



Problema 3

ANÁLISIS PREDICTIVO

[English version](#)

[Spanish version](#)

¿Cuáles son opciones innovadoras para analizar información disponible en las nuevas tecnologías y por tanto predecir y luego prevenir los brotes?

El enfoque – Los brotes son generalmente impredecibles y por lo tanto difíciles de prevenir.

Una corta descripción del problema:

Mecanismo de un brote

De acuerdo a la Organización Mundial de la Salud, “un brote de enfermedad es la ocurrencia de un número excesivo de casos de una enfermedad normalmente registrados en una comunidad, una zona geográfica o durante una estación dada. Un brote puede producirse en una zona geográfica restringida o extenderse a varios países”.¹ Por definición, se trata de acontecimientos difíciles de anticipar y, por ende, de prevenir con modelos analíticos estándar.

Para las Enfermedades Transmitidas por Mosquitos (ETM), el movimiento de las personas virémicas es más causativo en la propagación de la enfermedad y la creación de nuevos brotes que el movimiento de los mosquitos,² pero para que se produzca un nuevo brote en una zona geográfica, un cierto número de factores deben concurrir. Para algunas enfermedades, como el dengue, estos factores son conocidos,³ pero el virus del zika tiene muchas formas de transmisión y algunas aún no están claras,⁴ lo que lo vuelve más difícil de controlar. Actualmente, sin la existencia de una vacuna efectiva, la prevención de estas enfermedades depende de la eliminación y control de los vectores (en este caso el mosquito Aedes), lo que implica constantes e intensos esfuerzos en territorios extensos y puede demandar ingentes gastos financieros y recursos humanos.

Big Data (macrodatos o datos masivos)

El concepto se refiere a la recolección, agregación, almacenamiento, análisis y uso de datos a gran escala. La creciente literatura sobre el big data argumenta que este procedimiento genera valor pues crea transparencia; posibilita la experimentación para descubrir las necesidades, exponer la variabilidad y mejorar el desempeño; permite segmentar poblaciones para adaptar acciones; reemplaza/apoya la toma humana de decisiones por algoritmos automatizados; e innova modelos de negocios,

¹ [Disease Outbreaks](#), WHO Health Topics

² [“Disease Surveillance and Outbreak Prevention and Control”](#), Dengue Haemorrhagic Fever: Diagnosis, Treatment, Prevention and Control-Chapter 6, WHO, 1997

³ [Centers for Disease Control and Prevention. Dengue](#): Epidemiology

⁴ [“How does Zika virus spread? What we know so far”](#), Welch A., CBS News, 20 de febrero



productos y servicios.⁵ Antes de la era digital, la información tradicional se producía y recolectaba de fuentes de transacción de productos, registros financieros y canales de interacción, como los call centers y los puntos de venta. En nuestra era actual, transformadores avances tecnológicos han hecho posible recolectar y almacenar cantidades de datos antes inimaginables.

Dado lo intensiva en datos que es la industria de la salud, no sorprende que este sector haya sido un semillero de experimentación del big data. De hecho, desde hace varios años los expertos vienen abogando por el potencial del big data –y los conocimientos basados en evidencia que aporta– para el logro de beneficios masivos en el sector de la salud.⁶

La información que fluye a través del sector de la salud es a menudo generada por una de cuatro fuentes: proveedores de atención médica, pagadores públicos y privados, proveedores de servicios complementarios –desde farmacias hasta laboratorios– y consumidores de atención médica. Las herramientas de búsqueda de información médica, las aplicaciones que pueden ser usadas como podómetros para medir qué tan lejos camina una persona o servir para contar calorías, medir el ritmo cardíaco, controlar la presión sanguínea y los niveles glicémicos e incluso los riesgos ambientales de la zona donde vive la persona, están cambiando la forma en que los investigadores y médicos ven y comprenden las enfermedades, sus causas y consecuencias. Todo ello, más los registros médicos y de seguros, la información genética e incluso el uso de redes sociales, incrementan las bases de datos, siempre crecientes, sobre el estado de salud individual de las personas y del público en general.

Algunos de los usos paradigmáticos del big data para mejorar la atención en salud son:

- A. *Prevención de enfermedades mediante análisis de tendencias*: los análisis de tendencias son estudios longitudinales que pueden ofrecer una vista dinámica del estado de salud de una población. Datos sobre enfermedad o muerte y sobre comportamientos como el de fumar, se recolectan de una población a lo largo del tiempo para buscar tendencias y cambios. Estos datos son a menudo utilizados por los profesionales de la salud pública para asesorar las evaluaciones las necesidades de atención en salud, la planificación de servicios y el desarrollo de políticas. El uso de modelos predictivos con estos datos permite dar luces sobre las frecuencias futuras y las tasas de ocurrencia.
- B. *Perfeccionamiento de los tratamientos médicos*: los registros de salud electrónicos (EHR por sus siglas en inglés) optimizan y centralizan la información del paciente y generan datos que al despersonalizarse, agregarse y analizarse pueden proporcionar información valiosa. Juntar la información clínica de los laboratorios, la radiología y los registros médicos electrónicos así como la información del paciente, sus antecedentes y los factores sociales en una sola plataforma posibilita analizar y ampliar la precisión predictiva para perfeccionar los tratamientos. Mediante algoritmos es posible analizar,

⁵ [The GovLab Selected Readings on Big Data](#), The GovLab, enero 2014

⁶ Véase, por ejemplo: [“The Promise and Peril of Big Data”](#), Bollier D., The Aspen Institute, 2010



basándose en todos esos datos, con qué tratamiento se logrará el mejor resultado posible.

- C. *Innovación a nivel industrial*: los análisis de datos ya se han probado útiles tanto para medir la seguridad de los fármacos como para encontrarles nuevos usos. Por ejemplo, el laboratorio Atul Butte de la Facultad de Medicina de Stanford descubrió que el anticonvulsivo topiramato podía ser efectivo en el tratamiento del trastorno intestinal inflamatorio Crohn.⁷ Otras áreas de la industria de la salud se han estado desarrollando con el big data: el análisis de imágenes (Rayos X, resonancias magnéticas, etc.), dispositivos portátiles, herramientas de monitoreo casero, registros de salud electrónicos.
- D. *Orientación de la elección del paciente y de la toma de decisiones*: poner a disposición del público información gubernamental o de la industria a través de plataformas y herramientas puede contribuir a mejorar la toma de decisiones y la salud del paciente. En Uruguay, A tu servicio proporciona una diversidad de datos procesados y visualizados para informar al público sobre su selección de proveedores de salud.⁸ De modo similar, una herramienta que ofrece información sobre el precio de los medicamentos y el cambio de precios en Sudáfrica ayuda a los pacientes a tomar decisiones más informadas sobre por qué casa farmacéutica optar.
- E. *Focalización de las intervenciones en salud pública*: en Singapur y crecientemente en Paraguay, el gobierno recolecta información sobre los “puntos calientes” del dengue, elabora mapas y los pone a disposición de los ciudadanos y los interesados para aumentar la vigilancia en las zonas de probable transmisión y dirigir los esfuerzos tendientes a minimizar la reproducción del mosquito.⁹ Los esfuerzos basados en datos para focalizar los recursos y las acciones preventivas también entraron en juego durante la respuesta al ébola en África occidental, donde múltiples actores, incluidas organizaciones internacionales, gobiernos nacionales y ONG, compartieron información para ayudar a las diversas partes interesadas que operaban en el terreno a coordinar y asignar de modo efectivo los limitados recursos.¹⁰

Small Data

Un fenómeno relacionado basado en datos también se está consolidando en la atención en salud. Los *small data*, “las pistas digitales que las personas generan como un subproducto de las actividades diarias, tales como el envío de correos electrónicos o los ejercicios con pulseras de monitoreo (fitness trackers)”¹¹, ayuda a las personas a tomar decisiones más saludables como resultado de una mejor comprensión de su propia conducta y sus impactos en la salud. Los proyectos

⁷ [“King of the Mountain- Digging data for a healthier world”](#), Goldman B., Stanford Medicine, Summer 2012

⁸ [“Uruguay’s A Tu Servicio: Empowering citizens to make data-driven decisions on health-care”](#), Sangokoya D. et al., The GovLab- Open Data’s Impact, enero 2016

⁹ [“Singapore’s Dengue Cluster Map: Open Data for Public Health”](#), Young A. et al., The GovLab- Open Data’s Impact, enero 2016

¹⁰ [“Battling Ebola in Sierra Leone: Data sharing to improve crisis response”](#), Young A. and Verhulst S.G., The GovLab- Open Data’s Impact, enero 2016

¹¹ [Workshop on Small Data Experimentation: Opportunities, Challenges, and Infrastructure](#), New York, junio 2015



desarrollados en el Cornell Tech's Small Data Lab, por ejemplo, están permitiendo que la gente decida mejor acerca de la nutrición, la mitigación del dolor crónico sin medicación y la detección del cáncer.¹²

Análisis predictivo

Para extraer un valor significativo del big data es necesario procesar los datos y analizarlos. El análisis predictivo es la utilización de esos datos con algoritmos estadísticos para identificar la probabilidad de resultados futuros basados en información disponible. El análisis se basa en la captación y aprovechamiento de relaciones entre las variables explicativas y las variables predictivas de ocurrencias pasadas. El objetivo es ir más allá de saber qué ha sucedido para evaluar mejor lo que sucederá en el futuro.

Tener acceso a los datos y los modelos predictivos ayuda a responder preguntas como: a cuánta gente infectará potencialmente el brote; hasta dónde y cuán rápidamente se propagará la enfermedad; qué zonas y personas están en mayor riesgo, y cuándo lo están. Esto ayudaría a los gobiernos a comprender cómo usar de mejor modo los limitados recursos para prevenir y hacer frente a los brotes.

Tomemos Chicago como ejemplo. Se recolectaron datos meteorológicos, demográficos, de pruebas anteriores y sitios de fumigación, y luego se usó un algoritmo para predecir la presencia futura del virus del Nilo Occidental en el mosquito. Con esa información los funcionarios municipales de salud pudieron focalizar las fumigaciones con insecticida y otras medidas de contención en las zonas afectadas.¹³

Desafíos:

- *Incertidumbre:* desarrollar predicciones significativas requiere de profesionales con una sólida comprensión de qué tipo de incertidumbre está en juego en el proceso de análisis de datos. La procedencia de un flujo de big data puede estar plagada de incertidumbre. En algunos casos esa incertidumbre se entiende y puede tomarse en cuenta en el análisis, pero en otros casos los investigadores “no saben lo que no saben” y, en consecuencia, es casi imposible conducir el análisis de datos de un modo sensible a potenciales problemas.
- *Error de medición:* otra vez en relación con la procedencia de los datos por analizarse, incluso cuando son aparentemente de alta calidad, las diferencias contextuales en el proceso de recolección de datos pueden introducir errores y afectar negativamente el análisis. Por ejemplo, una sección de bajos ingresos en una ciudad podría ser más propensa a los brotes de zika, pero también reportar más lentamente los casos a través de la maquinaria burocrática del gobierno. Por otro lado, las zonas de ingresos más altos pueden ser menos propensas a los brotes, pero más inclinadas a reportar situaciones de zika. Tales diferencias en la generación de datos y los procesos de manejo podrían

¹² [Publications: The Small Data Lab](#)

¹³ [“How Predictive Analytics Can Combat the Zika Virus”](#), Mingle D., HIT Consultant, febrero 2016



llevar a evaluaciones imprecisas de las necesidades y condiciones sobre el terreno.

- *Variables predictivas:* de alguna manera, la principal ventaja del big data —por ejemplo la habilidad para analizar conjuntos de datos increíblemente grandes y diversos— también supone grandes desafíos para generar predicciones sólidas. Cuando el análisis se hace sobre conjuntos de datos que tienen un número considerable de observaciones (por ejemplo, casos de transmisión del zika), también hay exponencialmente más características o variables basadas en el tipo de análisis que podría realizarse (por ejemplo, el nivel de ingreso del paciente, la proximidad a una masa de agua, el tamaño de la familia, etc.). Con tantas variables predictivas potenciales en juego —y la probabilidad de positivos falsos basados en inferencias más correlativas que causales— establecer modelos predictivos útiles es un desafío.
- *Responsabilidad con los datos:* particularmente cuando las técnicas del big data potencian información personal identificable, se corre el riesgo, por ejemplo, de sacar a la luz la privacidad y seguridad. Los riesgos están presentes en cada fase del ciclo de vida de los datos: desde la recolección hasta el almacenamiento y el análisis que habrá de usarse, y es esencial disponer de una estrategia o marco inicial para enfrentar estos riesgos.¹⁴
- *Coordinación entre múltiples actores:* para utilizar en la práctica un análisis predictivo, se requiere de un plan de acción común entre los muchos actores —incluidos los gubernamentales— que serán responsables de implementar el plan. Ponerse de acuerdo en el flujo de trabajo, las audiencias objetivo y los procesos cotidianos es esencial para traducir las ideas en acción.
- *Capacitación y disposición:* del mismo modo, usar el big data y los análisis predictivos no es tarea común para mucha gente que podría tener a su cargo la implementación de un programa de análisis predictivo para la prevención del zika. El personal deberá ser rápidamente capacitado en las nuevas metodologías y estar preparado para implementarlas sobre la marcha. Esa capacitación podría requerir de inversión de tiempo y recursos.¹⁵

Enfoques interesantes de otros lugares y campos comparativos:

- Un estudio financiado por la NASA utiliza datos de satélite sobre la temperatura y la vegetación para ayudar a rastrear y predecir si el virus del Nilo Occidental se está propagando en Norteamérica. Los científicos y funcionarios de salud pública esperan poder usar un día mapas cercanos en tiempo real para focalizar recursos y evitar la enfermedad más eficientemente.¹⁶
- El proyecto de previsión del dengue (Dengue Forecasting Project): varios departamentos del gobierno federal de Estados Unidos (los departamentos de Salud y Servicios Humanos, Defensa, Comercio y el departamento de

¹⁴ Véase, por ejemplo: “[Building data responsibility into humanitarian action](#)”, Raymond N. et al., OCHA Policy and Studies Series, mayo 2016

¹⁵ “[4 Essential Lessons for Adopting Predictive Analytics in Healthcare](#)”, Crockett D., Health Catalyst, 2013

¹⁶ [Tracking and Predicting the Spread of Mosquito-borne Encephalitis Using NASA Data](#)”, NASA- Applied Sciences Program and Public Health, 2009



Seguridad Nacional) se han juntado, con el apoyo del Pandemic Prediction and Forecasting Science and Technology Interagency Working Group del Consejo Nacional de Ciencia y Tecnología, se han juntado para diseñar un proyecto de previsión de enfermedades infecciosas con el objetivo de galvanizar esfuerzos para predecir la epidemia de dengue.¹⁷

- El Laboratorio de Física Aplicada de la Universidad Johns Hopkins desarrolló un nuevo método, conocido como Predicting Infectious Disease Scalable Model/PRISM (Modelo Escalable de Predicción de Enfermedades Infecciosas), que relaciona datos clínicos, meteorológicos, climáticos y sociopolíticos para ayudar a los formuladores de política y planificadores a evaluar el riesgo futuro de una enfermedad que ocurre en una zona geográfica específica en un tiempo específico.¹⁸
- Una asociación entre Apple e IBM con dos compañías que colaboran con una plataforma de big data de salud permitirá a los usuarios de iPhone y de relojes Apple compartir datos en el servicio analítico de atención de salud en nube Watson Health de IBM. El objetivo es descubrir nuevos conocimientos médicos a partir del procesamiento de datos en tiempo real y de datos biométricos de millones de usuarios potenciales.¹⁹
- En un estudio piloto de casos de post cirugía colorrectal, la Clínica Mayo redujo las complicaciones a la mitad, acortó la estadía del paciente y ahorró USD 10 millones usando un programa que identificaba las mejores prácticas de atención y luego medía y monitoreaba esas mediciones en tiempo real.²⁰
- Intel está trabajando con la Universidad de California Santa Bárbara (UCSB) en la utilización de imágenes de satélite y tecnologías de mapeo de nieve para medir el nivel de nieve en las montañas de Sierra Nevada y hacer predicciones más precisas sobre la sequía, y mejorar los esfuerzos preparatorios.
- Se utilizaron líneas directas de emergencia de salud pública para un estudio en Pakistán, que revisaron 300.000 llamadas en dos años, analizaron los síntomas y direcciones de la gente y lograron despachar equipos de control del mosquito a los barrios correctos.²¹

Lecturas clave: <http://www.who.int/csr/resources/publications/dengue/060-66.pdf> // <http://www.jhuapl.edu/newscenter/pressreleases/2013/130320.asp> // [Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie or Die](#) // [The Legal And Ethical Concerns That Arise From Using Complex Predictive Analytics In Health Care](#) // [Forecasting Hotspots—A Predictive Analytics Approach](#)

Preguntas para las contrapartes gubernamentales:

A. ¿Tienen un plan de predicción de brotes?

¹⁷ [Dengue Forecasting](#), National Oceanic and Atmospheric Administration, USA

¹⁸ [Predicting Infectious Diseases Scalable Method \(PRISM\)](#), Applied Physics Laboratory (APL)

¹⁹ [IBM Watson](#)

²⁰ ["Mayo Clinic Saves \\$10M in Colorectal Surgery Proof-of-Concept"](#), Slabodkin G., Health Data Management, agosto 2014

²¹ [Researchers Show Phone Calls Can Forecast Dengue Fever Outbreaks](#), New York University Press Release, julio 2016.



SMARTER CROWDSOURCING
ZIKA

- B. ¿Cómo se creó?
- C. ¿Qué datos se usaron?
- D. ¿Considera que su institución tiene un enfoque basado en datos? ¿Tiene su personal competencia en materia de datos y capacidad para procesarlos?
- E. ¿El plan es nacional o regional?
- F. ¿Estaría interesado en invertir en un modelo predictivo e implementarlo?