

Redes sociales, aprendizaje automatizado y cladística en tiempos de COVID-19

Social networks, machine learning and cladistics in the time of COVID-19

Álvaro Díaz-Badillo^{1,2}, Carlos Ramírez-Pfeiffer² y Juan C. López-Alvarenga^{1,2*}

¹Departamento de Genética Humana, University of Texas Rio Grande Valley, Edinburg, Texas, EUA; ²Departamento de Investigación, Universidad México-Americana del Norte, Reynosa, Tamaulipas, México

*I falli commessi in un lontano passato sono molto più
facili da condannare che da correggere.*

Boccaccio

La mayor causa de mortalidad a lo largo de nuestra historia no han sido los conflictos bélicos, sino las enfermedades infecciosas. En el año 2009, la influenza H1N1 nos alertó sobre las pandemias, pero al parecer no logramos aprender mucho. La pandemia de COVID-19 nos muestra una vez más nuestra vulnerabilidad y nos invita a reflexionar sobre las herramientas que tenemos para contener pandemias similares en el futuro.

El virus causante de la COVID-19, agente infeccioso de la presente pandemia, pertenece a la gran familia de coronavirus que habitan en muchas especies de animales, incluidos camellos, vacas, gatos y murciélagos. En ocasiones, los coronavirus animales infectan a personas, como lo sucedido con el MERS-CoV (síndrome respiratorio de Medio Oriente por sus siglas en inglés) o con el SARS-CoV (síndrome respiratorio agudo grave por sus siglas en inglés), y ahora con este nuevo virus, llamado SARS-CoV-2.

El SARS-CoV-2 es un beta-coronavirus, como el MERS-CoV y el SARS-CoV, que tienen su origen en los murciélagos. La idea de que este quiróptero es el reservorio natural se apoya en las recientes secuencias genéticas del virus, obtenidas de pacientes de los Estados Unidos y que son muy similares a las obtenidas en China, cuyo cuadro clínico de COVID-19

se confunde con un amplio espectro de enfermedades con síntomas que varían de leves a graves. En una publicación reciente informaron enfermedad grave en el 16% de las personas infectadas en China¹, y en un informe semanal de morbilidad y mortalidad de los Centers for Disease Control and Prevention se menciona que el 80% de las muertes fueron entre adultos mayores de 65 años y que hubo una mayor frecuencia de casos graves en personas mayores de 85 años^{2,3}. Las personas con afecciones médicas subyacentes graves (afecciones cardíacas graves, enfermedades pulmonares crónicas y diabetes) también parecen tener un mayor riesgo de desarrollar COVID-19 grave.

Modelos naturales que se repiten en la historia

La «peste negra» durante la Edad Media (1347 y 1351) tuvo un efecto devastador sobre la cuarta parte de la población europea. En la historia se han registrado varias epidemias que han tenido un considerable impacto emocional en la población, pero en la actualidad este efecto emocional se ve exacerbado por los medios de comunicación masivos, que han comparado esta epidemia con «el fin del mundo», o que publican noticias de hospitales sobrepasados en su capacidad y el temor a la escasez de alimentos⁴. Los intentos para entender y predecir

Correspondencia:

*Juan C. López-Alvarenga

E-REBL 122

University Dr CP 78539

Edinburg, Texas, USA

E-mail: juan.lopezalvarenga@utrgv.edu

Fecha de recepción: 18-04-2020

Fecha de aceptación: 22-04-2020

DOI: 10.24875/CIRU.M20000065

Cir Cir. 2020;88(4):395-398

Contents available at PubMed

www.cirugiaycirujanos.com

0009-7411/© 2020 Academia Mexicana de Cirugía. Publicado por Permayer. Este es un artículo *open access* bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

las epidemias ha originado métodos disímiles para la obtención de datos y la generación de modelos de predicción. En el siglo xix, Sir Ronald Ross ya intentó hacer modelos lineales simples, y en el siglo pasado Kermack y McKendrick⁵ utilizaron ecuaciones diferenciales no lineales con grupos de individuos susceptibles, infectados y recuperados (funciones SIR), que mostraron predicciones más cercanas a la realidad.

La experiencia con pasadas epidemias, como la de influenza H1N1 y las de SARS y MERS, pueden resultar de gran ayuda para entender y manejar la COVID-19. La utilización de herramientas de bioinformática y de técnicas de aprendizaje automatizado (*machine learning*) puede predecir el diagnóstico de COVID-19 empleando la edad, los signos vitales y los síntomas clínicos. Otros modelos emplean resultados de gabinete y de laboratorio.

Muchos de los modelos diagnósticos y pronósticos que se han publicado no son fiables. Un problema real es que algunos modelos apresurados incluyen pacientes que pueden no ser representativos de la población en la que se quiere predecir, y por ello se han desarrollado guías metodológicas, como el *Reporte Transparente de Modelos de Predicción Multivariado para el Pronóstico o Diagnóstico de Individuos* (TRIPOD por sus siglas en inglés)⁶. En un metaanálisis⁷ con modelos predictivos se halló que la mortalidad predicha varió entre el 8 y el 59%; esta acotación tan amplia se debe a los sesgos introducidos al modelo. Otro problema que se presenta en estos estudios es la validación de los modelos, ya que apenas se cuenta con estadísticas de pacientes que se están recuperando y además hay retrasos en el diagnóstico de casos confirmados y en los reportes de casos fatales. Ante este problema, algunos investigadores consideraron los datos obtenidos de otras enfermedades infecciosas respiratorias para validar sus resultados. Cuando se introducen al modelo algunas condiciones estocásticas para acercarlo a la realidad, los investigadores se enfrentan con la falta de datos para validación.

La bioinformática como herramienta para la predicción de la infección

Cientos de equipos de investigación en todo el mundo están combinando sus esfuerzos para recopilar datos y desarrollar soluciones para esta epidemia de COVID-19. La bioinformática debería contribuir, con un aprendizaje automatizado, a identificar quiénes están

en mayor riesgo de contagio, diagnosticar pacientes de manera oportuna, desarrollar medicamentos más eficientes y de manera más rápida, predecir la propagación de la enfermedad y definir los grados de virulencia. Además, ayudaría a comprender mejor la biología de los virus, mapear los orígenes de donde estos provienen y predecir las próximas pandemias.

Como se puede ver en la literatura científica, el uso de la bioinformática en la actualidad ofrece herramientas que permiten seguir la trayectoria del virus desde sus posibles orígenes en el murciélago. Desde los estadios tempranos de la epidemia en Wuhan, China, se hizo la secuenciación del genoma y se encontraron siete dominios de proteínas conservados que dieron soporte a que este virus utiliza el mismo receptor de la enzima convertidora de angiotensina (ACE2)⁸ que la familia SARS-CoV. Además, la secuenciación en nueve pacientes de los primeros casos mostró similitud en el 99.98%, incluyendo la conservación del receptor ACE2⁹. Los estudios de variantes genéticas de ACE2 en humanos se asocian a alteraciones cardiovasculares, hipertensión y diabetes; sin embargo, los cambios estructurales, especialmente sobre los subdominios de la peptidasa M2, no parecen producir modificaciones importantes en la estructura. Los alelos rs73635825 (S19P) y rs143936283 (E329G) muestran baja afinidad por las proteínas *spike* del SARS-CoV2¹⁰. Esto va acorde con una reciente descripción de agregación familiar de gravedad de la COVID-19 que mostró una disminución de los linfocitos CD4+ con un aumento de la relación del conteo de neutrófilos/linfocitos¹¹. La cladística del virus sugiere que existen variantes en las bases nitrogenadas según las diferentes zonas geográficas, y que algunas de estas variantes se encuentran en las proteínas encargadas de la replicación. Así, la combinación de los genomas del virus y las variantes del hospedero pueden ser funciones que expliquen en parte la susceptibilidad.

El mejor abordaje que considera las variantes genéticas y su interacción con el ambiente para entender el comportamiento de esta pandemia es la cladística bajo la luz de la teoría de la evolución. La cladística tiene importancia por su posible predicción de los efectos de una segunda oleada de infecciones por SARS-CoV2 e incluso para futuras infecciones. Además, clasifica organismos y predice lo que aún no se ha observado. La cladística es privilegiada, ya que aparte de producir hipótesis sobre las relaciones de los organismos, también predice sus propiedades. Esto puede ser especialmente importante para las compañías farmacéuticas o de laboratorio

interesadas en mejorar los métodos diagnósticos, el aislamiento de los agentes o el desarrollo de tratamientos. De esta forma, aprendiendo de la presente pandemia, tenemos la obligación histórica de destinar más recursos y promover la investigación para prepararnos mejor para la próximos resurgimientos o las nuevas epidemias infecciosas.

Búsqueda de nuevos modelos integrativos

Bernoulli inició el modelado matemático de enfermedades infecciosas en 1760, pero el trabajo de Kermack y McKendrick⁵ sistematizó el marco de modelado.

El modelado puede comenzar con preguntas básicas que pueden investigarse: ¿qué condiciones son necesarias para que ocurra una epidemia?, ¿qué fracción de una población se infectará en una epidemia?, ¿puede la vacunación parcial en una población proteger contra el agente infeccioso? La validación del modelo con nuevas preguntas ayuda a acercarnos a la mejor predicción.

Falta ahora considerar las consecuencias de esta epidemia, integrando lo que vamos conociendo en la marcha de la investigación. Se mantiene el aforismo de Box: «Todos los modelos están equivocados, pero algunos modelos son útiles»¹². Las herramientas de modelado deben considerar la biología combinada con los efectos sociales, económicos y culturales, que han sido modificados por las decisiones de diferentes gobiernos en el mundo, y que han afectado el flujo comercial, las relaciones interpersonales y la ecología del ambiente.

Un modelado integrativo debe analizar los impactos sociales, culturales y económicos que pueden derivar en pérdidas aseguradas significativas. Las compañías de seguros de salud y vida, y otras empresas relacionadas con la salud, deben evaluar el riesgo de la mortalidad y la morbilidad derivadas de la propagación de este tipo de enfermedades infecciosas con el fin de determinar el capital de riesgo necesario para escenarios extremos, ya que las pandemias son inevitables hasta ahora.

Como se puede apreciar, a partir de otros escenarios (por ejemplo, la influenza H1N1 en 2009), los modelos analíticos y estadísticos son cruciales para comprender y manejar el riesgo de enfermedades infecciosas. La pandemia de 2009 fue casi un fracaso para el mundo y existe el peligro de asumir que las futuras pandemias de influenza serán tan «benignas» como aquella. El tiempo de aparición y la virulencia

de la próxima pandemia son inciertos, y además hay otras enfermedades infecciosas emergentes y reemergentes que comienzan a ocurrir (sarampión, dengue y hantavirus, entre otras), o que ya hemos visto en las últimas décadas, previamente desconocidas o poco caracterizadas, como el SARS, la infección por el virus de la inmunodeficiencia humana y la fiebre hemorrágica por el virus Ébola.

La desinformación en el siglo de la información

Las redes sociales son en la actualidad la primera línea de información sobre la pandemia de COVID-19. Aunque las plataformas digitales actualizan constantemente los datos de casos y fatalidades, también alimentan una ola de información errónea sobre el coronavirus, en la medida en que los usuarios se centran en la popularidad de lo que publican y no en la precisión.

Estudios recientes indican que la mayoría de las personas en todos los países del mundo dicen que los medios de comunicación les han ayudado a comprender la crisis y lo que pueden hacer. Sin embargo, aproximadamente una de cada tres personas también dice que los medios de comunicación han exagerado la pandemia. La regulación de la veracidad de la información de la pandemia ha hecho que se den a conocer las fuentes para decidir la confiabilidad de las noticias y las páginas de eventos, además de la prohibición de ciertos anuncios durante la pandemia.

Parece que el virus más peligroso del siglo xxi son las *fake news*, o contenido pseudoperiodístico difundido a través de portales de noticias, prensa escrita, radio, televisión y redes sociales, cuyo objetivo es la desinformación. Y al final del día, este fenómeno solo se puede contener con buenas noticias, reales, fiables y contrastadas por los modelos integrativos.

En conclusión, los esfuerzos deben encaminarse a la búsqueda de modelos integrativos que proporcionen un marco cuantitativo para evaluar el riesgo y tomar decisiones «informadas». Para ello debemos tomar en cuenta las lecciones aprendidas de la pandemia de influenza H1N1 de 2009, y las que aprenderemos de la de COVID-19 que actualmente nos afecta (abril de 2020). Esto implica mejorar la forma de obtención de datos en nuestra población, considerando el desfase que hay con la confirmación de diagnósticos, y hacer esfuerzos para evitar la propagación de

información errónea en las redes sociales. Además, son necesarios foros multidisciplinarios en los que se discutan las estrategias para futuras epidemias. Si no contamos con esto, estaremos haciendo cien cuentos picarescos al estilo del Decamerón.

Bibliografía

1. Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72314 cases from the Chinese Center for Disease Control and Prevention. *JAMA*. 2020. Feb 24. doi: 10.1001/jama.2020.2648. [Epub ahead of print]
2. Centers for Disease Control and Prevention. Coronavirus Disease 2019. CDC; 2019. Disponible en: <https://www.cdc.gov/coronavirus/2019-ncov/hcp/index.html>.
3. World Health Organization. Coronavirus disease (COVID-2019) situation reports. WHO; 2020. Disponible en: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/>.
4. Lima CKT, Carvalho PMM, Lima I, Nunes JVAO, Saraiva JS, de Souza RI, et al. The emotional impact of Coronavirus 2019-nCoV (new Coronavirus disease). *Psychiatry Res* 2020;287:112915.
5. Kermack WO, McKendrick AG. Contributions to the mathematical theory of epidemics - I. 1927. *Bull Math Biol*. 1991;53:33-55.
6. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med*. 2015; 162:55-63.
7. Wynants L, Van Calster B, Bonten MMJ, Collins GS, Debray TPA, De Vos M, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ*. 2020; 369:m1328.
8. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;579:270-3.
9. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet*. 2020;395:565-74.
10. Hussain M, Jabeen N, Raza F, Shabbir S, Baig AA, Amanullah A, et al. Structural variations in human ACE2 may influence its binding with SARS-CoV-2 spike protein. *J Med Virol*. 2020 Apr 6. doi: 10.1002/jmv.25832. [Epub ahead of print]
11. Xia XY, Wu J, Liu HL, Xia H, Jia B, Huang WX. Epidemiological and initial clinical characteristics of patients with family aggregation of COVID-19. *J Clin Virol*. 2020;127:104360.
12. Box GEP. Empirical model-building and response surfaces. New York: John Wiley & Sons; 1986.