



Pontificia Universidad
JAVERIANA
Bogotá

MAESTRÍA EN 
EPIDEMIOLOGÍA
CLÍNICA

DISEÑOS EN EPIDEMIOLOGÍA

MÓDULO III

Tema 4

Estrategias para su manejo

Licencia: Creative Commons: Atribución/Reconocimiento 4.0 Internacional

Mapa de sesgos y fundamentos de QBA

Para ver estructuras que el ojo humano no puede ver, ¿qué instrumento se usa? ¡Una lupa! Así son **los estudios epidemiológicos, son como lupas con las que se mira la realidad** y funcionan de la misma manera: usted diseña un estudio epidemiológico para ver un fenómeno de interés (por ejemplo, la posible relación entre una causa y un desenlace de interés) de manera que le permita ver ese fenómeno que está ocurriendo y que usted no puede ver simplemente observando a las personas.

Hacer uso de esta lupa **implica medir** los elementos del fenómeno de **interés** (la causa, el desenlace), **seleccionar a un grupo de personas** en quienes pueda medir estos elementos (su población de estudio) y **observarlos** de manera tal que usted pueda hacerse una idea de la manera como se **relacionan entre sí** (donde, entre otros, es crítico asegurarse de que la causa estaba antes de que el desenlace de interés ocurriera).

Sin embargo, como todos sabemos, la vida real es compleja y, con frecuencia, es difícil de capturar con este o cualquier otro instrumento. La distancia que hay entre lo que está pasando en la vida real y lo que se puede ver en un estudio epidemiológico es producto de **amenazas a la validez**.



¡Tome nota!

¿Qué son las amenazas a la validez?

Las amenazas a la validez implican que **usted llegue a conclusiones erróneas** sobre la relación que espera observar. Es decir, que **sus hallazgos no sean válidos**.

Las amenazas a la validez son sesgos. Los sesgos se hacen evidentes porque introducen una distorsión en la información obtenida en el estudio, de manera que la medida de asociación que se observa no corresponde a la relación que existe en la vida real.

Esta falta de correspondencia puede presentarse de diversas maneras: la medida de asociación puede sobrestimar la relación que se está explorando, puede subestimarla, puede sugerir que existe una relación que en realidad no existe o puede sugerir que no existe una relación que en realidad sí existe.

En la práctica, no tenemos una forma de validar en qué dirección y/o magnitud está distorsionada la medida de asociación observada, pues no tenemos una manera de medir la realidad tal y como es, sino a través de los estudios epidemiológicos.

La aproximación experimental nos aproxima, pero incluso la experimentación implica un contexto controlado que puede distanciarnos de lo que pasa en la vida real. Por lo tanto, **solo podemos reflexionar sobre el potencial impacto del sesgo en dicha medida de asociación**, teniendo en cuenta el diseño planteado y las limitaciones conocidas.

En general, las **fuentes de sesgo** que nos preocupan al reflexionar sobre las amenazas a la validez son de tres tipos: **de información**, en donde el error se introduce en la medición de las variables de interés; **de confusión**, donde la relación observada entre la exposición y el desenlace puede ser explicada por otras variables confusoras; y el **de selección**, donde el error se introduce por la forma en que ingresa o permanece la población en el estudio.

Cuando usted está planteando su estudio, o después, al interpretar sus resultados, las **preguntas** que se le pide que se haga son:

- 1.** ¿Qué posibles errores puedo estar introduciendo en mi estudio?
- 2.** ¿Cómo puedo minimizarlos?
- 3.** ¿Qué impacto tienen en la medida de asociación que estoy calculando?

Para responder las dos primeras preguntas, usted tiene diversas opciones a partir del conocimiento que tenga tanto del fenómeno en estudio como de la epidemiología y la bioestadística. Estos elementos le permitirán identificar las amenazas a la validez (es decir, la distancia que hay entre el fenómeno que usted cree que está ocurriendo en la vida real vs. el que usted puede observar con su estudio) e implementar estrategias en el diseño o en el análisis para reducir el impacto de estas fuentes de error.

Por ejemplo:

Para reducir el riesgo de	Algunos aspectos que tiene que establecer en el diseño del estudio	Algunas preguntas útiles
Error de medición	Definir claramente cuáles son los fenómenos que quiere medir.	¿Cuál es el fenómeno? ¿Qué definiciones existen? ¿Cuál es la definición que quiero usar para mi estudio?
	Establecer cómo se miden los eventos de interés en relación con su pregunta de investigación.	¿Qué estrategia de medición me permite observar el evento de interés tal como ha sido definido? ¿Qué estrategias puedo implementar para que la medición sea válida y precisa?
	Determinar la calidad y estandarización de los instrumentos o fuentes de datos.	¿He evaluado la sensibilidad, especificidad o confiabilidad de los instrumentos? ¿Cómo garantizo uniformidad entre observadores o sitios?

Para reducir el riesgo de	Algunos aspectos que tiene que establecer en el diseño del estudio	Algunas preguntas útiles
Confusión	Establecer claramente cuál es la pregunta de investigación y la relación causal que se quiere establecer.	¿Cuál es la exposición? ¿Cuál es el desenlace que estaría causado por dicha exposición? ¿Cómo se definen tanto exposición como desenlace?
	Identificar posibles variables confusoras a partir de la literatura, la teoría causal o diagramas de DAG.	¿Qué factores están asociados simultáneamente con la exposición y el desenlace, pero no son parte de la cadena causal?
	Planear estrategias para medir y controlar los confusores (aleatorización, restricción, emparejamiento, ajuste).	¿Tengo mediciones adecuadas de los confusores? ¿Qué estrategia de control es factible en mi contexto?
Sesgo de selección	Definir cuidadosamente los criterios de inclusión y exclusión de la población objetivo.	¿A quién quiero observar realmente? ¿Quién queda fuera por mis criterios y qué implicaciones tiene eso?
	Anticipar posibles patrones de no participación, pérdida o exclusión durante el seguimiento.	¿Qué tipo de personas es menos probable que entre o permanezca en el estudio? ¿Eso depende de la exposición o del desenlace?
	Diseñar estrategias para minimizar pérdidas y asegurar representatividad.	¿Puedo ofrecer incentivos, seguimiento activo o múltiples fuentes de contacto para reducir pérdidas? ¿Cómo documentaré quién entra y quién se pierde?

Sin embargo, usted ya sabe que, en la práctica, no es factible diseñar el estudio perfecto y siempre habrá la necesidad de reconocer que el estudio tiene limitaciones por diversas razones que serán explicadas mediante ejemplos.

Ejemplos

Ejemplo 1. Depresión

Usted sabe que el estándar de oro para **medir la presencia de depresión** es la entrevista por un psiquiatra entrenado, pero con esta estrategia se demoraría demasiado tiempo para poder establecer la prevalencia de depresión **entre 5 000 personas**. Usted podría recurrir a un cuestionario de tamizaje, reconociendo que la clasificación que le proporciona la herramienta de tamizaje le llevará a una proporción de falsos positivos o negativos relacionada con el rendimiento operativo de la herramienta.



Ejemplo 2. Consumo de café y enfermedades cardiovasculares



Supongamos que usted está evaluando la **asociación entre el consumo de café y la enfermedad cardiovascular usando una base de datos que ya existe**. Usted sabe que el tabaquismo es una fuente de confusión clave, pero en la base de datos se midió tabaquismo únicamente como “tabaquismo actual” (opciones de respuesta: sí/no), sin información sobre hábitos previos, como si la persona fue fumadora en el pasado, pero ya no lo es. Usted tiene que pensar si necesita descartar estos datos y empezar de cero a recolectar la información o, por el contrario, si usa la base existente reconociendo que habrá confusión residual.

Ejemplo 3. Violencia doméstica y depresión

Imagine que está evaluando la **asociación** entre la **exposición a violencia doméstica y la depresión en mujeres** que asisten a un centro de salud. Muchas de las mujeres con mayor exposición y síntomas más severos podrían decidir no asistir a la consulta o abandonar el estudio por miedo o desconfianza. En cambio, las mujeres con síntomas leves o sin violencia pueden permanecer. El resultado es que la muestra final no representa bien a la población objetivo: el grupo expuesto queda subrepresentado en su forma más grave. Por lo tanto, si usted encuentra que no hay relación entre violencia y depresión, deberá considerar si esto es consecuencia de un sesgo de selección inherente a la estrategia de reclutamiento de la población de su estudio.



La reflexión sobre estos errores nos lleva a **responder la tercera pregunta**: ¿Qué impacto tienen en la medida de asociación que estoy calculando?

Es importante que tenga claro que reflexionar sobre estas fuentes de error no se trata de negociar la validez de su estudio, sino de entender cómo su medida de asociación puede estar distorsionada por estos errores. **Por lo tanto, en la medida de lo posible, usted y su equipo de investigación se deben anticipar a estas fuentes de error en el diseño del estudio y tomar decisiones para llegar al diseño de estudio más robusto y factible posible.** Por ejemplo, esa tensión entre robustez y factibilidad es lo que le va a llevar a decidir si usa el tamizaje para depresión o reduce el tamaño de muestra para asegurar la clasificación correcta de la depresión medida por un psiquiatra entrenado. No olvide que no se trata de hacer el estudio por hacerlo, sino de responder una pregunta de investigación lo más cerca a la realidad posible.

Sin embargo, digamos que, por razones logísticas, usted toma la decisión de hacer el tamizaje para depresión, pero sabe que tiene error de clasificación, ¿cómo hacer para saber qué tan distante está la medida de asociación que usted está viendo respecto a la que está ocurriendo en la vida real?

Cuantificación del sesgo: la base del análisis cuantitativo de sesgo

Supongamos que usted y su equipo de investigación ya han reflexionado cuidadosamente sobre las posibles fuentes de sesgo de su estudio y han tomado decisiones intentando minimizar su impacto.

Por razones logísticas, deciden usar una prueba de tamizaje en lugar del estándar de oro para clasificar la depresión. Saben que esta decisión introduce error de clasificación, pero buscan responder una pregunta crucial: **¿qué tan distante está la medida de asociación que observamos de la verdadera medida que ocurre en la vida real?**

El enfoque tradicional: el análisis cualitativo del sesgo

Tradicionalmente, lo que se ha hecho es hacer un análisis cualitativo del impacto del error que el investigador sabe que está ocurriendo. Usted identifica la fuente del error, reflexiona sobre el impacto en la clasificación de sus pacientes y pondera en qué dirección este error puede distorsionar su medida de asociación. Y, con frecuencia, esta reflexión se aborda en la discusión de su trabajo, mencionando la limitación que ha detectado y el potencial significado de esta limitación.

Ejemplo

Usted puede concluir que una limitación del estudio es que se usó una herramienta de tamizaje en vez de estándar de oro para clasificar la depresión. ¿Qué significa esto?

- **Que usted tiene algunos falsos positivos** (probablemente más falsos positivos que negativos, dado que como todos sabemos, las pruebas de tamizaje tienden a optimizar descartar a las personas que definitivamente no tienen el evento de interés, con la esperanza de que una segunda prueba clasifique mejor a los falsos positivos).
- Y que, en consecuencia, **probablemente la relación** que usted busca **esté subestimada** debido a que su población con el evento puede estar sobreestimada con personas que son falsos positivos.



Al final, usted incluirá en su artículo: “una limitación del estudio es el uso de un instrumento de tamizaje en lugar del estándar de oro, lo que podría haber generado falsos positivos y, en consecuencia, una subestimación de la asociación observada”.

Esta conclusión usted no solo la conoce como una heurística del error de clasificación (está en todos los libros) sino que puede verificarlo manualmente en una tabla de 2x2 y por el conocimiento del fenómeno.

El problema de no cuantificar

Sin embargo, la verdad es que la pregunta de qué tanto se desvía la medida de asociación que usted ve con respecto a la vida real no ha sido respondida porque **la cuantificación del sesgo no es una práctica usual**.

En la **práctica**, esto se convierte en un problema de la evidencia científica por diversas razones. Por ejemplo, el cuerpo de la evidencia aporta unas medidas de asociación que sabemos que están distorsionadas con respecto a la vida real, pero no podríamos saber a ciencia cierta cuánto. Para la **toma de decisiones**, esto es problemático, porque no se sabe cuál es la magnitud de la relación entre exposiciones y desenlaces y no podemos saber bien qué tan fuerte es la relación entre una exposición y un desenlace que se está explorando. Por esto, **han ganado valor diversas estrategias para cuantificar el nivel del sesgo** y, de esta manera, aproximarnos al verdadero valor de la medida de asociación de interés.

De la descripción a la cuantificación

Por eso han cobrado fuerza los métodos de **análisis cuantitativo del sesgo (Quantitative Bias Analysis, QBA)**.

Su propósito es **estimar la dirección y magnitud de la distorsión introducida por diferentes tipos de sesgo y, así, aproximarse al valor verdadero de la medida de asociación**.

En otras palabras,

Si sabemos que hay sesgo, el QBA nos permite ponerle un número.

Estos métodos permiten responder preguntas como:

- ¿En qué medida mi medida de asociación se aleja del valor real?
- ¿Qué tan fuerte tendría que ser el sesgo para anular mi hallazgo?
- ¿Cómo cambiaría mi conclusión si corrijo la estimación según parámetros plausibles?

Ejemplo práctico 1

Cuantificar el impacto del error de clasificación

Imaginemos un estudio que busca estimar la asociación entre exposición a violencia doméstica (sí/no) y presencia de depresión (sí/no).

Debido a limitaciones logísticas, el equipo utiliza una prueba de tamizaje para identificar casos de depresión en lugar de un diagnóstico clínico formal.

Sabemos que la prueba tiene una sensibilidad del 80 % y una especificidad del 90 %, lo que implica cierto nivel de error de clasificación.

La tabla observada (con error)

	Depresión "según tamizaje" (+)	Depresión "según tamizaje" (-)	Total
Expuestos a violencia	80	120	200
No expuestos	40	160	200

A partir de esta tabla:

Riesgo en expuestos: $80 / 200 = 0,40$

Riesgo en no expuestos: $40 / 200 = 0,20$

Riesgo Relativo (RR) observado = $0,40 / 0,20 = 2,0$

A simple vista, parece que la exposición a violencia duplica el riesgo de depresión, pero este valor está calculado con un instrumento imperfecto.

Reconstruyamos lo que ocurre en la vida real.

Sabemos que:

Sensibilidad (Se) = 0,80 → el 80 % de los verdaderos casos se detectan correctamente.

Especificidad (Sp) = 0,90 → el 90 % de los no casos son clasificados correctamente.

Usando las fórmulas inversas de corrección:

$$T+=O+-(1-Sp)*N \quad Se+Sp-1; \quad T-=N-T+T+=O+-(1-Sp)*N \quad Se+Sp-1; \quad T-=N-T+$$

Aplicando esto:

Expuestos:

$$T+=80-0,1*200,7=85,7 \quad T+=80-0,1*200,7=85,7$$

$$T-=200-85,7=114,3 \quad T-=200-85,7=114,3$$

No expuestos:

$$T+=40-0,1*200,7=28,6 \quad T+=40-0,1*200,7=28,6$$

$$T-=200-28,6=171,4 \quad T-=200-28,6=171,4$$

Tabla corregida (verdadera)

	Depresión "real" (+)	Depresión "real" (-)	Total
Expuestos a violencia	86	114	200
No expuestos	29	171	200

Riesgos corregidos

Riesgo en expuestos: $86 / 200 = 0,43$

Riesgo en no expuestos: $29 / 200 = 0,145$

RR corregido $\approx 3,0$

Interpretación

El uso del tamizaje subestimó la fuerza de la asociación:

RR observado = 2,0

RR verdadero \approx 3,0

Esto ocurre porque el error de clasificación (falsos positivos y negativos) **diluye la diferencia** entre los grupos.

En otras palabras, la lupa de su estudio tenía una pequeña distorsión: los “bordes” de la depresión estaban borrosos y eso aplanó la relación real.

Lección: en la práctica, solemos reconocer este tipo de error en la discusión (“probablemente subestimamos la asociación...”). Con el análisis cuantitativo de sesgo es posible medir esa desviación y mostrar qué tan lejos podríamos estar del valor real.

Este ejercicio manual es el punto de partida para los métodos sistemáticos:

- Determinísticos (corregir con Se y Sp conocidas).
- Probabilísticos (cuando Se y Sp son inciertas, se usan distribuciones para simular escenarios).

Conclusión

- El QBA no elimina el sesgo, pero nos muestra su tamaño. Es pasar del “sabemos que hay error” al “sabemos cuánto importa ese error”.

En los siguientes apartados, veremos las estrategias disponibles para cuantificar el sesgo de acuerdo con el tipo de sesgo que se identifica:

Confusión no medida: el sesgo

¿Qué es la confusión no medida?

Cuando hacemos un estudio epidemiológico, queremos observar la relación entre una exposición y un desenlace y concluir algo sobre la realidad.

Pero la vida es más compleja de lo que cualquier base de datos puede capturar. Siempre habrá variables que no conocemos o que, aunque sepamos que existen, no logramos medir.

Esas variables invisibles pueden alterar la relación que creemos ver. A eso lo llamamos **confusión no medida**.

La confusión ocurre cuando una tercera variable está asociada tanto con la exposición como con el desenlace y no forma parte de la cadena causal.

Cuando esa variable no se mide o no se ajusta adecuadamente, hablamos de confusión no medida.

Ejemplo clásico

Se observa que el consumo de café se asocia con mayor riesgo de cáncer de páncreas.

Sin embargo, los fumadores tienden a tomar más café y también a tener más cáncer de páncreas.

El tabaquismo es la variable confusora.

La confusión no medida ocurre cuando una tercera variable está asociada tanto con la exposición como con el desenlace, pero no forma parte de la cadena causal.

Es decir: no es un paso intermedio sino un factor externo que interfiere.

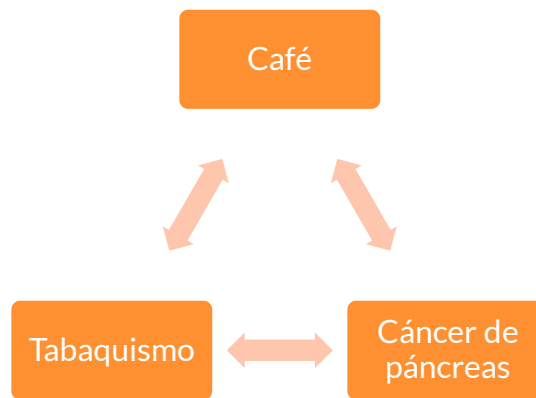
Por ejemplo, encontramos que quienes toman café parecen tener más riesgo de cáncer de páncreas.

Pero los fumadores tienden a tomar más café y a tener más cáncer.

En este caso, el cigarrillo es la variable confusora.

El estudio observa café y cáncer, pero la verdadera relación es entre cigarrillo y cáncer.

Lo que vemos no es falso, pero está distorsionado.



Cómo distorsiona la medida de asociación

La confusión no medida puede:

- **Inflar** una asociación (si el confusor está positivamente asociado con ambos factores).
- **Atenuarla o revertirla** (si actúa en direcciones opuestas).

La magnitud del sesgo depende de:

- Qué tan fuerte es la asociación entre la exposición y el confusor.
- Qué tan fuerte es la asociación entre el confusor y el desenlace.
- Qué tan prevalente es el confusor.

La confusión no medida puede inflar, atenuar o incluso invertir una asociación.

La magnitud del sesgo depende de tres cosas:

- Qué tan fuerte es la relación entre la exposición y el confusor.
- Qué tan fuerte es la relación entre el confusor y el desenlace.
- Qué tan común es el confusor en la población.

Por eso, cada estudio está expuesto a cierta distorsión y reconocerla aporta rigor.

Entonces, la **pregunta clave** es:

¿Cuánto tendría que influir esa variable no medida para explicar completamente mi hallazgo?

El **análisis cuantitativo** de sesgo nos permite responder exactamente.

El *E-value*, una forma simple de cuantificar

El *E-value* es un método propuesto por VanderWeele y Ding (2017) que responde a una pregunta intuitiva:

¿Qué tan fuerte tendría que ser la asociación entre el confusor no medido y tanto la exposición como el desenlace para explicar completamente la asociación observada?

El *E-value* es una forma práctica de medir la robustez frente a confusión no medida.

Lo propusieron VanderWeele y Ding en 2017 y se interpreta de manera muy intuitiva.

Imaginemos que encontramos que las personas físicamente activas tienen un riesgo relativo de 0,60 para depresión:

- El *E-value* para un RR de 0,60 es aproximadamente **2,5**.
- Eso significa que un confusor no medido tendría que estar asociado **al menos 2,5 veces** tanto con la actividad física como con la depresión para anular por completo el hallazgo.
- Si ningún confusor plausible tiene relaciones tan fuertes, nuestro resultado se considera robusto.
- En cambio, si un confusor conocido podría tener esa magnitud, nuestra evidencia es más débil.

Tipos de análisis de robustez

Otro enfoque similar es el análisis de **tipping point**: en lugar de asumir un solo valor, explora qué combinaciones de fuerza entre exposición–confusor y confusor–desenlace cambiarían la conclusión.

Es un mapa de escenarios posibles.

Esta estrategia es muy valiosa, pues en vez de intentar eliminar la incertidumbre, la hace visible.

Tipo	Qué pregunta responde	Herramienta
<i>E-value</i>	¿Qué fuerza de confusión anularía mi hallazgo?	Calculadora web o <i>E-value package</i>
<i>Tipping-point / tipr</i>	¿Qué combinación de relaciones (exposición–confusor, confusor–desenlace) invertiría mi conclusión?	tipr en R
<i>Omitted variable bias / sensemakr</i>	¿Cuánto varía el coeficiente si incorporo un confusor con fuerza determinada?	sensemakr en R

Interpretación práctica

Un *E-value* alto sugiere **robustez** del resultado frente a confusión no medida.

Un *E-value* bajo sugiere **vulnerabilidad** del hallazgo.

Ningún método “corrige” mágicamente el sesgo; su valor está en **cuantificar la incertidumbre**.

Ejemplo interpretativo para artículo

“El riesgo relativo ajustado para depresión fue de 0,60 (IC95 %: 0,45–0,80). El *E-value* correspondiente fue de 2,5, lo que indica que se requeriría un confusor no medido asociado al menos 2,5 veces tanto con la actividad física como con la depresión para explicar completamente el hallazgo.

No identificamos en la literatura variables con asociaciones de esa magnitud, por lo que consideramos que la confusión no medida es poco probable como explicación alternativa.”

Lecturas y herramientas recomendadas

- Mathur, M. B., Ding, P., Riddell, C. A., & VanderWeele, T. J. (2018). E-values: What they are, and why they matter. *BMJ*.
- VanderWeele, T. J., & Ding, P. (2017). Sensitivity analysis in observational research: Introducing the E-value. *Annals of Internal Medicine*, 167(4), 268–274. <https://doi.org/10.7326/M16-2607>

Packages:

- `EValue` (R) → cálculo directo.
- `tipr` → análisis de robustez flexible.
- `sensemkr` → omitted variable bias (para modelos lineales o logísticos).

EJEMPLO PRÁCTICO EN R – Confusión no medida (QBA)

0) Materiales

Datos: puede usar su propio estudio o el *dataset* de ejemplo incluido abajo.

Software: R (con los paquetes `EValue` y `tipr`).

Tiempo estimado: 45–60 minutos.

1) Caso guía

Pregunta: ¿la actividad física regular se asocia con menor depresión?

Medida observada: riesgo relativo (RR) = 0,60 (IC95 %: 0,45–0,80).

Supuesto: puede existir un confusor no medido (p. ej., nivel socioeconómico) asociado con la exposición y con el desenlace.

2) Paso a paso – E-value

¿Qué responde?: ¿qué tan fuerte tendría que ser la asociación de un confusor no medido con la exposición y con el desenlace para explicar completamente el efecto observado?

Instrucciones

Abra R y ejecute:

```
install.packages("EValue")  
library(EValue)  
# Reemplace RR, lo y hi por los de su estudio si aplica  
evalue(RR = 0.60, lo = 0.45, hi = 0.80, true = 1)
```

Anote el “Point estimate E-value” y el “Lower CI limit E-value”.

Interprete

Si el E-value es alto (p. ej. >2), se requeriría un confusor muy fuerte para anular el hallazgo → mayor robustez.

Si es bajo ($\approx 1-1,5$), el hallazgo es vulnerable a confusión plausible.

Plantilla de interpretación (copiar/pegar y completar)

El RR observado fue de __ (IC95%: -). El E-value para la estimación puntual fue __ y para el límite inferior del IC95% fue __. Esto indica que se requeriría un confusor no medido asociado al menos __ veces tanto con la exposición como con el desenlace para explicar completamente el hallazgo. Considerando la literatura disponible, __.

3) Paso a paso – Tipping point con tipr

¿Qué responde?: ¿qué combinaciones de fuerza (exposición–confusor; confusor–desenlace) cambiarían mi conclusión?”

Instrucciones

En R:

```
install.packages("tipr")  
library(tipr)  
# Use su RR e IC95% para estimar el error estándar (SE)  
est <- log(0.60)  
se <- (log(0.80) - log(0.45)) / (2 * 1.96)  
tp <- tip_analyze(est = est, se = se)  
plot(tp) # curva/superficie de tipping point  
summary(tp)
```

Observe el gráfico: identifique pares de fuerzas del confusor (RR_UC y RR_CD) que moverían el estimador a la nulidad o perderían significancia.

Anote un par de combinaciones “umbral” y contraste con la plausibilidad biológica/epidemiológica.

Plantilla de interpretación

El análisis de tipping point mostró que combinaciones alrededor de RR(Exposición–Confusor)=__ y RR(Confusor–Desenlace)=__ serían necesarias para invalidar el hallazgo. No identificamos confusores plausibles con asociaciones de esa magnitud; por tanto, la conclusión es relativamente robusta. Si existieran variables con fuerzas similares, la conclusión sería sensible a confusión no medida.

4) (Opcional) sensemakr – Sesgo por variable omitida (modelos lineales/logísticos)

¿Cuándo usarlo? Cuando el efecto proviene de un modelo de regresión y desea analizar la robustez del coeficiente frente a la omisión de una covariable.

Pseudopasos

Ajuste su modelo (OLS o logístico).

Estime la fuerza que tendrían las relaciones del confusor con exposición y desenlace (R^2 parciales).

Use `sensemkr` para obtener curvas de robustez y benchmarks usando covariables medidas.

5) Reporte mínimo para manuscrito

Incluya tabla y figura + un párrafo con:

- Estimación principal (con IC95 %, si es pertinente).
- Método QBA usado (E-value, tipping point, etc.).
- Valores obtenidos y su interpretación.
- Fuentes para juzgar la plausibilidad de confusores de esa magnitud (literatura/juicio experto).
- Declaración honesta de límites.

Ejemplo de párrafo

El RR ajustado fue 0,60 (IC95 %: 0,45-0,80). El E-value para la estimación puntual fue 2,48 (límite inferior: 1,81), indicando que se requeriría un confusor no medido asociado $\geq 2,5$ veces tanto con la exposición como con el desenlace para explicar completamente el hallazgo. El análisis de tipping point mostró que combinaciones de asociaciones de magnitud similar serían necesarias para invalidar la conclusión. Dado que no identificamos en la literatura confusores con esa fuerza, consideramos que la confusión no medida es poco probable como explicación alternativa; sin embargo, no puede descartarse por completo.

6) Dataset de ejemplo (si desea simular rápidamente)

Suponga 1 000 participantes. Depresión (sí/no) como desenlace y actividad física (sí/no) como exposición. Tasa de depresión en no activos: 20 %; RR observado: 0,60.

Generación rápida en R (opcional)

```
set.seed(123)
n <- 1000
p_exp <- 0.5
p_out_unexp <- 0.20
RR <- 0.60

exp <- rbinom(n, 1, p_exp)
prob <- ifelse(exp == 1, p_out_unexp*RR, p_out_unexp)
out <- rbinom(n, 1, prob)

# RR crudo
risk_exp <- mean(out[exp==1])
risk_un <- mean(out[exp==0])
RR_obs <- risk_exp / risk_un
RR_obs
```

Use los resultados obtenidos (RR e IC95%) para alimentar E-value y tipr.

Error de clasificación

En epidemiología, medir es observar la realidad a través de un instrumento. Pero ningún instrumento es perfecto. Cuando lo que observamos no corresponde exactamente con lo que ocurre en la vida real, hablamos de **clasificación errónea** o *misclassification*.

Es el sesgo que aparece cuando la forma de medir (la lupa) está un poco distorsionada.

La **clasificación errónea** ocurre cuando la variable de exposición o desenlace se mide incorrectamente.

- Si el error ocurre **por igual** en todos los grupos → **no diferencial**.
- Si el error **depende** de la exposición o del desenlace → **diferencial**.
- En ambos casos, puede sesgar el efecto estimado, pero **en direcciones distintas**.

En un estudio, podemos clasificar mal la **exposición**, el **desenlace**, o ambas. Esto puede deberse a cuestionarios, registros administrativos, pruebas diagnósticas o incluso a interpretaciones humanas.

Cuando la clasificación se equivoca, movemos personas entre casillas, y con ello, **distorsionamos la medida de asociación**.

No toda clasificación errónea produce sesgo:

Lo hace cuando el error **no ocurre al azar**, sino de manera **sistemática**.

Ejemplo simple

Supongamos que queremos estudiar la relación entre **consumo de alcohol** y **hipertensión**. Si usamos un cuestionario de autorreporte, algunas personas pueden **subestimar** su consumo.

- Si tanto hipertensos como no hipertensos subestiman igual → **no diferencial** → tiende a **diluir** el efecto ($RR \rightarrow 1$).
- Si los hipertensos tienden a subreportar más (por culpa o consejo médico) → **diferencial** → el sesgo puede ir **en cualquier dirección**.

Efecto sobre la validez

Cuando el error ocurre **por igual** en los grupos comparados —por ejemplo, si todos subestiman el consumo de alcohol por igual— hablamos de **clasificación no diferencial**.

En general, este tipo de error **diluye** el efecto: acerca el RR a 1.

Pero si el error **depende** de la exposición o del desenlace —por ejemplo, si los hipertensos reportan menos alcohol porque se lo prohibió su médico— hablamos de **clasificación diferencial**. En este caso, el sesgo puede **inflar o revertir** la asociación.

El error de clasificación es una forma de error de información. Se convierte en una fuente de sesgo cuando distorsiona la medida de asociación.

Tipo	Qué causa	Efecto típico
No diferencial	Error independiente de la otra variable.	Atenúa (lleva el RR hacia 1).
Diferencial	Error depende de la exposición o del desenlace.	Puede inflar o atenuar el efecto, impredeciblemente.

Piense en una balanza que pesa a todos con la ropa puesta: eso es error **sistemático** pero **no diferencial**.

Ahora imagine que la balanza de los hombres está en el gimnasio y la de las mujeres en el baño con el piso irregular: eso ya es **diferencial**.

En la investigación clínica, pasa lo mismo:

- El error igual para todos cambia la precisión, pero no tanto la dirección.
- El error que depende del grupo puede inventar relaciones o esconder las reales.

Cuando conocemos o podemos suponer la **sensibilidad (Se)** y **especificidad (Sp)** de nuestra medición, podemos **ajustar numéricamente** la tabla 2x2 y estimar un nuevo riesgo relativo o odds ratio.

Ejemplo

- Si la prueba para enfermedad respiratoria tiene $Se = 80\%$ y $Sp = 90\%$, podemos corregir la tabla observada y ver cómo cambia el RR.
- Con frecuencia, la medida corregida es **más extrema**, porque el error de clasificación tiende a diluir las diferencias.

Ejemplo numérico

Queremos estimar la asociación entre exposición a un contaminante y enfermedad respiratoria.

a) Datos observados (con error de clasificación en el desenlace).

	Enfermedad (+)	Enfermedad (-)	Total
Expuestos	50	150	200
No expuestos	20	180	200

RR observado = $(50/200) / (20/200) = 2.5$

Pero la prueba diagnóstica para la enfermedad tiene:

Sensibilidad = 80 %.

Especificidad = 90 %.

b) Corrigiendo la tabla

```
install.packages("episensr")
library(episensr)
misclassification(matrix(c(50,150,20,180), nrow=2, byrow=TRUE),
  type = "outcome", bias_parms = c(0.80, 0.90, 0.80, 0.90))
```

► RR corregido ≈ 3.0

Conclusión: el error de clasificación **dilúía** la fuerza de la asociación.

- Si no tenemos certeza del valor exacto de la sensibilidad o especificidad, podemos asignarles **distribuciones de probabilidad** (por ejemplo, Beta) y simular miles de escenarios. Así obtenemos un rango plausible de la verdadera medida de asociación.
- Este enfoque se llama **análisis probabilístico del sesgo de clasificación**, y permite mostrar cuánta incertidumbre hay... pero ahora medida con números.

Visualización conceptual

- Cada medida que usamos (cuestionario, test, variable administrativa) es una **ventana imperfecta** a la realidad. Mientras menor sea la sensibilidad o especificidad, más “borrosa” es la imagen. El QBA nos permite “ajustar el enfoque”.

Extensión probabilística

- Cuando no conocemos con exactitud la Se y Sp, podemos asignarles **distribuciones** (por ejemplo, Beta) y simular miles de escenarios para obtener un **rango plausible del RR verdadero**.

```
probsens(matrix(c(50,150,20,180), nrow=2, byrow=TRUE),  
          type = "outcome",  
          reps = 10000,  
          se = list("beta", 80, 20),  
          sp = list("beta", 90, 10))
```

El resultado: un intervalo de RR corregido con incertidumbre explícita.

Reporte para manuscrito

“Considerando la posible clasificación errónea del desenlace, asumimos una sensibilidad de 80 % y una especificidad de 90 %, valores comunes para el diagnóstico basado en historia clínica. Tras la corrección, el RR cambió de 2,5 a 3,0, indicando una subestimación del efecto en el análisis original.

Esta magnitud sugiere que el error de clasificación no altera la interpretación global, pero sí reduce la precisión de la estimación”.

GUÍA PRÁCTICA – SESGO POR CLASIFICACIÓN ERRÓNEA

1) Caso guía

Pregunta: ¿existe asociación entre exposición a un contaminante y enfermedad respiratoria?

Tabla observada:

	Enfermedad (+)	Enfermedad (-)	Total
Expuestos	50	150	200
No expuestos	20	180	200

RR observado = $(50/200) / (20/200) = 2.5$

Prueba diagnóstica (para el desenlace):

Sensibilidad = 0,80

Especificidad = 0,90

2) Corrección determinística

¿Qué responde?: ¿cómo cambia la estimación si conozco la sensibilidad y especificidad del instrumento?

Comando en R

```
install.packages("episensr")
library(episensr)
misclassification(matrix(c(50,150,20,180), nrow=2, byrow=TRUE),
  type = "outcome", bias_parms = c(0.80, 0.90, 0.80, 0.90))
```

Salida esperada:

Observed RR = 2,5

Corrected RR ≈ 3,0

Interpretación:

La corrección indica que el error de clasificación diluía la asociación.
El RR verdadero podría ser 3,0, no 2,5.

3) Análisis probabilístico (incertidumbre en Se/Sp)

¿Qué responde?: ¿cómo varía la estimación si Se y Sp no son valores fijos, sino inciertos?

Si no se conocen con certeza Se y Sp:

```
probsens(matrix(c(50,150,20,180), nrow=2, byrow=TRUE),  
          type = "outcome",  
          reps = 10000,  
          se = list("beta", 80, 20),  
          sp = list("beta", 90, 10))
```

El resultado entregará una distribución de posibles RR corregidos (media, mediana e IC simulados).

Interpretación

El RR corregido varió entre 2,8 y 3,3 en 95 % de las simulaciones, lo que confirma una ligera subestimación del efecto en el análisis original.

4) Plantilla de reporte para manuscrito

Asumiendo una sensibilidad del 80 % y una especificidad del 90 % para la clasificación del desenlace, el RR corregido fue de 3,0 frente al 2,5 observado.

Esto sugiere una leve subestimación de la asociación debido al error de clasificación no diferencial. En simulaciones probabilísticas, el RR osciló entre 2,8 y 3,3, lo que indica estabilidad del hallazgo frente a incertidumbre moderada en los parámetros del test."

Recursos complementarios

- Lash TL, Fox MP, MacLehose RF. *Good practices for quantitative bias analysis*. IJE 2014.
- episensr R package: documentación oficial en <https://cran.r-project.org/package=episensr>
- AppliSensr web app: <https://apps.epi-conf.org/apisensr/>

Sesgo de selección

Todo estudio epidemiológico implica una decisión: **a quién podemos observar**. Desde que definimos criterios de inclusión o seguimos una cohorte en el tiempo, delimitamos la porción de realidad que miramos. Pero, ¿qué pasa cuando esa porción no representa adecuadamente la población sobre la que queremos inferir? Ahí nace el **sesgo de selección**.

El sesgo de selección ocurre cuando la **probabilidad de entrar o permanecer en el estudio** depende (directa o indirectamente) de la exposición, del desenlace o de ambos.

En otras palabras, cuando el grupo que logramos observar **no es equivalente** al grupo que deberíamos observar para estimar la relación real.

A veces el problema surge **al ingresar** al estudio (quién acepta participar, quién califica). Otras veces aparece **durante el seguimiento** (quién se pierde, quién muere, quién cambia de grupo).

En ambos casos, lo que vemos puede no reflejar lo que ocurre en la vida real.

Imagine que hace una gota gruesa para ver esferocitos.

Si por la forma en que toma la muestra, los esferocitos no quedan en la lámina, usted concluye que el paciente no los tiene... aunque en realidad sí estén.

Algo similar pasa en investigación: **el patrón de entrada o permanencia** puede distorsionar lo que vemos.

En una cohorte, por ejemplo, si los pacientes más enfermos abandonan el estudio con mayor frecuencia, el grupo que permanece tiende a parecer más sano con el tiempo.

Si la probabilidad de abandonar el estudio está relacionada con la exposición o el desenlace, **el riesgo relativo se distorsiona.**

Tipo	Ejemplo	Efecto
Sesgo de participación	Solo responden personas con mayor interés o preocupación por el tema.	Muestra no representativa.
Pérdida en seguimiento (<i>attrition</i>)	Pacientes graves abandonan el estudio antes del desenlace.	Subestima el riesgo.
Sesgo de supervivencia	Solo se incluyen sobrevivientes de una exposición.	Sobreestima efectos protectores.
<i>Collider bias</i>	Se condiciona el análisis a una variable influida por exposición y desenlace.	Asociación espuria.

El impacto del sesgo de selección también puede explorarse con el análisis cuantitativo de sesgo. Si podemos estimar o suponer la probabilidad relativa de inclusión para distintos grupos, podemos ajustar la medida de asociación observada y aproximarnos al valor real.

Por ejemplo, si las personas con exposición y desenlace tienen **menor probabilidad de estar en el estudio**, el RR observado tenderá a **subestimar** la asociación.

Con herramientas como `episensr::selection()` o `mbias()`, podemos simular cuánto se desplaza la medida bajo distintos patrones de selección.

GUÍA PRÁCTICA – Sesgo de selección (*selection bias*)

Caso guía

Pregunta: ¿existe asociación entre la exposición a un fármaco y la aparición de un evento adverso grave?

	Evento (+)	Evento (-)	Total
Expuestos	60	140	200
No expuestos	30	170	200

$$\text{RR observado} = (60/200) / (30/200) = 2,0$$

Supongamos que los pacientes con evento adverso son **menos propensos** a permanecer en el estudio (probabilidad de seguimiento 0,8 en expuestos con evento, 0,95 en otros grupos).

Corrección determinística en R

```
install.packages("episensr")
library(episensr)
selection(matrix(c(60,140,30,170), nrow=2, byrow=TRUE),
           bias_parms = c(0.8, 0.95, 0.95, 0.95))
```

Salida esperada

Observed RR = 2,0

Corrected RR ≈ 2,3

Interpretación

El RR observado subestimaba el efecto.

El patrón de pérdida selectiva reducía la representación de los casos más graves entre los expuestos.

Simulación probabilística

Si la probabilidad de inclusión es incierta, se pueden usar distribuciones:

```
probsens(matrix(c(60,140,30,170), nrow=2, byrow=TRUE),  
          type = "selection",  
          reps = 10000,  
          p1 = list("beta", 80, 20),  
          p0 = list("beta", 95, 5))
```

El resultado mostrará un rango plausible de RR corregidos.

Interpretación

El RR corregido varió entre 2,1 y 2,5 en 95 % de las simulaciones, confirmando que las pérdidas diferenciales producen una subestimación moderada del efecto.

Plantilla de interpretación

Considerando una probabilidad de retención del 80 % entre expuestos con evento adverso y 95 % en otros grupos, el RR corregido fue de 2,3 (frente a 2,0 observado). Esto sugiere una leve subestimación del efecto debido a pérdidas diferenciales. En simulaciones probabilísticas, el RR osciló entre 2,1 y 2,5, mostrando moderada sensibilidad a este sesgo.

Recursos

- Lash TL et al. *Good practices for quantitative bias analysis*. IJE 2014.
- episensr package: funciones selection(), mbias().
- AppliSensr web app: <https://apps.epi-conf.org/apisensr/>

Módulo integrador: *quantitative bias analysis* (QBA) paso a paso

Tipo de sesgo	Qué distorsiona	Cuando ocurre	En qué dirección puede sesgar
Clasificación errónea	Medición de exposición/desenlace.	Cuando el instrumento no mide con precisión.	Tiende a atenuar (no diferencial) o a inflar (diferencial).
Confusión no medida	Comparabilidad entre grupos.	Cuando hay factores no controlados asociados a exposición y desenlace.	Cualquiera, depende del confusor.
Selección	Población observada.	Cuando la inclusión o permanencia depende de exposición/desenlace.	Cualquiera, usualmente subestima.

Estos tres sesgos rara vez ocurren de manera aislada. El análisis integrado permite ver cómo su combinación puede distorsionar la medida de asociación final.

Secuencia general del QBA

Paso	Pregunta clave	Ejemplo de herramienta
Identificar mecanismos de sesgo.	¿Qué puede estar distorsionando mis datos?	DAGs, marco conceptual.
Elegir tipo de sesgo a analizar.	¿Error de medición, confusión, o selección?	Clasificación / Confusión / Selección.
Definir parámetros plausibles.	¿Qué valores podrían tener Se, Sp, probabilidades, RR de confusores?	Literatura, validación, juicio experto.
Aplicar método determinístico.	¿Cómo cambia la medida si fijo esos valores?	misclassification(), selection(), evaluate().
Aplicar método probabilístico.	¿Qué ocurre si esos valores son inciertos?	probsens().
Integrar hallazgos y reportar.	¿Cómo afectan los sesgos combinados mi conclusión?	Tabla resumen y párrafo interpretativo.

Ejemplo integrador: exposición ambiental y salud mental

Pregunta: ¿la exposición a contaminación del aire se asocia con síntomas depresivos en adultos mayores?

Datos observados

	Depresión (+)	Depresión (-)	Total
Alta exposición	90	210	300
Baja exposición	60	240	300

$$\text{RR observado} = (90/300) / (60/300) = 1,5$$

Supuestos:

- El instrumento para depresión tiene **Se = 0,85**, **Sp = 0,90** → *clasificación errónea del desenlace*.
- Existe un confusor no medido: **nivel socioeconómico**, asociado 2× con exposición y 1,8× con desenlace → *confusión no medida*.
- Los participantes más expuestos con depresión tienen menor probabilidad de permanecer (0,85) → *sesgo de selección*.

Implementación combinada (en R)

```
library(episensr)
library(EValue)
# Paso 1: Clasificación errónea
outcome_corrected <- misclassification(matrix(c(90,210,60,240), nrow=2,
byrow=TRUE),
                                     type = "outcome",
                                     bias_parms = c(0.85, 0.90, 0.85,
0.90))
# Paso 2: Confusión no medida (E-value)
evaluate(RR = 1.5, true = 1)
# Paso 3: Sesgo de selección
selection(matrix(c(90,210,60,240), nrow=2, byrow=TRUE),
            bias_parms = c(0.85, 0.95, 0.95, 0.95))
```

Interpretación integrada

RR observado: 1,5

RR corregido por clasificación: $\approx 1,7$

E-value: $\approx 2,25$ (moderada robustez ante confusión no medida)

RR corregido por selección: $\approx 1,8$

Conclusión

El conjunto de sesgos sugiere una posible subestimación de la asociación. Aun considerando escenarios conservadores, la relación entre exposición a contaminación y depresión persiste con una magnitud moderada a fuerte.

Plantilla de reporte final

El análisis de sesgo cuantitativo (QBA) abordó tres posibles fuentes de distorsión: clasificación errónea del desenlace, confusión no medida por nivel socioeconómico y sesgo de selección por pérdida diferencial. Tras la corrección determinística y simulaciones probabilísticas, la medida de asociación se mantuvo entre RR = 1,6–1,8. Este rango indica que los resultados son consistentes frente a sesgos moderados y que la dirección del efecto se mantiene estable.

Recursos de referencia

- AppliSensr. (s.f.). *Web app for sensitivity and bias analysis*. <https://apps.epi-conf.org/apisensr/>
- Episensr, EVa1ue, tipr, & sensemakr R packages. (s.f.). *R packages for sensitivity analysis and quantitative bias assessment*. <https://cran.r-project.org>
- Lash, T. L., Fox, M. P., & Fink, A. K. (2009). *Applying quantitative bias analysis to epidemiologic data*. Springer.
- VanderWeele, T. J., & Ding, P. (2017). Sensitivity analysis in observational research: Introducing the E-value. *Annals of Internal Medicine*, 167(4), 268–274. <https://doi.org/10.7326/M16-2607>