

## Los casos clínicos y la historia clínica electrónica desde la perspectiva de la Epidemiología: *Big Data*, procesamiento de lenguaje natural y otras metodologías

La Medicina Interna y la Epidemiología no son disciplinas contrapuestas, todo lo contrario. Las estimaciones más precisas de las tasas epidemiológicas de incidencia, prevalencia o mortalidad de muchas enfermedades proceden de estudios basados en una gran colección de registros hospitalarios de pacientes. Las historias clínicas están indexadas originalmente para su recuperación por diagnóstico para su uso clínico en pacientes individuales. Sin embargo, la cantidad y calidad de información recogida en estas historias clínicas puede utilizarse para determinar la historia natural de enfermedades agudas y crónicas, su interrelación, así como la eficacia y seguridad de sus tratamientos.

La investigación en Medicina Interna (y en otras especialidades) se inicia cuando existe un problema de salud por resolver. El investigador (internista, biólogo, médico de Atención Primaria, diplomado en Enfermería, epidemiólogo, fisioterapeuta, gestor de servicios sanitarios, político u otros) identifica una necesidad o una pregunta, que puede ser bien de carácter práctico, interés local y aplicabilidad inmediata, bien de interés puramente científico, ámbito internacional, y aplicabilidad a largo plazo.

La Epidemiología, que se define habitualmente como “el estudio de la distribución y los determinantes de los sucesos y estados de salud en las poblaciones y la aplicación de este estudio para el control de los problemas de salud”, proporciona un método para llevar a cabo la investigación de problemas de salud. Se basa en principios estadísticos y en una metodología que incluye:

- Planteamiento de la pregunta de investigación.
- Selección del diseño.
- Selección de las variables a medir y los instrumentos de medida.
- Planteamiento de los análisis estadísticos a realizar.
- Consideración de los posibles sesgos que puede tener el estudio.

En algunos casos, la Epidemiología pretende describir la distribución de las frecuencias de enfermedades o de sus determinantes, clasificando los hechos según tiempo, lugar y persona (Epidemiología descriptiva). En otros casos, se pretende encontrar las causas o factores que se asocian a la aparición de enfermedad o a su evolución, planteando preguntas como “¿qué?, ¿quién?, ¿cuándo? y ¿dónde?” para responder con “¿cómo? y ¿por qué?” (Epidemiología analítica). Es por ello que, aunque los estudios epidemiológicos trabajen con grupos de personas y se basen en poblaciones, sus resultados permiten viajar tanto desde la población hasta el paciente, como desde la población hasta la molécula (componente traslacional de la Epidemiología).

Dado que las enfermedades infecciosas o crónicas que manejan los internistas representan una importante carga para la salud pública en cuanto a morbimortalidad, es necesario contar con el enfoque epidemiológico para estudiar la prevalencia y los determinantes de estas enfermedades, así como su evolución y control.

El uso y la funcionalidad de los registros de salud electrónicos, también llamados historia clínica electrónica (HCE), se han incrementado exponencialmente en la última década. Aunque el propósito principal de la HCE es clínico, los investigadores las han utilizado también para realizar investigaciones epidemiológicas, que van desde estudios transversales dentro de un hospital determinado, hasta estudios longitudinales o multicéntricos en pacientes distribuidos geográficamente a nivel nacional o internacional. Su uso en la investigación de salud de la población complementa y a veces supera las limitaciones de los métodos epidemiológicos tradicionales. Sus aplicaciones de investigación se benefician de los grandes tamaños de muestra y las poblaciones de pacientes generalizables que ofrecen la HCE. Éstos han incluido la reevaluación de hallazgos anteriores, una gama más amplia de enfermedades y comorbilidades, la evaluación en subgrupos y por fenotipos, epidemiología ambiental y social, condiciones estigmatizadas, modelos predictivos y evaluación de experimentos naturales. Aunque los estudios que utilizan métodos de recopilación de datos primarios pueden tener datos inicialmente más confiables y una mejor retención de la población, a priori los estudios basados en la HCE son menos costosos y requieren menor tiempo para completarse. La futura epidemiología de la HCE con una recolección mejorada de mediciones sociales y de hábitos y comportamientos, vinculación con registros vitales e integración de tecnologías emergentes como la detección probabilística personal/individualizada, podría ayudar a mejorar la atención clínica y la salud de la población.



**Figura 1.** Las herramientas de *Big Data* e inteligencia artificial pueden ya ser utilizadas por los internistas en la toma de decisiones clínicas

Los investigadores que tengan la capacidad para integrar información de diferentes disciplinas, y que también puedan interactuar con otros investigadores básicos y clínicos, junto con matemáticos y bioinformáticos, tendrán ventajas añadidas. Sólo se necesita una receta para superar con éxito estos nuevos desafíos: monitorizar las tendencias mediante la Epidemiología llamada de “suela

de zapato<sup>2,3</sup>, es decir la Epidemiología clásica, complementada por *Big Data* y otras nuevas tecnologías<sup>4</sup>, como la inteligencia artificial. La implementación de herramientas como las descritas recientemente será de gran ayuda<sup>5-7</sup>.

Esta nueva frontera en la investigación médica será probablemente más eficiente y potencialmente más precisa que el estándar de oro actual de los estudios de revisión manual de historias. Sin embargo, la HCE tal como existe actualmente tiene limitaciones significativas: faltan datos importantes o sólo se capturan en texto libre o en documentos PDF. El uso de información adicional en imágenes de radiología, bioquímica, estudios de función de órganos y sistemas u otras variables fisiológicas tampoco es directo ni sencillo<sup>8,9</sup>. Superar estos desafíos, conjugando algoritmos de in-

teligencia artificial y aprendizaje de máquinas (*machine learning*), con el reconocimiento del texto libre con la tecnología de procesamiento de lenguaje natural (NLP, en sus siglas en inglés) dará una verdadera "nueva vida" a los datos recogidos en las HCE, que el sistema de atención médica recopila por defecto, y las oportunidades que surgen de su reutilización pueden ser más que atractivas<sup>10</sup>.

Su explotación, combinada con datos ecológicos sobre meteorología, contaminación atmosférica o de aguas, o más inverosímiles como información fiscal o administrativa, y los esfuerzos empleados a favor de los cambios en los hábitos de vida saludables, pueden ayudar a identificar estrategias de progreso de la calidad de vida y la salud de nuestra sociedad<sup>11-12</sup>.

**Joan B. Soriano**

*Profesor Asociado de Medicina. Universidad Autónoma de Madrid. España  
Hospital Universitario de La Princesa. Madrid. España*

**Citar como:** Soriano JB. Los casos clínicos y la historia clínica electrónica desde la perspectiva de la Epidemiología: Big Data, procesamiento de lenguaje natural y otras metodologías. *Rev Esp Casos Clin Med Intern (RECCMI)*. 2019 (Ago); 4(2): 43-44. doi: 10.32818/reccmi.a4n2a1.

**Cite this as:** Soriano JB. Clinical cases and electronic medical history from the perspective of Epidemiology: Big Data, natural language processing and other methodologies. *Rev Esp Casos Clin Med Intern (RECCMI)*. 2019 (Ago); 4(2): 43-44. doi: 10.32818/reccmi.a4n2a1.

[jbsoriano2@gmail.com](mailto:jbsoriano2@gmail.com)

## Bibliografía

1. Last JM. Diccionario de Epidemiología, 4.ª edición. IEA, 2000.
2. Koo D, Thacker SB. In snow's footsteps: commentary on shoe-leather and applied Epidemiology. *Am J Epidemiol*. 2010; 172: 737-739.
3. Tong SY. Genomic polish for shoe-leather epidemiology. *Nat Rev Microbiol*. 2013; 11(1): 8.
4. Gené Badia J, Gallo de Puelles P, de Lecuona I. Big data y seguridad de la información. *Aten Primaria*. 2018; 50(1): 3-5.
5. Johnston IG, Hoffmann T, Greenbury SF, Cominetti O, Jallow M, Kwiatkowski D, Barahona M, Jones NS, Casals-Pascual C. Precision identification of high-risk phenotypes and progression pathways in severe malaria without requiring longitudinal data. *NPJ Digit Med*. 2019 Jul 10; 2: 63.
6. Balicer RD, Luengo-Oroz M, Cohen-Stavi C, Loyola E, Mantingh F, Romanoff L, Galea G. Using big data for non-communicable disease surveillance. *Lancet Diabetes Endocrinol*. 2018; 6(8): 595-598.
7. Burgner D, Jamieson SE, Blackwell JM. Genetic susceptibility to infectious diseases: big is beautiful, but will bigger be even better? *Lancet Infect Dis*. 2006; 6(10): 653-663.
8. Divita G, Carter M, Redd A, Zeng Q, Gupta K, Trautner B, Samore M, Gundlapalli A. Scaling-up NLP Pipelines to process large corpora of clinical notes. *Methods Inf Med*. 2015; 54(6): 548-552.
9. MacRae J, Darlow B, McBain L, Jones O, Stubbe M, Turner N, Dowell A. Accessing primary care Big Data: the development of a software algorithm to explore the rich content of consultation records. *BMJ Open*. 2015; 5(8): e008160.
10. Hernández-Medrano I, Carrasco G. El profesional de la salud ante el mundo del Big Data. *Rev Calid Asist*. 2016; 31(5): 250-253.
11. Biermann F, Abbott K, Andresen S, Bäckstrand K, Bernstein S, Betsill MM, et al. Science and government. Navigating the anthropocene: improving Earth system governance. *Science*. 2012; 335: 1306-1307.
12. Berger ML, Curtis MD, Smith G, Harnett J, Abernethy AP. Opportunities and challenges in leveraging electronic health record data in oncology. *Future Oncol*. 2016; 12(10): 1261-12