



Soluciones para implementar un buen método en la investigación clínica

ESCUELA SEMI DE INVESTIGACIÓN CLÍNICA



Área de Formación Online SEMI-FEMI
Sociedad Española de Medicina Interna (SEMI)
Fundación Española de Medicina Interna (FEMI)

© Sociedad Española de Medicina Interna (SEMI)
© Fundación Española de Medicina Interna (FEMI)

Todos los textos, imágenes y documentos presentes en esta publicación son propiedad intelectual de SEMI. Se autoriza a visualizar e imprimir estos materiales mientras sean respetadas las siguientes condiciones:

1. Los textos, imágenes y documentos solo pueden ser utilizados con fines informativos.
2. Los textos, imágenes y documentos no pueden ser utilizados para propósitos comerciales.
3. Cualquier copia de estos textos, imágenes y documentos, o de parte de los mismos, deberá incluir esta advertencia de derechos reservados y el reconocimiento de la autoría.

15 de julio de 2021

ISBN: 978-84-09-32407-1

Edita: Multimédica Proyectos, S.L.

C/. Menéndez Pidal, 27 bajo. 28036 Madrid

campussemi@multimedica proyectos.com

www.multimedica proyectos.com

Dr. Pablo Pérez Martínez

Director Científico del Instituto Maimónides
de Investigación Biomédica de Córdoba (IMIBIC)
Catedrático de Medicina, Universidad de Córdoba
Especialista de Medicina Interna
Hospital Universitario Reina Sofía (Córdoba)

Dr. Emilio Casariego Vales

Jefe de Servicio de Medicina Interna
Hospital Universitario Lucus Augusti (Lugo)
Expresidente de la Sociedad Española de Medicina Interna
y de la Fundación Española de Medicina Interna



directores
editores



prólogos

Es un gran honor presentar el segundo libro de la Escuela de Investigación de la Sociedad Española de Medicina Interna (SEMI), *Soluciones para implementar un buen método en la investigación clínica*. El avance de la ciencia es el motor más potente del progreso y la mejora de nuestra calidad de vida, y ha sido un elemento clave para transformar la medicina.

Es bien sabido que no es tarea fácil para el médico mantener actualizadas sus competencias profesionales. El mejor modelo para estar implicado en la generación de conocimiento es el de aquellos profesionales que complementan su actividad asistencial con la investigación clínica. Esa dualidad de utilizar el conocimiento aprendido de otros y de impulsar nuevos avances no es sencilla y exige unas cualidades y un esfuerzo especial para conseguirlo. Este tipo de médico permite que los nuevos conocimientos lleguen antes a sus pacientes. Por ello uno de los objetivos de esta Escuela, así como del libro que aquí presentamos, es aproximar a todos los internistas los medios e instrumentos que permiten generar investigación clínica e innovación de alta calidad orientada a los pacientes. En este escenario, el segundo libro de la Escuela ha sido diseñado para continuar enriqueciendo y afianzando los conocimientos adquiridos en el módulo II del Curso Online de Investigación Clínica, “Investigación traslacional de éxito”.

A lo largo de los once capítulos que componen la obra, el lector profundizará en ciertos aspectos relacionados con el diseño, el análisis y la interpretación de diferentes estudios clínicos, que le permitirán conocer las fortalezas y debilidades de los grandes tipos de estudios, cuyos resultados aplicamos en la práctica clínica diaria a la hora de tomar decisiones. El contenido se plantea de forma amena y concisa para el lector. Los editores creemos que el resultado final es brillante y queremos agradecer a los autores por su extraordinario trabajo.

Todo ello no hubiera sido posible sin el apoyo constante de Boehringer Ingelheim, siempre en su apuesta firme y continua por la formación de calidad y por desarrollar proyectos innovadores. Por ello, queremos expresar nuestro máximo reconocimiento a su equipo médico impulsor de esta obra. Nuestro agradecimiento también a la Junta Directiva de la SEMI, que sigue trabajando y dando su apoyo a diferentes acciones, como este libro, que permitan potenciar la investigación y la innovación en el seno de nuestra sociedad científica. Finalmente queremos reconocer la labor de Emilia Carretero, cuya colaboración inestimable ha permitido una vez más que podamos disfrutar de una nueva obra.

Dr. Pablo Pérez Martínez Dr. Emilio Casariego Vales

Directores y Editores

Escuela SEMI de Investigación Clínica



En Boehringer Ingelheim trabajamos para desarrollar soluciones que den respuesta a necesidades médicas no cubiertas. Así, hace más de 130 años que aportamos valor a través de nuestra investigación, que tiene como objetivo último poner a disposición de la sociedad medicamentos innovadores que cambien la vida de los pacientes, siempre manteniendo nuestro compromiso con la sostenibilidad.

Boehringer Ingelheim dedica gran parte de sus recursos a la investigación de fármacos: desde el descubrimiento de nuevas entidades químicas o biológicas hasta los ensayos clínicos en todas sus fases, sin olvidar el desarrollo galénico del medicamento. Una verdadera carrera de fondo.

La investigación clínica es, por tanto, un pilar fundamental de nuestra actividad llevada a cabo a nivel mundial en colaboración con investigadores de prestigio que desarrollan su trabajo con los más altos estándares de calidad. En este sentido, cabe destacar que España y sus investigadores contribuyen de manera muy importante en los ensayos clínicos internacionales promovidos por nuestra compañía en términos de aportación de conocimiento y de reclutamiento de pacientes.

Por este motivo, y por la estrecha e histórica relación de colaboración entre Boehringer Ingelheim y la Sociedad Española de Medicina Interna (SEMI), nos sentimos muy orgullosos de haber colaborado en la creación de la Escuela de Investigación Clínica SEMI y de haber puesto en marcha el primer curso para internistas dedicado íntegramente a la investigación clínica. Creemos que esta es una de nuestras contribuciones a la formación de los profesionales sanitarios que revertirá sin duda en una mayor calidad asistencial y, en consecuencia, en un beneficio para los pacientes.

Afianzamos así nuestro compromiso con la especialidad de Medicina Interna para recorrer juntos nuevos caminos de colaboración en algo tan importante y estratégico para ambas instituciones como es la investigación y el desarrollo de nuevos medicamentos y nuevas soluciones para mejorar la salud de las personas.

Dra. Elena Gobartt

Directora Médica de Boehringer-Ingelheim España

contenidos

CAPÍTULO 1

Ensayo clínico

Dra. Carmen Suárez Fernández 14

CAPÍTULO 2

Metaanálisis

Dra. María Rosa Bernal López
Dr. Ricardo Gómez Huelgas 28

CAPÍTULO 3

Revisión sistemática

Dra. Juana Carretero Gómez 44

CAPÍTULO 4

Análisis de mortalidad

Dr. José Pablo Miramontes González
Dr. Javier Martín Vallejo 58

CAPÍTULO 5

COVID-19 y factores pronósticos

Dr. José Manuel Casas Rojo 78

CAPÍTULO 6

Seguimiento a largo plazo de una cohorte (Framingham/UKPDS)

Dra. Pilar Román Sánchez 92

CAPÍTULO 7

Los estudios ecológicos para confirmar y establecer la evidencia de una hipótesis

Dr. Francisco Pérez Jiménez **110**

CAPÍTULO 8

Importancia de los subestudios

Dra. Isabel Martín-Garrido
Dra. Lourdes Moreno Gaviño **126**

CAPÍTULO 9

Cómo comparar grupos: *propensity score*

Dr. Martín Fabregate Fuente
Dra. Nuria Bara Ledesma
Dr. Luis Manzano Espinosa **138**

CAPÍTULO 10

BMJ, un artículo diferente

Dr. Javier García Alegría **150**

CAPÍTULO 11

Prevalencia de COVID-19 y tres series: datos del Ministerio de Sanidad en la epidemia

Dr. Manuel Crespo Blanco **162**

Ensayo clínico

Dra. Carmen Suárez Fernández
Jefe de Servicio de Medicina Interna
y de la Unidad de Hipertensión Arterial y Riesgo Vascular
Hospital Universitario de La Princesa (Madrid)

*La publicación seleccionada de este capítulo
es un trabajo ficticio basado en datos no reales,
creado por la autora como base para el desarrollo del tema del capítulo.*



capítulo 1

Publicación seleccionada

Dosis bajas de minibaxán en pacientes muy mayores con fibrilación auricular

Introducción

La prescripción de un tratamiento anticoagulante oral para la prevención del ictus en pacientes muy mayores con fibrilación auricular es un desafío, ya que, a pesar de que se benefician de este tratamiento, muchas veces no se les prescribe por miedo al sangrado.

Minibaxán, anticoagulante oral directo inhibidor del factor X activado, ha demostrado en un ensayo clínico de morbimortalidad su no inferioridad en eficacia y seguridad en comparación con la warfarina para la prevención del ictus y el embolismo sistémico en pacientes con fibrilación auricular a dosis de 40 mg/día, o 20 mg si está indicada una reducción de dosis, manteniendo su perfil de eficacia y seguridad incluso en el subgrupo de pacientes mayores. Se desconoce si dosis inferiores a las exploradas podrían ser eficaces y seguras en esta población de muy alto riesgo trombótico y hemorrágico.

El objetivo del presente ensayo es evaluar la eficacia y la seguridad de dosis inferiores a las aprobadas de minibaxán en pacientes mayores con fibrilación auricular y de alto riesgo trombótico y hemorrágico en comparación con la no administración de tratamiento anticoagulante.

Métodos

Ensayo clínico de fase 3, multicéntrico, aleatorizado, doble ciego y controlado con placebo para comparar el uso de 10 mg de minibaxán cada 24 horas frente a placebo en pacientes japoneses de 80 años o más con fibrilación auricular que no se consideraron candidatos adecuados para la terapia con anticoagulantes orales en las dosis aprobadas para la prevención del ictus por una o más de las siguientes razones: aclaramiento de creatinina en el rango de 15-30 mL/min, antecedente de sangrado en un área crítica o gastrointestinal, bajo peso (≤ 45 kg) y uso de antiinflamatorios no esteroideos o de antiagregantes plaquetarios.

Los pacientes fueron aleatorizados por bloques y estratificados según la escala CHADS₂ (≤ 2 o ≥ 3) en relación 1:1 para 10 mg de minibaxán o placebo.

Criterio de valoración principal de eficacia: Combinación de ictus y embolia sistémica.

Criterio de valoración principal de seguridad: Hemorragia mayor según la definición de la Sociedad internacional de Trombosis y Hemostasia (ISTH, por sus siglas en inglés).

VARIABLES SECUNDARIAS DE EFICACIA: Compuesto de ictus, embolia sistémica y muerte por causa cardiovascular; compuesto de ictus, embolia y muerte por cualquier causa; beneficio clínico neto (combinación de ictus, embolia sistémica, hemorragia mayor y muerte por cualquier causa).

VARIABLES SECUNDARIAS DE SEGURIDAD: Combinación de hemorragia mayor y hemorragia no mayor clínicamente relevante; sangrado no mayor clínicamente relevante; todo sangrado.

Análisis estadístico: Ensayo basado en eventos, con un objetivo de 65 eventos de ictus o embolia sistémica. De acuerdo con estudios previos de incidencia anual de ictus y embolia sistémica, se requerirían aproximadamente 400 pacientes por grupo para una potencia del 80% para mostrar la superioridad del minibaxán frente al placebo con un nivel de significación bilateral del 5%.

El análisis de eficacia principal se realizó en la población por intención de tratar.

La población de seguridad incluyó a todos los pacientes que recibieron al menos una dosis de minibaxán o placebo.

Para el análisis de tiempo hasta el primer evento de ictus o embolismo se utilizó un modelo de riesgos proporcionales de Cox con grupos de prueba y puntuación CHADS₂ (≤ 2 o ≥ 3) como covariables con un nivel de significación bilateral del 5%. Se estimó el riesgo relativo con el uso de ratios de riesgo con intervalos confianza al 95%. Los eventos se estimaron mediante el método Kaplan-Meier.

Resultados

Se inscribieron en 164 instituciones un total de 1086 pacientes, de los que 984 fueron aleatorizados (492 recibieron minibaxán, y 492 placebo). Un total de 102 pacientes fueron excluidos: 20 retiraron el consentimiento, 3 fallecieron y 79 no cumplieron con los criterios de elegibilidad.

La duración media de la participación en el ensayo fue de 466 días (rango intercuartil: 293,5-708). Dos pacientes en el grupo placebo se retiraron antes, tras recibir la primera dosis, y no fueron incluidos en la población de seguridad; 301 pacientes adicionales no completaron el ensayo (178 retiraron el consentimiento antes de la finalización del ensayo, 117 murieron durante el periodo de ensayo y 8 discontinuaron la participación por otras razones). Las razones más comunes para la discontinuidad fueron los eventos adversos no relacionados con el sangrado y la pérdida de motivación o falta de capacidad para seguir participando en el ensayo.

En la **tabla 1** se muestran las características de los pacientes incluidos.

Se comunicaron 66 eventos de ictus o embolia sistémica; 7 eventos no fueron adjudicados como primarios por el comité de eventos. Por tanto, se incluyeron 59 eventos

Tabla 1. Características clínicas y demográficas de los pacientes

Características	Todos (N = 984)	Minibaxán (n = 492)	Placebo (n = 492)
Edad años	86,6 ± 4,2	86,74,2 ± 4,2	86,4 ± 4,3
≤85 años n (%)	447 (45,4)	218 (44,3)	229 (46,5)
>85 años n (%)	537 (54,6)	274 (55,7)	263 (53,5)
Sexo masculino n (%)	419 (42,6)	212 (43,1)	207 (42,1)
Peso kilogramos	50,6 ± 11,0	50,6 ± 10,9	50,6 ± 11,1
Fibrilación auricular paroxística n (%)	463 (47,1)	237 (48,2)	226 (45,9)
Aclaramiento de creatinina mL/min	36,3 ± 14,4	36,3 ± 4,3	36,2 ± 14,5
≤50 mL/min n (%)	823 (83,6)	415 (84,3)	408 (82,9)
>50 mL/min n (%)	161 (16,4)	77 (15,7)	84 (17,1)
Escala CHADS₂	3,1 ± 1,1	3,0 ± 1,13	3,1 ± 1,1
≤2 n (%)	363 (36,9)	181 (36,8)	182 (37,0)
≥3 n (%)	621 (63,1)	311 (63,2)	310 (63,0)
Ictus o accidente isquémico transitorio previo n (%)	236 (24,0)	110 (22,4)	126 (25,6)
Insuficiencia cardiaca n (%)	533 (54,2)	259 (52,6)	274 (55,7)
Diabetes mellitus tipo 2 n (%)	225 (22,9)	115 (23,4)	110 (22,4)
Hipertensión arterial n (%)	810 (82,3)	412 (83,7)	398 (80,9)
Escala HAS-BLED	2,3 ± 0,9	2,3 ± 0,9	2,4 ± 0,9
Enfermedad coronaria n (%)	257 (26,1)	130 (26,4)	127 (25,8)
Historia de caídas n (%)	340 (34,6)	154 (31,3)	186 (37,8)
Pacientes frágiles n (%)	402 (40,9)	185 (37,6)	217 (44,1)
No pudieron ser evaluados n (%)	17 (1,7)	7 (1,4)	10 (2,0)
Falta de datos n (%)	23 (2,3)	11 (2,2)	12 (2,4)

en el análisis. En el análisis por intención de tratar, 15 pacientes (2,3% por paciente-año) en el grupo de minibaxán y 44 pacientes (6,7% por paciente-año) en el grupo de placebo presentaron un ictus o un embolismo sistémico (*hazard ratio* [HR]: 0,34; intervalo de confianza al 95% [IC95%]: 0,19-0,61; $p < 0,001$ para superioridad).

En la **tabla 2** se muestran las tasas de incidencia de las variables primarias y secundarias de eficacia y de seguridad.

Tabla 2. Variables primarias y secundarias de eficacia y seguridad

Variable	Número de pacientes con evento (porcentaje por paciente-año)		Hazard ratio (IC95%)	p
	Minoxabán 10 mg (n = 492)	Placebo (n = 492)		
Variable primaria de eficacia				
Ictus o embolismo sistémico	15 (2,3)	44 (6,7)	0,34 [0,19-0,61]	<0,001
Ictus	12 (1,8)	40 (6,0)	0,30 [0,16-0,57]	
Isquémico	12 (1,8)	39 (5,9)	0,31 [0,16-0,59]	
Hemorrágico	0	2 (0,3)		
Mortal	1 (0,1)	3 (0,4)	0,34 [0,04-3,30]	
Embolismo sistémico	3 (0,4)	6 (0,9)	0,50 [0,13-2,01]	
Variables secundarias de eficacia				
Ictus, embolismo sistémico o muerte de causa cardiovascular	52 (7,8)	72 (10,9)	0,72 [0,50-1,03]	
Eventos cardiovasculares mayores (MACE)	51 (7,7)	72 (11,0)	0,70 [0,49-1,01]	
Ictus, embolismo sistémico o muerte de cualquier causa	74 (11,1)	98 (14,8)	0,75 [0,56-1,02]	
Beneficio clínico neto	87 (13,5)	103 (15,6)	0,86 [0,65-1,15]	
Muerte por cualquier causa	66 (9,9)	69 (10,2)	0,97 [0,69-1,36]	
Variable primaria de seguridad				
Sangrado mayor	20 (3,3)	11 (1,8)	1,87 [0,90-3,89]	0,09
Hemorragia intracraneal	2 (0,3)	4 (0,6)	0,50 [0,09-2,72]	
Sangrado gastrointestinal	14 (2,3)	5 (0,8)	2,85 [1,03-7,88]	
Variables secundarias de seguridad				
Sangrado mayor o no mayor clínicamente relevante	97 (17,7)	62 (10,7)	1,65 [1,20-2,27]	
Sangrado no mayor clínicamente relevante	81 (14,5)	52 (8,9)	1,62 [1,14-2,30]	
Todos los sangrados	241 (63,0)	202 (45,0)	1,35 [1,12-1,63]	

Conclusiones

En pacientes muy mayores japoneses con fibrilación auricular no candidatos adecuados para dosis estándar de anticoagulantes orales, 10 mg de minibaxán una vez al día fue superior al placebo en la prevención del ictus y la embolia sistémica sin aumento significativo de hemorragia mayor con respecto a placebo. La incidencia de hemorragia mayor fue mayor con minibaxán que con placebo, aunque esta diferencia fue no significativa. Hubo más eventos de hemorragia gastrointestinal, así como sangrado, con minibaxán.

Planteamiento

Todo proyecto de investigación debe iniciarse revisando la literatura, actualizando los conocimientos sobre el tema e identificando las lagunas de conocimiento sobre las que se elaborará la pregunta de investigación.

En el caso del artículo presentado, los autores realizan el siguiente planteamiento: la fibrilación auricular es una arritmia muy frecuente, un importante factor de riesgo para ictus, cuya frecuencia aumenta con la edad. La anticoagulación reduce el riesgo de ictus y embolismo sistémico asociado a esta arritmia. Estos pacientes son los que tienen un mayor riesgo trombótico, aunque también tienen mayor riesgo hemorrágico, y se benefician claramente de la anticoagulación; sin embargo, muchos de ellos no reciben anticoagulantes por miedo al sangrado. El riesgo de sangrado se incrementa con la dosis de anticoagulante, pero no existen estudios en personas mayores que demuestre que dosis más bajas de anticoagulante, con inducción de menos sangrado, sean eficaces para reducir el ictus o la embolia sistémica. Y para evaluarlo diseñan el ensayo clínico aquí presentado.

Por lo tanto, los primeros pasos que se deben contemplar son los siguientes:

- Describir la situación actual del problema terapéutico.
- Describir brevemente los conocimientos actuales sobre el fármaco o la intervención.
- Justificar la necesidad del ensayo en su conjunto en relación con lo actualmente conocido.
- Justificar el diseño.
- Justificar el tipo de pacientes en que se realiza el estudio.

Formulación de la pregunta clínica

Una buena pregunta clínica debe ser factible, interesante, novedosa, ética y relevante.

Se ha propuesto un formato para la definición de la hipótesis y los objetivos, conocido con el acrónimo PICOT, que deben incluir los siguientes aspectos: la población diana (**p**opulation), la intervención (**i**ntervention), el grupo comparativo (**c**omparison group), la variable final (**o**utcome) y el tiempo (**t**ime).

La pregunta clínica bien formulada es si la administración de dosis más bajas de las recomendadas de un anticoagulante oral directo a pacientes mayores con fibrilación auricular y alto riesgo de sangrado puede reducir la incidencia de ictus y embolismo sistémico, sin incrementar significativamente el riesgo de hemorragia, frente a la no anticoagulación. Su justificación es que los pacientes mayores con fibrilación auricular y alto riesgo de sangrado no anticoagulados siguen teniendo un alto riesgo de ictus y embolismo sistémico.

Justificación del diseño y el análisis estadístico realizado

El diseño más adecuado para dar contestación a la pregunta formulada es un ensayo clínico. Un ensayo clínico se define, según S. J. Pocock, como "cualquier tipo de experimento planificado que involucra a pacientes y que es diseñado para dilucidar el tratamiento más apro-

piado de futuros pacientes con una enfermedad determinada". Se utilizan los resultados en una muestra limitada de pacientes para hacer inferencias en la población general de pacientes que requerirá tratamiento en el futuro.

La prueba de oro de la investigación clínica es el ensayo clínico controlado, que permitiría establecer la causalidad, y que corresponde al presentado.

El ensayo clínico controlado evalúa la respuesta terapéutica obtenida en un grupo de pacientes que reciben una intervención experimental (por ejemplo, un fármaco) en comparación con la respuesta obtenida en un grupo control que no recibe dicha intervención, como es nuestro caso, o recibe otro tipo de intervención o fármaco.

El uso del placebo como control presenta algunas ventajas:

- Permite evaluar la "eficacia absoluta" del tratamiento en estudio.
- Aísla el efecto del tratamiento de otros componentes de la respuesta terapéutica.

La asignación a los tratamientos (o intervención) de la comparación tiene como objetivo distribuir a los participantes en el ensayo clínico en dos grupos que sean lo más comparables posible. Existen distintos tipos de asignación, aunque la más empleada es la asignación aleatoria (aleatorización o *randomization*), un procedimiento sistemático y reproducible por el cual los sujetos participantes en un ensayo clínico son distribuidos al azar en los distintos grupos de tratamiento.

Sus objetivos son los siguientes:

- Disminuir los sesgos de selección (rechazar pacientes por conocer a qué tratamiento serán asignados).
- Evitar sesgos en la asignación a los tratamientos.
- Homogeneidad entre los grupos de la comparación.

A pesar de la aleatorización ciega, puede seguir existiendo mala distribución de las variables.

Tipos de asignación aleatoria:

1. Simple: sencilla y simple y totalmente imprevisible. Como inconvenientes presenta el riesgo de desigualdad entre grupos o centros y el riesgo de no homogeneidad entre grupos.
2. Por bloques, que asegura el balance entre grupos.
3. Estratificada, que permite asegurar la igualdad entre grupos respecto a variables pronósticas de interés. Los factores que influyen en el pronóstico deben estar bien identificados.

En nuestro estudio se ha utilizado una aleatorización por bloques y estratificada por la principal variable predictora de ictus o embolismo sistémico, como es la escala CHADS₂.

El enmascaramiento consiste en la serie de medidas que se toman con el fin de que a lo largo del estudio, ya sea el paciente, el médico o ambos, así como aquellos que evalúan la respuesta a los tratamientos, desconozcan el tratamiento administrado. Su objetivo es evitar los sesgos en la evaluación de la respuesta por el paciente y el médico y permite el uso de placebo.

Estadística en un ensayo clínico

Tiene tres objetivos fundamentales: la determinación del tamaño de la muestra, la definición del tipo de análisis estadístico que se va a realizar y la interpretación de los resultados. Implica varios procesos y tareas a lo largo de todo el estudio:

- Antes de empezar el ensayo (diseño):
 - Definición de la variable principal y las secundarias.
 - Cálculo del tamaño de la muestra.
- Durante la realización del ensayo:
 - Recogida de datos.
 - Análisis intermedios.
- Al final del ensayo
 - Comparación de las características basales.
 - Análisis de la variable principal y las secundarias.
 - Tipos de análisis: test de significación estadística, intervalos de confianza.
 - Efecto relativo frente a efecto absoluto.
 - Análisis por intención de tratar frente al análisis por protocolo.
 - Análisis de subgrupos.

Definición de las variables

Deben definirse dos tipos de variables:

- *Variable principal*: Es recomendable que exista una única variable principal adecuada para medir el objetivo principal.
- *Variables secundarias*: Todas las que se estimen convenientes para evaluar objetivos secundarios.

Tamaño de la muestra

Una muestra insuficiente no es capaz de detectar efectos clínicos relevantes y no tiene el poder estadístico necesario, por lo que puede llevar a resultados no válidos o conclusiones erróneas (intervalo de confianza muy amplio). Una muestra excesiva supone un derroche de tiempo y recursos y presenta problemas éticos al exponer a riesgo a más pacientes de los necesarios.

Existen diferentes programas para calcular el tamaño de la muestra (algunos son sencillos y de libre difusión):

- GRANMO (IMIM, Barcelona): <https://www.imim.es/ofertadeserveis/softwarepublic/granmo/index.html>.

- EPIDAT 4.0 (Xunta de Galicia): <http://www.sergas.es/Saude-publica/EPIDAT-4-2>.
- StatsToDo: https://www.statstodo.com/ResourceIndex_Categories.php.
- Aplicaciones para el móvil: SampSize, etc.
- Programas más complejos dirigidos a expertos en estadística.

Estadística al final del estudio

- Comparación de las características basales: comprobar que los dos grupos son homogéneos; si los factores pronósticos son distintos en un grupo en comparación con el otro puede ser necesario hacer algún ajuste estadístico para compararlos.
- Estadística descriptiva.
- Comparación de las variables del estudio. Es importante no solamente valorar la significación estadística, sino su relevancia clínica.
- Test de significación estadística. Su objetivo es rechazar la hipótesis nula (H_0), de no diferencia entre los tratamientos, y calcular la probabilidad (p) de que la diferencia observada sea explicada por el azar. Cuanto menor sea esta probabilidad mayor será la evidencia en contra de H_0 (se considera significativa si $p < 0,05$).

Tipos de análisis:

- Por intención de tratar, o según la asignación aleatoria aunque no cumplan o cambien de grupo (aproximación pragmática): es el más adecuado. Infraestima la eficacia del tratamiento, pero acerca más la evaluación a las condiciones habituales y preserva los beneficios de la aleatorización.
- Por protocolo o de casos válidos, es decir, aquellos pacientes que completan el estudio de acuerdo al protocolo (aproximación explicativa o de eficacia): es útil para conocer la eficacia en pacientes que no han presentado efectos adversos. Puede conducir a resultados erróneos (por confusión), sobre todo si el número de pacientes perdidos es importante.

Interpretación del artículo. Fortalezas y debilidades

La pregunta planteada es pertinente y relevante desde el punto de vista clínico.

El diseño elegido de ensayo clínico es el adecuado. Se trata de un ensayo clínico doble ciego controlado con placebo, un tipo de control que podría justificarse ante la falta de evidencia de un estándar de cuidado establecido para pacientes muy mayores con alto riesgo de ictus y sangrado, aunque las guías aconsejan anticoagular a estos pacientes de forma individualizada.

Los pacientes no tienen una contraindicación para la anticoagulación, sino que no son buenos candidatos por su alto riesgo de sangrado. Sin embargo, disponemos de evidencias que demuestran el beneficio clínico neto de la anticoagulación oral en estos pacientes.

Muchos de los pacientes incluidos en este ensayo serían tratados con dosis bajas aprobadas en condiciones de práctica clínica habitual, ya que no presentan contraindicación para ello.

Los autores no estratifican a los pacientes de acuerdo con el número de condiciones que desaconsejan la anticoagulación, aspecto pronóstico muy importante que condiciona que la población no sea homogénea y que habría que haberse tenido en cuenta para su estratificación, y no solo el CHADS₂.

Los pacientes asignados a placebo tenían más insuficiencia cardíaca, mayor riesgo de sangrado y mayor fragilidad, potentes factores pronósticos. Ello puede suponer un sesgo en la asignación de tratamientos que condicione la interpretación de los resultados.

Otro aspecto que se debe valorar es el sesgo por uso inadecuado de retiradas y abandonos. Se introduce cuando no se tienen en cuenta ni los pacientes ni las razones, similares o diferentes, por las que los pacientes son apartados o abandonan el estudio.

Una de las limitaciones del estudio es el alto número de pacientes que interrumpieron el ensayo debido a sus antecedentes de alto riesgo. Sin embargo, no se perdió ningún paciente durante el seguimiento, y solo seis retiraron el consentimiento debido a preocupaciones relacionadas con el sangrado. La mayoría de los pacientes que se retiraron lo hicieron debido a eventos adversos no relacionados con el sangrado o por mala funcionalidad. Para valorar este hecho sobre los resultados debe realizarse un análisis separado que tenga en cuenta el potencial de que la discontinuación del seguimiento influya en los resultados mostrados. Por lo tanto, aunque se realizó un análisis por intención de tratar, debe tenerse en cuenta el aspecto previo.

Los resultados muestran una reducción importante del ictus isquémico (reducción del riesgo relativo del 70 %) a expensas de incrementar por dos o tres todos los sangrados, particularmente el gastrointestinal.

Llama la atención que la incidencia de ictus o embolismo sistémico y de sangrado es muy parecida a la obtenida en el ensayo fundamental cuando se compara con los resultados del subanálisis de pacientes frágiles.

Aplicabilidad en la práctica clínica

El ensayo clínico presentado no tiene aplicabilidad clínica ni debe modificar las recomendaciones actuales. La pregunta de gran interés clínico no está contestada con este ensayo y sería interesante realizar un ensayo clínico bien diseñado para darle contestación.

Quizá habría sido más correcto establecer tres grupos para comparar la relación riesgo-beneficio: dosis baja aprobada (20 mg/día), dosis reducida no aprobada (10 mg/día) y placebo.

Adicionalmente, la población elegida no es extrapolable a otras poblaciones, ya que la población japonesa tiene mayor incidencia de ictus y de sangrado con el tratamiento anticoagulante que otras poblaciones de otras zonas geográficas.

Bibliografía

Higgins JPT, Altman DG, Gøtzsche PF, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011 Oct 18;343:d5928. doi: 10.1136/bmj.d5928.

Martín Andrés A, Luna del Castillo JD. *Bioestadística para las ciencias de la salud*. 4.^a ed. Las Rozas (Madrid): Ediciones Norma; 1994. Capítulo IX, Ensayos clínicos.

Preguntas de autoevaluación

1. Indique cuál de las siguientes no es una característica esencial de una pregunta clínica:

- a) Es factible.
- b) Es relevante.
- c) Es novedosa.
- d) Puede extrapolarse a una amplia población.

La respuesta correcta es la opción *d*.

Una buena pregunta clínica debe ser factible, interesante, novedosa, ética y relevante.

2. ¿Cuál de los siguientes contenidos es menos importante incluir en la formulación del objetivo?

- a) La población a la que va dirigida.
- b) El tipo de intervención.
- c) La dosis del fármaco.
- d) El comparador.

La respuesta correcta es la opción *c*.

Para la definición de la hipótesis y los objetivos deben incluirse los siguientes aspectos (formato PICOT): la población diana sobre la que se va a intervenir (*population*), el tipo de intervención (*intervention*), el grupo comparativo, es decir, con qué se va a comparar (*comparison group*), la variable final o la variable sobre la que se quiere medir el impacto de la intervención (*outcome*) y el tiempo (*time*). En las intervenciones farmacológicas no es necesario definir la dosis en el objetivo, sino en la metodología.

3. Indique cuál de las siguientes afirmaciones es incorrecta:

- a) La aleatorización por bloques asegura más el balance entre los grupos.
- b) La asignación aleatoria simple presenta el riesgo de desigualdad entre grupos, pero no entre centros, en el caso de un ensayo clínico multicéntrico.
- c) Los factores que influyen en el pronóstico deben estar bien identificados y han de utilizarse correctamente para realizar una asignación aleatoria estratificada, que es la que asegura una mayor igualdad entre grupos.
- d) La asignación aleatoria no siempre garantiza la homogeneidad de los grupos.

La respuesta correcta es la opción *b*.

La asignación aleatoria por bloques asegura el balance entre grupos. Es importante no realizar bloques muy pequeños ni muy grandes, y el investigador no debe conocer el tamaño de los bloques.

La asignación aleatoria simple es sencilla de realizar, pero totalmente imprevisible, lo que condiciona el riesgo de desigualdad entre grupos y también entre centros, así como el riesgo de no homogeneidad entre grupos.

La asignación aleatoria estratificada permite asegurar la igualdad entre grupos respecto a variables pronósticas de interés. Los factores que influyen en el pronóstico deben estar bien identificados para incluirse en la estratificación.

La asignación aleatoria no siempre garantiza la homogeneidad de los grupos. Puede seguir existiendo mala distribución de las variables a pesar de la aleatorización ciega.

Metaanálisis

Dra. María Rosa Bernal López

Research Coordinator of Clinical Management Unit of Internal Medicine
Biomedical Research Laboratory (IBIMA)
Hospital Regional Universitario de Málaga
Instituto de Investigación Biomédica de Málaga (IBIMA)
Universidad de Málaga (Málaga)

Dr. Ricardo Gómez Huelgas

Jefe de Servicio de Medicina Interna
Hospital Regional Universitario de Málaga
Expresidente
Sociedad Española de Medicina Interna
Fundación Española de Medicina Interna

*La publicación seleccionada de este capítulo
es un trabajo ficticio, basado en datos no reales y datos reales (véase la bibliografía),
creado por los autores como base para el desarrollo del tema del capítulo.*



capítulo 2

Publicación seleccionada

Introducción

Nos planteamos realizar un metaanálisis partiendo de una pregunta de interés: "Metformina y vitamina D: ¿es un buen tratamiento sinérgico para el control metabólico de la obesidad?". El objetivo es resumir y sintetizar la evidencia publicada sobre esta pregunta y evaluar el nivel de evidencia por resultados tras una revisión del tema.

Planteamiento

La vitamina D se obtiene mediante producción endógena en la piel y con la alimentación, por lo que podría intervenir en el desarrollo de la obesidad. Se ha documentado que individuos con niveles bajos de vitamina D presentan un mayor exceso de peso y grasa corporal que aquellos que tienen cantidades suficientes de esta hormona. Por otro lado, esta vitamina estimula la liberación de la insulina desde el páncreas, por lo que los individuos con concentraciones bajas de vitamina D aumentan su riesgo de desarrollar diabetes. Existen tratamientos para pacientes con diabetes mellitus tipo 2 (DM2) capaces de prevenir la obesidad. Uno de los antidiabéticos orales que ha demostrado que ofrece una importante protección contra la obesidad es la metformina.

Pregunta clínica

En los pacientes con obesidad y con DM2, ¿es la vitamina D una hormona importante *per se* para el control de la resistencia a la insulina a través de la activación de la expresión de su receptor en el tejido adiposo? ¿Podrían la metformina y la vitamina D actuar sinérgicamente en la producción de un efecto protector contra la DM2 en este tipo de pacientes?

Metodología

Para responder a esta pregunta clínica se plantea buscar publicaciones científicas recientes sobre el tratamiento que se va a estudiar (vitamina D y metformina en la obesidad) en bases de datos concretas (PRISMA, PubMed, etc.) con palabras clave que identifican la pregunta clínica (“obesidad”, “diabetes mellitus tipo 2”, “vitamina D”, “metformina”).

Resultados

Un total de 32 publicaciones reúnen los criterios de calidad científica para ser analizados en el metaanálisis. Tras analizar los resultados de estas publicaciones, se observa que se estudian los tratamientos, la vitamina D y la metformina, en un gran número de participantes, mayores de 18 años con obesidad, en presencia/ausencia de DM2, pero de distintas razas.

Conclusiones

La pregunta clínica planteada no ha quedado resuelta, ya que existe un sesgo de publicación (publicaciones con resultados negativos sobre el tratamiento con vitamina D y metformina) y la metodología utilizada en cada uno de los estudios seleccionados es heterogénea, por lo que los resultados no son extrapolables a otras poblaciones.

Concepto de metaanálisis

Para definir qué es un metaanálisis, primero debemos definir qué entendemos por investigar. Investigar es una actividad formal y ordenada de pensar orientada a la obtención de nuevos conocimientos y su aplicación para la solución a problemas o situaciones de interés. La investigación es la base del conocimiento científico, por lo que se hace necesario que la investigación se realice de forma sistemática, con unos objetivos claros y cuyos resultados, analizados de forma objetiva, puedan ser replicados.

Por ello, un metaanálisis es llevar a cabo una investigación sistemática y exhaustiva sobre un tema o pregunta científica muy específicos, unificando y evaluando todos los resultados seleccionados de la literatura que puedan responder a nuestra pregunta diana. Un metaanálisis es la estimación cuantitativa o cualitativa del efecto de una intervención o exposición a un determinado tratamiento en la que se valoran los sesgos, las fortalezas y las debilidades de los estudios existentes sobre dicha temática. Las técnicas utilizadas en el metaanálisis proporcionan un enfoque sencillo y estandarizado para evaluar, mediante análisis estadísticos, los hallazgos de los resultados que se van a estudiar.

Tipos de metaanálisis

Existen varias formas de abordar la realización de un metaanálisis:

- **Metaanálisis de datos individuales y de datos agregados.** Los metaanálisis se realizan a partir de los resultados publicados de estudios previos. Una alternativa a esta metodología es el metaanálisis de los datos individuales, donde se obtienen los datos de cada estudio y se sintetizan directamente. La fiabilidad de los resultados de metaanálisis mejora cuando los resultados se elaboran a partir de los datos individuales de todos los participantes de un estudio, pues esto permite emplear metodología estadística más sofisticada. Con los datos individuales es posible realizar análisis de subgrupos o *clusters*.
- **Metaanálisis acumulativo.** El metaanálisis acumulativo consiste en repetir el metaanálisis cuando existe un nuevo estudio acerca del tema diana. Este tipo de metaanálisis puede identificar de forma retrospectiva una pregunta de interés y analizar de forma prospectiva el problema tras realizarse nuevos estudios al respecto.
- **Network metaanálisis.** Es un tipo de metaanálisis en el que se aborda la efectividad relativa de dos problemáticas cuando estas no se han comparado directamente, pero sí se han comparado con otros estudios. Es decir, en este caso se comparan diferentes resultados entre sí a partir de una inferencia estadística.

La guía para autores de la Colaboración Cochrane establece que las comparaciones indirectas de este tipo no son aleatorizadas, sino que son estudios observacionales y, por tanto, pueden tener los sesgos de este tipo de estudios.

- **Metaanálisis en colaboración.** Este tipo de metaanálisis se realiza a partir de los estudios planificados o en desarrollo para un determinado problema. Normalmente se invita a los investigadores principales de cada estudio a proveer los resultados de los pacientes individuales una vez completado el trabajo. A veces estos metaanálisis colaborativos se realizan a partir de estudios ya publicados.

Formulación de la pregunta y justificación del diseño

La pregunta clínica se debe plantear según la prioridad del investigador, es decir, ¿se necesita una respuesta a corto plazo? ¿Es más prevalente el problema de salud que nos interesa?

En nuestro caso, se trata de prevenir o mejorar el control metabólico de la obesidad con dos tratamientos distintos. Nos planteamos por ello buscar evidencia científica publicada a partir de cierta fecha (últimos 5 años para que sea evidencia lo más reciente posible y se pueda evaluar lo más novedoso del tratamiento que se va a estudiar en nuestra pregunta clínica, el uso de la vitamina D y la metformina en la obesidad) y en bases de datos concretas (PRISMA, PubMed, etc.) con palabras clave que identifican la pregunta clínica (“obesidad”, “diabetes mellitus tipo 2”, “vitamina D”, “metformina”). Encontramos 32 publicaciones científicas.

Para nuestra pregunta clínica, se selecciona el metaanálisis de datos individuales, ya que se analizará la evidencia científica que presentan sus datos individualmente.

Tres estudios representativos de la búsqueda son los siguientes:

- Estudio longitudinal aleatorizado con tres grupos de estudio (placebo y dos grupos de tratamiento con dos dosis distintas de vitamina D) y de 5 años de duración en 7538 participantes mayores de 18 años, de raza caucásica, con obesidad y con modificación de estilo de vida.
- Estudio longitudinal de 12 meses de duración sobre 464 pacientes de 35 a 55 años de edad, de raza caucásica, con obesidad y DM2 en tratamiento con una única dosis de metformina.
- Ensayo clínico aleatorizado con cuatro grupos de estudio (placebo y tres grupos de tratamiento: vitamina D, metformina y vitamina D más metformina) en 25278 pacientes mayores de 18 años, de raza caucásica y oriental, con presencia y ausencia de obesidad y DM2.

Fortalezas y debilidades del metaanálisis

Debemos analizar las debilidades y limitaciones que presentan los metaanálisis.

Fortalezas

Varios investigadores han argumentado que el metaanálisis es una herramienta valiosa para mejorar nuestro entendimiento sobre un tema concreto y evaluar el grado en que la investigación apoya distintas hipótesis basadas en la teoría. Su gran ventaja es que nos permite trabajar con mucha información procedente de gran cantidad de estudios previos, de una forma bien estructurada y mediante el método científico.

Así, las fortalezas que caracterizan al metaanálisis son las siguientes:

- Los resultados obtenidos proporcionan una mejor estimación acerca de nuestra pregunta de interés sobre la población que la que ofrecen los estudios individuales.
- La precisión y validez de las estimaciones se pueden mejorar a medida que se utilizan más datos en un metaanálisis, es decir, según aumenta la cantidad de datos, aumenta el poder estadístico para detectar un efecto concreto.

- Se pueden analizar las incongruencias en los resultados entre los estudios, y se pueden analizar las bases de esas diferencias (sesgos de criterios de estudio, tamaño muestral, etc.).
- La prueba de hipótesis se puede aplicar a estimaciones resumidas.
- Se pueden identificar e incluir moderadores en los análisis para explicar la variación entre los estudios.

Debilidades

Aunque hemos enumerado las ventajas del metaanálisis, también existen varias debilidades o limitaciones asociadas con esta metodología. El metaanálisis es una técnica relativamente reciente, por lo que algunos de sus aspectos metodológicos aún son objeto de controversia. Por ello, la inexactitud de los procedimientos metaanalíticos aceptados puede dar lugar a inferencias erróneas sobre la relación entre variables.

Los problemas con el metaanálisis pueden estar relacionados con distintos aspectos:

- La selección de los estudios que se incluyen en el análisis (sesgo de selección). Es muy importante definir exhaustivamente los criterios de inclusión y exclusión para la búsqueda y revisión de estudios, y que estos sean lo más objetivos posible. El propio autor del metaanálisis puede sesgar los resultados, ya que los resultados de un estudio pueden haber influido en su criterio para incluirlo o excluirlo del análisis.
- La validez de los estudios seleccionados para el análisis.
- El sesgo de publicación, es decir, los estudios con resultado positivo tienen mayor probabilidad de acabar publicados. Esto condiciona los resultados de la búsqueda bibliográfica y da lugar a resultados sesgados en un metaanálisis.
- Tamaños muestrales pequeños.
- La heterogeneidad de los métodos (variabilidad clínica y sociodemográfica, entre otros) utilizados en los estudios incluidos en el análisis.

En resumen, la agregación de datos de diferentes estudios aumenta la precisión y exactitud de los resultados, pues se asume que la diferencia entre estos obedece al azar. Pero las diferencias en dirección o magnitud del efecto del tratamiento pueden deberse a otros factores, incluyendo ligeras diferencias en el tratamiento, las poblaciones, la variable de resultado, y el diseño y la calidad del estudio. Por lo tanto, los metaanálisis pueden generar resultados equívocos al ignorar una eventual heterogeneidad significativa entre estudios, al incorporar los sesgos de los ensayos clínicos aleatorizados individuales, así como introduciendo otros sesgos en el proceso de buscar los estudios.

Interpretación del artículo (fortalezas y debilidades)

Fortalezas

El metaanálisis que se ha elaborado se ha basado en evidencia científica con una definición exhaustiva de los criterios de inclusión y exclusión de la búsqueda [a través de las palabras claves "obesidad", "diabetes mellitus tipo 2", "vitamina D", "metformina"].

Los estudios realizados que se han escogido para realizar el metaanálisis no han sesgado los resultados negativos y han sido objetivos en sus conclusiones finales.

La evidencia científica seleccionada para la elaboración del metaanálisis ha analizado grupos de estudio aleatorizados, con grupo control y con un gran tamaño muestral (elevado número de pacientes estudiados). Esto conlleva una gran fortaleza estadística en el metaanálisis elaborado para que se pueda concluir una respuesta a la pregunta clínica (la metformina y vitamina D sobre el control metabólico en la obesidad).

Debilidades

Se ha encontrado solo evidencia científica cuyas conclusiones son positivas en cuanto a la efectividad de los tratamientos (vitamina D y metformina) sobre la pérdida de peso. Esto podría sobrevalorar las conclusiones del metaanálisis, ya que no se ha encontrado evidencia de la no efectividad del tratamiento. Es por ello que se cuenta con un sesgo de selección que se debe tener siempre presente en cuanto a las conclusiones del metaanálisis.

Se debe tener en cuenta la metodología empleada en cada evidencia científica seleccionada. En el metaanálisis realizado se incluyen estudios realizados en distintas razas (caucásica y oriental). Esto también debilita las conclusiones que se pueden extraer del metaanálisis, ya que los resultados obtenidos no se pueden extrapolar a otras razas no estudiadas (negra).

¿Cómo se hace un metaanálisis?

La realización de un metaanálisis es de gran ayuda cuando nos enfrentamos a la necesidad de investigar aspectos en los que los estudios previos muestran cierto grado de controversia entre sí, cuando queremos investigar qué grado de relación tienen distintas variables o cuando queremos valorar la efectividad de un programa o tratamiento. Así, en medicina se emplean este tipo de revisiones para explorar, investigar, priorizar y aplicar tratamientos basándose en la evidencia científica. Hay que tener en cuenta que su aplicación solo es viable cuando los estudios primarios que se escogen tienen una perspectiva cuantitativa y cuentan con muestras y condiciones homogéneas o muy similares entre sí, y obviamente que partan al menos del mismo objetivo y pregunta principal.

Realizar un metaanálisis no es sencillo y requiere un seguimiento estructurado. Las principales fases en la realización de metaanálisis son las siguientes:

- **Formulación del problema o pregunta de investigación.** Este primer paso es el más importante. Es la formulación del problema o la pregunta diana que queremos investigar y analizar. Esta debemos definirla de forma exhaustiva teniendo en cuenta las diversas variables que queremos evaluar. En función de esta pregunta se comienza a estructurar el metaanálisis y se elaboran los criterios de inclusión y exclusión de los estudios que se van a incluir.
- **Búsqueda bibliográfica y selección de estudios.** Este paso es uno de los más largos y tediosos. Es la búsqueda activa de la bibliografía que vamos a analizar de manera conjunta con el fin de hacer una buena revisión de nuestra temática de investigación. Para ello, siguiendo nuestros criterios de inclusión y de exclusión, seleccionaremos aquellos

estudios que cumplen los requisitos para poder ser valorados en el metaanálisis. Entre los criterios deben tenerse en cuenta el tipo de diseño que tenga cada estudio, cómo se han medido los resultados y las principales características de las muestras empleadas, así como las metodologías experimentales. Otro aspecto importante que se debe tener en cuenta es la fecha de publicación de los estudios seleccionados, ya que se recomienda no utilizar datos de estudios con más de diez años de antigüedad.

Para la búsqueda suele resultar recomendable emplear bases de datos indexadas tales como PubMed o Scopus, aunque se pueden emplear también otro tipo de fuentes documentales o sistemas de búsqueda.

- **Codificación de estudios.** Una vez seleccionados los estudios que van a formar parte del metaanálisis, es necesario realizar un manual de codificación en el que se registren la fiabilidad de los datos y las variables estudiadas. Debe incluirse toda variable moderadora que pueda influir en los resultados, así como indicar cómo se han valorado para que otros investigadores puedan reproducir los resultados obtenidos.
- **Análisis estadístico e interpretación.** El cuarto paso es el análisis estadístico de los resultados. En esta fase se valoran y se discuten las implicaciones o interpretaciones de los resultados. Este análisis estadístico suele evaluar el efecto de los diferentes estudios comparados, analizar si existe o no heterogeneidad entre ellos y, en caso de que esta exista, analizar qué variables pueden explicar dicha heterogeneidad a partir de pruebas como el análisis de varianza (ANOVA, paquete estadístico SPSS).
- **Publicación.** Tras la interpretación de los resultados, se redacta y publica el metaanálisis. Esta publicación debe constar de distintos apartados: introducción (con el marco teórico, la hipótesis y los objetivos del metaanálisis), método (exhaustivo y detallado para que otros autores puedan replicarlo; debe incluir las palabras clave, así como los sistemas de búsqueda de los estudios, el manual de codificación y el tipo de análisis realizado), resultados y discusión (donde se interpreta el metaanálisis de acuerdo con los resultados obtenidos).

Interpretación de los metaanálisis

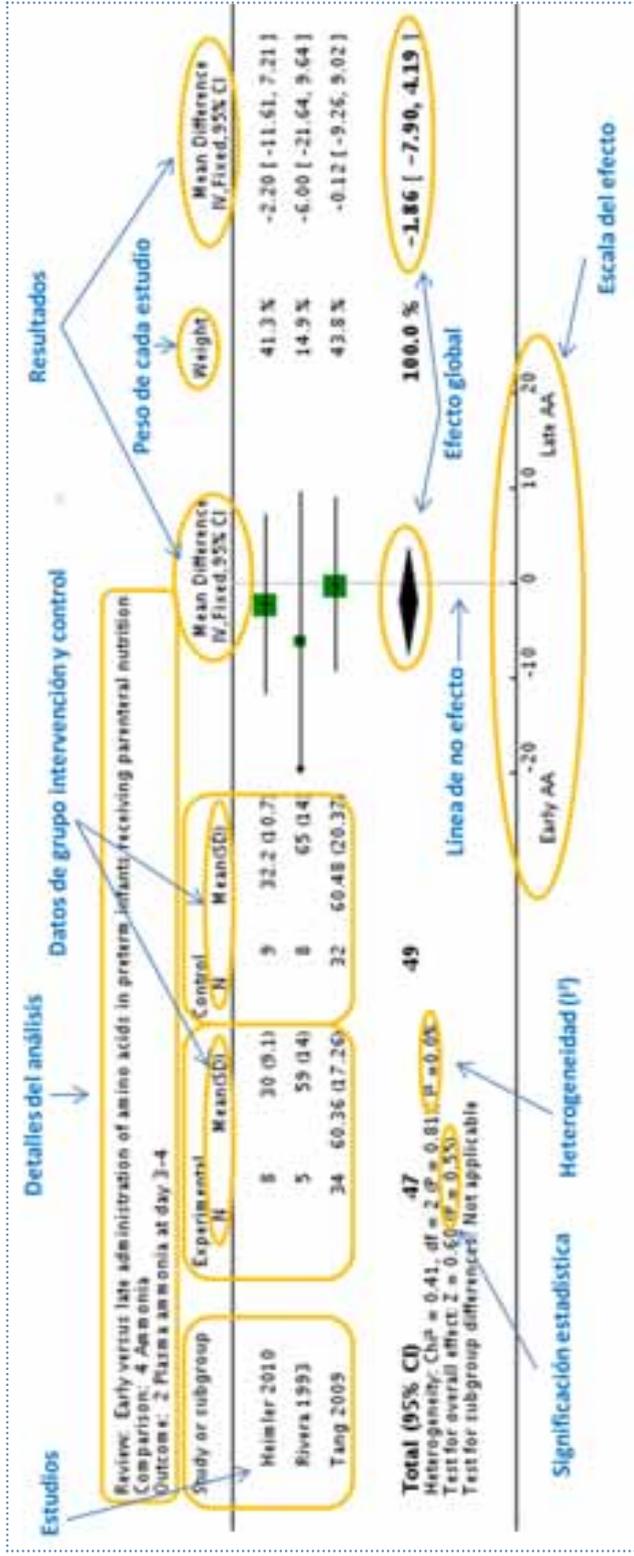
Los resultados de un metaanálisis se organizan a través de un gráfico llamado *forest plot* (diagrama de bosque), como el que se muestra en la **figura 1**. Se trata de una visualización gráfica de los resultados obtenidos en los estudios seleccionados que forman parte del metaanálisis, para comprender mejor cómo estructurar e interpretar los resultados del metaanálisis.

Normalmente, un *forest plot* se divide en seis columnas y los resultados individuales de cada estudio se disponen en filas sucesivas.

En la primera columna se identifican los estudios seleccionados e incluidos en el metaanálisis indicando el autor del estudio y el año de publicación. Todos los estudios incluidos deben haberse realizado con una metodología y un diseño experimental semejantes.

La segunda columna representa a los grupos de intervención y la tercera a los grupos de control. Estas dos columnas pueden presentar subcolumnas, que hacen referencia al

Figura 1. Descripción de los componentes de un forest plot



Fuente: Martínez Rodríguez R. Cómo interpretar un metaanálisis. [Actualizado el 4 de marzo de 2015]. En: Evidencia en Nutrición [Internet]. [Pamplona]: CAEC-FEDNI; 2015. Disponible en: <http://fedn.es/blog/evidencianutricion/como-interpretar-un-metaanalisis>.

tamaño muestral (N) y a la media y la desviación estándar (*mean*, *SD*) o los riesgos relativos (incluidos los valores mínimo y máximo), así como la significación estadística (p) de los resultados obtenidos en cada uno de los estudios que se valoran.

En la cuarta columna se representa un gráfico de los resultados de los distintos estudios. La línea vertical central de esta representación es la línea de no efecto, la cual simboliza que no hay diferencia entre los grupos de estudio (intervención frente a control, en este caso). A ambos lados de la línea se representa si los resultados favorecen o no la pregunta de estudio. Habitualmente, a la izquierda de la línea de no efecto se suele representar si se favorece el evento estudiado.

Dentro de la gráfica, los cuadrados representan el efecto evaluado en cada estudio y su tamaño es directamente proporcional al peso del estudio dentro del metaanálisis (expresado numéricamente en la quinta columna).

Por otro lado, la línea horizontal que atraviesa los cuadrados indica los intervalos de confianza. Cuanto más larga es la línea horizontal, mayores son estos intervalos, lo que indica la imprecisión de los resultados del estudio. Además, si existe una flecha en los extremos de la línea horizontal, se indica que el intervalo de confianza es mayor que el espacio disponible. Como se apuntaba antes, los efectos individuales (los cuadrados), se posicionarán a izquierda, derecha o sobre la línea de no efecto dependiendo del resultado de cada estudio. Si la mayoría de los efectos se posicionan a un lado de la línea de no efecto, se indica que la heterogeneidad interestudio es baja, pero si los efectos se distribuyen a ambos lados, significa que los estudios son contradictorios, lo que apunta que las conclusiones del metaanálisis pueden ser dudosas.

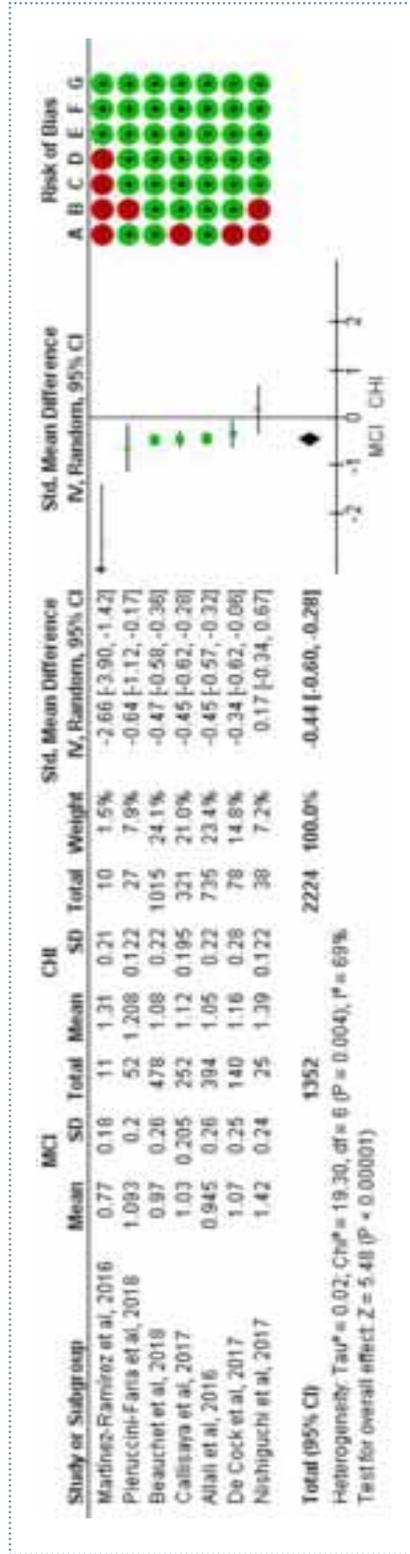
Los rombos representan los resultados globales del metaanálisis. El centro del rombo es el valor del efecto en conjunto y su ancho representa el intervalo de confianza general. Si el rombo se posiciona claramente a un lado de la línea de no efecto, indica que la diferencia entre los grupos estudiados puede considerarse estadísticamente significativa, pero si la cruza o simplemente la roza no se podrán sacar conclusiones que apunten en una dirección.

Por último, en la sexta columna se indican los resultados numéricos y los intervalos de confianza de cada uno de los estudios seleccionados. Un intervalo de confianza del 95% ($p = 0,05$) representa una seguridad del 95% de que la asociación estudiada no se da por azar.

Al final de esta presentación de datos, el *forest plot* incluye la prueba de heterogeneidad que evalúa la variabilidad entre estudios, e indica cómo de comparables son los estudios analizados. Finalmente, en el *forest plot* se muestra un apartado que presenta el test de efecto global, que mide la significación estadística del resultado del metaanálisis. Un resultado se considera estadísticamente significativo cuando el valor de la p es inferior a 0,05 (para un intervalo de confianza del 95%).

En otros metaanálisis, la disposición de las columnas puede variar, pero siempre muestran el mismo tipo de resultados mencionados anteriormente (véase como ejemplo el de la **figura 2**).

Figura 2. Ejemplo de forest plot con un diseño distinto



Fuente: Fuentes-Abolafio J, Stubbs B, Pérez-Belmonte LM, Bernal-López MR, Gómez-Huelgas R, Cuesta-Vargas A. Functional objective parameters which may discriminate patients with mild cognitive impairment from cognitively healthy individuals: a systematic review and meta-analysis using an instrumented kinematic assessment. Age Ageing. 2021;50(2):380-93. DOI: 10.1093/ageing/afaa135.

En resumen, en un *forest plot* hay que fijarse en los siguientes aspectos:

- La heterogeneidad de los efectos.
- El peso de los estudios.
- Los intervalos de confianza.
- La disposición del rombo.
- El test de efecto global.

Valoración de su aplicabilidad en la práctica clínica

El metaanálisis elaborado nos indica que no se puede aplicar a la práctica clínica diaria, ya que la pregunta clínica planteada no ha quedado resuelta, pues no se han incluido los resultados negativos del tratamiento (vitamina D y metformina) y no todos los estudios han desarrollado la misma metodología exhaustiva (pacientes de distintas razas), y los resultados no son extrapolables a otras razas (negra) que viven en otras zonas geográficas (África) con mayor abundancia de días de sol pero con carencias de vitamina D en la población.



Bibliografía

Fuentes-Abolafio IJ, Stubbs B, Pérez-Belmonte LM, Bernal-López MR, Gómez-Huelgas R, Cuesta-Vargas A. Functional objective parameters which may discriminate patients with mild cognitive impairment from cognitively healthy individuals: a systematic review and meta-analysis using an instrumented kinematic assessment. *Age Ageing*. 2021;50(2):380-93. DOI: 10.1093/ageing/afaa135.

Higgins JPT, Green S, eds. *Cochrane handbook for systematic reviews of interventions* 4.2.4. Chichester: John Wiley; 2005.

Lau J, Antman EM, Jimenez-Silva J, Kupelnick B, Mosteller F, Chalmers TC. Cumulative meta-analysis of therapeutic trials for myocardial infarction. *N Engl J Med*. 1992;327(4):248-54. doi: 10.1056/NEJM199207233270406.

Martínez Rodríguez R. Cómo interpretar un metaanálisis. [Actualizado el 4 de marzo de 2015]. En: *Evidencia en Nutrición* [Internet]. [Pamplona]: CAEC-FEDN; 2015. Disponible en: <http://fedn.es/blog/evidencianutricion/como-interpretar-un-metaanálisis>.

Ones D, Viswesvaran C, Schmidt F. Realizing the full potential of psychometric meta-analysis for a cumulative science and practice of human resource management. *Human Resource Manag Rev*. 2017;27(1):201-15. DOI: 10.1016/j.hrmmr.2016.09.011.

Sandoya E. Metaanálisis: fortalezas y debilidades. *Rev Urug Cardiol*. 2008;23:263-75.

Stone DL, Rosopa PJ. The advantages and limitations of using meta-analysis in human resource management research. *Human Resource Manag Rev*. 2017;27(1):1-7. DOI: 10.1016/j.hrmmr.2016.09.001.

Preguntas de autoevaluación

1. Indique cuál de las siguientes afirmaciones sobre el concepto de metanálisis es verdadera:

- a) Un metaanálisis es llevar a cabo una investigación sistemática y exhaustiva sobre un tema o pregunta científica muy específicos, unificando y evaluando todos los resultados seleccionados de la literatura que puedan responder a la pregunta científica.
- b) Para evaluar un metaanálisis, los análisis estadísticos no son importantes.
- c) En un metanálisis no existen sesgos en la selección de los estudios que se van a valorar si el motor de búsqueda es PubMed o Scopus.
- d) Los estudios deben ser siempre experimentales.

La respuesta correcta es la opción *a*.

Para llevar a cabo un metaanálisis se debe realizar una investigación sistemática y exhaustiva sobre la pregunta científica que se pretende responder.

2. Indique cuál de las siguientes afirmaciones sobre las debilidades o limitaciones de los metanálisis es incorrecta:

- a) El sesgo de selección y publicación de los estudios incluidos en un metanálisis puede contaminar las conclusiones.
- b) En un metaanálisis es una debilidad importante la heterogeneidad de los métodos utilizados en los estudios incluidos.
- c) Estudios con pequeños tamaños muestrales debilitan los resultados del metaanálisis.
- d) La precisión y la validez de las estimaciones se pueden mejorar a medida que se utilizan más datos en un metaanálisis; es decir, según se aumenta la cantidad de datos, independientemente de su homogeneidad, aumenta el poder estadístico para detectar un efecto concreto.

La respuesta correcta es la opción *d*.

Cuantos más datos homogéneos se incluyan en la realización de un metaanálisis, mayor es su fortaleza.

3. Indique cuál de las siguientes afirmaciones sobre la interpretación de los metaanálisis es correcta:

- a) Para una buena interpretación del metaanálisis se debe tener en cuenta que los estudios comparativos tengan amplios intervalos de confianza.
- b) La disposición del rombo en un *forest plot* indica la fiabilidad del metaanálisis.
- c) En un *forest plot* debemos fijarnos en la heterogeneidad de los efectos, el peso de los estudios, los intervalos de confianza, la disposición del rombo y el test de efecto global.
- d) La disposición de las columnas del *forest plot* en un metaanálisis es fija y nunca debe alterarse el orden.

La respuesta correcta es la opción c.

Es imprescindible tener en cuenta todas las premisas descritas para poder realizar una buena interpretación del metaanálisis de estudio.

Revisión sistemática

Dra. Juana Carretero Gómez
Servicio de Medicina Interna
Hospital Universitario de Badajoz (Badajoz)
Vicepresidenta primera
Sociedad Española de Medicina Interna
Fundación Española de Medicina Interna

*La publicación seleccionada de este capítulo
es un trabajo ficticio basado en datos no reales,
creado por la autora como base para el desarrollo del tema del capítulo.*



capítulo 3

Publicación seleccionada

Modificación de la composición corporal con los nuevos antidiabéticos. Una revisión sistemática

Introducción

Para 2040 se espera una prevalencia mundial de diabetes mellitus tipo 2 (DM2) de 642 millones de personas, mayor entre los individuos de 60 a 79 años. El 20% de las personas entre 70 y 79 años tienen diabetes. En España, la incidencia anual de DM2 es del 11,5%, y aumenta a partir de los 60 años. La prevalencia de obesidad en España ronda el 22% de la población (es mayor en varones de más de 65 años), y la de obesidad abdominal es del 64,7%. Tanto la DM2 como la obesidad son factores de riesgo para el desarrollo de enfermedad cardiovascular. La coexistencia de múltiples factores de riesgo cardiovascular, obesidad abdominal y DM2 es mayor en mujeres (23%) y en mayores de 65 años (42,8%). El 45% de las muertes y el 65% de la pérdida de años de vida ajustados a discapacidad se deben a enfermedades cardiovasculares en personas obesas con DM2. El exceso de tejido adiposo abdominal es predictor de enfermedad cardiovascular independientemente de la grasa corporal total. Los internistas atendemos mayormente a personas mayores de 65 años; de estas, el 30% son diabéticas y hasta en el 40% de los casos ingresan por enfermedad cardiovascular. Según el estudio OBEMI, en consultas de medicina interna la prevalencia de obesidad era del 23,6%, y la edad media de $62,47 \pm 15,27$ años.

Los análogos del receptor del GLP-1 (arGLP-1) y los inhibidores del cotransportador de sodio-glucosa tipo 2 (iSGLT-2) se han posicionado como tratamientos de primera elección para la DM2, especialmente en pacientes con alto o muy alto riesgo cardiovascular o con enfermedad cardiovascular establecida, enfermedad renal diabética o insuficiencia cardíaca. Además, no debe existir limitación del esfuerzo terapéutico en la diabetes por la edad del paciente, sino por su estado funcional o presencia de fragilidad.

Uno de los mecanismos beneficiosos de ambos grupos es la pérdida ponderal, mayor en el caso de los arGLP-1. Ambos, por mecanismos diferentes, inducen pérdidas ponderales a expensas fundamentalmente de la masa grasa, más evidente a nivel abdominal en el caso de los iSGLT-2. Sin embargo, no conocemos cuál es su efecto sobre la composición corporal total, es decir, sobre la masa magra o la masa no grasa corporal.

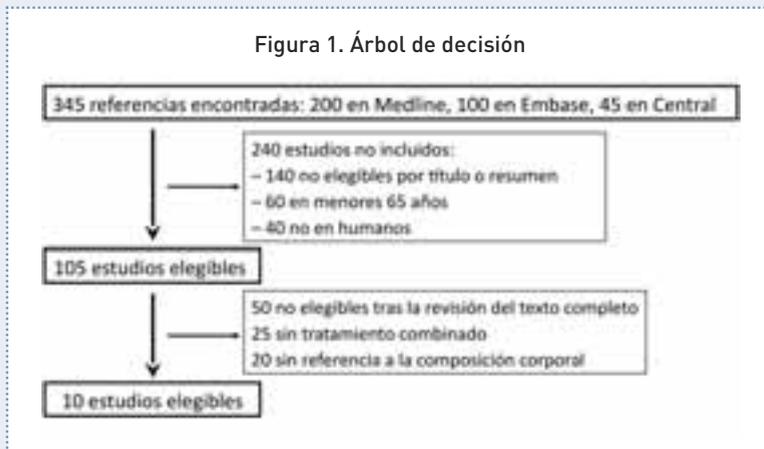
La prevalencia de sarcopenia aumenta con la edad y la inactividad, y llega a alcanzar el 20% de los mayores de 70 años. También es conocida la alta prevalencia de obesidad sarcopénica entre los mayores de 70 años o en pacientes que padecen enfermedad pulmonar obstructiva crónica (EPOC) o insuficiencia cardíaca. La pérdida de masa muscular es importante en ancianos, pues, aparte de la función de soporte

corporal, el músculo es el principal depósito de glucosa y determinante del gasto metabólico en reposo. Una baja masa muscular contribuiría a un peor control glucémico, alto riesgo de hipoglucemias, caídas y fragilidad, unido a la mayor prevalencia de fragilidad y dependencia en ancianos con DM2.

Asumiendo que dicha pérdida ponderal es igual en todos los grupos etáreos, nuestro objetivo es analizar cuál es la influencia sobre la composición corporal total del tratamiento combinado de arGLP-1 e iSGLT-2 en pacientes diabéticos mayores de 65 años.

Material y métodos

Realizamos una revisión sistemática de la literatura. Para ello, analizamos las bases de datos electrónicas Medline, Database y Central. En la **figura 1** se expone el flujo de selección de los artículos.



Resultados

Tras la revisión sistemática seleccionamos 10 artículos, de los que 7 son ensayos clínicos aleatorizados, 2 son metaanálisis y 1 es un estudio observacional. Los resultados, las características de los estudios, los pacientes y los resultados en la composición corporal se exponen en la **tabla 1**.

Discusión

El uso combinado de un arGLP-1 y un iSGLT-2 es cada vez más frecuente en la práctica clínica habitual y está recomendado por las guías de práctica clínica como la combinación de primera elección tras la metformina en prácticamente todas las situaciones

Tabla 1. Resultados poblacionales

Estudio	Fármacos	Edad (años)	Diseño	N	Duración (semanas)
Astrup <i>et al.</i> (2015)	Liraglutida, dapagliflozina	67,2	EC 3 brazos	60	24
Forgs <i>et al.</i> (2017)	Liraglutida, dapagliflozina	70	EC 3 brazos	100	24
Gómez <i>et al.</i> (2017)	Liraglutida, canagliflozina	66,8	EC 5 brazos	200	24
Liii <i>et al.</i> (2016)	Dulaglutida, canagliflozina	69,1	EC 4 brazos	160	26
Hang <i>et al.</i> 2018	Dulaglutida, empagliflozina	70,1	EC 3 brazos	220	24
López <i>et al.</i> (2019)	Semaglutida, canagliflozina	68	EC 3 brazos	180	24
Ohisoioni (2019)	Semaglutida, dapagliflozina	72	EC	200	26
Franks <i>et al.</i> (2018)	Exenatida LAR, dapagliflozina	66,2	EC	100	24
Alvárez <i>et al.</i> (2019)	Todos	71,7	MTA	587	26
Frias <i>et al.</i> (2020)	Todos	70	MTA	100	24
Yang <i>et al.</i> (2018)	Liraglutida, empagliflozina	75,1	OBS	340	26

BIA: bioimpedancia; DXA: densitometría corporal; EC: ensayo clínico; MG: masa grasa; MLG: m

clínicas independientemente de la edad del paciente. Nuestra revisión pone en evidencia que, efectivamente, la combinación de un arGLP-1 con un iSGLT-2 en personas mayores de 65 años con diabetes produce una importante pérdida ponderal, y que esta se produce a expensas de masa grasa frente a masa muscular o masa no grasa, en rangos entre el 64% y el 86% de pérdida ponderal total en forma de grasa. Además, observamos que los resultados son bastante homogéneos entre todos los estudios analizados, con datos más contundentes al asociar la canagliflozina al arGLP-1 frente

y de composición corporal en los trabajos seleccionados

	Técnica de cuantificación de la masa corporal	Peso perdido (kg)	Cambio total de la MG (kg)	Cambio total de la MLG (kg)	Cambio total del agua (kg)	Peso perdido en MG (%)	Peso perdido en MLG (%)
	BIA	15 (9-21)	12 (8-15)	2 (0-3)	1 (0-2)	80 (65-90)	20 (15-23)
	BIA	20 (12-28)	15 (10-22)	2 (0,5-3)	3 (1,2-4,3)	75 (60-82)	25 (18-30)
	BIA	25 (20-28)	18 (13-22)	2 (0,2-2,8)	5 (4,3-6,2)	72 (60-80)	28 (20-35)
	DXA	22 (14-33)	16 (10-22)	1 (0-2)	5 (3,2-6)	72,7 (62,3-80,2)	27,3 (20,1-33)
	DXA	20 (15-26,2)	16 (12-20)	2 (0-3)	2 (0,2-3)	80 (72-91)	20 (17-25)
	DXA, TAC	26 (22-30)	22 (20-25)	2 (1-4)	4 (2,5-5)	84,6 (70-98)	15,4 (10-17)
	BIA, TAC	25 (22-27)	20 (19-22)	2 (1-2,5)	3 (3-4,5)	80 (75-82)	20 (17-22)
	DXA	15 (12-17)	10 (8-11,3)	3 (2-4)	2 (1-2,5)	66,6 (55-70)	33,3 (30-36)
	DXA, BIA	22 (19-24,7)	16 (14,2-17,9)	2 (0,3-4,7)	4 (3-5,1)	72,7 (70-74)	27,3 (25-28,9)
	DXA, BIA	25 (22,5-27)	16 (5-17,7)	4 (2-5)	5 (3-5,9)	64 (62,8-66)	36 (34-37,8)
	BIA	23 (21-24,5)	20 (19-22,5)	0 (-0,03-1)	3 (1,7-3,9)	86,9 (85-88)	13,3 (11,1-15)

masa libre grasa; MTA: metaanálisis; OBS: observacional; TAC: tomografía axial computarizada.

a cualquier otro iSGLT-2, y al incluir en la combinación la semaglutida frente al resto de los arGLP-1. Aunque nuestro estudio no nos permite hacer inferencias ni sacar conclusiones con significación estadística, sí parece razonable pensar que, en el grupo de edad estudiado, la pérdida ponderal obtenida no sería a expensas de masa muscular y, por lo tanto, es esperable pensar que no aumentaría la aparición de sarcopenia entre nuestros pacientes. Por lo tanto, se trataría de una combinación segura para nuestros pacientes.

Planteamiento

Nos planteamos realizar una revisión sistemática de la literatura científica para evaluar cuál es el estado del conocimiento sobre la acción en la composición corporal total producida por los arGLP-1 y los iSGLT-2 usados en combinación en personas mayores de 65 años.

Justificación del diseño

Las revisiones sistemáticas son técnicas de investigación científica en las que la unidad de análisis utilizada son los estudios originales primarios, a partir de los cuales pretendemos contestar a una pregunta de investigación claramente formulada. Para ello, hemos de realizar un proceso sistemático, explícito y exhaustivo; esto es lo que las diferencia de las revisiones narrativas, en las que no se realiza dicho planteamiento, sino simplemente una escritura científica basada sobre todo en la opinión. La revisión sistemática se considera una revisión cualitativa, pues no combina estadísticamente los resultados, lo que a su vez la diferencia del metaanálisis, considerado revisión cuantitativa. Así pues, con nuestra revisión sistemática sintetizaremos los resultados de investigaciones primarias previamente realizadas mediante estrategias que limiten la aparición de posibles sesgos y errores aleatorios.

Estas estrategias serían las siguientes:

- Búsqueda sistemática y exhaustiva de todos los artículos potencialmente relevantes. En una revisión sistemática, los artículos que se utilizarían serían los ensayos clínicos, las revisiones sistemáticas y los metaanálisis. Sin embargo, si queremos evaluar la eficacia de intervenciones comunitarias o de salud pública o resultados de intervenciones a largo plazo, sobre todo para parámetros de seguridad, podemos utilizar estudios observacionales.
- Selección, mediante criterios previamente diseñados y reproducibles, de los artículos incluidos. Este punto es el que le confiere el carácter científico, a diferencia de las revisiones narrativas.
- Descripción del diseño y la ejecución de la revisión, síntesis de los datos obtenidos e interpretación de los resultados de los estudios originales seleccionados.

Además, es recomendable tener previamente realizado un protocolo con los métodos que vamos a utilizar, que además nos evitará tomar decisiones a posteriori en función de los resultados obtenidos. Para ello, debemos cuantificar cuáles son los efectos esperados en nuestra revisión. Así, si esperamos una respuesta binaria, esto es, enfermedad/no enfermedad o muerte/supervivencia, las medidas para cuantificarlos serán la diferencia de proporciones, el riesgo relativo o la *odds ratio*. Si por el contrario, como en nuestro caso, la variable es cuantitativa continua (la respuesta es un parámetro numérico), mediremos el efecto mediante la diferencia estandarizada de medias en los grupos de interés. Debemos haber fijado previamente la diferencia mínima en la variable respuesta que será clínicamente relevante.

Existen varias guías que informan de las etapas que debe tener toda revisión sistemática; entre ellas, la más importante es la guía MOOSE.

Etapas de una revisión sistemática

Las etapas de las que debe disponer una revisión sistemática son las siguientes:

1. Definición de la pregunta clínica de interés y los criterios de inclusión y exclusión de los estudios. Debemos tener en cuenta que unos criterios de inclusión excesivamente específicos pueden limitar la aplicabilidad externa de los resultados; y, al contrario, si son excesivamente amplios pueden hacer que nuestros resultados no tengan sentido clínico ni biológico.
2. Localización y selección de los estudios relevantes. Realizaremos una búsqueda exhaustiva, objetiva y reproducible de los trabajos originales sobre el tema, que incluirá bases de datos electrónicas, revistas relacionadas y búsquedas manuales en "literatura gris" (referencias bibliográficas, tesis doctorales, comunicaciones a congresos, informes de instituciones públicas o privadas, trabajos no publicados o publicados en revistas no indexadas, etc.). Este punto determinará en gran medida la calidad y la validez final de la revisión.
3. Extracción de datos de los estudios primarios.
En cada artículo seleccionado, buscaremos información de interés:
 - a) Diseño, criterios de inclusión y exclusión o selección de casos y controles, periodo de selección y periodo de seguimiento, aleatorización, tipo de intervención, etc.).
 - b) Características de la población de estudio.
 - c) Calidad metodológica (métodos de análisis estadístico utilizados) y sus resultados, especialmente en lo referido a las variables de interés. Podremos elaborar un formulario para la recogida de datos y que esta sea realizada por más de un investigador, para evaluar la fiabilidad de los resultados y consensuar las discrepancias.
4. Análisis y presentación e interpretación de los resultados.

Pregunta clínica de interés (pregunta de investigación)

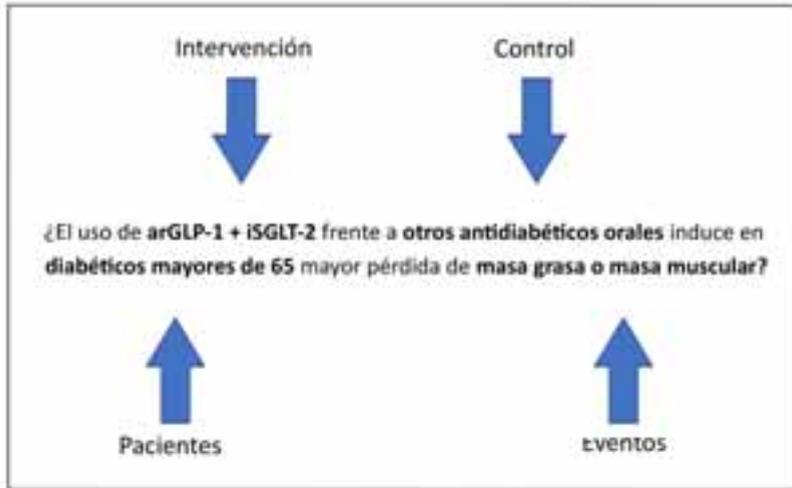
En general, será una pregunta explícita y estructurada que incluya la población objeto de estudio y el contexto clínico, la exposición de interés (intervención o tratamiento sobre nuestra población objeto de estudio y, a ser posible, una intervención de control) y el evento de interés que pretendamos contestar con nuestra pregunta.

Podemos representar gráficamente nuestra pregunta de investigación, junto los criterios de inclusión y exclusión, como se muestra en la **figura 2**.

Localización y selección de los estudios relevantes

En nuestro caso, realizaremos la búsqueda en bases de datos electrónicas, como Medline, Embase o Central. Es conveniente ayudarse de un software para la gestión de las citas bibliográficas, pues muchas de ellas aparecerán duplicadas. Hay que decidir el idioma de publicación de los artículos y el tipo de revista o publicación seleccionada. En nuestro caso, no

Figura 2. La pregunta clínica y los criterios de inclusión y exclusión



Criterios de inclusión

- Ensayos clínicos aleatorizados
- Estudios observacionales
- Revisiones sistemáticas y metaanálisis
- Pacientes mayores de 65 años con DM2
- Pacientes mayores de 65 años con DM2 en tratamiento con arGLP-1 + iSGLT-2 frente a otros antidiabéticos orales
- Referencia a la composición corporal
- Número mínimo de pacientes incluidos en el estudio: 150

Criterios de exclusión

- Pacientes menores de 65 años
- Pacientes en tratamiento con arGLP-1 o iSGLT-2 no en combinación
- No se recogen datos de composición corporal

DM2: diabetes mellitus tipo 2; arGLP-1: análogos del receptor del GLP-1;
iSGLT-2: inhibidores del cotransportador de sodio-glucosa 2.

limitaremos por idioma la búsqueda para evitar sesgos al no incluir artículos relevantes publicados en otro idioma distinto del inglés o el nativo, esto es, el español. Además, solo utilizaremos estudios publicados en revistas médicas que incluyan revisión por pares, corriendo el riesgo de sobreestimar el efecto buscado o esperado, pues los resultados no concluyentes no suelen publicarse.

En nuestro caso, una vez realizada la búsqueda, el resultado ha sido el siguiente: 345 referencias bibliográficas (200 en Medline, 100 en Embase y 45 en Central).

Nuestra hoja de recogida de datos queda como se muestra en la **figura 3**.

Figura 3. Hoja de recogida de datos para la selección de los artículos potenciales

Número de referencia del estudio:

Nivel de selección del estudio:
 Título ___ Resumen ___ Texto ___

Criterios de selección:

Pacientes
 ¿Incluye pacientes con DM2? Sí/NO
 ¿Incluye pacientes mayores de 65 años? Sí/NO

Intervención
 ¿Han recibido los pacientes tratamiento con arGLP-1 más iSGLT-2 en combinación? Sí/NO

Controles
 ¿Incluye pacientes que hayan recibido algún otro régimen de tratamiento? Sí/NO

Eventos
 ¿Se incluyen algunos de los siguientes resultados: masa grasa, masa no grasa, sarcopenia? Sí/NO

Diseño
 ¿Es un ensayo clínico aleatorizado? Sí/NO
 ¿Es un estudio observacional? Sí/NO
 ¿Es una revisión sistemática? Sí/NO
 ¿Es un metaanálisis? Sí/NO

ACCIÓN (Incluir SOLO en el caso de que la respuesta a TODAS las preguntas anteriores haya sido «Sí»):
 Incluir ___ Excluir ___ Dudoso ___

DM2: diabetes mellitus tipo 2; arGLP-1: análogos del receptor del GLP-1; iSGLT-2: inhibidores del cotransportador de sodio-glucosa 2.

Extracción y análisis de los datos

Es quizá la fase más importante de nuestro análisis. Debe hacerse con la máxima transparencia y rigor científico, por lo que es recomendable hacerlo por duplicado.

Podemos elaborar una hoja de extracción de datos, que debe incluir lo siguiente:

- Información sobre los pacientes incluidos (diabéticos mayores de 65 años), tipo de intervención (uso en combinación de un arGLP-1 y un iSGLT-2), intervención de control (uso de otro tipo de antidiabético oral), tipo de estudio (ensayo clínico aleatorizado, metaanálisis, estudio observacional).
- Información sobre los resultados obtenidos. En nuestro caso, al ser una variable cuantitativa continua (pérdida de masa grasa o masa no grasa) utilizaremos la media y la desviación estándar en cada grupo de tratamiento.
- La presentación resumida de los resultados obtenidos mediante una metodología sistemática y reproducible, como comentamos previamente, es lo que define una revisión sistemática como cualitativa y lo que la diferencia del metaanálisis, que, al combinar estadísticamente los resultados, constituye un análisis cuantitativo. Por ello, obtendremos una media aritmética de los resultados (que expresaremos junto a la desviación estándar), pero no tendremos en cuenta ni el peso específico de cada estudio incluido en el análisis ni la variabilidad intra- e interestudios. Por lo tanto, no hablaremos de heterogeneidad de resultados ni podremos asumir, propiamente hablando, una relación causa-efecto.

Interpretación del resultado. Fortalezas y debilidades

Nuestra revisión pone en evidencia que la combinación de un arGLP-1 con un iSGLT-2 en personas mayores de 65 años con diabetes induce una pérdida de peso corporal en forma de masa grasa sin inducir una pérdida significativa de masa no grasa o masa muscular, que podría ser perjudicial para la población en la que queremos aplicar nuestros resultados, esto es, mayores de 65 años con diabetes. Nuestros resultados son bastante homogéneos, luego podríamos pensar que el efecto es homogéneo y que no va a variar en exceso una vez que lo apliquemos a nuestros pacientes. Existen dos principios activos en los que el efecto parece mayor usados tanto juntos como asociados al resto de los arGLP-1 e iSGLT-2.

Entre las limitaciones de nuestro análisis se encuentra el número limitado de estudios analizados, puesto que en la mayoría de los ensayos clínicos el límite de edad para su inclusión son 65 años. Esto puede influir en la aplicabilidad externa de los resultados. En cambio, como fortaleza encontramos esta misma circunstancia: que no es necesario extrapolar en mis pacientes resultados de estudios que no los incluyan. Además, en la revisión sistemática hemos incluido siete ensayos clínicos con un número amplio de pacientes, cada uno con hasta cinco ramas de grupo control. Esto influye positivamente en la validez externa de los resultados.

Aplicabilidad en la práctica clínica

Nuestros resultados tienen una gran aplicabilidad en nuestra práctica clínica, puesto que hasta ahora en ese grupo de edad la tendencia era utilizar fármacos que priorizaban la seguridad sobre la eficacia. Con nuestro análisis, podemos considerar en pacientes con alto riesgo vascular o con enfermedad cardiovascular establecida, obesos sin fragilidad, el uso de ambas familias terapéuticas; con ello, conseguiremos una pérdida ponderal beneficiosa para el riesgo cardiovascular del paciente sin influir en la posible aparición de una sarcopenia que empeore la funcionalidad o la independencia de los pacientes.

Bibliografía

Carretero Gómez J, Arévalo Lorido JC, Gómez Huelgas R, et al.; Diabetes, Obesity, and Nutrition Spanish Working Group. Combination Therapy With Glucagon-Like Peptide-1 Receptor Agonists and Sodium-Glucose Cotransporter 2 Inhibitors in Older Patients With Type 2 Diabetes: A Real-World Evidence Study. *Can J Diabetes*. 2019;43:186-92. doi: 10.1016/j.jcjd.2018.09.001. Epub 2018 Sep 8.

Ferreira González I, Urrutia G, Alonso-Coello P. Systematic reviews and meta-analysis: scientific rationale and interpretation. *Rev Esp Cardiol*. 2011;64:688-96. doi: 10.1016/j.recesp.2011.03.029. Epub 2011 Jun 30.

Jüni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ*. 2001;323:42-6. doi: 10.1136/bmj.323.7303.42.

Lajara R. Combination therapy with SGLT-2 inhibitors and GLP-1 receptor agonists as complementary agents that address multi-organ defects in type 2 diabetes. *Postgrad Med*. 2019;131:555-65. doi: 10.1080/00325481.2019.1670017. Epub 2019 Oct 3.

Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol*. 2009;62:e1-34. doi: 10.1016/j.jclinepi.2009.06.006. Epub 2009 Jul 23.

Sargeant JA, Henson J, King JA, Yates T, Khunti K, Davies MJ. A Review of the Effects of Glucagon-Like Peptide-1 Receptor Agonists and Sodium-Glucose Cotransporter 2 Inhibitors on Lean Body Mass in Humans. *Endocrinol Metab*. 2019;34:247-62. doi: 10.3803/EnM.2019.34.3.247.

Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA*. 2000;283:2008-12. doi: 10.1001/jama.283.15.2008.

Preguntas de autoevaluación

1. Respecto a las revisiones sistemáticas, señale la afirmación incorrecta:

- a) La unidad de análisis utilizada son los estudios originales primarios.
- b) La revisión sistemática se considera una revisión cualitativa.
- c) El metaanálisis no se considera una revisión cuantitativa.
- d) La revisión sistemática no combina estadísticamente los resultados obtenidos.

La respuesta correcta es la opción *c*.

El metaanálisis se considera una revisión cuantitativa, pues conlleva un análisis estadístico de los resultados obtenidos en cada uno de los estudios incluidos. Esa es la principal diferencia entre el metaanálisis y la revisión sistemática, y por eso muchas veces se combinan ambas, pues la base del trabajo, esto es, la selección bibliográfica y la extracción de los datos, son muy similares.

2. En una revisión sistemática, los estudios incluidos deben ser...

- a) Metaanálisis.
- b) Ensayos clínicos
- c) Estudios observacionales.
- d) Todos.

La respuesta correcta es la opción *d*.

En una revisión sistemática, los artículos que se han de utilizar son los ensayos clínicos, las revisiones sistemáticas y los metaanálisis. Sin embargo, si queremos evaluar la eficacia de intervenciones comunitarias o de salud pública, o resultados de intervenciones a largo plazo, sobre todo para parámetros de seguridad, podemos utilizar estudios observacionales.

3. A la hora de plantear la pregunta de investigación, ¿qué debemos tener en cuenta?

- a) Que sea amplia para abarcar la mayor parte de la población.
- b) Que incluya la exposición de interés.
- c) No es necesario que incluya la intervención realizada sobre la población.
- d) La pregunta no debe pretender contestar al desenlace de interés.

La respuesta correcta es la opción *d*.

En general, la pregunta de investigación será una pregunta explícita y estructurada que incluya la población objeto de estudio y el contexto clínico en el que centramos la revisión, la exposición de interés (intervención o tratamiento sobre nuestra población objeto de estudio y, a ser posible, una intervención de control) y el evento o desenlace de interés que pretendamos contestar con nuestra pregunta.

Análisis de mortalidad

Dr. José Pablo Miramontes González
Servicio de Medicina Interna
Hospital Universitario Río Hortega (Valladolid)
Facultad de Medicina, Universidad de Valladolid (Valladolid)
Instituto de Investigación de Salamanca (IBSAL)

Dr. Javier Martín Vallejo
Departamento de Estadística
Universidad de Salamanca (Salamanca)

*La publicación seleccionada de este capítulo
es un trabajo real publicado (véase la bibliografía)*



capítulo 4

Publicación seleccionada

Resultados en seguridad cardiovascular de semaglutida en una cohorte de pacientes con diabetes mellitus tipo 2

Introducción

Desde hace unos años los reguladores sanitarios, especialmente la Administración de Medicamentos y Alimentos (FDA) de Estados Unidos y la Agencia Europea de Medicamentos (EMA), exigen estudios de seguridad cardiovascular para las distintas moléculas que se utilizan para el tratamiento de los pacientes diabéticos. Para ello se diseñan estudios con objetivos específicos orientados a solventar estos requerimientos de seguridad. Las variables que se fijan para poder establecer estas conclusiones suelen ser la mortalidad por causa cardiovascular y los eventos cardiovasculares, como son el infarto agudo de miocardio y el accidente cerebrovascular, con mayor o menor repercusión clínica.

Características del estudio

El estudio que proponemos para analizar la mortalidad, y por extensión la seguridad cardiovascular, es un estudio aleatorizado, doble ciego y controlado del fármaco contra placebo, con los grupos en paralelo. Se realizó en 230 centros de 20 países. Los pacientes se aleatorizaron para recibir 0,5 y 1,0 mg de semaglutida o placebo, con un periodo de seguimiento de 109 semanas para todos los pacientes.

Características de los pacientes

Como ocurre con otros tipos de análisis, el diseño del estudio y la población que se selecciona es muy importante para poder llegar a una conclusión que sea veraz. En el caso de los estudios de seguridad o supervivencia en una población con factores de riesgo para desarrollar eventos y sometida a una posible modificación, en este caso la administración de un fármaco, ayuda a tener unas conclusiones adecuadas. En el estudio que hemos seleccionado los pacientes eran todos diabéticos con hemoglobina glucosilada (HbA_{1c}) mayor de 7%, con edad mayor de 50 años y con enfermedad cardiovascular establecida (enfermedad cerebrovascular, enfermedad vascular periférica, enfermedad cardíaca, insuficiencia cardíaca clase II o III) o enfermedad renal en estadio 3 o superior; en caso de ser pacientes mayores de 60 años, era necesario tener factores de riesgo vascular. Como criterios de exclusión, entre otros, se descartó a los pacientes con antecedentes de un evento vascular en los 90 días previos.

Con estas características, se evaluó a un total de 4346 pacientes y se aleatorizó a 3297; de estos pacientes, 3232 (98%) asistieron a la última visita de seguimiento.

Objetivos

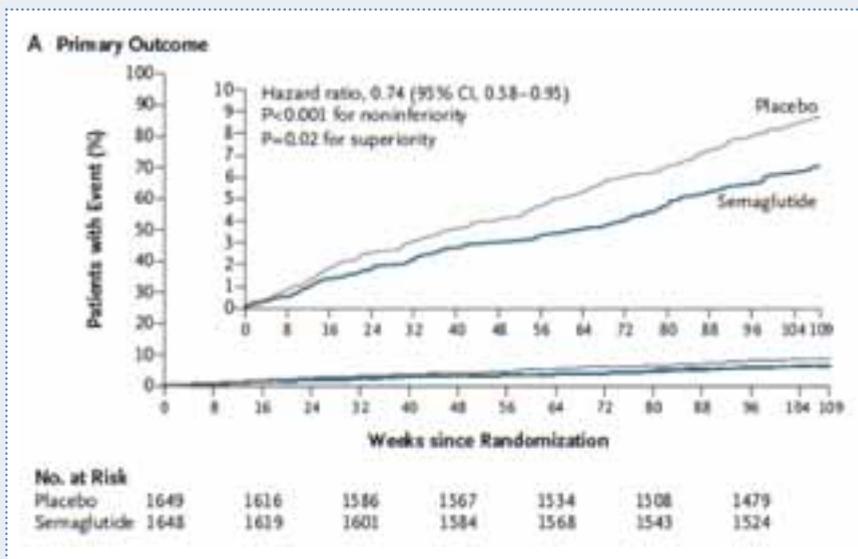
El objetivo compuesto primario fue el evento de muerte por causas cardiovasculares, infarto de miocardio no mortal (incluido el silente) y accidente cerebrovascular no

mortal. El resultado secundario preespecificado incluyó la primera aparición de una extensión del resultado cardiovascular compuesto (muerte de causas cardiovasculares, infarto de miocardio no mortal, accidente cerebrovascular no mortal, revascularización –coronaria o periférica– y hospitalización por angina inestable o insuficiencia cardíaca) y un resultado compuesto (muerte por todas las causas, infarto de miocardio no mortal o accidente cerebrovascular no mortal).

Resultados

El resultado primario compuesto ocurrió en 108 de 1648 pacientes (6,6%) en el grupo de semaglutida y en 146 de 1649 (8,9%) en el grupo placebo (*hazard ratio* [HR]: 0,74; intervalo de confianza del 95% [IC95%]: 0,58-0,95; $p < 0,001$ para no inferioridad; $p = 0,02$ para superioridad). Los pacientes tratados con semaglutida en el ensayo tenían un riesgo un 26% menor en relación con el resultado primario compuesto de muerte por causas cardiovasculares, infarto de miocardio no mortal, accidente cerebrovascular o accidente cerebrovascular no mortal que los que recibieron placebo (**figura 1**).

Figura 1. Resultado del objetivo primario (mortalidad) en el grupo de semaglutida frente a control



Fuente: Marso SP, Bain SC, Consoli A, et al.; SUSTAIN-6 Investigators. Semaglutide and cardiovascular outcomes in patients with type 2 diabetes. *N Engl J Med.* 2016 Nov 10;375(19):1834-44. DOI: 10.1056/NEJMoa1607141

Introducción

En el campo clínico hay tres modelos que son los más habituales a la hora de analizar los datos en las diferentes áreas del ámbito de las ciencias de la salud. Estos modelos son la regresión lineal, la regresión logística y los modelos de supervivencia. La principal diferencia entre los tres modelos es el tipo de variable de interés que se quiere explicar: en el modelo de regresión es una variable continua cuyos errores siguen una distribución normal; en el modelo logístico es una variable dicotómica (presencia o ausencia de un evento) cuyos errores siguen una distribución binomial; y en el modelo de supervivencia la variable dependiente es el tiempo hasta la presencia de un evento de interés, como ocurre en el artículo propuesto.

La historia del análisis de supervivencia parte de las inquietudes del hombre por estudiar el desarrollo en el tiempo de la vida de las personas por diferentes circunstancias. Las tablas de vida en el campo demográfico, las tablas actuariales en el campo de la economía y el trabajo pionero de Bernoulli sobre el efecto de la inoculación de la viruela en el tiempo de vida han marcado el desarrollo de las técnicas que estudian y describen el tiempo que transcurre desde un momento dado hasta la aparición de un evento que generalmente es la muerte en el campo clínico. Tradicionalmente, en medicina, el análisis de supervivencia estaba asociado a la muerte en el caso de enfermedades graves. Sin embargo, en la actualidad, y debido a la urgencia de resultados sobre la efectividad de los nuevos tratamientos, ha aparecido el interés por estudiar nuevos eventos que intentan sustituir al evento muerte. En este capítulo se va a tratar de dar las nociones básicas sobre el análisis de supervivencia sin adentrarse en aspectos técnicos profundos. En cualquier caso, para la interpretación precisa de los resultados de un análisis de supervivencia se deben conocer y comprender los aspectos teóricos en los que se basa este análisis.

Para la comprensión de los resultados más habituales que se reflejan en un artículo de investigación que aplica el análisis de supervivencia se ha elegido un artículo con los resultados del ensayo clínico SUSTAIN 6, pero hay otros similares en los que se puede observar este tipo de análisis. En estos ensayos se mide el efecto en términos de seguridad y beneficio de la semaglutida sobre los pacientes con diabetes mellitus tipo 2. En estos estudios, aparte del estudio de supervivencia (o mortalidad), se realizó un estudio longitudinal para analizar la seguridad y eficacia del tratamiento que está fuera del objetivo de este capítulo.

Características de los datos

En el análisis de supervivencia la variable que se analiza es el tiempo hasta que se produce un evento (tiempo hasta la muerte, tiempo hasta la aparición de un síntoma, tiempo hasta la aparición de metástasis, etc.). Las principales características de esta variable son las siguientes:

- Presenta distribuciones asimétricas y, por lo tanto, no cumple el supuesto de normalidad y la presencia de datos censurados.
- Presenta información parcial (censura), es decir, los individuos no presentan el evento por diversos motivos: finalización del ensayo, abandono del estudio, pérdida de seguimiento, etc.

El tipo de censura puede ser diferente en función de cuando se produce el evento en relación con el tiempo censurado. En la **figura 2** se muestra una clasificación del tipo de censura. La censura simple es la que se produce en un mismo momento y generalmente es cuando se acaba el experimento. Este tipo de censura es habitual en estudios experimentales con animales. La censura progresiva se produce en diferentes momentos del estudio. Dentro de la censura progresiva puede ser una censura a la derecha, cuando el tiempo de aparición del evento es posterior al último tiempo de revisión. Una censura a la izquierda ocurre cuando el tiempo de aparición del evento es anterior al tiempo de revisión, y censura por intervalo es cuando el tiempo de aparición del evento se produce entre dos tiempos de revisión. La mayoría de los estudios de supervivencia en el campo clínico consideran censuras a la derecha.

Figura 2. Clasificación de los tiempos de censura



Uno de los supuestos que deben cumplir los datos censurados para que las estimaciones de la función de supervivencia sean fiables es que los tiempos de supervivencia deben ser independientes de cualquier mecanismo que cause censura en los datos. En otras palabras, el proceso de censura debería operar de forma aleatoria. Si en un estudio la gran mayoría de los abandonos se deben al deterioro del paciente debido a la enfermedad, entonces no se cumpliría el supuesto de independencia. Cuando se reconoce un mecanismo de abandono del estudio, entonces se denomina *censura informativa* y debería evitarse en este tipo de estudios. Otro inconveniente relacionado con las censuras es la presencia de otro evento que impide (o modifica) la presencia del evento de interés que se está analizando y, por lo tanto, está relacionado con él. La presencia de este tipo de eventos se denomina *riesgos competitivos*. El efecto de los riesgos competitivos, cuando impiden que el evento de interés no se pueda producir, es la sobreestimación de la presencia del evento y, por lo tanto, subestima la función de supervivencia. Hay que estar alerta de la presencia en el estudio de este tipo de eventos.

El efecto de tener en cuenta el tiempo hasta la aparición del evento y también el tiempo hasta que un individuo es censurado por cualquier motivo es importante a la hora de valorar cómo se presenta un evento. Supongamos que se realiza un estudio de supervivencia de una determinada enfermedad durante un año (**figura 3**). Se estudia a 48 pacientes de los cuales 14 han abandonado el estudio por diversas razones; de los que han finalizado el estudio, 12 han fallecido por causa de la enfermedad. Si se analiza de manera clásica la proporción de supervivientes en el tiempo de estudio, es decir, considerando los abandonos como si no participaran en el estudio, el porcentaje de supervivencia sería de 64,7% (de los 34 individuos que completaron el estudio, 22 sobrevivieron). Se van a considerar dos escenarios extremos y ficticios sobre el desenlace de los 14 individuos que abandonaron:

1. Todos fallecieron: el porcentaje de supervivencia sería un 45%.
2. Todos sobrevivieron: el porcentaje de supervivencia sería un 75%.

Figura 3. Esquema de la simulación de un estudio de mortalidad

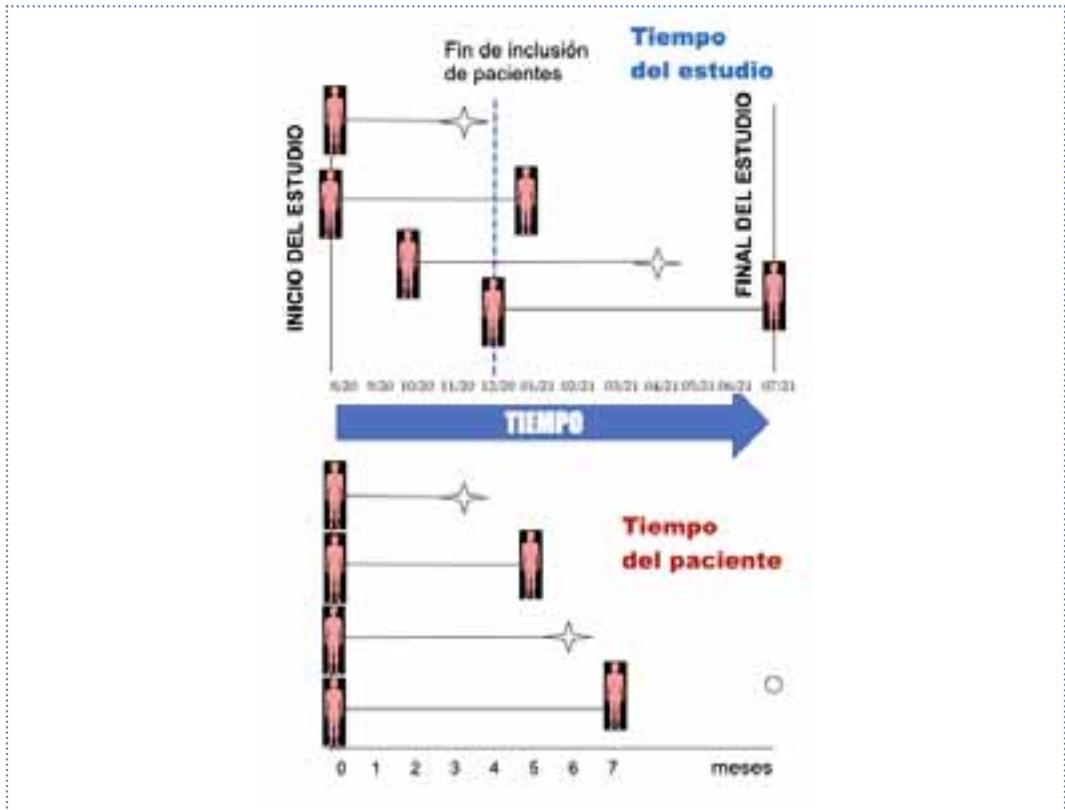


Entre los dos escenarios posibles se tendría una gran variedad de estimaciones de supervivencia que posiblemente estarían alejadas del valor estimado de la manera clásica. Además, si se aumenta el tiempo de estudio, las diferencias serían mucho mayores, con lo cual el porcentaje de supervivencia estaría más sesgado. Un análisis de la función de supervivencia en el que se registra el tiempo hasta que se produce el evento, la censura, es más real y las estimaciones más precisas¹.

Generalmente, en los ensayos clínicos los pacientes no son incluidos todos en el mismo tiempo. Los pacientes se van incorporando progresivamente según van apareciendo siempre que cumplan los requisitos de inclusión. Esto implica hacer una diferenciación entre el tiempo del estudio y el tiempo de paciente. El primero hace referencia al tiempo real según el calendario y el segundo es el tiempo de supervivencia del paciente desde su punto inicial hasta el final, tanto por la presencia del evento como por una censura (**figura 4**).

¹ Si se simula un estudio con esos datos y se plantean diferentes tiempos de censura (el tiempo hasta que abandona el estudio), la estimación más correcta de supervivencia al año sería del 70%.

Figura 4. Esquema de los registros de tiempo en un estudio de supervivencia



Una información importante para la comparación de estudios es el tiempo mediano de seguimiento del paciente.

En los estudios en los que el evento de interés es la muerte existen dos tipos de criterios de valoraciones de interés debido a la evolución de mayor conocimiento y mejor tratamiento que han sufrido las enfermedades:

- Los criterios clínicos o primarios (*clinical or primary outcome*): este tipo de criterios se consideran validados y objetivos a la hora de estimar el efecto en la supervivencia de una enfermedad.
 - Supervivencia global (*overall survival, OS*): el evento de interés que se valora es la muerte por cualquier causa.
 - Supervivencia por causa específica (*cause-specific survival, CSS*): el evento de interés valorado es la muerte por la enfermedad sometida a estudio. El problema del uso de este criterio es la presencia de riesgos competitivos; por ello, actualmente esta valoración tiene menos uso.

- Los criterios de valoración sustitutivos (*surrogate outcome*): biomarcadores o criterios de valoración intermedios que pretenden sustituir o predecir el resultado realmente de interés. Estos criterios se utilizan con mayor frecuencia en el análisis de supervivencia debido a que producen resultados con mayor rapidez que los criterios primarios. Los organismos reguladores de los medicamentos (FDA, EMA) apoyan este tipo de criterios de valoración para acelerar la aprobación de nuevos fármacos que puedan evitar muertes. Los criterios de valoración sustitutivos más utilizados son los siguientes:
 - Supervivencia libre de progresión (*progression-free survival, PFS*): tiempo hasta la progresión de la enfermedad o la muerte por cualquier causa.
 - Supervivencia libre de enfermedad (*disease-free survival, DFS*): tiempo hasta la recurrencia o la muerte por cualquier causa.
 - Tiempo hasta la progresión (*time to progression, TTP*): tiempo hasta la progresión de la enfermedad.

Dependiendo del tipo de enfermedad estudiada se pueden tomar unos criterios de valoración u otros debido a las características de supervivencia. La denominación es específica de los estudios de supervivencia con tumores y puede cambiar dependiendo de la enfermedad.

Los criterios de valoración son muy controvertidos porque, aunque están admitidos por las agencias del medicamento, hay varios autores que consideran que la relación con los criterios primarios no es muy grande y, por lo tanto, dudan de su validez.

Las variables estadísticas necesarias para analizar el criterio de valoración elegido en el estudio clínico son dos: una variable cuantitativa que mide el tiempo hasta la presencia del evento o censura, y otra dicotómica que indica si se ha producido el evento o una censura.

El ensayo clínico SUSTAIN 6 analiza como criterio primario la primera ocurrencia de uno de los tres siguientes eventos: muerte por causas cardiovasculares, infarto de miocardio no mortal o accidente cerebrovascular no mortal. Como criterio secundario definen también un criterio compuesto que es la primera ocurrencia de causas cardiovasculares, infarto de miocardio no mortal o accidente cerebrovascular no mortal, revascularización coronaria u hospitalización por angina inestable o insuficiencia cardíaca. Se establece un último criterio que se podría identificar también como criterio primario y que vendría definido por muerte por cualquier causa, infarto de miocardio no mortal o accidente cerebrovascular no mortal. Se supone que este último criterio se ha elegido para evitar el problema de riesgos competitivos (se supone que en los dos primeros criterios la muerte por otra causa se consideró como una censura aleatoria, supuesto que está en entredicho, como se ha comentado).

En el ensayo PIONEER 6 se definen los mismos criterios, con la pequeña diferencia de que en el criterio primario se añade un evento adverso mayor cardiovascular y que en las muertes cardiovasculares se añaden también las causas de muerte no determinadas, para evitar el problema de los riesgos competitivos.

En ambos se realiza un estudio de sensibilidad en el que se incluyen varias variables, como el sexo, la región, la edad en el inicio de estudio, la duración de la diabetes, la función renal basal, etc.

Métodos estadísticos

En el análisis de supervivencia los objetivos principales son los siguientes:

- Estimar la función de supervivencia (o la función de incidencia acumulada), estableciendo el tiempo mediano como medida descriptiva o cualquier otro cuartil de interés, además de estimar las proporciones de supervivencia en tiempos concretos del estudio que son de interés clínico.
- Comparar las curvas de supervivencia entre grupos, generalmente definidos por distintos tratamientos.
- Modelizar los tiempos de supervivencia para identificar variables pronóstico que definan diferentes grados de riesgo de que se produzca el evento de interés.

Para llevar a cabo estos dos objetivos se deben definir dos funciones esenciales para realizar los análisis de los datos.

- La función de supervivencia, que nos indica la probabilidad de que un individuo sobreviva más allá de un tiempo determinado. Esta función estima la probabilidad de supervivencia en los distintos tiempos observados. También se puede definir la función de incidencia acumulada, que expresa la probabilidad de que un individuo presente el evento antes de un tiempo determinado.
- La función de riesgos (*hazard function* o *hazard rate*): nos indica la probabilidad de que un evento se presente en un intervalo de tiempo muy pequeño; es decir, se puede definir como la tasa instantánea de presencia del evento. Es un concepto similar al de velocidad. Nos indicaría el ritmo de aparición del evento de interés en relación con el tiempo. No es una función de probabilidad, ya que los valores pueden ser superiores a 1.

Estimación de la curva de supervivencia

La estimación de las probabilidades de supervivencia si no hubiera censuras (escenario ficticio) se calcularía de acuerdo con la siguiente fórmula:

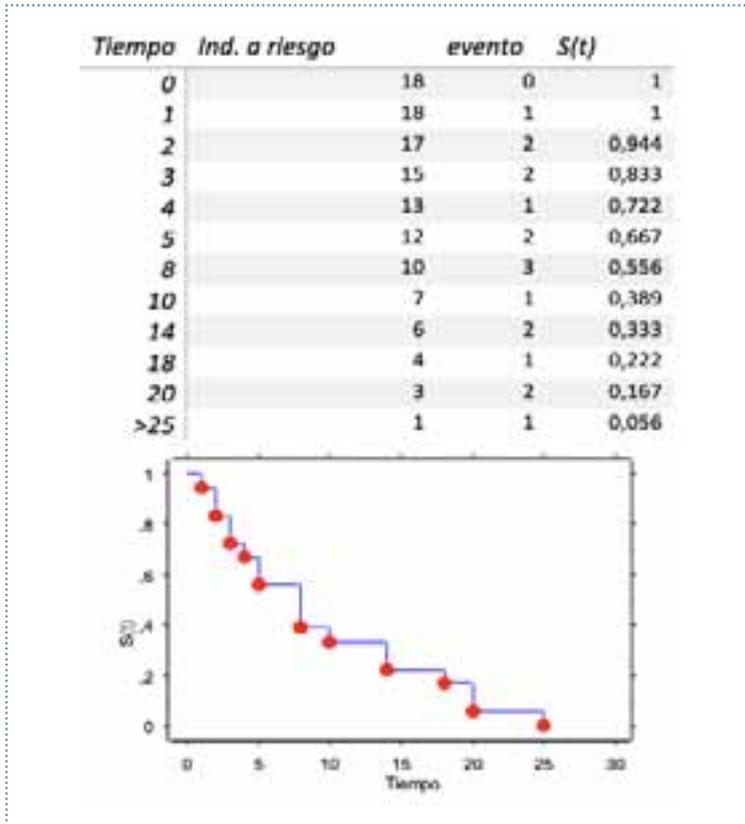
$$S(t) = \frac{\text{Número de individuos con tiempo de supervivencia } \geq t}{\text{Número de individuos total}}$$

Imaginemos que estamos analizando la función de supervivencia de individuos que padecen una enfermedad y no se han producido ninguna censura. Los resultados se muestran en la **figura 5**.

Cuando hay presencia de censuras, el denominador de la estimación de la función de supervivencia va cambiando y, por lo tanto, no coincide con la expresión mostrada anteriormente. Por eso, se plantea la estimación no paramétrica de Kaplan-Meier. Esta estimación está basada en la idea de calcular la probabilidad de sobrevivir a un tiempo dado t condicionado a que se haya sobrevivido los t_{-1} tiempos anteriores. Si los eventos aparecen de manera independiente, correspondería con esta expresión:

$$S(t) = p_1 \cdot p_2 \cdots p_{i-1} \cdot p_i$$

Figura 5. Tabla y representación de la función de supervivencia sin presencia de censuras



En el ejemplo de la **figura 5**, donde no existen censuras, el cálculo del valor de probabilidad acumulado de supervivencia en el tiempo 4 sería el siguiente:

$$S(4) = \frac{18}{18} \cdot \frac{17}{18} \cdot \frac{15}{17} \cdot \frac{13}{15} = \frac{13}{18} = 0,722$$

Sin embargo, si hay presencia de censuras, los numeradores no tendrían que coincidir con los denominadores, ya que los individuos expuestos al riesgo en cada tiempo cambian no solo porque se ha producido el evento, sino porque hay censuras (**figura 6**). Imaginemos el caso el ejemplo anterior con la inclusión de 7 tiempos censurados:

$$S(4) = \frac{25}{25} \cdot \frac{24}{25} \cdot \frac{22}{24} \cdot \frac{19}{21} = S(t=3) \cdot \frac{19}{21} = 0,796$$

Se parte n tiempos $\{t_1, t_2, \dots, t_n\}$, en los que se producen r eventos $\{t_1 < t_2 < \dots < t_r\}$, $r < n$; n_j es el número de individuos que no presentan el evento antes de t_j , el número de eventos que se producen en el tiempo j es d_j . La probabilidad de que se produzca el evento en el intervalo

de tiempo j es d_j/n_j y, por lo tanto, la probabilidad de supervivencia en ese intervalo es la siguiente:

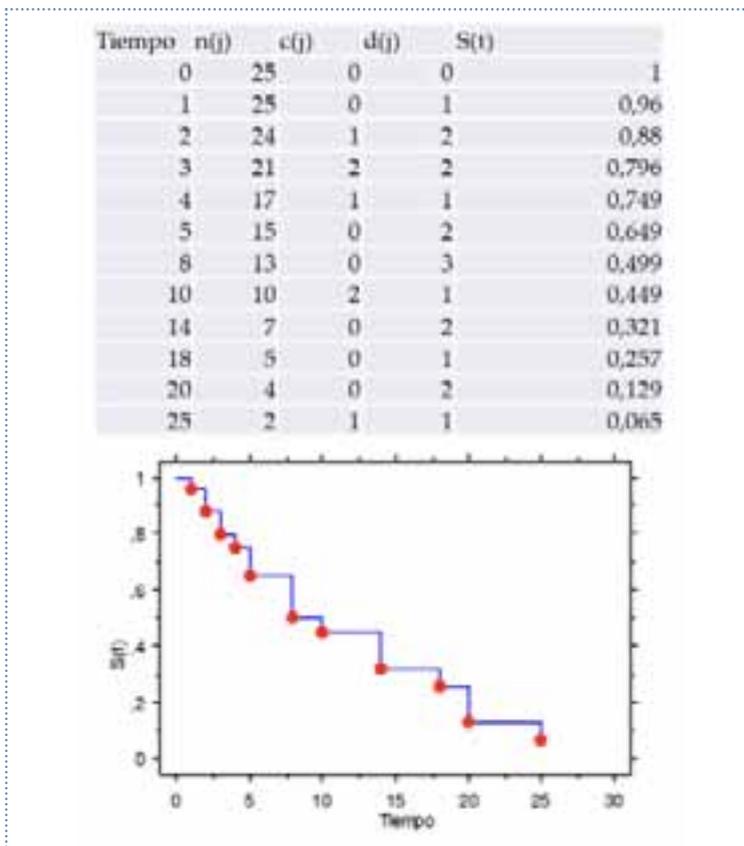
$$\frac{n_j - d_j}{n_j} = 1 - \frac{d_j}{n_j}$$

Por lo tanto, la estimación de Kaplan-Meier se calcula de la siguiente forma:

$$S(t_j) = S(t_{j-1}) \cdot \left(1 - \frac{d_j}{n_j}\right)$$

donde $t = 0$ y $S(0) = 1$. Se puede estimar la probabilidad de estar vivo en tiempos del estudio que pueden ser interesantes para los investigadores. Uno de los valores que se suelen uti-

Figura 6. Tablas y representación de la función de supervivencia en presencia de tiempos censurados



$n(j)$: individuos en riesgo en el tiempo j ; $c(j)$: número de censuras;
 $d(j)$: número de eventos en el tiempo j .

lizar para describir los tiempos de supervivencia es la mediana. Además, se puede estimar el intervalo de confianza al 95% para cada estimación. Hay varias propuestas de estimación del error de estándar para la estimación $S(t)$, pero el más utilizado es el calculado con la aproximación de Greenwood (Greenwood, 1926). Este intervalo de confianza se calcula para tiempos concretos, pero no muestran el proceso global de evolución en el tiempo. Por ello se han propuesto las bandas de confianza que se suelen mostrar en la curva de supervivencia. Los dos procedimientos más habituales para las estimaciones de las bandas de confianza son los propuestos por Hall y Wellner y por Nair. Estos procedimientos presentan mayor complejidad en el cálculo que los intervalos de confianza puntuales.

Calculada la función de supervivencia, se puede calcular la función de incidencia acumulada sin más que $F(t) = 1 - S(t)$. De la misma manera, se pueden calcular en la función de incidencia acumulada los diferentes cuartiles o la probabilidad de sufrir el evento a un tiempo determinado (**figura 7**).

En ambos ensayos revisados por el artículo se muestran la función de incidencia acumulada en lugar de la de supervivencia.

Comparación de curvas de supervivencia

En la mayoría de los estudios de supervivencia uno de los objetivos más habituales es la comparación de dos o más situaciones que generalmente están definidas por diferentes tratamientos. En nuestro estudio se compara el tratamiento con semaglutida en su preparación oral (PIONEER 6) o preparación estándar (SUSTAIN 6) frente a un placebo. Los procedimientos para comparar las curvas de supervivencia se basan en la integración de los resultados para cada tiempo concreto. Si se parte del caso más simple –la comparación de dos grupos, como ocurre en los ensayos analizados–, para cada tiempo se puede construir una tabla de contingencia 2×2 (**figura 8**) definida por los dos grupos experimentales (semaglutida en presentación oral o estándar frente a placebo) y la presencia o no del evento de estudio (cualquiera de los criterios primarios o secundarios de los estudios analizados).

Modelización de los tiempos de supervivencia

Hasta el momento se ha resuelto el problema cuando se pretende analizar la relación existente entre las curvas de supervivencia y un factor que define un conjunto de grupos. Pero hay otras situaciones que son interesantes de analizar y que con las herramientas explicadas hasta el momento no son posibles de realizar:

- Analizar la relación de la función de supervivencia con una variable cuantitativa, por ejemplo, el peso, la edad, los marcadores bioquímicos, etc. Una forma de analizar dicha relación sería recodificar la variable en grupos y el problema estaría resuelto. Muchas veces la recodificación no es aconsejable porque puede introducir sesgo o simplemente no se quiere simplificar la variable y, por lo tanto, las estimaciones de Kaplan-Meier y el test *log-rank* no se pueden calcular.
- Analizar la relación de la función de supervivencia con varias variables al mismo tiempo debido a que se quiere hacer un modelo pronóstico o por la inclusión de covariables que ajusten los valores del factor experimental que se está comparando. Hasta el

Figura 7. Función de supervivencia y de incidencia en un estudio de cáncer de mama

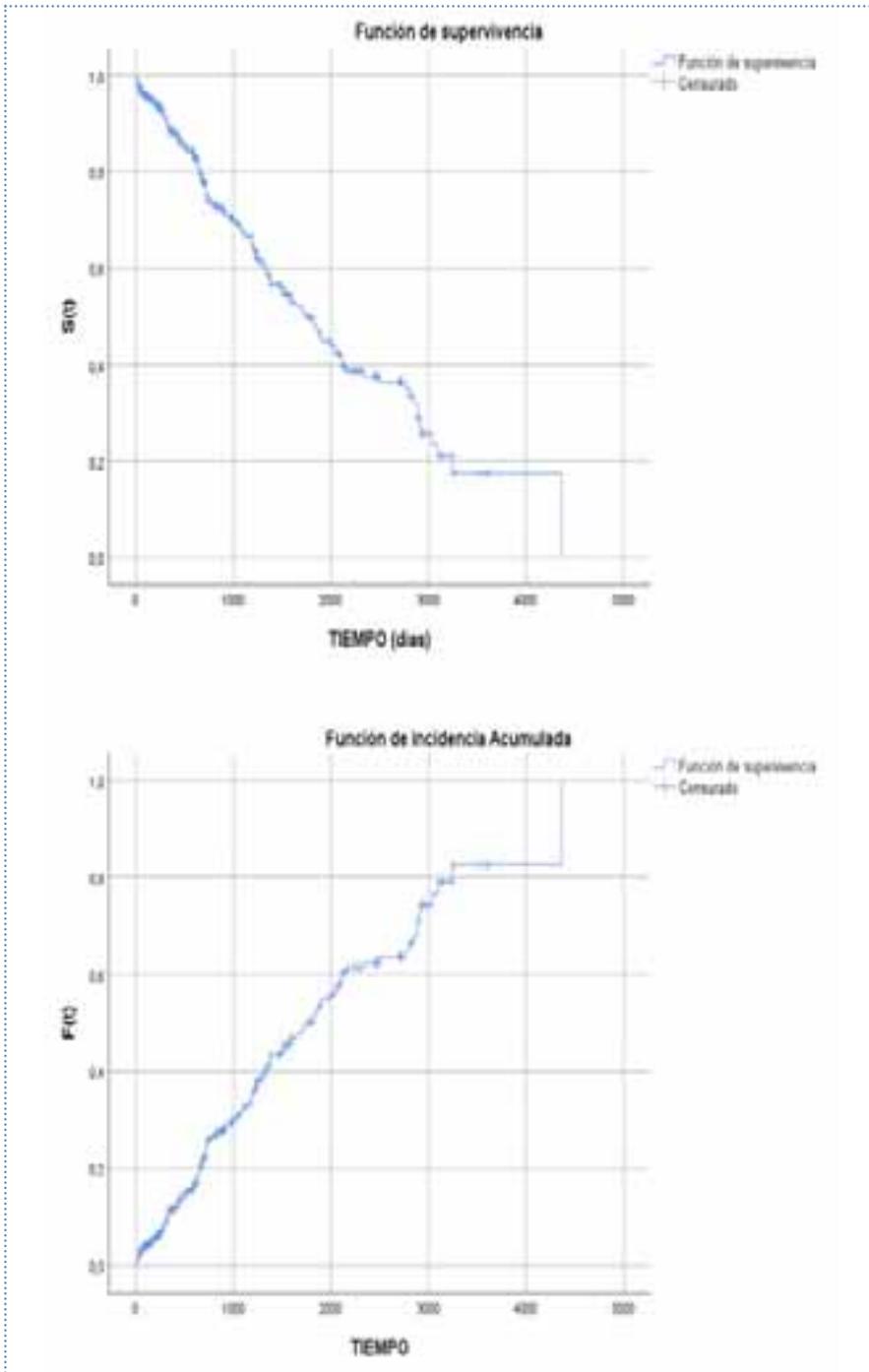
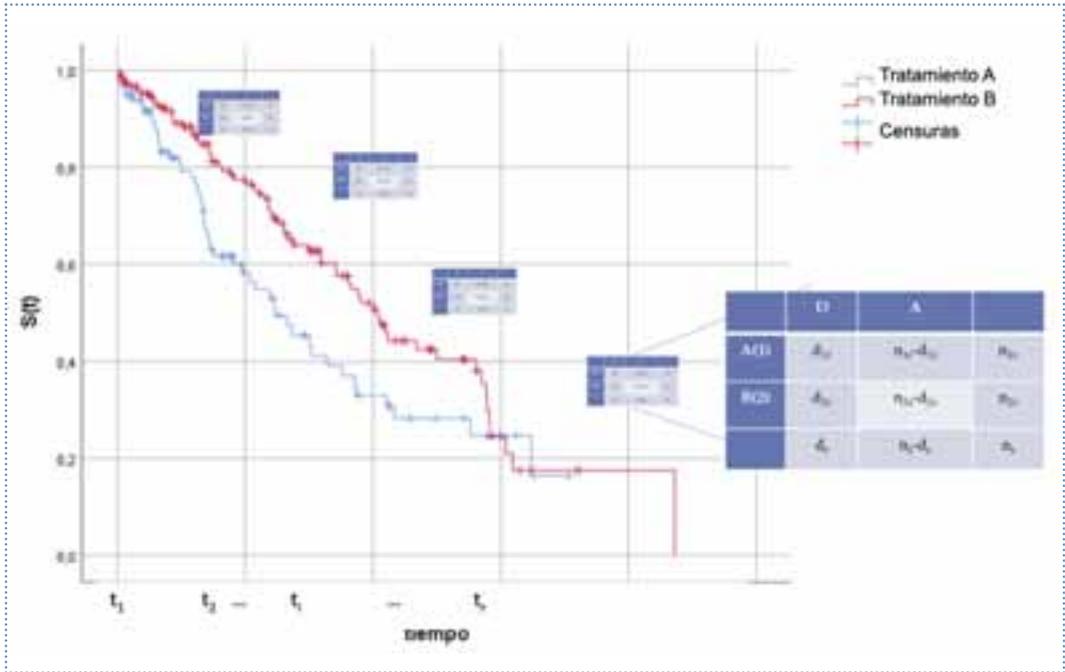


Figura 8. Representación del fundamento del test *log-rank* como integración de tablas 2x2 para cada tiempo *j*



D: evento; A: no evento; A(1): grupo experimental 1; B(2): grupo experimental 2.

momento solo se han analizado situaciones con una única variable y una segunda en el proceso de estratificación.

El modelo de Cox proporciona una solución a estas limitaciones. Este modelo parte del supuesto de proporcionalidad de riesgos que ya se planteó en el test *log-rank*. A partir de ese supuesto, plantea un modelo libre de distribuciones de probabilidad teóricas subyacentes sobre la función de supervivencia. Por ello, se considera que es un modelo semiparamétrico, ya que no necesita definir una función de probabilidad pero sí debe cumplir el supuesto de proporcionalidad de riesgos.

En este modelo se puede partir de la situación más simple de un conjunto de pacientes con una enfermedad en el que se pretende probar un nuevo tratamiento (podría ser también una variable pronóstica, como el grado del tumor, la edad, etc). Se va a definir teóricamente un riesgo basal que representa al "paciente medio" que presenta dicho cáncer. Se denota $h_0(t)$ y no tiene por qué estar especificada. A partir de esta función basal se puede expresar el riesgo de cualquier paciente con dicho cáncer en los siguientes términos:

$$h(t) = \psi \cdot h_0(t) \Rightarrow \psi = \frac{h(t)}{h_0(t)}$$

Por lo tanto, se considera el riesgo del paciente como un múltiplo, constante en el tiempo, del riesgo del paciente medio. Así, se puede calcular ese valor constante como el cociente de riesgos.

Si ψ es igual a 1, ambos individuos presentan el mismo riesgo; si es mayor de 1, el riesgo del paciente es mayor que el del paciente medio; y si es menor de 1, el riesgo es menor. Los riesgos pueden cambiar con el tiempo; sin embargo, ψ es constante (proporcionalidad de riesgos).

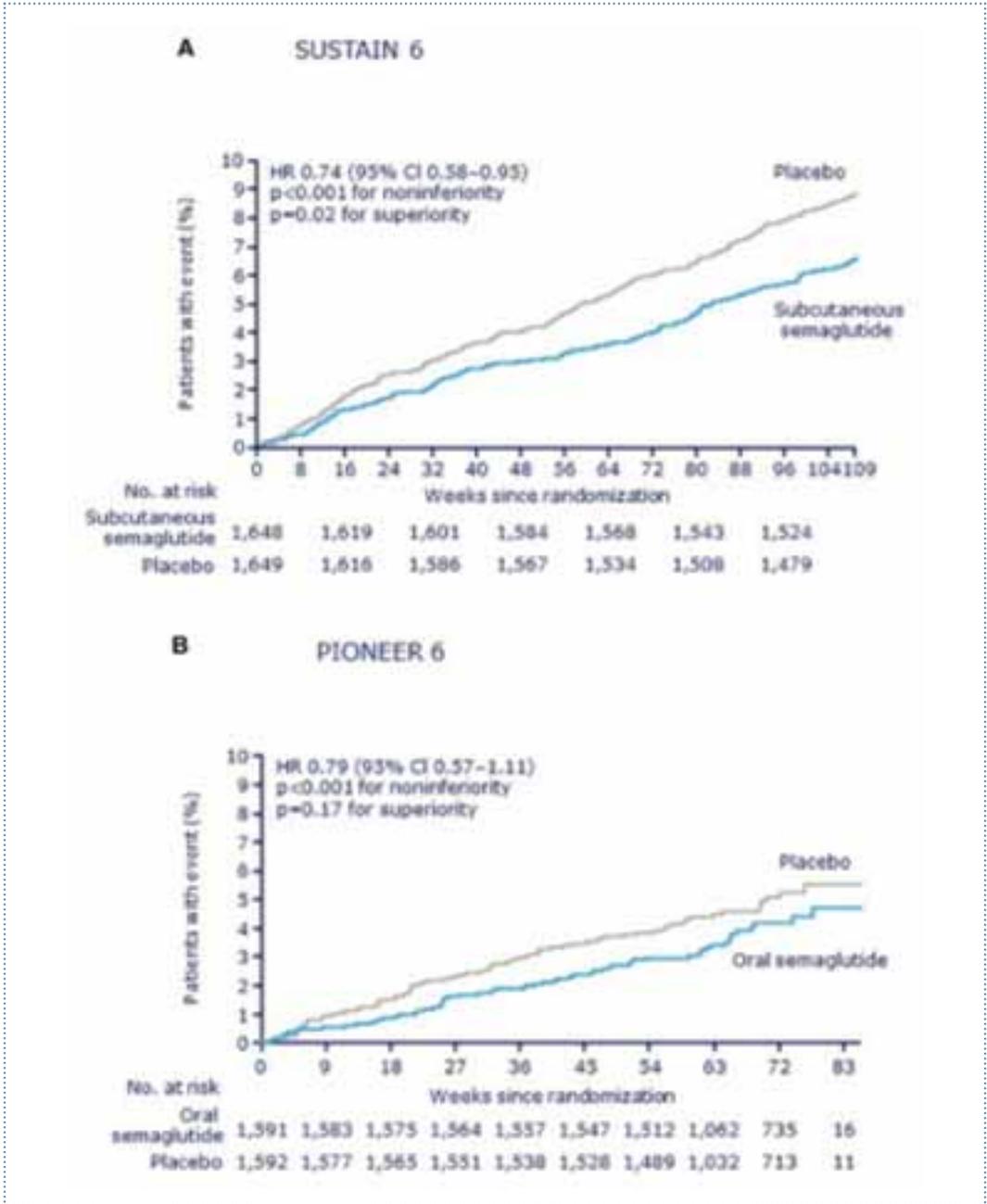
Tanto el ensayo SUSTAIN 6 como en el PIONEER 6 se realiza un estudio de incidencia acumulada de los criterios primarios y secundarios. Se comparan las curvas mediante el test *log-rank*, pero utilizando el modelo de Cox univariante, y se calculan las ratios de riesgo de las dos formas de administración del fármaco semaglutida frente a un placebo. Las ratios de riesgos se calculan en los artículos primarios para diferentes situaciones teniendo en cuenta los criterios de evaluación y diferentes variables pronóstico, aparte de la variable tratamiento. Como se ha indicado anteriormente, el test *log-rank* y su ratio de riesgos se calcularon llevando a cabo la estratificación para los factores que se controlaron en la aleatorización de los pacientes.

El protocolo es muy similar en ambos ensayos. El número de pacientes enrolados en el ensayo SUSTAIN 6 y que conforman el análisis es de 3297 (a 1648 se les administró semaglutida de manera subcutánea y 1649 recibieron un placebo). El tiempo de seguimiento fue de 2,1 años. En el ensayo PIONEER 6, el número de pacientes fue de 3183 (a 1591 se les administró el tratamiento de manera oral y 1592 recibieron un placebo). El tiempo mediano de seguimiento fue de 15,9 meses, menor que en SUSTAIN 6. La comparación directa de los resultados debería hacerse con cautela. En la **figura 9** aparecen los resultados de ambos ensayos en relación con el criterio primario definido por cada uno de los estudios.

La incidencia acumulada se estima mediante el método de Kaplan-Meier [recuérdese que $F(t) = 1 - S(t)$]. Por convenio, los gráficos de mortalidad o supervivencia presentan la escala de 0% al 100%, pero en estos casos la presencia de los eventos es tan rara que se ha acortado la escala para mostrar con mayor claridad las diferencias (véase la **figura 1**, en la que se muestra la gráfica original y la gráfica en la que se acorta la escala).

Debajo de la escala de tiempo en ambos ensayos se muestra la evolución de los individuos que está en riesgo de sufrir el evento cada 16 semanas. El descenso es debido tanto a la presencia del evento como a las censuras. El tiempo mediano de incidencia no está determinado porque sobreviven más del 50% de los individuos al finalizar el estudio; realmente, sobreviven por encima del 90% en ambos ensayos. La ratio de riesgos (HR) es de 0,74 en el caso del ensayo SUSTAIN 6 y 0,79 en el estudio PIONEER 6. El efecto es protector, dado que el riesgo de sufrir el evento es inferior con semaglutida en ambas formas de administración que con el placebo. Las curvas de incidencia del tratamiento están por debajo del placebo; si se presentara la supervivencia sería al revés, estarían por encima. La ratio de riesgos en SUSTAIN 6 es significativamente diferente de 1 y el intervalo de confianza no contiene al valor 1, mientras que en el ensayo PIONEER no es significativo porque el intervalo de confianza contiene al valor 1. En ambos casos no muestran el valor de p del *log-rank* debido a que realizan la evaluación de superioridad y no inferioridad que actualmente se lleva a cabo

Figura 9. Las curvas de incidencia y las ratios de riesgo de los ensayos clínicos SUSTAIN 6 y PIONEER 6



Fuente: Nauck MA, Quast DR. Cardiovascular safety and benefits of semaglutide in patients with type 2 diabetes: findings from SUSTAIN 6 and PIONEER 6. Front Endocrinol (Lausanne). 2021 Mar 29;12:645566.

en los ensayos clínicos. Esta evaluación cada vez es más habitual en los ensayos clínicos independientemente del análisis estadístico que se lleve a cabo. Surge de los inconvenientes que presenta la interpretación exclusiva del valor de p . En esta evaluación se debe considerar la ratio de riesgos que se supone que puede ser clínicamente relevante y partir de eso se llevan a cabo evaluaciones de superioridad, equivalencia y no inferioridad de tratamientos. En ambos ensayos se plantean las hipótesis de que el tratamiento es superior en eficacia al placebo, así como si no es inferior al placebo. A partir de la combinación de los resultados de ambas hipótesis se puede declarar si el tratamiento presenta mayor eficacia que el placebo, si el placebo es superior al tratamiento o si el tratamiento presenta la misma eficacia que el placebo. En el ensayo SUSTAIN 6, dado que ambos supuestos son significativos, hay evidencia de que la administración subcutánea de semaglutida es eficaz frente al placebo, mientras que en el caso de PIONEER 6, dado que el test de no inferioridad es significativo y el de superioridad no, nos indica que hay evidencia de que la administración oral de semaglutida no es inferior al placebo.

En los estudios originales se analizan también los criterios secundarios. Igualmente, dividen el criterio compuesto primario (ya que está formado por la presencia de varios eventos) en cada uno de los eventos. Estos análisis no mencionan el problema de riesgos competitivos ni la proporcionalidad de los riesgos, y también realizan un modelo de Cox multivariante incluyendo varias covariables, como el sexo, la región, la edad basal, la filtración glomerular, etc. Pretenden mostrar que el proceso de aleatorización ha funcionado y no existen factores confusores que interfieran en los resultados. La ratio de riesgo del modelo univariante con el factor tratamiento en ambos estudios no cambia drásticamente de la estimación con el modelo multivariante añadiendo las covariables; por lo tanto, se confirma que dichos factores no interfieren en los resultados.

Bibliografía

Husain M, Birkenfeld AL, Donsmark M, et al.; PIONEER 6 Investigators. Oral semaglutide and cardiovascular outcomes in patients with type 2 diabetes. *N Engl J Med*. 2019 Aug 29;381(9):841-51.

Marso SP, Bain SC, Consoli A, et al.; SUSTAIN-6 Investigators. Semaglutide and cardiovascular outcomes in patients with type 2 diabetes. *N Engl J Med*. 2016 Nov 10;375(19):1834-44. DOI: 10.1056/NEJMoa1607141.

Nauck MA, Quast DR. Cardiovascular safety and benefits of semaglutide in patients with type 2 diabetes: findings from SUSTAIN 6 and PIONEER 6. *Front Endocrinol (Lausanne)*. 2021 Mar 29;12:645566.

Preguntas de autoevaluación

1. En los estudios de supervivencia, una de las características del análisis de los datos es valorar las censuras. Respecto a las censuras, señale la afirmación incorrecta:

- a) El proceso de censura ha de ser aleatorio.
- b) Se denomina censura informativa cuando se reconoce un mecanismo de abandono del estudio.
- c) La mayoría de los estudios de supervivencia en el campo clínico consideran censuras a la derecha.
- d) La mayoría de los estudios de supervivencia en el campo clínico consideran censuras a la izquierda.

La respuesta correcta es la opción *d*.

La censura progresiva se produce en diferentes momentos del estudio. Dentro de la censura progresiva, puede tratarse de una censura a la derecha, cuando el tiempo de aparición del evento es posterior al último tiempo de revisión. En cambio, una censura a la izquierda ocurre cuando el tiempo de aparición del evento es anterior al tiempo de revisión, y censura por intervalo es cuando el tiempo de aparición del evento se produce entre dos tiempos de revisión. La mayoría de los estudios de supervivencia en el campo clínico consideran censuras a la derecha.

2. ¿Cuál de los siguientes no es uno de los criterios de valoración sustitutivos más utilizados?

- a) Supervivencia libre de progresión (*progression-free survival*, PFS): tiempo hasta la progresión de la enfermedad o la muerte por cualquier causa.
- b) Supervivencia por causa específica (*cause-specific survival*, CSS).
- c) Supervivencia libre de enfermedad (*disease-free survival*, DFS): tiempo hasta la recurrencia o la muerte por cualquier causa.
- d) Tiempo hasta la progresión (*time to progression*, TIP): tiempo hasta la progresión de la enfermedad.

La respuesta correcta es la opción *d*.

La supervivencia por causa específica es un criterio clínico o primario. Este tipo de criterio se considera validado y objetivo a la hora de estimar el efecto en la supervivencia de una enfermedad. El evento de interés valorado es la muerte por la enfermedad sometida a estudio. El problema del uso de este criterio es la presencia de riesgos competitivos, y por ello actualmente esta valoración tiene menos uso.

COVID-19 y factores pronósticos

Dr. José Manuel Casas Rojo
Servicio de Medicina Interna
Hospital Universitario Infanta Cristina (Parla, Madrid)

La publicación seleccionada de este capítulo es un trabajo basado en datos reales no publicado ni revisado por pares, creado por el autor como base para el desarrollo del tema del capítulo.



capítulo 5

Publicación seleccionada

Factores pronósticos de mortalidad en pacientes hospitalizados por COVID-19

Introducción

La investigación de los factores asociados a un mal pronóstico en los pacientes hospitalizados por COVID-19 debe permitir identificar a los pacientes más vulnerables para aplicar en ellos cuidados más estrictos y tratamientos más agresivos. Existe evidencia por otros estudios de que el sexo femenino puede tener un efecto protector sobre la mortalidad. También se sabe que la mortalidad es mayor en pacientes de edad avanzada. Dado que la edad media de las mujeres suele ser mayor que la de los hombres, se puede plantear si la edad es un factor de confusión o modificador del efecto del sexo sobre la mortalidad de los pacientes hospitalizados por COVID-19.

Métodos

Diseño: Estudio observacional multicéntrico retrospectivo sobre una cohorte de pacientes hospitalizados por COVID-19 en hospitales españoles.

Objetivo: Elaborar un modelo predictivo de mortalidad de los pacientes hospitalizados por COVID-19 basado en su comorbilidad. Como objetivo secundario, comprobar si el sexo femenino es un factor protector de mortalidad en estos pacientes.

Análisis de los datos: La edad se expresa como media y desviación estándar y se compara en los grupos mediante la prueba *t* de Student. Las variables categóricas se expresan mediante porcentajes y se comparan mediante la prueba de ji al cuadrado (χ^2).

Para la elaboración del modelo predictivo, se incluyen las variables cuya asociación con la mortalidad presenta un valor *p* menor o igual a 0,20, además de la interacción edad-sexo.

Se realiza una selección de variables mediante inclusión por pasos (*forward stepwise*), fijándose la significación para la inclusión de variables en 0,05 y para la exclusión en 0,10. Se calcula el área bajo la curva ROC (*receiver operating characteristic*) del modelo.

Adicionalmente, se estima otro modelo, basado en el mismo modelo máximo, para medir el efecto de la variable sexo femenino sobre el riesgo de mortalidad. La interacción sexo-edad se descarta por no ser significativa en una prueba de razón de verosimilitud. El ajuste de los factores de confusión se realiza estimando todas las combinaciones posibles de variables para elegir un modelo final que no modifique de forma importante (en más de un 10%) la estimación ajustada de la razón de *odds* (*odds ratio*, OR) del sexo femenino y mejore la precisión de la estimación.

Se calcula también el riesgo relativo ajustado del sexo femenino a partir de las estimaciones marginales de la probabilidad de cada evento.

Resultados

Se analizaron los datos de un total de 20289 pacientes ingresados en hospitales españoles por COVID-19. La mortalidad global de la serie fue del 18,8%. En la **tabla 1** se muestra la comparación de la mortalidad, la edad y la comorbilidad entre varones y mujeres, donde se observa que existe diferencia significativa en las medias y proporciones de diversas variables. La diferencia de mortalidad entre varones y mujeres fue de 3,76%; el riesgo relativo (RR) se situó en 0,82.

Variable	Varones	Mujeres	p
Fallecimiento %	20,43	16,67	<0,001
Edad media (desviación estándar) años	65,76 (15,59)	68,25 (16,6)	<0,001
Dislipemia %	38,46	38,38	0,903
Diabetes %	21,3	18,31	<0,001
Obesidad %	21,6	23,74	<0,001
Cardiopatía %	13,43	10,32	<0,001
Enfermedad respiratoria crónica %	16,46	13,85	<0,001
Enfermedad cerebrovascular %	2,99	2,79	0,41
Demencia %	6,75	11,93	<0,001
Enfermedad arterial periférica %	4,69	3,01	<0,001
Hepatopatía grave %	1,32	0,65	<0,001
Enfermedad renal grave %	6,06	5,71	0,293
Neoplasia maligna %	10,98	8,45	<0,001

En la **tabla 2** se muestra la asociación de diversas variables con la mortalidad. El sexo femenino tiene un efecto no ajustado protector frente a la mortalidad (OR: 0,78; intervalo de confianza al 95% [IC95%]: 0,73-0,84).

El modelo predictivo final de mortalidad, tras la selección de variables por pasos, excluye la interacción edad-sexo y la dislipemia. El modelo (**tabla 3**) tiene una buena capacidad predictiva, con un área bajo la curva ROC de 0,80 (0,79-0,80).

En el modelo máximo para la estimación del efecto del sexo sobre la mortalidad se incluyeron las mismas variables que en el modelo predictivo. La interacción sexo-edad se descartó por no ser significativa en la prueba de razón de verosimilitud. La OR ajustada del sexo para la mortalidad en este modelo fue de 0,69 (IC95%: 0,57-0,67). En la **tabla 4** se muestra el modelo reducido final de regresión logística elegido

Tabla 2. Asociación de la mortalidad con el sexo, la edad y la comorbilidad (N = 20 289)

Variable	Vivos	Fallecidos	OR	IC95% de la OR	p
Sexo mujer %	43,87	37,86	0,78	0,73-0,84	<0,001
Edad media (desviación estándar) años	64,07 (15,81)	78,7 (11,02)			<0,001
Dislipemia %	35,82	49,67	1,77	1,65-1,9	<0,001
Diabetes %	17,84	29,41	1,92	1,77-2,08	<0,001
Obesidad %	22,48	22,68	1,01	0,93-1,1	0,791
Cardiopatía %	9,28	24,27	3,13	2,86-3,43	<0,001
Enfermedad respiratoria crónica %	14,09	20,74	1,6	1,46-1,74	<0,001
Enfermedad cerebrovascular %	2,17	6,07	2,92	2,47-3,46	<0,001
Demencia %	6,11	21,29	4,16	3,76-4,6	<0,001
Enfermedad vascular periférica %	3,09	7,78	2,64	2,28-3,06	<0,001
Hepatopatía grave %	0,87	1,73	2,01	1,5-2,69	<0,001
Enfermedad renal grave %	4,21	13,22	3,46	3,07-3,91	<0,001
Neoplasia maligna %	8,31	16,73	2,22	2-2,45	<0,001

para valorar el efecto ajustado del sexo sobre la mortalidad, con una OR ajustada del sexo mujer de 0,613 (0,566-0,665). El riesgo relativo ajustado es de 0,716 (0,677-0,757).

Conclusión

La edad, el sexo y la comorbilidad permiten predecir el riesgo de mortalidad de los pacientes ingresados por COVID-19. El sexo femenino es un factor independiente de protección frente a la mortalidad por todas las causas en pacientes hospitalizados por COVID-19. El efecto protector del sexo femenino no se modifica significativamente por la edad.

Nota: La publicación que se ha presentado se basa en datos del Registro SEMI-COVID-19, modificados con la intención de que sirvan de ejemplo para la explicación del presente capítulo. No está revisado por pares ni debe considerarse como estudio de referencia.

Tabla 3. Modelo predictivo de mortalidad basado en el sexo, la edad y la comorbilidad (N = 20 289)

Fallecimiento	OR ajustada	IC95% de la OR	p
Sexo mujer	0,62	0,57-0,67	<0,001
Edad (años)	1,07	1,07-1,08	<0,001
Diabetes	1,24	1,13-1,36	<0,001
Cardiopatía	1,47	1,33-1,63	<0,001
Enfermedad respiratoria crónica	1,24	1,12-1,37	<0,001
Enfermedad cerebrovascular	1,38	1,14-1,66	0,001
Demencia	1,62	1,44-1,82	<0,001
Enfermedad vascular periférica	1,24	1,05-1,47	0,010
Hepatopatía grave	1,74	1,26-2,40	0,001
Enfermedad renal grave	1,79	1,56-2,04	<0,001
Neoplasia maligna	1,60	1,43-1,79	<0,001
Constante	0,00	0,00-0,00	<0,001

Tabla 4. Modelo reducido para medir el efecto del sexo sobre la mortalidad (N = 20 289)

Fallecimiento	OR ajustada	IC95% de la OR	p
Sexo mujer	0,61	0,57-0,66	<0,001
Edad (años)	1,08	1,08-1,08	<0,001
Cardiopatía	1,57	1,42-1,73	<0,001
Enfermedad respiratoria crónica	1,21	1,09-1,33	<0,001
Enfermedad vascular periférica	1,47	1,25-1,72	<0,001
Constante	0,00	0,00-0,00	<0,001

Introducción

Hemos visto un artículo ilustrativo sobre el análisis de factores pronósticos de una enfermedad, la COVID-19 en pacientes hospitalizados. Se obtienen medidas de asociación crudas y también ajustadas para dos modelos multivariantes que ejemplifican los dos propósitos para los que se suelen construir estos modelos: el predictivo y la medición de un efecto de una variable sobre un resultado. No es habitual incluir modelos con los dos propósitos en un mismo artículo, pero lo hemos hecho con fin didáctico para las explicaciones teóricas que siguen.

Modelos de análisis estadístico

Los estudios de investigación permiten realizar comparaciones y establecer asociaciones entre una exposición y una respuesta. Los modelos de regresión permiten utilizar múltiples predictores de la respuesta, tanto categóricos como cuantitativos. Cada tipo de respuesta tiene un modelo general de regresión:

- Para respuestas cuantitativas continuas, el modelo general es la regresión lineal múltiple. Se pueden explorar asociaciones mediante la comparación de medias (exposición binaria), análisis de la variancia (exposición categórica con más de dos categorías) o regresión simple (exposición continua).
- Para respuestas categóricas se realizan comparaciones de proporciones mediante pruebas de χ^2 ; el modelo general es la regresión logística binaria (o multinomial para exposiciones con más de dos categorías), que, como todos los modelos de regresión, admite predictores múltiples tanto cuantitativos como categóricos.
- Si la respuesta son datos de supervivencia, se pueden comparar curvas de supervivencia de grupos independientes; el modelo general es la regresión de riesgos proporcionales de Cox, también con predictores múltiples cuantitativos y categóricos.
- Finalmente, para recuentos (variables cuantitativas discretas), el modelo adecuado es la regresión de Poisson y binomial negativa.

En este trabajo nos referiremos fundamentalmente a las medidas de asociación entre una exposición y una respuesta binarias, mediante la comparación de proporciones y el modelo de regresión logística.

Medidas de asociación entre una exposición y una respuesta binarias

Diferencia de proporciones

La medida de asociación más básica entre una exposición y una respuesta (ambas binarias) es la diferencia de proporciones. Es poco informativa por sí sola porque su importancia depende de la proporción de casos en los sujetos no expuestos. En el ejemplo (**tabla 1**), siendo la exposición el sexo mujer, una diferencia de un 3,76% en la mortalidad se puede considerar importante para un riesgo en no expuestos (varones) del 20,43%.

Índices de razón

Sobre la clásica tabla de 2×2 explicamos las medidas de asociación de razón más habituales. La exposición es el sexo (tomando como referencia el sexo masculino) y el resultado es el fallecimiento: los casos son los fallecidos y los no casos son los no fallecidos.

	Varón (referencia)	Mujer	
Fallecido	a_0 (2373)	a_1 (1446)	m_1 (3819)
No fallecido	b_0 (9244)	b_1 (7226)	m_0 (16470)
	n_0 (11617)	n_1 (8672)	n (20289)

Riesgos:

- Riesgo de fallecimiento en varones: $a_0/n_0 = 2373/11617 = 0,204$.
- Riesgo de fallecimiento en mujeres: $a_1/n_1 = 1446/8672 = 0,167$.
- Riesgo relativo (mujeres respecto a varones): $0,167/0,204 = 0,818$.

Odds:

- Odds de fallecimiento en varones: $a_0/b_0 = 2373/9244 = 0,257$.
- Odds de fallecimiento en mujeres: $a_1/b_1 = 1446/7226 = 0,200$.
- Razón de odds (OR): $0,200/0,257 = 0,778$.

Riesgos y riesgo relativo

En estudios experimentales y de cohortes, a las proporciones de casos en expuestos y no expuestos las llamamos riesgos. El riesgo relativo (RR) es la medida de asociación más adecuada. Tiene una interpretación sencilla: un RR de 0,82 significa que el riesgo en expuestos (mujeres) se multiplica por 0,82 respecto a los varones. Su inverso ($1/0,82 = 1,22$) indica que el riesgo en los varones es 1,22 veces superior que en las mujeres.

Es frecuente expresar el RR como porcentaje de incremento relativo del riesgo:

$$(\text{RR} - 1) \cdot 100$$

Si la referencia son los varones, un RR de 0,82 indica que el riesgo en mujeres se reduce un 18% respecto al de los varones.

$$(0,82 - 1) \cdot 100 = -18\%$$

Si la referencia es el riesgo en mujeres, un RR de 1,22 significa que en varones el riesgo se incrementa un 22% respecto al de las mujeres.

$$(1,22 - 1) \cdot 100 = 22\%$$

En estudios transversales lo comentado es equiparable, pero en estos casos hablamos de prevalencias (en lugar de riesgos) y de razón de prevalencias (en lugar de RR).

Es importante destacar que el RR solo se puede calcular cuando la proporción entre casos y no casos es real, como en los estudios experimentales y en los de cohortes, puesto que se eligen los sujetos por la exposición y se comprueba cuántos desarrollan el resultado. Pero no es así en los estudios de casos y controles, donde el número de sujetos que presentan o no el evento está preestablecido y lo que se comprueba es cuántos de ellos presentaban la exposición.

Odds y razón de odds

La *odds* es el cociente entre la probabilidad de ocurrencia y de no ocurrencia del evento (en nuestro ejemplo, el fallecimiento). Es un término anglosajón utilizado en el lenguaje de las apuestas. El cociente entre la *odds* en expuestos y en no expuestos es la razón de *odds* u *odds ratio* (OR) y no se interpreta igual que el riesgo relativo. Una OR de 1,28 en varones respecto a mujeres significa que la *odds* (no el riesgo) de fallecimiento (razón fallecidos/no fallecidos) es 1,28 veces mayor en varones que en mujeres. También se puede expresar como porcentaje de incremento relativo, como en el caso del RR.

$$(OR - 1) \cdot 100$$

La OR tiene unas propiedades matemáticas muy útiles:

1. Se puede calcular sin conocer la proporción de casos frente a no casos. Esto tiene interés en los estudios de casos y controles, donde, como hemos comentado, no se conoce esa proporción.
2. Cuando los riesgos son pequeños, es decir, para eventos poco frecuentes, la OR coincide prácticamente con el RR y es una manera de estimarlo indirectamente.

La OR es la medida que se suele dar en los análisis multivariantes mediante regresión logística, porque se calcula fácilmente a partir del exponencial del coeficiente que indica el efecto de un factor pronóstico. En estudios experimentales y de cohortes es posible calcular también el RR ajustado a partir de las estimaciones marginales de la probabilidad de cada evento.

En los análisis de supervivencia, donde lo que se compara son tasas (eventos por sujeto y unidad de tiempo), la medida de asociación es la razón de tasas (*hazard ratio*, HR).

Ajuste multivariante de las medidas de asociación

Las medidas de asociación crudas solo tienen interés como medidas de efecto en estudios experimentales aleatorizados (ensayos clínicos), porque en ellos los grupos que se comparan son homogéneos para otras variables que pueden influir también en el resultado.

La ausencia de aleatorización en los estudios de cohortes y de casos y controles obliga a realizar un ajuste de la medida de asociación cruda por otras variables que también tienen efecto en el resultado. Esto se consigue mediante el análisis multivariante, que admite la introducción de variables explicativas tanto cuantitativas como categóricas. Ya se han comentado los diferentes modelos de regresión según el tipo de exposición.

Nuestro ejemplo es un modelo de regresión logística binaria. La variable dependiente (resultado) es el fallecimiento por todas las causas en pacientes hospitalizados por COVID-19. Para cualquiera de las variables incluidas en el modelo multivariante se estima una medida de efecto, la OR. Esta medida de efecto se interpreta como el factor por el que se multiplica la *odds* de fallecimiento cuando todas las demás variables permanecen constantes. Así, por ejemplo, una OR ajustada de 0,61 para el sexo femenino (**tabla 4**) indica que ser mujer multiplica por 0,61 la *odds* de fallecer (respecto a ser varón), es decir, tiene un efecto protector.

En los modelos de regresión logística (y en el resto de los análisis multivariante) se pueden introducir también interacciones entre variables para mejorar el ajuste. Su cálculo consiste en la estimación de un coeficiente para el producto de dos variables. El efecto de una variable que está incluida en una interacción debe calcularse teniendo en cuenta esta.

Los modelos multivariante se pueden estimar con dos propósitos distintos: para realizar predicciones o para medir el efecto de una variable concreta sobre el resultado.

Modelos para medir un efecto

En los modelos para medir un efecto de una exposición concreta sobre el resultado, el resto de los predictores se interpretan como variables de confusión y las interacciones como modificadores del efecto. Son dos conceptos importantes que pasamos a explicar sobre nuestro ejemplo.

Confusión

Una variable de confusión (por ejemplo, la edad), debe cumplir tres condiciones para considerarse como tal:

1. Debe mostrar asociación con el efecto en los sujetos no expuestos. En nuestro ejemplo, asumimos que la edad se asocia también con la mortalidad en los varones.
2. Debe estar desajustada respecto a la variable frente a la que sospechamos que actúa como factor de confusión. En nuestro ejemplo, las mujeres tienen más edad que los hombres, lo que quizá pueda confundir el efecto protector independiente del sexo femenino.
3. El posible factor de confusión debe estar presente antes de producirse el resultado. Es obvio que el efecto de la edad es previo al fallecimiento.

Las posibles variables de confusión se introducen como tales en el modelo, de manera que la OR de la variable cuyo efecto queremos medir queda ajustada por todas las demás presentes en el modelo. Este efecto ajustado puede reforzar o atenuar la medida cruda calculada previamente. Obtenemos así el efecto independiente de la variable estudiada.

Modificación del efecto (interacción)

Las interacciones se interpretan como modificadores del efecto. La edad actuaría como modificador del efecto del sexo femenino si a medida que aumenta se modifica el efecto protector del sexo femenino. Si existe una interacción, el cálculo de la OR ajustada se realiza a partir del valor de la variable estudiada, del producto entre este y el valor de la variable de interacción y de la OR de la variable estudiada y la interacción.

En nuestro ejemplo, la interacción sexo femenino-edad no es significativa, por lo que se debe excluir del modelo. A efectos explicativos, hemos estimado el modelo con la interacción, siendo la OR del sexo femenino (OR_{SF}) de 0,557 y la de la interacción sexo femenino-edad ($OR_{SF\#Edad}$) de 1,001. El cálculo se haría mediante la fórmula siguiente para un conjunto de edades representativas:

$$\text{Efecto "Sexo femenino"} = OR_{SF} \cdot OR_{SF\#Edad}^{(1 \cdot \text{Edad})}$$

Nótese que la OR de la interacción sexo femenino-edad se eleva al producto de los valores de las variables "Edad" y "Sexo" (que toma el valor 1 para el sexo femenino). Para un conjunto de edades representativas:

Edad (años)	$OR_{SF} \cdot OR_{SF\#Edad}^{(1 \cdot \text{Edad})}$
30	0,57
50	0,59
70	0,60
90	0,61
$OR_{SF} = 0,557; OR_{SF\#Edad} = 1,001$	

Si la interacción fuese significativa, se puede decir que, a medida que se incrementa la edad, el efecto protector del sexo femenino se reduce (la OR ajustada se acerca más a 1).

Las variables que forman parte de una interacción significativa deben incluirse en el modelo, sean o no variables de confusión (en este caso, la edad).

Los modelos estimados para medir un efecto se construyen partiendo de un conjunto máximo de variables que incluyen, aparte de la variable dependiente (fallecimiento en nuestro caso) y el factor cuyo efecto se pretende medir (el sexo femenino), todas aquellas variables que se considera que pueden ser factores de confusión y todas las interacciones con el factor de estudio que se considera que pueden producir modificación del efecto. Para obtener el modelo definitivo, se avanza de la siguiente manera:

- Se eliminan las interacciones que no sean estadísticamente significativas. Suele realizarse una prueba de razón de verosimilitud global (*chunk test*) entre el modelo máximo y el modelo sin interacciones. Si la prueba no da resultado significativo, pueden eliminarse todas las interacciones. Si es significativo, hay que ir probando con los diferentes modelos que se obtienen al eliminar cada interacción individual.
- Una vez eliminadas las interacciones no significativas, tenemos un modelo de referencia que contiene todos los posibles factores de confusión. Es posible obtener un modelo más sencillo excluyendo solo aquellas variables cuya ausencia en el modelo no produzca un cambio clínicamente importante en la estimación de la OR del efecto; habitualmente, el límite es un 10%. Si varios modelos cumplen esta condición, se

elige aquel que estime de forma más precisa la OR ajustada, es decir, con un intervalo de confianza más estrecho.

En nuestro ejemplo, tras el ajuste de la confusión, se ha elegido un modelo reducido en el que se prescinde de algunas variables sin repercutir de forma importante en la OR del sexo femenino en el modelo de referencia (**tabla 4**).

Como veremos, esta forma de seleccionar predictores es distinta de la utilizada en los modelos con propósito predictivo.

Modelos con intención predictiva

En los modelos estimados con intención predictiva se trata de incluir todos los predictores e interacciones que puedan tener influencia en la variable dependiente. En este caso no hablamos de confusión o modificación del efecto, porque todos los predictores tienen la misma importancia.

A partir del modelo máximo así construido, se elabora un modelo más sencillo en el que se incluyen solo las variables e interacciones cuyo efecto muestra significación estadística.

Existen varios métodos para obtener el modelo final: inclusión o exclusión secuencial de variables y selección paso a paso hacia adelante o hacia atrás. El método de elección es estimar todas las posibles ecuaciones combinando los predictores e interacciones y seleccionar el mejor de acuerdo con determinados criterios de ajuste. En regresión logística los criterios suelen ser el área bajo la curva ROC (AUC) y el criterio de información de Akaike (AIC).

El modelo predictivo de regresión logística se puede interpretar como una prueba diagnóstica. Se puede estimar la probabilidad del resultado (fallecimiento) dando un valor a cada una de las variables del modelo. También es posible estimar el AUC, que es un buen índice de validez del modelo. En nuestro ejemplo, un AUC de 0,80 (0,79-0,80) indica que el modelo clasifica bien a los pacientes fallecidos y no fallecidos en función del conjunto de variables predictoras.

Bibliografía

Doménech Massons JM. Fundamentos de diseño y estadística. 20.^a ed. Barcelona: Signo; 2019. Unidad 9, Comparación de dos proporciones. Medidas de asociación.

Doménech Massons JM, Navarro Pastor JB. Metodología de la investigación en ciencias de la salud. Regresión logística binaria, multinomial, de Poisson y binomial negativa. 13.^a. Barcelona: Signo; 2020.

García Garmendia JL, Maroto Monserrat F. Interpretation of statistical results. Med Intensiva (Engl Ed). 2018;42(6):370-9.

Kleinbaum DG, Klein M. Logistic regression [Internet]. New York, NY: Springer New York; 2010 [consultado el 26 de mayo de 2021]. Disponible en: <http://link.springer.com/10.1007/978-1-4419-1742-3>.

Miettinen O. Confounding and effect-modification. Am J Epidemiol. 1974;100(5):350-3.

Pita Fernández S, Valdés Cañedo FA. Determinación de factores pronósticos. [Actualizado el 18 de enero de 2001; consultado el 3 de junio de 2021] [Internet]. Madrid: Elsevier España; c2021. Disponible en: <https://www.fisterra.com/formacion/metodologia-investigacion/determinacion-factores-pronosticos>.

Preguntas de autoevaluación

1. En relación con los modelos de regresión logística para medir un efecto, señale la afirmación correcta:

- a) La elección de los factores de confusión se realiza según su significación estadística.
- b) Pueden eliminarse los términos de interacción que no modifiquen de forma importante la estimación de la *odds ratio* (OR) ajustada.
- c) Pueden eliminarse los términos de confusión que no modifiquen de forma importante la estimación de la OR ajustada.
- d) Todas las afirmaciones anteriores son incorrectas.

La respuesta correcta es la opción *c*.

En modelos con intención predictiva, tanto predictores como interacciones se eligen por su significación estadística, mediante inclusión o exclusión secuencial, por pasos o estimando todas las posibles ecuaciones y eligiendo el modelo mejor ajustado. En los modelos para medir un efecto, los factores modificadores del efecto (interacciones) se eligen por su significación estadística, pero las variables de confusión solo se eliminan si su ausencia no modifica de forma importante la estimación de la OR.

2. Señale la afirmación correcta sobre las medidas de asociación.

- a) El riesgo relativo (RR) es de elección como medida de asociación en estudios de casos y controles.
- b) La *odds ratio* (OR) coincide prácticamente con el RR para frecuencias de efecto pequeñas.
- c) El RR no se puede calcular en los estudios experimentales y de cohortes.
- d) Todas las afirmaciones anteriores son incorrectas.

La respuesta correcta es la opción *b*.

El RR no se puede calcular cuando la proporción de casos-no casos es artificial, como ocurre en estudios de casos y controles. Sí es correcto calcularlo en estudios experimentales (ensayos clínicos) y de cohortes, donde se elige a los pacientes por la exposición y las proporciones de casos-no casos son las que se encuentran en la población de origen de la muestra. Cuando las frecuencias son muy bajas, las *odds* coinciden prácticamente con los riesgos y la OR es equiparable al RR.

3. Señale la afirmación correcta en relación con la confusión y la modificación del efecto:

- a) Son términos aplicables a los modelos con propósito predictivo.
- b) Los factores de confusión no deben estar presentes antes de que se produzca el efecto.
- c) Los términos de interacción modifican el efecto ajustado que se investiga.
- d) Todas las afirmaciones anteriores son incorrectas.

La respuesta correcta es la opción c.

Una variable de confusión debe estar asociada al efecto, debe estar desajustada en los dos grupos que se comparan y debe estar presente antes de producirse el efecto. Una variable que forma parte de una interacción con el factor estudiado modifica su efecto, lo que hace necesario dar diferentes medidas del efecto para los diferentes valores de la variable modificadora. Tanto la confusión como la modificación son conceptos propios de los modelos estimados para medir un efecto. En los modelos con intención predictiva todos los predictores e interacciones tienen la misma importancia.

Seguimiento a largo plazo de una cohorte (Framingham/UKPDS)

Dra. Pilar Román Sánchez

Jefe de Servicio de Medicina Interna
Hospital Asociado Universitario de Requena (Valencia)

Expresidenta
Sociedad Española de Medicina Interna
Fundación Española de Medicina Interna



capítulo 6

Publicación seleccionada

Asociación del hábito tabáquico con la enfermedad de Alzheimer

Introducción

La esperanza de vida ha aumentado considerablemente en las últimas décadas y una de sus consecuencias es el aumento del número de personas que padecen demencia tipo Alzheimer. La demencia es una de las patologías con mayor impacto en la calidad de vida del paciente y su entorno. Supone un alto coste a la sociedad y a los sistemas sanitarios. No existe evidencia sólida sobre el efecto que el tabaquismo puede tener en la demencia tipo Alzheimer.

Objetivo

Conocer la asociación del consumo de tabaco con la demencia tipo Alzheimer en una cohorte multidimensional de 28 756 personas mayores de 55 años de la población general en Italia.

Material y métodos

Población: Los participantes fueron elegidos por el código postal entre la población italiana de mayores de 55 años y que manifestaron su conformidad con el estudio. Fueron excluidos los sujetos con ictus o patologías psiquiátricas. Se seleccionó a 14 881 personas con historia de consumo de tabaco (actual o en el pasado) y 13 875 personas que no habían fumado nunca.

Consumo de tabaco: El consumo de tabaco activo o pasado en años-paquete fue notificado por los participantes en entrevistas presenciales periódicas a lo largo de 10 años.

Función cognitiva: La función cognitiva se valoró al inicio del estudio, periódicamente durante su desarrollo y al final de este, mediante test de memoria (recuerdo inmediato y remoto de 10 palabras, la suma de las correctas en puntuación de 1 a 20), función ejecutiva (contar hacia atrás de 7 en 7, un punto por cada acierto; copiar pentágonos entrecruzados, 3 puntos si se resuelve, 0 si no) y orientación (4 preguntas sobre el día de la semana, el día del mes, el mes y el año, un punto por cada acierto) por personas ajenas al estudio. Se utilizó un test *Z-score* para comparar las distintas evaluaciones cognitivas.

Covariables analizadas: Sexo, edad, índice de masa corporal, ejercicio físico, consumo de alcohol, tensión arterial, diabetes, enfermedades cardiovasculares, enfermedades respiratorias, nivel de educación, nivel de ingresos, soledad, escala de depresión, portadores del gen APOE ε4.

Análisis estadístico: Los resultados se presentan como porcentaje en las variables discretas y como media en las continuas. La tasa de incidencia de eventos por personas-año se calculó con razón de tasas al 95% de intervalo de confianza. Se ajustaron las variables analizadas mediante la regresión de Poisson.

Resultados

Se analizaron los datos de 24 665 personas (243 presentaron deterioro cognitivo previo, patología cerebrovascular o psiquiátrica; 553 fallecieron, 331 de causa cardiovascular; 3 295 se perdieron).

Edad media de $64,1 \pm 9,2$ años; el 55,3% eran mujeres; la media de seguimiento fue de 9,1 años. Durante las 100 000 personas-año de seguimiento, los fumadores de 27-40 años-paquete (*hazard ratio* [HR]: 2,55; intervalo de confianza al 95% [IC95%]: 1,22-5,58) y los de más de 40 (HR: 2,92; IC95%: 1,37-6,53) tenían un significativo mayor riesgo de padecer demencia tipo Alzheimer que los fumadores de menos de 27 años-paquete. Estos últimos presentaron mayor riesgo que los no fumadores (HR: 1,82; IC95%: 1,26-2,57). No hubo diferencias significativas entre los fumadores en activo y los exfumadores.

Conclusión

Los fumadores, activos o en el pasado, tienen significativamente mayor riesgo de padecer demencia tipo Alzheimer que quienes no han fumado nunca. El riesgo aumenta directamente con la intensidad del consumo.

Estudios de cohortes

Una cohorte es un grupo de individuos con alguna característica común que se sigue a lo largo del tiempo; de este modo, los estudios de cohortes valoran la ocurrencia de eventos a lo largo del tiempo. Facilitan la comprensión de cómo múltiples factores que actúan a lo largo del tiempo pueden determinar la etiología, la historia natural y el curso clínico de una enfermedad o evento de interés. Así, permiten estudiar la naturaleza dinámica de diversos factores de riesgo en el tiempo para la aparición de una o varias enfermedades o eventos de interés con fines descriptivos o con el propósito de probar hipótesis relacionadas con la enfermedad o el evento de interés.

En un estudio de cohortes se recluta a un grupo de individuos que no presentan en ese momento el evento de interés. Todos los integrantes deben estar en riesgo de padecer o presentar dicho evento. Para su inclusión los individuos se clasifican de acuerdo con las características (factores de riesgo) que podrían guardar relación con el resultado. Posteriormente, estos individuos son observados a lo largo del tiempo para cuantificar cuáles de ellos manifiestan el resultado.

Los estudios de cohortes son observacionales (el investigador no interviene en los sujetos estudiados excepto en su selección), longitudinales (se sigue a los sujetos a lo largo del tiempo) y analíticos (se puede analizar la asociación de los factores de riesgo con el evento de interés y calcular la magnitud de dicha asociación). Miden la incidencia del evento de interés según el factor de riesgo.

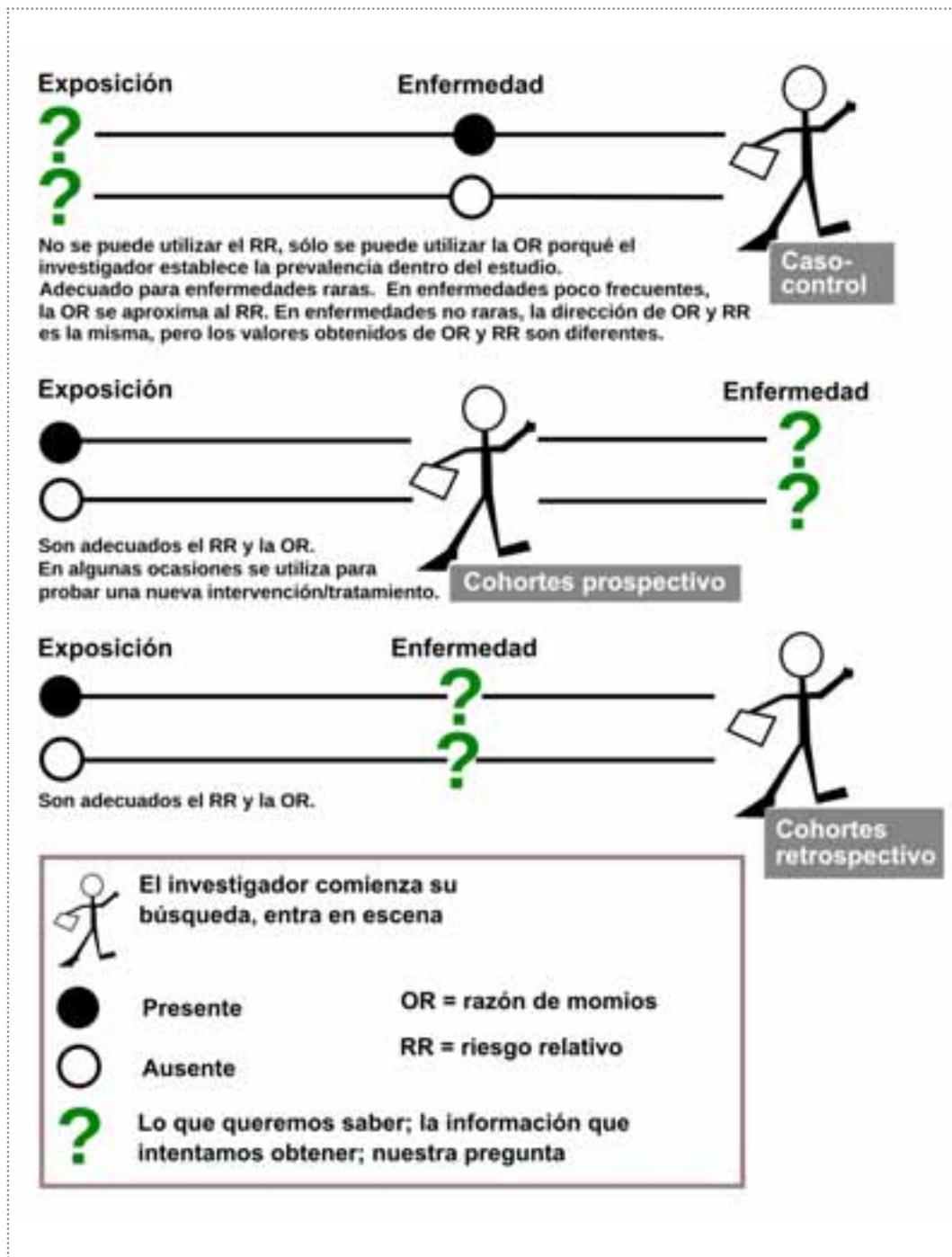
El otro tipo de estudio observacional, longitudinal y analítico es el de casos y controles. En ellos se parte de un grupo de individuos enfermos (casos) y se busca un grupo de individuos sanos (controles) para comparar la posible presencia de determinados factores de riesgo. Los controles deben ser comparables a los casos. No se puede estimar la incidencia, y se calcula la magnitud del riesgo por medio de la *odds ratio*. Son más rápidos y baratos que los estudios de cohortes, permiten estudiar enfermedades poco frecuentes y pueden evaluar múltiples factores de riesgo para una misma enfermedad. Los problemas más importantes en los estudios de casos y controles son los sesgos de memoria y encontrar un grupo de control adecuado (**figura 1**).

Historia de los estudios de cohortes

Los estudios de cohortes de hoy provienen de las tablas de vida construidas a partir de datos de mortalidad que surgieron entre 1600 y 1700.

En 1927, Wade Frost, en la Universidad Johns Hopkins, acuñó el concepto de cohorte para estudiar los factores asociados a la mortalidad por tuberculosis. En 1940 se aplicaron estos métodos para estudiar las consecuencias del tabaquismo en los médicos; y en la década de los 50 se estudiaron los riesgos de radiación en supervivientes de las bombas atómicas de Hiroshima y Nagasaki y se inició el famoso estudio de Framingham para conocer las causas de la enfermedad cardíaca. Todos conocemos famosos estudios de cohortes, como el UKPDS, el Nurses' Health Study, el Olmstead County Minnesota, etc. Han aportado conocimiento en la epidemiología clínica, molecular, genética, etc.

Figura 1. Diseños de estudios observacionales: casos y controles frente a estudios de cohortes



Actualmente, debido al desarrollo de los análisis estadísticos e informáticos y al *big data*, se ha evolucionado a tal nivel que se ha logrado realizar estudios integrados epidemiológicos y clínicos, por ejemplo, sobre el cáncer.

Por otra parte, los grandes registros prospectivos de pacientes están proporcionando valiosos datos epidemiológicos con menor dificultad que los estudios de cohortes. A diferencia de estos últimos, que son poblacionales, los registros se basan en personas enfermas. Ambos son muy útiles para las enfermedades poco frecuentes. La validez de los resultados de los registros dependerá de la fiabilidad de los datos.

Tipos de estudios de cohortes

Hay distintos tipos de estudios de cohortes según la perspectiva temporal y la población con la que se trabaja:

- Según la relación temporal del inicio del estudio con la ocurrencia del evento:
 - **Prospectivo o concurrente:** Los participantes están libres del evento de interés al inicio del estudio.
 - **Retrospectivo o histórico:** El evento de interés se ha presentado antes del estudio. Se utiliza información pasada de registros clínicos. El análisis estadístico se realiza como en un estudio de cohorte típico. Su validez dependerá de la calidad de los registros utilizados.
- Según el tipo de población:
 - **De cohortes cerradas o fijas:** No se incorporan individuos más allá del periodo de reclutamiento.
 - **De cohortes dinámicas:** Consideran la entrada y salida de nuevos sujetos del estudio durante la fase de seguimiento cuando cumplen criterios de elegibilidad, y se incorpora la aportación años-persona desde el momento de inclusión en el estudio.

Los objetivos en los estudios de cohortes

Los estudios de cohortes tienen como objetivo determinar la asociación entre un factor de exposición y el desarrollo de alguna enfermedad o evento de interés a partir de la estimación de riesgos, así como conocer la historia natural o el curso clínico de un evento de interés, determinar la supervivencia y estudiar los factores protectores y pronósticos (**tabla 1**).

Ventajas y desventajas de los estudios de cohortes

Los estudios de cohortes presentan las ventajas y desventajas que se muestran en la **tabla 2**.

Lectura crítica de un estudio de cohortes

Se han desarrollado varios instrumentos para evaluar los estudios de cohortes. En España, el grupo CASPe ha desarrollado un programa de lectura crítica, basado en la realización de

Tabla 1. Objetivos de los estudios de cohortes

<i>Características en común</i>	<i>Aspecto que se evalúa</i>	<i>Ejemplo</i>
Edad	Edad	Esperanza de vida de los individuos de 70 años (con independencia de cuándo nacieron)
Fecha de nacimiento	Cohorte	Tasa de cáncer cervical para mujeres nacidas en 1910
Exposición	Factor de riesgo	Cáncer de pulmón en individuos que fuman
Enfermedad	Pronóstico	Tasa de supervivencia de mujeres con cáncer cervical
Intervención preventiva	Prevención	Disminución de la incidencia de cáncer de hígado después de la vacunación frente a la hepatitis B
Intervención terapéutica	Tratamiento	Supervivencia similar para mujeres con cáncer de ovario epitelial unilateral a las que se les practicó cirugía conservadora para mantener su fertilidad

Fuente: Modificado de Lazcano-Ponce E, Fernández E, Salazar-Martínez E, Hernández-Avila M. Estudios de cohorte. Metodología, sesgos y aplicación. Salud Publica Mex. 2 de mayo de 2000;42(3):230-41.

Tabla 2. Ventajas y desventajas de los estudios de cohortes

Ventajas	Desventajas
Único método para determinar la incidencia. Diseño idóneo para estudiar la historia natural y el curso clínico y determinar el riesgo	Alto coste económico y de tiempo (cohorte concurrente)
Existe una clara secuencia temporal de exposición y evento de interés (por ende, permiten estudiar la causalidad)	Requieren un periodo de seguimiento Las pérdidas de seguimiento pueden alterar o invalidar los resultados
Permiten estudiar exposiciones poco frecuentes, enfermedades con resultados letales o eventos de interés poco frecuentes	Se requiere un gran número de sujetos para estudiar eventos de interés poco frecuentes
Se pueden evaluar de forma simultánea resultados múltiples, potencialmente asociados a una exposición	Los resultados pueden verse afectados por factores confundentes
Se puede determinar la incidencia del evento de interés para los expuestos y los no expuestos	Susceptibles de sesgos (de selección, de mal clasificación y de seguimiento)

Fuente: Modificado de Salazar P, Manterola C, Quiroz G, et al. Estudios de cohortes. 1ª parte. Descripción, metodología y aplicaciones. Rev Cir. 2019 Oct;71(5):482-93.

una serie de preguntas. Es el que se expone a continuación, aplicado a la publicación presentada al comienzo del capítulo.

¿Son válidos los resultados del estudio?

Las preguntas 1 y 2 mostradas a continuación son eliminatorias. Solo si la respuesta es afirmativa a ambas merece la pena continuar con las restantes.

1. ¿El estudio se centra en un tema claramente definido?
 - Se han definido claramente el factor de riesgo que se va a estudiar y otros que pueden influir en el resultado.
 - El resultado que se va a estudiar está bien definido (demencia tipo Alzheimer) y se ha establecido claramente cómo medirlo (test de evaluación cognitiva).
 - La población de origen de la que se han derivado las dos cohortes (población general seleccionada aleatoriamente por código postal) y los criterios de inclusión y exclusión están bien definidos.
2. ¿Se reclutó la cohorte de la manera más adecuada? Se trata de buscar sesgos de selección que comprometan los resultados (**tabla 3**).
 - ¿La cohorte es representativa de una población definida? El número de participantes es lo suficientemente elevado para representar a las personas de mediana edad.
 - ¿Hay algo "especial" en la cohorte? Las cohortes son similares en todos sus aspectos excepto en la condición de fumadores. Ningún individuo presentó el evento de interés al inicio del estudio y han salido los que lo han presentado durante el estudio.
 - ¿Se incluyó a todos los que debería haberse incluido en la cohorte? No se ha excluido a participantes por ningún factor de riesgo. No hay personas que debieron ser elegidas y no lo fueron.
 - ¿La exposición se midió de forma precisa con el fin de minimizar posibles sesgos? Se valoró el tabaquismo y el resto de factores de riesgo de confusión de forma precisa a lo largo del tiempo. El distinto grado de exposición está claramente diferenciado y definido.
3. ¿El resultado se midió de forma precisa con el fin de minimizar posibles sesgos? Se trata de buscar sesgos de medida o clasificación.
 - ¿Se utilizaron variables objetivas o subjetivas? La variable principal de interés fue objetiva. Las covariables, comoquiera que no se conoce un factor de riesgo claro para el desarrollo de demencia tipo Alzheimer, fueron subjetivas en la consideración de potenciales relaciones con el evento de interés.
 - ¿Las medidas reflejan de forma adecuada aquello que se supone que tienen que medir? Los test utilizados para medir el resultado están validados. La medida de la exposición depende de la sinceridad en las respuestas del participante, pero no existe otro método factible.

Tabla 3. Tipos de sesgos que pueden ocurrir en un estudio de cohortes

Etapas	Tipos de sesgo	
Antes de comenzar el estudio	Sesgo de selección	Durante la fase de reclutamiento. Al seleccionar pacientes deben utilizarse criterios rigurosos para evitar resultados confusos. Los pacientes deben proceder de la misma población general. Se asignan pacientes para estudiar cohortes utilizando criterios rigurosos.
Durante la realización del estudio	Sesgo del entrevistador o reclutador	Se debe estandarizar la interacción del entrevistador con el paciente. Un ejemplo es el encuestador ciego al estado de exposición.
	Sesgo en la cronología	Los estudios prospectivos pueden eliminar el sesgo cronológico. Evite utilizar controles históricos (confusión por tendencias seculares).
	Sesgo de información	Utilice fuentes de datos objetivas siempre que sea posible. Al utilizar fuentes de datos subjetivos, corrobore con la historia clínica.
	Sesgo de seguimiento	Se recomienda diseñar cuidadosamente el plan para los pacientes con pérdida de seguimiento antes del estudio. Definir claramente la exposición antes del estudio. Considere la estratificación de los grupos.
Después de la recopilación de datos	Sesgo de confusión	Los factores de confusión conocidos pueden controlarse con el diseño del estudio (diseño de casos y controles o aleatorización) o durante el análisis de datos (regresión).

Fuente: Modificado de Salazar P, Manterola C, Quiroz G, et al. Estudios de cohortes. 1ª parte. Descripción, metodología y aplicaciones. Rev Cir. 2019 Oct;71(5):482-93.

La medida de la exposición se ha expresado en términos de personas-año. Se valora el factor de riesgo en todo su espectro de exposición.

- ¿Se ha establecido un sistema fiable para detectar todos los casos? Se entrevista a todos los individuos periódicamente y se realizan las mediciones del mismo modo y de manera ciega en relación con el factor de riesgo.
 - ¿Se clasificaron todos los sujetos en el grupo de exposición utilizando el mismo método? Sí.
 - ¿Los métodos de medida fueron similares en los diferentes grupos? Sí.
 - ¿Eran los sujetos o el evaluador de los resultados ciegos a la exposición? Sí.
4. ¿Se ha tenido en cuenta el potencial efecto de los factores de confusión en el diseño o el análisis del estudio?
- Se ha realizado un estudio de regresión de Poisson para obviar los factores de confusión en los resultados con las variables estudiadas (**tablas 4 y 5**).

Tabla 4. Estrategias para el análisis de los estudios de cohortes

	Análisis de supervivencia	Análisis tiempo-persona
Tamaño de la muestra	Relativamente pequeño	Relativamente grande
Cantidad de eventos	Frecuentes	Poco frecuentes
Escala temporal	Única	Única o múltiple
Medidas de incidencia	Probabilidad	Tasa
Análisis unifactorial	Comparación de curvas de supervivencia Prueba de <i>log-rank</i> Razón de riesgos	Comparación de tasas Razón de tasas
Análisis multifactorial	Regresión de Cox	Regresión de Poisson

Fuente: Modificado de Salazar P, Manterola C, Quiroz G, et al. Estudios de cohortes. 1ª parte. Descripción, metodología y aplicaciones. Rev Cir. 2019 Oct;71(5):482-93.

5. ¿Fue lo suficientemente largo y completo el seguimiento de los sujetos?

- Diez años es, quizás, un periodo corto para la aparición de demencia tipo Alzheimer. No obstante, se produjo un número suficiente de eventos para la validez del estudio.
- Las pérdidas son del 14 % (cifra óptima, inferior al 5%; máxima aceptable, 20 %). Cuando el número de perdidos en el seguimiento es alto se suelen efectuar análisis de sensibilidad con el "peor de los casos" para evaluar la consistencia de los resultados principales.
- Las causas de las pérdidas están justificadas (se ha presentado un diagrama de flujo en el desarrollo del artículo). Es interesante presentar un diagrama de flujo del estudio: es útil conocer el número de participantes que se incluyeron en el estudio, los que se mantuvieron hasta el final o hasta el desarrollo del desenlace y los que abandonaron o se perdieron en el seguimiento, con sus motivos.
- La medición de la exposición se ha hecho en términos de personas-año. Cuando el estudio presenta pérdidas en el seguimiento (abandono, cambio de domicilio, muerte por causa diferente al evento del estudio, finalización del estudio por llegada al tiempo establecido por el protocolo o por agotamiento de recursos financieros), los tiempos de seguimiento son desiguales. Una forma de tratar periodos de seguimiento variables es con el análisis basado en tiempo-persona. En estos casos, se puede utilizar el promedio de tiempo contribuido por la totalidad de sujetos de la cohorte; es decir, se agregan seis meses por el año de entrada, seis meses más por el año de salida y doce meses por todos los años de entrada y salida.

Tabla 5. Resumen de los principales métodos para el análisis de datos procedentes de estudios de cohortes

Variable de interés	Exposición	Medida resumen	Comparación	
			Medida de asociación	Modelo de regresión
Episodios en personas-año		Tasa de incidencia	Razón de tasas	Regresión de Poisson
Tiempo al episodio	Fija	Método de Kaplan-Meier	Riesgo relativo	Regresión de Cox
		Estimador de máxima verosimilitud	Tiempo relativo	Modelo paramétrico
	Cambiante	Extensión del método de Kaplan-Meier	Riesgo relativo	Regresión de Cox, incorporando entradas retardadas
Tiempo al episodio en presencia de riesgos competitivos		Curva de incidencia acumulada	Razón de riesgos acumulados	Regresión de Cox para la subdistribución del riesgo
Marcador medido repetidamente en el tiempo		Tasa de cambio en el tiempo	Diferencias en la tasa de cambio en el tiempo	Modelo de efectos aleatorios Modelo conjunto de la trayectoria del marcador y del tiempo al episodio de interés

Fuente: Jarrín I, Geskus R, Pérez-Hoyos S, Del Amo J. Principales métodos de análisis en los estudios de cohortes de sujetos diagnosticados de infección por el virus de la inmunodeficiencia humana (VIH). *Enferm Infecc Microbiol Clin.* 2010;28(5):298-303.

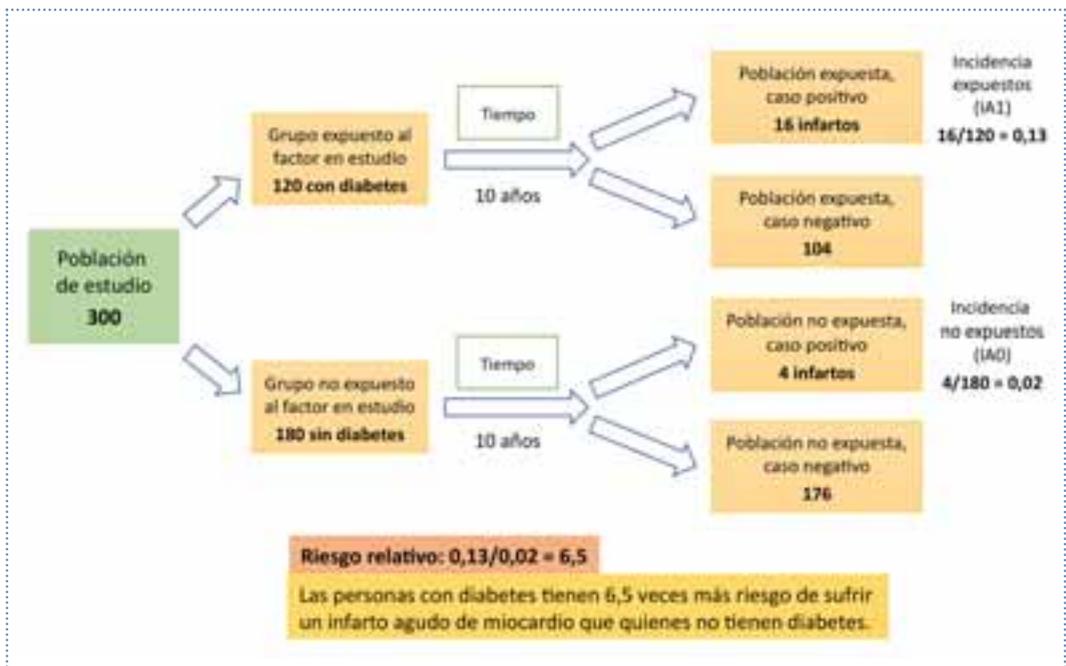
¿Cuáles son los resultados?

6. ¿Cómo son los resultados?

- ¿Cuáles son los resultados netos? El tabaquismo aumenta significativamente el riesgo de padecer demencia tipo Alzheimer. Una de las fortalezas de los estudios de cohortes es la posibilidad de calcular medidas de asociación, que son indicadores epidemiológicos que evalúan la fuerza con la que una determinada enfermedad o evento de interés se asocia con un determinado factor de exposición que se presume como su causa. Las medidas de asociación más sólidas se calculan utilizando la incidencia, ya que esta permite establecer que el efecto del evento de interés es posterior a la exposición.

- ¿Se especifica la tasa o proporción entre expuestos y no expuestos? Sí, se especifican los riesgos relativos (se supone que en el desarrollo del artículo). Las medidas de asociación también se pueden considerar como comparaciones de incidencias: incidencia del evento de interés en expuestos al factor en estudio frente a la incidencia de dicho evento en sujetos no expuestos. Las medidas de asociación cuantifican la relación entre el factor de riesgo y el evento de interés, es decir, la magnitud de la diferencia observada. Para ello, se ha de construir una tabla de contingencia, en la que se detallen los datos del desarrollo o manifestación del evento de interés en las cohortes de expuestos y no expuestos. Un sencillo ejemplo puede verse en la **figura 2**.
- ¿Cómo de fuerte es la relación de asociación entre el factor de riesgo y el evento de interés? Se ha estimado la razón de riesgos proporcionales (*hazard ratio*) entre los grupos. La magnitud cuantitativa de fuerza de asociación entre dos variables se puede medir con los siguientes instrumentos:
 - Medidas de efecto: Se basan en el cálculo de un cociente, por lo que permiten cuantificar discrepancias en la ocurrencia de un evento de interés en grupos que difieren en la existencia de cierta variable. Estas medidas son la razón de riesgos proporcionales (*hazard ratio*), la razón de riesgos, la razón de tasas, la diferencia de riesgos y la diferencia de tasas.

Figura 2. Riesgo relativo de sufrir un infarto de miocardio en las personas que tienen diabetes en comparación con las que no padecen esa enfermedad



Como los valores de probabilidad (valores de p) transmiten menos información que las medidas de efecto y son muy sensibles al tamaño de la muestra, lo idóneo es no describirlas cuando se comparan unos grupos con otros, o al menos no deben constituir la base de la comparación.

- Medidas de impacto: Se basan en las diferencias. En general, indican la contribución de un determinado factor a la producción del evento de interés entre los que están expuestos a él. Por este motivo, se dice que las medidas de impacto indican el riesgo de enfermar que podría evitarse si se eliminara la exposición: reducción absoluta de riesgo o reducción atribuible del riesgo o riesgo atribuible; reducción relativa del riesgo; número que es necesario tratar; número necesario para dañar; índice del número que es necesario tratar/número necesario para dañar.

En un estudio de cohortes, la comparación entre los resultados de los grupos de expuestos y no expuestos al factor de riesgo debe expresarse mediante medidas de efecto (**figura 3**).

7. ¿Cuál es la precisión de los resultados? Se ha especificado con el intervalo de confianza. Para expresar la incertidumbre del resultado, se emplea el intervalo de confianza al 95%, que representa el rango de valores en que se encontrarían los resultados si se efectuaran 100 estudios iguales pero en muestras diferentes. Cuanto más estrecho sea el intervalo de confianza, mayor será la precisión de la estimación. Las medidas de efecto deben acompañarse siempre de sus intervalos de confianza correspondientes.

¿Son los resultados aplicables a tu medio?

8. ¿Parecen creíbles los resultados?

- ¿Es potente el efecto del factor de riesgo sobre el evento de interés? Sí, existe un *hazard ratio* considerable.
- ¿Pueden deberse al azar, sesgos o confusión? No, se ha realizado un estudio multivariante para eliminarlos.
- ¿Son creíbles el diseño y los métodos? Sí.

9. ¿Los resultados de este estudio coinciden con otra evidencia disponible? Debería validarse en un grupo independiente de pacientes con la misma enfermedad objeto de estudio para confirmar el poder predictivo y la reproducibilidad de los resultados.

10. ¿Se pueden aplicar los resultados en nuestro medio? Los sujetos y el medio en que se desarrolla son similares a los nuestros.

11. ¿Va a cambiar el resultado tu decisión clínica?

Figura 3 (Continuación). Cálculo de la razón de riesgo en un estudio de cohortes (arriba) y aplicación al riesgo de desarrollar morbilidad posoperatoria en sujetos con obesidad o sobrepeso sometidos a cirugía gastrointestinal electiva (abajo)

		Exposición	
		Sí	No
Evento de interés	Sí	a	b
	No	c	d
Total		m_1	m_0

Inciencia de expuestos o riesgo absoluto de expuestos (RA-E) (m_1)	$a/(a + c)$
Inciencia de no expuestos o riesgo absoluto de no expuestos (RA-NE) (m_0)	$b/(b + d)$
Riesgo relativo (RR)	m_1/m_0
Reducción absoluta del riesgo (RAR) o diferencia de incidencia acumulada	$m_1 - m_0$
Reducción relativa del riesgo (RRR)	RAR/m_0
Número que es necesario tratar (NNT)	$1/RAR$
Número necesario para dañar (NND)	$1/AAR^*$

* Aumento absoluto del riesgo

		Índice de masa corporal		Total
		Sí	No	
Morbilidad posoperatoria	Sí	56	32	88
	No	12	28	40
Total		68	60	128

Inciencia de expuestos o riesgo absoluto de expuestos (RA-E) (m_1)	$a/(a + c) = 0,64$
Inciencia de no expuestos o riesgo absoluto de no expuestos (RA-NE) (m_0)	$b/(b + d) = 0,30$
Riesgo relativo (RR)	$m_1/m_0 = 2,13$
Reducción absoluta del riesgo (RAR) o diferencia de incidencia acumulada	$m_1 - m_0 = 0,34$
Reducción relativa del riesgo (RRR)	$RAR/m_0 = 1,13$
Número que es necesario tratar (NNT)	$1/RAR = 2,9$
Número necesario para dañar (NND)	$1/AAR = 3,0$

Fuente: Modificado de Salazar P, Manterola C, Quiroz G, et al. Estudios de cohortes. 1ª parte. Descripción, metodología y aplicaciones. Rev Cir. 2019 Oct;71(5):482-93.

Bibliografía

Cabello JB; por CASPe. Plantilla para ayudarte a entender estudios de cohortes. En: CASPe. Guías CASPe de lectura crítica de la literatura médica. Alicante: CASPe; 2005. Cuaderno II. p. 23-27.

Lazcano-Ponce E, Fernández E, Salazar-Martínez E, Hernández-Avila M. Estudios de cohorte. Metodología, sesgos y aplicación. Salud Publica Mex. 2 de mayo de 2000;42(3):230-41.

Salazar P, Manterola C, Quiroz G, et al. Estudios de cohortes. 1ª parte. Descripción, metodología y aplicaciones. Rev Cir. 2019 Oct;71(5):482-93.



Preguntas de autoevaluación

1. ¿Cuál de las siguientes afirmaciones es correcta si comparamos los estudios de casos y controles con los estudios de cohortes?

- a) La exposición o factor de riesgo está presente en los sujetos al inicio del estudio en los estudios de casos y controles, pero no en los estudios de cohortes.
- b) El evento de interés o resultado del estudio está presente en los sujetos al inicio del estudio en los estudios de casos y controles, pero no en los estudios de cohortes.
- c) Solamente en los estudios de casos y controles se puede valorar la incidencia y la causalidad.
- d) Es más difícil encontrar un grupo de control en los estudios de cohortes que en los de casos y controles.

La respuesta correcta es la opción *b*.

Ambos diseños de estudio son observacionales, longitudinales y analíticos.

En un estudio de cohortes se recluta a un grupo de individuos que no presentan en ese momento el evento de interés. Los sujetos se clasifican de acuerdo con las características (factores de riesgo) que podrían guardar relación con el resultado. Todos los integrantes deben estar en riesgo de padecer o presentar dicho evento. Posteriormente, estos individuos son observados a lo largo del tiempo para cuantificar cuáles de ellos manifiestan el resultado.

En los estudios de casos y controles se parte de un grupo de individuos enfermos (casos) y se busca un grupo de individuos sanos (controles) para comparar la posible presencia de determinados factores de riesgo (véase la figura 1 en el texto del capítulo).

2. Los estudios de cohortes tienen ventajas e inconvenientes. ¿Cuál de las siguientes es una de las ventajas de estos estudios?

- a) Están exentos de sesgos si se selecciona adecuadamente la población que se va a estudiar.
- b) Son los únicos en los que se puede determinar la incidencia de una enfermedad o evento de interés.
- c) Son fáciles de llevar a cabo.
- d) No se ven afectados por otros factores de riesgo que puedan producir confusión.

La respuesta correcta es la opción *b*.

En los estudios de cohortes los sujetos no presentan el evento de interés al inicio del estudio, a diferencia del resto de estudios. Por ello, es el único en el que se puede calcular la incidencia de un evento.

El sesgo de selección puede producirse, también, si se producen muchas pérdidas durante el seguimiento. Existen muchos más sesgos que se pueden producir además de los de selección, como el de confusión por la presencia de otros factores de riesgo que puedan influir en el resultado (véase la tabla 3 en el texto del capítulo).

Son estudios largos y costosos.

3. ¿Cómo se ha de especificar la precisión de los resultados en un estudio de cohortes?

- a) Con el valor de probabilidad de p .
- b) Con el intervalo de confianza.
- c) Con la razón de tasas.
- d) Con la razón de riesgos.

La respuesta correcta es la opción *b*.

Para expresar la incertidumbre del resultado, se emplea el intervalo de confianza al 95%, que representa el rango de valores en que se encontrarían los resultados si se efectuaran 100 estudios iguales pero en muestras diferentes. Cuanto más estrecho sea el intervalo de confianza, mayor será la precisión de la estimación. Las medidas de efecto deben acompañarse siempre de sus intervalos de confianza correspondientes.

La probabilidad de p , la razón de tasas y la razón de riesgos son medidas de efecto de asociación. Como los valores de probabilidad (valores de p) transmiten menos información que las medidas de efecto y son muy sensibles al tamaño de la muestra, lo más adecuado es que no se describan cuando se comparan unos grupos con otros, o al menos no deben constituir la base de la comparación entre los grupos de una cohorte.

Los estudios ecológicos para confirmar y establecer la evidencia de una hipótesis

Dr. Francisco Pérez Jiménez

Catedrático emérito de Medicina
Universidad de Córdoba (Córdoba)

Instituto Maimónides de Investigación Biomédica de Córdoba (IMIBIC)

*La publicación seleccionada de este capítulo
es un trabajo basado en datos reales no publicado ni revisado por pares,
creado por el autor como base para el desarrollo del tema del capítulo.*



capítulo 7

Publicación seleccionada

Correlación negativa entre el PIB per cápita de las provincias españolas y la prevalencia de mortalidad por infarto agudo de miocardio. Estudio ecológico

Introducción

La enfermedad cardiovascular es un proceso de etiopatogenia compleja que con frecuencia se aborda con enfoques reduccionistas. De hecho, el médico clínico la asocia sistemáticamente con la presencia de los factores de riesgo y, en consonancia, aborda su tratamiento con cambios en el estilo de vida y el empleo de fármacos. Pero ese enfoque obvia otros factores muy importantes, que quedan marginados en la comprensión de la auténtica realidad de dicho proceso. La publicación seminal de Walter Willet et al. (véase la bibliografía) sobre la sindemia global supone un avance clave al abrir una perspectiva totalmente nueva en el enfoque de las enfermedades crónicas no transmisibles, incluidas la enfermedad cardiovascular, la obesidad o la diabetes. Según él, la creciente prevalencia de obesidad mundial está vinculada al calentamiento del planeta y a una inadecuada alimentación, hechos ambos que repercuten en los grupos más pobres y, por tanto, más vulnerables con un deterioro de su modelo de dieta de consecuencias negativas para su salud y un aumento del riesgo de padecer obesidad, enfermedad cardiovascular y el resto de las enfermedades crónicas de las sociedades modernas. El presente trabajo es un estudio ecológico en el que se investiga si las diferencias en PIB per cápita entre 16 provincias españolas se correlaciona con la prevalencia de mortalidad cardiovascular en dichas poblaciones, para establecer si el fenómeno observado en países pobres también es patente en un país de elevado bienestar social, como España.

Pregunta clínica

¿Es cierto que la pobreza económica condiciona un mayor riesgo cardiovascular en las regiones más pobres, incluso en un país en el que el desarrollo industrial lo hace ser considerado como un entorno de alta protección social?

Hipótesis

Nuestra hipótesis es que el nivel económico de una población favorece la mortalidad por enfermedad cardiovascular, a pesar de que dicha población esté integrada en un país considerado de alto nivel económico y con un escudo social desarrollado. La hipótesis nula sería que los factores económicos no influyen en la mortalidad por procesos cardiovasculares.

Objetivo principal

Estudiaremos en 16 provincias de España, de tamaño y de localización geográfica diferente, la relación entre su producto interior bruto (PIB) per cápita y la tasa anual de defunciones (casos/100 000 habitantes) por infarto agudo de miocardio (IAM) en la población total, en hombres y en mujeres, en el año 2018.

Objetivos secundarios

1. Analizaremos, en la misma población, la posible existencia de diferencias en la mortalidad por enfermedades circulatorias según el PIB per cápita.
2. Analizaremos, en la misma población, la posible existencia de diferencias en la mortalidad por enfermedades cerebrovasculares (accidentes cerebrovasculares) según el PIB per cápita.

Selección de la población

Como población se han elegido 16 provincias de España de distintas autonomías, tanto peninsulares como de las islas, que cubren norte, sur, este y oeste: Girona, Córdoba, Cádiz, Cantabria, Cáceres, Navarra, Las Palmas, Valladolid, Zaragoza, Baleares, Coruña, Cuenca, Rioja, Vizcaya, Asturias y Madrid. Totalizan una población de 18 554 767 personas según los datos del año 2018, de las cuales 9 022 672 fueron hombres, y 9 532 095 mujeres. Los datos individualizados de cada provincia no se muestran.

Elección de las variables y recogida de datos

La variable independiente, como criterio del estatus económico de las poblaciones, es el PIB per cápita del año 2018. El PIB per cápita se define como la media por habitante de los rendimientos totales íntegros (antes de impuestos) producidos en un determinado territorio. Este será el factor de exposición cuya posible relación con las variables dependientes, los indicadores de salud, queremos conocer. Estos indicadores son las defunciones por IAM, enfermedades circulatorias y enfermedades cerebrovasculares del año 2018 según el informe anual del Instituto Nacional de Estadística. A partir de dichos datos se ha calculado la tasa de defunciones por 100 000 habitantes.

Análisis estadístico

Inicialmente contrastamos la normalidad de los datos con el test de Shapiro-Wilks. Para los que no cumplieron el supuesto de normalidad se empleó el coeficiente de correlación de Spearman (ρ , test no paramétrico), correlacionando cada variable con el valor del PIB per cápita, mientras que para las que cumplieron la normalidad se empleó el coeficiente de correlación de Pearson (test paramétrico). Se consideraron significativas las correlaciones con $p < 0,05$. El análisis estadístico se hizo con el paquete SPS, versión Statistic 20.

Resultados

Las defunciones totales, por los tres procesos, fueron 73 160, de las cuales 54 160 fueron por enfermedades circulatorias, 6741 por IAM y 11 809 por enfermedades cerebrovasculares. En la **tabla 2** se incluyen la tasa provincial de defunciones (eventos por 100 000 habitantes) por dichos procesos en hombres y mujeres, así como el PIB per cápita de cada provincia. Los datos relativos al IAM no cumplieron el supuesto de nor-

Tabla 2. Tasa de defunciones totales por 100 000 habitantes a causa de enfermedad en hombres y mujeres por provincias, y producto

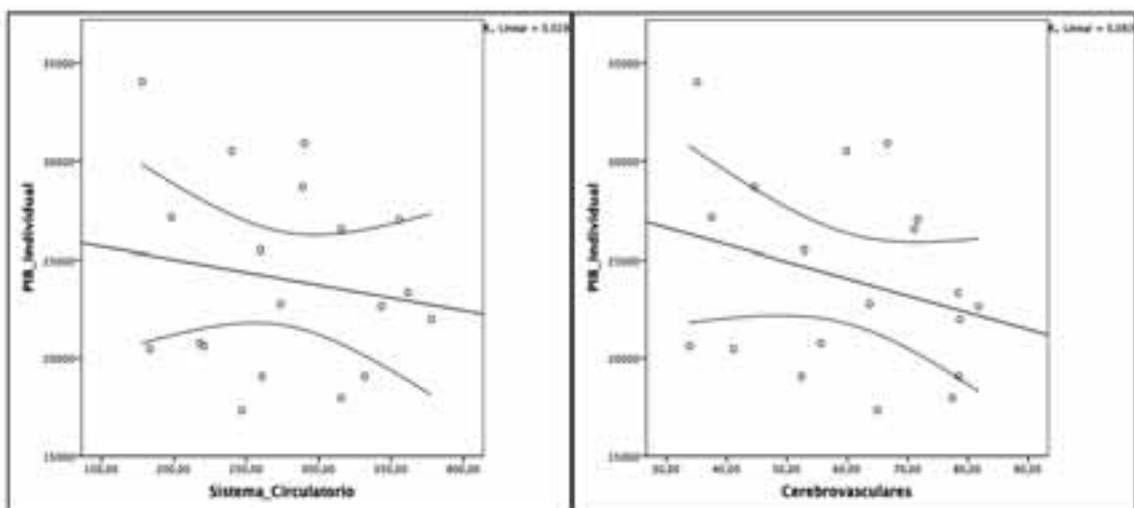
Provincia	Enfermedad cardiovascular			Infarto agudo	
	Total	Hombres	Mujeres	Total	Hombres
Córdoba	315,44	278,63	350,86	36,16	35,48
Girona	288,88	219,91	237,89	26,51	35,40
Cádiz	247,07	231,61	262,70	33,06	39,67
Cantabria	273,66	265,72	282,15	23,40	31,44
Cáceres	331,89	306,09	357,57	50,12	59,89
Navarra	240,12	227,35	233,23	30,94	36,79
Las Palmas	220,87	227,80	213,98	39,07	46,88
Valladolid	259,80	240,15	278,57	29,42	41,77
Zaragoza	355,63	268,51	331,49	34,47	43,82
Baleares	198,19	183,53	212,86	20,06	26,58
Coruña	361,81	340,33	382,13	36,30	46,46
Cuenca	343,50	308,90	378,78	30,50	38,61
Rioja	315,70	304,54	326,58	41,02	41,55
Vizcaya	290,12	284,77	295,07	27,42	31,74
Asturias	377,99	350,50	48,69	56,18	71,48
Madrid	178,11	164,00	192,75	25,30	29,47

malidad, por lo que empleamos el coeficiente de correlación de Spearman, mientras que para el resto se empleó el coeficiente de correlación de Pearson. Observamos que, aunque en el análisis de las tres variables dependientes existió una correlación inversa, entre el PIB per cápita y el riesgo de sufrir uno de los eventos cardiovasculares solo fue significativa en el caso del IAM (**figuras 1 y 2**).

Tabla 1. Mortalidad cardiovascular, infarto agudo de miocardio y enfermedades cerebrovasculares y producto interior bruto (PIB) per cápita de cada provincia

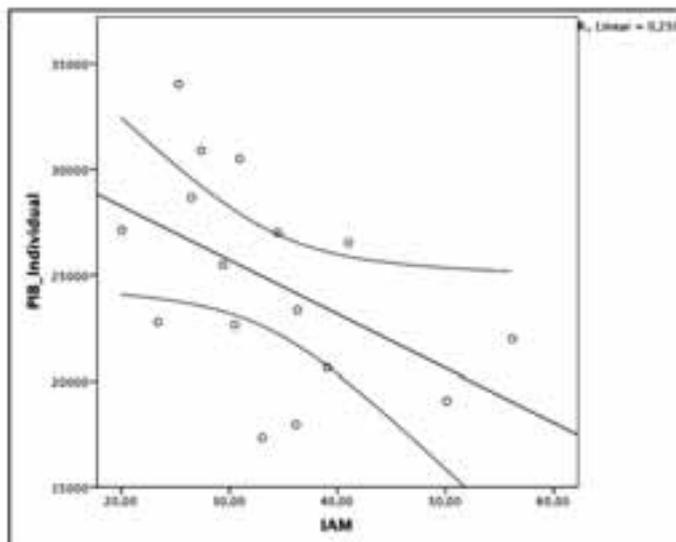
Provincia	de miocardio	Accidentes cerebrovasculares			PIB per cápita
	Mujeres	Total	Hombres	Mujeres	
Almería	77,42	77,42	68,03	86,46	17 964
Asturias	17,57	44,62	45,53	47,90	28 674
Baleares	25,59	65,01	56,77	73,24	17 343
Barcelona	15,82	63,68	53,58	71,71	22 767
Bizkaia	40,40	78,48	68,52	88,38	19 077
Burgos	23,29	59,87	50,62	63,63	30 508
Cantabria	31,32	33,83	34,58	33,09	20 614
Castellón	22,18	52,88	50,66	60,52	25 482
Cataluña	27,19	71,66	60,63	82,41	27 011
Extremadura	13,55	37,47	34,13	40,82	27 134
Galicia	26,97	78,41	66,17	89,86	23 343
Granada	22,22	81,79	65,34	98,98	22 639
Guipúzcoa	40,50	71,15	69,48	72,78	26 528
Huelva	23,42	66,66	59,44	73,34	30 901
Jaén	42,16	78,67	62,93	93,09	21 981
León	21,10	35,13	28,64	41,12	34 041

Figura 1. Correlación entre el producto interior bruto (PIB) per cápita y las defunciones por enfermedades del sistema circulatorio y cerebrovasculares en la población total, en hombres y en mujeres (el grado de significación fue en todos los casos >0,05)



	Sistema circulatorio	Sistema circulatorio (hombres)	Sistema circulatorio (mujeres)	Cerebrovasculares	Cerebrovasculares (hombres)	Cerebrovasculares (mujeres)
Pearson Correlation	-,166	-,327	-,118	-,289	-,320	-,275
Sig. (2-tailed)	,497	,172	,631	,231	,182	,254
N	19	19	19	19	19	19

Figura 2. Correlación entre la tasa de defunciones por infarto agudo de miocardio y el producto interior bruto (PIB) per cápita en la población total, en hombres y en mujeres (la significación fue en todos los casos $<0,05$)



	IAM_total	IAM_hombres	IAM_mujeres
Coefficiente de correlación	-,568*	-,538*	-,541*
Sig. (bilateral)	,022	,031	,030
N	16	16	16

Planteamiento, pregunta de investigación y objetivos

Nuestro proyecto es una investigación original, hecha para desarrollar este capítulo. Aunque hay información sobre el tema, es insuficiente y hemos pensado que sería más interesante basarse en una investigación diseñada *ad hoc*, adaptada a los objetivos de esta exposición.

Objetivos

La formulación de los objetivos es clave en cualquier investigación y puede no ser una tarea sencilla, porque hay que conjugar que sirvan para confirmar la hipótesis y que además sean realistas, para poder abordarlos según los recursos experimentales disponibles. Es importante saber que no todos los resultados obtenidos en la investigación se admitirán cómo científicamente probados, sino solo aquellos que se obtengan al desarrollar el objetivo principal, en torno al cual se elaborará el diseño del estudio. Por ello el objetivo principal debe ser claro y explícito, indicando lo que se desea hacer con concreción, ya que de su formulación dependerá la selección de la población, el tamaño de la muestra, el diseño metodológico y la estrategia de análisis de resultados. Es un error elaborar múltiples objetivos, ambiciosos y mal jerarquizados, que generan una gran confusión a la hora del diseño de la investigación. Por eso, aunque puedan plantearse objetivos secundarios, deben hacerse sabiendo que no tendrán suficiente evidencia para considerarlos científicamente probados. Son meros resultados de segundo nivel y, aunque pueden ser interesantes, su utilidad se limita a poder formular nuevas hipótesis.

Diseño del estudio, población y métodos

Cuando se piensa en un proyecto de investigación hay que decidir qué modalidad seleccionaremos, según la hipótesis que deseamos probar y los objetivos que deseamos alcanzar. En nuestro caso partimos de una hipótesis original, sobre un tema del que hemos dicho que tenemos insuficiente conocimiento previo. Esta incertidumbre desaconseja seleccionar un proyecto que sea muy costoso en recursos materiales, humanos y en tiempo, lo que nos orienta a optar por un estudio observacional. De este tipo de estudios, los más potentes son los estudios de cohortes, para lo que estudiaríamos, a lo largo del tiempo, una o varias poblaciones, investigando si los eventos que eligiéramos como objetivo final tienen una incidencia diferente según el nivel económico de los participantes. Pero realizar este tipo de estudio sería costoso en recursos materiales, humanos y de tiempo, sobre todo con una hipótesis llena de incertidumbre. Por ello optaremos por un estudio ecológico, que son menos costosos, aunque aportan menor nivel de evidencia. Concluido el estudio, si la hipótesis se confirma, será el momento de decidir si se puede avanzar en el conocimiento de esta cuestión con estudios de mayor evidencia.

Los estudios ecológicos

Los estudios ecológicos son los estudios epidemiológicos más sencillos y frecuentemente empleados para la descripción de la situación de salud o investigar la consecuencia de nuevas exposiciones en poblaciones humanas. De hecho, son los que suelen iniciar el proceso de investigación epidemiológica ante una hipótesis nueva, con la ventaja de que son de bajo

Los estudios ecológicos para confirmar y establecer la evidencia de una hipótesis

coste porque se utilizarán datos disponibles, recogidos con otras finalidades, bien a partir de la recogida de datos rutinarios o por razones administrativas y legales. Su peculiaridad fundamental es que no estudian a individuos por separado, sino que las unidades de análisis son poblaciones o grupos de personas, y se obtienen conclusiones por el posible efecto sobre ellas de riesgos individuales presentes en la población. Estos hechos hacen que, si se demuestra la asociación de dos o más variables a nivel del grupo, no se puede extrapolar al nivel individual, lo que se denomina falacia o sesgo ecológico. Tienen además limitaciones basadas en la heterogeneidad de la recogida de datos, al haber sido recogidos para otros propósitos, con métodos variables, lo que implica la existencia de posibles sesgos y baja calidad de la información. En síntesis, los datos no se han generado y recogido con criterios preestablecidos para nuestro estudio. Aun así, son idóneos para generar nuevos conocimientos y nuevas hipótesis, que se pueden refrendar posteriormente mediante estudios analíticos. Sus ventajas y limitaciones más importantes se esquematizan en la **tabla 1**.

Tabla 1. Ventajas y limitaciones de los estudios ecológicos

Ventajas	Sencillez
	Bajo coste
	Utiliza datos ya existentes
	Se pueden utilizar poblaciones diferentes
Limitaciones	Bajo nivel de evidencia
	La calidad de los datos puede ser limitada
	Más posibilidades de sesgos
	Falacia ecológica

Fuente: Adaptado de Morgenstern H. Uses of ecologic analysis in epidemiologic research. Am J Public Health. 1982;72(12):1336-44.

Existen cuatro modalidades de estudios ecológicos, a saber:

- Estudios exploratorios.** Se comparan las tasas de enfermedad entre distintas regiones durante un mismo periodo, o la frecuencia de la enfermedad a través del tiempo en una misma región, pero solo se estudiaría la tasa de enfermedad. El único propósito es buscar patrones espaciales o temporales que podrían sugerir hipótesis sobre las causas.
- Estudios múltiples.** Evalúan la asociación entre los niveles de exposición promedio de un agente y la frecuencia de la enfermedad entre grupos, habitualmente geopolíticos. La fuente de datos suelen ser las estadísticas de morbilidad y mortalidad rutinarias.

- c) **Estudios de series de tiempo.** Se comparan las variaciones temporales de los niveles de exposición del agente (por ejemplo, consumo de productos, conductas o concentraciones de contaminantes) con otra serie diferente que refleja los cambios en la frecuencia de la enfermedad en una población de un área geográfica durante ese periodo de tiempo.
- d) **Estudios mixtos.** En esta categoría se incluyen los estudios de series de tiempo combinadas con la evaluación de grupos múltiples.

En nuestro caso elegiremos la modalidad más utilizada, fácil de realizar y adecuada para nuestro propósito: la de estudios múltiples. Para su diseño se siguieron las etapas expuestas en el artículo.

Una característica de los estudios ecológicos es que, al ser poblacionales por definición, no requieren el cálculo del tamaño muestral, puesto que la unidad de análisis no es el individuo, sino que se utilizan grupos específicos, como grupos poblacionales, comunidades, regiones, estados, países, etc. Las medidas de exposición y resultados a nivel grupal son generalmente variables continuas, y para cuantificar su relación se puede calcular el coeficiente de correlación, que determina en qué medida la relación entre la exposición y el resultado es lineal.

Discusión

Comentarios

La discusión debe ser ordenada y sistemática, evitando reiteraciones y comentarios prolijos. Se inicia con la exposición de los resultados sobre el objetivo principal, haciéndolo de forma clara y concisa; a continuación, se comentarán dichos resultados, profundizando en cuál es su aportación sobre la información previamente conocida. Posteriormente se expondrán los resultados secundarios, las limitaciones del estudio y se propondrán, si las hay, nuevas hipótesis y posibles estudios de futuro.

Discusión

Nuestros datos demuestran que existe una correlación negativa entre el nivel económico de una población, valorado por el PIB per cápita, y la tasa de mortalidad por IAM. Este hallazgo lo observamos tanto en hombres como en mujeres a partir de la información correspondiente al año 2018 de 16 provincias de las distintas autonomías españolas que suman una población de más de 18 millones personas. Nuestra aportación tiene gran interés, desde la perspectiva de la salud pública, porque implica que el abordaje individual, en la prevención del IAM, será insuficiente si no se toman medidas orientadas a corregir las diferencias de nivel de vida en las poblaciones con mayor mortalidad. Es bien conocido que los individuos con mejor estatus socioeconómico tienen una salud global superior a la de los de un estatus peor, y que depende de múltiples factores sociales, como la situación laboral y el nivel de ingresos, que influyen en la educación, el acceso a cuidados de salud, los hábitos higiénicos o la capacidad para gestionar aspectos relacionados con el riesgo de enfermar. Aunque ya era conocido que estos factores son importantes en la incidencia de enfermedades transmisibles, nuestros datos sugieren que también son extensibles a

las enfermedades crónicas no transmisibles, entre las que se incluyen las enfermedades cardiovasculares, procesos que están adquiriendo una relevancia inesperada con el aumento de la expectativa de vida. Pero, además, nuestro estudio señala que este hecho no es exclusivo de países pobres, sino que afecta también a las poblaciones más pobres dentro de los países ricos, como España.

Nuestros resultados, por ser fruto de un estudio observacional ecológico, no nos permiten explicar la etiopatogenia de nuestros hallazgos. En un documento reciente, antes mencionado, se ha señalado que las poblaciones pobres sufren lo que se ha denominado inseguridad alimentaria, debido a una disponibilidad limitada en cantidad y calidad de los alimentos necesarios para cubrir los requerimientos nutricionales adecuados. Dicha situación favorece una alimentación insana, más barata y calórica, con consumo preferente de hidratos de carbono simples, poco nutritivos. Ese hecho se da no solo en países pobres, sino en los sectores de población más pobres dentro de un país con buen nivel económico. De esta forma, este tipo de situación alimentaria, que favorece la obesidad y sus complicaciones, podría explicar la mayor mortalidad por IAM en nuestras provincias más pobres. De hecho, la prevalencia de obesidad en España, según los datos mostrados en el mapa de obesidad del estudio ENPE, es mayor en las comunidades económicamente más desfavorecidas, hecho similar al encontrado desde hace años en relación con la prevalencia de diabetes mellitus tipo 2. Por tanto, aunque las limitaciones de nuestro estudio no nos permitan establecer de forma definitiva ese nexo de obesidad y mayor muerte por IAM, lo hacen bastante probable.

Nuestros datos sobre mortalidad por enfermedades circulatorias y cerebrovasculares siguieron una tendencia similar a los del IAM, pero sin alcanzar significación. Entre las razones que expliquen este resultado pueden estar el que no hicimos un estudio ajustado por edad o por otros factores de riesgo, que pueden ser diferentes entre provincias, y ello podría dificultar que la comparación entre ellas sea significativa. Por otra parte, la identificación de las enfermedades analizadas se hizo a partir de los resultados del INE, que utiliza un sistema de codificación estandarizado (Clasificación Internacional de Enfermedades, CIE), en el que la selección de la causa básica de defunción se define con reglas descritas. No obstante, en procesos poco definidos nosológicamente, como las enfermedades cardiovasculares y cerebrovasculares, puede haber diferencias de catalogación entre los profesionales, lo que da lugar a la inclusión bajo el mismo término de procesos diferentes y puede ser causa de error en el análisis estadístico. Precisamente, esta debilidad es una de las más obvias de los estudios ecológicos, por la falta de protocolización al recoger la información, al utilizar datos de distintos estudios hechos con criterios diferentes. En conclusión, nuestros datos confirman la hipótesis de que el PIB per cápita por provincias en España se correlaciona de forma negativa con la mortalidad por IAM, por lo que sería deseable plantear estudios de cohortes que permitan profundizar en dicha evidencia e incluso conocer si existe una relación causal entre dichas variables.

Conclusiones

En toda publicación la discusión debe acabar con una conclusión y una propuesta de futuro, tal como hemos incluido en el apartado anterior.

Interés social y para el sistema de salud

La enfermedad cardiovascular es un proceso multifactorial cuyo manejo ha sido abordado de una forma reduccionista, ya que se ha apoyado en la identificación de los factores de riesgo y en su tratamiento en los individuos aislados. En los últimos años se ha identificado la gran importancia que tienen los factores sociales y ambientales en relación con la distinta prevalencia del proceso entre poblaciones, cuyo abordaje debería realizarse desde la perspectiva de la salud pública. Por tanto, la demostración de la importancia del nivel económico en el riesgo de morir por enfermedades crónicas obliga a conocer la dimensión real de este factor, para diseñar políticas de salud que sean eficaces, en especial sobre los colectivos más vulnerables, asociados a una mayor pobreza.

Agradecimientos

En toda publicación se debe incluir la colaboración de personas que sin ser autores han participado en el estudio, así como las fuentes de financiación y los conflictos de interés de los autores. En este caso agradezco a los doctores Juan Francisco Alcalá Díaz (Hospital Universitario Reina Sofía de Córdoba) y Francisco José Gómez Delgado (Hospital Universitario de Jaén) su ayuda en el análisis estadístico de los datos.



Bibliografía

Aranceta-Bartrina J, Pérez-Rodrigo C, Alberdi-Aresti G, Ramos-Carrera N, Lázaro-Masedo S. Prevalence of General Obesity and Abdominal Obesity in the Spanish Adult Population (Aged 25-64 Years) 2014-2015: The ENPE Study. *Rev Esp Cardiol*. 2016; 69:579-87.

Carrillo J, González MR. Iris: Codificador automático internacional de causas de muerte. Madrid: INE; 2016 [Documento de trabajo 1/2016].

Hernández Ávila M. Epidemiología: diseño y análisis de estudios. México, DF: Editorial Médica Panamericana; 2009. ISBN: 978-968-7988-87.

Instituto Nacional de Estadística [Internet]. Madrid: INE; s.f. Defunciones según la causa de muerte 2018; [consultado el 3 de julio de 2021]. Disponible en: <https://www.ine.es/jaxi/Tabla.htm?path=/t15/p417/a2018/l0/&file=03001.px&L=0>.

Instituto Nacional de Estadística [Internet]. Madrid: INE; s.f. Población residente por fecha, sexo y edad. Resultados por provincias [datos correspondientes al año 2018]; [consultado el 3 de julio de 2021]. Disponible en: <https://www.ine.es/jaxiT3/Datos.htm?t=9687>.

Maresova P, Javanmardi E, Barakovic S, et al. Consequences of chronic diseases and other limitations associated with old age, a scoping review. *BMC Public Health*. 2019 Nov 1;19(1):1431.

Morgenstern H. Uses of ecologic analysis in epidemiologic research. *Am J Public Health* 1982;72(12): 1336-44.

Wikipedia [Internet]. St. Petersburg, FL: Wikimedia Foundation, Inc. 2001- . Anexo:Provincias de España por PIB [consultado el 3 de julio de 2021]. Disponible en: https://es.wikipedia.org/wiki/Anexo:Provincias_de_Espa%C3%B1a_por_PIB.

Willett W, Rockström J, Loken B, et al. Food in the Anthropocene: the EAT-Lancet Commission on healthy diets from sustainable food systems. *Lancet*. 2019 Feb 2;393(10170):447-92.

Preguntas de autoevaluación

1. ¿Qué significa *falacia o sesgo ecológico*?

- a) Que los resultados de los estudios ecológicos no tienen el mayor nivel de evidencia.
- b) Que se pretenden extrapolar los resultados poblacionales a individuos específicos.
- c) Que se necesita comparar grupos de individuos a lo largo del tiempo para evitar los sesgos de los estudios transversales.
- d) Que hay que recoger los datos protocolizando dicha recogida para que todos los observadores utilicen la misma metodología.

La respuesta correcta es la opción *a*.

Estos estudios consisten en comparar poblaciones o grupos, por lo que los resultados no se pueden atribuir a individuos aislados. Cuando esto se hace estamos realizando un sesgo ecológico. El resto de las respuestas no están relacionadas con este hecho.

2. ¿Qué afirmación es válida sobre el análisis estadístico de los datos?

- a) El coeficiente de correlación de Spearman se utiliza cuando los datos son normales.
- b) El coeficiente de correlación de Pearson se utiliza cuando los datos no son paramétricos.
- c) El test de Shapiro-Wilks se utilizar para saber si una muestra de datos sigue una distribución normal.
- d) En los estudios ecológicos el análisis estadístico es independiente de la normalidad de los datos.

La respuesta correcta es la opción *c*.

El test de Shapiro-Wilks es obligado, antes de hacer un estudio de correlación, para saber si los datos son o no normales, hecho que también sucede en los estudios ecológicos. Una vez conocido si son o no paramétricos, se procede a aplicar el coeficiente de correlación de Pearson cuando son paramétricos, o el coeficiente de correlación de Spearman (ρ) cuando no lo son.

3. Indique cuál de las siguientes afirmaciones es verdadera en relación con los estudios ecológicos:

- a) Como en todo estudio experimental, es importante definir antes el tamaño de muestra de la población participante.
- b) Los estudios ecológicos tienen un nivel de evidencia superior al de los estudios de cohortes.
- c) Los estudios ecológicos se deben realizar siempre analizando los datos a lo largo del tiempo.
- d) Los estudios ecológicos múltiples utilizan como fuente de datos las estadísticas de morbilidad y mortalidad rutinarias.

La respuesta correcta es la opción c.

Los estudios ecológicos se realizan sin hacer previamente un análisis del tamaño de la muestra. Tienen bajo nivel de evidencia y son útiles para establecer nuevas hipótesis. Cuando se hacen análisis a lo largo del tiempo se denominan estudios ecológicos de series de tiempo, pero eso no es obligado en las distintas modalidades de estudios ecológicos. La modalidad de estudios múltiples utiliza siempre datos de mortalidad o morbilidad rutinarias.

Importancia de los subestudios

Dra. Isabel Martín-Garrido

Dra. Lourdes Moreno Gaviño

Servicio de Medicina Interna

Hospital Universitario Virgen del Rocío (Sevilla)

La publicación seleccionada de este capítulo es un trabajo real en el que las autoras han realizado una modificación de datos como base para el desarrollo del tema del capítulo



capítulo 8

Publicación seleccionada

Introducción

En los pacientes con diabetes mellitus tipo 2 (DM2) que no pueden alcanzar el objetivo de hemoglobina glucosilada (HbA_{1c}) con metformina, con o sin medicamentos antidiabéticos orales adicionales, se puede iniciar terapia con insulina basal. Existen dos nuevas insulinas basales de segunda generación, insulina Moon e insulina Sunset, con perfiles de acción suave y riesgo reducido de hipoglucemias. A pesar de que existen ensayos clínicos controlados que demuestran que la insulina Moon permite un mejor control glucémico con menos riesgo de hipoglucemias en comparación con la insulina basal Sun, no tenemos datos de vida real en los que se compare con otras insulinas basales de segunda generación ni valoración de la utilización de los recursos sanitarios.

Objetivos

Evaluar los resultados clínicos, la necesidad de atención médica en centros sanitarios y los costes en pacientes con DM2 que utilizan insulina basal y cambian a insulina Moon u otra insulina basal (Sun/Sunset) en entornos clínicos del mundo real.

Diseño del estudio

Estudio retrospectivo de cohortes, utilizando datos de los registros médicos electrónicos obtenidos de la base de datos de DIRAYA del Servicio Andaluz de Salud. Se incluyó a pacientes con DM2 tratados con insulina basal y que hicieron un cambio a otras insulinas. Seguimiento de 6 meses.

Los criterios de inclusión fueron los siguientes: diagnóstico de DM2, tiempo de enfermedad de al menos 1 año y seguimiento posterior, y prescripción de insulina Moon, Sun o Sunset en esa fecha. Los criterios de exclusión fueron la presencia de diabetes mellitus tipo 1 y el uso de más de una insulina.

Se recogieron datos sobre la efectividad: cambio en la HbA_{1c} desde el inicio al seguimiento y proporción de pacientes que alcanzan los valores predeterminados objetivo de HbA_{1c} inferior a 7% (53 mmol/mol) y 8% (64 mmol/mol) durante el periodo de seguimiento de 6 meses; y sobre seguridad: incidencia y tasas de eventos de hipoglucemia o concentración de glucosa plasmática de 70 mg/dL o menos e hipoglucemia asociada con la hospitalización o visita al servicio de urgencias durante el periodo de seguimiento de 6 meses, y proporción de pacientes que requieren hospitalización, urgencias y servicios ambulatorios (para hipoglucemias y eventos relacionados con la diabetes y por todas las causas).

Resultados

Se identificaron a 2496 pacientes que habían cambiado a insulina Moon y 3837 que cambiaron a insulina Sun o Sunset. Después del emparejamiento quedaron compuestas dos cohortes de 1619 pacientes. La HbA_{1c} media basal fue del 8,9% (cohorte de cambio a insulina Moon) y 8,93% (cohorte de cambio a otras insulinas basales), y se redujo al final del seguimiento al 8,4% en ambas ($p < 0,001$ para ambas cohortes); el porcentaje de pacientes que sufrió algún episodio hipoglucémico fue significativamente inferior en la cohorte de cambio a insulina Moon ($p = 0,015$). La tasa de eventos en los pacientes que cambiaron a insulina Moon fue un 19% inferior al de los pacientes que cambiaron a otra insulina basal, una vez ajustado por la incidencia basal de hipoglucemia. Esto se tradujo en una disminución de uso de recursos sanitarios. La cohorte de cambio a insulina Moon se asoció con un menor riesgo de hospitalización por hipoglucemia (-33%), de visita a urgencias (-38%) y de consultas externas (-23%) que la cohorte de cambio a otras insulinas basales.

Adicionalmente, se realizó un subestudio en el que se incluyó solo a los pacientes de más de 65 años. La cohorte de cambio a insulina Moon quedó integrada por 1176 pacientes, y la cohorte de cambio a otras insulinas basales por 1142 pacientes. Después del cambio de insulina basal, las reducciones de HbA_{1c} fueron mayores o similares con insulina Moon frente a otras insulinas basales (seguimiento variable: $-0,45\% \pm 1,40\%$ frente a $-0,29\% \pm 1,57\%$; $p = 0,021$; seguimiento fijo: $-0,48\% \pm 1,49\%$ frente a $-0,38\% \pm 1,59\%$; $p = 0,114$), mientras que el logro de la meta de HbA_{1c} fue similar en ambas cohortes. El uso de insulina Moon se asoció con menos hipoglucemia (razón de eventos ajustada: 0,63; intervalo de confianza del 95% [IC95%]: 0,53-0,75; $p < 0,001$) e hipoglucemia asociada a hospitalización o visita a urgencias (cociente de riesgo ajustado: 0,58 [IC95%: 0,37-0,90, $p = 0,016$]; reducción absoluta del riesgo: 0,43 [IC95%: 0,31-0,60, $p < 0,001$]) por variable de seguimiento. Con un seguimiento fijo de 6 meses, las incidencias de las visitas a centros médicos fueron similares en ambas cohortes, independientemente de si estaban o no relacionadas con la diabetes o hipoglucemias. Las tasas de eventos también fueron similares en las cohortes. El número de días de ingreso fue menor en la cohorte de insulina Moon, y alcanzó significación estadística para los días de ingreso relacionados con la hipoglucemia (0,27 frente a 0,61 días; $p = 0,048$).

Se realizó otro subanálisis de esta cohorte con pacientes de más de 75 años. La HbA_{1c} media disminuyó significativamente desde el inicio hasta los 6 meses en ambas cohortes, con reducciones comparables (seguimiento fijo: $-0,3\% \pm 1,3\%$ y $-0,3\% \pm 1,4\%$; $p = 0,98$). El logro de HbA_{1c} por debajo de 8% también fue comparable en ambas cohortes, mientras que el logro de HbA_{1c} inferior a 7% fue significativamente mejor en la cohorte con insulina Sun o Sunset. Con un seguimiento fijo de 6 meses, los pacientes mayores de 75 años tenían en general más probabilidades

de tener hipoglucemia que la población general; sin embargo, debido al número insuficiente de pacientes (135 en cada cohorte tras emparejarlos) solo se puede proporcionar una estadística descriptiva.

Discusión

La media de HbA_{1c} disminuyó significativamente desde la basal hasta el seguimiento a 6 meses en ambos grupos, con reducciones significativamente mayores en la cohorte de insulina Moon, aunque sin diferencias estadísticamente significativas. Los objetivos de HbA_{1c} inferior a 7% u 8% no presentaron diferencias significativas entre la insulina Moon y las otras insulinas basales. Respecto a la seguridad, todos los resultados de hipoglucemia fueron significativamente favorables a la insulina Moon, sobre todo en las hipoglucemias nocturnas. En el subgrupo de pacientes mayores de 65 años, el cambio a insulina Moon frente a las insulinas Sun y Sunset se asoció con mejoras mayores y similares, respectivamente, en el control glucémico y, en general, menos hipoglucemia, sobre todo nocturna, así como con menor número de días de ingreso por hipoglucemia en el grupo de insulina Moon. Sin embargo, no se pueden extrapolar datos a pacientes mayores de 75 años por la escasa representación muestral, y sería necesario realizar un estudio de cohortes en este grupo de edad para establecer la seguridad.

Introducción

Un subestudio es el realizado sobre una parte del grupo de sujetos incluidos en el ensayo clínico o estudio observacional. Estos subestudios pueden estar relacionados o no con el objetivo principal de estos y pueden estar prediseñados desde el inicio del estudio. Además, pueden precisar un consentimiento informado adicional, como es el caso de los subestudios farmacocinéticos o los farmacogenéticos.

Los subestudios tienen la ventaja de analizar situaciones específicas sobre grupos de pacientes concretos (de una determinada edad, con una función renal concreta, estadiaje determinado o patología...). En definitiva, permiten ver resultados de un determinado ensayo clínico en pacientes con circunstancias clínicas concretas.

Así se pueden llegar a detectar diferencias de eficacia o de seguridad en diferentes grupos que en el estudio principal pasan desapercibidos. En este sentido, los resultados de los subestudios pueden ser congruentes con los resultados del estudio principal o no.

Pero por este motivo, puesto que el diseño del estudio no se realizó para responder a esa pregunta clínica concreta (sobre ese grupo de pacientes determinados), hemos de ser cautelosos en las conclusiones que extrapolemos.

Planteamiento

Realizar un estudio en vida real nos permite obtener información más fiable y representativa del alcance de los cambios de una exposición en la práctica clínica habitual (efectividad) que no siempre coincide con lo observado en los ensayos clínicos (eficacia). Si además nos interesa el impacto económico de una medida implementada, estaríamos hablando de eficiencia. En este caso, se realiza un estudio de cohortes retrospectivo para evaluar la efectividad y la seguridad de una nueva insulina basal frente a las ya comercializadas en vida real. Se quieren conocer los cambios en la HbA_{1c}, la tasa de hipoglucemias y los eventos relacionados con estas hipoglucemias (ingresos hospitalarios, visitas al servicio de urgencias o al ambulatorio).

Además, dentro de este estudio se realiza un subanálisis del subgrupo de pacientes mayores de 65 años y otro en mayores de 75 años. La importancia de realizar estos subanálisis radica en que en España el 63% de los pacientes con diabetes (de los que más del 90% presentan DM2) tienen más de 65 años y más de un tercio de la población mayor de 75 años padece diabetes. Además, el objetivo del tratamiento en el subgrupo de más de 75 años (ancianos) es diferente al de la población más joven. La Asociación Americana de Diabetes (ADA) en su último documento de consenso, así como otras sociedades científicas, debido a que las características clínicas y funcionales de esta franja de la población son diferentes a las de las personas más jóvenes, cambia en alguna medida el foco de control estricto de la glucemia y desvía la atención a la mejora de la calidad de vida, la preservación de la funcionalidad de estos pacientes y la protección frente a los efectos adversos, sobre todo las hipoglucemias, ya que estas conllevan un aumento del deterioro cognitivo, síndrome confusional, caídas y fracturas, ictus, arritmias cardíacas y mayor mortalidad.

Pregunta clínica

La pregunta de este estudio sería saber si en vida real el cambio a insulina basal de segunda generación tipo insulina Moon es similar en cuanto a resultados clínicos y necesidad de atención médica en pacientes con DM2 con respecto a otras insulinas basales tipo Sun o Sunset. Si utilizamos la propuesta PICOT (pacientes, intervención, comparación, *outcomes*/resultados, tiempo) para ver el diseño del estudio podemos decir que están bien definidos, salvo la variable del tiempo, ya que, a pesar de que está estimado el seguimiento a 6 meses, probablemente sea insuficiente para ver efectos de más peso, como cambios en la mortalidad, los eventos cardiovasculares mayores y los propios que se miden en el estudio.

Los investigadores, a su vez, se preguntan si en el subgrupo de pacientes de más de 65 años esta nueva insulina basal, insulina Moon, tiene un comportamiento similar al de las otras. Todos conocemos que los pacientes mayores de 65 años, y más aún los de 75 años o con pluripatología, presentan más riesgo de hipoglucemia y repercusiones más graves de estas, lo que conlleva una pérdida de calidad de vida, deterioro funcional y yatrogenia. Por este motivo, sería interesante saber si existe alguna diferencia, peculiaridad o especificidad para poder recomendar alguna insulina basal en este subgrupo específico de pacientes.

Justificación del diseño

El estudio es un estudio retrospectivo de cohortes, con distribución emparejada de pacientes para limitar los factores de confusión entre cohortes y maximizar la comparabilidad.

La distribución emparejada de pacientes (*propensity score matching*, PSM, por sus siglas en inglés) consiste en una selección emparejada de un paciente de cada grupo de tratamiento cuyas características clínicas basales y sociodemográficas sean similares hasta obtener dos poblaciones de estudio comparables. Los datos utilizados en el estudio proceden de bases de historias clínicas electrónicas, lo que permite analizar un volumen de datos mucho mayor y con un enfoque de poblaciones en vez de un enfoque de muestras.

Para la medición de las reducciones de HbA_{1c} en el seguimiento frente a las iniciales dentro de cada cohorte se utilizaron pruebas T pareadas. Las reducciones de HbA_{1c} se compararon entre cohortes mediante la prueba *t* de Student, mientras que el logro de la meta de HbA_{1c} se comparó entre cohortes mediante el test χ^2 . Los cocientes de riesgo ajustados para la hipoglucemia inicial se calcularon para el primer evento de hipoglucemia utilizando un modelo de regresión de Cox. Se calcularon las proporciones de tasas ajustadas para la hipoglucemia basal para todas las tasas de eventos de hipoglucemia utilizando la regresión de Poisson. Mediante regresión logística se calcularon las *odds ratio* (OR) para la incidencia de hipoglucemia. Se calcularon las diferencias de medias de mínimos cuadrados ajustados por hipoglucemia basal para las tasas de eventos de hipoglucemia mediante un procedimiento de modelo lineal generalizado.

Sobre el enfoque inicial del estudio se realiza un subanálisis en el que lo que se quiere evaluar es un subgrupo de pacientes específico. Este subgrupo de pacientes son los pacientes mayores de 65 años y, entre ellos, los mayores de 75 años, realizando las mismas medidas que en el estudio principal.

Interpretación del artículo. Fortalezas y debilidades

Fortalezas

Se han diseñado las cohortes de pacientes mediante el método de distribución emparejada (PSM) y se ha obtenido, por tanto, una situación clínica y sociodemográfica similar. Se ha realizado en vida real, con obtención de los datos de las bases electrónicas poblacionales, y la posibilidad de representar todos los subgrupos de pacientes en condiciones reales. En cuanto a las medidas de intervención, se realizan comparaciones con las insulinas basales comercializadas con la nueva insulina basal, y se comparan, por tanto, fármacos de la misma familia y clase, con la posibilidad de discernir si alguna de ellas es superior a las otras. El tamaño muestral para el estudio inicial, que es todos los pacientes con DM2 tratados con insulina, es adecuado para el impacto que se espera. Así mismo, el tamaño muestral del subgrupo de pacientes mayores de 65 años es adecuado, pero no así para el subgrupo de más de 75 años.

Debilidades

El periodo de seguimiento del estudio es muy corto, tan solo de 6 meses. Pueden existir sesgos al utilizar como fuente de bases de datos la historia de salud DIRAYA, por las malas codificaciones de los ingresos o el número de consultas, por la falta de registros de estas situaciones, por no registrarse en los informes médicos, etc. No se recoge el porqué del cambio de insulina basal, si han sido efectos adversos mayores con las otras insulinas, o si se trata de una decisión del médico o del paciente, y eso es importante a la hora de saber qué está ocurriendo en la vida real.

Los objetivos primarios del estudio son poco exigentes. Habitualmente, los estudios enfocados en la diabetes y otras enfermedades cardiovasculares tienen objetivos robustos y potentes: mortalidad, evento cardiovascular mayor, calidad de vida, entre otros; en este sentido, aunque se miden los eventos relacionados con hipoglucemias (ingreso hospitalario, atención en urgencias o centro médico) como estándar de seguridad, deberían haberse incluido esos otros objetivos más fuertes. Por otra parte, los puntos de corte de HbA_{1c} son <7% y <8% sin tener en cuenta el tipo de paciente por edad y comorbilidad en el análisis global, y se debería haber estratificado por edades los cambios en la HbA_{1c}. Otro de los factores no recogido son los cambios que pudiesen ejercer los procesos de educación diabético-lógica, si los hubiera, o la adición de otro fármaco oral hipoglucemiante.

El subanálisis en pacientes de más de 75 años, a pesar de hacerse de forma adecuada el emparejamiento, tiene un tamaño muestral pequeño, por lo que no se pueden extraer conclusiones, y es este subgrupo el que más se beneficiaría de la reducción del número de hipoglucemias, ya que es el más vulnerable.

Valoración de su aplicabilidad en la práctica clínica

Aunque los ensayos clínicos aleatorizados proporcionan información fiable, las condiciones especiales y los criterios estrictos de inclusión y exclusión pueden no reflejar la práctica clínica ni a los pacientes del mundo real. Por lo tanto, este estudio explora la práctica clínica habitual y proporciona información más ajustada a la realidad del día a día y puede

ser útil para pacientes y médicos implicados en el tratamiento de la DM2. Los resultados de este estudio tienen que tomarse en consideración según la pregunta inicial del estudio, es decir, pacientes con DM2. En cuanto al subgrupo de pacientes mayor de 65 años, los resultados son favorables para la insulina Moon en cuanto al menor número de hipoglucemias, sobre todo nocturnas, y la disminución de los días de ingreso por estas. Sin embargo, no tenemos datos de mortalidad ni eventos cardiovasculares mayores. En el subgrupo de pacientes mayores de 75 años no se han encontrado diferencias entre la insulina Moon y el resto de las basales, pero el bajo número de sujetos limita la potencia del análisis, por lo que sería necesario realizar un segundo estudio y no generalizar el uso de insulina Moon hasta que no se tengan más datos de estudios dirigidos a contestar esta pregunta de investigación.



Bibliografía

Bailey TS, Wu J, Zhou FL, et al. Switching to insulin glargine 300 units/mL in real-world older patients with type 2 diabetes (DELIVER 3). *Diabetes Obes Metab*. 2019;21(11):2384-93. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/31264346>.

Gómez-Huelgas R, Gómez Peralta F, Rodríguez Mañas L, et al. Treatment of type 2 diabetes mellitus in elderly patients. Tratamiento de la diabetes mellitus tipo 2 en el paciente anciano. *Rev Clin Esp (Barc)*. 2018;218(2):74-88. DOI: 10.1016/j.rce.2017.12.003.

Gordis L. *Epidemiology*. 5.ª ed. Elsevier Saunders; 2014. ISBN: 978-1-4557-3733-8.

Preguntas de autoevaluación

1. Si diseñamos un estudio para estimar los efectos de una intervención en vida real, ¿cuál de los siguientes diseños elegiría?

- a) Un diseño transversal.
- b) Un estudio ecológico.
- c) Un ensayo clínico aleatorizado.
- d) Un estudio de cohortes prospectivo con distribución emparejada de pacientes.

La respuesta correcta es la opción *d*.

A pesar de que tanto los ensayos clínicos aleatorizados como los estudios de cohortes podrían medir el efecto de una intervención, el modelo idóneo sería el estudio de cohortes prospectivo con distribución emparejada de pacientes, ya que se acercaría más a la realidad de los pacientes. El diseño transversal y el ecológico son observacionales sin control en el diseño de las variables.

2. En un estudio en el que se compara una nueva insulina (grupo experimental) frente al tratamiento habitual con insulina basal (grupo control) se han obtenido los siguientes resultados en la prevención de hipoglucemias a los 2 años de tratamiento: nuevo tratamiento, 25 hipoglucemias sobre 500 pacientes; tratamiento habitual, 50 hipoglucemias sobre 500 pacientes. ¿Cuál es el número de pacientes que es necesario tratar (NNT) que obtenemos para evitar una hipoglucemia con la nueva insulina?

- a) 50.
- b) 100.
- c) 20.
- d) 25.

La respuesta correcta es la opción *c*.

El NNT se calcula como el inverso de la reducción absoluta de riesgo, y esta reducción es la diferencia de riesgo relativo. Con los mismos datos que en la pregunta anterior, la reducción absoluta de riesgo será $(25/500) - (50/500)$, es decir, 0,05. Por lo tanto, el NNT será $1/0,05 = 20$.

3. ¿Qué debemos tener en cuenta si realizamos un subestudio en una cohorte?

- a) El análisis de los datos consiste en determinar si la proporción de expuestos en el grupo de pacientes con la enfermedad difiere de la de los controles.
- b) Se obtiene la prevalencia como medida de la frecuencia de la enfermedad.
- c) El azar decide la distribución de la exposición en los sujetos del estudio.
- d) Puede ser retrospectivo.

La respuesta correcta es la opción *d*.

En un estudio de cohortes no compararemos la proporción de pacientes expuestos frente a los no expuestos, ni tampoco obtendremos la prevalencia, sino la incidencia. El azar no decide la distribución de la exposición. La única afirmación correcta es la *d*: los estudios de cohortes (o longitudinales) pueden ser retrospectivos, así como prospectivos y ambispectivos (seguimiento retrospectivo y prospectivo de la misma cohorte).

Cómo comparar grupos: *propensity score*

Dr. Martín Fabregate Fuente

Dra. Nuria Bara Ledesma

Servicio de Medicina Interna

Investigación Biomédica. Gestión I+D+i. Grupo de Enfermedades Multisistémicas, IRYCIS
Hospital Universitario Ramón y Cajal (Madrid)

Dr. Luis Manzano Espinosa

Jefe de Servicio de Medicina Interna

Hospital Universitario Ramón y Cajal (Madrid)

Catedrático de Medicina

Universidad de Alcalá (Alcalá de Henares, Madrid)

*La publicación seleccionada de este capítulo
es un trabajo real publicado (véase la bibliografía)*



capítulo 9

Publicación seleccionada

Efecto del tratamiento con betabloqueantes sobre la mortalidad en pacientes con insuficiencia cardiaca y fracción de eyección preservada tras un ingreso por insuficiencia cardiaca aguda

Introducción

La insuficiencia cardiaca con fracción de eyección preservada (ICFEp) es un síndrome clínico en el que los pacientes presentan síntomas y signos de insuficiencia cardiaca como resultado de una presión de llenado ventricular elevada, con una fracción de eyección del ventrículo izquierdo (FEVI) normal ($\geq 50\%$). La ICFEp es más frecuente entre las mujeres de edad avanzada con hipertensión, diabetes mellitus, enfermedad arterial coronaria y fibrilación auricular.

Al menos la mitad de los pacientes ingresados por insuficiencia cardiaca aguda presentan una FEVI preservada. Sin embargo, los datos disponibles hasta la fecha sobre el tratamiento de esta enfermedad son escasos. Los ensayos clínicos aleatorizados para evaluar la eficacia de diferentes terapias farmacológicas en pacientes con ICFEp no han aportado una evidencia lo suficientemente sólida en este grupo de población, ya que el tamaño muestral incluido era escaso o bien los resultados derivados de ellos no han sido concluyentes.

Objetivo

El objetivo del presente estudio fue investigar la asociación entre el tratamiento con betabloqueantes (administrados durante al menos tres meses antes del ingreso) y la mortalidad por todas las causas a los dos años en pacientes con ICFEp hospitalizados por insuficiencia cardiaca aguda.

Métodos

Se realizó un estudio de cohorte retrospectivo de 500 pacientes adultos con ICFEp hospitalizados por insuficiencia cardiaca aguda en un hospital terciario entre 2001 y 2005. De ellos, 150 (30%) recibieron tratamiento con betabloqueantes durante al menos 3 meses antes del ingreso. Se excluyeron del análisis los pacientes ingresados con infarto agudo de miocardio, los pacientes con neoplasias avanzadas, enfermedades hepáticas y pulmonares en fase terminal, así como pacientes con ICFEr, valvulopatías significativas y aquellos sin estudio ecocardiográfico. Para los sujetos con múltiples ingresos se incluyó el primer ingreso dentro del periodo de estudio. El resultado primario fue la mortalidad por todas las causas a los dos años (seguimiento máximo disponible en todos los sujetos).

Se compararon dos grupos de pacientes: los que recibieron betabloqueantes en los 3 meses anteriores al ingreso (grupo BB) y los que no (grupo NBB). Para ajustar un posible desequilibrio en las características basales entre los dos grupos, el análisis principal se restringió a un subconjunto de pacientes ($n = 300$): los 150 pacientes en tratamiento previo con betabloqueantes y 150 pacientes que no recibieron betabloqueantes, pero que tenían una probabilidad o *propensity score* similar de recibir betabloqueantes al inicio del estudio.

Se estimó el *propensity score* para el tratamiento con betabloqueantes para cada paciente utilizando un modelo de regresión logística multivariable, en el que la toma de betabloqueantes se modeló utilizando todas las características basales de los pacientes. A continuación, se utilizó el *propensity score* para emparejar a cada paciente del grupo BB con un paciente del grupo NBB con un valor similar (diferencia de *propensity score* inferior al 10%).

Los resultados se presentaron como la media y la desviación típica para las variables continuas y como el total de pacientes y el porcentaje para los datos categóricos. Se utilizó la prueba T para la comparación de las variables continuas y la prueba de χ^2 para los datos categóricos, con el uso de la prueba exacta de Fisher si era necesario. Las curvas de supervivencia se calcularon mediante el método de Kaplan-Meier y la comparación entre los grupos de pacientes con y sin tratamiento con betabloqueantes se realizó mediante una prueba de Mantel-Cox o *log-rank*. Un valor de p bilateral menor de 0,05 se consideró estadísticamente significativo.

Resultados

En la **tabla 1** se comparan las características basales de los pacientes con ICFEP según el uso de betabloqueantes, tanto antes como después del emparejamiento mediante *propensity score*. Antes del emparejamiento, los pacientes en tratamiento con betabloqueantes previo al ingreso tenían menor edad media, mayores tasas de hipertensión y diabetes, y recibían más frecuentemente tratamiento con inhibidores de la encima convertidora de angiotensina (IECA) y antagonistas del receptor de angiotensina II (ARA-II), bloqueantes de los canales de calcio o estatinas. Tras el emparejamiento, los pacientes de ambos grupos fueron similares en cuanto a las covariables basales medidas.

Durante el periodo de seguimiento de dos años tras la hospitalización, se produjeron 75 fallecimientos (sobre 150 pacientes) en el grupo BB (mortalidad del 50%). Entre los pacientes sin tratamiento con betabloqueantes la tasa de mortalidad fue del 60% (210 sobre 350), mientras que tras el emparejamiento por *propensity score* la tasa de mortalidad en el grupo NBB fue del 62% (93 sobre 150). Se observaron diferencias en la supervivencia entre los grupos BB y NBB durante los dos años de seguimiento, que fueron estadísticamente significativas (prueba de *log-rank*, $p = 0,016$).

Tabla 1. Características basales de los pacientes por uso de betabloqueantes antes y después de emparejar por *propensity score*

	Grupos no pareados			Grupos pareados mediante PS		
	Con BB n = 150	Sin BB n = 350	p	Con BB (Grupo BB) n = 150	Sin BB (Grupo NBB) n = 150	p
Edad media y desviación típica (años)	76 (10)	78 (11)	0,03	76 (10)	76 (11)	0,40
Mujeres	93 (62 %)	196 (56 %)	0,28	93 (62 %)	90 (60 %)	0,56
Comorbilidades						
Hipertensión	118 (79 %)	217 (62 %)	0,01	118 (79 %)	117 (78 %)	0,79
Diabetes	75 (50 %)	119 (34 %)	<0,01	75 (50 %)	72 (48 %)	0,62
EPOC	9 (6 %)	70 (20 %)	<0,01	9 (6 %)	12 (8 %)	0,49
Fibrilación auricular	60 (40 %)	161 (46 %)	0,32	60 (40 %)	66 (44 %)	0,42
Insuficiencia renal	20 (13 %)	49 (14 %)	0,87	20 (13 %)	21 (14 %)	0,70
Cardiopatía isquémica	78 (52 %)	168 (48 %)	0,46	78 (52 %)	75 (50 %)	0,55
Tratamiento farmacológico						
IECA/ARA-II	100 (67 %)	119 (34 %)	<0,01	100 (67 %)	99 (66 %)	0,68
Bloqueantes de los canales de calcio	57 (38 %)	49 (14 %)	<0,01	57 (38 %)	54 (36 %)	0,52
Estatinas	63 (42 %)	35 (10 %)	<0,01	63 (42 %)	57 (38 %)	0,39
PS: <i>propensity score</i> ; BB: betabloqueantes; EPOC: enfermedad pulmonar obstructiva crónica; IECA: inhibidores de la enzima convertidora de angiotensina; ARA-II: antagonistas de los receptores de la angiotensina II.						

Conclusiones

En este estudio, el tratamiento con betabloqueantes, al menos tres meses antes del ingreso, se asoció a una mejora de la supervivencia a los dos años en los pacientes hospitalizados por descompensación aguda de la ICfEp. El tratamiento con betabloqueantes puede tener un efecto protector sobre la supervivencia de los pacientes con ICfEp.

Introducción

Un ensayo clínico aleatorizado (ECA) es un estudio experimental en el que los participantes se distribuyen de manera aleatoria en distintos grupos de intervención: un grupo experimental, que recibe la intervención que quiere estudiarse; y un grupo control, que recibe una intervención distinta y sirve de referencia para comparar el resultado obtenido. La asignación aleatoria de los participantes a cada grupo de intervención permite establecer relaciones de causalidad entre intervención y resultado, y es el diseño óptimo para evaluar el efecto de una intervención sanitaria.

Sin embargo, la realización de un ECA no siempre es viable por motivos éticos, pues no es admisible, por ejemplo, imponer el hábito tabáquico a una persona para estudiar si desarrollará cáncer. Además, la realización de un ECA, al tratarse de un estudio prospectivo, suele conllevar un elevado coste, habitualmente superior al de estudios observacionales (cohorte o casos y controles), especialmente cuando es posible utilizar datos retrospectivos. Debido a estas limitaciones, existe un interés creciente por utilizar estudios observacionales (o no aleatorizados) para la estimación del efecto del tratamiento. No obstante, en este tipo de estudios la asignación del tratamiento para cada paciente no es aleatoria, ya que suele verse influenciada por las características del sujeto. Como resultado, los participantes de los diferentes grupos pueden diferir sistemáticamente en algunas de sus características. Estos desequilibrios en la distribución de las variables basales pueden conducir a estimaciones sesgadas en el efecto de la intervención estudiada.

Para subsanar estas limitaciones de los estudios observacionales, se han desarrollado diferentes métodos estadísticos con el fin de controlar la influencia de las variables de confusión al estimar el efecto del tratamiento. La aproximación más común es utilizar un modelo de regresión multivariante para ajustar el efecto de los factores de confusión conocidos. Sin embargo, el ajuste mediante modelos multivariante puede presentar limitaciones cuando el número de predictores es elevado en relación con el número de eventos en la variable principal de resultado. Un método de ajuste alternativo para comparar grupos es la utilización de índices o puntuaciones de propensión, más conocidos por su nombre en inglés, *propensity score*. Esta metodología, objeto del presente capítulo, fue desarrollada por Rosenbaum y Rubin en la década de 1980.

El *propensity score* es la probabilidad que tiene cada participante del estudio de recibir un determinado tratamiento dadas sus características basales. Para predecir estas probabilidades a partir de los datos disponibles se utilizan modelos estadísticos, como, por ejemplo, modelos de regresión logística. Al parear a los pacientes según los *propensity scores* estimados podemos diseñar estudios observacionales que tengan grupos aproximadamente equilibrados en las covariables observadas, tratando así de imitar en cierto modo el resultado de la asignación de tratamiento aleatoria realizada en los ECA. Un aspecto esencial que hay que tener en cuenta al usar *propensity score* para el diseño de estudios es que las estimaciones deben crearse sin tener acceso a los resultados estudiados, de manera análoga a lo que ocurre en un ECA.

Para entender mejor el *propensity score*, analizaremos a continuación un ejemplo de su uso para parear dos grupos de pacientes con insuficiencia cardiaca con el objetivo de estudiar el efecto del tratamiento con betabloqueantes en su supervivencia tras un ingreso.

Planteamiento

En el momento en que se planteó el estudio, el tratamiento con betabloqueantes se consideraba un tratamiento estándar para los pacientes con ICfEr. En estos pacientes, el uso de betabloqueantes se había asociado a un menor riesgo de hospitalización por insuficiencia cardiaca aguda.

En los datos previos del estudio SENIORS, el tratamiento con nebivolol, un betabloqueante, mostró un efecto similar en la variable combinada de mortalidad por todas las causas y hospitalización por causa cardiovascular tras 21 meses de seguimiento en pacientes ancianos con ICfEp e ICfPr.

Sin embargo, de acuerdo a la evidencia disponible hasta la fecha, ningún tratamiento había demostrado de forma convincente reducir la morbilidad y la mortalidad en pacientes con ICfEp. No obstante, una gran proporción de estos pacientes recibían betabloqueantes para el tratamiento de otras patologías, como la hipertensión o la cardiopatía isquémica.

Formulación de la pregunta clínica

Los autores se plantearon abordar la pregunta clínica de si el tratamiento con betabloqueantes previo a un ingreso por insuficiencia cardiaca aguda tiene un efecto sobre la mortalidad durante el seguimiento tras el alta en pacientes con ICfEp.

Para ello, el objetivo del estudio fue investigar si existía una asociación entre el tratamiento con betabloqueantes, administrados durante al menos tres meses antes de un ingreso por insuficiencia cardiaca aguda, y la mortalidad por todas las causas a los dos años en pacientes con ICfEp.

Tanto la pregunta clínica como el objetivo del estudio estaban justificados por la evidencia previa disponible, presentada en el planteamiento.

Justificación del diseño y el análisis estadístico

Se diseñó un estudio observacional de cohorte retrospectivo a partir de datos previos de pacientes con ICfEp hospitalizados por una descompensación aguda y seguidos posteriormente durante 2 años para recoger la mortalidad.

Dado que se trataba de un estudio observacional, y la asignación de los tratamientos (betabloqueantes frente a no betabloqueantes) no había sido aleatoria, la distribución de las características basales entre ambos grupos podía diferir notablemente. Para tratar de controlar el efecto de estas covariables en la relación entre el uso de betabloqueantes y la mortalidad, los autores utilizaron la técnica de emparejamiento mediante *propensity score*. Para ello describieron en primer lugar las diferencias basales en las variables medidas antes de parear, y posteriormente realizaron un *propensity score* con estas covariables. Para ello, asignaron a cada paciente tratado con betabloqueantes un control sin betabloqueantes con características similares (*propensity score* similar). Así, tras el emparejamiento por *propensity score*, ambos grupos resultaron comparables en cuanto a las características basales. Finalmente, se realizó la comparación de la tasa de fallecimientos

entre ambos grupos pareados por *propensity score* mediante análisis de supervivencia y la prueba de *log-rank*.

Interpretación del artículo. Fortalezas y debilidades

El estudio tiene varias limitaciones que hay que considerar a la hora de interpretar los resultados. Al tratarse de un estudio retrospectivo realizado en un único centro, presenta las conocidas limitaciones en cuanto a la generalización de los resultados. Dada la naturaleza observacional del estudio, a pesar de la utilización de la metodología de *propensity score* para conformar grupos que estaban bien equilibrados en las variables basales medidas, podría quedar un sesgo oculto debido a la influencia de posibles factores de confusión no medidos. Por ejemplo, en este estudio podría existir algún factor de confusión residual asociado al denominado *efecto de usuario sano*. Así, los pacientes que habían recibido betabloqueantes antes del ingreso podrían estar recibiendo un tratamiento óptimo y, por tanto, el tratamiento previo con betabloqueantes se constituye como un marcador de paciente más sano y mejor tratado.

También cabe señalar que no se analizaron el tipo y las dosis de betabloqueantes, ni el cumplimiento terapéutico. Tampoco se dispuso de datos sobre el tiempo de progresión desde el diagnóstico de la insuficiencia cardiaca en el momento de la hospitalización, por lo que no fue posible descartar que las diferencias en las tasas de supervivencia entre los grupos BB y NBB se debieran a una evolución desigual de la insuficiencia cardiaca antes del ingreso. Otra limitación importante es que, al no realizarse una prueba ecocardiográfica ni medirse los péptidos natriuréticos al ingreso, podrían haberse incluido pacientes hospitalizados por estados edematosos o disnea por motivos distintos de la insuficiencia cardiaca aguda.

El principal punto fuerte de este estudio es que se ha investigado una población de pacientes del "mundo real" con unos criterios de exclusión mínimos en comparación con los estudios anteriores. Además, el uso de una metodología clara y rigurosa ha permitido responder a la pregunta clínica planteada sobre la asociación entre el uso de betabloqueantes y la mortalidad en estos pacientes. Así, a partir de datos de un estudio de cohorte retrospectivo se pudo estudiar dicha asociación controlando el efecto de posibles confusores y desequilibrios en la distribución de las características basales en ambos grupos.

Valoración de su aplicabilidad en la práctica clínica

Al tratarse de un estudio observacional, retrospectivo y monocéntrico, los resultados derivados de él deberían examinarse cuidadosamente antes de decidir sobre su aplicabilidad en la práctica clínica. Sería necesario, por tanto, disponer de nuevos estudios prospectivos llevados a cabo en otros centros para asegurar la reproducibilidad de estos resultados.

Sin embargo, la flexibilidad en los criterios de selección del estudio puede permitir aportar información valiosa sobre el efecto de los betabloqueantes en esta población, que podría servir de base para el diseño posterior de un ECA que permitiera establecer si existe una relación causal entre exposición y resultados.

Resumen

Este ejemplo demuestra cómo el *propensity score* es un instrumento útil para reducir los efectos de la confusión entre grupos de tratamiento en estudios observacionales. Para interpretar con exactitud los resultados es necesario conocer a fondo el modelo de *propensity score* y sus posibles limitaciones. Por tanto, los métodos de *propensity score* no deben interpretarse como una solución universal para abordar los factores de confusión. Sin embargo, utilizados con cuidado, pueden proporcionar a los investigadores una herramienta valiosa para optimizar el diseño y el análisis de un estudio observacional.



Bibliografía

D'Agostino RB Jr. Propensity scores in cardiovascular research. *Circulation*. 2007; 115(17):2340-3.

Dickstein K, Cohen-Solal A, Filippatos G, et al; ESC Committee for Practice Guidelines (CPG). ESC guidelines for the diagnosis and treatment of acute and chronic heart failure 2008: the Task Force for the diagnosis and treatment of acute and chronic heart failure 2008 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association of the ESC (HFA) and endorsed by the European Society of Intensive Care Medicine (ESICM). *Eur J Heart Fail*. 2008;10:933-89.

Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70:41-55.

Van Veldhuisen DJ, Cohen-Solal A, Böhm M, et al. SENIORS Investigators. Beta-blockade with nebivolol in elderly heart failure patients with impaired and preserved left ventricular ejection fraction: data From SENIORS (Study of Effects of Nebivolol Intervention on Outcomes and Rehospitalization in Seniors With Heart Failure). *J Am Coll Cardiol*. 2009;53:2150-8.

Preguntas de autoevaluación

1. ¿Qué característica de un ensayo clínico aleatorizado asegura la distribución homogénea de las características basales entre los grupos de intervención?

- a) El enmascaramiento.
- b) El tamaño muestral.
- c) La asignación aleatoria.
- d) La firma del consentimiento informado.

La respuesta correcta es la opción *c*.

La asignación aleatoria de los participantes a cada rama de intervención permite que los grupos sean homogéneos en la distribución de sus características basales, factores pronósticos conocidos o desconocidos, tratamiento concomitante, etc., y que se diferencien únicamente en la intervención que reciben.

2. ¿Qué es el *propensity score*?

- a) Una escala de riesgo para predecir mortalidad y rehospitalizaciones en pacientes con insuficiencia cardiaca.
- b) La probabilidad que tiene cada participante del estudio de recibir un determinado tratamiento dadas sus características basales.
- c) Una prueba estadística para evaluar si las diferencias en las características basales de dos grupos son significativas.
- d) Un tipo de diseño de los estudios de investigación.

La respuesta correcta es la opción *b*.

El *propensity score* es la probabilidad que tiene cada participante del estudio de recibir un determinado tratamiento dadas sus características basales. Al parear a los pacientes según los *propensity scores* estimados podemos diseñar estudios observacionales que tengan grupos aproximadamente equilibrados en las covariables observadas, tratando así de imitar el resultado de la asignación aleatoria de tratamiento realizada en los ensayos clínicos aleatorizados.

3. ¿Cuál de los siguientes aspectos es fundamental a la hora de utilizar el *propensity score*?

- a) Las estimaciones deben crearse sin tener acceso a los resultados estudiados.
- b) Hay que ajustar los resultados por edad y sexo.
- c) Revisar la literatura para identificar las potenciales variables de confusión.
- d) Utilizar la prueba de *log-rank*.

La respuesta correcta es la opción *a*.

Un aspecto esencial que hay que tener en cuenta al usar *propensity score* para el diseño de estudios es que las estimaciones deben crearse sin tener acceso a los resultados estudiados, de manera análoga a lo que ocurre en un ensayo clínico aleatorizado.

BMJ, un artículo diferente

Dr. Javier García Alegría
Jefe de Servicio de Medicina Interna
Hospital Costa del Sol (Marbella, Málaga)
Expresidente
Sociedad Española de Medicina Interna
Fundación Española de Medicina Interna

*La publicación seleccionada de este capítulo
es un trabajo real publicado (véase la bibliografía)*



capítulo 10

Publicación seleccionada

Asociación entre los partidos de fútbol de alto perfil en Europa y los accidentes de tráfico en Asia

Introducción

Desde hace tiempo, la revista *BMJ*, en su último número del año y el primero del siguiente, coincidentes con la Navidad y el Año Nuevo, viene incluyendo artículos "diferentes", que tienen como característica especial el humor, la fina ironía o el franco sarcasmo. Durante esta época del año existe un clima general de cordialidad, relajación de costumbres y alegría que los editores quieren reflejar en esta prestigiosa publicación. El objetivo de esta serie, un verdadero divertimento, es aunar el entretenimiento y la enseñanza, reflejar con unos planteamientos desenfadados y, a veces, completamente descabellados que el aprendizaje y la dedicación a la ciencia pueden ser muy divertidos y que cualquier pregunta, por absurda que parezca, puede ser convertida en una cuestión científica. Otro mensaje que subyace en ellos es que debemos estar precavidos ante determinados artículos, pues bajo una apariencia de rigor llegan a conclusiones completamente estrambóticas o carentes de validez, tanto interna como externa, al no haber tenido en cuenta todos los pasos esenciales que debe tener una investigación científica.

Objetivo

Analizar si existe una relación entre los partidos jugados por los equipos de fútbol europeos de alto nivel que son muy populares y los accidentes de tráfico en Asia.

Métodos

Diseño: Estudio basado en la información sobre 41 538 accidentes de taxistas en Singapur y 1 814 320 accidentes de tráfico en Taiwán, y su relación con 12 788 partidos de fútbol europeos durante 7 años.

Participantes: La mayor compañía de taxis de Singapur, con registros pormenorizados de los accidentes de tráfico durante 3 años, y todos los accidentes de tráfico en Taiwán durante 7 años.

Exposición: Días de partidos de equipos europeos de fútbol de alto nivel.

Resultado principal: Número de accidentes de tráfico.

Resultados

La regresión basada en modelos de series temporales sugiere que los días que hay partidos de equipos de fútbol europeos de alto nivel hay más accidentes de tráfico que cuando juegan equipos con menos seguidores. Por cada 134 746 € de incremento medio en el valor de los equipos que juegan entre sí un día determinado, hay un accidente extra entre los taxistas de Singapur, y por cada 7996 € de aumento hay un accidente adicional entre todos los conductores de Taiwán. Esta asociación se mantiene tras controlar por las condiciones climáticas, la demografía de los conductores y las tendencias temporales subyacentes. La asociación es mayor en los accidentes diurnos que en los nocturnos, lo que sugiere que el exceso de accidentes no se puede atribuir a celebraciones o falta de atención durante la conducción nocturna. Anualmente, este incremento podría causar 371 accidentes adicionales entre los taxistas de Singapur y 41 079 accidentes entre todos los conductores de Taiwán, lo que da lugar a unas pérdidas económicas de 821 448 € y 13 994 409 €, respectivamente.

El impacto, tanto en salud como económico, es probablemente mucho más elevado debido a que el huso horario (GMT +8) es la zona más populosa del mundo, pues allí vive el 24 % de la población mundial.

Conclusiones

Los días de partido entre equipos de fútbol de alto nivel en Europa se asocian con más accidentes de tráfico en Singapur y Taiwán que en el caso de los partidos de equipos de menor nivel. Una causa potencial podría ser que los conductores asiáticos pierden horas de sueño viendo estos partidos, que se televisan a media noche en Asia.

Ya se sabía que la privación de sueño aumenta el riesgo de tener un accidente de tráfico. La aportación de este estudio es que ver los partidos de fútbol de los equipos europeos de alto nivel en zonas distantes de la hora local se asocia positivamente con la prevalencia local de accidentes de automóviles. Esto es especialmente problemático en Asia, pues tiene una diferencia horaria aproximada de 8 horas con Europa.

Fundamentos

El fútbol es el deporte mundial que genera mayor pasión y la mayoría de los equipos de alto nivel con más simpatizantes son europeos. Los campeonatos más seguidos en el mundo por los aficionados en las retransmisiones son la Premier League Inglesa, la Liga de España, la Liga de Francia, la Bundesliga y la Serie A de Italia. Los partidos en España o el Reino Unido se suelen jugar a las 8 o 9 de la tarde.

La privación de sueño es una consecuencia de esa afición y de ahí se ha derivado esta hipótesis de trabajo, pues se asocia a desatención, reacción enlentecida y falta de respuesta. Esto podría suceder en las mañanas siguientes a los partidos, con consecuencias sobre la salud y la economía.

También podría haber otras explicaciones plausibles, como las celebraciones o el seguir los partidos durante la conducción, o estar relacionado con otras variables. Para el estudio se ha usado la información de la compañía de taxis más importante de Singapur, de gran calidad, y otra más amplia de Taiwán, que no incluye datos como las características demográficas de los conductores o el tiempo local.

Metodología

La mejor medida de la popularidad de los equipos son los índices de audiencia televisiva. Al no tener disponible este dato en Singapur y Taiwán, como aproximación los autores usaron el presupuesto de cada equipo según Transfer Markt, y para cada partido el promedio entre los dos equipos implicados. Se incluyeron todos los partidos de las ligas mencionadas entre los años 2012 y 2018 y a partir de los octavos de la UEFA Champions League y de la UEFA Europa League ($N = 1379$).

Los modelos primarios analizaron la relación estadística entre el valor de mercado en millones de euros en los partidos de fútbol jugados y el número de accidentes el mismo día. Como estos se emiten en Asia a primeras horas del día, la predicción fue que el número de accidentes sería mayor el mismo día por la falta de horas de sueño, excluyendo las horas previas. Se hicieron tres tipos de análisis para probar la asociación. El primero fue una regresión de Poisson para predecir el número total de accidentes/día, controlando los siguientes factores: días de la semana, mes, aspectos climatológicos, edad y género de los conductores, años de permiso de conducción y porcentaje de automóviles de color amarillo (incrementa el riesgo de accidente). El segundo análisis replicó el anterior, segmentando la noche y el día, para controlar que los accidentes se pudieran deber a la visualización del partido mientras se conducía o a la celebración durante el partido o al finalizar este. En el tercero se utilizaron series temporales que permitieron descartar la posibilidad de que el valor de mercado promedio y el número de accidentes de tráfico estuvieran relacionados debido a una tendencia temporal subyacente.

Estudio de comportamientos

El análisis también se basó en dos supuestos:

1. Que los taxistas son una muestra representativa de los singapurenses como seguidores del fútbol televisado, dado que puede haber un sesgo de género y un menor

nivel educativo que les haga más propensos a ver los partidos; o que tengan turnos largos de trabajo que minimicen la probabilidad de ver partidos.

2. Que el valor de mercado es un buen indicador de la audiencia televisiva de los partidos de fútbol.

Para confirmar que estas premisas eran correctas se hizo un estudio adicional con encuestas acerca del seguimiento de partidos de fútbol europeos a 100 taxistas (99 hombres; edad media de 53 años) y 100 no taxistas (49 hombres; edad media de 34 años).

Resultados

Los resultados de la encuesta de comportamientos ratificaron las suposiciones previas necesarias para probar la hipótesis del estudio:

- Los taxistas de Singapur ven partidos de fútbol a altas horas de la noche y sus hábitos son similares a los de la población general.
- Los partidos entre los equipos europeos de alto nivel despiertan un mayor interés.

El valor de mercado de los equipos y los accidentes de tráfico tuvieron una asociación significativa en Taiwán (estimación: 0,00015; intervalo de confianza al 95% [IC95%]: 0,0001-0,0002; $p < 0,001$; modelo 1 Taiwán) y Singapur (estimación: 0,0002; IC95%: 0,0001-0,0003; $p < 0,001$; modelo Singapur 1). La asociación de Singapur replicó con un tamaño de efecto similar cuando se controlaban las covariables demográficas (modelo Singapur 2).

La relación simultánea entre el valor de mercado promedio y la tasa de accidentes de tráfico fue significativamente positiva tanto en Taiwán ($r = 0,10$) como en Singapur ($r = 0,11$), sin que otras correlaciones alcanzaran la misma magnitud en ninguna de las muestras ($r < 0,10$). Estos datos sugieren que los resultados no se debían a incrementos en la cantidad de tráfico el día previo a los partidos de equipos europeos (**figura 1**).

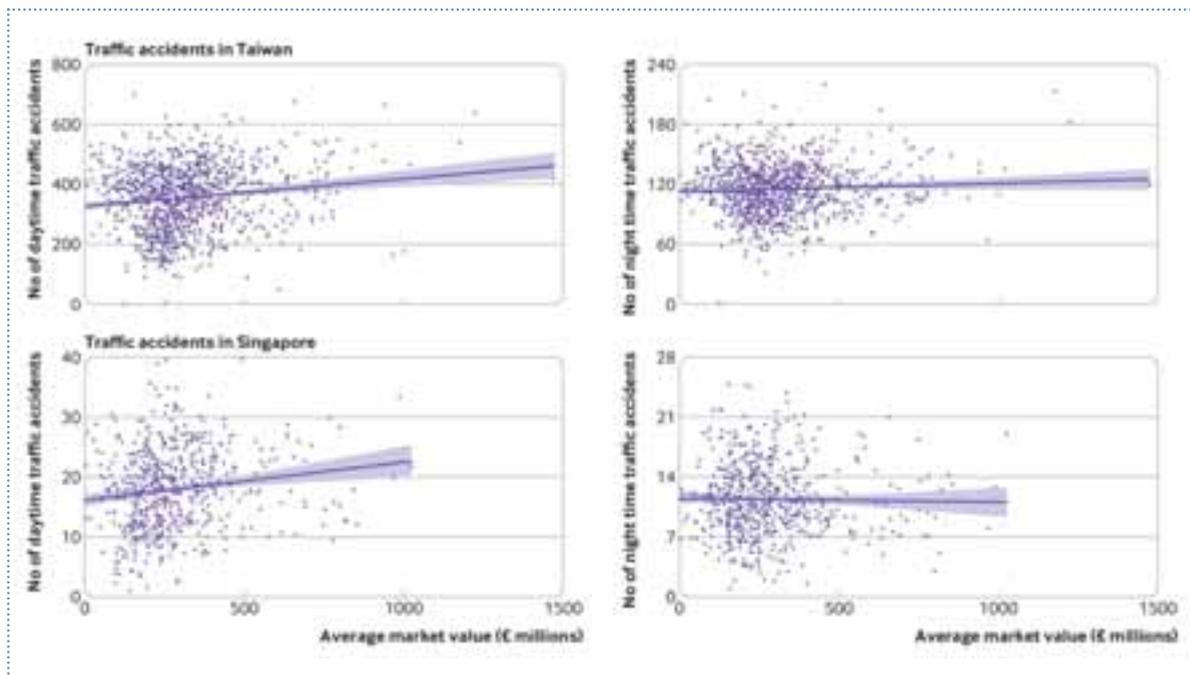
Análisis de series temporales. El resultado de la correlación entre el valor de mercado promedio de los partidos de fútbol y el número de accidentes de tráfico diarios en Taiwán y Singapur está representado en la **figura 2**. Cada barra representa una correlación con un desfase diferente en días (las barras positivas se corresponden con accidentes posteriores a los partidos, mientras que las barras negativas con los previos a los partidos).

Discusión

El análisis de accidentes de tráfico en Taiwán y Singapur apoya la hipótesis de que los días con partidos de fútbol de equipos europeos de alto nivel presentan tasas de accidentes de tráfico superiores a la media en Asia, posiblemente debido a que los conductores permanecen despiertos hasta las primeras horas de la mañana para ver los partidos, aunque el método usado no permite hacer inferencias de causalidad.

La relación se mantenía cuando se controlaba con los datos climáticos, día de la semana, efectos de fin de semana, mes, datos demográficos de los conductores y por las tendencias temporales subyacentes.

Figura 1. Relación entre el valor de mercado promedio de los partidos de fútbol y el número de accidentes de tráfico diurnos y nocturnos en Taiwán y Singapur



Fuente: Yam KC, Jackson JC, Lau J, Qin X, Barnes CM, Chong JK. Association of high profile football matches in Europe with traffic accidents in Asia: archival study. BMJ. 2020 Dec 16;371:m4465.

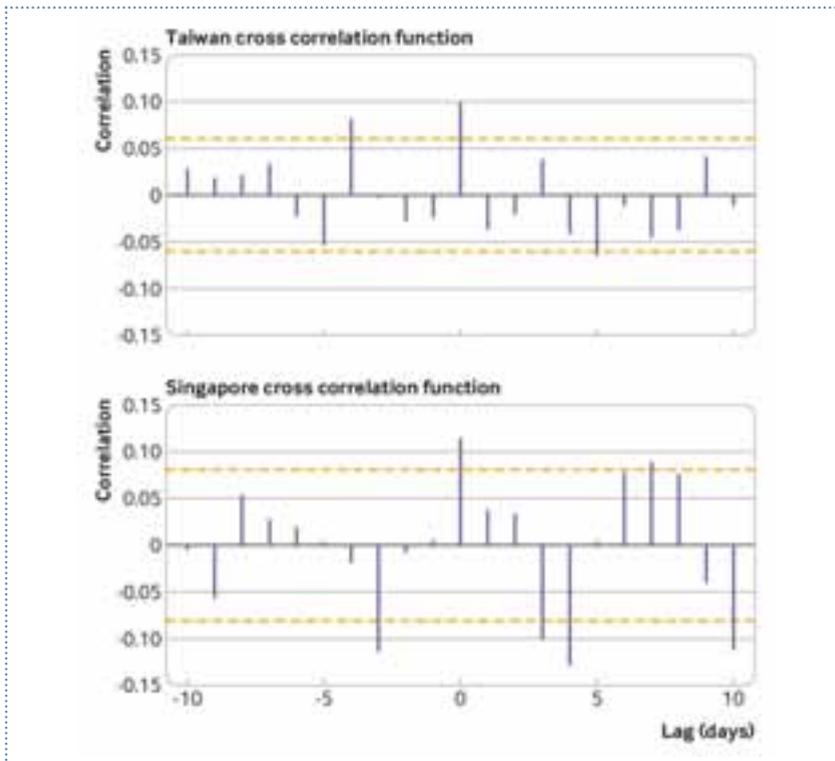
Entre las limitaciones del estudio están las siguientes:

- No se evaluó el impacto en la salud de los accidentes, que puede ser muy variable.
- Aunque la hipótesis de la falta de sueño es plausible, se debería corroborar con encuestas primarias a los accidentados.
- No hubo comparación entre los días que se juegan partidos y los que no.
- En Singapur y Taiwán hay que pagar para poder ver partidos de equipos de fútbol europeos de alto nivel y se desconoce si los conductores tenían esta suscripción, aunque hay numerosas plataformas ilegales muy accesibles.
- No se evaluaron las diferencias entre quienes permanecieron en vigilia y quienes se despertaron para verlo, pues la calidad del sueño puede ser diferente.

Conclusiones e implicaciones

Las estimaciones del modelo permiten calcular el potencial impacto económico entre los conductores de Asia. Los partidos de equipos de fútbol europeos de alto nivel pueden ser responsables de unos 371 accidentes anuales entre los taxistas de Singapur (821 448 €/año)

Figura 2. Correlación entre el valor de mercado promedio de los partidos de fútbol y el número de accidentes de tráfico diarios en Taiwán (arriba) y Singapur (abajo)



Fuente: Yam KC, Jackson JC, Lau J, Qin X, Barnes CM, Chong JK. Association of high profile football matches in Europe with traffic accidents in Asia: archival study. *BMJ*. 2020 Dec 16;371:m4465.

y de 41 079 accidentes entre los taiwaneses (13 994 409 €/año), aunque hay aspectos económicos positivos no analizados (por ejemplo, la facturación de los bares).

La población estudiada de esta zona horaria es el 1,73% del total, pues allí viven más de 1700 millones de personas (24% de la población mundial), por lo que el impacto real es mucho mayor. Esto podría tener implicaciones organizativas para las competiciones más seguidas (por ejemplo, programar los partidos entre los equipos de fútbol europeos de alto nivel en sábados o domingos) o prever más controles de tráfico.

La generación de la hipótesis

Hay numerosas anécdotas de cómo distintos científicos han producido una idea que ha resultado completamente revolucionaria. Las circunstancias concurrentes para que del ingenio surja un planteamiento realmente original, muchas veces, están incardinadas en la vida rutinaria, en la observación de la naturaleza con una mirada distinta, en una conversación al vuelo, en un comentario ajeno, en un hecho aparentemente irrelevante, o simplemente

durante la transición crepuscular entre el sueño y la vigilia. Por tanto, hay que estar atento al mundo que nos rodea, tener una mirada escrutadora y fomentar la curiosidad científica individual, pues, como decía Louis Pasteur, la suerte solo favorece a la mente preparada.

En el artículo comentado la idea surgió cuando el autor principal (Kai Chi Yam), un profesor de la Universidad de Singapur, se encontraba en Barcelona invitado para dar una conferencia el 7 de mayo de 2019. Ese día tuvo la oportunidad de ver en un bar de la ciudad un partido entre el Liverpool FC y el FC Barcelona, semifinal de la UEFA Champions League, rodeado de numerosos aficionados. Según describe él mismo, el ambiente general inicial era apasionado y festivo, pues el Barcelona había ganado el partido de ida por 3-0. Tras los 90 minutos, el Liverpool le había dado la vuelta a la eliminatoria al marcar cuatro goles, y en la final se alzaría con este campeonato, tras 13 años de sequía, con su victoria sobre el Tottenham. Al terminar el partido, sus amigos desde el otro confín del mundo comenzaron a mandarle mensajes para comentarlo. Se quedó sorprendido de que la mayoría de ellos estuviera despierto a tan altas horas de la madrugada, pues allí eran entre las 3 y las 4 de la mañana.

En ese momento formuló la hipótesis de trabajo: como casi todos los partidos de fútbol de alto nivel se juegan en Europa a última hora de la tarde, los aficionados asiáticos deben sufrir privación del sueño para poder verlos y esto podría tener consecuencias perjudiciales. La privación del sueño incrementa el riesgo de tener un accidente de tráfico, y se conocen diversas causas de alteración del sueño, algunas de ellas relacionadas con los estilos de vida actual, como ver la televisión durante las horas nocturnas.

Todos los trabajos científicos empiezan con una buena pregunta de investigación (hipótesis). Para hacer una pregunta correcta en ciencia se recomienda seguir los pasos del acrónimo FINER, que indican cómo ha de ser esa pregunta: **f**actible, **i**nteressante, **n**ovedosa, **é**tica y **r**elevante; junto con los siguientes elementos para redactarla o analizar la literatura científica (estrategia PICOT): **p**acientes/grupo de estudio, **i**ntervención, **c**omparación, **o**utcomes (resultados) y **t**empo.

En lo que respecta a los estudios observacionales, el consenso STROBE ha definido una serie de recomendaciones generales para mejorar el diseño y la evaluación de los estudios observacionales descriptivos y para aquellos que investigan la asociación de una o varias variables de exposición y los resultados en salud. Los puntos clave comienzan con un título y un resumen informativos, una explicación del contexto, los fundamentos y unos objetivos bien definidos. Una descripción meticulosa del método: participantes con los criterios de elegibilidad y de apareamiento, las fuentes de obtención de los datos, las variables de respuesta o exposición, predictoras o modificadoras del efecto, y los métodos de evaluación, así como las medidas adoptadas para controlar los sesgos, los criterios para determinar el tamaño muestral, el tratamiento estadístico (métodos, análisis de subgrupos, el tratamiento de datos ausentes y el análisis de sensibilidad). En los resultados hay que describir el número de participantes y el número de eventos, las estimaciones no ajustadas y, en su caso, ajustadas por factores de confusión y su precisión (intervalos de confianza), y, en caso necesario, las estimaciones de riesgo relativo y otros análisis (subgrupos, interacciones, sensibilidad). En la discusión hay que resumir los principales resultados, describir las limitaciones y las posibles fuentes de sesgos, ofrecer una interpretación prudente de los resultados, compararlos con estudios similares y ver la posibilidad de generalizar los resultados (validez externa). Por último, hay que especificar la fuente de financiación y los posibles conflictos de interés. El artículo seleccionado es un excelente ejemplo de rigor y de cumplimiento de las reglas STROBE y, por tanto, puede servir como un aprendizaje entretenido.

Este artículo ahonda en el conocimiento de los efectos perjudiciales para la salud de ver los partidos de fútbol por televisión de los equipos favoritos. Así, en un metaanálisis, Lin *et al.* demostraron que ver partidos de fútbol de la selección nacional del propio país en el Campeonato del Mundo se asoció con un mayor riesgo de enfermedad cardiovascular mortal tanto en hombres (riesgo relativo [RR]: 1,13; IC95%: 1,004-1,28) como en mujeres (RR: 1,08; IC95%: 1,01-1,15). El análisis de subgrupos mostró que el fracaso del equipo propio tiene un mayor riesgo de enfermedad cardiovascular mortal en general (RR: 1,29; IC95%: 1,15-1,45), aunque se observó un menor riesgo de enfermedad cardiovascular mortal cuando el equipo al que seguían los espectadores obtuvo una victoria (RR: 0,80; IC95%: 0,66-0,96). La conclusión de los autores es que ver partidos de fútbol se asoció con un mayor riesgo de enfermedad cardiovascular mortal y no mortal, especialmente en los varones, y cuando pierde el equipo propio. La explicación más plausible es que se debe a una súbita elevación del cortisol como respuesta al estrés.

Aunque los estudios de exposición a un riesgo, como el presentado, tan solo permiten generar hipótesis, esta debería confirmarse con estudios aleatorizados posteriores (por ejemplo, tres grupos de aleatorización de aficionados de un equipo determinado, con tres exposiciones distintas: ver el partido en directo, no ver el partido y ver el partido en diferido conociendo un resultado favorable). Dada la amplia afición mundial por este deporte, y la más que probable relación coste-beneficio, las autoridades sanitarias podrían plantearse imponer anuncios a las cadenas de televisión, similares a los que llevan las cajetillas de tabaco, con mensajes como “Ver un partido de fútbol de tu equipo favorito aumenta los riesgos para tu salud”, “Si la selección pierde en este partido estarás en mayor riesgo de sufrir un infarto de miocardio” o, en el caso los aficionados con una gran diferencia horaria que ven partidos durante la noche, “Ver partidos que alteran tu ritmo del sueño incrementa el riesgo de sufrir un accidente de tráfico”. Entre unas cosas y otras, los forofos del fútbol vamos a tener que buscar otras alternativas de menor riesgo. ¡Un poco de humor no está reñido con la ciencia!

Bibliografía

Lin LL, Gu HY, Yao YY, Zhu J, Niu YM, Luo J, Zhang C. The association between watching football matches and the risk of cardiovascular events: A meta-analysis. *J Sports Sci.* 2019 Dec;37(24):2826-34.

Newson M, Shiramizu V, Buhrmester M, et al. Devoted fans release more cortisol when watching live soccer matches. *Stress Health.* 2020 Apr; 36(2):220-7.

Vandenbroucke JP, von Elm E, Altman DG, et al.; en nombre de la Iniciativa STROBE. Mejorar la comunicación de estudios observacionales en epidemiología (STROBE): explicación y elaboración. *Gac Sanit [Internet].* 2009 [consultado el 6 de junio de 2021]; 23(2):158.e1-28. Disponible en: <https://scielo.isciii.es/pdf/gsv23n2/especial2.pdf>.

Yam KC. Does football viewership lead to unexpected traffic accidents in Asia? 17 de diciembre de 2020 [consultado el 6 de junio de 2021]. En: *BJM Opinion [Internet].* Londres: BMJ Publishing Group; c2021. Disponible en: <https://blogs.bmj.com/bmj/2020/12/17/does-football-viewership-lead-to-unexpected-traffic-accidents-in-asia>.

Yam KC, Jackson JC, Lau J, Qin X, Barnes CM, Chong JK. Association of high profile football matches in Europe with traffic accidents in Asia: archival study. *BMJ [Internet].* 2020 Dec 16 [consultado el 6 de junio de 2021];371:m4465. doi: 10.1136/bmj.m4465. Disponible en: <https://www.bmj.com/content/bmj/371/bmj.m4465.full.pdf>.

Preguntas de autoevaluación

1. Una de las siguientes afirmaciones es incorrecta acerca del riesgo asociado a ver un partido de fútbol en Asia de los equipos europeos de alto nivel que se juegan a las 8 o 9 de la tarde (hora europea):

- a) Se asocian a un mayor riesgo de accidentes de tráfico en la mañana siguiente.
- b) Se asocian a un mayor riesgo de accidentes de tráfico durante la noche.
- c) No hay mayor riesgo de accidentes de tráfico.
- d) La afición por el fútbol en Asia es baja y no ven los partidos europeos.

La respuesta correcta es la opción a.

Los accidentes se han relacionado con privación del sueño en la mañana siguiente a la celebración del partido, pues hay unas 8-9 horas de diferencia con Europa. No hay diferencias durante la noche y, por tanto, no se pueden atribuir a celebraciones.

2. En relación con el riesgo asociado a ver partidos por televisión de la propia selección en el Campeonato Mundial de Fútbol, señale cuál de las siguientes afirmaciones es correcta:

- a) Es mayor en Brasil y Argentina.
- b) Es mejor quedar eliminados en las primeras fases, pues así se reduce la exposición al riesgo (número de partidos).
- c) Aumenta la mortalidad cardiovascular, sobre todo cuando pierde el partido la selección nacional propia.
- d) Cuando la selección nacional propia gana el partido también se incrementa la mortalidad por el estrés, aunque es menor que cuando se pierde.

La respuesta correcta es la opción c.

Las dos primeras respuestas no se han comprobado, pues no hay estudios comparativos entre esos dos países y otros distintos, y por otra parte no se ha estudiado el impacto en el nivel de estrés y las consecuencias de quedar eliminados en las primeras fases en comparación con llegar a la final del campeonato. En el metaanálisis de Lin *et al.* (véase la bibliografía del capítulo) se demuestra que cuando la selección propia gana el partido se reduce la mortalidad cardiovascular, y cuando pierde se incrementa la mortalidad cardiovascular, de manera especial entre los varones.

3. ¿Cuál de las siguientes características de una pregunta de investigación está contemplada en las recomendaciones del acrónimo FINER para formular preguntas de investigación?

- a) Que sea entendible.
- b) Que sea financiable.
- c) Que sea ética.
- d) Que sea rigurosa.

La respuesta correcta es la opción c.

Las siglas FINER corresponden a factible, interesante, novedosa, ética y relevante. Puede obtenerse información adicional en <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6691636>.

4. En la estrategia PICOT para escribir preguntas de investigación o para evaluar la calidad de un artículo no figura uno de los siguientes elementos, ¿cuál es?

- a) Comparación.
- b) Pacientes.
- c) Intervención.
- d) Organización.

La respuesta correcta es la opción d.

La o del acrónimo corresponde a *outcomes* (resultados). El acrónimo PICOT se refiere a pacientes, intervención, comparación, *outcomes* (resultados) y tiempo. Puede obtenerse información adicional acerca de la estrategia PICOT en <http://www.scielo.org.co/pdf/recis/v15n3/1692-7273-recis-15-03-00309.pdf>.

Prevalencia de COVID-19 y tres series: datos del Ministerio de Sanidad en la epidemia

Dr. Manuel Crespo Blanco
Servicio de Medicina Interna
Hospital Universitario de Bellvitge (Barcelona)

*La publicación seleccionada de este capítulo
es un trabajo real publicado (véase la bibliografía)*



capítulo 11

Publicación seleccionada

Estudio de seroprevalencia de la infección por SARS-CoV-2 en España

Antecedentes

Durante las fases iniciales de la pandemia por COVID-19, la escasez de test de diagnóstico microbiológico y la sobresaturación de los servicios sanitarios condicionaron el conocimiento del grado de circulación del virus SARS-CoV-2 en la comunidad. En este contexto, el estudio serológico basado en la detección de anticuerpos frente al SARS-CoV-2 proporciona una información mucho más precisa de la extensión y la evolución de la epidemia.

Metodología

Diseño y población del estudio: ENE-COVID es un estudio epidemiológico poblacional, longitudinal, cuyos objetivos principales son estimar la prevalencia de infección por SARS-CoV-2 en España y evaluar la evolución temporal a lo largo de tres oleadas separadas en el tiempo entre abril y junio de 2020.

Debido a la distribución territorial heterogénea del virus y la posible asociación con el tamaño de la población, se selecciona la muestra entre la población no institucionalizada siguiendo un sistema de muestreo estratificado por provincias y, dentro de cada provincia, por el tamaño poblacional de cada municipio. En cada estrato provincia-tamaño municipal, se seleccionan aleatoriamente las unidades censales, y dentro de estas los domicilios. Todos los sujetos residentes en los domicilios seleccionados son invitados a participar. El estudio, que incluye una encuesta epidemiológica con antecedentes y síntomas relacionados con COVID-19, cuenta con la aprobación del Comité de Ética de la Investigación del Instituto de Salud Carlos III.

Tamaño de la muestra y análisis estadístico: Se necesita un mínimo de 900 participantes por provincia para estimar una prevalencia del 5% con una precisión de $\pm 2,5\%$ (error relativo, 50%) con un nivel de confianza del 95% y una participación de dos tercios. Se asume un efecto diseño de 2 (factor de inflación de la varianza por la correlación entre personas de una misma unidad censal o unidad familiar). En cada comunidad autónoma es necesario incluir un mínimo de 1980 sujetos para estimar una prevalencia del 5% con una precisión de $\pm 1,65\%$ (error relativo, 33%) asumiendo el resto de las premisas. Globalmente, se estima que es necesario invitar a 90 000 personas para obtener una muestra de 60 000 participantes (dos tercios de los sujetos invitados).

De acuerdo con los datos del Instituto Nacional de Estadística, el número medio de residentes en cada hogar español es 2,5, lo que determina una muestra media de

60 sujetos por unidad censal, compuesta por 24 domicilios. Se necesitan 1500 unidades censales para alcanzar los 90 000 participantes requeridos.

La asignación territorial de la muestra se realiza de modo que se incluye en cada provincia el número mínimo de participantes necesario para estimar la seroprevalencia con el menor nivel de precisión establecido. Este sistema de asignación, que comporta el sobremuestreo de las provincias menos pobladas, permite disminuir la variancia y aumentar la eficiencia de las múltiples estimaciones de seroprevalencia realizadas para las provincias.

Las estimaciones de seroprevalencia tienen en cuenta el efecto de la estratificación por provincia y tamaño municipal, y la agrupación por domicilios y secciones censales. Adicionalmente, se realiza un ajuste según el sexo, el grupo de edad y el nivel de renta relativo de la sección censal.

Análisis serológico: Se utilizan dos test diferentes para medir la presencia de anticuerpos IgG frente al SARS-CoV-2:

- Un test rápido de inmunocromatografía (Orient Gene Biotech COVID-19 IgG/IgM), que permite obtener resultados *in situ* y no requiere venopunción.
- Un inmunoensayo quimioluminiscente de micropartículas (ARCHITECT, de Abbott), con mayor precisión diagnóstica, aunque necesita venopunción, lo que puede condicionar una disminución de la tasa de participación.

En un análisis realizado con sueros disponibles de la primera ronda (30 % del total de participantes), la concordancia entre ambos test fue del 97 %. Debido a que no se dispone de los datos completos del inmunoensayo de las rondas 2 y 3, y a la menor sensibilidad y especificidad de la IgM del test rápido, los resultados presentados en este informe se refieren a la IgG del test rápido.

Resultados

Las tres rondas de ENE-COVID incluyen 68 296 participantes, el 69,2 % de los individuos elegibles; el 80,3 % han participado en las tres rondas.

La prevalencia estimada de anticuerpos IgG frente al SARS-CoV-2 en España ha sido del 5 % (intervalo de confianza al 95 [IC95%]: 4,7-5,4) en la primera ronda, de 5,2 % (IC95%: 4,9-5,5) en la segunda ronda y de 5,2 % (IC95%: 4,9-5,5) en la tercera ronda. En todas ellas, la prevalencia es similar en hombres y mujeres. Geográficamente, se observa un patrón mantenido a lo largo del estudio: prevalencias cercanas o superiores al 10 % en el entorno de Madrid, e inferiores al 5 % en las provincias de la costa –excepto Barcelona–, Andalucía y las islas.

La prevalencia es algo mayor en áreas urbanas (municipios de más de 100 000 habitantes). No se aprecian diferencias importantes por sexo, nacionalidad, renta ni número

de convivientes. Los niños muestran prevalencias más bajas en la primera ronda, menos evidentes en la última.

La prevalencia aumenta a medida que aumenta el número de síntomas, y es superior al 40 % en personas que refieren anosmia o ageusia. El diagnóstico previo (PCR positiva) se traduce en una seroprevalencia cercana o superior al 75 %. Las personas que han tenido contacto con un caso confirmado muestran prevalencias más altas, similares o superiores al 10 %, que alcanzan el 27-31 % en los convivientes con casos COVID-19.

La incidencia de seroconversión ha sido del 0,9 % entre la ronda 1 y la ronda 2 (IC95%: 0,8-1,0) y de 0,7 % entre la ronda 2 y la ronda 3 (IC95%: 0,6-0,8). Aparte de los antecedentes personales (clínica y diagnóstico de COVID-19), convivir con un paciente o un caso sospechoso es un factor de riesgo para la seroconversión.

Conclusiones

La prevalencia de anticuerpos frente al SARS-Cov-2 en España durante el periodo de estudio se sitúa en torno al 5 %, con una marcada variabilidad geográfica y una tasa de seroconversión inferior al 1 %. Se aprecia un incremento en la prevalencia de anticuerpos a medida que aumenta el número de síntomas. El diagnóstico previo de COVID-19 (PCR positiva) se traduce en cifras de seroprevalencia cercanas o superiores al 75 %. La seroprevalencia aumenta en personas con contacto previo con COVID-19 y se sitúa en torno al 30 % entre los convivientes con un caso confirmado.

Planteamiento

España ha sido uno de los países con mayor grado de afectación por la pandemia de COVID-19. Durante la primera oleada causó un grave problema de salud pública, que obligó a la adopción de medidas drásticas de confinamiento domiciliario. La saturación de los servicios asistenciales y la escasez de pruebas de diagnóstico virológico limitaron el conocimiento del impacto y la evolución de la epidemia a nivel poblacional. ENE-COVID es un estudio seroepidemiológico realizado en la población española entre abril y junio de 2020 cuyos objetivos principales son estimar la prevalencia y la evolución de la infección por SARS-CoV-2 durante la primera oleada de la pandemia mediante el análisis de anticuerpos frente al SARS-CoV-2.

El planteamiento del estudio se orienta asimismo a conocer los datos de prevalencia y la evolución temporal atendiendo a la diversidad territorial, las características sociales y demográficas y la información epidemiológica y clínica.

Formulación de la pregunta clínica

La pregunta del estudio es conocer el grado de afectación y la evolución de la pandemia por COVID-19 en la población española durante la primera oleada de la pandemia, mediante la determinación de anticuerpos frente al virus (seroprevalencia).

Durante el periodo de estudio, la ausencia de métodos de diagnóstico virológico y la dificultad para acceder a los servicios asistenciales dificultaron el diagnóstico microbiológico de la infección. Además, en un elevado porcentaje de pacientes, la infección puede cursar de forma oligosintomática o sin síntomas. Estos factores limitan el conocimiento del grado de extensión y la evolución de la epidemia, que se basa en el diagnóstico de los casos sintomáticos.

Los estudios de seroprevalencia permiten diagnosticar a los pacientes sintomáticos no diagnosticados en la fase aguda de la infección por la dificultad de acceso al sistema sanitario o por limitación de métodos de diagnóstico, y permiten detectar casos asintomáticos, descritos en el 20-35% de los pacientes infectados por SARS-CoV-2.

Justificación del diseño y el análisis estadístico realizado

Justificación del método de selección de la muestra. Se selecciona la muestra del estudio mediante un método de estratificación en dos etapas, de acuerdo con las recomendaciones de la Organización Mundial de la Salud. El primer nivel de estratificación está constituido por las 50 provincias y las ciudades autónomas de Ceuta y Melilla. El segundo nivel de estratificación tiene en cuenta el tamaño poblacional de los municipios dentro de cada provincia. Este método de selección aumenta la representatividad territorial de la muestra analizada (tiene en cuenta la distribución territorial heterogénea del virus) y permite estimar la variabilidad de la prevalencia en función del área geográfica y la densidad de población.

Dentro de cada estrato definido por la provincia y el tamaño municipal se eligen aleatoriamente las secciones censales (unidades de muestreo de primera etapa). Posteriormente, a partir del registro municipal de cada provincia (actualizado en enero de 2020), se seleccionan de forma aleatoria los hogares. Se invita a participar a todos los convivientes en las viviendas

seleccionadas, lo que permite analizar la asociación entre seroprevalencia y exposición a casos confirmados entre convivientes.

Se estima el tamaño de la muestra asumiendo una prevalencia del 5 % con una precisión de $\pm 2,5$ % (error relativo, 50 %) para la estimación provincial y de $\pm 1,65$ % (error relativo, 33 %) para la estimación de prevalencia en las comunidades autónomas. En ambos casos, el intervalo de confianza es del 95 % (indica el grado de confianza que se tendrá de que el verdadero valor de prevalencia en la población de referencia se encuentre dentro del intervalo obtenido en la muestra analizada).

El cálculo del tamaño de la muestra tiene en cuenta asimismo el efecto diseño del estudio y el muestreo agrupado por unidades censales y hogares: se produce un aumento de la variancia en la prevalencia estimada asociada con la agrupación de casos positivos entre los sujetos de un mismo hogar o de la misma sección censal.

La asignación territorial de la muestra se realiza de modo que se incluye en cada provincia el número mínimo de participantes necesario para estimar la seroprevalencia con el menor nivel de precisión establecido. La probabilidad de participación está inversamente relacionada con el tamaño de la población, lo que condiciona un sobremuestreo de las provincias menos pobladas. Este sistema de asignación permite disminuir la variancia y aumentar la eficiencia de las múltiples estimaciones de seroprevalencia realizadas para las provincias.

Por otra parte, el diseño longitudinal del estudio permite analizar la evolución de la pandemia, condicionada por factores virológicos (capacidad de transmisión), epidemiológicos y relacionados con las medidas de control de la infección (no farmacológicas) adoptadas por las autoridades sanitarias (por ejemplo, la limitación de la movilidad o el confinamiento domiciliario). La determinación seriada de anticuerpos en las tres rondas secuenciales del estudio (27 de abril a 11 de mayo; 18 de mayo a 1 de junio; 8 a 22 de junio) corrige el retraso temporal entre la fase aguda de la infección y la detección de los anticuerpos, y permite conocer la tasa de seroconversión (cambios de la prevalencia) y la extensión de la epidemia a nivel poblacional.

Justificación metodológica. Para el análisis serológico se utilizan dos test diagnósticos complementarios: por un lado, un test de quimioluminiscencia de alta fiabilidad diagnóstica, pero que requiere extracción de sangre y un proceso analítico más complejo, realizado habitualmente en los laboratorios de microbiología; y, por otro lado, un test rápido que permite el análisis en el punto de atención, que es más económico y fácilmente escalable en estudios epidemiológicos poblacionales. Debido a que la sensibilidad y la especificidad de los test diagnósticos rápidos empleados en las fase inicial de la pandemia eran bajos o desconocidos, se realizó un estudio comparativo entre ambas pruebas diagnósticas y se observó una concordancia diagnóstica del 97 %.

El estudio incluye una encuesta epidemiológica con información de antecedentes y síntomas relacionados con la COVID-19, lo que permite analizar las características clínico-epidemiológicas de la pandemia y analizar el riesgo de infección en los contactos y en los convivientes de un caso confirmado.

Con respecto al análisis estadístico, el diseño del estudio, estratificado por etapas, hace el análisis más complejo. Las estimaciones de seroprevalencia se realizan teniendo en cuenta la estratificación de la muestra por provincia y tamaño municipal, y la agrupación por secciones censales y domicilios. Adicionalmente, se realiza una ponderación de la prevalencia estimada según el sexo, el grupo de edad y el nivel de renta.

Fortalezas y debilidades del estudio

Entre las fortalezas del estudio destacan las siguientes:

- El rigor metodológico empleado en la selección de la muestra estudiada es un soporte sólido de su representatividad con respecto a la población de referencia (provincias y comunidades autónomas de España). En la medida en que la muestra seleccionada es representativa de la población diana, se garantiza la validez de la prevalencia estimada y que el resultado obtenido refleje de la forma más precisa posible la situación real de la pandemia. La representatividad provincial de la muestra analizada permite estimar la prevalencia en cada comunidad autónoma y en cada provincia y conocer el grado de extensión de la pandemia en los diferentes territorios.
- El diseño longitudinal del estudio permite analizar la evolución temporal de la pandemia durante el periodo estudiado.
- Las altas tasas de participación y de adherencia (el 80 % de las personas incluidas en el estudio han participado en las tres rondas) reducen el riesgo del sesgo de participación y permiten mantener la potencia estadística en los análisis realizados. Esto permite estimar con confianza la seroprevalencia ponderada por rangos de edad, sexo, estudios o nivel de renta.
- La combinación de dos test de determinación de anticuerpos con características complementarias (test rápido, realizado en el punto de atención, fácilmente escalable; y test de quimioluminiscencia, más preciso, aunque requiere venopunción y presenta mayor complejidad técnica) proporciona una mayor fiabilidad a los resultados y permite estimar una franja de prevalencia con valores que dependen de si realizamos una valoración más específica (ambas pruebas son positivas) o más sensible (alguna de las dos pruebas realizadas resulta positiva).
- La recogida de información epidemiológica adicional permite conocer las principales características clínicas y epidemiológicas de la pandemia en España y el riesgo de contagio asociado con la exposición a un caso confirmado.

Entre las limitaciones cabe destacar las que siguen:

- Los resultados del estudio son representativos en relación con la población española no institucionalizada. Las personas mayores y las personas dependientes que viven en residencias o en otro tipo de instituciones no están suficientemente representadas en este estudio. Tampoco están bien caracterizados en el estudio los trabajadores sanitarios o de centros sociosanitarios ni otros colectivos de trabajadores esenciales (profesionales de la educación, conductores de transporte público, fuerzas de seguridad, etc.). Aunque la muestra estudiada incluye a personas mayores, sanitarios y trabaja-

dores de centros sociosanitarios y de colectivos esenciales, el número de sujetos analizados pertenecientes a cada uno de estos grupos resulta insuficiente para una caracterización adecuada.

- En términos estadísticos, un número elevado de personas elegibles (más del 31 %) no participaron en el estudio. No disponemos de información relacionada con las características de estas personas (si eran similares o no a la población estudiada) y las potenciales implicaciones de su no participación en los resultados del estudio.
- Limitaciones de la fiabilidad diagnóstica de los test empleados. Es posible que en un futuro próximo dispongamos de técnicas analíticas más precisas con las que se pueda reevaluar la seroprevalencia y obtener resultados diferentes.
- Aunque se asume que los pacientes en los que no se detectan anticuerpos no han tenido exposición previa al virus, la metodología diagnóstica empleada puede haber sido insuficiente: es posible que existan casos de infección en los que los anticuerpos no estén presentes o lo estén con niveles bajos, no detectables, lo que supondría una infraestimación de la circulación real del SARS-CoV-2. Además, para valorar la inmunidad adquirida en la población expuesta es necesario conocer mejor la respuesta inmunitaria al SARS-CoV-2 y estudiar la inmunidad celular, que podría desempeñar un papel muy relevante.
- La información epidemiológica fue autorreportada por los sujetos incluidos en el estudio, por lo que existe un riesgo potencial de sesgo de recuerdo y de interpretación en las respuestas recogidas.
- Aunque las tasas de participación han sido muy altas y todas las estimaciones están ponderadas por la participación según la edad, el sexo y la renta relativa, no se ha podido realizar un análisis preciso de la prevalencia ponderada por otras variables potencialmente relevantes (por ejemplo, la ocupación laboral, la comorbilidad, etc.).
- No se ha analizado el potencial impacto de la heterogeneidad de los profesionales (sanitarios, administrativos, informáticos) encargados de realizar el trabajo de campo y de la recogida de la información clínico-epidemiológica.

Aplicabilidad en la práctica clínica

Los resultados de este estudio son fácilmente generalizables al estar obtenidos en una muestra representativa de la población diana a la que se dirige la inferencia.

El conocimiento de la extensión y la evolución de la circulación del virus en tiempo real puede ser útil para la adopción de medidas de salud pública (por ejemplo, la limitación de la movilidad o el confinamiento domiciliario) necesarias para controlar la pandemia.

Finalmente, el conocimiento de los datos desagregados por comunidades autónomas, provincia y municipio nos permite comparar los resultados obtenidos en territorios con prevalencia similar y puede ser útil para detectar déficits asistenciales y la necesidad de implementar planes de mejora.

Bibliografía

Instituto Nacional de Estadística. 2019 Continuous Household Survey [Nota de prensa]. Madrid: INE; 2020. Disponible en: https://www.ine.es/en/prensa/ech_2019_en.pdf.

Ministerio de Sanidad. Estudio ENE-COVID: Informe final. Estudio nacional de sero-epidemiología de la infección por SARS-CoV-2 en España. Madrid: Ministerio de Sanidad; 2020. Disponible en: <https://www.msbs.gob.es/gabinetePrensa/notaPrensa/pdf/INFOR060720134446500.pdf>.

Pollán M, Pérez-Gómez B, Pastor-Barriuso R, et al. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *Lancet*. 2020;396(10250):535-44. DOI: 10.1016/S0140-6736(20)31483-5. Disponible en: [http://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(20\)31483-5/fulltext](http://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)31483-5/fulltext).

Yanes-Lane M, Winters N, Fregonese F, et al. Proportion of asymptomatic infection among COVID-19 positive persons and their transmission potential: A systematic review and meta-analysis. *PLoS One*. 2020 Nov 3;15(11):e0241536. DOI: 10.1371/journal.pone.0241536.

Preguntas de autoevaluación

1. ¿Cuál de las siguientes no es una característica del estudio ENE-COVID?

- a) Es un estudio de seroprevalencia poblacional, con carácter longitudinal.
- b) Para el cálculo del tamaño de muestra se define la prevalencia estimada (5%), el grado de precisión mínimo (2,5%) y el nivel de confianza (95%).
- c) La selección de la muestra del estudio se realiza mediante un método de estratificación de la población diana en dos etapas.
- d) Se realiza una encuesta epidemiológica para conocer la asociación entre la vacunación previa y el riesgo de infección.

La respuesta correcta es la opción *d*.

Para el cálculo del tamaño de la muestra se tiene en cuenta la prevalencia estimada (5%) con un grado de precisión mínimo del $\pm 2,5\%$ (error relativo, 50%) para las provincias y de $\pm 1,65\%$ (error relativo, 33%) para el caso de las comunidades autónomas; en ambos casos, con un nivel de confianza del 95%. El tamaño de muestra se ajusta por la participación esperada (dos tercios de las personas invitadas) y por el efecto diseño del estudio (corrige la potencial agrupación de casos en unidades censales y domicilios).

Se selecciona la muestra del estudio mediante un método de estratificación en dos etapas. El primer nivel de estratificación está constituido por las provincias, y el segundo nivel de estratificación tiene en cuenta el tamaño poblacional de los municipios dentro de cada provincia. Este método de selección aumenta la representatividad territorial de la muestra analizada y permite estimar la variabilidad de la prevalencia en función del área geográfica y de la densidad de población.

El estudio incluye una encuesta epidemiológica con información de los antecedentes y síntomas relacionados con la COVID-19, lo que permite analizar las características clínico-epidemiológicas de la pandemia y analizar el riesgo de infección en los contactos y en los convivientes de un caso confirmado. No se analiza la asociación entre la vacunación previa (todavía no disponíamos de vacunas en la época en la que se realizó el estudio) y el riesgo de infección.

2. ¿Cuál de las siguientes no es una fortaleza del estudio ENE-COVID?

- a) La muestra analizada es representativa de la población general.
- b) La representatividad de la muestra permite estimar con precisión la prevalencia en las comunidades autónomas y en las provincias.

- c) La representatividad de la muestra permite estimar con confianza la prevalencia en el colectivo de personas institucionalizadas y en trabajadores de centros sociosanitarios.
- d) El alto nivel de participación permite estimar con confianza la prevalencia ponderada por rangos de edad y sexo.

La respuesta correcta es la opción *c*.

La muestra analizada se selecciona a partir de la población no institucionalizada. Aunque en el estudio se incluye a personas mayores y a trabajadores sanitarios, sociosanitarios y trabajadores esenciales, el número es limitado y no permite estimar con confianza la prevalencia en estos colectivos. Todas las demás son fortalezas del estudio.

3. ¿Cuál de las siguientes informaciones no es un resultado del estudio ENE-COVID?

- a) Marcada variabilidad geográfica en la seroprevalencia, inferior al 5% en la mayoría de las provincias costeras y superior al 10% en centro peninsular.
- b) Tasa de seroconversión a nivel nacional en torno al 5%.
- c) Elevada prevalencia en las personas con síntomas compatibles con COVID-19.
- d) Prevalencia estimada alrededor del 30% entre los convivientes de un caso confirmado.

La respuesta correcta es la opción *b*.

Entre los resultados del estudio ENE-COVID destaca un patrón mantenido a lo largo del estudio: prevalencias cercanas o superiores al 10% en el entorno de Madrid, e inferiores al 5% en las provincias de la costa –excepto Barcelona–, Andalucía y las islas. La prevalencia también es algo mayor en las áreas urbanas (municipios de más de 100 000 habitantes).

Se observó un aumento de la prevalencia a medida que aumenta el número de síntomas, especialmente en personas con anosmia o ageusia (>40%). El diagnóstico previo (PCR positiva) se traduce en una seroprevalencia cercana o superior al 75%. Igualmente, las prevalencias estimadas fueron más altas (similares o superiores al 10%) en las personas que referían contacto previo con un caso confirmado, y alcanzaban el 27-31% en los convivientes con casos de COVID-19.

Por el contrario, la tasa global de seroconversión (anticuerpos IgG detectables en personas seronegativas en la ronda previa) fue inferior al 1%: del 0,9% (IC95%: 0,8-1,0) entre la ronda 1 y la ronda 2, y del 0,7% (IC95%: 0,6-0,8) entre la ronda 2 a la ronda 3.

Patrocinado por:

