

Estado de la publicación: El preprint no ha sido enviado para publicación

# El control de terceras variables: mediación, interacción y confusión

Rafael Del-Pino-Casado

<https://doi.org/10.1590/SciELOPreprints.16002>

Enviado en: 2026-04-30

Postado en: 2026-05-05 (versión 1)

(AAAA-MM-DD)

## **Título**

El control de terceras variables: mediación, interacción y confusión [Control of third variables: mediation, interaction and confusion].

## **Autor**

Rafael del-Pino-Casado. Catedrático de Universidad. Departamento de Enfermería, Facultad de Ciencias de la Salud. Universidad de Jaén, Jaén, España.

## **ORCID**

<https://orcid.org/0000-0003-2263-2415>.

## **Conflicto de intereses**

Ninguno.

## **Financiación**

Ninguna.

## **Disponibilidad de los datos de investigación**

Al ser un artículo metodológico, no tiene datos de investigación.

Este artículo está bajo una licencia Creative Commons CC-BY.



## **Resumen**

En ciencias de la salud, los fenómenos bio-psico-sociales y culturales no pueden explicarse mediante una relación entre dos variables o un grupo de predictores y un desenlace. Las relaciones entre variables conforman redes causales complejas donde múltiples factores interactúan de forma simultánea y multidimensional. En este escenario, resulta fundamental integrar estas interrelaciones en el diseño y análisis estadístico mediante el análisis de los procesos de mediación, interacción y confusión. Este artículo pretende servir de guía para los profesionales de la salud que deseen iniciarse en el análisis de dichos procesos con una doble finalidad: por un lado, dotar a los consumidores de información científica de las bases necesarias para su correcta comprensión e interpretación; por otro, facilitar a los investigadores las herramientas metodológicas para analizar, describir y controlar estos procesos, permitiendo obtener estimaciones más robustas e insesgadas.

## **Palabras clave**

Terceras variables; mediación; interacción; confusión; investigación cuantitativa; sesgos.

## **Abstract**

In health sciences, bio-psycho-social and cultural phenomena cannot be explained by a relationship between two variables or a group of predictors and an outcome. Relationships between variables form complex causal networks where multiple factors interact simultaneously and multidimensionally. In this scenario, it is essential to integrate these interrelationships into the design and statistical analysis by examining the processes of mediation, interaction, and confusion. This article aims to serve as a guide for health professionals who wish to begin analysing these processes with a twofold purpose: on the one hand, to provide consumers of scientific information with the necessary basis for their correct understanding and interpretation; on the other, to provide researchers with the methodological tools to analyze, describe, and control these processes, allowing for more robust and unbiased estimates.

## **Keywords**

Third variables; mediation; interaction; confusion; quantitative research; bias.

## Índice

1. Marco general del control de terceras variables .....	1
2. Recomendaciones para analizar el rol de la tercera variable.....	4
3. Mediación: el camino del efecto.....	5
3.1. El concepto de mediación y de mecanismo causal.....	5
3.2. Criterios de Baron y Kenny (1986) y su evolución metodológica .....	7
3.3. El Test de Sobel: inferencia de efectos indirectos.....	8
3.4. Bootstrapping (remuestreo) .....	8
4. Interacción frente a confusión .....	9
4.1. Aspectos conceptuales: la especificación del efecto frente a su distorsión .....	9
4.2 ¿Interacción o confusión? .....	10
5. Referencias .....	12

## 1. Marco general del control de terceras variables

En el ámbito de las ciencias de la salud, los fenómenos bio-psico-sociales y culturales difícilmente pueden explicarse mediante una relación entre dos variables o entre un grupo de predictores y un desenlace. La realidad suele presentar redes causales complejas, donde múltiples factores interactúan de forma simultánea y multidimensional. En este escenario, el análisis aislado de la vía causal de interés (variable independiente y dependiente) resulta insuficiente si no se consideran otras variables intervinientes. Así, para obtener estimaciones robustas e insesgadas, es necesario integrar estas interrelaciones en el diseño y análisis estadístico; este proceso de neutralización, ajuste y explicación de efectos externos relacionados con la vía causal de interés es lo que conocemos como control de terceras variables.

Antes de comenzar a desgranar los distintos contenidos de interés, es necesario advertir de que este tema se expone desde una óptica de causalidad, por eso utilizamos la palabra “efecto”. En un escenario de asociación o correlación (diseños que no permiten demostrar causalidad), habría que hablar de “posible efecto”.

Cuando incluimos una tercera variable en la relación de una variable independiente (VI) y otra dependiente (VD), pueden darse diversos tipos de escenarios. Los más básicos, que serán nuestro objeto de estudio, son (Figura 1):

- *Independencia* (también denominado escenario de covariación): La tercera variable no afecta la relación entre la VI y la VD, de manera que la VI y la tercera variable se relacionan con la VD, pero no entre sí (*covariable*). En este caso, incluir la tercera variable en el modelo explicativo de la VD disminuiría la cantidad de varianza no explicada.
- *Confusión*: Existe una distorsión en la estimación del efecto real entre la VI y la VD, provocado por la tercera variable (variable de confusión o confundidora). Esta distorsión puede consistir en ocultar un efecto real, un efecto espurio, un cambio de signo en el efecto o una infra o sobreestimación de la magnitud del efecto. Al incluir la tercera variable en el modelo explicativo de la VD (ajustar o controlar por la tercera variable), se “limpia” la distorsión.
- *Interacción (o moderación)*: El efecto de la VI sobre la VD es heterogéneo al variar en magnitud e incluso en dirección o significación estadística según los valores de la tercera variable (*variable de interacción o moderadora*). Aquí no puede hablarse de ajuste al incluir la variable de interacción mediadora en el modelo explicativo de la VD, ya que no hay nada que corregir o eliminar, al ser el efecto de interacción una propiedad real de las variables que se estudian, y no una distorsión (un error)<sup>1</sup>.
- *Mediación*: Existe un mecanismo de paso en el efecto de la VI sobre la VD, representado por la tercera variable (*variable mediadora*), de manera que la VD es causa de la mediadora y esta a su vez es causa de la VD.

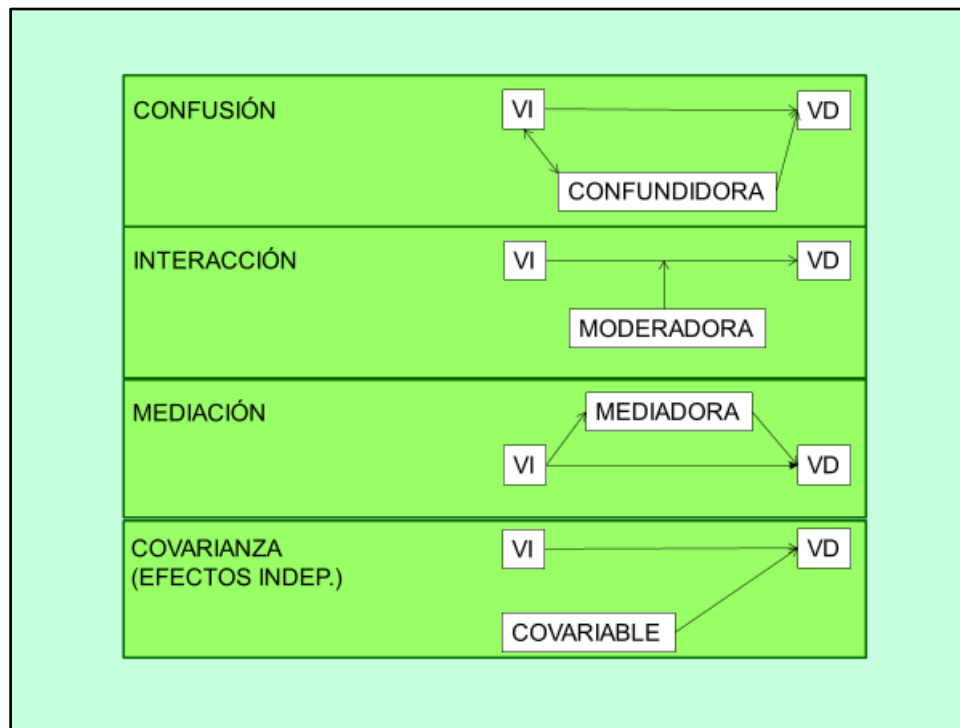


Figura 1. Escenarios básicos al introducir en el análisis una tercera variable.

Abreviaturas; VI: variable independiente; VD: variable dependiente.

Fuente: elaboración propia.

Hay que tener en cuenta que estos efectos no son todos excluyentes entre sí. Una variable de confusión o una covariable no pueden tener otro rol en un mismo modelo explicativo y set de datos, pero una variable mediadora puede ser también de interacción<sup>2</sup>. Aunque metodológicamente se suelen separar los roles para simplificar el análisis, en ciencias de la salud una variable puede ser, al mismo tiempo, el camino por el que transcurre un efecto (mediación) y el volumen al que ese efecto se reproduce (interacción). En este caso, hablaríamos de un caso particular denominado *mediación moderada*, en el que la misma variable hace de mediadora y moderadora de su propio efecto sobre la VD<sup>2</sup>.

El control y manejo estadístico de las terceras variables no es un mero ejercicio de cálculo, sino una necesidad para la integridad científica. Los resultados de controlar adecuadamente las terceras variables pueden resumirse en:

- En confusión: se limpia de distorsiones la relación entre la VI y la VD, dejando el efecto real (se corrige un sesgo).
- En mediación e interacción: se aumenta el conocimiento sobre el efecto de la VI sobre VD (no se persigue corregir un error sino aumentar el conocimiento científico):
  - Mediación: se establece el mecanismo causal de la VI sobre la VD. La variable mediadora es necesaria para que la VI tenga efecto sobre la VD, bien como único efecto [mediación completa] o como efecto añadido al efecto directo [mediación parcial].

- Interacción: se especifica en más detalle el efecto de la VI sobre la VD, estableciendo bajo qué condiciones (valores de la variable de interacción) el efecto varía en magnitud o dirección. Por tanto, las variables de interacción ayudan a explicar la variabilidad (heterogeneidad) del efecto en una investigación.

Por el contrario, las consecuencias de un inadecuado control serían:

- Confusión: el efecto de la VI sobre la VD está distorsionado (es erróneo).

Un ejemplo clásico de efecto de confusión es el del tabaco en la relación entre café y riesgo de cardiopatía isquémica. En la década de los 60, hubo estudios<sup>3</sup> que encontraron que las personas que eran grandes consumidoras de café tenían más riesgo de enfermedad coronaria. Estos resultados se debieron a que las personas que eran grandes consumidores de café (expuestas) también eran grandes fumadoras, que era realmente el verdadero factor de riesgo.

- Mediación: se pierde (oculta) el efecto de la VI sobre la VD (en mediación completa) o se infraestima (en mediación parcial).

Recientemente, López-Martínez y cols.<sup>4</sup> han puesto de manifiesto, en un estudio longitudinal prospectivo sobre personas cuidadoras familiares de mayores dependientes, que determinados estilos de afrontamiento como el replanteamiento positivo y la aceptación pueden tener efectos protectores sobre la ansiedad tanto directos como indirectos a través de la sobrecarga subjetiva (mediación parcial), de forma que estos estilos de afrontamiento por un lado hacen que los estresores se perciban como menos amenazantes y por otro suavizan la sensación de sobrecarga subjetiva, la cual, al disminuir, disminuye la ansiedad. Si no se hubiesen tenido en cuenta los efectos de mediación de la sobrecarga subjetiva, el efecto de estos estilos de afrontamiento se hubiese visto infravalorado al haberse recogido solo el efecto directo.

- Interacción: el efecto de la VI sobre la VD es verdadero pero incompleto (una verdad a medias), al no tenerse en cuenta que dicha relación no es uniforme para todos los participantes en el estudio.

Siguiendo con las personas cuidadoras familiares, en este caso de niños con enterostomía, Zhang y cols. Encontraron que la relación entre resiliencia y síntomas depresivos esta modulada por el apoyo social percibido, de manera que el posible efecto protector de la resiliencia sobre los síntomas depresivos era mayor en personas con bajo apoyo social percibido que en aquellas con alto apoyo social percibido. Esto puede ser debido a que un mayor nivel de apoyo social percibido se asocia con una menor dependencia de los recursos internos, como la resiliencia, lo que a su vez disminuye el efecto predictivo de la resiliencia sobre la depresión.

## 2. Recomendaciones para analizar el rol de la tercera variable

Cuando, como investigadores/as estamos planificando el estudio de la relación entre dos variables (VI y VD), debemos tener en cuenta todas aquellas variables que influyan en dicha relación tanto ampliando el conocimiento de dicha relación (mediación e interacción), para tenerlas en cuenta, como distorsionando dicha relación, para eliminar esta interferencia.

Así, el primer paso de esta planificación es consultar los marcos teóricos y el estado de la cuestión para identificar variables que puedan influir en nuestra relación principal, con el fin de definir las e incluirlas en nuestro proyecto de investigación.

Cuando ya tenemos nuestros datos recogidos, tenemos que comprobar si se cumple el rol asignado inicialmente a las terceras variables en nuestros datos, ya que el papel que desempeña una variable depende de su relación con la VI y la VD en un estudio particular. Por esta razón, por ejemplo, una variable que actúa moderadora en un estudio puede ser una covariable en otro.

Comprender la influencia potencial de parámetros que no son el foco de un estudio es importante para identificar qué exposiciones influyen o que intervenciones funcionan, para quién lo hacen, cuándo influyen más o funcionan mejor y en qué contextos tienen más impacto o son más útiles.

Para la comprobación del rol de una tercera variable en nuestros propios datos disponemos de criterios y pruebas estadísticas, como veremos más adelante, si bien es necesario tener una visión de conjunto de cómo tomar decisiones al respecto. En este sentido, se propone seguir el algoritmo de la Figura 2.

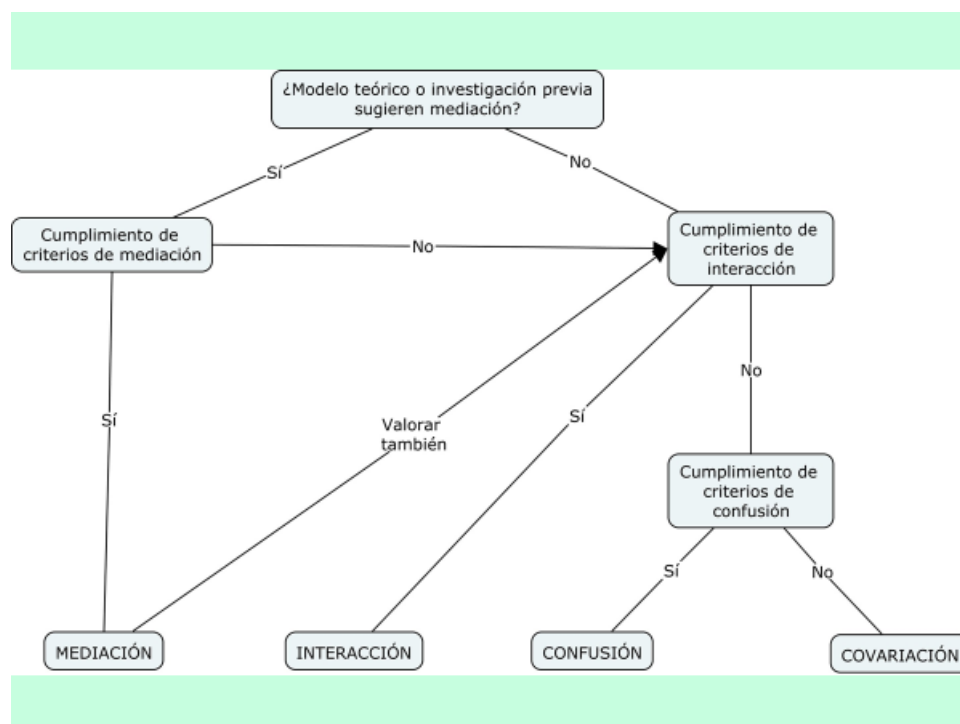


Figura 2. Algoritmo de identificación del rol de una tercera variable.

Fuente: elaboración propia.

Las bases conceptuales del algoritmo son: 1) indistinguibilidad estadística entre mediación y confusión: ambos roles provocan cambios similares en los coeficientes estadísticos, por lo que su distinción debe basarse en la teoría y la evidencia previa; lo ideal es el uso de diseños longitudinales que establezcan la secuencia temporal lógica (VI → Mediadora → VD)<sup>5</sup>; debido a esta equivalencia, el algoritmo solo contempla la mediación si existe un marco teórico o empírico sólidos que la sustenten; 2) dualidad funcional: como ya se ha comentado, una misma variable puede actuar simultáneamente como mediadora y moderadora<sup>2</sup>; el algoritmo contempla esta complejidad al evaluar la interacción de forma independiente, sin que el hecho de asumir o descartar una ruta mediadora interfiera en el análisis de la interacción; 3) primacía jerárquica de la interacción: Al evaluar una tercera variable, diversos autores recomiendan analizar primero la interacción<sup>6,7</sup>; si existe interacción, un estimador único ajustado por confusión resultaría engañoso, ya que ocultaría la heterogeneidad (que el efecto varía significativamente según los valores de la tercera variable) <sup>6</sup>; en presencia de interacción, la confusión pierde relevancia clínica, pues el foco debe desplazarse hacia la descripción de los efectos específicos en cada grupo o condición<sup>7</sup>.

Pasemos seguidamente a describir las principales relaciones expuestas en el orden de actuación recomendado.

### 3. Mediación: el camino del efecto

#### 3.1. El concepto de mediación y de mecanismo causal

La mediación describe el proceso o camino través del cual una exposición o intervención (VI) se relaciona con un problema o resultado (VD). A este camino se le llama *cadena o mecanismo causal*. Dentro de esta cadena causal, la *variable antecedente* es la que precede cronológica o lógicamente a una mediadora en una secuencia causal.

Como ya se ha comentado, el análisis de la mediación permite explicar el mecanismo causal por el que la VI influye en la VD. Si no se tiene en cuenta un efecto de mediación, se oculta el efecto de la VI sobre VD (en mediación completa) o se infraestima (en mediación parcial).

Los escenarios donde puede analizarse la mediación son muy variados en función del número de variables mediadoras (puede haber más de una variable mediadora en la cadena causal de la VI y la VD), del tipo de variables (tan la VI, la VD como la mediadora pueden ser cuantitativas, categóricas o una mezcla de todos los tipos) o del diseño del estudio (transversal, longitudinal, etc.). En este artículo nos dedicaremos al escenario más básico: un diseño transversal con una variable mediadora de la relación entre la VI y la VD, todas ellas cuantitativas, con sus dos posibilidades: 1) *mediación completa* (no hay efecto directo de la VI sobre la VD) y 2) *mediación parcial* (hay tanto efecto directo de la VI sobre la VD como efecto indirecto mediado por la mediadora).

Desde una perspectiva estadística, la inclusión de una variable mediadora en el modelo explicativo multivariante de la VD suele disminuir la magnitud de la asociación o efecto de la variable independiente sobre la VD (mediación parcial) o incluso eliminarla por completo (mediación completa). Es importante no interpretar esta reducción como un sesgo de confusión, sino como la confirmación empírica del mecanismo de mediación, como más adelante veremos. Siguiendo las directrices de autores de referencia como Hayes<sup>2</sup> y MacKinnon<sup>8</sup>, la variable mediadora debe integrarse en el modelo explicativo junto a la VI; su exclusión con el fin de preservar la significación de esta última solo conduciría a una comprensión superficial y fragmentada del fenómeno estudiado.

La demostración de un efecto de mediación permite desentrañar la 'caja negra' de la relación causal. La interpretación de la mediación depende de si es completa o parcial. La mediación completa indica que la VI no posee un efecto directo intrínseco sobre la VD, sino que su influencia se ejerce íntegramente a través de la variable mediadora. En términos clínicos, esto sugiere que el mecanismo de acción de la VI depende totalmente de la activación o modificación de la mediadora para producir un impacto en el desenlace. La mediación parcial sugiere que la VI ejerce su influencia sobre el desenlace a través de múltiples vías. Una parte del efecto se explica por la intervención de la variable mediadora analizada (vía indirecta), mientras que otra parte persiste de forma independiente a esta (vía directa).

Un ejemplo interesante de mediación parcial lo encontramos en el trabajo de Lorig y Holman<sup>9</sup>, donde se explica que una parte importante de efecto de un programa de educación diabetológica en el control de la hemoglobina glicosilada se produce por una mejora de la autoestima, mientras que otra parte se debe al aprendizaje inherente a la intervención educativa.

Cuando describimos el modelo básico de mediación, utilizamos los esquemas y notaciones que aparecen en la Figura 3. En dicha figura puede observarse que, en caso de mediación parcial, el efecto de la VI sobre la VD tiene dos componentes: el efecto directo de la VI sobre la VD ajustado por la variable mediadora ( $c'$ ) y el efecto indirecto (el producto de los coeficientes de la trayectoria del efecto indirecto:  $a*b$ ).

Los coeficientes  $a$ ,  $b$ ,  $c$  y  $c'$  son coeficientes de regresión lineal simple ( $a$  y  $c$ ) o múltiple ( $b$  y  $c'$ ) no estandarizados. Recordemos que la interpretación de los coeficientes de regresión es la siguiente: en cuantos puntos aumenta (coeficiente positivo) o disminuye (coeficiente negativo) la VD cuando la VI aumenta un punto. Por ejemplo, si tenemos apoyo social percibido y ansiedad, un coeficiente de regresión no estandarizado de  $-1,2$  significa que por cada punto que aumente el apoyo social, la ansiedad disminuirá en  $1,2$  puntos. En las salidas de los programas estadísticos o en las notaciones estandarizadas (APA, etc.), los coeficientes de regresión no estandarizados suelen representarse por la letra "b" minúscula o mayúscula en cursiva ( $b$  o  $B$ ). Sus valores posibles van de  $-\infty$  a  $+\infty$ . Y el contraste de hipótesis de dichos coeficientes se hace mediante una  $t$  de Student, donde la hipótesis nula es  $B=0$ .

Para poder comparar los coeficientes de regresión no estandarizados entre estudios que utiliza distintos instrumentos de medida se calcula el coeficiente de regresión estandarizada ( $\beta$ ), que es producto del

coeficiente no estandarizado por el cociente entre la desviación típica de VI entre la de VD y va de -1 a + 1, siendo su valor nulo el 0.

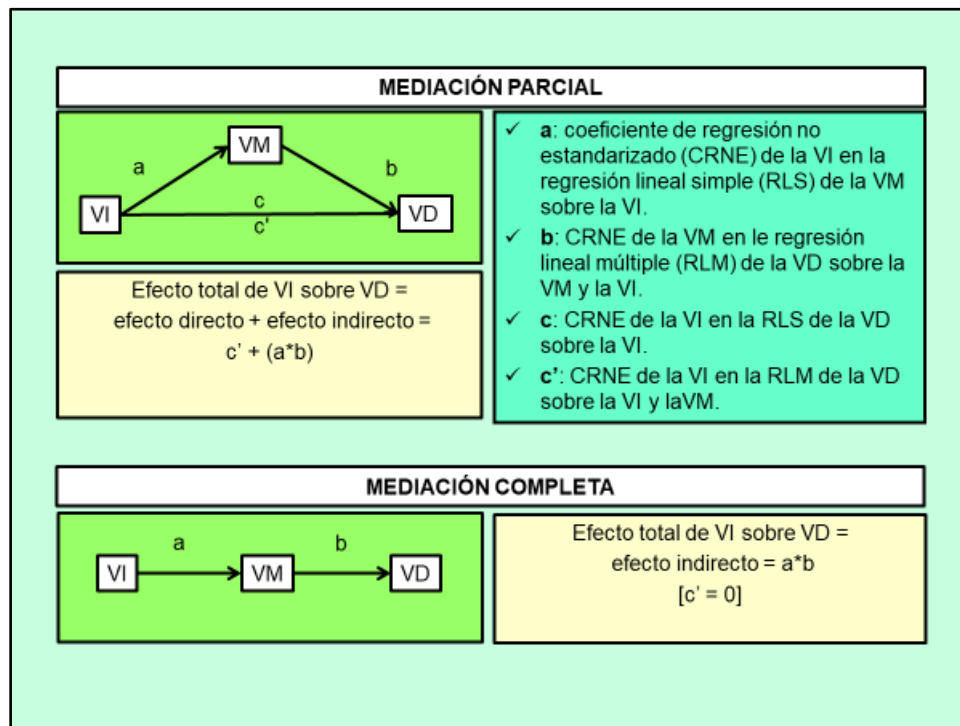


Figura 3. Esquemas y notación del modelo básico de mediación (variables cuantitativas con una mediadora)

Abreviaturas; VI: variable independiente; VM: variable mediadora; VD: variable dependiente.

Fuente: elaboración propia.

### 3.2. Criterios de Baron y Kenny (1986) y su evolución metodológica

Los primeros criterios para establecer mediación en el caso de variables cuantitativas fueron los de Baron y Kenny<sup>10</sup>. Estos autores establecieron que, para que se diese mediación, se tenían que cumplir las siguientes condiciones (ver Figura 3 para notaciones):

1. Path a (VI → VM): Existencia de asociación entre la VI y la mediadora.
2. Path c (efecto directo VD → VD bivalente): Asociación entre la VI y la VD. Cabe destacar que, según MacKinnon (2007), este criterio no es estrictamente necesario para confirmar mediación.
3. Path b (VM → VD): Asociación entre la Mediadora y la VD, ajustando por la presencia de la VI.
4. Path c' (efecto directo VD → VD ajustado): Asociación residual entre VI y VD tras ajustar por la mediadora. Si  $c' = 0$  o  $p > 0.05$ , hablamos de mediación completa (no hay efecto directo ajustado  $c'$ ).

Un criterio restrictivo sería considerar mediación completa solo si  $c' = 0$ , con independencia del valor de  $p$ , en casos de potencia insuficiente, ya que en estos casos la  $p$  no significativa no es concluyente.

Si  $c' < c$ , estamos ante una mediación parcial (efecto directo e indirecto).

Con posterioridad, MacKinnon<sup>8</sup> estableció que, de esas 4 condiciones, solo la 1 (path a) y la 3 (path b) son necesarias. La 2 (path c) puede no darse en un modelo de mediación<sup>11</sup> (con lo que, de entrada, tendríamos una mediación completa) y la 4 (path c') sirve para dilucidar el tipo de mediación cuando hay asociación bivalente entre VI y VD (path c significativo).

### 3.3. El Test de Sobel: inferencia de efectos indirectos

Los criterios anteriores permiten establecer la existencia del efecto indirecto ( $a*b$ ) confirmando la significación del path a y del path b, pero que a y b sean significativos no implica necesariamente que  $a*b$  lo sea. Para realizar un análisis más robusto, se utilizó el test de Sobel, que testa la significación estadística del path  $a*b$  (hipótesis nula:  $a*b=0$ ) mediante el cálculo de una puntuación z (y su correspondiente valor de p) a partir de a y b y de sus respectivos errores estándar ( $Sa$  y  $Sb$ )<sup>i</sup>. Existen diferentes calculadoras en Internet válidas para realizar el test de Sobel<sup>ii</sup>.

Una vez constatada la significación estadística del valor z del test de Sobel (existencia de mediación), si c no es significativo (o es 0 en situaciones de potencia insuficiente), concluimos mediación completa, y si es significativo, lo comparamos con c' para ver el tipo de mediación (como hemos descrito anteriormente). es significativo, podemos.

### 3.4. Bootstrapping (remuestreo)

El test de Sobel tiene la limitación de su baja potencia estadística al necesitar una distribución normal del efecto indirecto, lo cual solo se consigue con muestras grandes (de forma orientativa, mayores de 200 unidades. Una de las consecuencias de esto es que, entre otras cosas, con potencia estadística insuficiente algunos efectos indirectos podrían pasar desapercibido (falsos negativos).

Para superar esta limitación se utiliza el *remuestreo o bootstrapping*<sup>13</sup> para el cálculo del intervalo de confianza del efecto indirecto  $a*b$ . Diversos autores, como Hayes<sup>2</sup>, recomiendan utilizar esta técnica al ser un procedimiento de análisis del efecto indirecto que no asume una distribución normal y tiene mayor potencia estadística que los anteriormente vistos. El bootstrapping es una técnica estadística no paramétrica de remuestreo masivo con reposición. Su objetivo es estimar la distribución de probabilidad de un estadístico (como la media, la varianza o un efecto de mediación) cuando se desconoce su distribución teórica o no se cumplen los supuestos de normalidad. El remuestreo calcula los estadísticos mediante la generación de miles de muestras extraídas de la partir muestra original y de igual año que esta, donde cada observación puede ser seleccionada varias veces, una o ninguna en cada muestra.

---

<sup>i</sup> 
$$z = \frac{a*b}{\sqrt{b^2*Sa^2 + a^2*Sb^2}}$$

<sup>ii</sup> Por ejemplo: <https://www.danielsoper.com/statcalc/calculator.aspx?id=31>

Existen diversos programas estadísticos y desarrollos específicos que analizan mediación mediante remuestreo. Por citar algunos ejemplos, que también sirven para estudiar interacción, tenemos AMOS (programa de análisis de ecuaciones estructurales que trabaja sobre una base de datos de SPSS), PROCESS (macro disponible para SPSS, R o SAS) y Med4way (comando de Stata).

## 4. Interacción frente a confusión

### 4.1. Aspectos conceptuales: la especificación del efecto frente a su distorsión

Habiendo definido la mediación, es necesario contrastarla con la interacción. Esta no representa un error de estimación, sino una profundización en el conocimiento. Mientras que la confusión distorsiona la relación, la interacción la especifica y profundiza en su conocimiento. Su análisis permite determinar cómo varía el efecto de la VI sobre la VD en función de los valores de la variable moderadora. No considerar la interacción nos lleva a reportar a un conocimiento incompleto (verdades a medias) sobre la relación estudiada.

La variable de confusión altera y distorsiona la relación entre la VI y la VD, de manera que puede provocar una sobre o infraestimación del efecto, una asociación espuria (que realmente no existe), un cambio de signo del efecto u ocultamiento de una asociación real.

Para que una variable sea confundidora tiene que cumplir de forma simultánea los siguientes criterios<sup>14</sup>: 1) efecto sobre la VD: debe ser un predictor independiente de la VD, incluso en ausencia de la VI; Miettinen y Cook<sup>14</sup> advierten que este criterio debe verificarse mediante el marco teórico y la evidencia científica disponible, y no basarse únicamente en los datos de la muestra estudiada; 2) asociación con la VI: debe estar asociada estadísticamente con la VI en la muestra estudiada; en un estudio con variables independientes categóricas (comparación de grupos), esto implica que la variable de confusión se distribuye de forma desigual en los grupos de comparación; 3) no ser una variable intermedia: la variable no debe ser un paso intermedio en la cadena causal entre la VI y la VD (es decir, no debe ser una variable mediadora); en el algoritmo de decisión propuesto (Figura 2), podemos observar que para analizar la confusión es requisito haber descartado previamente la mediación; 4) criterio operativo (cambio en la estimación): la magnitud del efecto observado entre la VI y la VD debe modificarse de manera sustancial al ajustar o controlar por dicha variable. Respecto del último criterio, Miettinen y Cook<sup>14</sup> advierten de que cualquier cambio significativo de la magnitud del efecto entre la VI y la VD al ajustar por la tercera variable no equivale automáticamente a confusión; dicho cambio puede deberse a otros artefactos estadísticos y, por tanto, debe estar respaldado por una base teórica y los tres criterios anteriores.

El sesgo de confusión es el único que puede prevenirse o controlarse no solo en la fase de diseño del estudio, sino también en la de análisis, siempre que las variables potenciales de confusión hayan sido identificadas (en función de la teoría y la evidencia disponible), incluidas en el proyecto de investigación y recogida su información. Muy brevemente, los métodos de control del sesgo de confusión serían<sup>15</sup>: a) en la fase de diseño del estudio: intentando eliminar del estudio la variable de confusión mediante *restricción* (es decir, mediante

criterios de elegibilidad); o bien, si se comparan grupos (VI categórica), consiguiendo que la variable de confusión esté igualmente distribuidas en los grupos que se comparan, utilizando para ello: *apareamiento* (también conocido por su nombre en inglés: *matching*), *muestreo estratificado* o *asignación aleatoria*. b) En la fase de análisis: mediante técnicas de ajuste o control estadístico tales como: *análisis multivariante*, *análisis estratificado* o *estandarización*.

## 4.2 ¿Interacción o confusión?

En nuestro discurrir sobre la identificación del rol de una determinada tercera variable, una vez que hemos analizado la mediación, el siguiente paso es chequear primero la interacción y, si no se cumple, pasar a chequear la confusión. Para tal fin, se proponen los procedimientos y reglas de decisión que se detallan en la Tabla 1. En esta tabla se plantean los escenarios más frecuentes en función del tipo de variables y los análisis de primera elección.

En las reglas de decisión expuestas en la Tabla 1 anteriormente se ha habla de igualdad o diferencia de efectos pero, ¿cuándo podemos considerar que es significativa la variación del tamaño del efecto o asociación al estratificar o introducir la tercera variable en los modelos explicativos? Para contestar esta pregunta tenemos varias situaciones y opciones: 1) cambio en la significación estadística del valor de p: transición de un resultado estadísticamente significativo a uno no significativo (o viceversa); 2) cambio en el sentido o dirección: la magnitud del efecto o asociación cambia de signo (de un valor positivo a uno negativo o viceversa); 3) diferencia en la magnitud del efecto o asociación cruda (bivariante) de la VI y la VD y la ajustada por la tercera variable: con varias posibilidades, de menor a mayor complejidad y grado de robustez: 3.1) variación mayor al 10% entre la medida cruda y la ajustada<sup>1</sup>; 3.2) no solapamiento de los intervalos de confianza de la medida cruda y la ajustada, con suficiente potencia estadística; si no se trabaja con suficiente potencia estadística, es mejor utilizar la primera opción; 3.3) un valor de p estadísticamente significativo en test específicos de homogeneidad de efectos, donde la hipótesis nula asume la igualdad de efectos. Como ejemplo de estos test de homogeneidad de los efectos podemos citar el test de Breslow-Day y el test Q de Cochran, utilizados en el análisis de interacción (Tabla 1, escenario 5).

	Naturaleza de la VD	VI y 3ª variable	Técnica para interacción (moderación) (1)	Técnica para confusión (ajuste) (1)/ ME
1	Cuantitativa	Cualquier tipo (2)	Regresión lineal múltiple El factor de interacción (nueva variable creada mediante el producto de la VI por la 3ª variable) tiene un valor de p estadísticamente significativo al incluir la 3ª variable y dicho factor de interacción en el modelo de regresión	Regresión Lineal Múltiple / ME: Coeficientes de regresión ( $B$ o $\beta$ ) La ME bivalente de la regresión de VD sobre VI (regresión lineal simple) varía al incluir la 3ª variable en el modelo de regresión (regresión lineal múltiple)
2	Cuantitativa	Ambas categóricas (factores)	ANOVA de dos vías El valor de p del factor de interacción entre los factores es estadísticamente significativo	ANOVA / ME: $\eta^2$ La ME de la VI en el ANOVA sin la tercera variable (de un factor) varía al ejecutar el ANOVA de dos vías con la tercera variable (sin programar interacción, solo efectos principales)
3	Cuantitativa	VI categórica (factor) y 3ª variable cuantitativa	ANCOVA de una vía Valor de p del factor de interacción entre la VI y la tercera variable (covariable) es estadísticamente significativo	ANOVA y ANCOVA de una vía / ME: $\eta^2$ La ME de la VI en el ANOVA de una vía (sin la tercera variable) varía al ejecutar el ANCOVA de una vía (con la tercera variable).
4	Tiempo hasta evento	Cualquier tipo (2)	Regresión de Cox El factor de interacción (producto de la VI por la 3ª variable) tiene un valor de p estadísticamente significativo al incluir la 3ª variable y dicho factor de interacción en el modelo	Regresión de Cox / ME: HR La ME de la regresión de la VD sobre la VI varía al incluir la 3ª variable en el modelo
5	Dicotómica	VI dicotómica y tercera variable categórica (3)	Análisis Estratificado / ME: OR o RR. Criterio: valor en estratos (ajustado) distinto del valor global (crudo) y distinto entre estratos. Test de homogeneidad de los estratos: Breslow-Day para OR y Q de Cochran para RR (valor de p significativo: interacción)	Análisis Estratificado / ME: OR o RR. Criterio: valor en estratos (ajustado) distinto del valor global (crudo) y similar entre estratos. Método de Mantel y Haenszel para calcular el valor promedio de la ME ajustada por la tercera variable.
6	Dicotómica	Cualquier tipo (2)	Regresión Logística binaria múltiple El factor de interacción (producto de la VI por la 3ª variable) tiene un valor de p estadísticamente significativo al incluir la 3ª variable y dicho factor de interacción en el modelo de regresión	Regresión Logística binaria simple y múltiple / ME: OR. La ME bivalente de la regresión de VD sobre VI (regresión logística binaria simple) varía al incluir la 3ª variable en el modelo de regresión (regresión logística binaria simple)

Tabla 1. Análisis de la interacción y la confusión en siete escenarios distintos según tipo de variables.

Abreviaturas: VD: variable dependiente, VI: variable independiente, ME: magnitud del efecto o asociación,  $B$ : coeficiente de regresión lineal no estandarizado,  $\beta$ : coeficiente de regresión lineal estandarizado, ANOVA: análisis de la varianza,  $\eta^2$ : eta cuadrado, ANCOVA: análisis de la covarianza HR: hazard ratio, OR: odds ratio, RR: riesgo relativo.

Fuente: elaboración propia.

NOTAS:

(1) Análisis de primera opción (sólo válidos si se cumplen las condiciones para su aplicación).

(2) Si hay variables policotómicas, deben de introducirse en el modelo en forma de variables dummy, que son variables dicotómicas formadas a partir de la variable policotómica, donde la categoría de interés es una de las categorías de interés de la variable policotómica y la categoría de referencia es la categoría de referencia de la policotómica. Por tanto, en una variable policotómica de n categorías, se formarán n-1 variables dummy. Por ejemplo: Variable dicotómica: nivel: alto, medio y bajo (de referencia); variables dummy: nivel 1 (medio / bajo) y nivel 2 (alto / bajo).

(3) También puede analizarse como en el escenario 6.

## 5. Referencias

1. Rothman KJ, Greenland S, Lash TL. *Modern epidemiology*. Philadelphia: Lippincott Williams & Wilkins; 2008.
2. Hayes AF. *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. 3rd ed. New York: Guilford Press; 2022.
3. Paul O, Lepper MH, Phelan WH, Dupertuis GW, Macmillan A, Mc KH, et al. A longitudinal study of coronary heart disease. *Circulation*. 1963; 28:20-31.
4. Lopez-Martinez C, Orgeta V, Frias-Osuna A, Del-Pino-Casado R. Coping and anxiety symptoms in family carers of dependent older people: Mediation and moderation effects of subjective caregiver burden. *J Nurs Scholarsh*. 2024; 56:371-81.
5. MacKinnon DP, Krull JL, Lockwood CM. Equivalence of the mediation, confounding and suppression effect. *Prev Sci*. 2000; 1(4):173-81.
6. Kleinbaum DG, Sullivan KM, Barker ND. *A pocket guide to epidemiology*. New York, NY: Springer Science & Business Media; 2007.
7. Gordis L. *Epidemiology*. 6th ed. Philadelphia, PA: Elsevier; 2018.
8. MacKinnon DP, Fairchild AJ, Fritz MS. Mediation analysis. *Annu Rev Psychol*. 2007; 58:593-614.
9. Lorig KR, Holman H. Self-management education: history, definition, outcomes, and mechanisms. *Ann Behav Med*. 2003; 26(1):1-7.
10. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *J Pers Soc Psychol*. 1986; 51(6):1173-82.
11. Jose P. *Doing statistical mediation and moderation*. New York: The Guilford Press; 2013.
12. Chumbler NR, Rittman M, Van Puymbroeck M, Vogel WB, Qin H. The sense of coherence, burden, and depressive symptoms in informal caregivers during the first month after stroke. *Int J Geriatr Psychiatry*. 2004; 19(10):944-53.
13. Preacher KJ, Hayes AF. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behav Res Methods*. 2008; 40(3):879-91.
14. Miettinen OS, Cook EF. Confounding: essence and detection. *Am J Epidemiol*. 1981; 114(4):593-603.

15. Del Pino Casado R, Frías Osuna A, Palomino Moral P. El control de sesgos en la investigación cuantitativa enfermera. *Revista Iberoamericana de Enfermería Comunitaria*. 2010; 4(1):24-34.

## Este preprint fue presentado bajo las siguientes condiciones:

- Los autores declaran que se obtuvieron los términos necesarios del consentimiento libre e informado de los participantes o pacientes en la investigación y se describen en el manuscrito, cuando corresponde.
- Los autores declaran que la preparación del manuscrito siguió las normas éticas de comunicación científica.
- Los autores declaran que son conscientes de que son los únicos responsables del contenido del preprint y que el depósito en SciELO Preprints no significa ningún compromiso por parte de SciELO, excepto su preservación y difusión.
- Los autores declaran que los datos, las aplicaciones y otros contenidos subyacentes al manuscrito están referenciados.
- El manuscrito depositado está en formato PDF.
- Los autores declaran que la investigación que dio origen al manuscrito siguió buenas prácticas éticas y que las aprobaciones necesarias de los comités de ética de investigación, cuando corresponda, se describen en el manuscrito.
- Los autores declaran que una vez que un manuscrito es postado en el servidor SciELO Preprints, sólo puede ser retirado mediante solicitud a la Secretaría Editorial deSciELO Preprints, que publicará un aviso de retracción en su lugar.
- Los autores aceptan que el manuscrito aprobado esté disponible bajo licencia [Creative Commons CC-BY](#).
- El autor que presenta el manuscrito declara que las contribuciones de todos los autores y la declaración de conflicto de intereses se incluyen explícitamente y en secciones específicas del manuscrito.
- Los autores declaran que el manuscrito no fue depositado y/o previamente puesto a disposición en otro servidor de preprints o publicado en una revista.
- Si el manuscrito está siendo evaluado o siendo preparando para su publicación pero aún no ha sido publicado por una revista, los autores declaran que han recibido autorización de la revista para hacer este depósito.
- El autor que envía el manuscrito declara que todos los autores del mismo están de acuerdo con el envío a SciELO Preprints.