

## **Evolución histórica de las técnicas estadísticas y las metodologías para el estudio de la causalidad en ciencias médicas**

Historical changes of the statistical techniques and the methodologies for the study of the causality in medical sciences

Dra. Nelsa María Sagaró del Campo<sup>1\*</sup> <https://orcid.org/0000-0002-1964-8830>

Dra. C. Larisa Zamora Matamoros<sup>2</sup> <https://orcid.org/0000-0003-2210-0806>

<sup>1</sup>Universidad de Ciencias Médicas. Santiago de Cuba, Cuba.

<sup>2</sup>Universidad de Oriente. Santiago de Cuba, Cuba.

\*Autor para la correspondencia. Correo electrónico: [nelsa@infomed.sld.cu](mailto:nelsa@infomed.sld.cu)

### **RESUMEN**

Las técnicas estadísticas constituyen las herramientas matemáticas que permiten el análisis causal de los problemas de salud; de ahí la importancia de conocer cómo ha sido su desarrollo a lo largo del tiempo, por lo cual el objetivo de este trabajo fue elaborar una síntesis de la evolución histórica de las técnicas y metodologías empleadas en la investigación médica en relación con los modelos de causalidad que han primado en cada momento histórico. A tales efectos, se realizó una amplia revisión documental y, para una mejor comprensión, se dividió la evolución en cinco etapas, según el desarrollo alcanzado en las técnicas estadísticas. Asimismo, se describen los principales hechos y las características fundamentales de cada época y se destaca la regresión logística binaria como la técnica más empleada.

**Palabras clave:** historia de la estadística; causalidad en medicina; modelos de causalidad; técnicas estadísticas; análisis estadístico implicativo.

## **ABSTRACT**

The statistical techniques constitute the mathematical tools that allow the causal analysis of the health problems; that is why it is important to know about their development through time, reason why the objective of this work was to elaborate a synthesis of the historical course of the techniques and methodologies used in the medical investigation in connection with the causality models that have prevailed in each historical moment. To such effects, a wide documental review was carried out and, for a better understanding, the changes were divided in five stages, according to the development reached in the statistical techniques. Also, the main facts and the fundamental characteristics of each stage are described and the binary logistical regression is highlighted as the most used technique.

**Key words:** history of statistics; causality in medicine; causality models; statistical techniques; statistical implicative analysis.

Recibido: 14/01/2019

Aprobado: 29/03/2019

## **Introducción**

En las ciencias médicas se explica el llamado proceso salud-enfermedad con el apoyo de los paradigmas de la causalidad. Esta noción ha sido ampliamente abordada, dando explicaciones a los fenómenos de salud y creando diferentes modelos con los que se ha pretendido estructurar lo no observable de acuerdo con los acontecimientos científicos de la época, el contexto y las experiencias de cada autor.<sup>(1)</sup>

El concepto de causa tiene un significado especial en las ciencias médicas por su naturaleza probabilística. Es por ello que la inferencia estadística ha desempeñado una función protagónica en las investigaciones en cuanto a la identificación de posibles relaciones causales.

Las técnicas estadísticas aplicadas en las ciencias médicas como herramientas matemáticas para el análisis causal han presentado notables cambios con el transcurso del tiempo y el devenir de nuevos adelantos científico-técnicos; por ello, cuando se quiere investigar acerca de estas, constituye una premisa indispensable el estudio de su desarrollo histórico, para una mejor comprensión e interpretación de su esencia como objeto de investigación.

La bioestadística, área de la estadística aplicada al análisis de datos provenientes de las ciencias de la vida, brinda no solo las técnicas para el análisis de los datos, sino también la metodología para un buen diseño de la investigación, el tipo de estudio pertinente en cada caso, el diseño de muestreo a emplear, las precauciones para evitar el sesgo y dar mayor validez a los resultados, así como los métodos para interpretar y presentar dichos resultados.

De tal manera, los estudios de causalidad en medicina han ido evolucionando y así mismo las técnicas de análisis estadístico asociadas a ellos, desde la simple descripción de las condiciones de salud mediante técnicas univariadas, como las medidas de frecuencia absolutas y relativas, de comparación y de tendencia, y la explicación de las causas de la enfermedad con pruebas de hipótesis bivariadas y multivariadas, como las técnicas de regresión para identificar factores de riesgo o pronósticos; hasta la predicción, con la creación de modelos matemáticos y técnicas de simulación, así como muchas otras técnicas heredadas de la econometría, para ir eliminando posibles sesgos y llegar a resultados válidos.

El objetivo de este trabajo fue elaborar una síntesis de la evolución histórica de las técnicas y metodologías que han ido surgiendo y empleándose en la investigación médica en relación con los modelos de causalidad que han primado en cada momento histórico. Para ello, se realizó una amplia revisión documental y luego se determinó y seleccionó la información, teniendo en cuenta la claridad y actualización, así como la unificación y separación de aquella que era reiterada en más de una fuente, de forma interpretativa y evaluativa.

## **Desarrollo de los métodos y las técnicas estadísticas en investigaciones de causalidad**

La evolución histórica de las técnicas estadísticas para el estudio de la causalidad en las ciencias médicas se dividió en cinco etapas —a consideración de las autoras—, para que fuesen mejor comprendidas, según el desarrollo alcanzado de las técnicas estadísticas en cada una, aunque no existe un límite exacto en el tiempo que las divida, inclusive existe solapamiento entre ellas, a saber:

1. La aparición de las técnicas estadísticas
2. Las técnicas univariadas
3. Las técnicas bivariadas
4. Las técnicas multivariadas
5. Las técnicas avanzadas

A continuación, se presentan, en apretada síntesis, los hechos y las características principales de cada etapa.

### **Etapa 1. La aparición de las técnicas estadísticas**

Desde la era hipocrática hasta el siglo XVII no existían técnicas para el estudio de la causalidad en la medicina. En 1662, John Graunt publicó su estudio con la primera tabla de vida que se conoce, donde recopiló datos de más de 30 años y logró analizar las causas de natalidad y mortalidad y hacer predicciones de estas.<sup>(2)</sup>

En esta época comenzó el desarrollo de la teoría de la probabilidad. En 1713, se propuso el teorema de Bernoulli y la distribución binomial, y entre 1731 y 1736, Thomas Bayes utilizó la probabilidad inductivamente y estableció la base matemática para la inferencia probabilística; además, estudió el problema de la determinación de la probabilidad de las causas a través de los efectos observados, conocido como la probabilidad inversa. Luego, en 1733, De Moivre descubrió la función de densidad de probabilidad normal y, en 1738, propuso un caso particular del teorema central del límite, conocido como el teorema de De Moivre-Laplace, pero no fue hasta finales de ese siglo e inicios del XIX que

se empezaron a sentar las bases teóricas de las probabilidades, con los trabajos de Lagrange y Laplace, cuya aplicación se limitó en ese momento a los juegos de azar.<sup>(3,4)</sup> También a mediados del siglo XVIII comenzó el auge de la estadística descriptiva en asuntos sociales y económicos; sin embargo, existía una oposición al empleo de procedimientos numéricos para el análisis de datos clínicos, puesto que se consideraba la medicina como una disciplina fundamentada únicamente en la fisiología experimental de cada paciente de modo individual y no podían existir conclusiones generalizables.

## **Etapa 2. Las técnicas univariadas**

Esta etapa se extiende desde comienzos hasta finales del siglo XIX y se caracteriza por el censo exhaustivo de las poblaciones en sus inicios y los estudios parciales a partir de pequeñas muestras a finales de siglo, así como la continuación del desarrollo de la teoría de la probabilidad y la estadística por separado, el surgimiento del primer modelo de causalidad y la introducción de los métodos numérico y científico a la medicina. Hasta este momento la estadística empleada en la medicina era muy elemental; solo en unos pocos estudios a finales de siglo se utilizaron técnicas descriptivas univariadas.

Hasta mediados del siglo XIX, para estudiar las causas de las enfermedades o la efectividad de un tratamiento, se empleaban estudios de series de casos con descripciones de cada paciente. A finales de ese siglo surgió la idea de utilizar grupos comparativos en las investigaciones.

Aunque ya desde 1546 se había planteado la primera teoría del contagio, que describe tres vías para la transmisión de una enfermedad,<sup>(5)</sup> no fue hasta la mitad del siglo XIX que comenzaron a desarrollarse teorías relacionadas con la causalidad mediante el modelo de causa única, que trata la causa como necesaria o suficiente y exige la especificidad de la causa y del efecto. La unicausalidad primó durante todo ese mismo siglo, con la teoría del miasma, surgida a mediados del siglo XVIII, y luego, la del germen, con el descubrimiento de los microorganismos y establecimiento de estos como causa de las enfermedades, por Robert Koch, en 1876.

Por esta época, con el aumento de la población, los cambios en el medio ambiente y la revolución industrial, se había consolidado el pensamiento epidemiológico y la

necesidad de utilizar el método numérico para comprender los problemas de salud en relación con el ambiente. Se cita al médico francés Pierre-Charles Alexandre Louis como el primero en introducir el método numérico a la medicina, con el estudio de la neumonía, alrededor de 1830.<sup>(6)</sup>

En 1837, William Farr generalizó el uso de las tasas de mortalidad, los conceptos de población bajo riesgo, gradiente dosis-respuesta, inmunidad de grupo, direccionalidad de los estudios, valor “año-persona” y letalidad. También descubrió las relaciones entre la prevalencia, la incidencia y la duración de las enfermedades.<sup>(2)</sup>

Asimismo, entre los primeros estudios de causalidad en medicina con el empleo de métodos numéricos, se refieren el de John Snow, en 1855, quien usó un diseño similar al de casos y controles en una investigación sobre cólera, y el de Yule, en 1899, relacionado con las causas de la pobreza, en el cual se aplicaron pruebas de significación estadística rudimentarias.

Sin embargo, el mayor representante de los estudios sobre la regularidad estadística en el siglo XIX fue Quételet, quien en 1844 usó los estudios de Poisson y Laplace para identificar los valores promedios de múltiples fenómenos biológicos y sociales, al aplicar, por primera vez, el método científico a estos.<sup>(2,7)</sup>

### **Etapas 3. Las técnicas bivariadas**

Esta etapa se enmarca desde finales del siglo XIX, cuando se registraron cambios importantes en la estadística, al fusionarse con la teoría de la probabilidad, hasta la primera mitad del siglo XX, cuando se produjo un desarrollo vertiginoso de las técnicas estadísticas como consecuencia del intercambio entre las escuelas estadísticas de diferentes países.

La organización y tabulación de los datos se produjo primeramente en la sociología, sobre 1927, teniendo en cuenta las variables y sus categorías.<sup>(8)</sup>

Con la consolidación axiomática de la teoría de la probabilidad, en 1933, se establecieron las bases actuales de las probabilidades y surgió la era moderna de la estadística con la estadística inferencial.

En esta etapa surgieron las técnicas que permiten el estudio de relaciones entre dos variables (la causa y el efecto), como la correlación y la regresión lineal para las variables cuantitativas y la prueba de la  $X^2$  para el caso de las variables cualitativas.

Las pruebas de hipótesis fueron creadas entre 1915 y 1933 como resultado de dos visiones antagónicas, la de Ronald Fisher y la de Jerzy Neyman con Egon Pearson.<sup>(9)</sup> Dichas pruebas permitieron probar la asociación entre dos variables, al codificar y organizar los datos en una tabla de contingencia de dos filas por dos columnas (2x2), según la presencia o ausencia de ambas variables (la causa y el desenlace).

En el cuadro se muestran algunos de los aportes trascendentales a la estadística, casi todos los cuales se sucedieron en esta etapa.<sup>(2,9,10,11)</sup>

**Cuadro.** Aportes trascendentales a la estadística desde finales del siglo XIX hasta mediados del XX

Investigador	Aporte	Año
Andréi Andréyevich <b>Márkov</b> (Rusia, 1856-1922)	Estudió la serie hipergeométrica y sobre las cadenas, que llevan su nombre, en el ámbito de los procesos estocásticos, así como un método para ajustar modelos lineales.	1886-1905
Francis <b>Galton</b> (Inglaterra, 1822-1911)	Motivado por las teorías evolucionistas de la época, inició la aplicación de la estadística a la biología experimental. Sus principales aportes fueron la correlación y la regresión lineal, al comparar la talla de los hijos con la de los padres, e introdujo el concepto de mediana como medida de tendencia central.	1881-1889
Karl <b>Pearson</b> (Inglaterra, 1857-1936)	Creó un puente entre la estadística descriptiva de la época y las probabilidades. Fundó la Bioestadística. Introdujo el término de desviación estándar, creó el método de estimación de los momentos y aportó las primeras medidas de asociación y contingencia. Creó la prueba de bondad de ajuste de la $X^2$ y los coeficientes de correlación para variables cuantitativas y de asimetría para distribuciones uniformes, unimodales y moderadamente asimétricas.	1894-1900
William Sealy <b>Gosset</b> (Inglaterra, 1876-1937)	Dedujo la distribución de la t de Student para el estudio de muestras pequeñas y aplicó por primera vez el método de simulación de Montecarlo.	1908
Andrei <b>Kolmogórov</b> (Rusia, 1903-1987)	Definió el álgebra de las probabilidades e ideó asignar un espacio muestral de sucesos observables a cada experimento aleatorio y representarlos como subconjuntos del espacio muestral, dando una interpretación probabilística a las operaciones con sucesos.	1933
Ronal Arnold <b>Fisher</b> (Inglaterra, 1890-1962)	Considerado por muchos como "el Padre de la Estadística moderna", comenzó a plantear el uso de la Estadística para determinar el efecto causal en problemas agrícolas y luego incursionó en la Genética y la Biología. Ilustró en varias	1922-1935

		publicaciones la construcción de estimadores, los criterios de eficiencia del estimador, el método de estimación de máxima verosimilitud, los intervalos confidenciales, los contrastes de significación, la noción de grados de libertad, la teoría del diseño de experimentos, el análisis de varianza y covarianza y el análisis discriminante. Su método de aleatorización sirvió de base para realizar inferencias causales.	
Walter A. <b>Shewhart</b> (Estados Unidos, 1891-1967)		Delimitó el problema de la calidad en términos de variación de causas asignables y ocasionales e introdujo la gráfica de control como una herramienta para distinguir entre las dos.	1924
Harold <b>Hotelling</b> (Estados Unidos, 1895- 1973)		Generalizó la prueba de la t de Student a varias dimensiones, desarrolló el análisis de correlaciones canónicas y el análisis de componentes principales, esbozado antes por Pearson.	1933
Jerzy <b>Neyman</b> (Polonia, 1894- Estados Unidos, 1981)		Comenzó a desarrollar la teoría del muestreo para superar de manera científica la posibilidad de ahorrar recursos al no estudiar la población completa, y formuló los diseños de muestreo estratificado y por conglomerado. Dio un nuevo enfoque a las pruebas de significación, fortaleció la teoría sobre los intervalos de confianza y creó el lema de Neyman-Pearson, base de la teoría de la decisión. Introdujo el análisis causal mediante el concepto de efecto de un tratamiento.	1934
Charles <b>Spearman</b> (Inglaterra, 1863- 1945)		Desarrolló el análisis factorial exploratorio en el ámbito de la Psicometría para medir variables abstractas no observables (constructos), en especial la inteligencia. También desarrolló el diseño de experimentos factoriales y el coeficiente de correlación de Spearman.	1920- 1935
Maurice George <b>Kendall</b> (Inglaterra, 1907- 1983)		Aplicó el análisis factorial a la productividad de los cultivos. Realizó estudios de series temporales y desarrolló el coeficiente de correlación por rangos de Kendall.	1930- 1972
William Gemmell <b>Cochran</b> (Escocia, 1909 - 1980)		Realizó contribuciones al diseño de experimentos y a la teoría del muestreo. Desarrolló la prueba Q de Cochran para muestras pareadas.	1934
Abraham Wald (Rumanía, 1902- India, 1950)		Fundó la llamada teoría de la decisión para tomar decisiones bajo condiciones de incertidumbre, a diferencia de la escuela inferencista de Fisher. Introdujo las funciones de pérdida, función de riesgo, reglas de decisión admisibles, distribuciones <i>a priori</i> , teoría bayesiana de la decisión, y reglas minimax para la toma de decisión.	1939
Frank <b>Yates</b> (Inglaterra, 1902- 1994)		Trabajó en diseño experimental, realizó contribuciones a la teoría del análisis de la varianza, donde desarrolló los métodos computacionales para el análisis de experimentos a través de mínimos cuadrados de los bloques equilibrados incompletos. También trabajó en la teoría del diseño y análisis de encuestas.	1933- 1954
Frank <b>Wilcoxon</b> (Irlanda, 1892- Estados Unidos, 1965)		Introdujo las pruebas no paramétricas: de suma de rangos para dos muestras independientes y del rango con signo para dos muestras pareadas, alternativas a la prueba paramétrica de la t de Student, y los métodos secuenciales no paramétricos.	1945
George Edward Pelham <b>Box</b> (Inglaterra, 1919- Estados Unidos, 2013)		Realizó importantes contribuciones en las áreas de control de calidad, análisis de series temporales, donde aportó la metodología Box-Jenkins, el diseño de experimentos y la inferencia bayesiana. Acuñó el término robustez, para designar los	1953- 1985

---

métodos estadísticos que procuran asegurar resultados aceptables, cuando no se cumplen los supuestos estándares en que se basan los métodos estadísticos regulares.

---

John Wilder <b>Tukey</b> (Estados Unidos, 1905- 2000)	Introdujo mejoras en los procedimientos de los censos de población, el método de remuestreo Jackknife, para la estimación de sesgos y varianzas, y el análisis exploratorio de datos con los diagramas de tallos y hojas, de cajas y bigotes, radigramas y diagramas de ajuste. Introdujo las técnicas para estimar el espectro de las series temporales y el algoritmo de la transformada rápida de Fourier, fundamental para crear el procesamiento digital de datos.	1965- 1977
---	---	---------------

---

Entre 1930 y 1940 la estadística había alcanzado un elevado nivel matemático, y comenzaron a formularse las distribuciones de probabilidades multivariadas, debido a los progresos logrados en el cálculo de las probabilidades y la fusión de los trabajos de rusos, franceses, ingleses y americanos; también se iniciaron los primeros trabajos aplicando la estadística bayesiana. Después de la Segunda Guerra Mundial se desarrollaron técnicas, como el análisis de series temporales, la teoría de juegos, la teoría del caos, las redes neuronales; pero pasaron muchos años para que estas técnicas se emplearan en las ciencias médicas.

#### **Etapas 4. Las técnicas multivariadas**

Esta época empezó a mediados del siglo XX, cuando se inició la Estadística actual con dos pilares fundamentales: la teoría de la decisión y los métodos bayesianos, y se caracteriza por el establecimiento de los modelos multicausales, la búsqueda de métodos de estimación más eficientes, la aparición de nuevas pruebas no paramétricas y las técnicas multivariadas. Tal desarrollo fue posible por la aparición del computador, que hizo que la metodología estadística se desplazara hacia técnicas aplicadas en grandes masas de datos y procesos iterativos de búsqueda del modelo ideal. En esta etapa se gestó un verdadero cambio en la medicina, donde comenzaron a realizarse estudios de factores de riesgo y ensayos clínicos, con la introducción de técnicas estadísticas avanzadas.

Primero surgió la econometría, pionera en el uso de modelos de análisis de dependencia para el reconocimiento de la causalidad en estudios no experimentales, con la formulación del teorema de covariación, la correlación parcial y las ecuaciones estructurales, las cuales constituyen los primeros antecedentes de los modelos

matemáticos causales, puesto que estas ecuaciones representan una relación teórica de causa-efecto de la que parten los modelos causales jerárquicos o recursivos y de causación mutua.<sup>(8)</sup>

Entre los años 40 y 50, el reconocimiento de la influencia de varios factores en un desenlace dio paso al modelo multicausal determinista, que se consolidó en el año 1960. Bajo la influencia de los trabajos de Fisher, sobre el diseño de experimentos, comenzó la aplicación de la estadística en la investigación médica, que va adaptando las técnicas existentes para el estudio de la causalidad al contexto biomédico. Alrededor de 1940 se reconoció la bioestadística como parte de la investigación médica para el diseño y análisis en el laboratorio, la epidemiología y los ensayos clínicos. Los estudios de esta época tenían como objetivo representar de manera simplificada las relaciones causales o predecir, en base a lo observado, dentro de ciertos márgenes de error.

Las investigaciones sobre factores de riesgo trataban las causas como directas e indirectas y se enmarcaban en el enfoque lineal constante, proporcional y, por tanto, predecible entre dos o más variables (causa y efecto), el cual predominaba en las ciencias naturales. Así, el nexo causa-efecto se analiza, desde entonces, como una relación lineal, no compleja, unívoca y dimensionable. Se empezaron a aplicar la variable dependiente, representada por el desenlace, y tantas covariables como el investigador considerase que pudieran intervenir en dicho desenlace, así como la aleatorización para controlar la variabilidad experimental y eliminar el sesgo. Se desarrollaron con mayor precisión los conceptos de exposición, riesgo, asociación, confusión y sesgo, y se incorporó el uso franco de la teoría de la probabilidad y de un sinnúmero de técnicas estadísticas avanzadas.<sup>(2)</sup>

Doll y Hill incorporaron el método experimental y la asignación aleatoria de individuos en las ciencias médicas. En 1946 se realizó el primer ensayo clínico con la participación de Hilly; en 1950, Doll y Hill publicaron su investigación de casos y controles para el estudio del cáncer de pulmón en relación con el hábito de fumar. En esta época se sucedieron varios estudios prospectivos para la identificación de factores de riesgo, sobre todo de índole laboral.<sup>(11)</sup>

En 1945, Neumann y Morgenstern desarrollaron la teoría probabilística de la utilidad en el ámbito económico y, en 1947, McNemar propuso un test para probar la

homogeneidad de proporciones en dos muestras correlacionadas y Yerushalmy introdujo los términos de sensibilidad y especificidad como indicadores para evaluar la eficacia inherente a una prueba diagnóstica.

Posteriormente, en 1954, se llevó a cabo el mayor y más costoso estudio para evaluar la eficacia de la vacuna contra la poliomielitis. En 1956, Quenouille aportó técnicas que permiten reducir el sesgo y establecer nuevos entornos para la estimación, y en 1959, Mantel y Haenszel propusieron su técnica para controlar las variables confusoras.<sup>(12)</sup>

A partir de 1960 los científicos se dieron cuenta del peligro de realizar inferencias, cuando en los datos aparecieron valores extraños, y propusieron modelos y estimadores robustos como alternativas para estos casos.<sup>(13)</sup>

Entre las décadas de los 60 y 70, MacMahon propuso los métodos de las redes causales o la tela de araña, como los de la diferencia, la analogía y la concordancia. También introdujo el término de asociación causal. Luego propuso el modelo de las cajas negras o modelo histórico-social y aclaró el concepto de interacción cuando comenzaron las discusiones en torno a la metodología de las investigaciones de causalidad de las enfermedades crónicas.<sup>(2,14)</sup>

Otra variante del modelo multicausal fue el esquema propuesto por Leavell y Clark, denominado “historia natural de la enfermedad”, donde se distinguen tres períodos: prepatogénico, patogénico y de resultados.<sup>(15)</sup> En 1965, Hill propuso nueve consideraciones para establecer una relación de causalidad a partir de una relación de asociación, las cuales constituyen la primera propuesta teórica para orientar el proceso de inferencia causal, muy empleada hasta hoy, a pesar de las críticas.<sup>(1)</sup> En 1991, Susser propuso tres criterios: asociación, orden temporal y dirección, e introdujo el llamado modelo de las “cajas chinas” o ecoepidemiológico, donde los sistemas se encuentran separados y organizados jerárquicamente.<sup>(2,16)</sup>

En 1965, Zadeh introdujo la teoría de los conjuntos difusos o borrosos, que ofreció un marco adecuado para tratar con la causalidad imperfecta. Unos 10 años más tarde, Kosko incluyó los mapas cognitivos difusos. Sin embargo, en estos momentos no se visualizaba su utilidad en las ciencias médicas.<sup>(17)</sup>

En la década de los 70, Rothman elaboró su modelo multicausal determinista modificado, intentando superar los modelos anteriores. En este modelo no solo

consideró las múltiples causas, sino también las múltiples vías causales (combinaciones de causas que llevan a un efecto) y trató de determinar la fuerza de las asociaciones causales para conocer la importancia de cada factor causal dentro de esa cadena de multicausalidad. Definió la causa componente y redefinió la causa suficiente; facilitó, además, la explicación de la interacción entre causas, la proporción de enfermedad atribuible a una causa específica y el periodo de inducción.<sup>(1,18,19)</sup>

La idea de normalidad y valor medio dieron paso a nuevas concepciones, con las cuales el modelo multicausal determinístico cedió espacio al multicausal probabilístico-estadístico, que emplea la teoría de las probabilidades y las técnicas estadísticas para probar y estimar la magnitud de la relación observada. Debido a la incertidumbre en las relaciones entre variables, se utilizó el término factor de riesgo, en vez de causa, para indicar una variable supuestamente relacionada con la probabilidad de que un individuo desarrollara una enfermedad; también se introdujeron los términos de factor pronóstico y marcadores de riesgo y pronóstico. Se comenzaron a estimar el riesgo relativo y el riesgo atribuible. Asimismo, ya no se hablaba de causalidad, sino de asociaciones causales, y el empleo de los métodos estadísticos permitía, al utilizar variables continuas, describir curvas de dosis-efecto, en las cuales varios niveles del factor de estudio están relacionados con la probabilidad de desarrollar la enfermedad.<sup>(20)</sup>

En 1970, Cox difundió la regresión logística, cuyos antecedentes están en Bartlett, quien en 1937 utilizó la transformación para analizar proporciones. En 1938, Fisher y Yates sugirieron el uso de esa transformación  $\log(y/(1-y))$  para analizar datos binarios. En 1944, Berkson introdujo el término *logit* para designar esta transformación y, en 1961, Cornfield utilizó, por primera vez, la regresión logística para conocer el riesgo de padecer una enfermedad coronaria, a través del cálculo de los *odds ratio* como valores aproximados del riesgo relativo en un estudio de casos y controles.<sup>(13)</sup> En 1967, Walter y Duncan crearon el algoritmo para estimar la probabilidad de ocurrencia de un desenlace en función de múltiples variables, con la obtención de los estimadores de máxima verosimilitud; paso crucial para automatizar la técnica de la regresión logística. Gracias al surgimiento de los ordenadores fue posible poner en práctica los algoritmos para la obtención de los estimadores, lo cual propició la aplicación y el desarrollo de dicha técnica, llegando a ser la más empleada hasta la actualidad, puesto que permite al

investigador cuantificar el riesgo a partir de múltiples características biológicas o no biológicas y la creación de modelos predictivos de fenómenos complejos, controlando el efecto de posibles variables confusoras y analizando la interacción entre diferentes covariables, siempre que se trate de un desenlace dicotómico.

Sobre los años 70 se desarrolló en la sociología, además del modelo *logit*, el modelo multinivel como una variante de este y el *log lineal* para probar asociación entre múltiples variables categóricas.<sup>(8)</sup> En la econometría se creó el análisis factorial confirmatorio y el primer programa estadístico, conocido como *Linear Structural Relations (LISREL)*, que permitió estimar y probar los modelos de ecuaciones estructurales.<sup>(21)</sup>

Entre 1974 y 1976, Miettinen introdujo el concepto de modificación del efecto (interacción) y estableció la concepción moderna de los diseños de casos y controles, presentando sus bases teóricas, apoyado en los estudios previos de Cornfield, Mantel y Haenszel.<sup>(22)</sup> En esta fecha, para incrementar la validez de los estudios observacionales, Rubin formalizó el modelo de respuestas potenciales o contrafactuales, siguiendo las ideas de Neyman. Este modelo, heredado de la econometría, consta de dos elementos fundamentales: las respuestas potenciales (lo que hubiese sucedido en la ausencia del supuesto factor causal) y el mecanismo de asignación del tratamiento.<sup>(1)</sup>

## **Etapa 5. Las técnicas avanzadas**

Esta época comenzó a finales del siglo XX y se caracteriza por la propuesta de múltiples enfoques, opuestos a la linealidad de pensamiento que reinó en las etapas anteriores. Con los avances en la física cuántica y la genética molecular, surgió el pensamiento de la complejidad como reacción necesaria ante la interpretación determinista y reductora que caracterizó la concepción tradicional de la ciencia hasta esos momentos.

Se estudió la distribución poblacional de genes que podrían explicar las variaciones en la presentación de diversos padecimientos, donde se apreció el componente aleatorio y surgió el enfoque de “curso de vida”, que sugiere la participación de modificaciones epigenéticas a lo largo de la vida, determinando la expresión o el silenciamiento de los genes sin que haya cambios en la estructura del genoma, así como la transmisión

generacional de las modificaciones del epigenoma, que parecen constituir las bases moleculares de muchas enfermedades. Este modo de pensar conduce al paradigma de la complejidad, que obliga a los investigadores a estudiar el proceso salud-enfermedad a través de modelos complejos, dinámicos, jerárquicos, que tienen en cuenta la función creativa del desorden, de las inestabilidades, del azar, de las asimetrías, de las teorías del caos, los fractales, los conjuntos borrosos y las catástrofes.<sup>(23)</sup>

El desarrollo computacional alcanzado después de los 80 permitió un avance extraordinario en los métodos estadísticos, con la aplicación de las técnicas multivariadas esbozadas desde principios del siglo, y en otros algoritmos iterativos complejos, que solo pueden llevarse a cabo con la ayuda del computador, como las técnicas de clasificación de remuestreo: el *bootstrap*, la simulación de Montecarlo y las cadenas de Markov.<sup>(1)</sup>

Otras técnicas desarrolladas fueron la lógica difusa, la estimación autosuficiente, los modelos bayesianos, los métodos basados en la G-estimación, los listados libres (*free-lists*) y las técnicas de la minería de datos (*data mining*), como las reglas de asociación, los árboles de clasificación y las redes neuronales, entre otras.<sup>(24,25,26)</sup>

En los años 70 cobró auge el uso de las técnicas de supervivencia en la medicina, que tienen sus raíces en la ingeniería, para determinar la duración y fiabilidad de los elementos de las maquinarias. A lo largo de la década de los 80, surgieron los métodos de bondad de ajuste para la regresión logística como el test de Hosmer y Lemeshow, el de Score y, por último, los pseudo-R<sup>2</sup> de Cox y Snell.<sup>(27)</sup>

También en los 70 se comenzó a aplicar la curva de características operacionales del receptor (conocida por las siglas ROC del inglés *Receiver Operating Characteristic*) a la evaluación de medios de diagnóstico médico, la cual había sido desarrollada en el contexto de la detección de señales electrónicas en los inicios de la década de los 50 y en 1967, usada por primera vez en el proceso de toma de decisiones médicas por el radiólogo Leo Lusted. De igual forma, las técnicas de inteligencia artificial, como las redes neuronales artificiales, los árboles de decisión y las redes bayesianas, comenzaron a ser empleadas en la investigación médica de causalidad.<sup>(28)</sup>

Regis Gras, en los años 80, utilizó los conceptos de similaridad, cuasi-implicación y cohesión para generar relaciones asimétricas basadas en reglas que conducen a

hipótesis de causalidad en el campo de la didáctica de las matemáticas, con lo que apareció una nueva forma de análisis estadístico, llamada análisis estadístico implicativo,<sup>(29)</sup> que se adaptó al contexto médico y ha sido empleada junto a la regresión logística para identificar factores de riesgo y pronósticos en varios estudios efectuados en Santiago de Cuba.<sup>(30,31,32,33)</sup>

Pearl,<sup>(34)</sup> uno de los primeros en dar un formalismo matemático al fenómeno de la causalidad en las ciencias empíricas, en 1985 propuso las redes de inferencia como una extensión del modelo probabilístico clásico basado en las redes bayesianas y, en 1993, formalizó métodos gráficos como el grafo acíclico dirigido, basado en las probabilidades condicionales, para representar las relaciones causales, y retomó algunos métodos, como las variables instrumentales empleadas desde 1920 en econometría, para controlar el efecto confusor residual, y los índices de susceptibilidad, para lograr la comparabilidad de los grupos en diseños no experimentales.<sup>(35)</sup>

En 1982 se inició la difusión oficial, en la revista JAMA, de un nuevo enfoque para la práctica médica, la medicina basada en la evidencia. Su mayor impulsor fue Archie Cochrane quien, en 1970, reclamó una revisión crítica de los ensayos clínicos, relevantes y periódicos, según la especialidad. Estas revisiones se apoyan en el metaanálisis para resumir los resultados de la literatura en una valoración cuantitativa, la cual fue empleada por primera vez por Karl Pearson, en 1904, para superar el problema del reducido poder estadístico de los estudios con pequeños tamaños muestrales. El metaanálisis fue aplicado, inicialmente, en las ciencias sociales y en psicología; a partir de los 80, comenzó a aplicarse de forma creciente en medicina.

En 1986, Robins presentó un enfoque gráfico para la identificación y estimación de parámetros causales en estudios de cohorte ocupacional, donde es frecuente el sesgo del trabajador saludable. Luego formalizó su teoría contrafactual, que es una generalización de las anteriores, pero que puede aplicarse a estudios longitudinales con exposiciones cambiantes. El primer producto de su teoría fue la fórmula g, un método causal no paramétrico, la generalización del ajuste por estandarización, en contraste con los métodos tradicionales que ajustan mediante estratificación de la medida de asociación y que permite calcular efectos causales de exposiciones cambiantes, bajo la condición de no confusión residual, incluso en presencia de confusores cambiantes afectados por la

exposición. Luego, como la fórmula  $g$  requeriría cantidades enormes de datos y tiempo de cálculo, desarrolló dos clases de modelos semiparamétricos: los modelos estructurales marginales y los estructurales anidados.<sup>(36,37)</sup>

Más adelante, en el 2000, se consolidó el modelo causal estructural, el cual unificó las teorías previas sobre los métodos gráficos, los resultados potenciales o contrafactuales, los confusores, las causas componentes y suficientes, y las ecuaciones estructurales; proporcionando perspectivas complementarias y mejorando las interpretaciones causales de los resultados estadísticos convencionales. Estos constituyen una alternativa a los modelos clásicos de regresión cuando existe una variable confusora dependiente del tiempo. La Escuela de Salud Pública de Harvard trabaja en el perfeccionamiento de estos modelos.<sup>(37)</sup>

Estudios recientes han abordado también métodos para tratar datos perdidos (*missing value*) y los resultados potenciales perdidos, como la imputación múltiple y los  $g$ -estimadores paramétricos,<sup>(38,39)</sup> o los diagramas causales canónicos,<sup>(40)</sup> aunque, desde 1978, Rubin había propuesto el método de imputación múltiple de cadenas de Markov Montecarlo, como alternativas ante la ausencia de valores en las bases de datos.<sup>(33)</sup>

También se han realizado notables esfuerzos por mejorar el diseño metodológico de estas investigaciones, estableciendo guías (actualizables de manera periódica), donde se pautan los métodos estadísticos y todos los aspectos que avalan la calidad del informe de la investigación.<sup>(41,42)</sup>

Se continúa trabajando en la búsqueda de algoritmos para mejorar la obtención de diagramas causales,<sup>(43)</sup> de métodos que identifiquen confusores entre los modelos multivariados,<sup>(44)</sup> técnicas para incrementar la validez de los estudios no experimentales, como la aleatorización mendeliana, creada en el 2013, como un tipo particular de análisis entre las variables instrumentales, con lo cual se busca una aproximación a los estudios experimentales mediante el uso de variantes génicas relacionadas con algún factor de exposición, y además se evita el sesgo de causalidad inversa.<sup>(45,46)</sup>

El avance en la genética molecular ha llevado a la creación de grandes bases de datos a nivel internacional, con la que se llevan a cabo múltiples investigaciones de causalidad, y se siguen creando herramientas informáticas para el procesamiento de la información

genómica, la determinación de modelos y su relación causal con enfermedades comunes, para fomentar la medicina personalizada.<sup>(47)</sup>

El avance en la informática, la nanotecnología y las comunicaciones, también han permitido el empleo de satélites y sensores remotos en el ámbito de la salud. En el estudio de la causalidad medioambiental se entrelazan ciencias que aplican técnicas estadísticas espacio-temporales avanzadas, como la geografía médica, la geología médica, la biometeorología y la ecología sanitario-ambiental.

Otros aspectos que se deben considerar son el apoyo de las investigaciones de causalidad cuantitativas con técnicas cualitativas,<sup>(48)</sup> la evaluación indispensable por parte de los comités de ética de la investigación en salud;<sup>(49)</sup> y, dada la premura en conseguir mejores resultados, se sugiere a los autores mostrar a la comunidad científica sus artículos, junto a sus datos, para que se puedan replicar los estudios reutilizando los mismos datos.<sup>(50)</sup>

Para ampliar esta información, se puede consultar a Rodríguez-Villamizar<sup>(1)</sup> y López-Moreno,<sup>(2)</sup> quienes exponen detalladamente todos los modelos de causalidad ya referidos, así como Celik<sup>(3)</sup> y Almenara,<sup>(11)</sup> que ofrecen los detalles de la evolución de la estadística y la bioestadística, en particular.

## **Consideraciones finales**

En los últimos años el desarrollo de la estadística ha sido acelerado y vertiginoso, tanto que a los especialistas de la rama les cuesta apropiarse de las múltiples técnicas existentes. No obstante esa variedad, desde mediados del siglo XIX, la técnica más empleada para el estudio de la causalidad en las ciencias biomédicas ha sido la regresión logística binaria.

Hoy día es preciso aprovechar los descubrimientos científicos, el gran avance en la bioestadística, la computación con los múltiples paquetes de software estadísticos, los sistemas de información, la inteligencia artificial, los procesos de análisis en línea, el acceso y tratamiento de grandes volúmenes de datos de estudios poblacionales de diferentes fuentes, con la identificación de modelos válidos, potencialmente útiles y

entendibles, mediante la aplicación de las técnicas de descubrimiento del conocimiento en bases de datos, la colaboración internacional y la transdisciplinariedad, donde confluyen matemáticos, filósofos, cibernéticos, biólogos, epidemiólogos, económicos, sociólogos, geógrafos; a fin de lograr un óptimo aprovechamiento de los saberes con la aplicación de las más disímiles técnicas.

## Referencias bibliográficas

1. Rodríguez Villamizar L. Inferencia causal en epidemiología. Rev Salud Pública. 2017;19(3):409-15.
2. López Moreno S, Garrido Latorre F, Hernández Avila M. Desarrollo histórico de la epidemiología: su formación como disciplina científica. Salud Pública de México. 2000 [citado 11/09/2018];42(2). Disponible en: <http://saludpublica.mx/index.php/spm/article/viewFile/6221/7398>
3. Celik MY. The Miracle of Biostatistics in Medical Research. International Journal of Basic and Clinical Studies (IJBCS). 2013;2(2):1-6.
4. A Brief History of Statistics (Selected Topics). ALPHA Seminar. 2017, 29 Ago [citado 11/09/2018]. Disponible en: <http://homepage.divms.uiowa.edu/~dzimmer/alphaseminar/Statistics-history.pdf>
5. Malagón Oviedo R. Epidemiología, saberes y prácticas: un análisis crítico. Rev Salud Pública. 2017 [citado 11/09/2018];19(3):416-22. Disponible en: <http://www.scielo.org.co/pdf/rsap/v19n3/0124-0064-rsap-19-03-00416.pdf>
6. Rosser Matthews J. History of Biostatistics. Medical Writing. 2016 [citado 11/09/2018];25(3). Disponible en: <https://journal.emwa.org/statistics/history-of-biostatistics/>
7. Tanton JS. History of Probability and Statistics. En: Encyclopedia of Mathematics. New York: Facts On File; 2005.
8. Lizón MA. Estadística y causalidad en la sociología empírica del XX. Papers. 2006 [citado 11/09/2018];(80):223-55. Disponible en: <https://core.ac.uk/download/pdf/13270350.pdf>

9. Historia de la Estadística. Revista Varianza. 2011 [citado 11/09/2018];(8). Disponible en: [http://www.revistasbolivianas.org.bo/scielo.php?pid=S9876-67892011000100011&script=sci\\_arttext&tlng=es](http://www.revistasbolivianas.org.bo/scielo.php?pid=S9876-67892011000100011&script=sci_arttext&tlng=es)
10. Rivas MD, Araujo A, Martínez L, Flores P. Historia de la Estadística y la Probabilidad. Ciudad Bolívar: Universidad Pedagógica Experimental Libertador; 2014 [citado 11/09/2018]. Disponible en: [https://www.academia.edu/7476667/HISTORIA\\_DE\\_LA\\_ESTADISTICA\\_Y\\_LA\\_PROBABILIDAD](https://www.academia.edu/7476667/HISTORIA_DE_LA_ESTADISTICA_Y_LA_PROBABILIDAD)
11. Almenara Barrios J, Silva Ayçaguer LC, Benavides Rodríguez A, García Ortega C, González Caballero JL. Historia de la Bioestadística, la génesis, la normalidad y la crisis. Rev Esp Salud Pública. 2004 [citado 11/09/2018];78(1). Disponible en: [http://scielo.isciii.es/scielo.php?script=sci\\_arttext&pid=S1135-57272004000100011](http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1135-57272004000100011)
12. Silva Ayçaguer LC. Causalidad y predicción: diferencias y puntos de contacto. Medwave. 2014 [citado 11/09/2018];14(8). Disponible en: <https://www.medwave.cl/link.cgi/Medwave/Series/TyC-Estadistica/6016>
13. Molinero LM. Historia del razonamiento estadístico. Madrid: SEH-LELHA; 2004 [citado 11/09/2018]. Disponible en: <https://www.seh-lelha.org/historia-del-razonamiento-estadistico/>
14. Parascandola M. Causes, risks and probabilities: probabilistic concepts of causation in chronic disease epidemiology. Prev Med. 2011 [citado 11/09/2018]; 53(4-5):232-4.
15. Sánchez Pérez MP, Velasco Benítez CA. El modelo de Leavell y Clark como marco descriptivo dentro de las investigaciones sobre el virus de la hepatitis B en niños con infección por VIH/sida del grupo de investigación Gastrohnutp de la Universidad del Valle de Cali, Colombia. Revista Gastrohnutp. 2017 [citado 11/09/2018];15(3 Sup. 1):6-9. Disponible en: <http://revistas.univalle.edu.co/index.php/gastrohnutp/article/view/1316/1425>
16. Vallverdú J. Causalidad y epidemiología. En: Estany A, Puyol A. Filosofía de la epidemiología social. Madrid: CSIC; 2016.

17. Leyva Vázquez M, Pérez Teruel K, Febles Estrada A, Gulín González J. Técnicas para la representación del conocimiento causal: un estudio de caso en Informática Médica. Rev Cuba Inf Cienc Salud. 2013 [citado 11/09/2018];24(1). Disponible en: [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S2307-21132013000100006](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2307-21132013000100006)
18. Wensink M, Westendorp RG, Baudisch A. The causal pie model: an epidemiological method applied to evolutionary biology and ecology. Ecol Evol. 2014 [citado 11/09/2018];4(10):1924-30. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4063485/>
19. Hakeem Alrawahi A. Classical causation research practices and sufficient-component cause model – Appraisal and pitfalls. Epidemiology Biostatistics and Public Health. 2017 [citado 11/09/2018];14(3). Disponible en: <https://ebph.it/article/download/12576/11418>
20. Rodríguez López JF, Rodríguez González B. Epidemiología: El cambio de paradigmas. Medident Electrón. 2014 [citado 11/09/2018];18(3):93-9. Disponible en: [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S1029-30432014000300002](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1029-30432014000300002)
21. Manzano Patiño AP. Introducción a los modelos de ecuaciones estructurales. Inv Ed Med. 2017 [citado 11/09/2018];7(25):67-72. Disponible en: <http://www.scielo.org.mx/pdf/iem/v7n25/2007-5057-iem-7-25-67.pdf>
22. Lazcano Ponce E, Salazar Martínez E, Hernández Avila M. Estudios epidemiológicos de casos y controles. Fundamento teórico, variantes y aplicaciones. Salud Pública Méx. 2001 [citado 11/09/2018];43(2). Disponible en: <https://www.scielosp.org/article/spm/2001.v43n2/135-150/>
23. Bacallao Gallestey J, Alerm González A, Ferrer Arrocha M. Paradigma del curso de la vida. Implicaciones en la clínica, la epidemiología y la salud pública. La Habana: Editorial Ciencias Médicas; 2016.
24. Wang H, Tu W. Bootstrap Methods: The Classical Theory and Recent Development. Wiley Online Library, 2018 [citado 11/09/2018]. Disponible en: <https://doi.org/10.1002/9781118445112.stat04579.pub2>
25. Idrovo AJ. Determinación social del proceso salud enfermedad: una mirada crítica desde la epidemiología del siglo XXI. Rev Salud Pública. 2017;19(3):404-8.

26. Bacallao Gallestey J. Enfoques modernos del sesgo y la causalidad en la investigación epidemiológica. Rev Cubana Salud Pública. 2012 [citado 11/09/2018];38(Supl. 5). Disponible en: [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S086434662012000500003&lng=es&nrm=iso&tlng=es](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S086434662012000500003&lng=es&nrm=iso&tlng=es)
27. Iglesias Cabo T, Escabias Macucha M, Aguilera del Pino AM. Métodos de Bondad de Ajuste en Regresión Logística [tesis de maestría]. Universidad de Granada; 2013 [citado 11/09/2018]. Disponible en: [http://masteres.ugr.es/moea/pages/tfm-1213/tfm\\_iglesiascabo\\_tania/](http://masteres.ugr.es/moea/pages/tfm-1213/tfm_iglesiascabo_tania/)
28. Monsalve Torra AE, Ruiz Fernández D, Soriano Payá A. Sistema de ayuda a la decisión clínica en enfermedades de diagnóstico complejo [tesis doctoral]. Alicante: Universidad de Alicante; 2017 [citado 21/11/2018]. Disponible en: [https://rua.ua.es/dspace/bitstream/10045/65334/1/tesis\\_monsalve\\_torra.pdf](https://rua.ua.es/dspace/bitstream/10045/65334/1/tesis_monsalve_torra.pdf)
29. Couturier R, Pazmiño R. Use of Statistical Implicative Analysis in Complement of Item Analysis. International Journal of Information and Education Technology. 2016 [citado 21/11/2018];6(1). Disponible en: <http://www.ijiet.org/vol6/655-DL0026.pdf>
30. García Mederos Y, Zamora Matamoros L, Sagaró del Campo N. Análisis estadístico implicativo en la identificación de factores de riesgo en pacientes con cáncer de pulmón. MEDISAN. 2015 [citado 21/11/2018];19(8). Disponible en: [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S1029-30192015000800003](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1029-30192015000800003)
31. Moraga Rodríguez A, Zamora Matamoros L, Sagaró del Campo NM, Moraga Rodríguez A, Rodríguez Griñán A. Análisis estadístico implicativo para la identificación de factores pronósticos de la mortalidad por cáncer de pulmón. MEDISAN. 2016 [citado 21/11/2018];20(3). Disponible en: [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S1029-30192016000300010](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1029-30192016000300010)
32. Moraga Rodríguez A, Zamora Matamoros L, Sagaró del Campo NM, Moraga Rodríguez A, Rodríguez Griñán A. Análisis estadístico implicativo para la identificación de factores pronósticos de la mortalidad por cáncer de mama. MEDISAN. 2017 [citado 21/11/2018];21(4):413. Disponible en: [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S1029-30192017000400003](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1029-30192017000400003)

33. Paez Candelaria Y, Sagaró del Campo NM Zamora Matamoros L. Análisis estadístico implicativo en la determinación de factores pronósticos del estado nutricional del paciente grave al egreso. MEDISAN 2018 [citado 21/11/2018];22(6):650. Disponible en: [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S1029-30192018000600007](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1029-30192018000600007)
34. Pearl J. An Introduction to Causal Inference. Int J Biostat. 2010 [citado 21/11/2018];6(2):7. Disponible en: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2836213/pdf/ijb1203.pdf>
35. Weihua An. Instrumental variables estimates of peer effects in social networks. Social Science Research. 2015;50:382-94.
36. Hernán MA. Conocimiento experto, confusión y métodos causales. Gac Sanit. 2001;15(Supl. 4):44-8.
37. Hernan MA, Robins JM. Causal inference. Boca Raton: Chapman & Hall; 2017.
38. Harel O, Mitchell EM, Perkins NJ, Cole SR, Tchetgen EJ, Sun B, Schisterman EF. Multiple Imputation for Incomplete Data in Epidemiologic Studies. Am J Epidemiol. 2018 [citado 21/11/2018];187(3):576–84. Disponible en: <https://academic.oup.com/aje/article-abstract/187/3/576/4642952>
39. Perkins NJ, Cole SR, Harel O. Principled approaches to missing data in epidemiologic studies. Am J Epidemiol. 2018;187(3):568–75.
40. Moreno-Betancur M, Lee KJ, Leacy FP, White IR, Simpson JA, Carlin JB. Canonical Causal Diagrams to Guide the Treatment of Missing Data in Epidemiologic Studies. Am J Epidemiol. 2018; 187(12):2705-15.
41. Hung M, Bounsanga J, Wright Voss M. Interpretation of correlations in clinical research. Postgrad Med. 2017; 129(8):902–6.
42. Castellanos-Olivares A, Rojas-Peñaloza J, Vásquez-Márquez PI. ¿Cómo evaluar artículos de riesgo o causalidad en medicina? Revista Mexicana de Anestesiología. 2016;39(Supl. 1):232-5.
43. Oates CJ, Kasza J, Simpson JA, Forbes AB. Repair of Partly Misspecified Causal Diagrams. Epidemiology. 2017;28(4):548–52.
44. Janzing D, Schölkopf B. Detecting confounding in multivariate linear models via spectral analysis. Journal of Causal Inference. 2017 [citado 21/11/2018];6(1). Disponible en: <https://arxiv.org/pdf/1704.01430.pdf>

45. Holmes MV, Lange LA, Palmer T, Lanktree MB, North KE, Almqvist B, et al. Causal effects of body mass index on cardiometabolic traits and events: a Mendelian randomization analysis. *Am J Hum Genet.* 2014;94:198-208.

46. Iturrieta-Zuazo I, Stefan Walter S. Aleatorización mendeliana: presente y futuro de los estudios epidemiológicos en cardiología. *Rev Esp Cardiol.* 2015 [citado 21/11/2018];68(2):87-91. Disponible en:

<http://www.revespcardiol.org/es/aleatorizacion-mendeliana-presente-yfuturo-delosestudios/articulo/90376679/>

47. Laperle J, Hébert-Deschamps S, Raby J, de Lima Morais DA, Barrette M, Bujold D, et al. The epi Genomic Efficient Correlator (epiGeEC) tool allows fast comparison of user data sets with thousands of public epigenomic datasets. *Bioinformatics.* 2019;35(4):674-6.

48. Palinkas LA. Causality and Causal Inference in Social Work: Quantitative and Qualitative Perspectives. *Res Soc Work Pract.* 2014;24(5):540-7.

49. De Lecuona I. Evaluación de los aspectos metodológicos, éticos, legales y sociales de proyectos de investigación en salud con datos masivos (big data). *Gaceta Sanitaria.* 2018 [citado 21/11/2018];32(6):576-8. Disponible en: <http://www.gacetasanitaria.org/es-evaluacion-los-aspectos-metodologicos-eticos-articulo-S0213911118300864>

50. Ebrahim S, Ferrie J. E, Smith G. D. The future of epidemiology: methods or matter? *International Journal of Epidemiology.* 2016;45(6):1699-716.



Esta obra está bajo una [licencia de Creative Commons Reconocimiento-NoComercial 4.0 Internacional](https://creativecommons.org/licenses/by-nc/4.0/).