

## 11. Multicausalidad: el fenómeno de confusión

### *Justificando la naturaleza multicausal de la enfermedad – asociaciones secundarias y su control*

#### **Introducción**

Cuando la “epidemiología moderna” se desarrolló en la década de los 70, Olli Miettinen organizó las fuentes de sesgo en tres categorías principales: sesgo de selección, sesgo de información, y sesgo de confusión. Si nuestro interés es la asociación bruta entre dos factores, el sesgo de selección puede llevarnos a observar una asociación diferente de la que existe en la población que creemos estar estudiando (la población blanco). De la misma manera, el sesgo de información puede provocar que la asociación aparezca diferente de cómo lo es en la realidad. El sesgo de confusión difiere de los otros tipos de sesgo porque el fenómeno de confusión no altera la asociación bruta. Lo que preocupa en el fenómeno de confusión es la interpretación de la asociación observada.

Ya hemos considerado el fenómeno de confusión sin llamarlo por ese término, en el capítulo sobre estandarización por edad. La comparación de tasas brutas de mortalidad puede ser engañosa, no porque las tasas estén sesgadas, sino porque son afectadas en forma importante por las distribuciones por edad de los grupos que son comparados. Así, para poder interpretar la comparación de tasas de mortalidad necesitamos examinar las tasas específicas por edad y las tasas estandarizadas por edad para evitar o igualar la influencia de la edad. Si hubiéramos intentado interpretar las tasas brutas, nuestra interpretación hubiera sido **confundida** por las diferencias de edades en las poblaciones comparadas. Por lo tanto **controlamos** los efectos de la edad para quitar el efecto de confusión. En este capítulo ahondaremos en los procesos del fenómeno de confusión y revisaremos el repertorio de estrategias para evitarlo o controlarlo.

#### **Razonamiento contrafactual**

La investigación epidemiológica, sea descriptiva o analítica, etiológica o evaluativa, habitualmente busca llegar a interpretaciones causales. Una asociación entre dos factores promueve la interrogante sobre cuál será la causa del fenómeno (o en la situación opuesta, por qué no aparece una asociación esperada). El razonamiento causal sobre las asociaciones, aún en aquellas que no son el objetivo de la investigación, es parte del proceso de encontrar el significado de los datos. De manera que es fundamental la capacidad para inferir relaciones causales a partir de las asociaciones observadas.

En el “mundo ideal de los epidemiólogos”, podríamos inferir la causalidad comparando resultados en salud para una persona expuesta al factor de interés con lo que hubiera sido el resultado en ausencia de la exposición. Una comparación de lo que ocurriría en presencia de la exposición y en ausencia de la exposición es llamada contrafactual, porque un lado de la comparación va en contra de los hechos (ver Rothman y Greenland, pág.49, que atribuyen este concepto a los trabajos de Hume del siglo XVIII). Esta comparación contrafactual provee de una base lógica robusta para

inferir causalidad, porque el efecto de la exposición puede ser aislado de la influencia de otros factores.

En el mundo factual, sin embargo, nunca observaremos la misma situación dos veces, salvo talvez en el caso de la “repetición instantánea (replay)”, que de cualquier manera no nos permite alterar la situación de exposición. La plétora de factores que pueden influir en un resultado varían de persona a persona, lugar a lugar, y de un tiempo a otro. La variación de estos factores es responsable de la variabilidad en los resultados que observamos, de manera que un objetivo clave tanto en la investigación experimental como en la observacional es minimizar todas las fuentes de variabilidad salvo aquella cuyos efectos están siendo observados. Las diferencias entre los resultados de aquellos con y sin la exposición se pueden atribuir a la exposición sólo cuando se han logrado controlar adecuadamente todas las fuentes de variabilidad.

## ***Ciencias experimentales***

Las ciencias experimentales minimizan la variabilidad no deseada controlando los factores relevantes a través del diseño experimental. Las oportunidades para el control que se encuentran en la experimentación de laboratorio son una de las razones de su poder y su éxito en obtener hallazgos repetibles. Por ejemplo, los experimentos de laboratorio pueden usar cultivos de tejidos o animales de laboratorio de la misma cepa genética y mantener idéntica temperatura, iluminación, manejo, alojamiento, alimentación, y demás. Dado que no todas las fuentes de variabilidad pueden ser controladas, los experimentos también emplean grupos control o condiciones que reflejan la influencia de factores que el investigador no puede controlar. La comparación de las condiciones experimentales y de control permite al investigador controlar analíticamente los efectos de las influencias no deseadas.

Los experimentos pueden lograr un alto nivel de seguridad de la equivalencia de las condiciones experimentales y de control con respecto a todas las influencias que no sean la exposición de interés porque pueden manipular el objeto de estudio. El experimentador puede hacer una comparación antes y después midiendo el resultado antes y después de implementar una “exposición”. Cuando es importante controlar los cambios que ocurren con el tiempo (envejecimiento), se puede usar un grupo control. Con la asignación aleatoria de la exposición, se puede estimar la probabilidad de cualquier diferencia entre los grupos experimental y control y puede hacerse tan pequeña como uno quiera aleatorizando un gran número de participantes. Si la exposición no tiene efectos persistentes, se puede usar un diseño cruzado en que la exposición es aplicada primero a la mitad de los participantes seleccionados al azar y luego a la otra mitad. La comparación antes y después controla las diferencias entre los grupos, y la comparación entre los grupos controla los cambios que ocurren a lo largo del tiempo. Si las mediciones pueden llevarse a cabo sin conocimiento del estado de exposición, los efectos del observador pueden también ser reducidos. Con suficiente control se puede lograr un acercamiento al ideal, la comparación contrafactual.

## ***Grupos de comparación***

En epidemiología los estudios antes y después y los cruzados no son comunes, en parte porque la exposición a menudo no puede ser manipulada por el investigador; por un lado por el tiempo

prolongado que requieren los procesos en estudio; por otro lado, también porque la exposición, el proceso de observación, o ambos, a menudo tienen efectos persistentes a largo plazo. La aproximación habitual a una comparación contrafactual utiliza un grupo de comparación, a menudo llamado “grupo control” en forma análoga con el modelo experimental. El grupo de comparación sirve de sustituto para el “grupo expuesto sin la exposición” contrafactual. Así, el hecho de que el grupo control sea adecuado depende de su capacidad para producir una estimación precisa de qué resultados hubieran habido en el grupo de exposición en ausencia de la exposición.

## ***Ensayos aleatorios***

El diseño de estudio epidemiológico que se acerca más al modelo experimental es el ensayo aleatorizado de gran dimensión. Sin embargo, el grado de control que se puede ejercer sobre seres humanos es muchísimo menor al que se puede ejercer sobre cultivos de células. Por ejemplo, consideremos el Estudio de la Salud de los Médicos (Physicians Health Study) en que el Dr. Charles Hennekens y sus colegas de la Universidad de Harvard incluyeron médicos norteamericanos (incluyendo varios profesores de mi departamento) en un ensayo para ver si la aspirina y/o los beta-carotenos disminuyen el riesgo de un infarto agudo de miocardio y/o un cáncer. El estudio empleó un diseño factorial en que se solicitaba a los médicos participantes que tomaran diferentes píldoras en días alternativos. Un grupo de médicos alternaba entre aspirina y beta-carotenos; otro grupo alternaba entre aspirina y un placebo diseñado para parecerse a la cápsula de beta-caroteno; el tercer grupo alternaba entre un comprimido que se parecía a la aspirina y beta-carotenos; y el cuarto grupo alternaba entre los dos placebos. De esta manera los investigadores podían estudiar los efectos de cada sustancia, sola y con la otra – dos experimentos separados llevados a cabo simultáneamente.

Con 20,000 participantes, este diseño de estudio aseguraba que los cuatro grupos eran virtualmente idénticos en términos de las características basales. Pero claramente había menos control sobre los médicos durante el período de seguimiento que lo que hubiera sido posible, por ejemplo, con ratas de laboratorio. Los médicos participantes podrían haber aumentado sus niveles de ejercicio, cambiado sus dietas, empezado meditación, o haber hecho otros cambios que pudieran afectar sus riesgos para la enfermedad. Dichos cambios pueden hacer que un estudio no produzca información útil.

## ***La debacle de MRFIT***

Una situación desafortunada del estilo de lo descrito en el párrafo anterior aparentemente se desarrolló en el Ensayo de Intervención sobre Múltiples Factores de Riesgo [Multiple Risk Factor Intervention Trial (MRFIT)] un estudio de gran envergadura (12,000 participantes, más de US\$100 millones de dólares) auspiciado por el Instituto Nacional Norteamericano de Corazón, Pulmón y Sangre (NHLBI) de los Institutos Nacionales Norteamericanos de Salud (U.S. National Institutes of Health, NIH). Al aumentar la evidencia de que la colesterolemia era un factor de riesgo etiológico para múltiples formas de enfermedad cardiovascular, particularmente la enfermedad cardíaca coronaria, en el ensayo se exploró intensamente la posibilidad de que la disminución de los niveles de colesterol disminuiría la magnitud de enfermedad cardíaca coronaria. Sin embargo, a fines de los años 60 no habían medicamentos disponibles; la única intervención que disminuía el nivel de

colesterol era la modificación de la dieta. Un ensayo de “dieta cardiaca” exigiría más de un millón de participantes y duraría muchos años – un escenario poco atrayente.

Sin embargo, la idea de un ensayo de dieta cardiaca perduró, transformándose en una investigación para verificar que las tasas de enfermedad cardiovascular podían ser disminuidas modificando los tres factores de riesgo cardiovascular más comunes: el hábito de fumar, el colesterol sérico elevado, y la hipertensión arterial. Así nació MRFIT.

El ensayo fue lanzado a principios de la década de los 70. Se seleccionaron hombres (porque tienen tasas de enfermedad cardiovascular más altas) con factores de riesgo que los colocaba en alto riesgo de sufrir enfermedad cardiovascular y fueron asignados aleatoriamente a un grupo de “Intervención Especial” (IE) o a un grupo de Atención Habitual (AH). Los participantes de IE recibían una intervención intensiva, de vanguardia, basada en teorías del momento, para mejorar la dieta y promover el abandono del hábito de fumar. Los pacientes hipertensos en el grupo de IE eran tratados con un protocolo sistematizado para controlar su presión arterial. Los participantes de AH eran examinados, enviándose un resumen a sus médicos personales, pero no recibían ningún tratamiento a través del estudio. En esta era pre-“saludable” (promoción de salud / prevención de la enfermedad a través de los cambios de conducta), los diseñadores del ensayo proyectaron cambios modestos en los factores de riesgo de los participantes del grupo IE, y casi ningún cambio en los participantes del grupo de AH. Aunque los médicos personales de los participantes de AH recibieran copias de los resultados de los exámenes practicados a los pacientes, en aquellos años no era habitual que los médicos practicantes se involucraran en cambios dietéticos, cese del hábito de fumar, o aún control de la presión arterial en los pacientes sanos.

El tamaño muestral planificado de unos 12,000 hombres, aproximadamente 6,000 en IE y 6,000 en AH, se logró y se mantuvo en seguimiento durante siete años. Al final del período de seguimiento, los niveles de los factores de riesgo en el grupo de IE habían alcanzado los niveles esperados, y 46% de los fumadores habían dejado de fumar. Pero para la sorpresa (y la consternación) de los investigadores de MRFIT, los niveles de colesterol y la presión sanguínea también disminuyeron entre los participantes de AH, y 29% de los fumadores de AH habían abandonado el hábito. Durante los años del ensayo, la dieta, la HTA y el hábito de fumar se habían incorporado a las prioridades de atención de los médicos y del público (presumiblemente ayudado por otra iniciativa del NHLBI, El Programa Nacional Norteamericano de Control de Presión Arterial Elevada). La mortalidad entre los participantes AH no sólo era mucho menor que la proyectada por el estudio de Framingham sino que era hasta (levemente) menor que la de los participantes IE. Obviamente, hubieron muchos epidemiólogos incómodos cuando los resultados salieron a la luz.

## ***Estudios no aleatorizados***

La mayor parte de los estudios epidemiológicos no tienen oportunidad de comparar grupos formados por un procedimiento de asignación aleatoria. Sea que estudiemos el hábito de fumar, de beber alcohol, posesión de arma de fuego, alimentación, ejercicio, sobrepeso, uso de una medicación particular, exposición a agentes tóxicos, colesterolemia, presión arterial, contaminación del aire, o lo que sea, no hay ninguna seguridad de que el grupo de comparación (los participantes no expuestos) sea igual al de los participantes expuestos salvo por la exposición en estudio. De hecho, lo opuesto

es más probable, dado que todo tipo de factores están relacionados con los ámbitos familiares y físicos, la ocupación (p.ej., exposiciones laborales), estilos de vida (p.ej., nutrición, actividades físicas), influencias sociales (p.ej., apoyo social, injusticias), atención de salud, condiciones de salud (p.ej., medicaciones), dotación genética, y otras características.

## ***El fenómeno de confusión***

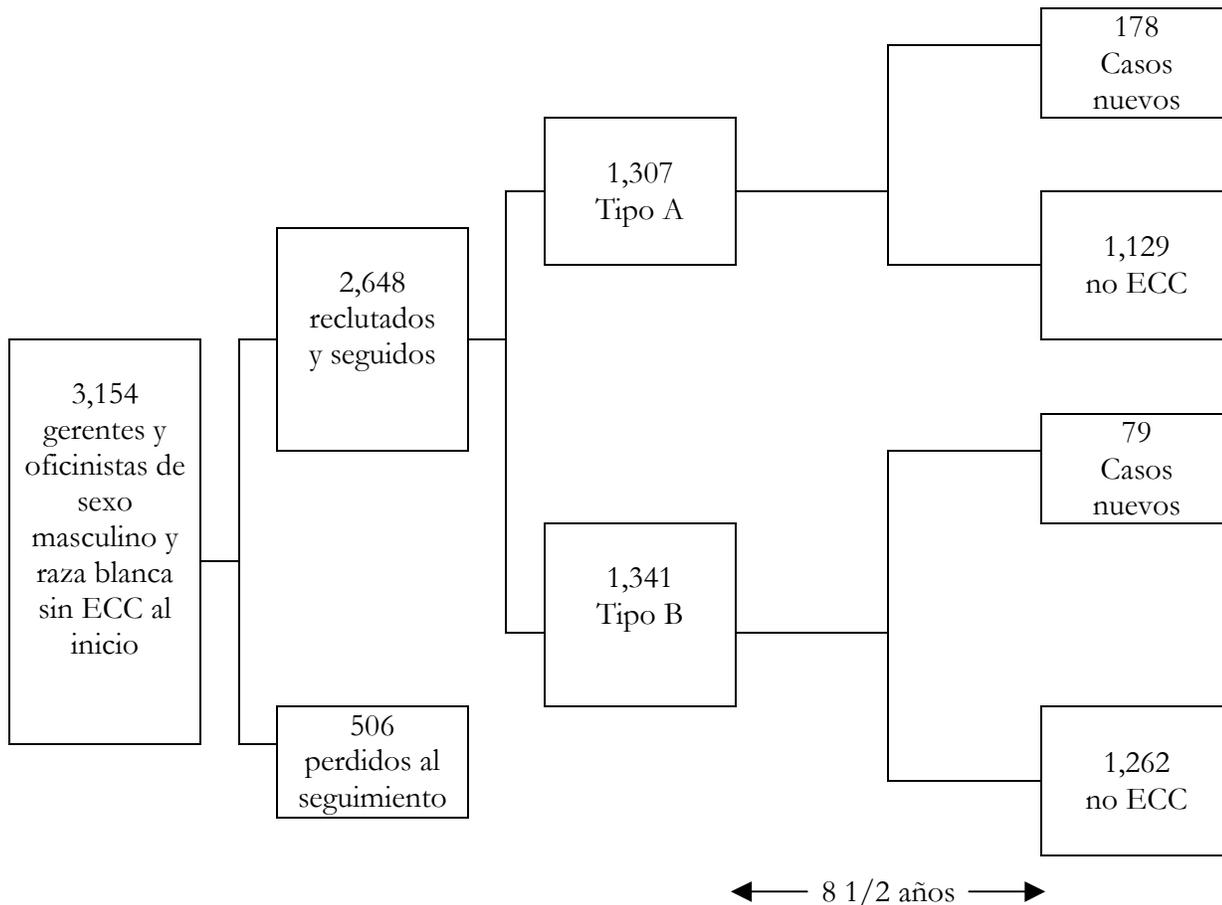
Así, siempre que comparemos grupos con respecto a los factores de interés, debemos tener en consideración que diferencias entre los grupos con respecto a otros factores “extraños” pueden ser responsables de lo que observamos (o no observamos) (factores extraños = factores diferentes a los involucrados en la relación en estudio). ***El fenómeno de confusión*** (en inglés, ***confounding***) (del latín, *confundere*, mezclar) puede ser definido como una “situación en que la medida del efecto de una exposición sobre el riesgo es distorsionada por la asociación de la exposición con otro factor o factores que influyen sobre el resultado en estudio” (Last, *A Dictionary of Epidemiology*). El fenómeno de confusión es un problema de comparación, un problema que surge cuando factores extraños pero importantes, se distribuyen en forma desigual en los grupos que están siendo comparados. El punto central del concepto del fenómeno de confusión y su control en epidemiología deriva de las limitadas oportunidades para el control experimental.

## ***Un ejemplo hipotético [con disculpas al Grupo de Estudio Colaborativo del Oeste (Western Collaborative Group Study)]***

Para investigar como surge el fenómeno de confusión y como se puede manejar, consideremos los siguientes datos hipotéticos basados en el estudio del riesgo de enfermedad cardíaca coronaria (ECC) en gerentes y oficinistas que tienen un patrón de comportamiento propenso a ECC del Grupo de Estudio Colaborativo del Oeste. Este patrón, a menudo llamado de tipo A de comportamiento, es descrito como exigente, apurado e hiper-agresivo. En contraste las personas Tipo B se consideran más relajadas y de trato fácil.

En este estudio, Meyer Friedman, Raymond Rosenman, y sus colegas reclutaron 3,154 gerentes masculinos de raza blanca, de entre 35 y 59 años, empleados en 10 compañías de California. Los hombres fueron sometidos a exámenes médicos para detectar ECC y una entrevista estandarizada, estructurada, para determinar su patrón de comportamiento. El tipo de comportamiento era determinado revisando videos grabados de las entrevistas. Los 2,648 participantes considerados libres de ECC al inicio fueron seguidos con exámenes físicos anuales para la detección de nuevos casos de ECC durante los siguientes 8 ½ años. Los resultados (verdaderos) del estudio se presentan en el siguiente esquema y son tabulados en la Tabla 1.

**Figura 1**  
**Seguimiento de 8 ½ años del Estudio del Grupo Colaborativo del Oeste**



**Tabla 1**  
**Relación de la ECC con el tipo de comportamiento**

	Tipo de comportamiento		
	A	B	Total
Casos de ECC	178	79	257
No manifiestan ECC	1,129	1,262	2,391
Total	1,307	1,341	2,648

Dado que estos datos provienen de un estudio de cohortes, deben ser analizados estimando la incidencia de la ECC para los grupos de Tipo A y Tipo B. Aunque el período de riesgo para la ECC se extiende mucho más allá del período de seguimiento, usaremos la incidencia acumulada (IA) para simplificar. En estos datos, la IA es  $178/1307 = 0.14$  para el grupo de Tipo A, y  $79/1341 = 0.06$  para el grupo de Tipo B. El riesgo relativo (razón de riesgos, razón de incidencias acumuladas) es por lo tanto  $0.14/0.06 = 2.3$ .

## **Preguntas a hacer:**

Hay muchos aspectos del diseño y la ejecución de este proyecto sobre los cuales quisiéramos tener más información. Por ejemplo:

- ¿Cuáles fueron los criterios para clasificar los participantes como de Tipo A o de Tipo B?
- ¿Cuántos participantes se perdieron al seguimiento?
- ¿Cómo se define y diagnostica la ECC?
- ¿Sabían los médicos que determinaban que un paciente era un caso a qué grupo de comportamiento pertenecía el paciente, o no?

Pero dado que nuestro tema hoy es el fenómeno de confusión, nuestro interés principal es la pregunta:

Los grupos de Tipo A y de Tipo B, ¿difieren en otros factores que pudieran haber afectado sus tasas observadas de ECC?

O, en forma equivalente,

¿Existen factores distintos al patrón de comportamiento que pueden ser responsables de las tasas observadas?

(Puede ser agregado aquí que la misma pregunta sería relevante se hubiera observado o no una diferencia entre el Tipo A y el Tipo B.

## **Escenario hipotético**

Probablemente la mayoría de ustedes saben que en el Grupo de Estudio Colaborativo del Oeste, no habían otros factores que explicaran las diferencias en la incidencia de ECC entre los grupos de tipo A y de Tipo B. Así que ahora nos alejaremos del verdadero estudio para crear un escenario en que la diferencia entre las incidencias observadas para los participantes del Tipo A y del Tipo B se debe a diferencias en el hábito de fumar de los participantes.

Supongamos que hubiéramos obtenido los datos de la Tabla 1. ¿Cómo hacemos para ver si la incidencia entre los grupos de Tipo A y de Tipo B deben ser atribuidas a las diferencias en el hábito de fumar más que al tipo de comportamiento? El enfoque tradicional y más habitual para contestar esta pregunta es la estratificación de los datos por la condición de fumadores o no de los participantes. La Tabla 2 muestra los resultados de dicho análisis estratificado (con datos hipotéticos).

**Tabla 2**  
**Relación de ECC con el Patrón de Comportamiento,**  
**Análisis Estratificado controlando por el hábito de fumar [DATOS HIPOTÉTICOS ]**

	Fumadores		No Fumadores	
	Tipo A	Tipo B	Tipo A	Tipo B
ECC	168	34	10	45
<hr/> ECC	880	177	249	1,085
Total	1,048	211	259	1,130

Esta tabla muestra la relación entre tipo de comportamiento y ECC, estratificada por la experiencia de fumar. Podemos ahora calcular la incidencia (acumulada) de ECC entre los no fumadores de Tipo A y compararlos con la de los no fumadores de Tipo B, lo cual nos dirá el efecto del tipo de comportamiento cuando el hábito de fumar no puede ser responsable de los resultados (sin contar el humo de tabaco ambiental). También podemos ver la incidencia para los fumadores de Tipo A y los fumadores de Tipo B, cuando hemos creado (de cierta manera) grupos que son más comparables.

¿Qué vemos cuando hacemos estos cálculos? La incidencia de ECC entre los no fumadores de Tipo A es  $10/259 = 0.04$ , exactamente igual que entre los no fumadores de Tipo B ( $45/1130 = 0.04$ ). Por lo tanto llegamos a la conclusión de que, por lo menos entre los no fumadores, el patrón de comportamiento no influyó. De manera similar, la incidencia acumulada es la misma (0.16) para los fumadores de Tipo A y los fumadores de Tipo B. De nuevo, el patrón de comportamiento no influyó. Aparentemente, el hábito de fumar es el que sí influye de manera importante. Esta variable clave “extraña” aparentemente estaba distribuida en forma muy desigual entre los dos grupos de tipos de comportamiento y nos llevó a observar una diferencia que casi atribuimos a un patrón de comportamiento.

### ***El fenómeno de confusión – una discrepancia entre los valores crudos y los controlados***

Este ejemplo ilustra el fenómeno de confusión. En la tabla sin control o “cruda”, vimos una asociación (razón de incidencias acumuladas de 2.3). Cuando controlamos para el hábito de fumar (que supondremos por ahora que es la única variable extraña relevante), encontramos que no había asociación (RIA de 1.0) entre nuestro factor de estudio (patrón de comportamiento) y el daño (ECC). Esta discrepancia entre la RIA cruda (2.3) y las RIA específicas por estrato (1.0) indica que hay confusión por estado de fumador. La estratificación es un método para controlar el efecto de confusión por el hábito de fumar. [Permítanme enfatizar aquí que el ejemplo anterior **no** es de la vida real. En el estudio de verdad de Friedman y Rosenman, el comportamiento de tipo A se asociaba con ECC aún cuando se controlaban los efectos de fumar.] También vale la pena mencionar que la confusión también puede ocurrir en forma inversa, es decir, podríamos no ver asociación en el análisis crudo pero encontrar una cuando hacemos el análisis estratificado. De manera que la confusión puede crear una asociación aparente o enmascarar una verdadera.

## **El fenómeno de confusión surge de la distribución desigual de un factor de riesgo**

¿Cómo puede ocurrir el fenómeno de confusión? Como indicamos antes, las condiciones necesarias para producir confusión (en esta situación algo simplificada) son que un verdadero factor de riesgo para el daño se distribuya en forma desigual entre los grupos que se comparan. Para ver esto en el ejemplo anterior, he reorganizado las columnas de la Tabla 2. Este arreglo enfatiza que la mayor parte de los participantes de Tipo A eran fumadores y la mayor parte de los de Tipo B no lo eran.

**Tabla 3**  
**Relación entre ECC, Patrón de comportamiento, y hábito de fumar**  
**[DATOS HIPOTÉTICOS]**

	Patrón de comportamiento Tipo A			Patrón de comportamiento Tipo B			Ambos Gran total
	Fumadores	No fumadores	<b>Total</b>	Fumadores	No fumadores	<b>Total</b>	
ECC	168	10	<b>178</b>	34	45	<b>79</b>	257
<u>ECC</u>	<u>880</u>	<u>249</u>	<u><b>1,129</b></u>	<u>177</u>	<u>1,085</u>	<u><b>1,262</b></u>	2,391
Total	1,048	259	<b>1,307</b>	211	1,130	<b>1,341</b>	2,648

Aunque esta tabla fue creada redistribuyendo las columnas de la Tabla 2, puede ser más interesante pensar que nos está mostrando la historia por detrás de los datos no controlados (crudos) de la Tabla 1. Señalemos que la Tabla 1 está incluida en esta tabla como los marginales para cada una de las sub-tablas (las columnas en negritas). Las sub-tablas muestran como están compuestos los grupos de Tipo A y de Tipo B. Claramente la mayor parte ( $1048/1307 = 80\%$ ) de los participantes del grupo A son fumadores mientras que la mayor parte de los participantes del grupo B no son fumadores. Con esta desigualdad tan marcada, no es sorprendente que un factor de riesgo como el hábito de fumar, pueda distorsionar la asociación global (no controlada). Los atributos de un factor de confusión son entonces, que sea un factor de riesgo independiente para el daño y que se asocie con el factor de riesgo en estudio.

## **El fenómeno de confusión – atribución equivocada de una asociación observada**

El exceso de casos del grupo de Tipo A se debe, claramente, a la mayor proporción de fumadores en el grupo de Tipo A que en el de Tipo B. Si nos hubiésemos quedado con el valor crudo, hubiéramos atribuido la diferencia entre los grupos equivocadamente al patrón de comportamiento en vez de al hábito de fumar. El fenómeno de confusión puede ser definido como una distorsión en la medida de asociación debido a la distribución desigual de un determinante del daño.

Señalemos, sin embargo, que la asociación cruda de cualquier manera es “real”. Los participantes de Tipo A *tenían* una mayor incidencia de ECC. El fenómeno de confusión surge cuando *atribuimos* la incidencia elevada al hecho de ser de Tipo A, dado que la mayor incidencia se debe en realidad al hecho de que tienen más fumadores (en este ejemplo). Pero los hombres Tipo A como grupo tenían una incidencia mayor de ECC. Hay situaciones en que sigue siendo importante considerar las asociaciones crudas.

### **Otra perspectiva –promedios ponderados**

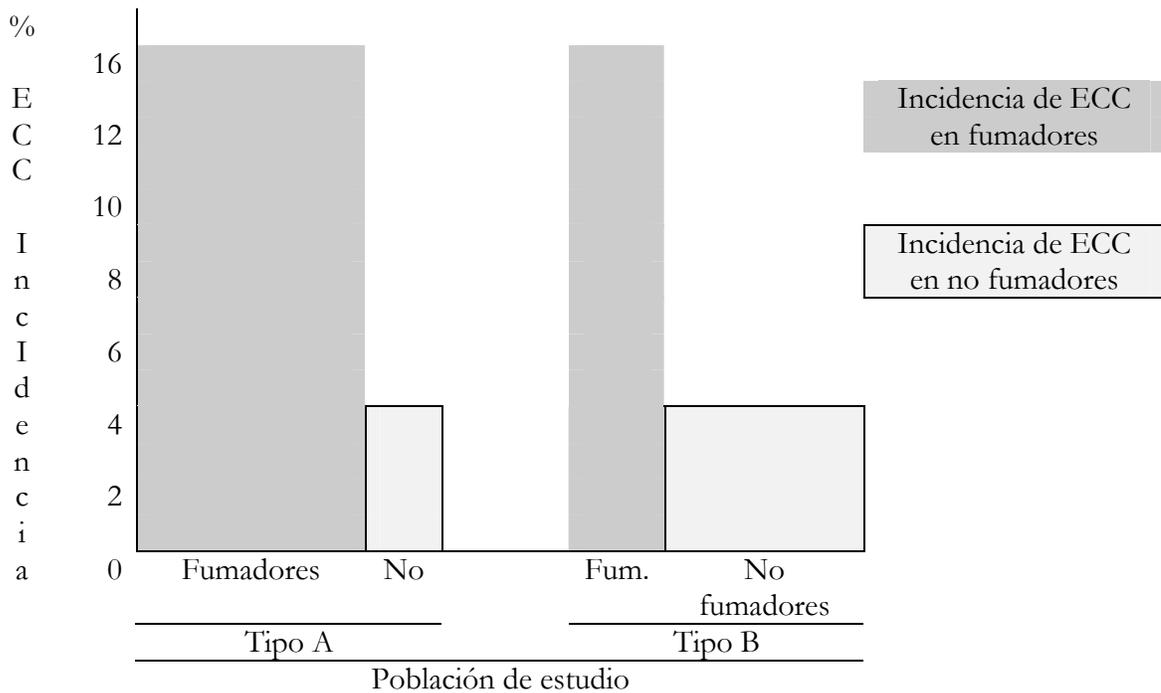
Una tabla resumen enfatiza las incidencias y hace muy evidente el patrón.

**Tabla 4**  
**Incidencia de ECC por Tipo de Comportamiento y según hábito de fumar**  
**[DATOS HIPOTÉTICOS]**

Patrón de comportamiento	Hábito de fumar		Total	
	Fumador	No fumador		
Tipo A	0.16	0.04	0.14	← (incidencias ← obtenido de la tabla 1)
Tipo B	0.16	0.04	0.06	
Total	0.16	0.04		

Aquí se ve claramente que cuando mantenemos constante el hábito de fumar (i.e., miramos hacia abajo en cualquiera de las dos primeras columnas de incidencias), no hay ningún efecto del tipo de comportamiento. Cuando mantenemos constante el tipo de comportamiento (i.e., leemos horizontalmente cualquiera de las dos primeras filas), vemos que el hábito de fumar se asocia con un aumento de la incidencia de cuatro veces. Los marginales de la tabla son, en efecto, promedios ponderados de las incidencias al interior de la tabla. Las incidencias de la última fila son las mismas que al interior de la tabla – tienen que serlo, porque un promedio ponderado de dos cifras idénticas es siempre la misma cifra. Sin embargo, las incidencias en la columna de la extrema derecha, pueden ser cualquier cifra entre 0.16 y 0.04 – dependiendo de la ponderación utilizada en promediar 0.16 y 0.04. Estos conceptos se pueden ver gráficamente.

**Incidenca de ECC según Patrón de Comportamiento y por hábito de fumar  
[HIPOTÉTICO]**



Como lo muestra el diagrama, la población de estudio puede considerarse como consistiendo de cuatro subgrupos diferentes, cada uno con una combinación distinta de tipo de comportamiento y hábito de fumar. Si estos fueran los únicos subgrupos relevantes, las tasas de incidencia para cada uno representarían el indiscutible estado “verdadero” en la población de estudio. La tasa en la población de estudio en conjunto y para cualquier grupo contenido en ella, p.ej., todos los Tipos A, puede ser considerada un promedio ponderado de las incidencias en los subgrupos componentes, donde los pesos son los tamaños proporcionales de los subgrupos componentes. Así la tasa en los de Tipo A es:

$$0.14 = \frac{178}{1,307} = \frac{1,048}{1,307} \times \frac{168}{1,048} + \frac{259}{1,307} \times \frac{10}{259}$$

o simbólicamente,

$$IA_{ECC|A} = P_{F|A} \times IA_{FA} + \bar{P}_{F|A} \times \bar{IA}_{FA}$$

donde:

IA es incidencia (acumulada)  
P es prevalencia o proporción

F indica fumadores ( $\bar{F}$  indica no fumador)

A indica comportamiento Tipo A

y la notación  $F|A$  significa “fumadores entre (o dado) el comportamiento de Tipo A”.

## ***El fenómeno de confusión – comparación de promedios ponderados usando distintos pesos***

La incidencia para cualquier grupo (p.ej., los de Tipo A) puede variar desde la incidencia menor de cualquiera de sus subgrupos (p.ej., Tipo A) hasta la incidencia más alta de cualquier subgrupo (p.ej., fumador Tipo A). El lugar, en este rango donde cae la incidencia global del grupo, es determinado por el tamaño de cada subgrupo (fumadores Tipo A, no fumadores Tipo A) como proporción del grupo total (todos los de tipo A). Puede resultar un fenómeno de confusión cuando estas proporciones difieren para los grupos que son comparados.

Dado que hay muchas maneras posibles en que estas proporciones pueden ser diferentes, el fenómeno de confusión puede hacer que una medida de asociación cruda (global) sobreestime, subestime, disimule por completo, o hasta invierta la asociación que se apreciaría en comparaciones llevadas a cabo dentro de los subgrupos. Como ejemplo conocido, si dos poblaciones tienen diferentes distribuciones por edad, una comparación de sus tasas de mortalidad totales (crudas) puede sobreestimar o subestimar la imagen que se apreciaría comparando dentro de grupos específicos por edad, aún al extremo de que la comparación de tasas crudas aparente favorecer la población que tienen las tasas más altas (peores) dentro de cada estrato por edad. La estandarización por edad es un caso especial de la estrategia más general llamada análisis estratificado, que es un recurso fundamental para controlar el fenómeno de confusión.

### ***Los límites del fenómeno de confusión***

Hay límites sobre la fuerza de la asociación (secundaria) que puede ser producida por el fenómeno de confusión. Por ejemplo, dados los datos de la Tabla 1, un fuerte efecto del hábito de fumar y un desequilibrio muy llamativo entre los dos grupos de tipo de comportamiento fue necesario para que el hábito de fumar justificara completamente el efecto aparente del comportamiento de Tipo A. Esa es una de las razones por las cuales la fuerza de asociación es un criterio de inferencia causal. Cuanto más fuerte la asociación observada entre la enfermedad y el factor estudiado, menos probable es que un factor completamente extraño pueda justificar toda la asociación observada.

### ***Estudios caso-control***

Hasta ahora en nuestra discusión nos hemos limitado a los estudios de tipo cohorte. Cuando enfocamos el tema del fenómeno de confusión en los estudios caso-control, hay unas complejidades adicionales a consecuencia de la manera en que la población de origen es representada en la población de estudio. Para comprender las características del fenómeno de confusión en un estudio caso-control, generemos un estudio de ese tipo a partir de la cohorte que consideramos anteriormente.

La cohorte original consistía de 2,648 individuos con un seguimiento completo y produjo 257 casos. En forma ideal, nuestro estudio caso-control detectaría todos los casos incidentes y una muestra de los no casos a medida que ocurrieran los casos (lo que se llama “muestreo de densidad”). Para simplificar nuestra ilustración, sin embargo, saquemos la muestra de controles de aquellos individuos que eran libres de ECC al final del período de seguimiento. La siguiente tabla muestra los

mismos casos, con la distribución de controles que se esperaría si se obtuviera una muestra representativa de los no casos, de doble el tamaño del grupo de casos (i.e., supongamos 514 controles con la misma proporción de personas Tipo A y fumadores que se encuentra en los no casos del estudio de cohortes). (Los números en la fila “No ECC” se obtienen multiplicando la fila “No ECC” en la Tabla 1 (i.e., todos los no casos) por 514/2391 (0.21) de manera que los 2,391 no casos se convierten en 514 controles. De esta manera, los 1,129 sujetos Tipo A sin enfermedad ECC manifiesta se convierten en 243 controles Tipo A, y los 1,262 sujetos de Tipo B sin ECC manifiesta se convierten en 271 controles Tipo B.)

**Tabla 5**  
**Resultados Esperados del Estudio Caso-Control [HIPOTÉTICO]**

	Patrón de Comportamiento			
	Tipo A	Tipo B	Total	
Casos ECC	178	79	257	
ECC no manifiesta	243	271	514	← Esta fila es simplemente 0.21 veces la fila correspondiente de la Tabla 1
Total	421	350	771	

El odds ratio para esta tabla es [2.5], algo mayor que la razón de riesgos en el estudio de cohortes. [La diferencia entre el odds ratio y la razón de riesgos refleja la incidencia de ECC en la cohorte – cuanto menor la incidencia, más se acercaría el odds ratio a la razón de riesgos.]

Ahora crearemos, de la misma manera, la tabla esperada para el hábito de fumar y el patrón de comportamiento en un análisis estratificado:

**Tabla 6**  
**Resultados esperados para el Estudio Caso-Control, Estratificado por hábito de fumar [HIPOTÉTICO]**

	Fumadores		No Fumadores		
	Tipo A	Tipo B	Tipo A	Tipo B	
ECC	168	34	10	45	
ECC	189	38	54	233	← Esta fila es simplemente 0.21 veces la fila correspondiente en Tabla 2.
Total	357	72	64	278	

El odds ratio de cada tabla es 1.0, de manera que de nuevo encontramos presente el fenómeno de confusión. De nuevo aquí vemos que el factor de confusión está asociado con el resultado: el odds ratio para el hábito de fumar y ECC en el grupo Tipo B es 4.6. También encontramos que el fumar

está asociado con el tipo de comportamiento: la proporción de fumadores entre los no casos de Tipo A es de 0.78 mientras que en los no casos de Tipo B es de solamente 0.14 [verifica estos números].

La razón del énfasis planteado sobre las asociaciones *condicionales* (“en grupo de Tipo B”, “entre no casos”) más que sobre las asociaciones crudas o *no condicionales* es que una variable de confusión debe estar asociada con la exposición en estudio en la población de la cual surgieron los casos (ver Rothman y Greenland). Es el grupo control que provee la estimación de la prevalencia de la exposición en la población de origen. También, en un estudio caso-control, los totales para los distintos grupos de exposición (p.ej., el total de los fumadores Tipo A) no son cantidades que tengan mucho sentido, por lo menos con el propósito de comparación. La razón es que las relaciones entre estos totales reflejan fundamentalmente la razón (arbitraria) de casos a controles. De manera que la asociación de exposición que es relevante para el fenómeno de confusión en un estudio caso control es la asociación entre la exposición y el potencial factor de confusión entre los controles.

La razón para no morar dentro del grupo de Tipo A es que la asociación en este grupo podría reflejar la modificación del efecto entre la exposición (comportamiento Tipo A) y la covariable, más que la confusión como tal. Profundizaremos sobre este tema cuando veamos la modificación del efecto, en el próximo capítulo.

## ***El fenómeno de confusión – una característica de la base de estudio***

Hemos dicho que el fenómeno de confusión requiere dos asociaciones: (1) el factor de confusión debe ser un factor de riesgo para el daño o su detección y (2) el factor de confusión debe estar asociado con la exposición. Esta última asociación debe existir dentro de la base de estudio (ver Rothman y Greenland). Vale la pena elaborar sobre este punto.

### **Estudios de seguimiento**

En un estudio de seguimiento, la base de estudio, de la cual provienen los casos, es simplemente la población en seguimiento, la población de estudio. Para que ocurra el fenómeno de confusión, el potencial factor de confusión y la exposición deben estar asociados en esta población. La asignación aleatoria de una intervención tiende a distribuir los potenciales factores de confusión en forma uniforme entre los grupos de intervención y de control. Si la asignación al azar es exitosa, ninguna variable extraña se asocia con la intervención por lo cual el fenómeno de confusión no podrá ocurrir. Sin embargo, si la aleatorización no es “efectiva”, de manera que existe una falta de balance para un potencial factor de confusión, entonces puede ocurrir confusión con respecto a ese potencial factor de confusión. Cuanto mayor el número de participantes, menos probable es que ocurra una falta de balance por azar.

### **Estudios caso-control**

En un estudio caso-control, la base de estudio es la población subyacente que está siendo seguida a través de la ventana del diseño caso-control. Para que ocurra el fenómeno de confusión, la exposición y el potencial factor de confusión (factor de riesgo) deben estar asociados en esa

población subyacente (población fuente de la cual surgen los casos). Pero dado que el investigador observa esa población sólo indirectamente, el asunto es más complicado. Sin embargo, si no hay asociación entre el potencial factor de confusión y la exposición en la población base del estudio, entonces el fenómeno de confusión no ocurre aún cuando encontramos que el potencial factor de confusión y el de exposición están asociados dentro del grupo control de nuestro estudio caso-control (Miettinen y Cook, citados en Rothman, página 93).

Este resultado algo sorprendente se demuestra fácilmente. Supongamos que estamos observando una población a través del tiempo para examinar una asociación entre un carcinógeno ocupacional sospechoso y un cáncer que también está fuertemente relacionado (RIA=10) con el hábito de fumar cigarrillos. Supongamos también que la exposición ocupacional es de hecho un carcinógeno y que en esta población el hábito de fumar no está asociado con la exposición ocupacional. Si suponemos una tasa básica de 3 casos/1,000 años-persona y una RIA de 3.3 para el carcinógeno ocupacional, el seguimiento de la población podría producir la siguiente tabla.

**Tasas de incidencia, tamaño poblacional y número de casos  
para los datos hipotéticos sobre la exposición ocupacional y el hábito de fumar**

	Fumadores		No fumadores	
	Expuestos	No expuestos	Expuestos	No expuestos
	(1)	(2)	(3)	(4)
1. Número de casos	300	90	70	21
2. Tamaño poblacional (personas-año)	3,000	3,000	7,000	7,000
3. Densidad de Incidencia por 1,000 a-p	100	30	10	3
	RDI = 3.3		RDI = 3.3	

Con una observación hipotética de 7,000 años-persona para los no fumadores que además no están expuestos al carcinógeno, la tasa de incidencia de base prevista de 3/1,000 a-p se espera que se produzcan unos 21 casos incidentes. Si la cantidad de tiempo-personas entre los no fumadores expuestos es también de 7,000 a-p, entonces esperaríamos  $3.3 \times 3/1,000 \text{ a-p} \times 7,000 \text{ a-p} \approx 21$  casos para ese grupo. Si el tiempo-persona para los fumadores expuestos y no expuestos es de 3,000 a-p para cada grupo, esperaríamos 300 y 90 casos incidentes, respectivamente, si la RDI para la exposición ocupacional es la misma entre no fumadores y fumadores y la RDI para fumadores es 10 no importa si tiene o no la exposición ocupacional.

Señalemos que esta población hipotética ha sido construida de manera que la proporción de años-personas expuestos es de 50% entre los fumadores (columnas 1 y 2), entre no fumadores (columnas 3 y 4), y globalmente, i.e., no hay asociación entre el hábito de fumar y la exposición ocupacional. De igual manera, las proporciones de años-personas para los fumadores entre los expuestos (columnas 1 y 3) y los no expuestos (columnas 2 y 4) son de 30% en cada caso ( $3,000/[7,000+3,000]$ ). La RDI

cruda para el carcinógeno ocupacional es por lo tanto 3.3 (asegúrate de poder derivar esta RDI), que es idéntica a la RDI para la exposición entre fumadores y no fumadores. De esta manera, el fenómeno de confusión no está presente.

Supongamos ahora que llevamos a cabo un estudio caso-control en esta población durante el mismo período de tiempo. Si hay un registro de cáncer podríamos identificar e incluir la totalidad de los 481 casos (ver la fila 1 en la siguiente tabla, que es idéntica a la fila 1 de la tabla anterior). Si obtenemos una muestra representativa de la población de 5% de la misma como nuestro grupo control, entonces la distribución del hábito de fumar y del carcinógeno ocupacional en nuestro grupo control (la fila 2 en la siguiente tabla) será igual que la distribución de estas variables en la fila de tiempo-población de la fila 2 de la tabla precedente (30% de fumadores y 50% expuesto al carcinógeno ocupacional, sin asociación entre los dos). El OR (asegúrate que puedes calcularlo) será idéntico a la RDI de 3.3, ya vista. En este estudio caso-control con un grupo control (**no sesgado**) que es directamente proporcional a la población de origen, no hay fenómeno de confusión.

### Distintos grupos control para un estudio caso-control hipotético sobre exposición ocupacional y el hábito de fumar

Fila	Fumadores		No fumadores	
	Expuestos	No expuestos	Expuestos	No expuestos
#	(1)	(2)	(3)	(4)
1. Número de casos	300	90	70	21
2. Controles proporcionales	150	150	350	350
	(OR = 3.3)		(OR = 3.3)	
3. Controles sesgados	250	150	250	350
	(OR = 2.0)		(OR = 4.7)	

Supongamos, sin embargo, que los controles son seleccionados en una manera sesgada generando un **grupo control sesgado** (fila 3 de la segunda tabla) en que el hábito de fumar y la exposición **están** asociados (verifica este hecho; intenta, por ejemplo, calcular el OR para el hábito de fumar en relación a la exposición). Reflejando el grupo de control sesgado, las RDI específicas por estrato ya no son 3.3. Sin embargo en este capítulo estamos enfocando la **asociación cruda** y si representa en forma exacta la verdadera situación (que en este caso hemos construido, más que teniendo que considerar las asociaciones estratificadas como la verdadera situación). El OR crudo de la tabla anterior, usando los casos en la fila 1 y los controles de la fila 3, es (trata de calcularlo antes de ver la respuesta)

Así, aún cuando este grupo control sesgado el OR se mantiene sin fenómeno de confusión. Pero, el potencial factor de confusión (el hábito de fumar, un factor de riesgo causal para el daño) de hecho

está asociado con la exposición en los controles (sesgados). [Varias maneras de ver esta asociación son:

El odds de exposición entre los fumadores (cols. 1 y 2) es de  $250/150$ , muy diferente del odds de exposición entre los no fumadores (cols. 3 y 4:  $250/350$ ) produciendo un odds ratio entre el hábito de fumar y la exposición de  $OR = 2.3$ ).

Proporcionalmente más fumadores están expuestos [ $250/(250 + 150) = 0.63$ ] que no fumadores [ $50/(250 + 350) = 0.42$ ].

El odds de fumar entre los expuestos (cols. 1 y 3) es de  $250/250$ , muy distinto al odds de fumar entre los no expuestos (cols 2 y 4) :  $150/350$ ), produciendo por supuesto, el mismo odds ratio,  $2.3$ ).

Proporcionalmente más expuestos son fumadores [ $250/(250+250) = 0.5$ ] que no expuestos lo son [ $150/(150 + 350) = 0.3$ ].

La variable de confusión potencial, el hábito de fumar, también se asocia con el daño en los no expuestos (p.ej., RDI = 30 por mil a-p en la población de base,  $OR = (90 \times 350) / (21 \times 150)$  en el estudio caso-control con cualquiera de los dos grupos de control. Así es posible tener un factor de riesgo que está asociado con la exposición entre los que no son casos y sin embargo que no haya fenómenos de confusión.

Se puede tener una comprensión mayor tomando en cuenta los mecanismos que causan el fenómeno de confusión, como fue ilustrado en el ejemplo de comportamiento de Tipo A. La confusión resulta de una falta de balance entre los grupos de expuestos y no expuestos con respecto a los determinantes de la enfermedad. Si el factor potencial de confusión aumenta el riesgo de la enfermedad y el factor potencial de confusión se asocia con la exposición, entonces la incidencia de la enfermedad en los expuestos se verá aumentada con respecto a la de los no expuestos, debido al factor de confusión. Este aumento desproporcionado en la incidencia, y por lo tanto en los casos, aumentará el odds de la exposición en un grupo representativo de los casos. Si el factor de confusión no es controlado en el análisis, este odds aumentado causará el fenómeno de confusión en la asociación de la exposición y la enfermedad.

El OR (de exposición) para el daño es simplemente la razón del odds de exposición en el grupo de casos dividido por el odds de exposición en el grupo control. El odds de exposición en el grupo de casos obviamente no se ve afectado por algo que le ocurre al grupo control (incluyendo el apareamiento, por cierto). De manera que una distorsión del OR crudo tendrá que venir de un cambio en el odds de exposición en el grupo control. Mientras el sesgo del grupo control no produzca que el odds **crudo** de exposición difiera de aquel del de la población de origen, (p.ej.,  $0.5/0.5 = 1.0$  en nuestro ejemplo del carcinógeno ocupacional), el OR crudo será igual al de la población de origen, i.e., sin elementos de confusión.

En la mayor parte de los estudios caso-control solemos tener poca información independiente sobre la población de origen del estudio, de manera que el grupo control nos muestra una ventana en ella. Si el grupo control está sesgado, nuestra visión de la población de origen estará distorsionada, y podemos concluir que la condición para el fenómeno de confusión se cumple (i.e., un factor de riesgo para la enfermedad se asocia con la exposición en los no casos). Debido a lo sesgado del

grupo control, el control de potenciales factores de confusión introducirá un sesgo en este análisis (p.ej., en el ejemplo anterior, los OR específicos por estrato son diferentes del valor correcto de 3.3). Sin embargo, un promedio ponderado de los OR específicos por edad puede acercarse al valor crudo.

## ***Pruebas estadísticas para el fenómeno de confusión***

Dado que el fenómeno de confusión requiere una asociación entre la variable potencial de confusión y la exposición, los investigadores a veces utilizan pruebas estadísticas para las diferencias en las distribuciones de los potenciales factores de confusión entre los distintos grupos de exposición. Si los grupos no difieren en forma significativa, los investigadores concluyen que el fenómeno de confusión no ocurrirá. Esta práctica a menudo lleva a la conclusión correcta, aunque es algo imprecisa.

Las pruebas estadísticas de significancia están dirigidas a resolver el tema de si hay o no una asociación entre la exposición y los potenciales factores de confusión con mayor probabilidad que la que surgiría sólo por azar. Pero el fenómeno de confusión depende de la magnitud de la asociación (p.ej., odds ratio, razón de prevalencias), más que de la fuerza de la evidencia de que no surgió por el azar. De manera que una diferencia grande pero “no significativa” puede tener más posibilidad de causar el fenómeno de confusión que una diferencia pequeña pero “muy significativa”. Las razones de esta aseveración aparentemente paradójica son que la significancia estadística depende de la magnitud del número de participantes expuestos y no expuestos, de manera que cualquier asociación será estadísticamente significativa si el estudio es suficientemente grande y no significativa si es suficientemente pequeño. La presencia o magnitud del fenómeno de confusión, sin embargo, no se ve afectado por el aumento o disminución de participantes.

Por lo tanto, el fenómeno de confusión es una función de la magnitud de las asociaciones, más que de sus significancias estadísticas. Dado que las asociaciones fuertes más probablemente sean estadísticamente significativas, las pruebas estadísticas que comparan los grupos expuesto y no expuesto pueden ser una técnica conveniente para identificar asociaciones que puedan ser suficientemente fuertes para causar el fenómeno de confusión, razón por la cual el procedimiento a menudo lleva a las conclusiones correctas sobre la necesidad de controlar para el fenómeno de confusión. Algunos (ver Rothman y Greenland) han sugerido usar las pruebas estadísticas con un valor de alfa (probabilidad de error Tipo I) de 0.20, para aumentar la potencia de detectar diferencias que pueden ser importantes con respecto al fenómeno de confusión. Pero, como guía para la probabilidad del fenómeno de confusión, las pruebas estadísticas no tienen en realidad mucho que ver. (Hay una distinción sutil pero valiosa entre las pruebas estadísticas para evaluar el fenómeno de confusión y las pruebas estadísticas para evaluar si la adjudicación aleatoria al tratamiento o al control fue “exitosa”. Dado que la asignación aleatoria intenta operacionalizar “el azar”, el número y el tamaño de las diferencias observadas entre los grupos de tratamiento y control no deberían exceder con frecuencia lo que esperaríamos del azar, que es justamente lo que las pruebas estadísticas están diseñadas para evaluar. Si hay más diferencias que las que “debieran haber”, podría indicar algún problema en la implementación de la aleatorización. Se supone que un control de estas diferencias sería necesario.)

## Componentes del riesgo relativo crudo

Hay varios otros aspectos del fenómeno de confusión que sería útil considerar. El primero de estos es un método, adjudicado a Miettinen (Miettinen OS: Componentes del riesgo relativo crudo (Components of the crude risk ratio). *Am J Epidemiol* 1972; 96:168-172) para asignar una asociación observada a un componente debido al fenómeno de confusión y a un componente debido al factor de estudio de interés. Según Miettinen, el riesgo relativo crudo (o el odds ratio) puede ser considerado como el producto de un “verdadero” riesgo relativo y un componente debido al fenómeno de confusión. En los ejemplos que hemos considerado hasta ahora, la totalidad de la asociación observada ha sido debido al fenómeno de confusión, i.e., al efecto del hábito de fumar. Pero también es posible tener una asociación que permanece, aunque más fuerte o más débil, después de que se han quitado los efectos de la variable de confusión.

Los siguientes datos hipotéticos ilustran el concepto de Miettinen. Supongamos que estamos llevando a cabo un estudio caso-control para investigar si los hidrocarburos halogenados que se encuentran en el agua potable tratada con cloro que contiene materia orgánica, pueden aumentar la incidencia del cáncer de colon. Recolectamos datos sobre todos los casos en una región de múltiples condados durante varios años y armamos un grupo control usando la selección por discado telefónico de dígitos al azar. Entrevistamos los casos y controles preguntando por la fuente de agua potable (agua tratada versus agua de pozo o embotellada) y dado que otros estudios han sugerido que algún factor desconocido en la vida urbana aumenta la incidencia del cáncer de colon, también recogemos datos sobre la residencia urbana y rural. El análisis crudo de los datos produce la siguiente tabla:

**Tabla 7a**  
**Cáncer de colon y agua potable (datos hipotéticos del estudio caso-control)**

	Exp	Exp	Total
Casos de cáncer colon	170	80	250
Controles	80	170	250
Total	250	250	500

El OR crudo para esta tabla es  $(170 \times 170) / (80 \times 80) = 4.5$ . ¿Será que hay confusión por la residencia urbano-rural?

Podemos investigar el fenómeno de confusión estratificando los datos por la residencia urbano-rural y examinando los OR específicos por estrato:

**Tabla 7b**  
**Cáncer de colon y agua potable (datos hipotéticos del estudio caso -control)**

	Rural		Urbano		Crudo			
	Exp	$\overline{\text{Exp}}$		Exp	$\overline{\text{Exp}}$		Exp	$\overline{\text{Exp}}$
Enf	20	30	Enf	150	50	Enf	170	80
$\overline{\text{Enf}}$	50	150	$\overline{\text{Enf}}$	30	20	$\overline{\text{Enf}}$	80	170

El OR tanto del estrato rural como del urbano es de 2.0, de manera que sabemos que hay un fenómeno de confusión en el OR crudo – sobrestima el “verdadero” OR, haciendo que una asociación moderada parezca una fuerte. ¿Cuánto del OR crudo puede ser atribuido al fenómeno de confusión? Miettinen sugiere que el OR debido a la confusión es el OR de la asociación que sería observado aún si la exposición (hidrocarburos halogenados) no tuviera efecto sobre el resultado (cáncer de colon). Si la exposición no tiene efecto sobre el resultado, entonces cualquier asociación remanente en el análisis crudo puede ser que se deba completamente al fenómeno de confusión.

De manera que para obtener el OR atribuible al fenómeno de confusión, podemos eliminar la verdadera asociación entre los hidrocarburos halogenados y el cáncer de colon. En el ejemplo anterior, consideramos que las tablas específicas están mostrando la verdadera relación (i.e., estamos suponiendo que no hay sesgo de selección, ni sesgo de información y que el único factor de confusión potencial es la residencia urbano-rural como variable dicotómica medida sin error). De manera que "eliminamos" la verdadera asociación de las tablas específicas por estrato. Entonces podemos combinar las tablas específicas por estrato modificadas en una nueva tabla cruda y calcular un OR bruto nuevo. Este OR debe reflejar totalmente el fenómeno de confusión, porque la verdadera asociación ya no existe más.

Dado que el OR es la razón de productos cruzados de las cuatro celdas de una tabla, podemos modificar el OR cambiando cualquier celda de la tabla. Por convención, cambiamos la celda "a" (casos expuestos) a lo que sería si no hubiera asociación entre el factor en estudio y la enfermedad. En este caso si la celda Enf, Exp en el estrato rural contuviera 10 en vez de 20, el OR del estrato rural sería 1.0, i.e., no hay asociación. De igual manera, si la celda Enf, Exp en el estrato urbano contuviera \_\_\_\_ (¿tu adivinanza?) en vez de 150, el OR de ese estrato también sería 1.0. Las tablas revisadas pueden apreciarse a continuación:

**Tabla 7c**  
**Tablas modificadas para el Cáncer de Colon y el agua potable**

	Rural		Urbano		Crudo modificado		Crudo original		
	Exp	$\overline{\text{Exp}}$	Exp	$\overline{\text{Exp}}$	Exp	$\overline{\text{Exp}}$	Exp	$\overline{\text{Exp}}$	
Enf	10	30	Enf	50	Enf	85	Enf	170	80
$\overline{\text{Enf}}$	50	150	$\overline{\text{Enf}}$	30	$\overline{\text{Enf}}$	80	$\overline{\text{Enf}}$	80	170

El OR de la tabla cruda modificada, y por lo tanto el componente atribuible al fenómeno de confusión, es de 2.25. Es interesante señalar que esta cifra es igual al cociente entre el OR bruto original (4.5) y el OR controlado (2.0). De hecho, esta relación se aprecia habitualmente: el OR bruto es igual al producto del OR controlado y el componente atribuible al fenómeno de confusión:

$$\text{Odds ratio (o razón de riesgos) bruto} = \text{Componente debido al factor de estudio} \times \text{Componente debido al fenómeno de confusión}$$

De manera que el componente (de la razón bruta) atribuible al fenómeno de confusión es el grado de asociación "esperada" a partir de la distribución del factor de confusión potencial (en este caso, residencia urbano-rural), i.e., a partir del hecho de que el potencial factor de confusión se distribuye en forma diferente entre las personas expuestas y no expuestas de la población de estudio.

Otra forma de ver esta relación es que el componente atribuible al efecto del factor de estudio, i.e., la asociación sin confusión, puede ser re-formulada de la siguiente manera:

$$\text{Componente debido al factor de estudio} = \frac{\text{Odds ratio (o razón de riesgos) bruta}}{\text{Componente debido al fenómeno de confusión}}$$

De manera que el componente (de la razón bruta) atribuible al factor de estudio i.e., la asociación sin confusión, o "verdadera", puede ser considerada como la razón de una asociación "observada" a una asociación "esperada". La expresión de la relación de esta manera nos recuerda la razón estandarizada de mortalidad (REM), que también es una razón de "esperado" a "observado". De hecho, Miettinen se refiere al OR controlado antes mencionado (i.e., el componente debido al factor de estudio) como un "odds ratio internamente estandarizado", que es simplemente la versión odds ratio de la REM. También es interesante señalar que los OR específicos por estrato son también razones de "observado" a "esperado", en el sentido de que estos OR son equivalentes a la razón del número observado de casos expuestos (el contenido de la "celda a") y el número esperado en ausencia de una verdadera asociación.

En este momento te estarás preguntando cuánto de esto necesitas saber para ejercer la especialidad de epidemiología o controlar el fenómeno de confusión. La respuesta es que este enfoque no es esencial, pero la apreciación del fenómeno de confusión desde esta perspectiva es una ayuda más para comprender los conceptos interrelacionados de confusión, análisis estratificado, estandarización y hasta el marco contrafactual de la inferencia causal. Si la verdadera comparación causal es entre la experiencia en un grupo expuesto y lo que hubiera sido su experiencia en ausencia de la exposición, la REM puede ser considerada la medida de efecto ajustada más relevante, dado que es la razón de la tasa observada en el grupo expuesto a la tasa que se esperaría para ellos si no hubieran estado expuestos (suponiendo que las tasas en la población de estudio difieran de aquellas en la población estándar debido sólo al factor de estandarización y al de exposición).

## **Estudios apareados**

Dado que el fenómeno de confusión es un problema de comparación, uno de los objetivos principales del diseño de estudio es el de obtener grupos que son comparables con respecto a los determinantes del resultado. En diseños experimentales, este objetivo es tal vez el motivo principal de la asignación al azar del factor de estudio. Dado que la asignación al azar no garantiza la distribución equitativa de todos los factores relevantes (aunque en estudios muy grandes la probabilidad de una distribución equitativa es muy alta), se puede utilizar la **pre-estratificación** (también llamada "**bloqueo**") para forzar distribuciones iguales cuando el tamaño muestral es pequeño. La pre-estratificación implica colocar a los participantes, primero en grupos según su configuración de factores de riesgo, y luego la asignación al azar dentro de cada grupo. Este procedimiento generalmente aumenta la eficiencia estadística (grado de precisión por participante del ensayo) (ver Rothman y Greenland, pág. 161).

## **Estudios de seguimiento**

En un estudio no aleatorio, en que el investigador no tiene oportunidad para asignar el factor de estudio, el procedimiento análogo a la pre-estratificación es el **apareamiento**. Al aparear, los participantes en el grupo de comparación (i.e., el grupo no expuesto en un estudio de seguimiento o el grupo control en un estudio caso control) son seleccionados de manera de que se parezcan al **grupo índice** (el expuesto en un estudio de seguimiento o los casos en un estudio caso-control) en cuanto a uno o más factores relevantes. Cuando el grupo no expuesto en un estudio de seguimiento ha sido apareado al grupo expuesto con respecto a todos los factores relevantes, de manera que los dos grupos se diferencian sólo en términos de la exposición al factor de estudio de interés, las incidencias en los dos grupos pueden ser comparadas sin peligro de confusión por parte de las variables de apareamiento. En la práctica, sin embargo, los riesgos que compiten y/o las pérdidas al seguimiento pueden introducir diferencias. Por esta y otras razones (ver Rothman y Greenland pág. 160), los estudios de cohorte apareados no se realizan frecuentemente.

De cualquier manera, ni la pre-estratificación ni el apareamiento son imprescindibles para evitar el fenómeno de confusión, dado que puede ser controlado en el análisis de los resultados del estudio – siempre y cuando exista una superposición adecuada de la distribución de los factores de riesgos entre los grupos. Por esta razón, el objetivo principal del apareamiento es el aumento de la eficiencia estadística asegurando una superposición suficiente (que por lo tanto ayuda indirectamente en el control del fenómeno de confusión).

## Estudios caso-control

En un estudio caso-control, la situación no es, como siempre, tan sencilla. Dada la naturaleza del diseño de estudio caso control, el apareamiento no impide el fenómeno de confusión por el/los factor(es) de apareamiento. Es más, cambiando la composición del grupo control, el apareamiento en un estudio caso control puede hasta hacer que el análisis crudo (sin controlar) sea sesgado. ¿Cómo puede ocurrir esto?

Dado que en un estudio caso-control los participantes son seleccionados en base a la presencia o ausencia de la enfermedad, el apareamiento significa que los grupos caso y control son iguales con respecto a los potenciales factores de confusión. Sin embargo, como vimos anteriormente, el fenómeno de confusión depende de la comparabilidad de los grupos expuesto y no expuesto en la población base del estudio, no entre los casos y los controles de la población de estudio. Aunque el asegurar que los casos y los controles sean similares con respecto a los potenciales factores de confusión puede facilitar el control de la confusión (a través de una mayor eficiencia estadística), el apareamiento de los controles con los casos no cambia la población origen del estudio y por lo tanto no puede modificar el odds de exposición entre los casos. Pero el fenómeno de confusión surge por que el odds de exposición en los casos se ve influido por un desbalance poblacional de alguno de los determinantes del resultado.

Es más, al seleccionar el grupo control de manera que se asemeje al grupo de casos en cuanto a edad, sexo, lugar de atención, u otros factores, el investigador puede forzar al grupo control global a tener una prevalencia de exposición diferente a la que existe en la población base del estudio, que el grupo control busca reflejar. Por supuesto, un grupo control apareado puede igual dar una estimación correcta de la prevalencia de exposición dentro de cada configuración de factores de riesgo. De manera que no tiene porqué haber problemas siempre y cuando el análisis tome en cuenta el apareamiento. Si el análisis apareado y el análisis no apareado dan los mismos resultados, se puede utilizar un análisis no apareado, y para simplificar, a menudo es el utilizado, salvo que el análisis apareado provea mayor precisión.

### ***Ejemplo de apareamiento en un estudio caso-control***

El siguiente ejemplo puede ayudar a hacer más claros estos conceptos. Consideremos otro estudio de cáncer de colon y agua potable que se presenta en la siguiente tabla. Esta vez los tamaños de las poblaciones específicas por estrato y las prevalencias de exposición al agua clorada se presentan, junto al número de casos y la prevalencia de exposición entre los casos.

### Cáncer de colon y agua potable (datos hipotéticos)

Residencia	Tamaño de la población	% de la pob. Total con agua potable clorada	# de casos de cáncer de colon	% de casos con agua clorada
Rural	400,000	20 %	30	40 %
Urbana	600,000	80 %	90	90 %
Total	1,000,000	56 %	120	___ %

Un investigador que lleve a cabo un estudio caso control en esta población y seleccione controles comunitarios sin apareamiento, observaría una prevalencia de exposición de 56% (i.e., un promedio de las prevalencias de exposición específicas rural y urbana, ponderadas por sus respectivos tamaños poblacionales:  $[0.20(400/1000) + 0.80(600/1000)]$ ). Por lo contrario, un grupo control apareado con la distribución de casos tendría una prevalencia de exposición de 65%  $[0.20(30/120) + 0.80(90/120)]$ , dado que en este caso las dos prevalencias son ponderadas por la proporción de casos rurales y urbanos, más que la proporción de habitantes rurales y urbanos en la población.

La prevalencia de exposición en el grupo control apareado, 65%, es una estimación distorsionada de la prevalencia global de exposición en la población como un todo. Pero la estimación no es un problema cuando nuestro análisis toma en cuenta la residencia urbana-rural, dado que las prevalencias de exposición específicas por estrato igual son correctas y conocemos las proporciones de habitantes rurales y urbanos en la población. Si la prevalencia de exposición (columna de la extrema derecha) es 40% en los casos rurales y 90% en los casos urbanos, el odds ratio es de 2.67 (rural) y 2.25 (urbano), 2.70 (bruto, controles no apareados) y 1.85 (bruto, controles apareados). Así ni los controles apareados ni los no apareados dan el OR correcto para un análisis bruto. Por el contrario, un análisis estratificado que toma en cuenta la residencia producirá una estimación del OR válido con cualquiera de los grupos control. (Sugerencia: calcula todos estos OR).

Para una explicación más completa de apareamiento ver el capítulo 10 de Rothman y Greenland. De acuerdo con estos autores, aunque hay circunstancias en que es claramente deseable o no deseable, el valor del apareamiento en estudios caso control es un tema complejo.

### **Potenciales factores de confusión versus verdaderos factores de confusión**

Un tema de importancia práctica y teórica es el de la elección de las variables para investigar como factores de confusión. Como vimos antes, para ser un factor de confusión una variable debe estar asociada tanto con la enfermedad como con la exposición. Así cuando a través del apareamiento en un estudio de cohortes aseguramos que la distribución de los potenciales factores de confusión es idéntico en ambos grupos de exposición (i.e., no hay asociación entre estas variables y la exposición), entonces el primero no puede producir confusión en nuestros resultados (suponiendo que no hay sesgos por otras causas de muerte y otros mecanismos de pérdidas). Aparte de esa situación,

debemos controlar los potenciales factores de confusión en el análisis del estudio para ver si han distorsionado o no la asociación observada (que implica que hemos recordado medirlas!).

La investigación sobre si una variable es un potencial factor de confusión o un verdadero factor de confusión es por lo tanto un tema de determinación empírica en nuestros datos. En la práctica, por lo tanto, el tema de si la variable X es o no un potencial factor de confusión es un tema secundario. Nuestra preocupación principal es la de obtener una estimación válida de la relación entre el factor de estudio y el daño. Si tenemos que controlar, lo haremos; si no necesitamos hacerlo, podemos no hacerlo. En cualquiera de estos casos no tenemos una particular preocupación, en general, en concluir que tal o cual variable es un factor de confusión.

Pero lo que sí es de primordial importancia es la selección de las variables que deben ser consideradas potenciales factores de confusión, i.e., qué variables deben ser medidas y posiblemente controladas para poder obtener una estimación válida de la asociación entre el factor de estudio y el daño. En este caso, nuestra decisión depende de nuestra comprensión de cuáles son las variables, además del factor de estudio, que pueden explicar o responsabilizarse de una asociación observada (o la falta de la asociación). Así, la decisión sobre si una variable debe ser tomada en cuenta para ser controlada como potencial factor de confusión recae en primer lugar y fundamentalmente sobre nuestro modelo conceptual.

En principio y de fundamental importancia, está el hecho de que el factor de confusión debe tener algún tipo de relación con la ocurrencia de la enfermedad u otro resultado. El potencial factor de confusión debe aumentar la probabilidad de que la enfermedad ocurra o debe disminuir el tiempo hasta la ocurrencia de la enfermedad. Si esto no ocurre, ¿porqué le atribuiríamos una asociación observada a esa variable en vez de a nuestro factor de estudio? (Dado que la ocurrencia de la enfermedad debe ser observada, los factores que afectan la detección de la enfermedad pueden también incluirse entre estos factores). Es más, si la variable relevante ocupa una posición intermedia en la cadena causal hipotética que vincula el factor de estudio a la enfermedad, entonces de nuevo, ¿cómo puede ser esa variable y no el factor de estudio la "verdadera" causa de la asociación observada? (Si yo convengo a Jorge que robe un banco y la policía se entera, ¿puedo convencer al juez que me libere porque aparte de lo que hizo Jorge, yo no robé nada?) Al estratificar por condición de fumador en nuestro ejemplo de comportamiento tipo A – ECC, estamos suponiendo que la asociación entre el comportamiento Tipo A y el hábito de fumar surge debido a una causa antecedente común (p.ej. la dificultad para manejar ambientes laborales de alta presión) o debido al efecto del hábito de fumar sobre el patrón de comportamiento, pero no debido a un efecto del patrón de comportamiento sobre el hábito de fumar, lo cual significaría que el hábito de fumar es una variable interviniente y por lo tanto no apropiada para el control de esta manera (Kaufman and Kaufman, 2001).

Sin embargo, en la práctica es a menudo difícil tomar decisiones definitivas sobre cuáles variables son verdaderos factores de riesgo, cuáles son variables intermediarias, y así sucesivamente, de manera que una actitud precavida sería la de obtener datos sobre la mayor parte de las variables potencialmente relevantes posible, explorar los efectos de controlarlas en el análisis del estudio, y luego intentar darle sentido a los resultados. Consideremos por ejemplo, un estudio del efecto del sobrepeso sobre la incidencia de la ECC. Dado que el sobrepeso aumenta el nivel de colesterol y de

la presión arterial, siendo ambos factores de riesgo causales de la ECC, la asociación cruda entre sobrepeso y ECC reflejará alguna combinación de:

1. un efecto directo del sobrepeso sobre la ECC, si existiera,
2. un efecto indirecto del sobrepeso sobre la ECC debido al efecto del sobrepeso sobre el colesterol y la presión arterial, que a su vez aumentan el riesgo de ECC
3. posible confusión, si los niveles de colesterol y de presión sanguínea son más altos en las personas que tienen sobrepeso, no por el efecto del sobrepeso sino por alguna otra razón (p.ej., dieta, estilo de vida sedentario, factores genéticos).

¿Deberíamos controlar para la presión arterial y el colesterol cuando estimamos la asociación entre sobrepeso y ECC? Si no lo hacemos, nuestra medida de asociación estará distorsionada al punto que el fenómeno de confusión estará presente. Si controlamos para ellos con los métodos habituales, sin embargo, nuestra medida de asociación estará distorsionada al punto que el sobrepeso tendrá su efecto sobre la ECC a través de los aumentos en la presión arterial y el colesterol.

Como otro ejemplo, consideremos el problema de estudiar el tema de si las enfermedades de transmisión sexual, como la gonorrea, aumentan el riesgo de adquirir VIH y si el uso de preservativo disminuye ese riesgo. ¿Deberíamos controlar la relación entre ETS y seroconversión para VIH para el uso de preservativo? ¿Debería la relación entre uso de preservativo e incidencia de VIH ser controlada para ETS? Tanto el uso de preservativos como las ETS parecen afectar el riesgo de adquirir una infección por VIH pero los preservativos son también un medio para prevenir las ETS, que en este sentido pueden ser consideradas como una variable ubicada en el camino causal de los preservativos al VIH.

Es más, un factor causal obligatorio para tanto las ETS adquiridas por relaciones sexuales como la infección por VIH es el contacto sexual con un compañero infectado. Los compañeros "de alto riesgo" tienen una probabilidad mayor de estar infectados, y cuantos más compañeros hayan, mayor es el riesgo de exposición a la infección. ¿Deberíamos controlar para el número de compañeros de alto riesgo al investigar la relación entre preservativos, ETS, y VIH? Pero los compañeros de alto riesgo son también un factor de riesgo para las ETS, de manera que las ETS pueden ser consideradas una variable intermedia entre las relaciones sexuales con compañeros de alto riesgo y la infección por VIH. Por lo tanto, el sólo pensar en qué variables controlar en una red de causalidad puede producir de por sí mucha confusión! Greenland, Pearl, y Robins (1999) presentan un sistema de diagramas causales para describir y analizar redes causales para identificar las variables que deben ser controladas. Entre otros puntos, explican que el control para una variable puede, en algunas situaciones, crear un fenómeno de confusión que de otra manera no ocurriría. En general, el control del fenómeno de confusión (y la interpretación de datos en general) se fundamenta en los supuestos de las relaciones causales involucrando las variables medidas y no medidas. Los datos aislados no son adecuados para resolver los temas de causalidad sin estos supuestos. La comprensión de la metodología en esta área está aumentando (ver los artículos por Greenland, Pearl, y Robins y Kaufman y Kaufman). Sin embargo, el limitado conocimiento de las relaciones causales además de la que está en estudio y la probable existencia de variables no medidas pero importantes permanece como uno de los escollos para la investigación observacional.

## **Controlando para las variables sociodemográficas**

Casi todas las investigaciones epidemiológicas controlan de una u otra manera para las variables sociodemográficas (p.ej., edad, género, raza, nivel socioeconómico). Como vimos en el capítulo sobre Estandarización, las comparaciones que no controlan para esas variables pueden ser muy engañosas. Sin embargo, hay dificultades en la interpretación del ajuste para factores sociodemográficos, en parte porque las asociaciones con factores sociodemográficos probablemente reflejen los efectos de factores asociados con algunos de ellos, y algunos de estos factores pueden ser variables intermediarias. Por ejemplo, los estudios de inequidades étnicas en salud a menudo tratan de controlar para las diferencias en nivel socioeconómico. Sin embargo se ha argumentado que el nivel socioeconómico es una variable intermediaria entre la etnia y los resultados de salud, de manera que controlarla por los métodos habituales es muy problemático (Kaufman y Kaufman, 2001). El problema de la interpretación se ve agravado por la persistencia de una asociación con la etnia, a pesar del control para otros factores, que lleva a que el investigador llegue a inferencias no confiables de que debe haber un factor genético involucrado. También vale la pena señalar que la asociación bruta presenta la situación tal cual existe. Aún cuando la explicación causal indica que hay otros factores responsables, el hecho de un estado de salud desigual sigue siendo un tema que hay que interpretar. Es más, una solución puede no requerir tratar con la "verdadera" causa.

### **“Colapsabilidad” versus “comparabilidad”**

Aunque el problema del fenómeno de confusión y la necesidad de controlarlo hace mucho tiempo que forman parte de la epidemiología y otras disciplinas, la comprensión teórica del fenómeno de confusión se ha desarrollado fundamentalmente a partir de los artículos de Miettinen publicados desde mediados de los años 70. Dos definiciones o perspectivas opuestas eran discutidas en ese momento, una llamada de "comparabilidad" y la otra llamada de "colapsabilidad".

En la definición bajo la teoría de la **comparabilidad**, respaldada por Sander Greenland, James Robins, Hal Morgenstern, y Charles Poole, entre otros [ver en la bibliografía el artículo "Identifiability, exchangeability, and epidemiological confounding" ( Posibilidad de Identificación, intercambio y el fenómeno epidemiológico de confusión) y las cartas "RE: Confounding confounding"(Re: confundiendo el fenómeno de confusión)], el fenómeno de confusión se define en relación al modelo contrafactual de la inferencia causal, descrito al principio de este capítulo. El fenómeno de confusión resulta de la falta de comparabilidad, i.e., una diferencia entre la distribución de resultados para el grupo no expuesto con respecto a lo que se hubiera observado en el grupo expuesto si no hubiera sido expuesto. Dado que esta última situación es hipotética y no puede ser observada, la definición bajo la teoría de la comparabilidad no puede ser aplicada directamente, aunque tiene algunas ventajas teóricas además de implicancias prácticas.

En la definición bajo la teoría de la **colapsabilidad**, promovida por D.A. Grayson (*Am J Epidemiol* 1987;126:546-53) y otros, el fenómeno de confusión está presente cuando la medida de asociación cruda difiere del valor de la misma medida cuando son controladas las variables extrañas por medio de la estratificación, ajuste, o modelado matemático. Si las tablas 2 x 2 para los distintos estratos de un factor de riesgo (como en el ejemplo de la personalidad Tipo A) producen medidas de asociación (p.ej., RR) que son esencialmente equivalentes a la medida de asociación para la tabla 2 x 2

"colapsada" (sin tener en cuenta el factor de riesgo utilizado para la estratificación), entonces no hay fenómeno de confusión con respecto a esa medida de asociación. La definición bajo la teoría de la colapsabilidad se aplica fácilmente en la práctica y es de uso generalizado. Las desventajas de esta definición son que hace que el fenómeno de confusión sea específico de la medida de asociación utilizada y las particulares variables que son controladas.

Afortunadamente para los epidemiólogos en ejercicio, las dos definiciones generalmente están de acuerdo en cuanto a la presencia o ausencia del fenómeno de confusión cuando la medida de efecto es una razón o una diferencia de incidencias (proporciones o tasas). El problema práctico principal surge cuando la medida de asociación es el odds ratio (salvo que la situación sea una en que el odds ratio estima estrechamente una razón de riesgos o un riesgo relativo, p.ej., un resultado muy poco frecuente). Hay más explicaciones de este y temas relacionados en el Apéndice (y en Rothman y Greenland).

## **Control del fenómeno de confusión**

Ahora que estamos todos impresionados con la importancia y el valor de tomar en cuenta los efectos de las múltiples variables, ¿cuáles son las técnicas analíticas disponibles? Las principales son las siguientes:

- Restricción
- Apareamiento
- Análisis estratificado
- Aleatorización
- Modelado

### **Restricción**

Cuando adoptamos la restricción como técnica, estamos en realidad optando por no intentar un análisis de múltiples variables – simplemente restringimos o limitamos nuestro estudio a participantes con características particulares (p.ej., hombres, 40 a 50 años de edad, no fumadores, de edad promedio para la altura, sin enfermedades conocidas o presión arterial elevada) de manera de no tener que preocuparnos sobre los efectos de los diferentes valores de estas variables. Algún tipo de restricción está casi siempre presente en el diseño de estudio, dado que prácticamente todos los estudios involucran un área geográfica delimitada, un rango específico de edad, y así sucesivamente, aunque el motivo sea uno de factibilidad más que de evitar el fenómeno de confusión. Si se sabe o se sospecha que una asociación es más fuerte en un subconjunto particular de la población, puede tener sentido enfocar los estudios en ese grupo. O, si hay pocos datos disponibles que tienen que ver con una población particular, puede tener sentido restringir los participantes del estudio a personas de esa población. La restricción también es útil cuando una variable importante (particularmente un factor de riesgo que sospechamos que es fuerte) se distribuye en forma muy irregular en la población blanco, de manera que es difícil y costoso obtener suficientes participantes de los niveles menos comunes de esa variable.

Consideraciones como estas a menudo han sido citados como razones por las cuales tantos estudios en EEUU han incluido sólo participantes de raza blanca, a menudo solo hombres blancos. Por ejemplo, en muchas posibles poblaciones de estudio (definidas por la geografía, el empleo, o la pertenencia a algún grupo), hay (o habían) demasiado pocos integrantes de los grupos minoritarios para proveer suficientes datos para estimaciones confiables. Si razonamos que en semejante situación el análisis estratificado por raza/etnia es esencialmente equivalente a la restricción a sólo blancos, los investigadores a menudo simplemente limitan el análisis de los datos o la recolección de datos a los blancos. El razonamiento detrás de la limitación de las investigaciones a los hombres ha sido que las tasas de enfermedad en hombres y mujeres son muy diferentes de manera que investigaciones de, por ejemplo, ECC en personas adultas, requiere muchas, muchas más mujeres que hombres, para obtener un número dado de casos (y por lo tanto alcanzar la potencia estadística o precisión requeridas). Puede también haber influido el hecho de que hasta la década de los 80 el número de mujeres epidemiólogas y su participación en la elaboración de políticas era bastante reducido.

Las razones para restringir los participantes del estudio según raza/etnia son más complejas. Si la raza/etnia (el término "raza" resulta virtualmente imposible de definir con precisión) no es un factor de riesgo para el resultado que se está investigando, entonces no hay necesidad de estratificar o restringir por raza/etnia para controlar por el fenómeno de confusión. Pero la división por raza siempre presente en EEUU y la discriminación que la acompaña, frecuentes injusticias, recurrentes atrocidades, y el hecho de que la sociedad dominante sigue ignorando el tema, han creado barreras intelectuales, actitudinales, políticas y logísticas a la investigación neutral con respecto a raza (ver la bibliografía, los estados del Sur de EEUU mantuvieron un apartheid legalmente impuesto hasta fines de la década de los 60, y el apartheid no legalmente impuesto continúa hasta el día de hoy). Muchos de estos temas también han surgido en otras poblaciones norteamericanas que provienen de otros continentes aparte del Europeo.

El concepto de la raza como potente variable biológica capaz de producir el fenómeno de confusión en muchas asociaciones de exposiciones-daños tiene sus raíces históricas en la "carrera de la ciencia", del siglo XIX, en que varias características anatómicas, fisiológicas y de comportamiento que se suponían determinadas genéticamente, eran interpretadas como que demostraban la relativa superioridad/inferioridad de grupos poblacionales y justificando la subordinación de las personas de color por parte de los blancos (Bhopal, Raj. Manuscrito en preparación, 1996; ver la bibliografía para referencias adicionales). Varias enfermedades y condiciones fueron vinculadas con los grupos raciales (incluyendo "drapetomanía", el deseo irracional y patológico de los esclavos de escapar, y "disestesia Aethopica" ["picardía"]). Se puede leer en los libros médicos del período que los negros son "una raza exótica".

La mayor parte de estas ideas han sido desacreditadas en general, aunque no están extintas [ver Carles Muntaner, F. Javier Nieto, Patricia O'Campo "The Bell Curve: on race, social class, and epidemiologic research". ("La curva en campana: sobre raza, clase social e investigación epidemiológica") *Am J Epidemiol*, September 15, 1996;144(6):531-536]. Pero hasta hace poco tiempo, la gran mayoría de las poblaciones de estudio epidemiológicas han sido personas blancos, de habla inglesa, urbanas o suburbanas, y no pobres. La base científica vinculando la raza a los resultados en salud ha surgido para unas pocas condiciones (siendo las más llamativas el cáncer de piel, y anemia de células falciformes). Pero la sospecha de que la raza puede ser un factor de riesgo es difícil de desechar, en parte porque se ve reforzada por las muchas diferencias relacionadas con la raza de los resultados en salud. Estas diferencias presumiblemente surgen por las diferencias en dieta

y nutrición, ambiente físico y social, experiencias tempranas en la vida, recursos económicos, atención médica, características del vecindario, interacciones sociales, experiencias de discriminación, estilo de vida, y una multitud de otros factores que afectan la salud y el bienestar, pero la raza es un sustituto más fácil de medir (aunque no de definir) para el riesgo.

Además, muchas de estas diferencias presentan desafíos logísticos (p.ej., falta de familiaridad de investigadores y personal [fundamentalmente blancos y de clase media] al estudiar personas y comunidades con otros antecedentes, distantes de las instituciones de investigación, con infraestructura limitada, escasez de cuestionarios y otras herramientas de medición que han sido validadas en múltiples grupos étnicos/raciales, entre otros). Los aspectos prácticos de los estudios epidemiológicos típicamente exigen gran cantidad de tiempo, esfuerzo y costo, de manera que es natural buscar maneras de reducirlos.

Cualquiera sean los motivos y sus méritos, el impacto global de enfocar la investigación en poblaciones blancas, de habla inglesa, urbanas/suburbanas, no pobres, es la escasez de conocimiento, experiencia de investigación, herramientas de recolección de datos, y beneficios adicionales de la participación en la investigación (p.ej., acceso a tratamientos de última generación, vinculaciones entre los efectores de la asistencia médica y los centros de investigación universitarios) para otras poblaciones – aún para condiciones para las cuales estas poblaciones tienen tasas más altas o para las cuales hay razones científicas o de salud pública para cuestionar la pertinencia de los hallazgos en norteamericanos de origen europeo en personas de color o de origen latino.

Desde aproximadamente mediados de la década de los 80, en parte en respuesta a la estimulación de las políticas inspiradas por el Congreso Norteamericano de los Institutos Nacionales Norteamericanos de Salud (National Institutes of Health) y el Centro de Prevención y Control de Enfermedades (Centers for Disease Control and Prevention) que ahora exigen que todas las solicitudes de financiación presenten una justificación importante para no incluir cantidades significativas de mujeres y minorías en las investigaciones propuestas, la investigación de las poblaciones poco estudiadas ha aumentado considerablemente. Estas políticas y las medidas tomadas para que se cumplieran han creado nuevos desafíos para los epidemiólogos y en muchos casos han aumentado la complejidad de los estudios epidemiológicos. Sin embargo, la epidemiología no puede, por un lado reclamar ser una estrategia esencial para mejorar la salud pública y por otra parte, ignorar un cuarto (minorías) o cinco octavos (las mujeres, más los hombres de las minorías) de la población.

Hace varios años el Colegio Norteamericano de Epidemiología (American College of Epidemiology) emitió la "Declaración de Principios sobre Epidemiología y Poblaciones Monoritarias" ("Statement of Principles on Epidemiology and Minority Populations") (*Annals of Epidemiology*, Noviembre 1995;5:505-508; comentario 503-504; también se encuentra bajo "declaraciones de políticas" [policy statements] en el sitio web del Colegio, [www.acepidemiology.org](http://www.acepidemiology.org)) reconociendo la importancia de la salud de las minorías para la salud pública, de mejorar los datos epidemiológicos sobre poblaciones minoritarias, y de aumentar la diversidad étnica en la profesión epidemiológica. La Declaración ha sido endosada por los cuerpos dirigentes de varias organizaciones de epidemiología y salud pública, incluyendo el Consejo sobre Epidemiología y Prevención de la Asociación Norteamericana del Corazón, Asociación Norteamericana de Registros Centrales de Cáncer, Asociación de Profesores de

Medicina Preventiva, Colegio Norteamericano de Medicina Preventiva, Sector de Estadísticas en Epidemiología de la Asociación Norteamericana de Estadística, Asociación Norteamericana de Salud Pública, (Council on Epidemiology and Prevention of the American Heart Association, North American Association of Central Cancer Registries, Association of Teachers of Preventive Medicine, American College of Preventive Medicine, American Statistical Association Section on Epidemiology in Statistics, American Public Health Association) y las escuelas de epidemiología en numerosas instituciones (p.ej., Harvard, UNC, Universidad de Massachusetts en Amherst y el Centro de Ciencias de la Salud de la Universidad de Texas). En enero de 2000, El Departamento de Salud y Servicios Humanos anunció el objetivo de eliminar las desigualdades raciales/étnicas en salud para el año 2010. Este desafío y otro relacionado, el de incluir la diversidad racial/étnica a la profesión epidemiológica, son fundamentales para la Salud Pública, por lo menos en EEUU.

## **Apareamiento**

Como se discutió anteriormente, el fenómeno de confusión producido por uno o más factores de riesgo puede ser evitado en un estudio de seguimiento asegurando a través del apareamiento que los múltiples grupos de exposición tengan la misma distribución (conjunta) de los factores de riesgo. Así en un estudio de cohortes o un ensayo clínico, podemos seleccionar los participantes en un nivel de exposición y luego seleccionar los participantes para otro nivel de exposición (incluyendo los "no expuestos") a partir de una población candidata mayor según la distribución de los factores de riesgo seleccionados en el primer grupo.

Por ejemplo, consideremos un estudio de cohortes retrospectivo para investigar si los jugadores de deportes de nivel universitario (p.ej., fútbol americano, basketball) cuando llegan a la edad de 60, tienen más probabilidades de tener potenciales evocados alterados frente a estímulos auditivos, sugiriendo alteraciones de las funciones neurológicas. La cohorte expuesta podría consistir de ex-jugadores de basketball de los listados de equipos de varias universidades; la cohorte de comparación (no expuesta) podrían ser ex-estudiantes de las mismas universidades de la misma época.

Dado que la medición de los potenciales evocados es un proceso costoso y de prolongada duración, queremos que cada participante aporte la máxima información posible, para minimizar el número total de participantes necesarios para el estudio. Si seleccionamos los participantes no expuestos totalmente al azar, es probable que diferirán de los jugadores de basketball de muchas maneras (medido durante sus años universitarios) que pueden afectar los potenciales evocados – altura, salud física, fuerza, agilidad, coordinación, edad (por ejemplo, los jugadores de basketball es poco probable que fueran estudiantes maduros que vuelven para completar su grado después de haberse tomado el tiempo de formar una familia), educación de los padres, puntajes en la prueba de aptitud escolástica (porque los atletas pueden ser incluidos por sus talentos aunque sus registros académicos sean menos competitivos), y género (estos deportes son, o eran por lo menos, masculinos). Algunas de estas características pueden afectar los potenciales evocados. Por lo tanto, las comparaciones de los potenciales evocados a los 60 años entre los jugadores de basketball y los otros alumnos podría sufrir confusión por las distribuciones diferentes de estas y otras variables.

Cuando intentamos controlar estas diferencias, podemos encontrar que son tan importantes que hay jugadores de basketball (p.ej., aquellos de más de 2 metros de altura) para los cuales hay muy pocos o

ningún sujeto de comparación y sujetos de comparación (p.ej., aquellos de menos de 1.60 metros de altura) para los cuales hay muy pocos, si es que hay algún jugador de basketball. Pero los estratos con pocos jugadores de basketball o con pocos sujetos control dan menos información para la comparación de los potenciales evocados que los estratos en que los dos grupos tienen más o menos la misma cantidad de sujetos. Los hallazgos del análisis que controla para el fenómeno de confusión serán por lo tanto menos "eficientes", en términos de información por sujeto, que si los jugadores de basketball y los sujetos control tuvieran distribuciones similares de los factores de riesgo que son controlados. Con el mismo número total de sujetos y la misma razón de riesgos, el estudio con grupos de comparación más similares dará un intervalo de confianza más estrecho (mayor precisión estadística) además de un valor p menor (mayor significancia estadística).

Una forma de obtener un mejor balance de los factores de riesgo entre los jugadores de basketball y el grupo de comparación es aparear el grupo de comparación con el grupo de los jugadores de basketball con respecto a los factores de riesgo más importantes. Por ejemplo, podemos estratificar los jugadores de basketball por altura y puntaje académico promedio durante la universidad. Una estratificación doble puede tener un total de 16 estratos. Entonces seleccionamos los sujetos de comparación de manera que tengan la misma distribución en estos 16 estratos. La selección de esta manera del grupo de comparación se denomina apareamiento por categoría o frecuencia. (Esta investigación puede ser una ocasión lógica para usar la restricción, p.ej., incluir sólo hombres de 18-22 años, sin ningún trastorno físico o médico).

El método descrito anteriormente del apareamiento por frecuencia requiere conocer la distribución del factor de riesgo en el grupo índice antes de reclutar los sujetos control, de manera que estos últimos puedan ser elegidos de manera de tener la misma distribución. Otro método de lograr el apareamiento por frecuencia es el muestreo apareado. Con el muestreo apareado, un sujeto de comparación es seleccionado secuencialmente o al azar de un grupo de potenciales sujetos control que tienen los mismos valores de covariables que el sujeto índice. Por ejemplo, cada vez que se ingresa un ex-jugador de basketball en la investigación, buscamos un sujeto control que pertenece al mismo estrato de altura y puntaje académico promedio que el jugador. Cuando dejamos de incluir sujetos, los dos grupos tendrán idénticas distribuciones en los 16 estratos.

Similar al muestreo apareado es el apareamiento por parejas. Al aparear en parejas seleccionamos cada sujeto de comparación según las características de un sujeto índice, características que no se comparten en general con otros sujetos índice (i.e., los estratos son muy pequeños, conteniendo tal vez un sólo sujeto índice cada uno). Por ejemplo, podemos decidir usar los hermanos de los sujetos índices como controles. O, podemos decidir que queremos que la distribución conjunta de altura y puntaje académico promedio sea tan similar entre los jugadores y el grupo comparación que no queremos tener que categorizar las variables. En este caso elegiríamos cada sujeto de comparación para que fuera de una altura dentro de un cierto rango (p.ej., 2 centímetros) de la altura del sujeto índice y su puntaje académico promedio dentro de un pequeño rango del del sujeto índice (el apareamiento por parejas de esta manera también se conoce como "apareamiento calibrado", aunque ha sido criticado (ver Rothman y Greenland)).

Lo que diferencia el apareamiento por parejas del muestreo por parejas y otras formas de apareamiento por frecuencia es la estrechez del vínculo entre el sujeto índice y el control. Si hay

múltiples pares de sujetos índice-control en cada estrato, de manera que los pares pueden disolverse, mezclarse, y formarse de nuevo, sin mayor efecto siempre y cuando todos los sujetos se queden en su estrato, entonces la situación es una de apareamiento por frecuencia. Si por el contrario, los sujetos de comparación en general no son intercambiables con otros sujetos de comparación, si cada sujeto de comparación se considera completamente comparable sólo con el sujeto índice con quien está apareado, entonces la situación es una de apareamiento por parejas. (Para una discusión de muestreo por parejas versus apareamiento por parejas, ver MacMahon y Pugh, 1970, págs. 252-256. Además, aunque la discusión actual se ha concentrado en pares, todos estos conceptos se aplican a tres, cuatro y n controles o número variable de sujetos de comparación para cada sujeto índice, p.ej., los hermanos del sujeto índice).

En estudios caso-control, como vimos anteriormente, la arquitectura del estudio nos impide asegurar que los grupos de exposición son similares con respecto a otros factores de riesgo aún en la población de estudio, y mucho menos en la población base del estudio. Por lo tanto, el apareamiento en los estudios caso-control no impide el fenómeno de confusión. De cualquier manera el apareamiento puede ser beneficioso, ya que si los potenciales factores de confusión más importantes se distribuyen de manera similar en los casos y los controles, la comparación de los dos grupos puede ser más estadísticamente eficiente – con el mismo número de participantes, el intervalo de confianza para la estimación del odds ratio será más estrecho (i.e., la estimación será más precisa).

Lamentablemente, el tema de si es o no beneficioso aparear los controles con los casos resulta ser difícil de contestar, ya que en algunos casos el apareamiento puede llevar a disminuir la eficiencia estadística. Si la variable(s) de apareamiento está fuertemente asociada con la exposición, entonces la prevalencia de exposición en los controles apareados será más parecido a aquella en los casos, que lo que ocurriría para un grupo de control no apareado, disminuyendo de esta manera la fuerza de asociación observada entre la exposición y la enfermedad. Si los factores de apareamiento no son factores de riesgo fuertes para la enfermedad, ha ocurrido un "sobre-apareamiento" y una verdadera asociación puede pasar completamente desapercibida.

El consejo actual para los estudios caso-control es el de aparear sólo los determinantes importantes del resultado en estudio, especialmente si es probable que se distribuyan en forma muy diferente en los casos y los controles. Además, por supuesto, no se debe aparear una variable cuya relación con el resultado es de interés estudiar. Una vez que se han apareado los casos con los controles con respecto a una variable, su odds ratio será uno. Aunque el apareamiento en un estudio de seguimiento no tiene los problemas que pueden surgir en los estudios caso-control, en cualquier diseño de estudio el uso de apareamiento puede presentar dificultades prácticas y logísticas, sobretodo si la reserva de potenciales sujetos de comparación es pequeña o si la identificación o evaluación de potenciales pares es costosa.

## **Aleatorización**

La aleatorización, la asignación al azar de participantes a la "exposición" o "tratamiento" y al grupo de comparación, sólo se consigue en los ensayos clínicos. La aleatorización asegurará que, en promedio, los grupos índice y de comparación tendrán una distribución y proporciones similares de todos los factores. Por supuesto, en cualquier estudio particular los grupos pueden diferir (y a

menudo lo hacen) en algún aspecto u otro (i.e., la aleatorización no "funcionó", aunque en realidad sí funcionó –simplemente no cumplió con todo lo que nos hubiera gustado que hiciese). Así, a menudo los grupos de intervención y de control se obligarán a ser similares (a través del apareamiento, también denominado "pre-estratificación") o se analizarán utilizando el análisis estratificado.

Una consideración importante con respecto a la aleatorización – y su ventaja clara sobre cualquiera de los otros métodos disponibles – es que en promedio la aleatorización controla los efectos de variables que no pueden ser medidas o ni siquiera se sospecha que son factores de riesgo. Salvo que una variable sea identificada como relevante y pueda ser medida, ninguno de los otros enfoques descritos anteriormente (o más adelante) pueden ser utilizados. Con la aleatorización, tenemos la seguridad de que por lo menos hemos tenido en cuenta la influencia de factores de riesgo desconocidos y no sospechados.

## **Análisis estratificado**

El análisis estratificado es la descomposición en partes componentes de un conjunto, de manera de poder observar cada sub-componente individual. Si el hábito de fumar es un factor relevante para la enfermedad en estudio, simplemente decimos, "muy bien, miraremos a los fumadores primero y a los no fumadores después." La mayor parte de los ejemplos del fenómeno de confusión y de modificación del efecto que hemos usado han sido presentados en término de análisis estratificado.

El análisis estratificado tiene sentido intuitivo y es ampliamente utilizado. Sirve fundamentalmente para el control de variables nominales (variables cuyos valores no tienen un orden en relación unos con otros, como por ejemplo, las regiones geográficas [norte, este, oeste]) y las variables ordinales que tienen pocas categorías (e.g., severidad de lesiones [leve, moderada, severa]). El análisis estratificado da una "imagen" de lo que está ocurriendo en los datos, se presenta y explica fácilmente, y no requiere supuestos restrictivos sobre un modelo estadístico.

Por otro lado, el análisis estratificado requiere que las variables continuas sean categorizadas, lo cual introduce un grado de arbitrariedad y produce pérdida de información. No es posible controlar para más que unas pocas variables al mismo tiempo porque a medida que aumenta el número de estratos, la comprensión e interpretación de los resultados presenta un importante desafío, especialmente si los resultados varían de un estrato a otro sin un patrón obvio. A pesar de estas desventajas, el análisis estratificado es un pilar de los enfoques de análisis epidemiológicos.

Cuando hay múltiples estratos, puede ser difícil describir y resumir los resultados, particularmente porque muchos estratos tendrán relativamente pocos sujetos, de manera que las diferencias que existan pueden ser debidas a la variación aleatoria. En casos como esos existen varias medidas de resumen– generalmente diferentes formas de promedios ponderados de medidas específicas por estrato. Una medida de resumen es una medida única, global, de asociación de todos los estratos (o un subgrupo de estratos), controlando para las variables sobre las cuales se ha realizado la estratificación. La razón de riesgo estandarizada (RRE) presentada en la sección sobre estandarización por edad es una de esas medidas de resumen. Otras se presentarán en el capítulo de

"Análisis e Interpretación de Datos". Por supuesto, igual que para cualquier medida de resumen, si hay diferencias importantes entre los estratos, un promedio global puede no tener mayor sentido.

## **Modelado**

Dado un número ilimitado de participantes, y tiempo, paciencia, y capacidad de interpretación de los datos, podemos enfrentar cualquier problema de análisis multi-variable por medio de la estratificación. Pero consideremos las dimensiones del desafío: si tenemos tres variables, cada una dicotómica, hay ocho estratos únicos posibles; si tenemos 6 variables, cada una dicotómica, hay 64; si tenemos seis variables dicotómicas y tres variables con tres niveles cada una, el número de estratos llega a 1728! Imagínate tratando de interpretar 1728 odds ratio, suponiendo que tenemos suficientes participantes para cada uno.

Dado que a menudo tenemos unas cuantas variables que queremos utilizar, y variables (p.ej., edad, presión arterial, peso corporal) que son continuas, de manera que podemos llegar a perder información categorizándolas en un pequeño número de niveles, hay una necesidad obvia de tratamientos más sofisticados que no nos exija examinar cada combinación posible de niveles de los factores para descubrir los efectos de cada variable. Hay un tratamiento así – modelado matemático – pero su uso implica un costo, en términos de ciertos supuestos que debemos hacer para poder simplificar la situación. Otro precio que pagamos es que los datos mismos no están a la vista. En las palabras de Sir Richard Doll (entrevista con Beverly M. Calkins publicada en la Carta de Otoño de 1992 del Colegio Americano de Epidemiología):

"Han habido importantes adelantos en el camino: estudios de mayor escala, técnicas estadísticas más potentes, y el desarrollo de computadoras que permiten utilizarlas. Temo, sin embargo, que la facilidad de utilizar paquetes estadísticas a veces vuelve ciegas a las personas a lo que realmente está ocurriendo. No se puede tener un profundo conocimiento de lo que son las relaciones cuando incluye los componentes ambientales y todos los demás de la historia juntos en una regresión logística que permite quince cosas diferentes. Soy un gran creyente en la simple estratificación. Uno sabe lo que hace, y realmente puedes ver los pasos intermedios y no tener todos los datos en la computadora."

## **Limitaciones en la capacidad de controlar los potenciales factores de confusión**

Habitualmente los epidemiólogos no conocen todos los determinantes de las condiciones de salud que estudian. Otros determinantes pueden ser conocidos pero no pueden ser medidos, sea en general o bajo las circunstancias del estudio. Los factores de confusión desconocidos y no medidos pueden ser controlados sólo a través de la aleatorización. Esta ventaja única que poseen los diseños aleatorizados es la razón primordial de su fuerza particular.

Hasta los potenciales factores de confusión que son controlados a través de la restricción, apareamiento, análisis estratificado, o modelado, existen limitaciones o errores en la conceptualización, medición, codificación, y especificación del modelo que comprometen la efectividad del control. Dichos controles incompletos resultan en el "fenómeno de confusión

residual" por el factor potencial de confusión. El fenómeno de confusión residual, como el fenómeno de confusión no controlado, pueden llevar a un sesgo en cualquier dirección (positiva o negativa, alejándose del valor nulo y acercándose a él) en la medida de efecto ajustada entre el factor de estudio y el daño. Aún si el error de medición en el potencial factor de confusión es no-diferencial (i.e., independiente del factor de estudio y del daño), el sesgo en la asociación de interés primaria puede darse en cualquier dirección.

Es importante estar al tanto de estas limitaciones, pero no son motivo para desanimarse. La epidemiología ha dado y continúa dando valiosas percepciones y evidencias a pesar de estos y otros obstáculos. Las limitaciones surgen fundamentalmente del tema tratado – fenómenos relacionados con la salud en poblaciones humanas que viven en libertad – más que de la disciplina. Estar al tanto de las limitaciones, minimizándolas cuando es posible, y evaluando perspicazmente su posible impacto en la interpretación de los datos son lo que distinguen un epidemiólogo bien entrenado.

### ***El fenómeno de confusión y la modificación del efecto***

Como señalamos en el capítulo sobre Inferencia Causal, el enfoque de variable única de la epidemiología, el método de un-factor-a-la-vez que subyace la evolución de la comprensión epidemiológica, es la base de los conceptos del "fenómeno de confusión" y de la "modificación del efecto". Hay también similitudes en la manera en que son investigados en el análisis de los datos. Para hacer clara la distinción, haremos un contraste con estas dos implicancias de la multicausalidad.

Si observamos una asociación entre una enfermedad y algún factor nuevo – pero no tomamos en cuenta los posibles efectos de causas conocidas de la enfermedad – podemos atribuirle erróneamente al factor nuevo la asociación observada cuando en realidad podemos estar apreciando los efectos de factores conocidos. El fenómeno de "confusión" se refiere a aquella situación en que un exceso observado de enfermedad puede ser equivocadamente atribuido a la exposición de interés cuando, de hecho, algún otro factor – relacionado tanto con la enfermedad como con la exposición – es responsable del exceso observado. Por ejemplo, la tasa de mortalidad bruta es mayor en Florida (EEUU) que en Alaska. Si atribuyéramos la tasa de mortalidad más alta en Florida al efecto de la industria de los cítricos, estaríamos contribuyendo al fenómeno de confusión. La "verdadera" razón subyacente de las tasas de mortalidad más altas en Florida es una distribución por edad mayor de la población de Florida.

Cuando consideramos el fenómeno de confusión estamos haciendo la pregunta "¿la asociación observada entre el uso de anticonceptivos orales y el riesgo de infarto de miocardio, se debe a un efecto de los anticonceptivos orales o se debe en realidad a los efectos de otros factores de riesgo para el IAM, como el hábito de fumar, la presión arterial elevada, el colesterol sanguíneo elevado, y diabetes, que resultan asociados al uso de anticonceptivos orales?" Para contestar esta pregunta, trataremos de asegurar que los grupos que se están comparando son iguales con respecto a estos "potenciales factores de confusión" y/o examinaremos la relación ACO-IAM dentro de las categorías de los "potenciales factores de confusión" intentando "mantener otros factores constantes".

"Modificación del efecto" se refiere a la variación en la relación entre la exposición y el daño, variación que se debe a las acciones de algún otro factor (llamado modificador del efecto). Por ejemplo, la relación entre estrógenos exógenos y el cáncer de endometrio parece ser más débil en presencia de obesidad. La relación entre los anticonceptivos orales y el infarto agudo de miocardio parece ser más fuerte en mujeres que fuman cigarrillos que en las que no fuman.

Cuando consideramos la modificación del efecto, estamos haciendo la pregunta "¿la asociación observada entre el uso de anticonceptivos orales y el riesgo de IAM se ve influida en forma importante por otros factores de riesgo del IAM, como el hábito de fumar, la presión arterial elevada, el colesterol elevado, o hasta por factores que, por sí mismos, no afectan el riesgo de IAM?" Para contestar esta pregunta, examinaremos la relación ACO-IAM en las categorías de estos "potenciales modificadores". Buscaremos también las explicaciones biológicas y/o de comportamiento para las posibles influencias modificadoras.

Con el fenómeno de confusión estamos preocupados por determinar si la relación entre nuestra exposición y nuestro daño existe o no existe. Con la modificación del efecto, estamos preocupados por definir las especificaciones de la asociación entre la exposición y el daño. Es decir, estamos interesados en identificar y describir los efectos de los factores que modifican la asociación exposición-daño. La interrogante sobre el fenómeno de confusión es fundamental en el establecimiento de factores de riesgo. La interrogante sobre modificación del efecto tiene importantes implicancias para definir la etiología de las enfermedades y para las intervenciones. El fenómeno de confusión es una molestia. La modificación del efecto, aunque difícil de evaluar por razones estadísticas, es de considerable interés potencial.

Una mnemotécnica que puede ser útil es la siguiente. Una evaluación del fenómeno de confusión es una investigación de la "culpabilidad" o "inocencia". Una evaluación de la modificación del efecto es una investigación de la "conspiración".

## PUNTOS PRINCIPALES

- El fenómeno de confusión es una distorsión o atribución errónea de un efecto a un factor de estudio particular. Resulta de la falta de comparabilidad del grupo control.
- Un factor de confusión es un determinante del daño o su detección, o posiblemente un factor correlacionado con un determinante, que se distribuye en forma desigual entre los grupos que son comparados.
- Un determinante de la enfermedad debe presentarse como un factor de riesgo independiente, i.e., no uno cuya asociación con la enfermedad resulta de la asociación con el factor de estudio.
- Un factor de confusión potencial (i.e., un determinante de la enfermedad) no es necesariamente un verdadero factor de confusión – un verdadero factor de confusión debe estar asociado con el factor de estudio.
- El fenómeno de confusión puede ser controlado en el diseño de estudio y/o el análisis.

- El control a través del diseño de estudio se cumple a través de restricción, apareamiento (pre-estratificación), o aleatorización.
- El control en el análisis se cumple a través del análisis estratificado y/o el modelado matemático.
- La aceptabilidad del control se ve comprometida por errores conceptuales, en la medición, codificación, y especificación del modelo para los factores de confusión potenciales.
- El fenómeno de confusión trata con la "culpabilidad" o "inocencia"; la modificación del efecto trata con la "conspiración".
- El descubrimiento de que una asociación surge del fenómeno de confusión no lo hace menos "real", pero sí cambia su interpretación.
- La asociación cruda es real y para algunos fines es la medida relevante.

## Bibliografía

Rothman and Greenland (see index); Rothman, *Modern epidemiology*, pp. 177-181 and 226-229.

W. Dana Flanders and Muin J. Khoury. Indirect assessment of confounding: graphic description and limits on effect of adjusting for covariates. *Epidemiology* 1990; 1:239-246.

Greenland, Sander; Hal Morgenstern, Charles Poole, James M. Robins. Re: "Confounding confounding". Letter and reply by D.A. Grayson. *Am J Epidemiol* 1989; 129:1086-1091

Savitz, David A.; Anna E. Baron. Estimating and correcting for confounder misclassification. *Am J Epidemiol* 1989; 129:1062-1071.

Mickey, Ruth M; Sander Greenland. The impact of confounder selection criteria on effect estimation. *Am J Epidemiol* 1989; 129:125-37.

Stellman, Steven D. Confounding. *Preventive Medicine* 1987; 16:165-182 (from Workshop on Guidelines to the Epidemiology of Weak Associations)

Greenland, Sander and James M. Robins. Identifiability, exchangeability, and epidemiological confounding. *International Journal of Epidemiology* 1986; 15:412-418. (advanced)

Schlesselman, J.J.: Assessing effects of confounding variables. *Am J Epidemiol* 108(1):3-8, 1979.

Schlesselman, J.J. *Case-control studies*. Pp. 58-63.

Kleinbaum, Kupper and Morgenstern. *Epidemiologic research: principles and quantitative methods*. Chapter 13, Confounding.

Boivin, Jean-Francois; Sholom Wacholder. Conditions for confounding of the risk ratio and of the odds ratio. *Am J Epidemiol* 1985; 121:152-158.

Greenland, Sander; James M. Robins. Confounding and misclassification. *Am J Epidemiol* 1985; 122:495-506.

Kupper, Larry L. Effects of the use of unreliable surrogate variables on the validity of epidemiologic research studies. *Am J Epidemiol* 1984; 120:643-8.

Greenland, Sander. The effect of misclassification in the presence of covariates. *Am J Epidemiol* 1980; 112:564-9.

Greenland, S. and Neutra R.: Control of confounding in the assessment of medical technology. *International J Epidemiology* 9:361-367, 1980.

## **Raza/etnia**

Amott T, Matthaei J. *Race, gender, and work: a multicultural economic history of women in the United States*. Boston, South End Press, 1991.

Crow JJ, Escott PD, Hatley FJ. *A history of African Americans in North Carolina*. Raleigh, Division of Archives and History, N.C. Department of Cultural Resources, 1992.

Franklin, John Hope. *From slavery to freedom: a history of African Americans*. 7th ed. NY, McGraw-Hill, 1994.

Freeman HP. The meaning of race in science – considerations for cancer research. *Cancer* 1998;82:219-225.

Gamble, Vanessa N. Under the shadow of Tuskegee: African Americans and health care. *Am J Public Health* 1997;87:1773-.

Greenland, Sander; Judea Pearl, James M. Robins. Causal diagrams for epidemiologic research. *Epidemiology* 1999;10:37-48.

Hacker A. *Two nations: black and white, separate, hostile, unequal*. NY: Macmillan; 1992.

Kaufman, Jay S.; Sol Kaufman. Assessment of structured socioeconomic effects on health. *Epidemiology* 2001;12:157-167.

Krieger ND, Rowley DL, Herman A. Racism, sexism and social class: implications for studies of health, disease, and wellbeing. *Am J Prev Med* 1993;9(6[Suppl]):82-122.

Morbidity/mortality gap: is it race or racism? American College of Epidemiology Tenth Annual Scientific Meeting, *Ann Epidemiol* 1993;3:119-206.

Moss, Nancy. What are the underlying sources of racial differences in health? Editorial. *Ann Epidemiol* 1997;7:320.

Osborne NG, Feit MD. The use of race in medical research. *JAMA* 1992;267:275-279.

Stepan N. *The idea of race in science*. London, Macmillan, 1982.

Senior P, Bhopal RS. Ethnicity as a variable in epidemiologic research. *British Med J* 1994;309:327-330.

Varma J. Eugenics and immigration restriction: lessons for tomorrow. *JAMA* 1996;275:734.

Warnecke, Richard B., et al. Improving question wording in surveys of culturally diverse populations. *Ann Epidemiol* 1997;7:334-.

Williams, David R. Race and health: basic questions, emerging directions. *Ann Epidemiol* 1997;7:322-333.

## Apéndice

La siguiente discusión es para los estudiantes más avanzados (sea actualmente o cuando tomes un curso más avanzado de metodología) – los demás pueden saltar esta sección.

### ***El fenómeno de confusión : “Comparabilidad” versus “colapsibilidad”***

Como ya ha sido discutido, la teoría de la comparabilidad define una situación como presentando el fenómeno de confusión cuando la distribución del daño para el grupo no expuesto difiere de la distribución (contrafactual) para el mismo resultado en el grupo expuesto si éste pudiera ser observado sin la exposición. La definición bajo la teoría de la colapsibilidad considera al fenómeno de confusión como una situación en que la medida de asociación cruda difiere del valor de esa medida cuando se controla para variables extrañas (por medio de estratificación, ajuste o modelado matemático). Las dos definiciones llevan a la misma conclusión en muchas situaciones, siendo una excepción fundamental el caso en que la medida de asociación es un odds ratio que no estima una razón de riesgos (no cumple con el supuesto de enfermedad rara) o una razón de tasas (no se cumplen los supuestos para estimar la RDI).

La razón por la cual el odds ratio es diferente a las razones de tasas y de riesgo en estos casos se relaciona con el hecho de que, a diferencia de las proporciones y las tasas, el odds para un grupo no es un simple promedio de los odds individuales de los integrantes (Greenland S. *Am J Epidemiol* 1987;125:761). El análisis estratificado simplemente coloca a los integrantes individuales de un grupo en un reducido número de estratos. Dado que la incidencia para un grupo sí es igual al simple promedio de los riesgos de los integrantes individuales, la incidencia global (en expuestos, noexpuestos, or todos) también será igual al promedio de los riesgos o tasas específicas por estrato, ponderados por el tamaño del estrato (número de expuestos, número de no expuestos, o número total) como proporción del total (i.e., la distribución de participantes entre los estratos).

Para un riesgo o una tasa por lo tanto, la comparación (por medio de una razón o una diferencia) de la incidencia global en los expuestos con respecto a la incidencia global en los no expuestos es una comparación de promedios ponderados. Si los pesos en los grupos expuesto y no expuesto son los mismos, la comparación es válida (i.e., no hay fenómeno de confusión). En este caso, la razón de incidencia global (o la diferencia) es un promedio ponderado de las razones de incidencia (o de las diferencias) entre los estratos, una condición para evitar el fenómeno de confusión propuesto por Boivin y Wacholder (*Am J Epidemiol* 1985;121:152-8) e implica colapsabilidad. Dado que los pesos son las distribuciones de los participantes expuestos y no expuestos entre los estratos, pesos iguales significan distribuciones idénticas, lo cual a su vez significa que la exposición no está relacionada con el factor de riesgo utilizado para crear los estratos.

Si las distribuciones de los participantes expuestos y no expuestos entre los estratos son diferentes (i.e., la exposición está relacionada con la variable de estratificación), la incidencia global en los participantes expuestos y en los no expuestos son promedios basados en diferentes pesos, de manera que la razón y la diferencia entre ellos no será igual a un promedio ponderado de las razones o diferencias de incidencias específicas por estrato. La comparabilidad y la colapsibilidad por lo tanto

no están presentes, y la comparación entre incidencia globales sufre el efecto del fenómeno de confusión por el factor de estratificación. Sin embargo, dado que el odds para el grupo no es un simple promedio del odds para los integrantes individuales, nada de lo dicho anteriormente se mantiene para el odds ratio salvo que sea suficientemente raro como para aproximarse a la razón de riesgos o ha sido obtenido a partir de un diseño que produce un odds ratio que estima la razón de densidad de incidencia.

Algunas de las relaciones que acabamos de presentar pueden ser fácilmente demostradas utilizando álgebra simple. Aceptemos que  $a_i, b_i, c_i, d_i, n_{1i}$ , and  $n_{0i}$  en cada estrato tomen los valores que se detallan en la siguiente tabla, y que sus respectivos totales de todos los estratos sean representados por  $a, b, c, d, n_1$ , y  $n_0$  (i.e.,  $a$  = todos los casos expuestos,  $b$  = todos los casos no expuestos,  $c$  = todos los sanos expuestos,  $d$  = todos los sanos no expuestos,  $n_1$  = todas las personas expuestas,  $n_0$  = todas las personas no expuestas).

Enfermedad	Exposición		Total	
	Si	No		
Si	$a_i$	$b_i$	$m_1$	$(a_i + b_i)$
No	$c_i$	$d_i$	$m_2$	$(c_i + d_i)$
Total	$n_{1i}$	$n_{0i}$	$n_i$	
	$(a_i + c_i)$	$(b_i + d_i)$		

La incidencia en las personas expuestas es  $a_i/n_{1i}$  dentro de cada estrato y  $a/n_1$  cuando se ignoran los estratos (i.e., el total, o tabla bruta). La incidencia promedio (promedio) en los expuestos en los estratos es:

$$\sum \left( \frac{n_{1i}}{n} \times \frac{a_i}{n_{1i}} \right) = \sum \left( \frac{a_i}{n} \right) = \frac{a}{n}$$

en que la sumatoria se realiza sobre todos los estratos.  $a/n$  es simplemente la incidencia cruda en los expuestos. De igual manera, el promedio ponderado de las razones de riesgo específicas por estrato se puede expresar como la suma a través de los estratos de:

$$\frac{w_i}{W} \times \frac{a_i/n_{1i}}{b_i/n_{0i}} = \frac{w_i}{W} \times \frac{a_i n_{0i}}{b_i n_{1i}}$$

donde  $w_i$  son los pesos para cada estrato y  $W$  es la suma de las  $w_i$ . Si dejamos que  $w_i = b_i n_{1i} / n_{0i}$ , entonces tenemos la suma de todos los estratos de

$$\frac{b_i n_{1i} / n_{0i}}{W} \times \frac{a_i n_{0i}}{b_i n_{1i}} = \sum \left( \frac{a_i}{W} \right) = \frac{a}{W}$$

Mientras tanto,  $W$  es la suma de todos los estratos de:

$$w_i = \frac{b_i n_{1i}}{n_{0i}}$$

Si la exposición no se relaciona con las variables de estratificación, de manera que la distribución de los expuestos  $n_{1i} / n_1$  es igual que la distribución de los no expuestos  $n_{0i} / n_0$ , en todos los estratos, entonces la razón de expuestos a no expuestos en todos los estratos debe ser igual que en la tabla global,  $n_1 / n_0$ . Por lo tanto

$$w_i = \frac{b_i n_{1i}}{n_{0i}} = \frac{b_i n_1}{n_0}, \quad \text{cuya suma es simplemente} \quad W = \frac{b n}{n}$$

Por lo tanto, la suma de  $\frac{a_i}{w_i}$  es  $\frac{a}{b n_1 / n_0}$ , es igual a

$$\frac{a / n_1}{b / n_0}, \quad \text{la razón de riesgos global.}$$

De manera que, cuando no hay fenómeno de confusión, las siguientes tres medidas de resumen son todas equivalentes:

$$\text{Razón de riesgo (o tasa) global} = \frac{\text{Incidencia global en expuestos}}{\text{Incidencia global en no expuestos}}$$

$$\begin{aligned}
& \text{Promedio ponderado de la incidencia en los expuestos, en} \\
& \text{todos los estratos} \\
= & \frac{\hspace{10em}}{\hspace{10em}} \\
& \text{Promedio ponderado de la incidencia en no expuestos, en} \\
& \text{todos los estratos} \\
= & \text{Promedio ponderado de las razones de riesgo (o tasas)}
\end{aligned}$$

Con el odds de incidencia y el odds ratio, sin embargo, lo anterior no se puede utilizar. El odds global de incidencia es simplemente  $a/c$ . Por lo contrario, el promedio de los odds específicos por estrato, ponderados por el número de expuestos, es la suma en todos los estratos de:

$$\frac{n_{1i}}{n_1} \times \frac{a_i}{c_i}$$

Es posible construir un odds ratio de incidencia que es un promedio ponderado de los *odds ratio* de incidencia específicos por estrato, y por lo tanto un odds ratio resumen. Sin embargo, este odds ratio de incidencia resumen no será igual a una razón de *odds* de incidencia promedio específico por estrato para los expuestos sobre el *odds* de incidencia promedio específico por estrato para los no expuestos.