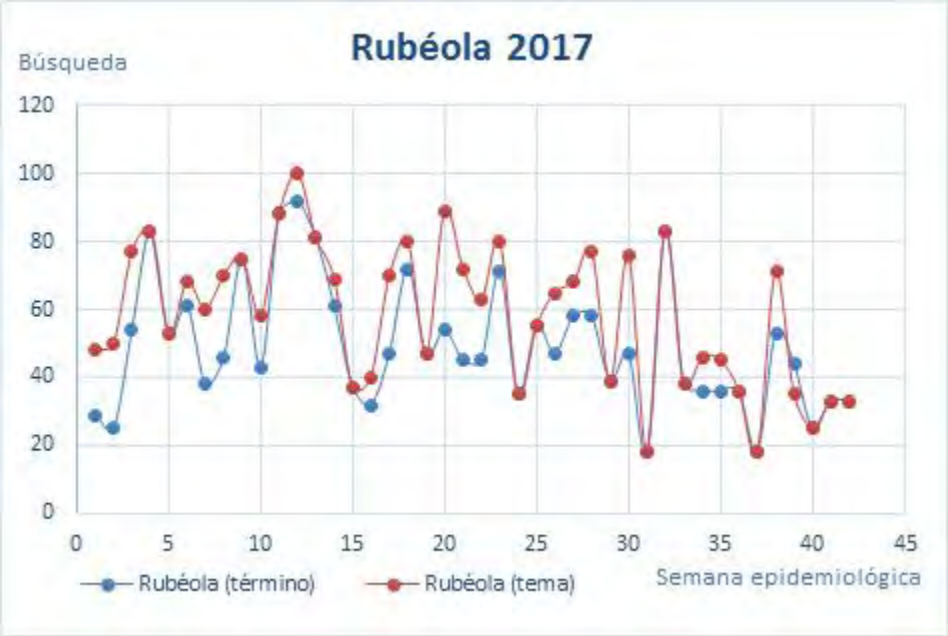

Figura 21. Gráfico epidemiológico de datos informales de Rubéola en Venezuela 2017

Tabla 15. Interés por subregión en "Rubéola" en 2017 según tendencias de Google

Región	Rubéola: (termino)	Rubéola: (tema)
Mérida	89	100
Guárico	0	89
Aragua	66	86
Bolívar	48	70
Zulia	53	68
Miranda	49	68
Anzoátegui	55	67
Distrito Capital	45	59
Lara	44	59
Carabobo	53	59
Táchira	48	52
Monagas	0	17

Figura 22. Gráfico epidemiológico de datos informales de Diarrea en Venezuela 2017

Tabla 16. Interés por subregión en "Diarrea" en 2017 según tendencias de Google

Región	Diarrea: (termino)	Diarrea: (tema)
Apure	100	100
Cojedes	92	92
Trujillo	81	81
Anzoátegui	78	78
Falcón	74	74
Guárico	73	73
Táchira	73	73
Mérida	72	72
Miranda	72	72
Nueva Esparta	70	70
Barinas	69	69
Bolívar	69	69
Zulia	65	65
Distrito Capital	64	64
Carabobo	63	63
Aragua	63	63
Yaracuy	62	62
Vargas	60	60
Portuguesa	58	58
Lara	56	56
Monagas	52	52
Sucre	44	44

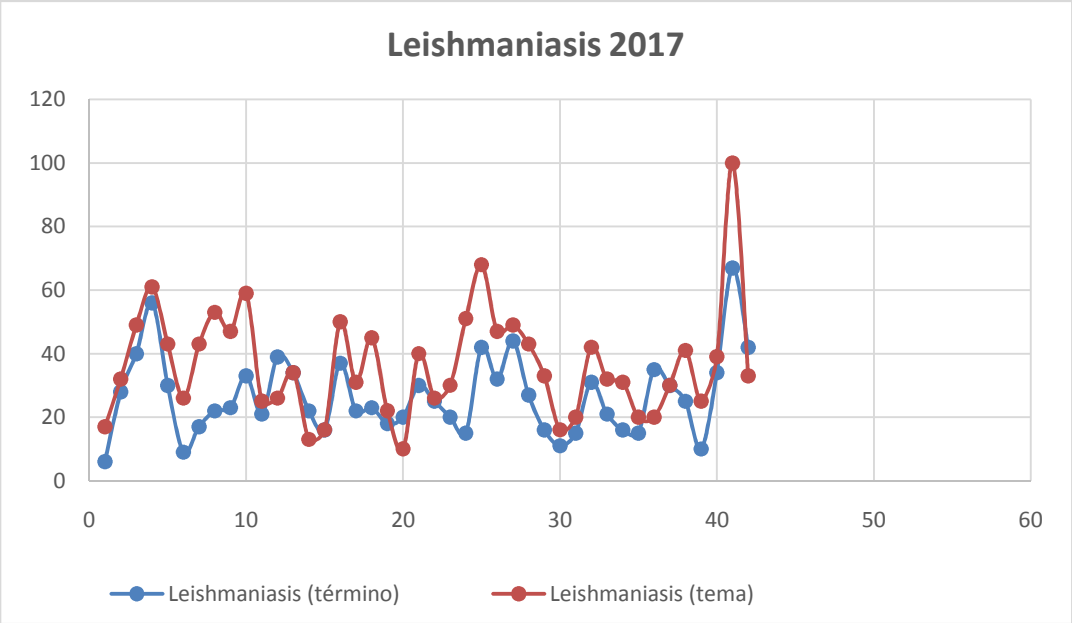

Figura 23. Gráfico epidemiológico de datos informales de Leishmaniasis en Venezuela 2017

Tabla 17. Interés por subregión en "Leishmaniasis" en 2017 según tendencias de Google

Región	Leishmaniasis (termino)	Leishmaniasis: (tema)
Trujillo	54	100
Mérida	29	37
Lara	29	35
Aragua	23	30
Táchira	21	30
Monagas	16	25
Miranda	18	25
Distrito Capital	19	24
Zulia	19	24
Sucre	0	23
Carabobo	19	21
Anzoátegui	16	20
Guárico	16	16
Bolívar	13	16
Vargas	13	13

CAPÍTULO V - **Discusión**

V.1. Análisis estadístico

En la tabla 11, se observan los valores de correlación entre los datos informales y formales, y también de los datos informales entre sí. Se incluyen algunas de las enfermedades para tres años con la intención derepresentar el comportamiento de los datos.

En el año 2006 se puede observar un buen grado de correlación entre los datos formales e informales para dengue, rubéola y sarampión; las demás enfermedades tienen valores que indican que hay poca o ninguna correlación entre las variables, comportamiento que se intensifica a medida que aumenta el número de años, y como se observa, en 2016 ya no hay correlación positiva entre los datos formales e informales. Este resultado podía esperarse, ya que como se observa en los gráficos de dispersión (figuras 11, 13 y 15) los datos formales, proporcionados por el Ministerio del Poder Popular para la salud, no se adaptan correctamente a ninguno de los modelos de dispersión, evento que se repite para casi todos los años y enfermedades, lo cual podría deberse a un muestreo incorrecto de los datos o al manejo inadecuado de los mismos por parte de las instituciones de vigilancia epidemiológica tradicional de Venezuela. Es importante resaltar que la información aportada por dichas instituciones es muy escasa y no mantiene un formato sistemático en las distintas publicaciones a lo largo del tiempo, lo que dificulta su análisis.

También puede observarse en la tabla 11, que la correlación entre dos fuentes de datos informales es muy buena, con valores de *p* que indican una excelente significancia. En los gráficos12, 14 y 16, se observa que, al identificar un modelo de dispersión, los datos informales se adaptan generalmente a un modelo de distribución normal o logística, La función logística, o curva en forma de S es una función matemática que aparece en diversos modelos de crecimiento de poblaciones, propagación de enfermedades epidémicas y difusión en redes sociales. y la distribución normal es una de las distribuciones de probabilidad que con más frecuencia aparece aproximada en fenómenos reales. [47] Esto indica que el comportamiento de los datos informales se ajusta a un modelo de propagación de enfermedades.

V.2. Modelado epidemiológico

Luego del desarrollo de esta investigación, no fue posible elaborar un modelo predictivo como se había planteado, ya que como se explica en la metodología, cada uno de los modelos epidemiológicos propuestos presenta algunas limitaciones. Sumado a esto, Witten-Poulter 2006, explican que además del tamaño epidémico, tamaño y distribución de los brotes, hay una serie de medidas de una epidemia necesarias para llevar a cabo una simulación real. Por un lado, es

importante conocer la distribución de los nodos infectados: ¿están concentrados o dispersos? ¿La infección se propaga en forma de anillo, dejando atrás al resto de los nodos? En el contexto del VIH, es importante saber la proporción de nodos de alto grado que están infectados para determinar su importancia en la propagación de la enfermedad. Para rastrear el progreso de la enfermedad, la tasa de nuevas infecciones es una medida importante, así como la proporción de susceptibles que se mantuvieron intactos durante el impacto de la epidemia, como una red con nodos de bajo grado, entre otros factores, y los boletines epidemiológicos usados reflejan únicamente el número de casos semanales de cada enfermedad para el país, y como se mencionó previamente, la mayoría no se ajustan correctamente a ningún modelo de dispersión.

En 2015 Zhao L. y colaboradoresproponen que para tener modelos que permitan acercase a la realidad, no solo debe incluirse la simulación de la epidemia basada en el individuo, sino que también la inferencia del estado de salud del individuo según redes sociales, teniendo así la ventaja de incluir el enfoque de la epidemiología computacional pero también del *social mining*. En este caso no se cuenta con información suficiente para ninguno de los enfoques, por la deficiente información oficial acerca de las enfermedades de notificación obligatoria, y no fue posible obtener información suficiente de las redes sociales Twitter y Facebook debido a recientes cambios en sus políticas de seguridad, para construir redes que permitan tener idea de la forma en la que están ocurriendo las interacciones entre los individuos. A partir de Twitter fue posible descargar datos, pero abarcaban un máximo de 12 días; en el caso de Facebook no fue posible obtener datos. La literatura explica que la estructura de la red proporciona muchas propiedades que influyen en la propagación de una epidemia. Con la mayoría de las enfermedades, las personas con las que tiene contacto un individuo probablemente tengan contacto entre ellas o, en el contexto del VIH, un pequeño número de los nodos de alto grado podrían ser una influencia significativa en su propagación [5]. También podemos esperar que la estructura de propagación de la enfermedad sea diferente para una enfermedad que se transmite a través del aire, que para las que se transmiten por contacto o sexualmente. [44]

Por otro lado, como se observa en la tabla 9, las enfermedades quedan ubicadas en distintos modelos de compartimiento por sus características epidemiológicas propias, como el número de estados que atraviesa un individuo infectado, y se ha demostrado que una clasificación exhaustiva de las enfermedades es fundamental para poder establecer un modelo. Esto indica que es necesario elaborar un modelo predictivo para cada una de las enfermedades, tomando en cuenta las características epidemiológicas de la enfermedad y además las características de la población estudiada, tanto a nivel de contactos entre los nodos, como la influencia de los contactos geográficos en la propagación de un brote.

V.3. Alerta epidemiológica temprana

En vista de que no hubo correlación de los datos formales e informales, pero hubo muy buena correlación entre los informales entre sí, se corroboraron los datos con información de la organización mundial de la salud (OMS).

En la figura 17 están representadas las tendencias de Google (término y tema) hasta la semana 44 para “Difteria 2017”, y se puede observar que entre la semana 5 y 9 hay un pico, que representa un aumento en el interés de la población por esta enfermedad, que podría ser un indicador del inicio de un brote, y según el gráfico el interés por esta enfermedad sigue aumentando a lo largo del año. Esta información coincide con la actualización epidémica de difteria de la OMS del 22 de agosto de 2017, [49] que indica queen Venezuela, entre la SE (semana epidemiológica) 28 de 2016 y la SE 24 de 2017 se notificaron 447 casos sospechosos de difteria (324 en 2016 y 123 en 2017) confirmados por laboratorio 51 casos, incluidas 7 defunciones en Anzoátegui (2 casos), Bolívar (1 caso), Monagas (3 casos), y Sucre (1caso); teniendo entre los casos confirmados una tasa de letalidad del 20%. En la tabla 11 se observa el interés en difteria por subregión de acuerdo a las tendencias de Google, y las regiones que reportaron casos sospechosos según la OMS. Se puede evidenciar que hay una relación importante entre los estados con mayor búsqueda y los estados que reportaron casos sospechosos.

En este informe se reportan 324 casos sospechosos de difteria para 2016, y el boletín epidemiológico de 2016 se reporta un total de 0 casos para difteria en el país, mientras que las tendencias tuvieron una curva que también representa un aumento de la población por la enfermedad para esta fecha (figura 24).

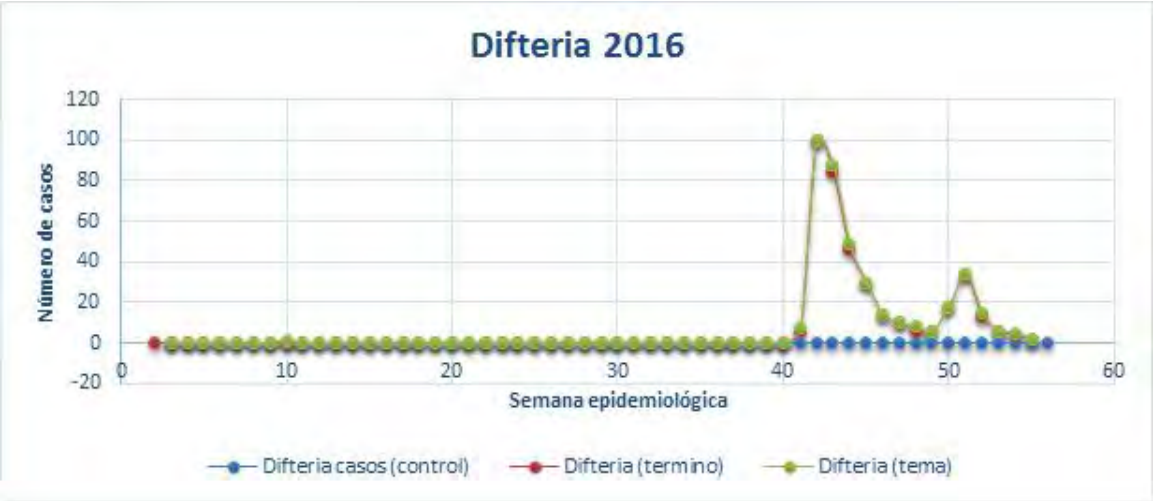

Figura 24. Comparación temporal de tendencias de Google y reportes de boletines epidemiológicos para difteria 2016.

Otro ejemplo está representado en la figura 18, el gráfico epidemiológico con datos informales de sarampión 2017, donde se observan pequeños picos de indican que en el trascurso del año 2017 ha habido interés acerca del tema, seguidos de un pico más pronunciado entre las semanas 33 y 40. El 22 de septiembre de 2017 la OMS publica que entre la SE 26 y la SE 35 de 2017se notificaron 84 casos sospechosos de sarampión en 10 parroquias en el municipio de Caroní, estado Bolívar, Venezuela. Del total, 34 casos fueron confirmados por laboratorio, 42 están bajo investigación y 8 fueron descartados. El 79% (n=27) de los casos confirmados tienen una edad ≤ 9 años. [50] Esto permite inferir que el interés de la población puede estar relacionado con la presencia de un brote de sarampión, cuyos casos han ido en aumento en varios países de América y Europa según este mismo reporte.

En el gráfico epidemiológico de datos informales de malaria 2017 (figura 19) se observa que las tendencias de búsquedas de malaria han sido frecuentadas a lo largo del año y tienden a aumentar, información que coincide con el comportamiento de la malaria en Venezuela.La Alianza Venezolana de la Salud, junto con la Red Defendamos la Epidemiología y el Observatorio Venezolano de la Salud emitió una carta abierta a los asistentes al foro Malaria en las Américas 2017, evento que se realiza en Washington D.C. En ella indican que en un periodo de 16 años la malaria en Venezuela ha crecido en un 709%, con un aumento del 521% de muertes relacionadas con la malaria y un aumento del 540% en la incidencia parasitaria anual (Figura 25). [51], [52]

Figura 25. Cambio en los porcentajes de casos notificados de malaria. Venezuela, 1988-2016.

Además, señalan que en la semana epidemiológica número 26 (datos no publicados) de este año, se evidenció un aumento del 70% de los casos comparados con el mismo periodo del año anterior, así

como un aumento de dos veces en el número de muertes relacionadas con esta enfermedad.El 6 de noviembre de 2017, el periódico el nacional publica un artículo en el que se afirma que hasta el 21 de octubre se reportaron 206.240 casos de la enfermedad solo en el estado Bolívar, lo que representa un incremento de 42% con respecto al mismo lapso de tiempo reflejado en la semana 41 en esa región cuando se produjeron 144.762 casos; también se han reportado casos en Amazonas y Sucre. Se puede observar una coincidencia con el gráfico de datos informales de 2017, donde se evidencia un aumento de interés entre las SE 30 y SE 40, y también se observa en la tabla 13, una coincidencia con un mayor número de búsquedas en los estados Amazonas, Bolívar y Sucre. [52], [53]

Los demás gráficosde tendencias de 2017 se muestran para tener idea del interés que tiene la población acerca de ciertas condiciones de salud, que como sugieren los comparados con los reportes de la OMS, aportan información acerca de eventos epidemiológicos, por esta razón se propone el uso de esta información como herramienta para la alerta temprana y respuesta, definida por el CDC como un mecanismo establecido para detectar lo antes posible cualquier acontecimiento anormal o cualquier alteración de la frecuencia habitual. Las fuentes de información que se pueden usar para la alerta temprana van mucho más allá de la vigilancia tradicional basada en enfermedades que incluye la confirmación del laboratorio. Abarcan desde vigilancia ambiental y ecológica, densidad de vectores, calidad del agua y el aire, datos climáticos, entre otros; hasta la información sobre el comportamiento relacionado con la salud, seguimiento del ausentismo escolar o laboral, venta de medicamentos y productos paramédicos como repelentes de insectos, actividades en Internet o en las redes sociales, etc. En tal sentido la información obtenida de las tendencias puede ser usada cabalmente para este fin, puesto que permite recopilar información antes de que ocurran casos humanos o antes de que un evento se detecte o se notifique a través de los sistemas de registro y notificación convencionales. [54]

Como se reconoce en el mandato de la OMS en el Artículo 9 del RSI referido al uso de otras fuentes de información, el mecanismo de alerta temprana y respuesta nacional debe integrar la recopilación y el análisis de información de cualquier fuente más allá de la generada por el sistema de salud. Este tipo de vigilancia se denomina "**vigilancia basada en eventos**". La vigilancia basada en eventos aumenta significativamente la sensibilidad del sistema de vigilancia. Una función eficaz de alerta temprana garantiza una respuesta rápida a los eventos agudos de salud pública de todos los orígenes, y por ende la mitigación del impacto en la salud pública. Esto requiere mayor coordinación y la colaboración estrecha con todos los interesados directos dentro y fuera del sector de la salud. [54]

Los procesos de recopilación de datos y análisis del mecanismo de alerta temprana y respuesta se deben sistematizar y formalizar para garantizar la eficiencia. A este respecto, el mecanismo dependerá de un proceso, inteligencia epidémica, que se define como la recopilación sistemática, el

análisis y la comunicación de cualquier información, para detectar, comprobar, evaluar e investigar eventos y riesgos para la salud con un objetivo de alerta temprana (en contraposición con la vigilancia de tendencias o carga de enfermedad). La inteligencia epidémica integra ambas fuentes de información, la vigilancia basada en indicadores y la vigilancia basada en eventos, para detectar eventos agudos o riesgos de salud pública. [54]

Por todo lo antes expuesto, se plantea que la información obtenida en las tendencias de Google puede ser útiles como información para la vigilancia de eventos, la cual es necesaria en el país para dar un paso adelante en un sistema epidemiológico completo, sensible, actualizado y además obtener las ventajas que ofrece la vigilancia de eventos a salud pública.

Por otro lado, como se plantea en la metodología, la representación gráfica aporta información valiosa, que permite tener idea del evento ocurrido, como el patrón de propagación, tendencias en el tiempo y magnitud de la epidemia, origen de la infección; factores que permiten profundizar el estudio de una epidemia, recopilar información, podría ser útil para proyectos futuros, como la construcción de un modelo predictivo.

CAPÍTULO VI - **Conclusiones**

No hay correlación entre los datos informales (tendencias de Google) y los formales (boletines epidemiológicos de Venezuela).

Los datos informales de 2017 se corresponden con la información publicada en las actualizaciones epidemiológicas de la OMS del mismo año.

Se propone el uso de datos informales, como herramienta útil para alertas epidemiológicas tempranas.

Los medios digitales y las redes sociales pueden servir de herramienta para la vigilancia de eventos y complementar el sistema de vigilancia epidemiológica clásico, para disminuir el impacto en salud pública y optimizar los recursos.

La elaboración de un modelo epidemiológico requiere de una amplia variedad de información epidemiológica y de la estructura de la población, información que se podría obtener de las redes sociales en línea.

No fue posible elaborar un modelo epidemiológico predictivo con la información disponible.

Perspectivas

Es necesario mejorar el sistema de vigilancia epidemiológica de Venezuela, y garantizar la publicación de información más extensa, donde además del número de casos semanales para las enfermedades, se publique también el porcentaje de individuos susceptibles e infectados en la población, el patrón de propagación del brote, entre otros, y que su publicación contenga un formato estructurado y constante a lo largo del tiempo, para facilitar el análisis de los mismos para investigaciones epidemiológicas y la posterior toma de decisiones.

Por otro lado, lo más indicado sería actualizar dicho sistema, y plantear un sistema de inteligencia epidemiológica, donde no solo se incluya la vigilancia tradicional por indicadores, sino incluir la vigilancia de eventos que certifica la información acerca de un brote lo antes posible, permitiendo así tomar acciones preventivas para disminuir el tamaño del brote, que de ser descubierto en un tiempo más prolongado, sería no solo más difícil de controlar, sino que la población afectada sería mayor.

Es importante promover en el país la educación acerca del valor de los datos epidemiológicos, para garantizar recursos y formación de recursos humanos, que permitan incluir el uso de herramientas computacionales para obtener información acerca del comportamiento de la población, la estructura de su red social, cuales son los nodos más prominentes en ella, para tomar decisiones conscientes en las campañas de vacunación, estudios focalizados en poblaciones de riesgo, entre otras cosas que no solo optimizan los recursos, también mejorarían la calidad del sistema de vigilancia que se ha tenido hasta ahora.

De igual forma es importante incluir el uso de herramientas computacionales para el análisis de los datos obtenidos, que como se mencionó anteriormente, proporcionan un análisis más eficiente, en tiempo real, mayor objetividad y una población de estudio más amplia.

Anexos

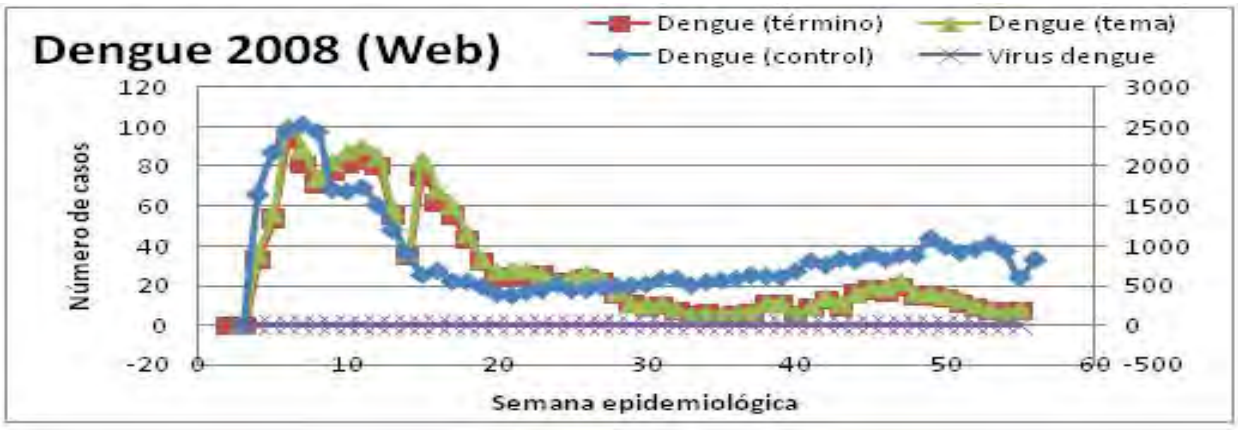

Figura 26. Comparación temporal de tendencias de Google y reportes de boletines epidemiológicos para dengue en 2008 (Web)

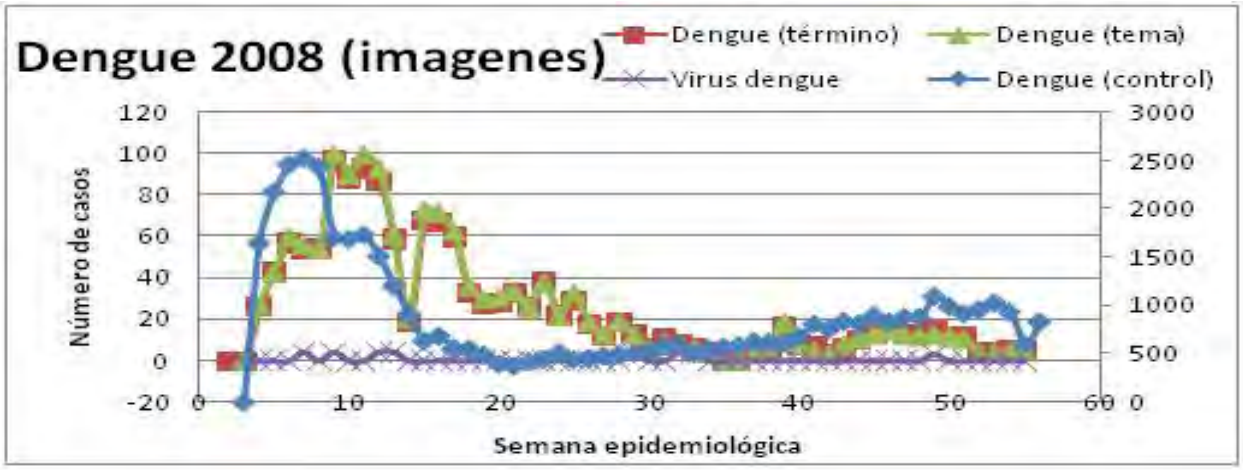

Figura 27. Comparación temporal de tendencias de Google y reportes de boletines epidemiológicos para dengue en 2008 (imágenes)

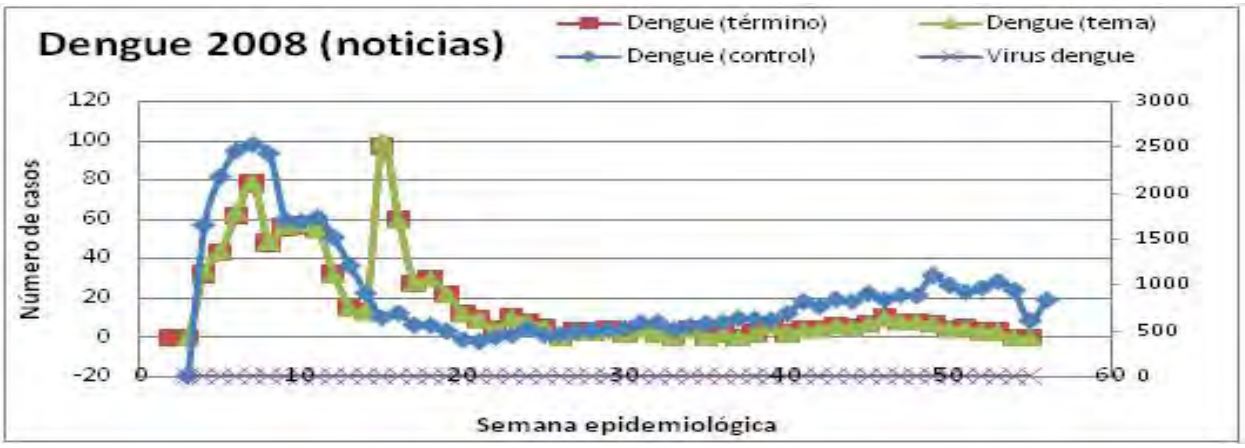

Figura 28. Comparación temporal de tendencias de Google y reportes de boletines epidemiológicos para dengue 2008 (noticias

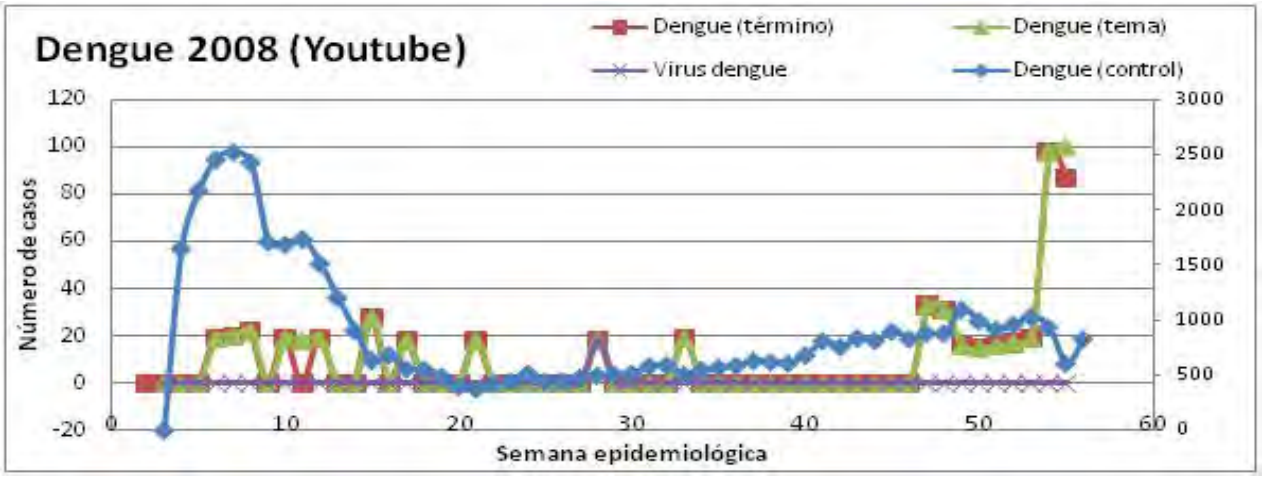

Figura 29. Comparación temporal de tendencias de Google y reportes de boletines epidemiológicos para dengue 2008 (Youtube)

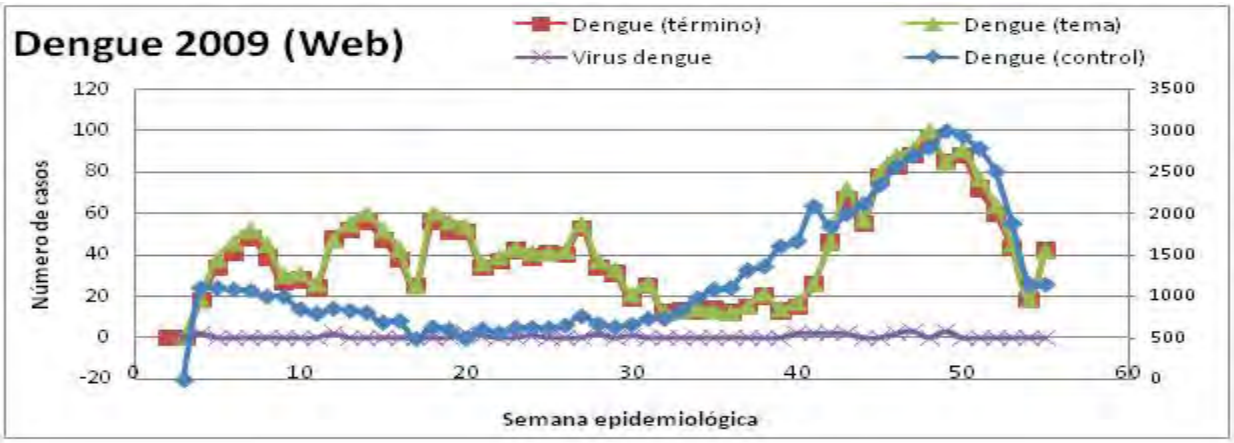

Figura 30. Comparación temporal de tendencias de Google y reportes de boletines epidemiológicos para dengue 2009 (Web)

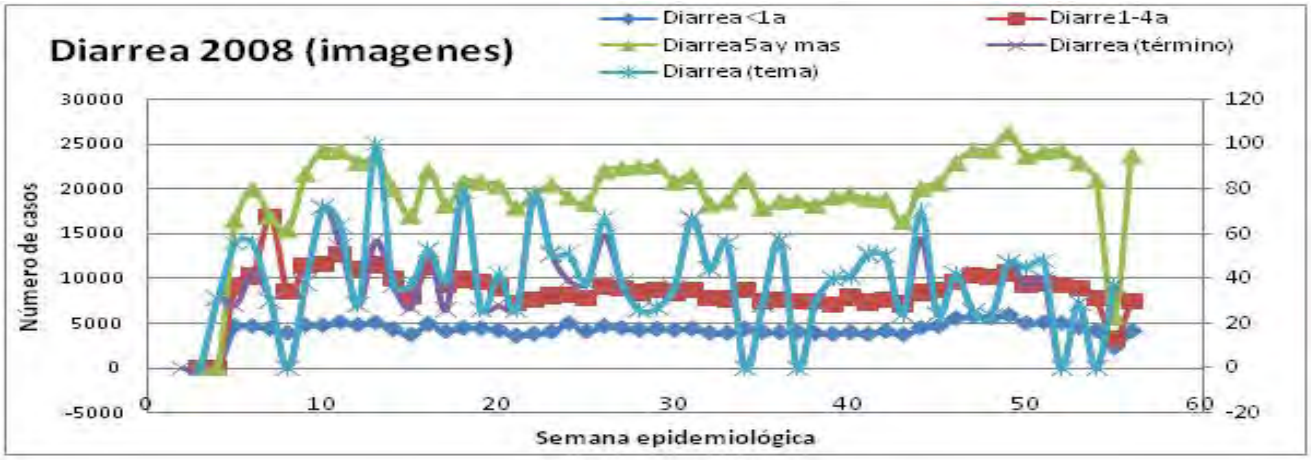

Figura 31. Comparación temporal de tendencias de Google y reportes de boletines epidemiológicos para diarrea en 2008 (imágenes)

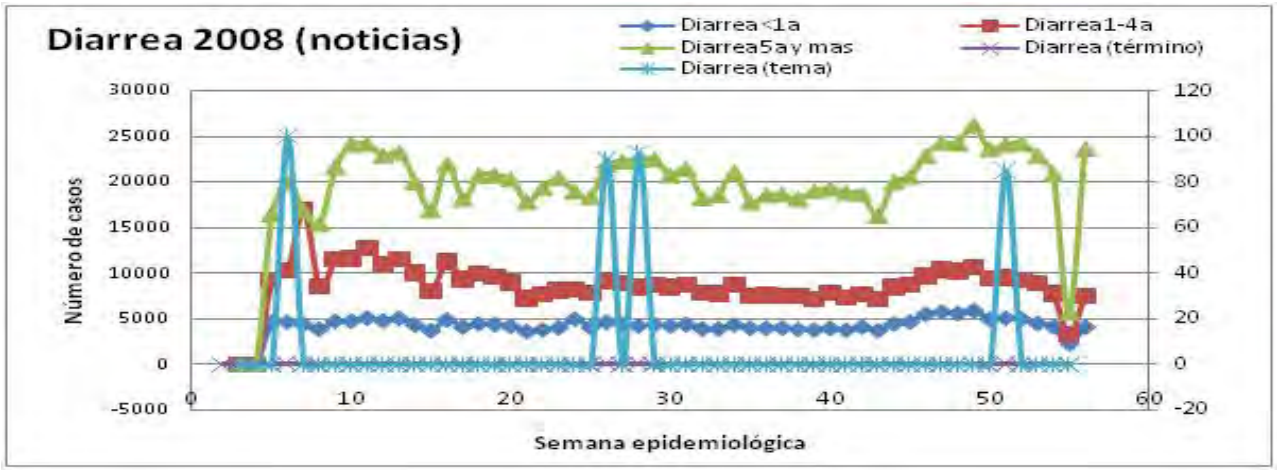

Figura 32. Comparación temporal de tendencias de Google y reportes de boletines epidemiológicos para diarrea en 2008 (noticias)

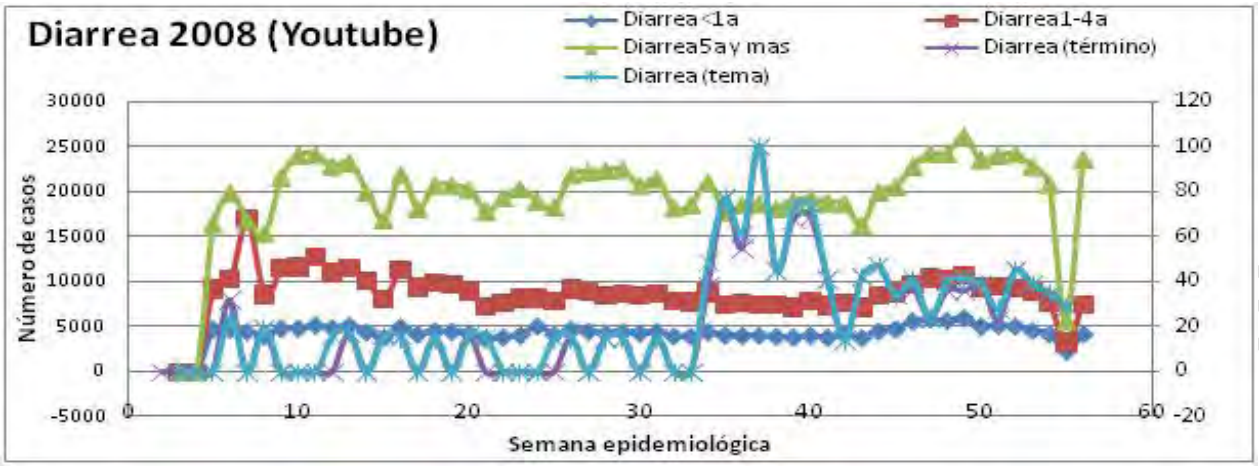

Figura 33. Comparación temporal de tendencias de Google y reportes de boletines epidemiológicos para diarrea en 2008 (Youtube)

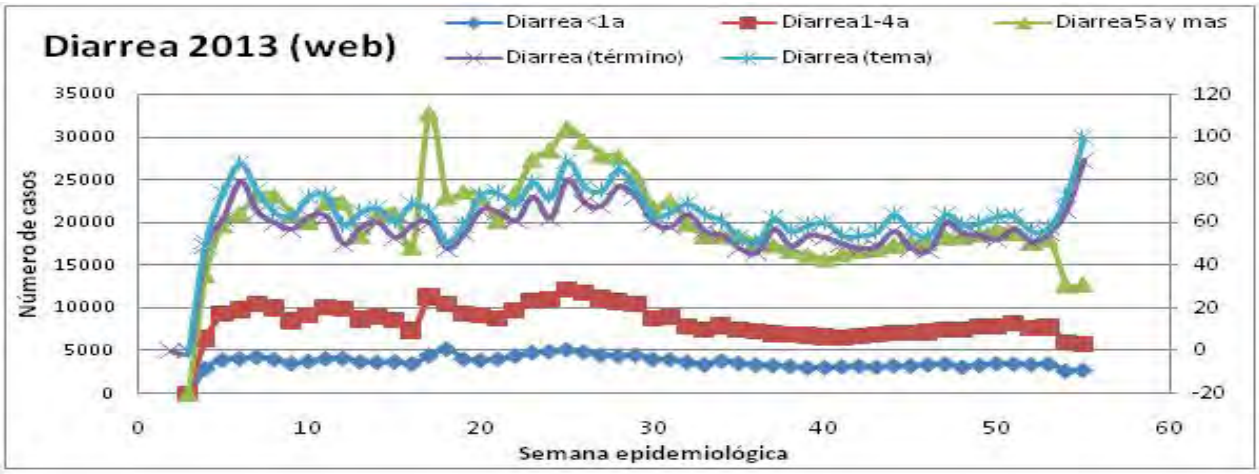

Figura 34. Comparación temporal de tendencias de Google y reportes de boletines epidemiológicos para diarrea en 2013 (Web)

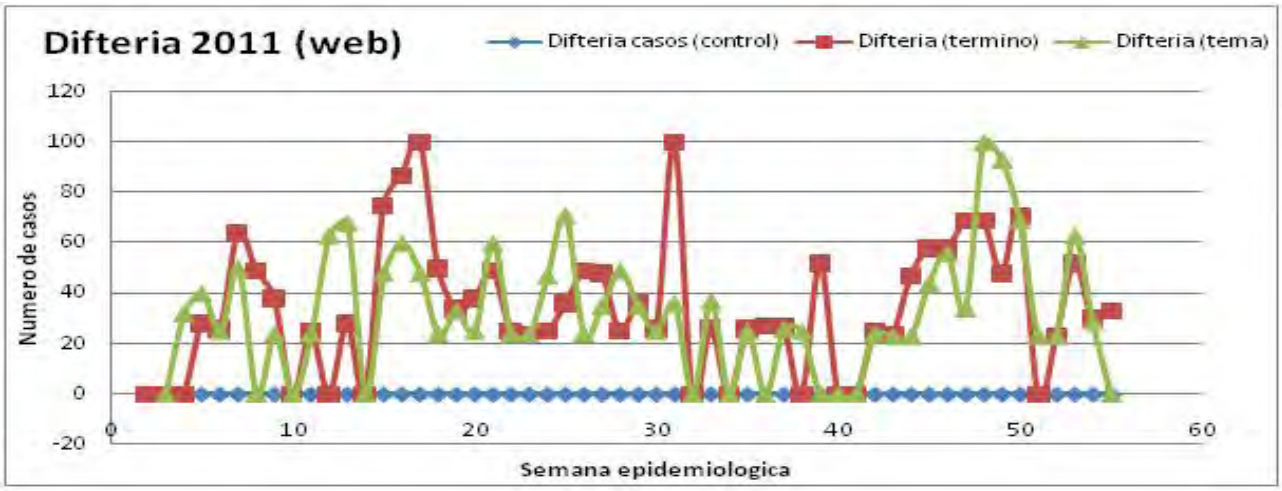

Figura 35. Comparación temporal de tendencias de Google y reporte de boletines epidemiológicos para malaria 2011 (Web)

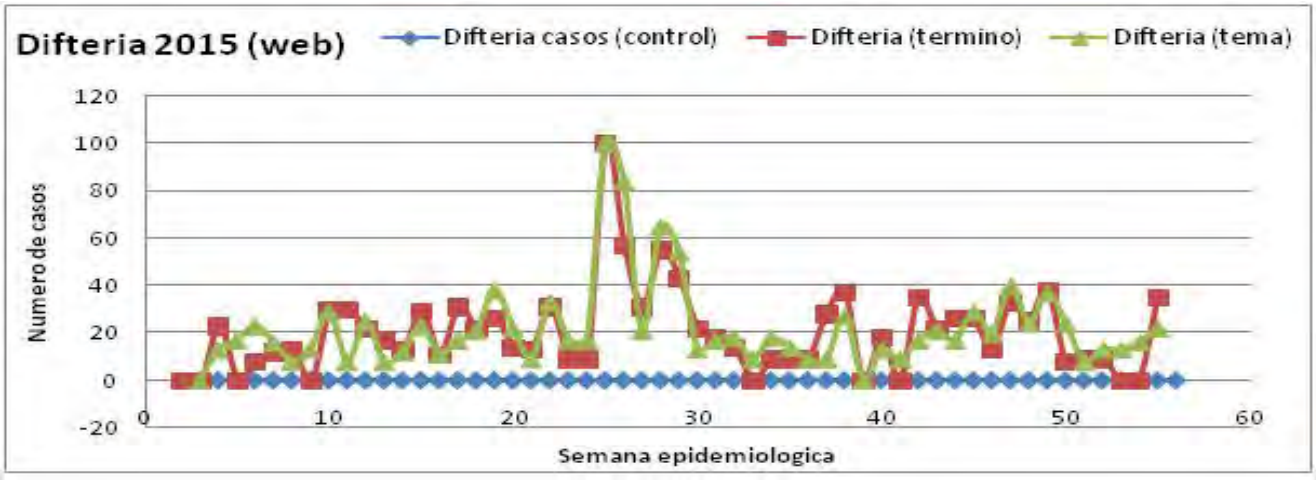

Figura 36. Comparación temporal de tendencias de Google y reporte de boletines epidemiológicos difteria 2015 (Web)

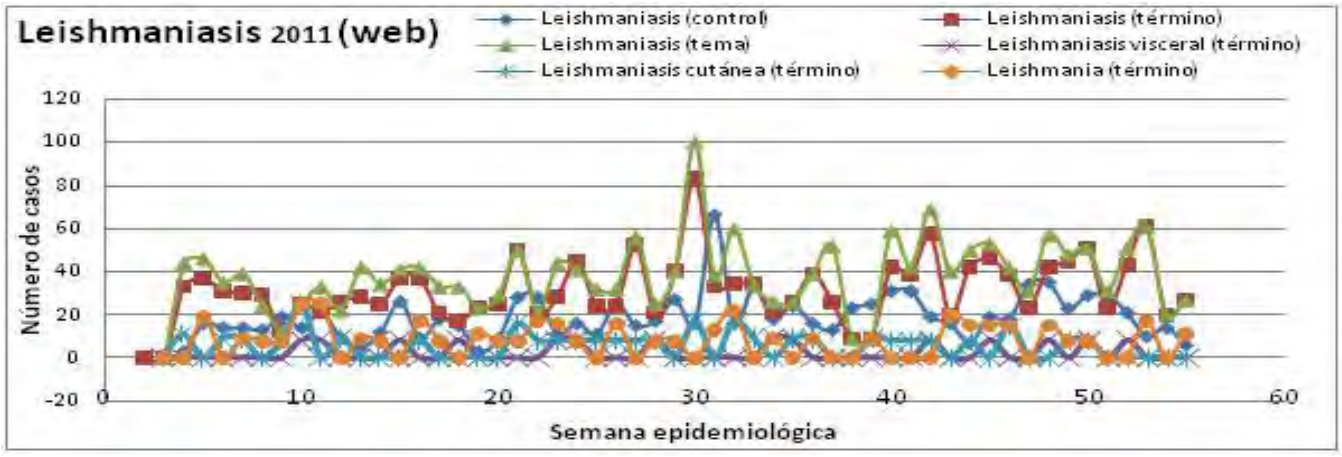

Figura 37. Comparación temporal de tendencias de Google y reporte de boletines epidemiológicos para leishmaniasis en 2011 (Web)

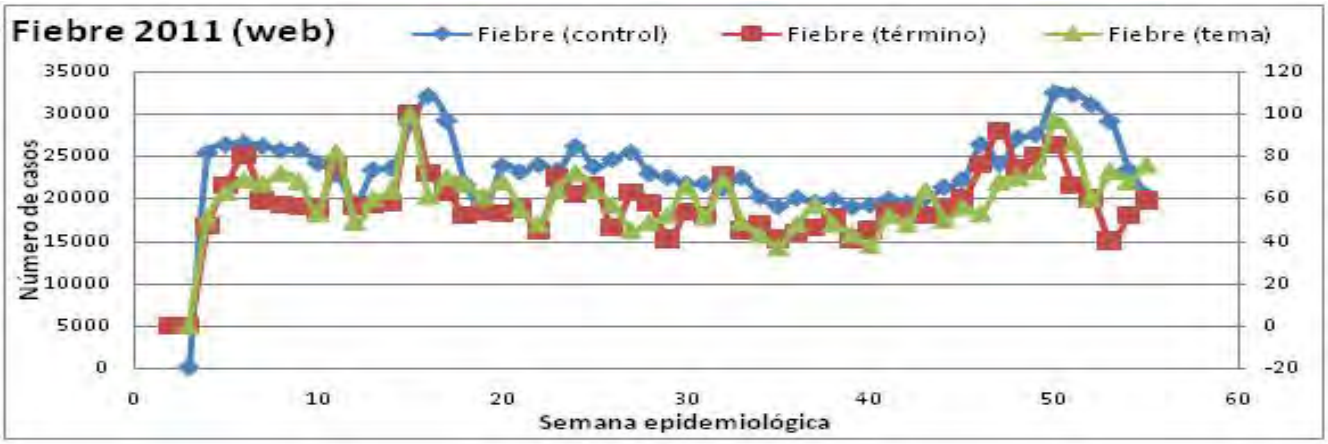

Figura 38. Comparación temporal de tendencias de Google y reporte de boletines epidemiológicos para fiebre en 2011 (Web)

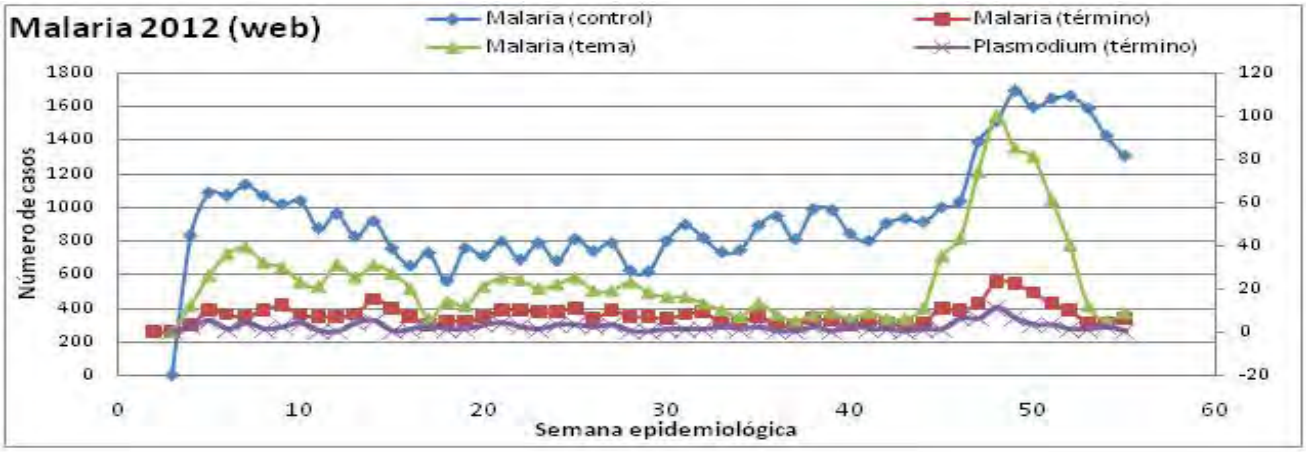

Figura 39. Comparación temporal de tendencias de Google y reporte de boletines epidemiológicos para malaria en 2012 (Web)

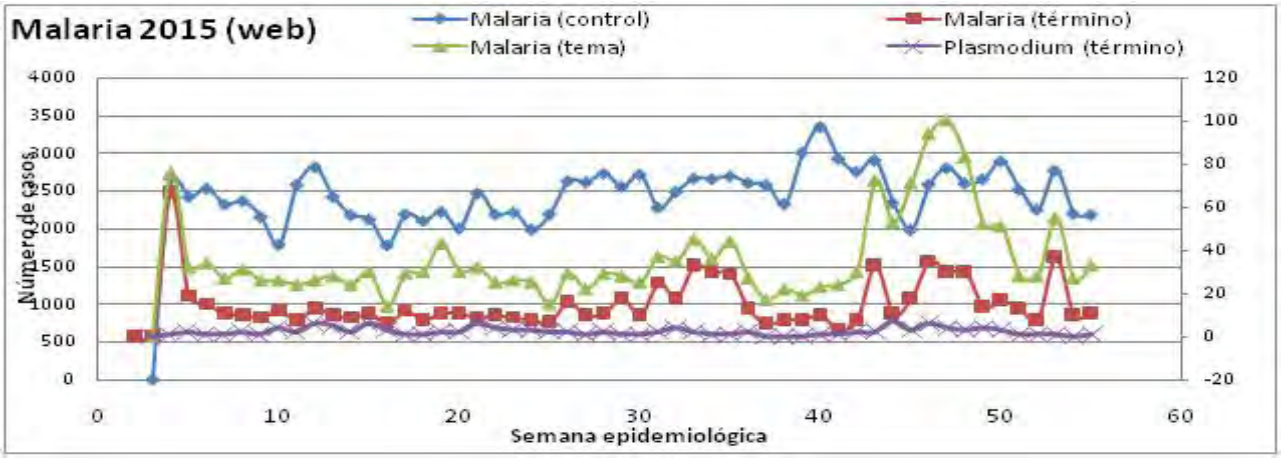

Figura 40. Comparación temporal de tendencias de Google y reporte de boletines epidemiológicos para malaria en 2015 (Web)

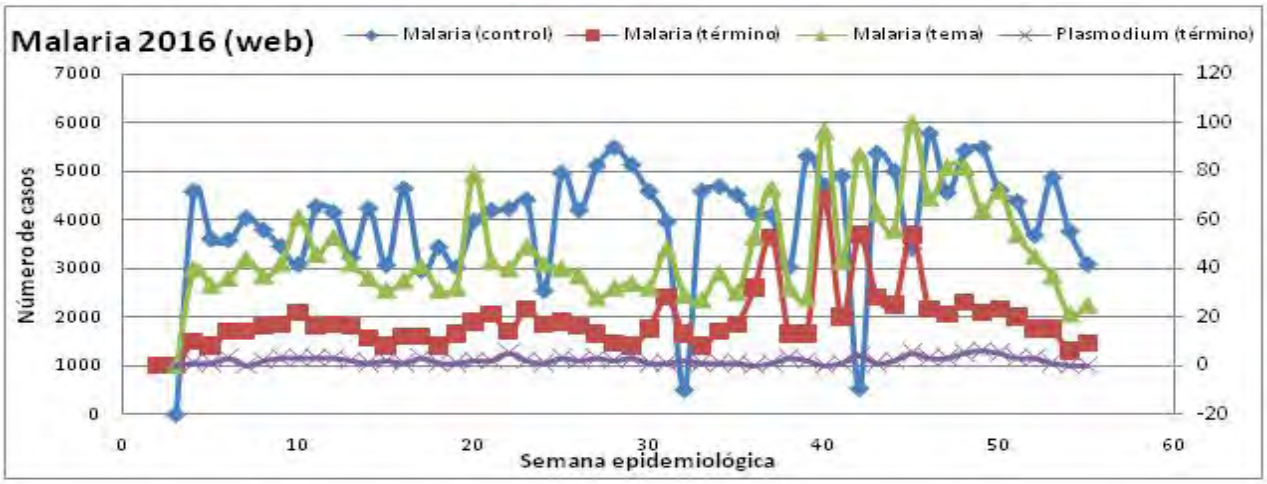

Figura 41. Comparación temporal de tendencias de Google y reporte de boletines epidemiológicos para malaria 2016 (Web)

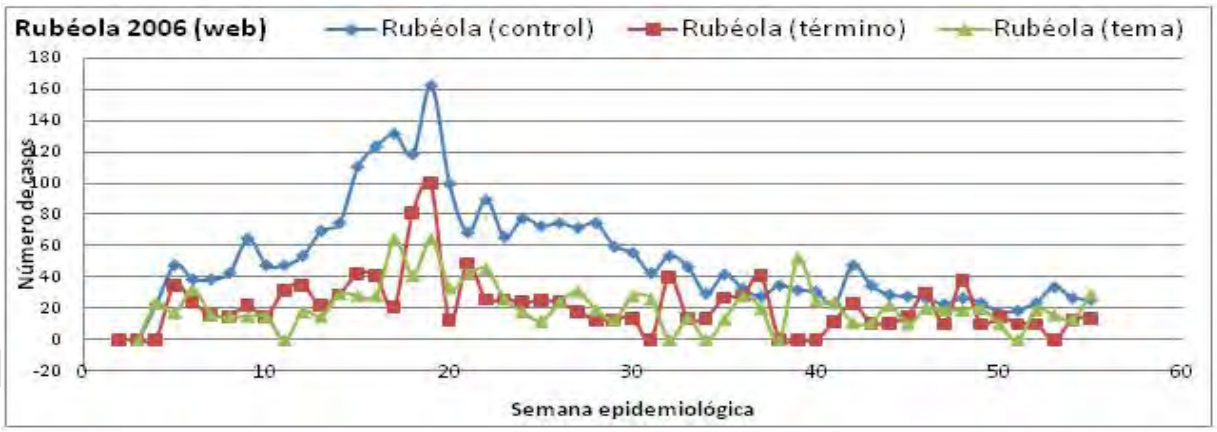

Figura 42. Comparación temporal de tendencias de Google y reporte de boletines epidemiológicos para rubéola en 2006 (Web)

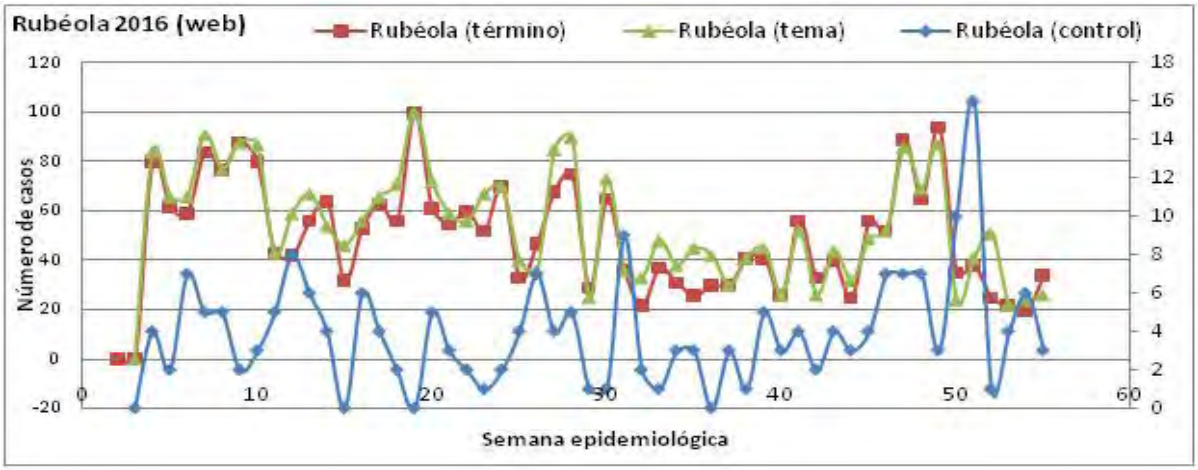

Figura 43. Comparación temporal de tendencias de Google y reporte de boletines epidemiológicos para rubéola en 2016 (Web)

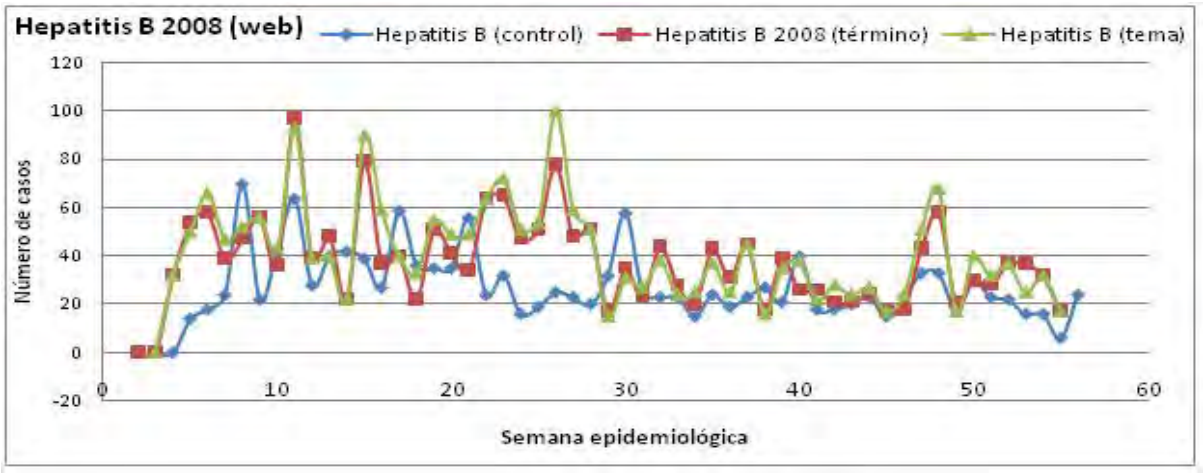

Figura 44. Comparación temporal de tendencias de Google y reporte de boletines epidemiológicos hepatitis B 2008 (Web)

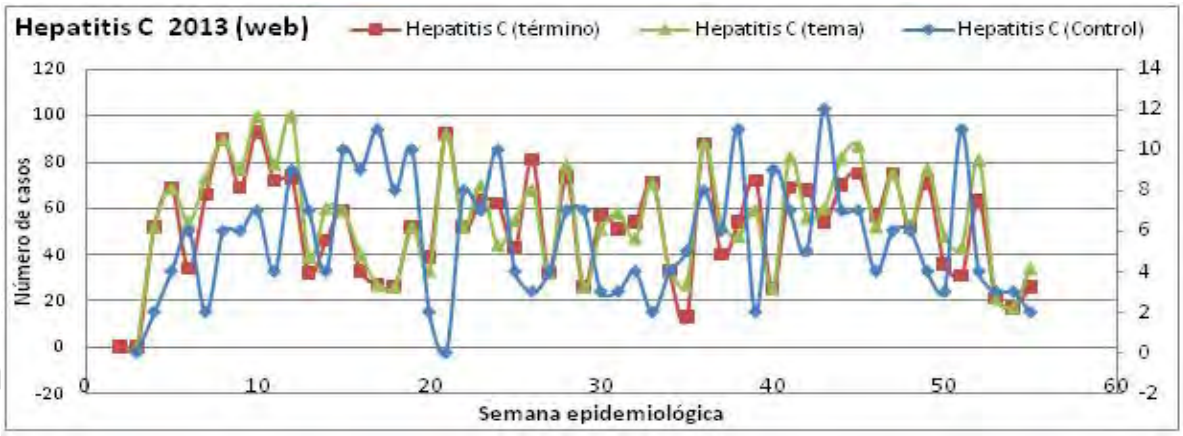

Figura 45. Comparación temporal de tendencias de Google y reporte de boletines epidemiológicos para hepatitis C 2013 (Web)

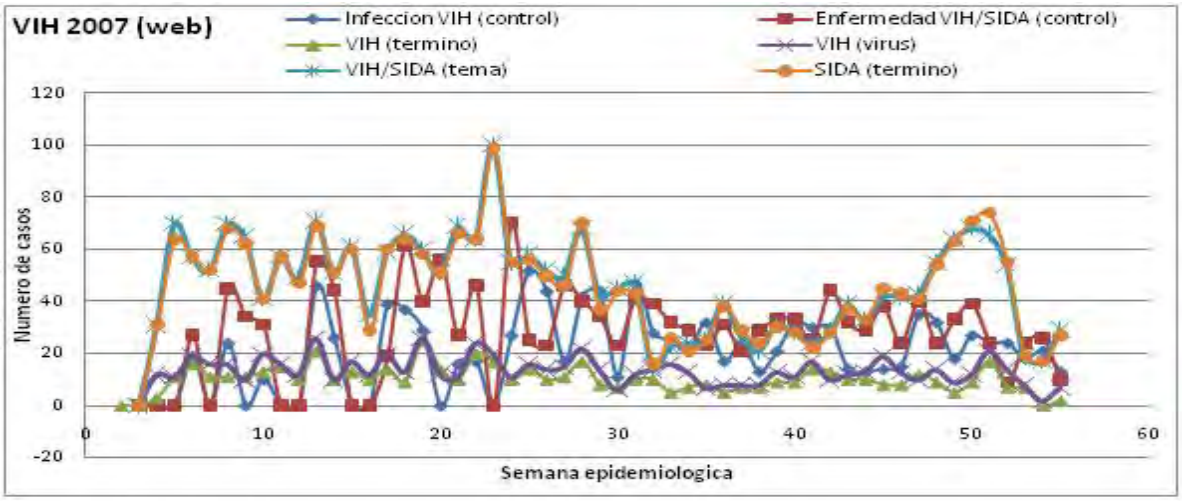

Figura 46. Comparación temporal de tendencias de Google y reporte de boletines epidemiológicos para VIH 20007 (Web)

Referencias bibliográficas

1. Principios de EPIDEMIOLOGÍA. Segunda Edición 1992. Una Introducción a la Epidemiología y la Bioestadística Aplicadas. Departamento de Salud y Recursos Humanos de los Estados Unidos de América, Centros para el Control y Prevención de Enfermedades (CDC), Oficina del Programa de Epidemiología, Oficina del Programa de Práctica de la Salud Pública. Atlanta, Georgia: Instituto Nacional de Salud, División de Publicaciones y Biblioteca, República de Colombia.

2. Salathé M, Bengtsson L, Bodnar TJ, Brewer DD, Brownstein JS, et al. (2012) Digital Epidemiology. PLoS Comput Biol 8(7): e1002616. doi: 10.1371/journal.pcbi.1002616

3. Fayyad M, Piatetsky-Shapiro G, Smyth P, (1996). From Data Mining to Knowledge Discovery in Databases. American Association for Artificial Intelligence, Menlo Park, CA, USA.

4. Rumi Chunara y col., (2012). Social and News Media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak, The American Society of Tropical Medicine and Hygiene

5. . Jiawei H, Jian P, Micheline K, (2012) Data Mining Concepts and Techniques – 3rd ed. Amsterdam, Holanda, Elsevier.

6. Avilán J, (2008). El Boletín Epidemiológico Semanal, Gaceta Médica de Caracas, Venezuela. Recuperado de http://www.scielo.org.ve/scielo.php?script=sci_arttext&pid=S0367-47622008000100001

7. Bracho C, (2016). El país tiene 57 semanas sin cifras epidemiológicas. El Panorama. Recuperado de http://www.panorama.com.ve/ciudad/El-pais-tiene-57-semanas-sin-cifras-epidemiologicas-20160207-0062.html

8. Bonita R, Beaglehole R, y Kjellstrõm T, (2008), Epidemiología básica. Segunda edición, Washington, D.C, USA: Organización Panamericana de la Salud.

9. Christakis NA, Fowler JH (2010) Social network Sensors for Early Detection of Contagious Outbreaks. PLoS ONE 5(9): e12948. doi: 10.1371/journal.pone.0012948.

10. Ayuda de tendencias de búsqueda, (s.f). Support.google.com. Recuperado de https://supportgoogle.com/trends/?hl=es#topic=6248052.

11. About | HealthMap, (s.f). Healthmap.org, Recuperado de http://www.healthmap.org/site/about.

12. Flu Prediction: About, (s.f). Flu-prediction.com Recuperado de http://www.flu-prediction.com/about.

13. Google Flu trends, (s.f). En.wikipedia.org. Recuperado de https://en.wikipedia.org/wiki/Google_Flu_Trends.

14. Chew C, Eysenbach G (2010) Pandemics in the Age of Twitter: Content Analysis of Tweets during the 2009 H1N1 Outbreak. PLoS ONE 5(11): e14118.doi: 10.1371/journal.pone.0014118

15. Xie Y, Chen Z, Cheng Y, Zhang K, (2013). Detecting and tracking disease outbreaks by mining social media data, Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence.

16. Chunara R, Bouton L, Ayers J, Brownstein J (2013). Assessing the Online Social Environment for Surveillance of Obesity Prevalence, PLOS ONE.

17. Jiawei H, Jian P, Micheline K, (2012) Data Mining Concepts and Techniques – 3rd ed. Amsterdam, Holánda, Elsevier.

18. Tendencias digitales 2916. Penetración y usos de internet en Venezuela 2016. Recuperado de http://tendenciasdigitales.com/penetracion-y-usos-de-internet-en-venezuela-2016/

19. Centro de prensa OMS, (2017). Nota descriptive: Dengue y dengue grave. Recuperado de http://www.who.int/mediacentre/factsheets/fs117/es/

20. Centro de prensa OMS, (2017). Nota descriptive: Enfermedades diarreicas. Recuperado de http://www.who.int/mediacentre/factsheets/fs330/es/

21. Ibáñez C., (2008). Epidemiología de la Difteria. Madrid Blogs Fundación para el conocimiento Madrid más. Recuperado de http://www.madrimasd.org/blogs/salud_publica/2008/09/12/100764

22. Fiebre (s.f). En.wikipedia.org. Recuperado de https://es.wikipedia.org/wiki/Fiebre

23. Centro de prensa OMS, (2017). Nota descriptiva: hepatitis A. Recuperado de http://www.who.int/mediacentre/factsheets/fs328/es/

24. Centro de prensa OMS, (2017). Nota descriptive: hepatitis B. Recuperado de http://www.who.int/mediacentre/factsheets/fs204/es/

25. Centro de prensa OMS, (2017). Nota descriptiva: hepatitis C. Recuperado de http://www.who.int/mediacentre/factsheets/fs164/es/

26. Centro de prensa OMS, (2017). Nota descriptive: Paludismo. Recuperado de http://www.who.int/mediacentre/factsheets/fs094/es/

27. Centro de prensa OMS, (2017). Nota descriptive: Rubéola. Recuperado de http://www.who.int/mediacentre/factsheets/fs367/es/

28. Centro de prensa OMS, (2017). Nota descriptive: Sarampión. Recuperado de http://who.int/mediacentre/factsheets/fs286/es/

29. Centro de prensa OMS, (2017). Nota descriptive: VIH/sida. Recuperado de http://www.who.int/mediacentre/factsheets/fs360/es/

30. "Data source: Google Trends (www.google.com/trends)."

31. Michelle T., (s.f). Curvas Epidémicas. Enfoque en Epidemiología de Campo, Volumen 1, número 5. North Carolina Center for Public Health Preparedness - The North Carolina Institute for Public Health.

32. Vallverdú J., (2011). “Sentiment analysis”: ¿Qué es, ¿cómo funciona, para qué sirve? Recuperado de: https://www.tecnonews.info/opiniones/sentiment_analysis_que_es_como_funciona_para_que_sirve

33. R (lenguaje de programación), (s.f). En.wikipedia.org. Recuperado de https://es.wikipedia.org/wiki/R_(lenguaje_de_programaci%C3%B3n)

34. Gentry J., (2016). Package ‘twitteR’. Repository CRAN.

35. Barbera P., (2017). Package ‘Rfacebook’. Repository CRAN.

36. MySQL, (s.f). En.wikipedia.org. Recuperado de https://es.wikipedia.org/wiki/MySQL

37. French J., (2014). Using R with MySQL Databases. Northwestern University Blog. Recuperado de http://www.jason-french.com/blog/2014/07/03/using-r-with-mysql-databases/

38. Apache HTTP Server, (s.f). En.wikipedia.org. Recuperado de https://en.wikipedia.org/wiki/Apache_HTTP_Server

39. phpMyAdmin, (s.f), En.wikipedia.org. Recuperado de https://es.wikipedia.org/wiki/PhpMyAdmin

40. PHP, (s.f), En.wikipedia.org. Recuperado de https://es.wikipedia.org/wiki/PHP

41. WAMP, (s.f), En.wikiperdia.org. Recuperado de https://es.wikipedia.org/wiki/WAMP

42. Soporte Minitab 18 (2017). Relaciones lineales, no lineales y monótonas. Recuperado de https://support.minitab.com/es-mx/minitab/18/help-and-how-to/statistics/basic-statistics/supporting-topics/basics/linear-nonlinear-and-monotonic-relationships/

43. Soporte Minitab 18. Interpretar los resultados clave para Correlación. Recuperado de https://support.minitab.com/es-mx/minitab/18/help-and-how-to/statistics/basic-statistics/how-to/correlation/interpret-the-results/key-results/

44. Witten G., Poulter G., (2006). Simulations of infectious diseases on networks / Computers in Biology and Medicine, Elsevier.

45. Charles-Smith LE, Reynolds TL, Cameron MA, Conway M, Lau EHY, Olsen JM, et al. (2015). Using Social Media for Actionable Disease Surveillance and Outbreak Management: A Systematic Literature Review. PLoS ONE 10(10): e0139701. doi: 10.1371/journal.pone.0139701

46. Stattner E., Vidot N., (s.f), Social Network Analysis in Epidemiology: Current Trends and Perspectives, LAMIA Laboratory, University of the French West Indies and Guiana FRANCE.

47. Verhulst, Pierre-François (1838). Notice sur la loi que la population poursuit dans son accroissement. Correspondance mathématique et physique

48. Zhao L, Chen J, Chen F, Wang W, Lu CT, Ramakrishnan N., (2015). SimNest: Media Nested Epidemic Simulation via Online Semi-supervised Deep Learning.Social

49. Organización Panamericana de la Salud / Organización Mundial de la Salud. Actualización Epidemiológica: Difteria. 22 de agosto de 2017, Washington, D.C. OPS/OMS. 2017 Organización Panamericana de la Salud • www.paho.org• © OPS/OMS, 2017

50. Organización Panamericana de la Salud / Organización Mundial de la Salud. Actualización Epidemiológica, Sarampión. 27 de octubre de 2017, Washington, D.C. OPS/OMS. 2017

51. Organización Panamericana de la Salud / Organización Mundial de la Salud. Alerta Epidemiológica: Aumento de casos de malaria, 15 de febrero de 2017, Washington, D.C. OPS/OMS. 2017

52. Herrera I., (2017). La malaria repuntó en Bolívar con 206.240 casos, El Nacional. Recuperado de http://www.el-nacional.com/noticias/salud/malaria-repunto-bolivar-con-206240-casos_210308

53. Ostos P., (2017). Casos de malaria se han incrementado en un 50% en el estado Bolívar, El Universal. Recuperado de http://www.eluniversal.com/noticias/venezuela/casos-malaria-han-incrementado-estado-bolivar_676329

54. Early detection, assessment and response to acute public health events: Implementation of Early Warning and Response with a focus on Event-Based Surveillance. Interim Version. World Health Organization, 2014, WHO/HSE/GCR/LYO/2014.4

55. Reportajes OMS, (2016). Las alertas tempranas de brotes de enfermedad ayudan a guiar la respuesta de la OMS en el nordeste de Nigeria. Recuperado de http://www.who.int/features/2016/early-warning-nigeria/es/

56. Ciencia ciudadana (s.f), Enwikipedia.org. Recuperado de https://es.wikipedia.org/wiki/Ciencia_ciudadana