

Epidemiología para *dummies*

Dr. John Jairo Zuleta Tobón

Especialista en Obstetricia y Ginecología

Magister en Epidemiología

Docente Universidad de Antioquia

Introducción

En New England Journal of Medicine se publicó: “En comparación con mujeres que nunca han utilizado anticoncepción hormonal, el riesgo relativo de cáncer de mama entre todas las mujeres que en la actualidad usan o recientemente han usado anticoncepción hormonal fue 1.20 (IC95% 1.14 a 1.26). Este riesgo incrementa [...] a 1.38 (IC95% 1.26 a 1.51) con más de 10 años de uso ($P=0.002$)”. De manera casi que simultánea, Newsweek “tradujo” esta información para sus lectores como “Cáncer de mama: el control de los nacimientos puede incrementar el riesgo hasta 38%”; por su parte, El Espectador, en Colombia, titula “Todos los anticonceptivos hormonales aumentan riesgo de cáncer de mama”. Una colega escribe: “¿Vamos a suspender la anticoncepción hormonal a todo el mundo?” y, más extremo, un mastólogo de la ciudad de Medellín, muy acucioso, suspende los anticonceptivos orales combinados a una mujer de 28 años porque “mamas densas y anticonceptivos ya son dos factores de riesgo”.

Relativo vs Absoluto

Aunque antes de interpretar los resultados de una investigación es necesario evaluar la validez de la metodología utilizada, es innegable que son ellos los que se llevan la primera impresión, y si adicionalmente se presentan de una manera atractiva, vendedora, con mayor razón: es bien

conocido que la forma de presentar los resultados puede afectar las decisiones de las personas en general y del personal de salud en particular.

Para afirmar que existe una asociación entre un factor o exposición (ejemplo, tener antecedente familiar de cáncer de mama) o una intervención (un medicamento hormonal) con un desenlace (cáncer de mama) es necesario comparar los resultados obtenidos en al menos dos grupos, uno expuesto y otro no expuesto al factor de interés (hormonas), si se trata de un estudio de cohortes, o un grupo con el desenlace (cáncer de mama) y otro sin él, si se trata de estudios de casos y controles o un ensayo clínico. Cuando una investigación evalúa la aparición o no aparición de un desenlace o resultado (cáncer de mama o ausencia de cáncer de mama), la investigación que aplica el método epidemiológico puede presentar los resultados como una resta (términos absolutos) o una división (términos relativos) de los resultados obtenidos en ambos grupos.

Una misma magnitud de efecto tiene impactos bien diferentes dependiendo de cómo se presenta. En términos relativos se puede expresar, por ejemplo, que un tratamiento logró una reducción en la frecuencia de un desenlace negativo para la salud en un 50%, y esto inmediatamente impacta positivamente y la intuición llevará a la mayoría de personas a aceptar la intervención. Sin embargo, esa disminución del 50% puede representar presencia del desenlace en 90% del grupo sin la intervención y 45% en el

grupo con ella, pero también puede ser que se pase de dos por cada millón de personas a una por cada millón. Aunque ambos representan la misma reducción en términos relativos (reducción en 50%), los juicios serán diferentes si se analizan en términos absolutos (reducción del 50% en el primer caso, reducción del 0.0000001%, en el segundo).

Situación similar sucede con los incrementos del riesgo. Los anticonceptivos hormonales en el estudio en mención incrementaron hasta un 38% el riesgo de cáncer de mama en el subgrupo de mujeres que los usaron por más de 10 años. Como estimado general del estudio, el incremento fue del 20% (RR 1.2) y este 20% para la generalidad de las personas se interpretará como dos mujeres de cada 10, o una de cada cinco, lo que convertiría en prohibitivo el uso de estos medicamentos. Una mujer que lea el titular de Newsweek entenderá que casi cuatro de cada 10 usuarias desarrollarán cáncer, y está demostrado que la interpretación en médicos, incluso especialistas, no es mejor. La pregunta que siempre se debe hacer el clínico, o cualquier persona que quiera entender esta forma de presentar el riesgo, es: ¿Es un incremento de 38% a partir de cuánto? ¿Es un incremento de 20% a partir de cuánto? Es diferente ganarse el 80% del valor de un tiquete del metro de Medellín a ganarse un 1% de lo recibido por el Plan Colombia. Posiblemente es mejor un descuento del 10% en una camisa de un almacén de ropa italiana que un 40% en un almacén del centro-centro de la ciudad. Este es el cálculo que se debe realizar siempre. ¿Incremento o disminución a partir de cuánto?

Específicamente, ¿qué significa un incremento del riesgo de cáncer de mama en Colombia por el uso de anticonceptivos hormonales? Según Globocan, en Colombia se presentan 35.9 casos por cada 100.000 mujeres. Obviamente, la frecuencia es menor en las susceptibles de utilizar anticoncepción: 8.5 casos por cada 100.000 mujeres entre 15 y 39 años, que asciende a 47.4 por 100.000 mujeres entre 40 y 44 años. Una simple regla de tres permite calcular que un aumento del 20% desde 8.5 del grupo de 15 a 39 años es 1.7, lo que significa que en las usuarias de anti-

concepción hormonal no serán 8.5 cánceres sino 10.2 por el uso de anticoncepción hormonal, pero por cada 100.000 mujeres. En ellas ese aumento relativo del 20% corresponde a un aumento absoluto de 1.7 por 100.000, o lo que es lo mismo, del 0.000017%. Para las mujeres entre 40 y 44 años se aumentarán 9.5 mujeres, es decir, se pasará de 47.4 a 56.9 por cada 100.000. Es distinto pensar en un incremento del casi 40% como cuatro de 10 a pensarlo como 9.5 por cada 100.000 y en estas condiciones es más fácil hacer una evaluación de riesgos y beneficios. Existe una medida de más fácil interpretación: el Número Necesario a Tratar NNT, técnicamente definido como “el inverso de la reducción o aumento absoluto del riesgo”, en este caso $1/0.000569 - 0.000474 = 1/0.000095 = 10526$, que significa que sería necesario que 10.526 mujeres usaran anticoncepción hormonal durante el mismo tiempo evaluado en el estudio para que se presente un caso más de cáncer de mama de los que se presentarían si no usaran estos medicamentos. En el artículo, gracias a que el riesgo basal de cáncer de mama es mayor, esta cifra fue de 7.960.

¿De dónde sale el porcentaje de aumento o de disminución del riesgo?

El informe de la investigación plantea que el riesgo relativo (RR) de cáncer de mama entre las todas las mujeres que en la actualidad usan o recientemente han usado anticoncepción hormonal fue 1.20. El incremento del 20% referido se obtiene de ese RR, así como el incremento del 38% del periódico se obtiene del RR 1.38. Si se entiende cómo se obtiene esta relación será más fácil asimilar esta interpretación. El riesgo relativo se obtiene de dividir la proporción de incidencia del desenlace en el grupo expuesto al factor o a la intervención por la proporción de incidencia del desenlace en el grupo no expuesto. En el estudio, por cada 100.000 mujeres año hubo 55 casos de cáncer de mama entre las no usuarias y 68 entre las usuarias. Si se busca una relación entre esas dos proporciones con

una simple regla de tres, se encuentra que por cada una mujer que no usó anticoncepción hormonal y tuvo cáncer, se presentaron $(1 \times 68) / 55 = 1.24$ cánceres en el grupo de usuarias [Se aclara que se trae el concepto de regla de tres para facilitar la comprensión, pero en vida real se aplican fórmulas técnicas].

Por otro lado, el resultado real fue de 1.20 porque en el análisis se tienen en cuenta otros factores que influyen (ajuste, que luego se retomará) y se tiene en cuenta el tiempo de exposición. Un RR de 1.38 significa que por cada un cáncer en mujeres no usuarias de anticoncepción hormonal se presentaron 1.38 en usuarias de anticoncepción hormonal, que también se puede expresar que por cada 100 no usuarias hubo 138 usuarias con cáncer, con lo cual es obvio ver de dónde se obtiene el incremento del 38%. En conclusión, para conocer cuánto es el aumento del riesgo representado en un RR, se resta uno al valor de ese RR y se puede expresar en porcentaje. En casos en que el RR es un número menor que uno, significa que hubo una disminución del riesgo y, para obtener la proporción de disminución, se extrae de uno el valor obtenido. Por ejemplo, el RR para la anticoncepción con terapia transdérmica (parches) se obtuvo un RR de 0.85, lo que significaría que por cada mujer no usuaria de anticoncepción hormonal que desarrolla cáncer, lo desarrollarían 0.85 usuarias de parche, por lo tanto $1 - 0.85 = 0.15$, o lo que es lo mismo, reducción de un 15% (**Nota:** esta información se debe complementar con la evaluación del valor de p o con el intervalo de confianza para hacer el juicio completo).

El RR también se puede interpretar como el número de veces (o veces más) que se presenta el desenlace en los expuestos con respecto a los no expuestos; el HR para más de 10 años de ciproterona fue dos, lo que significa que el riesgo de cáncer de mama es dos veces el de las no usuarias o que es una vez más frecuente en usuarias que en no usuarias (esto es así porque uno significa que el riesgo es igual entre los dos grupos). También significa que incrementa el riesgo en un 100%, según las interpretaciones previas. Dependiendo del tipo de estudio y de las pruebas

estadísticas utilizadas, el riesgo también se puede expresar en términos de Odds Ratio OR o de Hazard Ratio HR.

Entonces, ¿qué significado y qué importancia tiene el intervalo de confianza?

El intervalo de confianza hace parte de la estadística inferencial, es decir, la que permite inferir a partir de una muestra lo que sucede en toda la población. Su importancia está en que cuantifica la magnitud de la medición o tamaño del efecto cuando se evalúa un tratamiento o un factor, permite aproximarse a la importancia clínica del hallazgo, identifica la dirección del efecto y aporta un rango de valores compatibles con el resultado obtenido. Esto último es importante porque siempre existe incertidumbre con las mediciones puntuales obtenidas en un único estudio por la variación natural de los muestreos.

Si se tuviera la oportunidad de estudiar toda la población, se obtendría el verdadero valor del parámetro que se quiere conocer. Como no es factible estudiarla toda, se estudian muestras obtenidas de ella para estimar ese parámetro. Asumiendo que se realizan estudios respetando los principios metodológicos, la variación dada por diferentes fuentes, entre ellas el azar, hará que se obtengan valores diferentes en cada uno de esos estudios, unos valores estarían cercanos alrededor del verdadero, otros más alejados e, incluso, es posible encontrar, en pocas ocasiones, unos valores extremos y poco representativos del parámetro.

La definición técnica del IC 95% contempla que, al repetir el muestreo, aproximadamente el 95% de los intervalos construidos incluirá, a la larga, el verdadero valor existente. Cuando se tiene un único resultado de un estudio es imposible saber si se obtuvo un valor cercano o muy alejado al verdadero, y ante hechos ya consumados, solo existen dos opciones: ese valor verdadero está dentro o está por fuera del IC calculado para este estudio, pero solo con esa

información es imposible saber dónde está. Las fórmulas y los insumos que se utilizan para el cálculo del IC permiten afirmar con un grado de confianza (por ejemplo, 95%) que el valor verdadero está entre esos límites. Queda un porcentaje (5% si la confianza se calculó del 95%) de que el valor verdadero no esté comprendido entre ellos. Dado que el IC 95% informa que de cada 100 estudios que se realicen en iguales o muy similares condiciones, 95 de los IC construidos incluirán el valor verdadero, es aceptable tener más confianza de que el estudio puntual que se realizó pertenezca a los 95 que contienen el valor verdadero a que pertenezca a los cinco que no lo incluyen. Un supuesto para que esto se cumpla es que los estudios se realizan en condiciones similares, con una correcta aplicación de la metodología para controlar potenciales sesgos que tergiversen los resultados, con lo cual la variabilidad en las mediciones es explicada por el azar del muestreo. Por otro lado, es posible calcular IC con confianzas diferentes, por ejemplo, 90% o 99% y el significado y la interpretación serán similares, obviamente, varía el valor asignado a esa confianza.

En el caso que se está analizando, se encontró un riesgo relativo de 1.20, con un intervalo de confianza del 95% (IC95%) que va desde 1.14 hasta 1.26. Debido que no se analizó todo el universo de mujeres sino una muestra, el RR 1.2 obtenido (aumento de 20% en el riesgo de desarrollar cáncer de mama) variará en estudios similares, por lo tanto, ese aumento de 20% es estadísticamente compatible con un aumento tan bajo como 14% ($1.14 - 1 = 0.14$), tan alto como 26% ($1.26 - 1 = 0.26$) o con cualquiera de todos los posibles valores existentes entre esos dos límites. Ese es un rango de valores plausibles compatibles con el modelo evaluado y según los datos obtenidos. Este IC 95% entre 1.14 y 1.26 implica que es menos probable que la verdadera diferencia sea menor de 1.14 o mayor de 1.26 y, por lo tanto, se acepta que es más probable que la verdadera diferencia esté entre 1.14 y 1.26. No se asegura que el verdadero valor se encuentre allí, el verdadero valor está adentro o no está adentro y lo que queda es mayor confianza de que esté dentro del IC calculado.

El concepto técnico ortodoxo es complejo y difícil de aceptar para quienes no tenemos la suficiente formación estadística, situación que ha llevado a que incluso estadísticos y textos de estadística reconocidos propongan interpretaciones erróneas para que los no estadísticos lo “comprendamos”, como expresar que “se tiene una confianza del 95% de que el verdadero valor se encuentra entre los dos límites del intervalo”. Esta interpretación es cuestionada en la literatura, generadora de un número alto de artículos que llaman a evitar esas interpretaciones “prácticas”, porque llevan a tomar conclusiones erróneas no respaldadas de manera objetiva con los datos, que es la intención del método científico. Existen propuestas para llegar a una interpretación válida de conocer cuál es la probabilidad de que el verdadero valor se encuentre en el IC, condicionado por los datos obtenidos en el estudio, específicamente en los análisis con enfoque bayesiano, que tampoco están libres de cuestionamientos. La estimación de los intervalos de credibilidad, denominación que toman en este enfoque, requieren de conocimientos en estadística bayesiana que no están al alcance de estas notas.

¿Cómo interpretar el valor de p?

Es un concepto que, aunque es altamente cuestionado en la literatura para muchos usos en investigación clínica, persiste como elemento central. Esto se complica más porque es bien reconocido que la mayoría de las personas lo interpretamos mal. Existe un llamado general a que para las decisiones de inferencia estadística se prefiera el IC, pero se reconoce que, en muchas situaciones, propias de la estadística, es necesario y correcto presentar resultados en términos de valor p.

El valor de p se utiliza en el contexto de prueba de significación y en el contexto de una estrategia que pretende corregir algunas de las deficiencias de ella, conocida como prueba de hipótesis. En esta última, el investigador plantea una hipótesis llamada alterna (H_a) que es la que él quiere demostrar como cierta y una hipótesis nula (H_0)

que propone lo contrario a lo que se quiere demostrar. Contrario a lo que se esperaría, la investigación no se planea para confirmar H_0 sino que toda la estrategia estadística se planea para rechazar H_0 . Si no se logra confirmar H_0 , se le rechaza y, sin ninguna comprobación adicional, se acepta H_a (**Nota:** no es un error de redacción, aunque suene anti-intuitivo para un clínico, simplemente, así funciona el sistema). Se denomina nula porque es la que se pretende rechazar y para rechazarla o no se evalúa la probabilidad de que los valores que arroja el estudio (o unos valores más extremos) sean consistentes desde el punto de vista estadístico con el modelo que se propuso en la hipótesis nula. Si la probabilidad de que los valores obtenidos sean consistentes con la H_0 es baja, se rechaza H_0 y se acepta H_a como cierta. Por tradición, de una manera posiblemente arbitraria, se acepta que un valor de 0.05 -5%- o menos es bajo. Se aclara que otra causa para rechazar H_0 es que no se haya cumplido con otros supuestos subyacentes con la hipótesis (validez metodológica del estudio).

El estudio encontró una diferencia en la frecuencia de cáncer de mama entre usuarias y no usuarias de anticoncepción hormonal. La calidad metodológica o validez interna del estudio es buena porque su riesgo de sesgo en general es bajo, es decir, se puede aceptar que se cumplen muchos de los supuestos que acompañan H_0 del estudio. La prueba estadística realizada para evaluar H_0 de que no existe diferencia arrojó un valor de $p=0.02$, el cual se puede interpretar como que la probabilidad de haber encontrado esa diferencia o una mayor, aceptando que H_0 es cierta, es del 2%. Si se parte del supuesto de que H_0 es verdadera, una probabilidad de solo 2% de haber encontrado esa diferencia o una mayor lo convierte en un hallazgo infrecuente para los límites tradicionales (0.05), por lo tanto, se concluye que es poco probable que los resultados obtenidos sean consistentes con esa hipótesis, en consecuencia, se rechaza H_0 y se acepta H_a , que propone que sí existe una diferencia en el riesgo de cáncer de mama con el uso de anticoncepción hormonal.

Técnicamente la definición del valor p es: “probabilidad, bajo un modelo estadístico específico, de que un resumen estadístico de los datos (por ejemplo, la diferencia de medias de la muestra entre dos grupos comparados) sea igual o más extremo que su valor observado” (American Statistical Association Statement on Statistical Significance and P-Values. 2016). Proporciona un enfoque para resumir la incompatibilidad entre un conjunto particular de datos y un modelo propuesto para estos datos. El contexto más común es un modelo, construido con un conjunto de supuestos, junto con una «hipótesis nula». Algunas definiciones erróneas encontradas con frecuencia en la literatura médica y en algunos textos de bioestadística que pretenden facilitar la comprensión del concepto son: “probabilidad de que la hipótesis nula sea verdad” o “probabilidad de que los resultados hayan sido producidos solo por el efecto del azar”; y algunas interpretaciones erróneas frecuentes son “un valor de p igual o menor de 0.05 significa que la hipótesis nula es falsa” o “un resultado no significativo ($p>0.05$) que la hipótesis en prueba es verdadera”. Es más recomendable intentar comprender el concepto tal como es y no caer en el facilismo de estas falsas interpretaciones, aunque lo primero sea más difícil.

Estadísticamente significativo no es sinónimo de clínicamente significativo

Alvan Feinstein, un maestro de la epidemiología, afirmó que la abdicación de responsabilidad intelectual más impresionante de los clínicos del siglo XX (situación que no ha cambiado en estos inicios del XXI) ante los retos clínicos fue la relegación de las decisiones sobre significación e importancia a los modelos matemáticos. Afirmó que, a diferencia de los juicios para definir “alto”, “viejo”, “gordo”, “grande”, en los que la respuesta juiciosa de todos, o la mayoría, sería “depende”, cuando se presenta un estudio para definir qué es significativo, la respuesta inmediata es

0.05. Otra figura de la epidemiología, Kennet Rothman, afirmó que las pruebas de significación en lugar de servir como herramienta para un pensador se convirtieron para algunos en un sustituto torpe para el pensamiento. El mismo padre de las pruebas de significación, Ronald Fischer, pensaba que ningún trabajador científico tiene un nivel fijo de significación en el que, de año en año, en todas las circunstancias, rechaza las hipótesis; y que este trabajador más bien prefiere concentrarse en cada caso particular para su evidencia y sus ideas. A pesar de estos y cientos de llamados de atención similares, se puede afirmar que la generalidad en la interpretación es aceptar o rechazar una hipótesis simplemente por un valor de p o por la inclusión o exclusión de la igualdad en un intervalo de confianza y, con ello, dar recomendaciones clínicas.

Siempre será posible encontrar una significación estadística: la estrategia más frecuente es aumentar el tamaño de la muestra. Cualquier diferencia por insignificante que sea desde el punto de vista clínico podrá llegar a ser estadísticamente significativa (es decir, tendrá valores de p menores a 0.05) si se aumenta de manera suficiente el tamaño de la muestra. Por otro lado, si se quiere sustentar que dos intervenciones son similares o que un factor no tiene efecto, se reduce el tamaño de la muestra y, con esto, una diferencia numérica y clínicamente importante, se puede clasificar como no significativa ($p > 0.05$). Lo correcto es que los investigadores planearan sus estudios pensando en la relevancia clínica; sin embargo, la gran cantidad de conflictos de interés hace que no siempre sea así. La defensa contra esto es pensar cuál es la diferencia mínima para considerar que un factor o una intervención son importantes si se quiere demostrar superioridad o presencia de un efecto, por un lado, o cuál es la diferencia máxima para considerar que dos intervenciones son similares o un factor no tiene efecto para estudios que buscan equivalencia, no inferioridad o demostrar que un factor no tiene efecto y después, más que centrarse en los valores de p o en utilizar los IC de confianza como pruebas de hipótesis, utilizar toda la información que aportan los IC e integrar con el conocimiento adicional disponible del tema.

Asociación vs Causalidad

Las pruebas de significación estadística confirman una asociación, pero esa presencia no es suficiente para asegurar que existe una relación de causalidad. No hay duda de que existe asociación entre el incremento del perímetro abdominal y el infarto de miocardio, pero no hay una asociación causal. La asociación es la cuantificación del grado de la relación entre un factor y un efecto y, para configurar una relación de causalidad, es necesario que se cumplan otras condiciones más allá de la asociación y que se resuman en los criterios de Bradford Hill. No es necesario que una asociación los cumpla todos para configurar el nexo de causalidad y el único obligatorio es el de temporalidad, que en el estudio en cuestión se exploró con la valoración del momento del uso y con la identificación de cuáles, con seguridad, habían usado los anticonceptivos antes del inicio del cáncer. La aplicación de los otros criterios al estudio: constancia o consistencia (existen múltiples estudios, con diferentes metodologías, reunidos en meta-análisis, que han encontrado el incremento del riesgo del cáncer de mama con los anticonceptivos orales, incluso en una magnitud similar); especificidad (es más fácil aceptar una relación causa-efecto cuando para un efecto solo se plantea una única etiología; sin embargo, esto aplica más en infecciones y actualmente se acepta la multicausalidad y la multiplicidad de efectos de un factor, como es el caso del cáncer de mama); relación dosis-respuesta (los autores hacen evaluación sobre el efecto de los incrementos de la dosis y la duración del uso y se confirma que la fuerza de la asociación se aumenta con el aumento de ellos, lo cual refuerza ese papel causal de los anticonceptivos hormonales); plausibilidad biológica (desde hace varios años se reconoce el papel cancerígeno tanto de los estrógenos como de los progestágenos, reconocido por el Centro Internacional de Investigaciones sobre el Cáncer IARC); coherencia (los hallazgos del estudio no contradicen lo conocido en la historia natural de la enfermedad, en la que se aceptan otros factores hormonales ni con la evolución histórica de la frecuencia de la enfermedad, según las fluctuaciones en el uso de hormonas por las mujeres); experimentación (aun-

que no es ético realizar un ensayo clínico para confirmar el daño de una intervención, encontrar el efecto deletéreo en ensayos clínicos diseñados con otro objetivo refuerza el papel causal de la asociación. Acá se podría extrapolar del efecto encontrado en los ensayos clínicos de terapia de reemplazo hormonal en la posmenopausia, que igualmente han encontrado asociación entre hormonas y cáncer de mama); analogía (si un factor de riesgo produce un efecto a la salud, otro con características similares debería producir el mismo impacto, en nuestro caso no tengo un ejemplo para sustentar este criterio, pero es claro que no es necesario cumplirlos todos). Como se puede ver, la relación evaluada tiene alta probabilidad de ser causal. Adicional a estos criterios, los autores deben garantizar desde el diseño y luego en el análisis de los datos que la relación encontrada no sea explicada por sesgos o deficiencias metodológicas, y una de esas precauciones es evaluar el efecto de otros factores de riesgo que, de manera simultánea, pudieran estar actuando en estas mujeres.

El fenómeno de la confusión

En el estudio en cuestión se llega a la conclusión de que las mujeres que usaron anticoncepción hormonal desarrollaron cáncer de mama con mayor frecuencia que las no usuarias; sin embargo, si se encuentra que las usuarias de anticoncepción hormonal a la vez tenían más edad, más antecedente familiar de cáncer de mama, mayor frecuencia de obesidad, menor edad de menarca, entre otras, se podría cuestionar si efectivamente el incremento del riesgo es dado por el uso de la anticoncepción hormonal o lo es por la presencia de estos otros factores de riesgo. Si por el contrario, esas características fueran más frecuentes en el grupo de no usuarias, se podría pensar que el riesgo realmente pudiera ser mayor en las usuarias porque, a pesar de tener menos riesgo, desarrollaron con más frecuencia el cáncer. Este fenómeno es el denominado confusión y todas estas otras características que pudieran enmascarar (confundir) el efecto real de la exposición se denominan variables confusoras.

Es un reto para el investigador controlar la confusión en investigación. Se puede controlar desde el diseño mediante restricción, cuando se excluyen personas con determinados factores de riesgo para evitar la influencia de esos factores; mediante emparejamiento, especialmente en estudios de casos y controles en los que, para cada caso expuesto a un factor, se selecciona un control sin la enfermedad, pero expuesto al mismo factor; y mediante asignación al azar de las intervenciones, estrategia exclusiva de los ensayos clínicos. Excepto para la asignación al azar, es imposible controlar todos los potenciales factores confusores durante el diseño, por lo tanto, es necesario hacerlo en el análisis de los datos, lo cual se denomina ajuste, que se logra mediante la estandarización, que solo permite controlar una variable mediante estratificación y, con mayor frecuencia, mediante el modelamiento matemático con modelos multivariados.

Los métodos de ajuste comparan la frecuencia del resultado entre subgrupos de participantes en igualdad de condiciones por los factores confusores, por ejemplo, mayores de 40 años usuarias y no usuarias, por un lado, y menores de 40 años usuarias y no usuarias, por otro (análisis estratificado), pero esto se complica cuando se tienen varios factores, en cuyo caso se presentarán situaciones como mayores de 40 años, con antecedente familiar, obesas, usuarias y no usuarias y otras menores de 40 años, con antecedente familiar, obesas, usuarias y no usuarias y así habría que proceder con una inmensa cantidad de combinaciones de factores que crean subgrupos con riesgos diferentes, lo que no resulta práctico para un análisis estratificado y, entonces, se acude a la aplicación de modelos multivariados como la regresión lineal, la regresión logística, la regresión de Cox u otros.

En el estudio que se está analizando los autores contaron con información para ajustar por edad, año calendario de la presentación del cáncer, nivel educativo de la mujer, antecedente de síndrome de ovario poliquístico, endometriosis, paridad e historia familiar temprana de cáncer de mama u ovario. En estas condiciones, al lector le corresponde evaluar si se tuvieron en cuenta todos los potenciales factores

confusores conocidos para hacer un juicio con respecto a lo completo del ajuste y si se considera que faltaron algunos se pudiera cuestionar la validez del resultado. Los investigadores, conscientes de esa posibilidad, y dado que no siempre es posible medir todas las variables confusoras, hicieron un esfuerzo para estimar cuál es el efecto residual de la confusión no controlada. Los autores reconocieron que no tuvieron información y, por lo tanto, no pudieron controlar el efecto de algunas características como edad de la menarca, lactancia, consumo de alcohol y, de manera parcial, el índice de masa corporal. Por otro lado, siempre existe el efecto de los factores de riesgo (confusores) existentes pero que aún son desconocidos, los cuales solo son controlables en los ensayos clínicos porque el ensamble de los grupos de comparación, mediante distribución al azar, tiene la posibilidad de organizar dos grupos homogéneos en promedio para todas las variables confusoras medidas, no medidas y no conocidas. Los autores de la investigación hacen unas estimaciones y consideran que para que un factor confusor pudiera modificar los resultados obtenidos, sería necesario que fuera muy frecuente en la población y tener fuerte influencia en el desarrollo del cáncer, lo que es poco probable.

Conclusión

Existe suficiente evidencia para creer que la anticoncepción hormonal se asocia con incremento en el riesgo de desarrollar cáncer de mama; sin embargo, el riesgo en términos absolutos es bajo, por lo tanto, no existe indicación para suspender de manera masiva el uso de ellos, pero sí es necesario tenerlo en cuenta en el momento de hacer el balance de riesgos y beneficios. Posiblemente, para una mujer que va a obtener un beneficio más allá de la anticoncepción, la balanza está en favor de la anticoncepción hormonal, pero para otra que no tenga ese beneficio adicional posiblemente sea al contrario. Esta debe ser una decisión conjunta entre un proveedor de salud que entienda la información y una mujer o una pareja debidamente informadas. ■



Referencias bibliográficas

1. Mørch LS, Skovlund CW, Hannaford PC, Iversen L, Fielding S, Lidegaard Ø. Contemporary hormonal contraception and the risk of breast cancer. *N Engl J Med* 2017;377: 2228-39.
 2. Feinstein AR. Invidious comparisons and unmet clinical challenges. *Am J Med.* 1992 Feb;92(2):117-20.
 3. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping Doctors and Patients Make Sense of Health Statistics. *Psychol Sci Public Interest.* 2007 Nov;8(2):53-96.
 4. Forrow L1, Taylor WC, Arnold RM. Absolutely relative: how research results are summarized can affect treatment decisions. *Am J Med.* 1992 Feb;92(2):121-4.
 5. Greenland S, Senn S, Rothman K, Carlin J, Poole C, Goodman S, Altman D. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol* (2016) 31:337–350.
 6. Agoritsas T, Merglen A, Shah N, O'Donnell M, Guyatt G. Adjusted Analyses in Studies Addressing Therapy and Harm Users' Guides to the Medical Literature. *JAMA.* 2017;317(7):748-759.
-