

Estudios de asociación y riesgo

Juan Carlos López Abvarengo
Gregorio Tomás Obrador

Se pueden presentar dos razones radicalmente opuestas de por qué un problema es insoluble: una razón es que el problema es demasiado elevado para la inteligencia; la otra es que la propia formulación del problema asume puntos que le quitan el sentido.
John Dewey (1859-1952)

Riesgo se define como la probabilidad de desarrollar una enfermedad u otro evento de interés cuando se produce la exposición a un factor determinado. Cuando el factor se asocia casi invariablemente con el desarrollo de la enfermedad, se dice que la asociación es causal. Más frecuentemente, el factor es sólo un marcador o indicador, es decir, su presencia se asocia con un mayor riesgo de desarrollar la enfermedad, pero no es necesariamente la causa de la misma (factor de riesgo). Para determinar si una asociación es causal se deben cumplir una serie de criterios de causalidad, los cuales se explican en el capítulo 4. Como ejemplo de estos conceptos mencionaremos la asociación entre tabaquismo y cáncer de pulmón. Se ha demostrado que los fumadores tienen un riesgo 20 veces mayor de desarrollar cáncer de pulmón que los no fumadores. No obstante, sólo 2 de cada

1,000 fumadores importantes desarrollará cáncer de pulmón en los siguientes 10 años. Aunque no todos los sujetos que fuman desarrollan cáncer de pulmón, la asociación se considera causal porque cumple con criterios de causalidad.

Los estudios de riesgo o asociación se utilizan fundamentalmente para dilucidar factores causales o de riesgo para el desarrollo de una enfermedad. La identificación de dichos factores es importante con fines de prevención. También se utilizan para identificar factores pronósticos, es decir, aquellos que se asocian con una mejor o peor expectativa de sobrevivir a una enfermedad. El análisis de estudios pronósticos es diferente al de estudios sobre factores de riesgo y se explica en el capítulo 17, *Análisis de supervivencia y análisis multivariado*.

DISEÑO DE LOS ESTUDIOS DE RIESGO O ASOCIACIÓN

Para determinar si un factor se asocia al desarrollo de una enfermedad, lo ideal sería realizar un experimento (estudio clínico controlado) en el que exponemos a un grupo de individuos al supuesto factor de riesgo y comparamos la tasa de enfermedad con la de otro grupo que no ha sido expuesto a dicho factor. No obstante, esta estrategia es impracticable en muchos casos porque: 1) frecuentemente la exposición al factor de riesgo requiere un tiempo largo para que se desarrolle la enfermedad; 2) exponer a un grupo de individuos a supuestos factores de riesgo para el desarrollo de una enfermedad puede ser inaceptable desde el punto de vista ético. Por tanto, los estudios que examinan el riesgo o la asociación entre un factor y una enfermedad suelen ser observacionales, como los estudios de cohortes y los tipos de estudios. En el capítulo 5 se explican los aspectos esenciales de ambos tipos de estudios. En este capítulo se estudian en más detalle y se examinan las medidas de asociación.

ESTUDIOS DE COHORTES

Una cohorte se define como un grupo de individuos que tienen una o varias características en común (por ejemplo, exposición a un supuesto factor de riesgo) y que son seguidos en su evolución en el futuro en busca del desarrollo de una enfermedad u otro evento de interés. Es importante asegurar que al inicio los individuos no tengan la enfermedad que se quiere estudiar, que se defina con precisión si están o no expuestos al supuesto factor de riesgo, y que se les de seguimiento a todos por un tiempo suficientemente largo para determinar si la enfermedad efectivamente se produce o no (figura 8-1).

En los estudios de cohortes se compara la incidencia de una enfermedad en individuos expuestos y en los no expuestos al supuesto factor de riesgo, con lo que

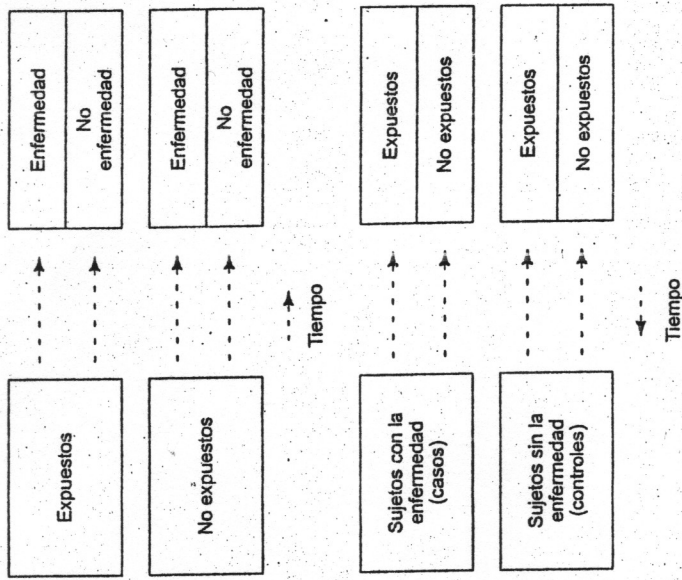


Figura 8-1. Diseño de estudios de cohortes y de casos y controles.

se obtiene el riesgo relativo o la diferencia de riesgo (ver más adelante). A los estudios de cohortes también se les llama longitudinales (en contraposición a los transversales), prospectivos (puesto que generalmente se sigue su evolución en el futuro) o de incidencia (porque ésta es la medida de frecuencia de la enfermedad).

Hay dos tipos de estudios de cohortes: prospectivos (o concurrentes) y retrospectivos (o históricos). En los prospectivos se determina la exposición a un factor determinado en el presente y se sigue la evolución de los individuos de la cohorte en el futuro, para ver si se desarrolla la enfermedad en los expuestos y en los no expuestos al factor.

En los históricos, tanto la exposición al factor como el desarrollo de la enfermedad han ocurrido en el pasado, pero la obtención de la información se realiza en el presente.

Tabla 8-1. Tipos de sesgos en los estudios observacionales

Sesgo de susceptibilidad	Es un tipo de sesgo de selección que consiste en que hay un imbalance de factores que influyen en el desarrollo de la enfermedad en los grupos que se están comparando. En otras palabras, un grupo (cohorte, caso o control) puede ser más susceptible a desarrollar la enfermedad debido a la existencia de factores distintos del factor de riesgo que se está estudiando.
Sesgo de sobrevivencia	Es un tipo de sesgo de selección que puede observarse en estudios de pronóstico y que consiste en que preferentemente los individuos que sobreviven son seleccionados para el estudio, por lo que su pronóstico no representa la historia natural de la enfermedad. Para evitarlo se debe intentar captar a los pacientes al inicio o en un momento bien definido de la evolución de la enfermedad.
Sesgo de migración	Ocurre cuando un individuo sale del estudio por diversas razones (muerte, recuperación, cambio de ciudad, intolerancia a los medicamentos, etc.) o "migra" de un grupo a otro (de una cohorte a la otra, de caso a control o viceversa). Dichos individuos que salen o migran suelen tener características diferentes a las de los individuos que continúan en el estudio; si dichas características influyen en la probabilidad de desarrollar la enfermedad o su pronóstico, su salida o migración introducirá sesgo en los resultados.
Sesgo de información	Ocurre cuando la probabilidad de detectar la enfermedad o el evento de interés es mayor en un grupo que en otro. Para evitarlo se debe asegurar que los investigadores no conozcan a que cohorte pertenece el individuo y que los criterios para definir la enfermedad o el evento de interés sean definidos <i>a priori</i> con toda claridad. También se pueden buscar fuentes alternativas de información.

Aunque los estudios de cohortes son el mejor sustituto de los ensayos clínicos controlados, tienen la desventaja de que son poco prácticos para enfermedades raras, frecuentemente requieren de un tiempo de seguimiento largo y los costos suelen ser altos. También, como otros estudios observacionales, están sujetos a posibles sesgos, los cuales se resumen en la tabla 8-1. Dichos sesgos se pueden controlar en la fase de diseño o de análisis del estudio utilizando una serie de métodos, los cuales se resumen en la tabla 8-2.

■ ESTUDIOS DE CASOS Y CONTROLES

Los estudios de casos y controles se inician con la identificación de un grupo de "casos" que tienen la enfermedad u otro evento de interés y un grupo de "controles" que no la tienen; después se determina retrospectivamente la exposición a uno o varios factores de riesgo que se desea investigar y se compara su magnitud en ambos grupos (figura 8-1). Por ejemplo, si queremos determinar si el cáncer cervicouterino está asociado al uso de estrógenos, número de parejas sexuales e historia de circuncisión, podríamos escoger un grupo de casos que tengan la

Tabla 8-2. Métodos para controlar el sesgo de selección

Durante el diseño del estudio: ¹	
Restricción	Consiste en limitar los criterios de inclusión a aquellos individuos que cumplan ciertas características, con el objeto de formar grupos homogéneos que sólo se distinguen por la exposición al factor de riesgo o al factor pronóstico que se está estudiando. Aunque la restricción aumenta la validez interna del estudio, puede disminuir la generalización de las conclusiones (ver capítulo 4).
Pareamiento	Consiste en que los grupos que se están comparando sean similares ("estén pareados") en características que se sabe pueden influenciar la probabilidad de desarrollar la enfermedad o su pronóstico (por ejemplo, la edad y el sexo pueden influenciar la aparición y/o el pronóstico de muchas enfermedades).
Durante el análisis del estudio: ²	
Estratificación	Consiste en comparar las tasas de incidencia de la enfermedad en subgrupos o "estratos" con características similares, en los que la probabilidad de desarrollar la enfermedad o el pronóstico sea parecida, excepto por la influencia del factor en estudio.
Análisis multivariado	Consiste en ajustar o controlar por varios factores que pueden influenciar el desarrollo de la enfermedad mediante la estimación de modelos matemáticos de regresión (ver capítulo 17).

1. La aleatorización es otro método para controlar el sesgo de selección, pero sólo se aplica en los ensayos clínicos controlados (ver capítulo 9). Consiste en que los individuos son escogidos al azar de modo que tengan la misma probabilidad de pertenecer a cualquiera de los grupos que se están comparando (experimental o control).

2. La estandarización, aunque menos utilizada, es otro método de controlar el sesgo de selección durante el análisis. Es un método matemático que permite darle un "peso" parecido a estratos que tengan un pronóstico similar.

enfermedad con base en criterios bien definidos, y un grupo de controles que no la tenga. Posteriormente, podríamos determinar mediante un cuestionario o la revisión de expedientes clínicos la magnitud de la exposición a los posibles factores de riesgo de cáncer cervicouterino en los casos, comparado con los controles.

Los estudios de casos y controles tienen la ventaja de ser más prácticos y económicos que los estudios de cohortes, porque no requieren seguir una cohorte grande de individuos por un tiempo largo hasta que se produzca la enfermedad, especialmente si ésta es poco frecuente. Con el diseño de casos y controles, se identifican individuos que ya tienen la enfermedad y se compara simultáneamente la exposición a factores de riesgo con un grupo que no la tiene. A pesar de estas ventajas, la posibilidad de sesgos es grande, por lo que la evidencia de la asociación entre enfermedad y factores de riesgo suele ser menos fuerte en estudios de casos y controles que en los de cohortes.

Para la selección de los casos se deben definir con precisión los criterios diagnósticos de la enfermedad o del evento que se quiera investigar. Es preferible seleccionar casos incidentes (nuevos), pues la frecuencia de los casos prevalentes depende tanto de la incidencia como de la duración de la enfermedad. Una preva-

lencia alta no permite distinguir si existe un alto riesgo de desarrollar la enfermedad o si los pocos que la desarrollan permanecen enfermos un largo tiempo, o ambas (sesgo de prevalencia). En cuanto a la selección de los controles, se debe asegurar que no tienen la enfermedad; también que potencialmente tienen la misma probabilidad de exposición a los factores de riesgo que los casos. Es por esto último que los controles deben provenir de la misma población que los casos. Por ejemplo, si los casos fueron seleccionados de un hospital, los controles deberán ser hospitalarios también; en investigaciones de campo los investigadores pueden pedir a los casos que lleven a un amigo, vecino o familiar como control, para aumentar la probabilidad de que ambos estén expuestos a los mismos factores ambientales.

Como en este tipo de diseño se parte de la identificación de casos, no se puede determinar la incidencia acumulada de la enfermedad y por tanto el riesgo relativo entre los expuestos al factor de riesgo y los no expuestos. Es por ello que la medida de asociación más utilizada es la razón de momios u *odds ratio*, la cual es matemáticamente equivalente al riesgo relativo (ver más adelante). Como el valor de referencia es 1, valores por arriba de 1 se consideran de riesgo y por debajo de 1 de protección.

Por ejemplo, si se obtiene una razón de momios de 3.0, le restamos 1 interpretamos que los individuos expuestos al factor de riesgo tienen tres veces más riesgo de padecer la enfermedad que los no expuestos.

Por el contrario, si se obtiene una razón de momios de 0.25 y queremos saber cuantas veces más protección tienen los expuestos al factor protector comparados con los no expuestos, simplemente obtenemos el valor inverso ($1.0 \div 0.25 = 4.0$), le restamos 1 interpretamos que los individuos expuestos al factor tienen cuatro veces más protección que los no expuestos.

Los estudios de casos y controles son más susceptibles que los de cohortes a sesgos de selección e información, debido a que la exposición al factor en estudio se hace retrospectivamente, una vez que la enfermedad o la ausencia de enfermedad ha ocurrido. En cuanto al sesgo de selección, cabe la posibilidad de los casos se seleccionen por presentar una historia positiva de exposición al factor de riesgo, y que los controles se seleccionen por presentar una historia negativa de exposición al mismo. Por ejemplo, si se está realizando un estudio sobre factores de riesgo para el desarrollo de tromboembolia pulmonar, es probable que mujeres que toman anticonceptivos sean seleccionadas para el estudio porque se sabe que los estrógenos pueden ser un factor de riesgo. En cuanto al sesgo de información, también la presencia o ausencia de la enfermedad pueden influenciar la forma como el paciente recuerda la exposición a posibles factores de riesgo, o como el investigador busca o documenta la información acerca de los mismos. Por ejemplo, es más probable que un sujeto que padece la enfermedad recuerde, e incluso

exagerar, la exposición a un posible factor de riesgo que un sujeto que no la padece (sesgo de memoria). Por otro lado, el investigador puede indagar con más insistencia acerca de un factor de riesgo en los casos que en los controles si sabe que dicho factor está relacionado con el desarrollo de la enfermedad.

Para disminuir la posibilidad de sesgos de selección, lo primero es que tanto los casos como los controles provengan de la misma población o cohorte. También se pueden utilizar métodos como el apareamiento (*matching*), la restricción, la estratificación y el análisis multivariado (tabla 8-2). Uno de los más usados es el apareamiento de características en los casos y en los controles, que se sabe influyen en el desarrollo de la enfermedad (por ejemplo, la edad, el sexo, el lugar de residencia, etc.), independientemente del factor de riesgo en estudio. Es importante que el apareamiento no incluya características que están directamente relacionadas con la exposición al factor de riesgo, ya que así pueden anularse las diferencias en la exposición (*overmatching*). También que se limite a 4 o 5 características, porque el sesgo de selección no suele mejorar si seorean más variables confusoras. Cuando se opta en el diseño por el apareamiento, el análisis estadístico también debe ser pareado para no perder eficiencia estadística. Otra estrategia para disminuir el riesgo de sesgo de selección consiste en escoger más de un grupo control de diferentes fuentes, siempre y cuando los casos y los controles provengan de la misma población o cohorte. Si la razón de momios que se obtiene es similar en los diferentes grupos control, es improbable que haya habido un sesgo de selección. Por último, la posibilidad de sesgos de información se puede disminuir buscando fuentes alternativas de exposición al factor (por ejemplo, registros en las farmacias) y limitando la información que se proporcione a los sujetos enrolos en el estudio y a los investigadores.

MEDIDAS DE ASOCIACIÓN

Las medidas de asociación o efecto son estimadores de la relación entre la exposición a un factor (de riesgo en estudios de riesgo, o pronóstico en estudios pronósticos) y la incidencia o el pronóstico de una enfermedad. Las medidas más comúnmente utilizadas son el riesgo relativo, la razón de momios (en estudios de casos y controles) y el riesgo atribuible.

RIESGO RELATIVO

El riesgo relativo es el cociente de la incidencia de enfermedad en los individuos expuestos entre los no expuestos al factor en estudio. Consideremos la siguiente tabla de contingencia:

		Enfermedad	
		Sí	No
Exposición	Sí	a	b
	No	c	d
		a + c	b + d
		a + b	c + d
		N	

La celda a representa aquellos individuos que tienen la enfermedad y también la exposición; la celda b , aquellos que no tienen la enfermedad, pero sí la exposición; la celda c , los no expuestos, pero con la enfermedad; y la d , los que no tienen la exposición ni la enfermedad. Los marginales muestran a los individuos con enfermedad ($a+c$) y sin enfermedad ($b+d$), los expuestos al factor ($a+b$) y los no expuestos ($c+d$). La suma de todas las celdas es la muestra total (N).

En un estudio de cohortes se obtienen las incidencias acumuladas de enfermedad en el grupo de los expuestos (CI_E) y en el de los no expuestos (CI_0). La incidencia acumulada en los expuestos se obtiene dividiendo el número de casos con la enfermedad y la exposición entre todos los expuestos [$CI_E = a/(a+b)$]. La incidencia acumulada en los no expuestos se obtiene dividiendo el número de casos con la enfermedad y sin exposición entre todos los no expuestos [$CI_0 = c/(c+d)$]. El riesgo relativo (RR) se calcula de la siguiente manera:

$$RR = CI_E/CI_0 = \frac{a/(a+b)}{c/(c+d)}$$

Como ejemplo, en un estudio reciente de 3,867 pacientes diabéticos del Reino Unido (UKPDS) que fueron seguidos por un promedio de 10 años, se comparó la eficacia del tratamiento convencional (mantener la glucemia por debajo de 270 mg/dL) con la del tratamiento intensivo (mantenerla por debajo de 108 mg/dL) de la *diabetes mellitus*. Se definieron 21 complicaciones de la diabetes como eventos finales, incluyendo infarto del miocardio, enfermedad vascular cerebral, amputación de un miembro inferior, ceguera en un ojo, diálisis peritoneal y muerte relacionada a la diabetes, entre otros. Los resultados fueron los siguientes:

	Alcanzaron algún evento final	No alcanzaron algún evento final	Total
Tratamiento convencional	438	700	1,138
Tratamiento intensivo	963	1,766	2,729
	1,401	2,466	3,867

La incidencia acumulada de algún evento final en el grupo con tratamiento convencional fue de $438 \div 1,138 = 0.38$, y en el grupo con tratamiento intensivo fue de $963 \div 2,729 = 0.35$. El RR de alcanzar un evento final en el grupo que recibió tratamiento conservador comparado con el que recibió tratamiento intensivo fue de $0.38 \div 0.35 = 1.08$. Esto significa que los pacientes con tratamiento convencional presentan 0.08 veces más riesgo de presentar alguna complicación de la diabetes que los pacientes con tratamiento intensivo. Otra forma de expresarlo es que los pacientes que reciben tratamiento convencional tienen 8% más riesgo de padecer alguna de las complicaciones de la diabetes que los que reciben tratamiento intensivo.

■ RAZÓN DE MOMIOS (OR: ODDS RATIO)

Como en los estudios de casos y controles se parte de la identificación de casos con la enfermedad y se determina retrospectivamente la exposición a factores de riesgo al mismo tiempo, no se pueden calcular tasas de incidencia acumulada, y por tanto el riesgo relativo. Afortunadamente se ha demostrado que otra medida de asociación o efecto, la razón de momios u *odds ratio*, es matemáticamente equivalente al riesgo relativo y se puede calcular con la información obtenida en estudios de casos y controles. Momios u *odds* se define como la relación o cociente de la probabilidad de que un evento ocurra entre la probabilidad de que no ocurra.

La razón de momios se calcula dividiendo los momios de enfermedad, dado que se está expuesto, (a/c = los sujetos expuestos que tienen la enfermedad entre los no expuestos que también tienen la enfermedad) entre los momios de ausencia de enfermedad, dado que se está expuesto (b/d = los sujetos expuestos que no tienen la enfermedad entre los no expuestos que tampoco tienen la enfermedad).

$$\frac{a/c}{b/d} = \frac{ad}{bc}$$

Como la ecuación de los momios de enfermedad entre los de ausencia de enfermedad, dado que se tiene la exposición, se puede simplificar a ad/bc , se le ha dado el nombre de razón de productos cruzados, porque que es como formar una cruz en la tabla de contingencia.

Como ejemplo, González-Barranco y colaboradores realizaron un estudio de casos y controles en 433 mujeres obesas (índice de masa corporal > 25) de diversos estados de la República Mexicana para determinar la influencia de factores ambientales en el desarrollo de obesidad de tipo androide (relación de la circunferencia cintura-cadera mayor de 0.85). Este tipo de obesidad, en contraposición a la ginecoide, se ha asociado con una mayor frecuencia de enfermedad cardiovascular en mujeres. Aunque se ha postulado que la obesidad de tipo androide tiene una base genética im-

portante, se realizó el estudio para determinar la influencia de factores ambientales, en concreto, la migración de zona rural a zona urbana, en el desarrollo de este tipo de obesidad. Los resultados que se obtuvieron fueron los siguientes:

	Obesidad de tipo androide	Obesidad de tipo ginecoides	Total
Migró de zona rural a urbana	69	52	121
Siempre ha vivido en zona urbana	126	186	312
	195	238	433

La razón de momios se calcula dividiendo $(69 \times 186) \div (52 \times 126) = 1.96$. Esto significa que las mujeres que migraron de zona rural a urbana tienen casi dos veces más riesgo de desarrollar obesidad androide que las que siempre han vivido en zona urbana, lo que sugiere que factores ambientales juegan un papel modesto en el desarrollo de este tipo de obesidad.

■ RIESGO ATRIBUIBLE

Como no es infrecuente que sujetos no expuestos al factor que se considera de riesgo desarrollen la enfermedad (los no fumadores también desarrollan cáncer de pulmón), la diferencia de la incidencia de la enfermedad en los expuestos y no expuestos es el exceso de riesgo de desarrollar la enfermedad que se puede atribuir al factor. A esta medida de asociación se le denomina riesgo atribuible (RA) o exceso de riesgo, porque expresa el exceso de riesgo debido a la exposición al factor, por encima del riesgo que confiere la ausencia de exposición al mismo. También se le conoce como diferencia de riesgo, puesto que resulta de restar la incidencia acumulada de enfermedad en los expuestos (I_E) y los no expuestos (I_0).

$$RA = I_E - I_0$$

La proporción de riesgo atribuible (RA%) es otra medida de asociación que se calcula con la siguiente fórmula:

$$RA\% = \frac{RA}{I_E} \times 100 = \frac{I_E - I_0}{I_E} \times 100$$

Ambas medidas de asociación expresan el número de casos (RA) o la proporción de casos de enfermedad (RA%) que se podrían evitar si se eliminara el factor de riesgo.

Volviendo al estudio UKPDS antes mencionado, ¿cuál es el RA y la proporción de riesgo atribuible de utilizar tratamiento convencional en vez de tratamiento

intensivo? El RA o diferencia de riesgo es $0.38 - 0.35 = 0.03$, lo que significa que el riesgo de complicaciones graves disminuiría en 0.03 si se utilizara el tratamiento intensivo en vez del convencional. El RA% se calcula dividiendo $0.03 \div 0.35 = 0.09 \times 100 = 9\%$, lo que significa que se podrían disminuir las complicaciones graves de la diabetes en un 9% si se usara el tratamiento intensivo en vez del convencional.

■ INTERPRETACIÓN DE LAS MEDIDAS DE ASOCIACIÓN

Una vez que se obtiene la medida del riesgo de desarrollar la enfermedad dada la exposición a un factor, se debe calcular el valor de p y del intervalo de confianza para determinar la significancia estadística del resultado (ver capítulos 4 y 15). En el estudio de las mujeres que migraron y tuvieron casi 2 veces más riesgo de tener obesidad androide que las que permanecieron en la zona urbana, el valor de p fue significativo ($p = 0.02$) y el intervalo de confianza a 95% fue de 1.29 a 3.09, lo que indica que el aumento del riesgo encontrado no es debido al azar.

También se debe analizar la relación entre las distintas variables o factores de riesgo estudiados. Un factor confusor es una variable que se relaciona con otro factor de riesgo, y que independientemente del mismo, se asocia con el desarrollo de la enfermedad (ver capítulo 4). En ocasiones el efecto de un factor de riesgo aumenta o disminuye dependiendo de la presencia de otro, lo que se conoce como modificación del efecto. Para determinar el efecto de un factor en el desarrollo de la enfermedad independientemente del efecto de otros, se utilizan técnicas como la estratificación y el análisis multivariado (ver capítulo 17).

Por último, cuando se hace análisis estratificado, en vez de calcular una medida de asociación para cada estrato se suele reportar un solo valor ajustado para todos los estratos. El RR ajustado para los diferentes estratos se obtiene aplicando la técnica de Mantel-Haentzel a los estudios de cohorte:

$$RR_{MH} = \frac{\sum \frac{a(c+d)}{N} - \frac{c(a+b)}{N}}{\sum \frac{c(a+b)}{N}}$$

Para los de casos y controles:

$$OR_{MH} = \frac{\sum \frac{ad}{N}}{\sum \frac{bc}{N}}$$

Donde los numeradores y los denominadores representan la suma de los productos de cada estrato.