

## 7. Relacionando los factores de riesgo con los resultados de salud

*Cuantificando las relaciones entre dos factores o un factor y la ocurrencia, presencia, severidad o evolución de la enfermedad*

### **La imagen “grande”**

En este punto del curso, sería bueno ver donde estamos y a donde vamos. Después de una breve visión panorámica de la población y la salud, hemos considerado cuidadosamente el fenómeno de la enfermedad con respecto a la manera en que los epidemiólogos estudian la enfermedad. En ese capítulo examinamos los temas de definición, clasificación e historia natural. Luego nos dedicamos al asunto de cómo medir la frecuencia de la enfermedad y su extensión en la población. También examinamos algunos aspectos generales de la utilización de los números y de la estadística descriptiva, y luego nos dedicamos a las medidas epidemiológicas fundamentales, la prevalencia y la incidencia, con esta última enfocada como proporción o como tasa. De ahí seguimos con el tema de la estandarización, que facilita las comparaciones de las prevalencias e incidencias entre poblaciones con distinta composición demográfica, y vimos como estas medidas y conceptos diversos relacionados se usan en epidemiología descriptiva y vigilancia epidemiológica.

En la siguiente sección del curso nos preocuparemos por ver como investigar las asociaciones entre los resultados de salud y los potenciales factores de riesgo. Esa tarea involucra cuestiones de diseño de estudio, medidas de asociación, validez, inferencia e interpretación. Los temas de diseños de estudio y medidas de asociación están tan entrelazados que no importa con cual comencemos, siempre parece que tendríamos que haber comenzado con el otro! Los estudios analíticos proveen los datos para estimar las medidas de asociación y de impacto, pero las medidas de asociación y de impacto son las que motivan el diseño de los estudios.

Sin embargo, el enfoque epidemiológico básico que relaciona los factores de riesgo a los resultados de salud es más general que lo específico de cada uno de los temas. Consideremos una población en que una enfermedad u otra condición ocurre en toda la población pero más frecuentemente en personas con la característica A. Es probable que nos interese como se compara la existencia (prevalencia) o ocurrencia (incidencia) de la enfermedad en las personas con la característica A con la de la población general y con personas con alguna otra característica B (que puede ser simplemente la ausencia de A). Para hacer esta comparación:

- a. Medir la frecuencia - prevalencia, IA, DI - de la enfermedad o la condición en cada grupo (y talvez en la población entera);
- b. Comparar las frecuencias (en forma justa! - p.ej., después de estandarizar si es necesario))
- c. Cuantificar la comparación con una medida de asociación
- d. Cuantificar el impacto potencial de la característica en la condición, si estamos dispuestos a plantear una relación causal.

Ya hemos discutido las medidas de frecuencia y de extensión. Ahora veremos las medidas de asociación e impacto.

## ***Midiendo la fuerza de una relación***

La pregunta que resume los temas anteriores podría ser planteada como “¿Cuánto hay del factor? o “¿Cuán frecuente es la enfermedad (u otro fenómeno)?” Sin embargo, mucha de la epidemiología se preocupa por la relación entre factores, particularmente con el efecto de una “exposición” sobre una “enfermedad”. Por lo tanto el capítulo actual se aboca a contestar la pregunta “¿Qué tan fuerte es la relación entre dos factores?” o “ ¿Qué tan fuerte es la relación entre un factor de estudio y un resultado?” Una relación puede ser “fuerte” sin ser “causal”, y viceversa. Sin embargo, dos factores que están fuertemente asociados tienen más probabilidad de estar causalmente relacionados.

Hay varias maneras de evaluar la fuerza de la relación entre dos variables. Podemos, por ejemplo, evaluar la magnitud de la variación en una variable cuando varía la otra variable o, en forma equivalente, la magnitud de la modificación de la distribución de una variable con respecto a los diferentes valores de la otra variable. Para esta evaluación el epidemiólogo usa una medida de asociación\* .

Una segunda perspectiva es hasta qué punto el nivel de uno de los factores puede deberse al valor del segundo factor, como en la pregunta de cuánto de una enfermedad es atribuible a un factor que influye en su ocurrencia. Los epidemiólogos usan medidas de impacto para contestar esta pregunta.

La mayor parte de las medidas que veremos en este capítulo tienen que ver con las relaciones entre un factor que es dicotómico (binario, de dos valores posibles) y una medida de frecuencia o alcance, en particular, una tasa, riesgo, u odds. Dichas medidas son las más frecuentemente usadas en epidemiología. También veremos brevemente medidas que se usan en otras situaciones.

## **Medidas de asociación**

Una medida de asociación indica qué tan fuerte es la variación conjunta de dos factores en estudio. Cuanto más fuertemente están ligados, mayor es la evidencia que están causalmente relacionados uno con el otro (aunque no necesariamente que uno causa el otro, dado que ambos pueden ser causados por un tercer factor).

*Asociación* - dos factores están asociados cuando la distribución de uno es distinta para algún valor del otro. El decir que dos factores están asociados significa, esencialmente, que conociendo el valor de uno, implica una distribución diferente del otro. Consideremos las siguientes dos tablas (hipotéticas)

---

\* Aunque este término y “medida del efecto” han sido usados frecuentemente en forma indistinta (p.ej., en este texto) Rothman y Greenland (2000:58-59) señalan la siguiente distinción: las asociaciones involucran comparaciones entre grupos o poblaciones; los efectos involucran comparaciones de la misma población (teóricamente) observada en dos situaciones diferentes; las medidas de asociación se usan clásicamente para estimar las medidas de efecto.

Enfermedad cardiaca coronaria (ECC) y anticonceptivos orales (ACO) en mujeres de 35 años y más			Cáncer de mama y ACO				
	ACO	No ACO	Total		ACO	No ACO	Total
CHD	30	20	50	Cáncer	15	35	50
No-case	30	70	100	No-caso	30	70	100
Total	60	90	150	Total	45	105	150

Consideremos la primer tabla de la izquierda (ECC y ACO). La proporción general de usuarias de ACO es de  $60/150 = 0.40$ , pero entre los casos de ECC esta proporción es  $30/50 = 0.60$  mientras que entre las no enfermas es  $30/100 = 0.30$ . La distribución de los valores de usuarias de ACO (“usuarias” y “no usuarias”) es por lo tanto diferente para los distintos valores de presencia de ECC (“caso”, “no caso”). De igual manera, la distribución de los valores de presencia de ECC es distinta para los diferentes valores de ACO (30/60 de las usuarias de ACO tienen ECC; 20/90 de las no usuarias de ACO tienen ECC).

Si nos pidieran que estimáramos la proporción de usuarias de ACO en una muestra de 40 mujeres seleccionadas al azar de la tabla a la izquierda, ¿querríamos saber cuántas en la muestra tienen ECC y cuántas no? Seguro que sí.

Sabemos que la proporción de usuarias de ACO no debe ser menor de 0.30 (si la muestra fuera toda de no casos) y no mayor de 0.60 (si la muestra es toda de casos). En ausencia de esta información, nuestra mejor estimación sería la proporción global de la población, 0.40. Pero si supiéramos la proporción de casos en la muestra, podríamos ajustar nuestra estimación hacia un valor mayor (si más de un tercio fueran casos) o a valores menores (si menos de un tercio fueran casos). [Ahora, verifica que para estimar la proporción de casos de ECC en una muestra al azar, nos gustaría saber la proporción de usuarias de ACO. ¿Cuál es la mejor estimación de la proporción con ECC si la muestra es de 22 usuarias de ACO y 18 no usuarias? – La respuesta está al final del capítulo.]

Así, en los datos en la tabla de la izquierda, hay una asociación entre el uso de ACO y la ECC. En contraste, en la tabla a la derecha (Cáncer de mama y ACO), la distribución de usuarias de ACO es la misma para los casos que para los no casos, y para todo el grupo. Por lo tanto, los datos de la tabla de la derecha no muestran asociación entre el cáncer de mama y el uso de ACO.

## Correlación y Concordancia

**Asociación** es un término general que engloba muchos tipos de relaciones. Otros términos son utilizados para indicar tipos específicos de asociación. Dos de ellos que son importantes son:

**Correlación** es un tipo de asociación en que la relación es monotónica, i.e., va en una sola dirección - cuanto más hay de un factor, más hay del otro (correlación directa o positiva), o cuanto más hay de un factor, menos hay del otro (correlación negativa o inversa). La correlación lineal (medida por el

coeficiente de correlación producto-momento de Pearson) evalúa en que medida la relación puede ser resumida en una línea recta. Los coeficientes de correlación no-paramétricos, como el coeficiente de correlación de rango de Spearman, evalúan en qué medida dos factores están correlacionados pero sin tomar en cuenta la magnitud del cambio en uno que acompaña el cambio en el otro, sólo considera la dirección del cambio.

**Concordancia** es un tipo de correlación en que los dos factores (generalmente dos mediciones del mismo fenómeno) no sólo están directamente correlacionados uno con el otro sino que tienen los mismos valores. Por ejemplo, dos esfigomanómetros deben dar los mismos valores cuando son usados en la misma persona en la misma oportunidad, no solamente valores correlacionados. Dos mediciones de una fenómeno estable deben estar de acuerdo una con otra, no meramente correlacionarse. Si se sabe que una de las medidas es muy precisa y la otra está siendo evaluada, podemos evaluar la validez de esta última, más que simplemente la concordancia entre las dos.

## APARTADO

### **Algunos comentarios sociológicos**

Dado que los factores estudiados por los epidemiólogos son a menudo la ocurrencia de enfermedad y la presencia de exposición, las principales medidas epidemiológicas son proporciones y tasas de enfermedad entre distintos grupos de exposición. De hecho, como estas medidas son tan familiares para epidemiólogos y clínicos, aún cuando la enfermedad (p.ej., presión arterial) y/o la exposición no están representados por variables dicotómicas (de dos categorías), es habitual convertirlas en proporciones o tasas para, por lo menos, algunos análisis. Por lo tanto pasaremos la mayor parte de nuestro tiempo con las medidas de asociación y de impacto que involucran tasas y proporciones. Hay que tener en cuenta, sin embargo, que en el caso de los fenómenos que son capaces de ser medidos cuantitativamente (p.ej., medidas fisiológicas, consumo de nutrientes, exposiciones ambientales) es más correcto analizarlos sin dicotomizarlos.

La preferencia por las tasas y las proporciones es una razón para los distintos enfoques del análisis estadístico usado por los epidemiólogos y los científicos sociales que también estudian datos sobre poblaciones. Pero hay otras diferencias en el enfoque que presumiblemente tienen una base diferente, talvez los enfoques de los epidemiólogos sobre las relaciones biológicas.

Una fuente potencial de confusión – y hasta de conflicto! – es la diferencia en la manera que los epidemiólogos por un lado, y los científicos sociales y los bioestadísticos, miran las asociaciones. Los epidemiólogos tienden a mirar la fuerza de la asociación como un tema separado de la evidencia numérica de que la asociación no hubiera surgido fácilmente por azar (i.e., su “significancia estadística”). Otras profesiones, sin embargo, miran primero la significancia estadística de una asociación antes de considerar cualquier otra característica. Así un bioestadístico o un psicólogo pueden descartar completamente una asociación que un epidemiólogo puede caracterizar como “fuerte aunque potencialmente debida al azar”. A la inversa, un psicólogo o un bioestadístico pueden caracterizar una asociación como “altamente significativa” cuando los epidemiólogos la descartan por ser demasiado débil para

tener significado biológico. Como veremos más adelante, varias medidas de asociación usadas en estadística (p.ej., pruebas de Chi cuadrada, coeficientes de correlación) están en una categoría de medidas de asociación diferente a la de las medidas de asociación que veremos a continuación.

## FINAL DEL APARTADO

### **Algunas medidas básicas**

Antes de profundizar nuestra discusión sobre como medir asociaciones, deseamos empezar con algunas cosas básicas. Supongamos que se le pidiera a un epidemiólogo que investigue el posible riesgo de un sistema inadecuado de filtrado de aire en un gran edificio que aloja una escuela pública en un barrio urbano de bajos recursos. La preocupación principal son los niños con asma, de los cuales 400 van a la escuela (escuela A). Al epidemiólogo se le informa que un día particular 12 niños tuvieron una crisis de asma, mientras que en una escuela cercana (escuela B) con 500 niños asmáticos, solo 5 sufrieron una crisis asmática ese mismo día.

El epidemiólogo primero resume los datos en una tabla  $2 \times 2$  :

#### **Incidencia acumulada de crisis asmáticas durante una jornada escolar**

Tuvieron una crisis asmática	Escuela A	Escuela B	Total
Si	12	5	17
No	388	495	883
Total	400	500	900

El primer paso es calcular la incidencia en cada escuela:

Incidencia acumulada de 1 día en escuela A:  $12 \text{ casos} / 400 \text{ niños en riesgo} = 0.03$  o 3%

Incidencia acumulada de 1 día en escuela B:  $5 \text{ casos} / 500 = 0.01$  o 1%

La escuela A de hecho tiene una incidencia mayor de crisis asmáticas el día investigado.

Para evaluar la fuerza de asociación entre la escuela y la incidencia de asma, el siguiente paso es calcular una medida de la fuerza de asociación. La medida más frecuentemente calculada en esta situación es la razón de dos incidencias acumuladas (la razón de “incidencias acumuladas”, RIA también conocida como “razón de riesgos”) [en inglés “cumulative incidence ratio”, CIR, también conocido como “risk ratio”]. La RIA es simplemente  $0.03/0.01 = 3.0$ , que a menudo se interpreta como que es una asociación “moderadamente fuerte”. El epidemiólogo puede también calcular la diferencia entre las incidencias acumuladas (“diferencia entre incidencias acumuladas” DIA) [ en inglés “cumulative incidence difference”, CID], e informar que el tener un filtro de aire inadecuado

se asocia con dos puntos porcentuales más en la incidencia de asma durante el días escolar de 7 horas. Con este ejemplo básico, examinemos los conceptos que apoyan estas medidas.

## ***Efectos absolutos versus relativos***

Cuando tenemos proporciones o tasas de incidencia de dos poblaciones diferentes (p.ej., usuarios de PC y usuarios Mac), es fácil decir cual de las tasas es mayor. Pero cuantificar cuánto más plantea el problema de cómo comparar las dos tasas. Una pregunta básica es si la magnitud por la cual la tasa mayor excede la menor debe ser, o no, relativa al tamaño de una de las tasas.

Si le preguntamos a una niña de 10 años cuanto mayor es que su hermanito de 5 años, probablemente contestará “5 años”. Pero si tiene facilidad para las matemáticas, ella talvez conteste que “tiene el doble de su edad” o que es “100% mayor”. Ambas respuestas cuantifican con precisión la magnitud de la diferencia de edad, sin embargo tienen distintos “sentidos”. ¿Tenemos razones para preferir uno sobre el otro?

Podríamos inclinarnos a la respuesta “5 años”. “Podríamos”, porque la selección de una medida depende de nuestro objetivo, y no hemos especificado un objetivo. Pero se nos ocurren dos razones por las cuales podríamos preferir la diferencia absoluta (5 años) a la diferencia relativa (100% mayor) o la razón (el doble de su edad).

Por un lado, “5 años” será preciso indefinidamente, mientras que “el doble” (o “100% más”) son precisos sólo este año. En ese sentido “5 años” es un mejor resumen de la relación entre las edades respectivas de los niños. Por otro lado, el crecimiento y envejecimiento humano, por lo menos desde el punto de vista social y talvez también desde el punto de vista biológico, son procesos que están marcados por aumentos absolutos, no relativos. Por ejemplo, generalmente consideramos el ingreso a la escuela y la graduación, la pubertad, la edad de obtener la libreta de conducir, la presbicia, y la jubilación en términos de rangos específicos de edad, no de aumentos proporcionales. Decimos “en 15 años probablemente necesites bifocales”, y no “cuando tu edad sea 50% mayor”. Al contrario, cuando ajustamos una receta para un número mayor o menor de invitados, multiplicamos o dividimos las cantidades de cada ingrediente por un factor común, y no restamos una cantidad común de cada uno. Para adaptar la receta, estamos interesados en aumentos proporcionales (relativos).

De manera similar, cuando cuantificamos la comparación de dos incidencias (o dos prevalencias), podemos tomar la diferencia absoluta (diferencia de incidencia) o la diferencia relativa, (riesgo excesivo). ¿Cuál, absoluta o relativa, es de mayor interés para nosotros en cuantificar la comparación de las dos medidas de ocurrencia o alcance? Esta pregunta ha inspirado mucho debate en los primeros tiempos de la epidemiología moderna (circa 1955) y, como pasa con tanta frecuencia, se puede argumentar a favor de cualquiera de las dos posiciones. La opción depende de nuestro objetivo, nuestro concepto del fenómeno, y la disponibilidad de datos.

Un problema con usar la diferencia absoluta (llamada también “diferencia de riesgo”, “diferencia de tasas”, “diferencia de incidencias acumuladas”, “riesgo atribuible”, según la moda, el grupo de epidemiólogos con el cual el epidemiólogo quiere identificarse, la década en la cual aprendió

epidemiología, o según que la comparación involucre tasas de incidencia, proporciones de incidencia, prevalencias, o tasas de mortalidad) como medida de asociación es que si las propias incidencias son pequeñas, como siempre será el caso para una enfermedad rara, entonces la diferencia también será pequeña. Por ejemplo, si la tasa de mortalidad anual de una enfermedad rara como el cáncer de esófago es de 60/100,000 en personas con bajo consumo de vitamina C y 20/100,000 en las personas con alto consumo de vitamina C, la diferencia es sólo 40/100,000. Por el contrario, la diferencia en una asociación que involucra una enfermedad más frecuente, como el consumo de vitamina E y ECC, puede ser de 1,200/100,000 para el bajo consumo y 800/100,000 para el alto consumo de vitamina E = 400/100,000, una diferencia de mayor magnitud.

El tamaño mucho mayor de la segunda diferencia indica que si estas dos vitaminas son factores causales se salvarían muchas más vidas aumentando el consumo de vitamina E que el de vitamina C. La vitamina E parece tener un mayor impacto de salud pública. Pero, ¿sería lógico concluir a partir de la mayor diferencia para la vitamina E que su asociación con ECC es más fuerte que la de la vitamina C con cáncer de esófago? En primer lugar, si llegáramos a esa conclusión implicaría que casi cualquier asociación que tiene que ver con una enfermedad frecuente debe ser más fuerte que todas las asociaciones que involucran enfermedades muy raras. En segundo lugar, dado que la incidencia verdadera de la mayor parte de las condiciones varía según todo tipo de factores (edad, género, recursos económicos, hábito de fumar, consumo de alcohol, actividad física, dieta, genética, cofactores) la diferencia absoluta muy probablemente varíe, posiblemente en forma importante, entre poblaciones (sin embargo, la diferencia relativa también puede variar).

Al contrario, la expresión de la diferencia de incidencias en relación con el tamaño de las incidencias mismas, produce medidas de asociación que parecen comparables. De esta manera podemos calcular una diferencia relativa en la incidencia de la mortalidad por cáncer de esófago en relación con la vitamina C como  $(I_1 - I_0)/I_0 = (0.00060 - 0.00020)/0.00020 = 2.0$  y una diferencia relativa para la mortalidad por ECC en relación con la vitamina E es  $(I_1 - I_0)/I_0 = (0.01200 - 0.00800) / 0.00800 = 0.50$ . En base a esto, la asociación que involucra la vitamina C es sustancialmente mayor que la que involucra la vitamina E. Esta medida de diferencias relativas a menudo se conoce como exceso de riesgo (o “exceso de tasa”, dado que los datos son tasas, no proporciones). Si agregamos 1.0 al exceso de riesgo o de tasa, obtenemos una medida relativa aún más simple,  $I_1/I_0$ , que se conoce como riesgo relativo, razón de riesgos, razón de tasas, razón de incidencias acumuladas, razón de densidades de incidencia, o, en el caso de la prevalencia, razón de prevalencias.

### ***Medidas de asociación relativas versus absolutas***

Aquí encontramos dos ejemplos de la vida real que muestran el contraste entre las medidas de asociación relativas y absolutas. El primero se basa en los datos de un estudio de seguimiento por Mann y cols, (presentado en un seminario en la Universidad de Carolina del Norte por Bruce Stadel):

**Incidencia de Infarto Agudo de Miocardio (IAM)  
en usuarias de ACO, por 100,000 años-mujer.**

Edad (años)	Cigarrillos/día	Usuarias de ACO			
		No usuarias	RR**	RA***	
30-39	0-14	6	2	3	4
	15 +	30	11	3	19
40-44	0-14	47	12	4	35
	15 +	246	61	4	185

Notas:

\* RR= riesgo relativo (razón de tasas)

\*\* RA= riesgo atribuible (diferencia de tasas, diferencia absoluta)

En esta tabla la incidencia de IAM es claramente mayor para las usuarias de ACO, dado que en cada estrato de calidad de fumadora las usuarias de ACO tienen una mayor incidencia (DI) que las no usuarias. Es más, la razón de las dos incidencias (el RR) es casi constante en los estratos, una propiedad deseable para una medida de resumen, mientras que la diferencia de tasas (RA) varía de manera importante. Según Breslow y Day, la razón de tasas tiende a ser más estable entre los estratos, apoyando su atractivo como medida de asociación. No todos los epidemiólogos cuantitativos están de acuerdo con esta aseveración.

El segundo ejemplo viene de un estudio de seguimiento de cáncer de pulmón y enfermedad de las arterias coronarias con relación al hábito de fumar:

**Tasas de mortalidad por 100,000 años-persona por cáncer de pulmón  
y enfermedad de las arterias coronarias para fumadores y no fumadores**

	Fumadores	No fumadores	Razón	Diferencia
Cáncer de pulmón	48.3	4.5	10.8	44
Enfermedad de las arterias coronarias	294.7	169.5	1.7	125

Fuente: 1964 *Surgeon General's Report on Smoking and Health*, página 110, citado en Joseph Fleiss, *Statistical methods for rates and proportions*, 2nd edición, página 91

La razón de tasas para la relación entre el hábito de fumar y la mortalidad por cáncer de pulmón es mucho mayor que entre el hábito de fumar y la mortalidad por coronariopatía. Estas cifras habitualmente se interpretan como que el cáncer de pulmón está más asociado con el hábito de fumar que con la mortalidad por coronariopatía; la eliminación del hábito de fumar llevaría a una disminución proporcionalmente mayor de cáncer de pulmón que la disminución proporcional en la mortalidad por enfermedad coronaria. Sin embargo, la disminución en el número de muertes por cáncer de pulmón sería más pequeña en magnitud que las muertes por coronariopatía.

Profundizaremos en estos detalles más adelante en este capítulo en la sección de Medidas de Impacto.

## **Concepto de riesgo relativo**

No obstante, en general usamos el riesgo relativo como la medida básica de fuerza de asociación entre una característica y el desarrollo de una condición.

El concepto de riesgo relativo es operacionalizado por :

- a. La razón de incidencias acumuladas, (RIA), también llamada razón de riesgo
- b. Razón de densidad de incidencia, (RDI), también llamada razón de riesgos
- c. Odds ratio (OR), que en ciertas circunstancias estima la RIA y la RDI.

Fórmula general:

$$\text{Razón de incidencia} = \frac{\text{Incidencia en "expuestos"}}{\text{Incidencia en "no expuestos"}} = \frac{I_1}{I_0}$$

Talvez recuerden que vimos en el capítulo de estandarización, que la RME puede ser considerada una razón de tasas de mortalidad “observadas” sobre “esperadas”. De hecho, podemos usar el concepto de observado y esperado aquí también. Cuando contrastamos las tasas de incidencia en los grupos expuestos y no expuesto, estamos usando, típicamente, la incidencia de los no expuestos como el barómetro de la incidencia que podríamos encontrar en el grupo expuesto si la exposición no tuviera efecto. En ese sentido, la incidencia en los no expuestos hace de “esperado”, mientras que la incidencia en los expuestos hace de “observado”.

El concepto de riesgo relativo también puede ser usado en situaciones en que las estimaciones de incidencia no están disponibles o ni siquiera son de mayor interés. Por ejemplo, se puede obtener una estimación directa de la razón de incidencia de un estudio caso control con casos incidentes (que recién ocurren), si los controles son seleccionados de manera adecuada (como se explica en el capítulo sobre Diseños de Estudios Analíticos). En situaciones en que queremos razones de incidencia pero solo hay disponibles datos de prevalencia, la razón de prevalencia o el odds ratio de prevalencia pueden darnos una solución. La razón para ello es la relación entre la prevalencia, la incidencia, y la duración que presentamos en el capítulo sobre Medición de la enfermedad y la Exposición (en una población estable, odds de prevalencia = incidencia × duración promedio o en el caso de un resultado raro, prevalencia ≈ incidencia × duración promedio). Una pregunta clave que hay que hacerse es si la duración es igual en todos los grupos que están siendo comparados, dado que si no fuera así, una comparación de prevalencias daría una imagen distorsionada de la comparación de incidencias.

La razón de prevalencias puede también ser la elección lógica para cuantificar asociaciones entre las exposiciones y las condiciones cuya duración es tanto o más importante que su incidencia. Por ejemplo, una proporción importante de la población experimenta emociones o condiciones como ansiedad, fatiga, o tristeza de vez en cuando. Dado que la prevalencia puntual va a contar

fundamentalmente las personas en las cuales la condición persiste, la prevalencia puede ser tan útil o más útil que la incidencia como medida de frecuencia en dichos casos. (La razón de prevalencias es también la elección simple para afirmaciones descriptivas sencillas como “el hábito de fumar era dos veces más frecuente en las personas con educación menor a secundaria”.)

## Interpretación del riesgo relativo

Ejemplo: razón de incidencia de 2.0 significa que:

- “La incidencia en la población expuesta es el doble de la de la población no expuesta”
- “La exposición esta asociada con un aumento del 100% en la incidencia.”
- “La exposición está asociada con una incidencia dos veces mayor” (aunque frecuente, esta interpretación probablemente deba ser evitada ya que “dos veces mayor” puede también ser interpretado como 200% mayor, que corresponde a una razón de incidencia de 3.0).

Adjetivos descriptivos para la magnitud de asociación (como se usan habitualmente )

1.0	No hay asociación (valor nulo)
1.1-1.3	Débil
1.4-1.7	Modesto
1.8-3.0	Moderado
3-8	Fuerte

Para las asociaciones inversas (razón de incidencia menor a 1.0), toma el recíproco y consúltalo en la tabla más arriba, p.ej. el recíproco de 0.5 es 2.0, que corresponde a una asociación “moderada”.

## Tablas dos por dos

La disposición de datos más básica en epidemiología es la tabla dos por dos:

Enfermedad	Exposición		Total	
	Si	No		
Si	a	b	m <sub>1</sub>	(a + b)
No	c	d	m <sub>2</sub>	(c + d)
Total	n <sub>1</sub>	n <sub>0</sub>	n	
	(a + c)	(b + d)		

Una de las controversias mayores en epidemiología es si la enfermedad se debe mostrar en las filas, como en la tabla anterior, o en las columnas. Kleinbaum, Kupper, y Morgenstern usan el formato ya descrito. Hennekens y Buring ponen las categorías de enfermedad en las columnas y la exposición en las filas. Algunos autores usan una presentación para los estudios de cohortes y el otro para los

estudios caso control. Como podrás ver, la epidemiología no es aún verdaderamente una disciplina (o no está aún disciplinada).

La forma anterior de tabla  $2 \times 2$  se usa para presentar los datos de un estudio (p.ej., de cohortes, de corte, caso control) con datos de conteo. Cuando el estudio usa datos de años-persona (p.ej., para estimar la densidad de incidencia), entonces se quita la columna de “no enfermedad” y los totales de tiempo-persona ( $PY_1$ ,  $PY_0$ ) ocupan el marginal de la derecha:

Enfermedad	Exposición		Total
	Si	No	
Si	a	b	$m_1$
Tiempo-persona	$PY_1$	$PY_0$	PY

Una vez que tenemos nuestras tablas (cualquiera sea su orientación), definiremos ahora las tres principales medidas de riesgo relativo, sobre las cuales hay mucho menos controversia:

1. Razón de Incidencia Acumulada (RIA)
2. Razón de Densidad de Incidencia (RDI)
3. Odds ratio (OR)

**Razón de densidad acumulada (también llamada “razón de riesgos” o “riesgo relativo”)**

La razón de incidencia acumulada (RIA) responde la pregunta: “¿cuántas veces excede el riesgo en la personas expuestas, el riesgo de la personas no expuestas?” Si la RIA es 3, podemos decir que las personas expuestas tienen 3 veces el riesgo de las personas no expuestas. También podemos decir que el individuo expuesto promedio tiene tres veces el riesgo de enfermarse que el individuo no expuesto promedio. Esto es a veces justo lo que queremos saber. La definición matemática es:

$$\text{Razón de Incidencia Acumulada} = \frac{\text{Incidencia acumulada en los “expuestos”}}{\text{Incidencia acumulada en los “no expuestos”}} = \frac{IA_1}{IA_0}$$

Dado que la RIA se basa en estimaciones de incidencia acumulada o riesgo, la RIA puede ser estimada directamente sólo a partir de estudios de cohorte. Es posible, sin embargo, estimarla indirectamente en otras situaciones.

## **Razón de densidad de incidencia (también llamada “razón de tasas”)**

La razón de densidad de incidencia (RDI) responde a la pregunta “¿cuántas veces excede la tasa de enfermedad en las personas expuestas la de las personas no expuestas?”. Si la RDI es 3 podemos decir que la tasa en los expuestos es 3 veces la de los no expuestos. No hay una interpretación obvia a nivel individual, pero la RDI es de primordial importancia para los estudios de poblaciones dinámicas y cohortes prolongadas. La definición matemática es:

$$\text{Razón de densidad de incidencia} = \frac{\text{Densidad de incidencia en los "expuestos"}}{\text{Densidad de incidencia en los "no expuestos"}} = \frac{DI_1}{DI_0}$$

La RDI se usa en situaciones en que el resultado es el tiempo hasta la ocurrencia de un evento (p.ej., muerte) y es matemáticamente equivalente a la razón de riesgos instantáneos (“hazard ratio” en inglés) del análisis de sobrevida. La RDI puede ser estimada directamente en un estudio de seguimiento (de cohorte fija o una población dinámica).

## **(Riesgo) odds ratio**

El odds ratio (OR) es una razón de “odds”, que son transformaciones de riesgos o probabilidades.

$$\text{odds} = p/(1-p), \text{ donde } p = \text{probabilidad}$$

El OR responde a la pregunta “¿cuántas veces mayor es el “odds” de la enfermedad para las personas expuestas que para las no expuestas?” Dado que los odds tienen una escala de medida diferente a la del riesgo, la respuesta a la pregunta puede a veces ser diferente a la respuesta a la pregunta correspondiente sobre el riesgo. A menudo, sin embargo, estamos preocupados con enfermedades raras, para las cuales el riesgo y el odds son muy parecidos y las RIA y OR (y la RDI) también son muy parecidas. Dado que el OR puede ser definido en términos de odds de la enfermedad entre los expuestos o odds de exposición entre los casos, hay dos fórmulas matemáticas:

$$\text{Odds ratio} = \frac{\text{Odds en "expuestos"}}{\text{Odds en "no expuestos"}}$$

El odds es simplemente una transformación algebraica de la probabilidad, de manera que cualquier probabilidad (que debe, por supuesto, ser menor a 1.0) puede ser expresado como un “odds”. La probabilidad de que algo pueda pasar, sobretodo algo desagradable, a menudo es llamado un “riesgo”. Los odds derivados de un riesgo se denominan, apropiadamente odds de riesgo, de manera que una razón de dos odds de riesgo se llama una **razón de odds de riesgo**, o **ROR** (en inglés risk odds ratio, o ROR)

## Odds ratio (de exposición)

La prevalencia a menudo se conoce como una estimación de probabilidad (p.ej., de exposición). Este uso está justificado por el hecho de que si fuéramos a seleccionar un individuo al azar del grupo, la probabilidad de que el individuo tenga una cierta característica es estimada por la prevalencia en el grupo. El odds que corresponde a la probabilidad de exposición se llama “odds de exposición”, de manera que la razón es un **odds ratio de exposición**, o **ORE** (en inglés, exposure odds ratio, o EOR). Aunque desde el punto de vista conceptual estos dos odds ratio son muy distintos, para la tabla dos-por-dos son algebraicamente idénticos como veremos más adelante. Por lo tanto, nuestra posibilidad de estimar un odds ratio (de exposición) en una situación en que no conocemos la incidencia de la enfermedad es una herramienta potente para examinar *asociaciones* que involucran la incidencia de enfermedad aún cuando no tenemos datos de incidencia, como fue presentado por primera vez en un trabajo clásico por Jerome Cornfield (para más detalles ver el capítulo sobre los Diseños de Estudios Analíticos).

$$OR_r = \text{odds ratio de riesgo} = \frac{\text{Odds de riesgo en "expuestos"}}{\text{Odds de riesgo en "no expuestos"}} = \frac{odds_1}{odds_0} = \frac{IA_1 / (1-IA_1)}{IA_0 / (1-IA_0)}$$

$$OR_e = \text{odds ratio de exposición} = \frac{\text{Odds de exposición en "casos"}}{\text{Odds de exposición en "no casos"}} = \frac{odds_1}{odds_0}$$

## Relación de la razón de odds (odds ratio) con la razón de riesgos

Cuando las incidencias son pequeñas (i.e., el resultado en estudio es raro en la población), el odds ratio se aproxima mucho tanto a la razón de riesgos como a la razón de densidad de incidencia. La guía convencional para clasificar una enfermedad como “rara” es una incidencia por debajo de 10%. Una buena manera de evaluar el alcance de la divergencia entre el odds ratio y la razón de riesgos es la observación de una hoja de cálculo con ejemplos de incidencias y calcular los riesgos relativos y odds ratios (p.ej, las directivas sugeridas por Zhang y Yu [1998] para la incidencia por debajo de 10% y razón de riesgos por debajo de 2.5 permite que el odds ratio sea solamente 20% mayor que la razón de riesgos).

Si uno cree que el OR exagera la fuerza de asociación en forma inaceptable, es fácil derivar una estimación de la razón de riesgos correspondiente **si** uno tiene información adicional – prevalencia general de exposición, incidencia global de la enfermedad, incidencia de la enfermedad en los expuestos, o incidencia de la enfermedad en los no expuestos (Hogue, Gaylor, y Schulz, 1983). La conversión más simple se consigue si uno conoce la incidencia en el grupo no expuesto, p.ej.:

$$RR = \frac{OR}{(1 - IA_0) + (IA_0 \times OR)}$$

Donde  $IA_0$  es la incidencia en el grupo no expuesto [Zhang y Yu (1998), adaptado a la nomenclatura usada aquí]. Un odds ratio de prevalencia puede ser convertido en una razón de prevalencia usando la misma fórmula y sustituyendo la prevalencia en los no expuestos en lugar de  $IA_0$ . La divergencia entre el OR y la RDI generalmente será menor que la que hay entre OR y la RIA. Las razones para todo esto es que las tres medidas de incidencia (DI, IA, odds) tienen el mismo numerador (casos nuevos), pero a medida que aumenta la incidencia los denominadores de la DI y el odds disminuyen, mientras que el denominador de la IA no cambia.

### **Razones de proporciones versus razones de odds**

En los estudios caso control sin información adicional, el OR es a menudo la única medida de asociación que puede ser estimada. También, cuando el resultado es raro, las tres medidas de riesgo – el OR, la RIA y la RDI – tienen aproximadamente el mismo valor. En otras situaciones (i.e., datos de cohortes o de estudios de corte con resultados no raros), el hecho de que el OR sea una medida epidemiológica de asociación ha sido tema de mucha discusión.

Los que apoyan el OR señalan varias propiedades matemáticas deseables en relación a la razón de riesgos, incluyendo el hecho de que la fuerza de asociación no es afectada invirtiendo la definición del resultado (Walter, 2000). Por ejemplo, en un ensayo de cesación de fumar, el OR para el éxito será el recíproco del odds ratio del fracaso; la razón de “riesgos” (RIA) para el éxito, sin embargo, será muy distinta de la RIA del fracaso. Además, el odds ratio de prevalencia puede, en principio ser usado para estimar la razón de tasas de incidencia a partir de datos de un estudio de corte transversal, suponiendo que la duración de la enfermedad no está relacionada con la exposición y que las incidencias y la duración en los grupos expuesto y no expuesto han sido constantes durante suficiente tiempo como para adquirir una condición de estado estable. Es más, la popularidad de la regresión logística múltiple, que estima el OR controlando para múltiples variables (ver el capítulo sobre Análisis e Interpretación de Datos), ha sido una fuerte motivación para que muchos investigadores estimen odds ratios aún en estudios de cohortes en que la incidencia puede ser estimada directamente.

Sin embargo, a medida que han aparecido programas de computación para estimar la razón de incidencia acumuladas y la razón de prevalencias (p.ej., SAS PROC GENMOD), el uso del odds ratio en los estudios de cohortes o de corte transversal se ha vuelto menos aceptable, sobretodo para resultados no raros (Thompson, Myers, y Kriebel, 1997). Su valor en los datos provenientes de estudios de corte transversal está algo minimizado por la dificultad de aceptar que se mantiene el supuesto de la población estacionaria (estado quiescente).

Los críticos han llamado al OR “incomprensible” (Lee, 1994:201) y han dicho que le falta “inteligibilidad” (Lee y Chia, 1994). De hecho, después de que surgiera una controversia sobre el informe en las noticias de un estudio hecho por Kevin Schulman (Schulman y cols., 1999), los editores del *New England Journal of Medicine* se disculparon por haber permitido el uso del OR en el

resumen del trabajo (*New Engl J Med* 1999;341:287). Un informe de seguimiento en Brillscontent.com citaba a uno de los autores del trabajo que dijo que “salvo que seas un bioestadístico es poco probable que tengas la más mínima noción de lo que significa un OR. La verdad es que es confuso para mucha gente, incluyendo los médicos”.

En el estudio de Schulman *y cols.*, médicos de atención primaria que asistían a reuniones profesionales veían entrevistas grabadas en video de pacientes hipotéticos (en realidad actores) y recibían información médica adicional, y luego tenían que decidir si referirían o no los pacientes para cateterización cardiaca. Un hallazgo principal fue que los médicos recomendaron la cateterización en el 84.7% de los casos cuando el actor era afroamericano comparado con el 90.6% de los casos cuando el actor era americano-europeo. El hallazgo se presentó como un OR de 0.6, que fue informado en la prensa como que los pacientes de raza negra tenían “40% menos probabilidad” de ser referidos que los pacientes de raza blanca (ver tabla 2 en *Schwartz y cols.*, 1999 para un resumen de las noticias de prensa).

*Schwartz y cols.* (1999) explicaron que como el resultado era tan común, la razón de riesgos real (0.93, que indica una asociación débil) fue muy sobrestimada por el OR, que a su vez contribuyó a que la prensa sobrestimara la asociación. Sin embargo, la razón de riesgos para la **no** referencia es también 0.6 (0.09/0.15), indicando que los pacientes de raza blanca tenían un 60% de probabilidad de la de los pacientes negros de **no** ser referidos o que los pacientes negros tenían 60% más probabilidad de no ser referidos que los pacientes blancos (RR de 1.6 = 1/0.6). Así, si la impresión que dieron los medios de prensa fue exagerada o no, es discutible, por lo menos con respecto al OR (ver *Schwartz y cols.* por otras limitaciones del estudio).

Greenland (1987) asegura que la relevancia del OR para la epidemiología proviene exclusivamente de su capacidad para estimar la razón de tasas o la razón de incidencias acumuladas. Su objeción al OR como medida de efecto descansa en la falta de una correspondencia simple entre el odds para una población y el odds para un individuo. Mientras que la “proporción de incidencia” (p.ej., incidencia acumulada) es equivalente a un promedio simple del riesgo de cada individuo de la población y la densidad de incidencia es equivalente a un promedio sencillo del “peligro” (hazard) de cada individuo de la población, el odds de incidencia no es equivalente a un promedio simple de los odds de enfermedad de cada individuo de la población (Greenland, 1987). Así, el OR no es una razón de promedios interpretable a nivel individual. Resulta que esta propiedad del OR (la “no capacidad de colapsarse”) puede hacer que su uso sea engañoso cuando se intenta examinar una asociación controlando para otros factores (ver capítulo sobre Análisis e Interpretación de Datos).

Aunque uno podría refugiarse en la aseveración de que los “juicios cualitativos basados en la interpretación de odds ratio como si fueran riesgos relativos es poco probable que esté muy errada” (Davies, Crombie, y Tavolaki, 1998:991), es más seguro evitar el OR cuando las razones de incidencia o prevalencia pueden ser estimadas.

## **Dos supuestos típicamente no explicitados**

### **Estado de exposición estable**

La discusión anterior supone que la población en estudio es razonablemente estable con respecto a el estado de exposición. Cuando esto no es así, puede ser necesario cambiar el estado de exposición de los individuos durante el período de observación, asignando su seguimiento a uno u otro grupo de exposición, si se cree que el efecto de la exposición no persiste. Por ejemplo, un sujeto puede realizar actividad física, dejar de hacerlo y comenzar de nuevo. Si se cree que el efecto del ejercicio se termina poco después de dejar de realizar actividad física y a comenzar de nuevo poco después de retomar la actividad, el tiempo de seguimiento (tiempo-persona) puede ser sumado a la categoría apropiada de práctica de actividad física para cada parte del período de seguimiento de una medida de densidad de incidencia. (Un enfoque alternativo es poner a los que se “cambian” de categoría de exposición en un grupo propio.)

### **Ausencia de “contagio”**

La discusión anterior también supone que la exposición y el resultado son independientes, i.e., la enfermedad de una persona no afecta el riesgo de otra. Este supuesto se ve violado, por supuesto, en las enfermedades contagiosas, como las enfermedades de transmisión sexual y para los patógenos transportados por artrópodos, como por ejemplo la malaria, en que los humanos sirven de reservorio. En este caso, la diseminación de la enfermedad aumenta la exposición de los individuos no afectados, de manera que su riesgo aumenta. Estos llamados “sucesos dependientes” pueden resultar en una distorsión, o por lo menos una variabilidad marcada a través del tiempo, en las medidas de asociación mencionadas anteriormente (ver, por ejemplo, Koopman JS *y cols.*, 1991). Los sucesos dependientes no se limitan de ninguna manera a las enfermedades transmisibles, dado que los comportamientos personales y comunitarios a menudo se ven influidos por lo que otras personas y otras comunidades están haciendo. Algunos ejemplos son el cese de fumar, cambios en la dieta, intentos de suicidio, comportamiento al conducir, reglamentación de seguridad en el tránsito, y la intensidad de detección y comunicación de enfermedades.

## **Más sobre riesgo y riesgo relativo**

El **exceso de riesgo** nos da el aumento proporcional en la incidencia (una medida análoga puede ser construida usando la densidad de incidencia o el odds). Es una leve modificación de la razón de incidencias acumuladas muy útil en diversas circunstancias incluyendo medidas de impacto relativo, que discutiremos en breve. La definición algebraica es:

$$\text{Exceso de riesgo} = \text{RIA} - 1 = \frac{IA_1}{IA_0} - 1 = \frac{IA_1 - IA_0}{IA_0}$$

Para enfermedades con un período de riesgo extendido, a medida que la duración del seguimiento aumenta, el riesgo y la incidencia acumulada se hacen mayores. Siendo acumulada y basada en la población que sigue estando en riesgo, la incidencia acumulada es una función en aumento cuyo límite es 1.0 – si permanecemos en riesgo para siempre, eventualmente todos nos convertiremos en casos. A medida que  $IA_1$  y  $IA_0$  ambas aumentan hacia su límite de 1.0, la RIA también se acerca a

1.0. Por lo tanto el valor de la RIA puede cambiar a medida que se alarga el tiempo de seguimiento. También es posible que la razón de densidad de incidencias cambie con la duración del seguimiento, pero eso es una función de la historia natural de la enfermedad más que de las propiedades matemáticas de la razón de incidencias acumuladas.

Cuando la incidencia acumulada es baja, porque la enfermedad es rara y/o el período de seguimiento es corto:

$$IA \approx DI \times T \quad (\text{donde } T = \text{tiempo de seguimiento})$$

$$OR \approx RDI \approx RIA$$

Porque si IA es  $\approx DI \times T$ , entonces  $IA_1 = DI_1 \times T$  y  $IA_0 = DI_0 \times T$ ,

$$RIA \approx \frac{DI_1 \times T}{DI_0 \times T} = \frac{DI_1}{DI_0} = RDI$$

A medida que el tiempo de seguimiento se hace más corto, la Incidencia acumulada se hace más pequeña, llegando eventualmente a 0. Pero a medida que la IA se hace más pequeña su valor se hace cada vez más parecido a  $DI \times T$ . Por esta razón, el límite de la RIA a medida que el tiempo de seguimiento se hace mínimo ( $T \rightarrow 0$ ) es la razón de densidad de incidencia. Es por esto que la razón de densidad de incidencia a veces se denomina la “razón de incidencias acumuladas instantánea”.

En una población dinámica en estado estable (tamaño y distribución por edad constantes, densidad de incidencia, prevalencia, y duración de la enfermedad constantes):

$$\text{Odds de prevalencia} = \text{Incidencia} \times \text{Duración} \quad (\text{ver capítulo anterior})$$

A partir de esto podemos ver que la razón de odds de prevalencia (ROP) estima la RIA si la duración no está relacionada con la exposición, porque:

$$ROP = \frac{\text{odds}_1}{\text{odds}_0} = \frac{DI_1 \times T}{DI_0 \times T} = \frac{DI_1}{DI_0} = RDI$$

donde  $T$  aquí es la duración en los casos expuestos y no expuestos.

**Estimando el riesgo relativo (a través del odds ratio) a partir de datos de un estudio caso control.**

1. Construye una tabla (2x2, doble entrada)

Enfermedad	Exposición		Total	
	Si	No		
Si	a	b	m <sub>1</sub>	(a + b)
No	c	d	m <sub>2</sub>	(c + d)
Total	n <sub>1</sub> (a + c)	n <sub>0</sub> (b + d)	n	

2. Odds de exposición en los casos:

$$\text{Odds} = \frac{\text{Proporción de casos que son expuestos}}{\text{Proporción de casos que no son expuestos}} = \frac{a / (a + b)}{b / (a + b)} = \frac{a}{b}$$

3. Odds de exposición en los controles

$$\text{Odds} = \frac{\text{Proporción de controles que son expuestos}}{\text{Proporción de controles que no son expuestos}} = \frac{c / (c + d)}{d / (c + d)} = \frac{c}{d}$$

4. Odds ratio de exposición (OR<sub>e</sub>)

$$\text{OR}_e = \frac{\text{Odds de exposición en los casos}}{\text{Odds de exposición en los controles}} = \frac{a / b}{c / d} = \frac{ad}{bc}$$

Si los datos hubieran sido obtenidos de un estudio de cohortes o un estudio de corte transversal, hubiéramos estimado en vez el odds ratio de riesgo (OR<sub>r</sub>), como el odds de enfermedad en las personas expuestas dividido por el odds de enfermedad en las personas no expuestas. Algebraicamente, los odds ratio de exposición y de enfermedad son idénticos.

Señalemos que el odds ratio puede ser calculado a partir de proporciones o porcentajes igual que a partir de los números reales, dado que para el cálculo del odds ratio el primer paso es la conversión de los números en proporciones y luego se convierten las proporciones en odds (ver arriba).

### **Medidas de diferencia**

Las medidas basadas en la diferencia entre dos proporciones o tasas son la otra manera principal de comparar tasas y proporciones. A menudo se utilizan como medidas de impacto, como discutiremos en la próxima sección. Las fórmulas y las denominaciones para las diferencias de incidencias acumuladas (o riesgos) y las tasas de incidencia son:

$$DIA = IA_1 - IA_0 \quad (\text{“Diferencia de incidencias acumuladas”},$$

también conocida como “Diferencia de riesgos” o “Riesgo atribuible” )

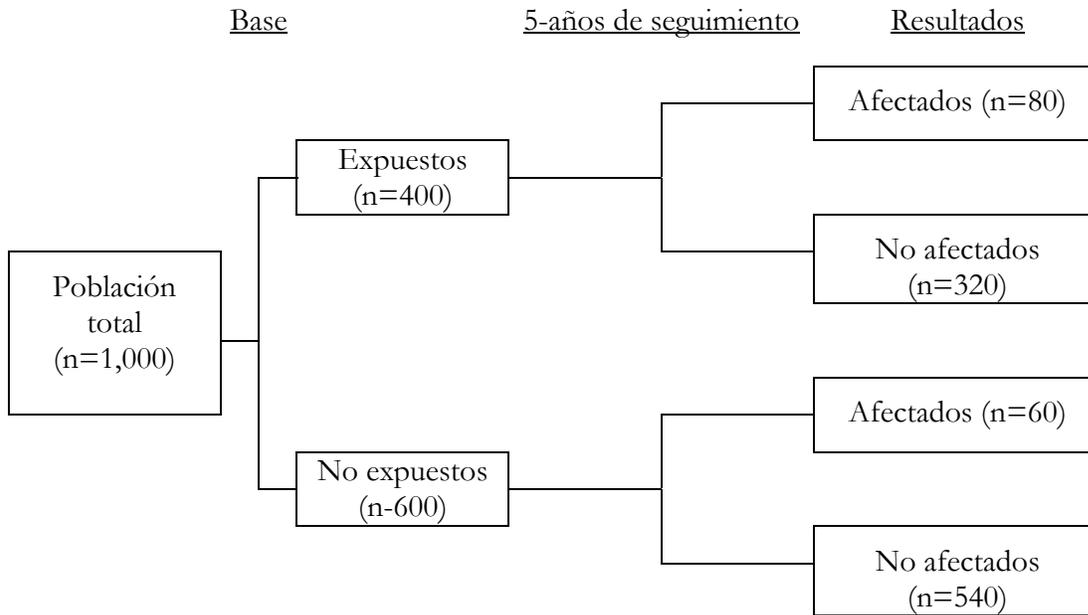
$$DDI = DI_1 - DI_0 \quad (\text{“Diferencia de densidades de incidencia”},$$

También conocida como “Diferencia de tasas”)

Estas medidas de diferencia, por supuesto, pueden ser directamente derivadas sólo de un estudio de cohortes o de seguimiento. Si nos falta información sobre el tamaño de la población en riesgo, como en un estudio caso control sin información adicional, no tenemos manera de estimar la incidencia acumulada o la densidad de incidencia, de manera que no podemos estimar diferencias de riesgos o de tasas. En un estudio de corte transversal, no podemos estimar incidencia de ninguna manera, aunque por analogía con la Diferencia de Incidencias Acumuladas y la Diferencia de Densidades de Incidencia podemos estimar la *diferencia de prevalencias*,  $P_1 - P_0$ .

### **Ejemplos de cálculos**

Seguimiento de una cohorte fija



(Supongamos que no hay pérdidas al seguimiento, incluyendo muertes.)

Los datos arriba expuestos se pueden resumir en una tabla  $2 \times 2$ :

### Incidencia de enfermedad de 5 años

Enfermedad	Exposición		Total
	Si	No	
Si	80	60	140
No	320	540	860
Total	400	600	1000

Nota sobre terminología: Los cuatro números (80, 60, 320, 540) en el interior de la tabla se llaman las “celdas”; los totales de las filas y las columnas (140, 860, 400, 600) son denominados los “marginales”. Las celdas a menudo se llaman “a”, “b”, “c”, y “d” en forma zigzagueante empezando con la celda de arriba a la izquierda.

IA (cruda) =  $140 / 1000 = .14$  (i.e., la incidencia acumulada global en los 5 años fue 14/100)

$IA_1 = 80 / 400 = .20$ ,  $IA_0 = 60 / 600 = .10$

$RIA = IA_1 / IA_0 = .20 / .10 = 2.0$  (la exposición se asoció con una duplicación del riesgo)

$DIA = IA_1 - IA_0 = .20 - .10 = .10$  (ver la interpretación a continuación)

Exceso de riesgo = RIA - 1 = 1.0 (i.e., la exposición se asoció con un aumento del 100% del riesgo)

$$OR_r = \frac{IA_1 / (1 - IA_1)}{IA_0 / (1 - IA_0)} = \frac{0.20 / 0.80}{0.10 / 0.90} = \frac{0.25}{0.11} = 2.25$$

Notemos que el OR es más extremo que la RIA.

Se podrían calcular medidas promedio de densidad de incidencia a partir de la tabla anterior haciendo la suposición de que los casos ocurrieron uniformemente durante todo el período, o de manera equivalente, que todos los casos ocurrieron en el punto medio del período de seguimiento, 2.5 años:

$$DI = \frac{\text{casos}}{\text{Años persona}} = \frac{140}{(860)(5) + (140)(2.5)} = \frac{140}{4300 + 350} = \frac{140}{4650} = 0.030 \text{ casos/años persona}$$

(El total de personas años en riesgo comprende 860 personas seguidas 5 años y 140 personas seguidas 2.5 años – una vez que ha ocurrido la enfermedad, se supone que el sujeto ya no está en riesgo. Si la situación no fuera así, los años persona se calcularían de una manera diferente.)

$$DI_1 = 80 / [(320)(5) + (80)(2.5)] = 80 / 1800 = 0.044 \text{ casos / años persona}$$

$$DI_0 = 60 / [(540)(5) + (60)(2.5)] = 60 / 2850 = 0.021 \text{ casos / años persona}$$

$$RDI = 0.044 / 0.021 = 2.095 = 2.1 \text{ (comparar con RIA de 2.0)}$$

$$DDI = 0.044 - 0.021 = .023 \text{ casos / años persona } \underline{\text{o}} \text{ 23 casos / 1000 años persona}$$

Señalemos que cada DI es muy aproximada a las IA correspondientes divididas por el número de años (5). Cuando la incidencia es baja, la IA es aproximadamente igual a la DI × (intervalo de tiempo).

### **Medidas de asociación – exposiciones no dicotómicas**

Las medidas de asociación de razón son adecuadas para las medidas dicotómicas (i.e., de dos categorías), como la presencia de enfermedad (si o no) o la presencia de la exposición (si o no). Si la exposición tiene múltiples categorías (por ejemplo, distintos tipos de solventes industriales o varios niveles de exposición), una medida de efecto de razón puede ser calculada para cada tipo o nivel con respecto al grupo no expuesto (sino hay un grupo no expuesto, entonces se puede seleccionar una exposición o nivel para ser la categoría de referencia). Consideren, por ejemplo, el estudio clásico

por Wynder y Graham (1950) sobre cáncer de pulmón y el hábito de fumar. En ese caso “ningún cigarrillo (menos de 1 por día)” fue seleccionado como la categoría de referencia, y el odds ratio fue calculado para cada nivel de fumador en relación al nivel de referencia.

### Historia de fumador de 605 pacientes con cáncer de pulmón de sexo masculino y 780 controles

Cantidad de cigarrillos fumados durante 20 y más años* (distribución porcentual)	Pacientes de cáncer de pulmón [N=605]	Controles [N=780]	OR
Ninguno (menos de 1 por día)	1.3	14.6	1.0**
Leve(1-9 por día)	2.3	11.5	2.2
Moderadamente intenso (10-15 por día)	10.1	19.0	6.0
Intenso (16-20 por día)	35.2	35.6	11.1
Excesivo (21-34 por día)	30.9	11.5	30.2
En cadena (35+ por día)	20.3	7.6	30.0

\* incluye fumadores de pipa y cigarros, mediante una fórmula de conversión.

\*\* categoría de referencia.

El odds ratio (OR) se obtiene formando la tabla 2x2 para cada nivel de exposición en relación con el nivel de referencia. Por ejemplo, para “Intenso (16-20 por día)” en comparación con “Ninguno” :

	Cáncer de pulmón	Control
Intenso	35.2	35.6
Ninguno	1.3	14.6

$$OR = \frac{35.2 \times 14.6}{35.6 \times 1.3} = 11.1$$

(Como se planteó anteriormente, el cálculo del OR puede ser hecho tan fácilmente a partir de los porcentajes como de los números reales de los casos y controles, dado que el primer paso es derivar las proporciones a partir de los números. El hecho de que estos porcentajes están ajustados por edad significa que los OR también están ajustados por edad.)

El odds ratio revela la existencia de una marcada relación dosis respuesta.

### Medidas de asociación – enfermedad no dicotómica

Cuando la enfermedad o variable de resultado no es dicotómica (p.ej., índice de masa corporal) pero la exposición sí lo es, la variable de resultado puede ser categorizada (p.ej., “por encima o por debajo del 30% del peso ideal”) para permitir el cálculo de medidas de asociación de razón.

Alternativamente, se puede calcular una medida de resumen (p.ej., masa corporal promedio) para cada categoría de la exposición, pero entonces no tenemos una medida que pueda ser interpretada como riesgo relativo.

Cuando tanto la enfermedad como la exposición tienen múltiples categorías ordenadas (p.ej., puntuación para la severidad de las lesiones (una variable ordinal), paridad (una variable discreta), o presión arterial (una medida continua), se puede imponer la categorización para obtener una medida de efecto de razón. Alternativamente, la relación entre resultado y exposición puede ser graficada, y usarse la pendiente como medida de la fuerza de la relación (p.ej., un aumento de 2 mmHg en la presión diastólica por cada 14 gramos de alcohol consumido es más fuerte que un aumento de 1 mmHg por cada 14 gramos). Se usan los coeficientes de regresión lineal para estimar la pendiente de la relación y dan un índice satisfactorio de fuerza de asociación para variables continuas, aunque no medidas que sean directamente comparables a medidas de riesgo relativo. Volveremos a los coeficientes de regresión más adelante en el curso.

Los coeficientes de correlación a menudo se usan como medidas de asociación entre variables ordinales o continuas, pero como se explicará más adelante, estas no son consideradas medidas epidemiológicas de fuerza de asociación.

### **Otras medidas de asociación**

“Cuando uso una palabra, significa exactamente lo que quiero que signifique, ni más ni menos” (Lewis Carroll, *Alicia en el País de las Maravillas* [*Alice in Wonderland*])

Como se mencionó anteriormente, un punto de confusión para el estudiante es la diferencia entre lo que los epidemiólogos llaman una medida asociación y lo que es medido por varios estadísticos que también son llamados medidas de asociación. Para aclarar esta situación poco feliz, discutiremos dos medidas ampliamente usadas tanto en epidemiología como en otras disciplinas, pero que los epidemiólogos consideran muy diferentes de las medidas de asociación que hemos discutido anteriormente.

### **Chi-cuadrado para asociación**

Un estadístico casi omnipresente en epidemiología es el chi cuadrado para asociación. El chi cuadrado y su valor p asociado responden a la pregunta sobre el grado al que una asociación observada en una muestra es probable que refleje una asociación en la población de la cual fue obtenida la muestra, más que el haber simplemente surgido por la variabilidad de muestreo. Los valores p estiman la probabilidad de que la variabilidad del muestreo al azar pueda resultar en que dos variables se asocien en la muestra aún cuando son totalmente independientes en la población. Aunque obviamente hay una conexión entre la pregunta planteada por el chi cuadrado y la pregunta planteada por el riesgo relativo, las dos preguntas no son de ninguna manera intercambiables. Por ejemplo, consideremos la tabla al principio de este capítulo

### ECC y ACO en mujeres de 35 años o más

	ACO	$\overline{\text{ACO}}$	Total
ECC	30	20	50
$\overline{\text{ECC}}$	30	70	100
Total	60	90	150

Tomando estos datos de un estudio caso control hipotético, seleccionamos el odds ratio (OR) como la medida de fuerza de asociación apropiada. Dado que la ECC es una enfermedad rara, el OR estimará la RIA además de la RDI. El OR de la tabla anterior es:

$$\text{OR} = \frac{30 \times 70}{20 \times 30} = 3.5$$

i.e., la asociación observada sugiere que el riesgo de ECC en las mujeres de 35 o más años que utilizan ACO es 3.5 veces la de las mujeres de similar edad que no usan ACO.

El estadístico chi cuadrado para esta tabla dará un valor p que se aproxima a la probabilidad de que una tabla con un OR de 3.5 o más fuerte surgirá de una selección al azar de 50 mujeres (que llamaremos “casos”) de una población de 60 usuarias de ACO y 90 no usuarias. Ese estadístico chi cuadrado es 12.4, que corresponde a una probabilidad muy pequeña – mucho menor a 0.0001, o 1 posibilidad en mil (los cálculos serán demostrados más adelante en el curso). Supongamos que el estudio que dio los resultados de la tabla anterior se hubiera basado en una población un quinto del tamaño de la de la tabla anterior. Manteniendo la misma proporción en cada una de las cuatro celdas, obtendríamos la siguiente tabla:

### ECC y ACO en mujeres de 35 años o más

	AOC	$\overline{\text{ACO}}$	Total
ECC	6	4	10
$\overline{\text{ECC}}$	6	14	20
Total	12	18	30

El odds ratio para esta tabla sigue siendo 3.5, pero el estadístico chi cuadrado es ahora sólo 2.42, que corresponde a un valor p de 0.12. Este valor mayor de p resulta del hecho de que es mucho más fácil obtener una asociación de 3.5 o más tomando al azar 10 “casos” de una habitación con 12 usuarias de ACO y 18 no usuarias que seleccionar al azar 50 “casos” de una habitación con 60 usuarias de ACO y 90 no usuarias.

Dado que el OR permanece igual pero el estadístico chi cuadrado y su valor p cambian dramáticamente, es claro que la medida epidemiológica de asociación y el chi cuadrado están midiendo distintas características de los datos. El estadístico chi cuadrado se usa para evaluar el grado de la evidencia numérica de que la asociación observada no fue un hallazgo fortuito. La medida epidemiológica de asociación se usa para cuantificar la fuerza de asociación como evidencia de una relación causal.

## Coefficientes de correlación

Los coeficientes de correlación son medidas de asociación lineales o monotónicas, pero de nuevo, no con el mismo significado que las medidas de riesgo relativo. El coeficiente de correlación lineal (de Pearson o correlación producto-momento, habitualmente abreviado como “r”) mide el grado en que la asociación entre dos variables es lineal. Una r de cero significa que las dos variables no están para nada relacionadas linealmente (pueden sin embargo estar asociadas de otra manera, p.ej., una relación en forma de U). Un r de +1 o -1 significa que cada par de observaciones de las dos variables corresponde a un punto en una línea recta dibujada en un papel de gráficas común. Sin embargo, conociendo que la relación es lineal o no, no nos dice nada sobre la pendiente de la recta, p.ej., cuanto aumento de la presión arterial resulta de un aumento de 5% en la masa corporal. Otros coeficientes de correlación (p.ej., Spearman) miden el grado de monotonía de la relación (i.e., las dos variables co-varían con respecto a si los pares de observaciones corresponden a una línea recta o una curva).

Los epidemiólogos piensan en las relaciones entre variables como indicaciones de procesos mecanicistas, de manera que para un epidemiólogo, la fuerza de asociación significa que tan grande es el cambio en el riesgo u algún otro resultado a partir de un cambio absoluto o relativo dado en la exposición. Si el supuesto es correcto, la fuerza no debería depender del rango de exposiciones medidas u otros aspectos de la distribución. En contraste, r se ve afectado por el rango y la distribución de las dos variables y por lo tanto no tiene una interpretación epidemiológica (Rothman, p.303). Los coeficientes de regresión tampoco son recomendados para el análisis epidemiológico por razones similares (ver Greenland, Schlesselman, y Criqui, 1986).

Coefficientes de correlación entre variables dicotómicas — Los coeficientes de correlación pueden ser particularmente problemáticos cuando se usan para cuantificar la relación entre dos factores dicotómicos (binarios), sobretodo cuando uno o ambos son raros. La razón es que los coeficientes de correlación entre variables binarias no pueden alcanzar los valores teóricos mínimo (-1) y máximo(+1) salvo en aquel caso especial en que ambos factores están presentes la mitad del tiempo y la otra mitad están ausentes (Peduzzi, Peter N., Katherine M. Detre, Yick-Kwong Chan. Límites superior e inferior para correlaciones en tablas  $2 \times 2$  - visitados de nuevo [Upper and lower bounds for correlations in  $2 \times 2$  tables—revisited.] *J Chron Dis* 1983;36:491-496). Si uno o ambos factores son raros, aún si las dos variables están fuertemente relacionadas, el coeficiente de correlación (p.ej., 0.15) puede en realidad ser grande en comparación con el valor máximo obtenido para las proporciones marginales dadas.

Por ejemplo, el coeficiente de correlación entre el hábito de fumar y el cáncer de pulmón no puede ser grande cuando la proporción de casos de cáncer de pulmón es pequeña pero la de fumadores es grande, como se demuestra en el siguiente ejemplo (Peduzzi PN, Detre KM, Chan YK. Upper and

lower bounds for correlations in  $2 \times 2$  tables—revisited. *J Chron Dis* 1983;36:491-496) basado en datos del Condado de Allegheny, Pennsylvania:

	Fumador	No fumador	Total
Cáncer de pulmón	20	2	22
No cáncer de pulmón	14,550	9,576	24,126
Total	14,570	9,578	24,148
Incidencia de cáncer de pulmón			0.001
Prevalencia del hábito de fumar			0.60
Odds ratio			6.6

	Coefficiente de Correlación (r)	R-cuadrado (R <sup>2</sup> )
Basado en los datos anteriores	0.019	0.00036
Si todos los casos fueran fumadores	0.024	0.00058
Si ninguno de los casos fuera fumador	-0.037	0.00157

En este caso el coeficiente de correlación (r) tiene un valor de apenas 0.019, con un R<sup>2</sup> correspondiente (“proporción de la varianza explicada) de 0.000356. Aún si los 22 casos de cáncer de pulmón fueran fumadores, el coeficiente de correlación puede aumentar solo a 0.024 (cuando R<sup>2</sup> = 0.0006), y si ningún caso de cáncer de pulmón fumara r cae sólo a 0.037. En contraste, el OR es 6.6, indicando una relación fuerte (el RR y la RDI son esencialmente similares, dado que el resultado es tan raro) Por lo tanto el coeficiente de correlación y la proporción de la varianza explicada no son inmediatamente aplicables a las relaciones entre variables dicotómicas, sobretodo cuando los totales de las columnas y las filas son muy diferentes.

## Medidas de Impacto

### Concepto

Las medidas de riesgo relativo comparan el riesgo (o tasa) en un grupo expuesto con aquel de un grupo no expuesto de manera de evaluar la fuerza de asociación entre la exposición y el resultado con el propósito de evaluar si la asociación es de naturaleza causal, como veremos en el capítulo sobre Inferencia Causal. Pero cuando hemos decidido (o supuesto) que la exposición produce el resultado, a menudo queremos evaluar la importancia a nivel individual y/o de Salud Pública de la relación, i.e.,

- ¿Cuánto de la enfermedad puede ser atribuida al factor causal?
- ¿Cuál es el beneficio potencial de intervenir para modificar el factor?

Las respuestas a estas preguntas conciernen la toma de decisión en políticas de salud pública y en principio las decisiones individuales, dado que indican la cantidad o la proporción de carga de enfermedad que puede ser prevenida eliminando el factor causal putativo (p.ej., control de polución) o llevando a cabo una intervención preventiva (p.ej., refuerzo de los alimentos). Ejemplos del tipo de preguntas que promueven el uso de medidas de impacto son:

1. Ahora que tengo 35 años de edad, mi riesgo de EEC por tomar anticonceptivos orales es el doble del que tenía cuando tenía 25 años de edad. ¿Pero *cuánto riesgo más tengo* por tomar los ACO?
2. En las parejas VIH discordantes que no utilizan preservativos y uno de la pareja tiene una enfermedad de transmisión sexual bacteriana, ¿cuánto del riesgo de la transmisión heterosexual de VIH es debida a la presencia de la enfermedad de transmisión sexual y por lo tanto podría ser eliminado a través de las medidas de control de las ETS?
3. ¿*Cuántos casos de asma* son debidos al dióxido de sulfuro ambiental?
4. ¿*Qué proporción de las muertes por vehículos de motor* pueden ser prevenidos por el uso obligatorio del cinturón de seguridad?
5. ¿Qué proporción de la transmisión perinatal de VIH puede ser prevenida por el uso de zidovudina prenatal, intraparto, y neonatal?

Para contestar estas preguntas empleamos las **fracciones atribuibles**, que son medidas de **impacto** o **riesgo atribuible**. El concepto de lo atribuible es de importancia primordial para la salud pública, dado que responden a la pregunta de “¿Y qué?”. Aunque algunos estudiantes encuentran que el tema de riesgo atribuible es una fuente de confusión, y por lo menos parte de su confusión es atribuible(i) tanto a la terminología como al concepto básico. Hay, sin embargo, varias sutilezas y fuentes legítimas de confusión relacionadas con el riesgo atribuible. Para introducir el concepto hacemos el supuesto simplificado de que la exposición estudiada tiene efectos adversos o beneficiosos, pero no ambos, que los grupos expuesto y no expuesto son idénticos salvo por esa exposición y algún otro factor causal (i.e., una persona que experimenta el resultado por otro factor causal no lo hará por la

exposición en estudio, y viceversa). También comenzamos por concentrarnos en los riesgos y las proporciones, más que en las tasas.

Una nota introductoria más: aunque parezca una tontería, quiero señalar que las preguntas sobre el riesgo atribuible surgen sólo en situaciones en que más de un factor puede causar el resultado en consideración. Cuando el daño tiene un único factor causal (típicamente cuando el daño es definido en términos del agente etiológico, como es el caso de las enfermedades infecciosas) todos los casos deben ser atribuibles a ese factor. Eliminando el factor se evita todo el riesgo. Si una causa necesaria (“C”) requiere un co-factor o un factor de susceptibilidad (“S”) para que el efecto se produzca, entonces todos los casos son atribuibles tanto a “C” como a “S”. Este último punto demuestra que las fracciones atribuibles no suman 1.0, aunque a menudo se expresan como porcentajes.

## **Perspectivas**

Hay una variedad de diferentes medidas de impacto, y por lo menos el doble de esa cantidad de nombres para ellas. (Por ejemplo, el término “riesgo atribuible” a veces es usado para indicar la diferencia de riesgo, a veces la proporción de riesgo atribuible de la población descrito más adelante, y a veces la clase de medidas de impacto. Ver Rothman y Greenland para los variados usos, con citas). Una razón para la cantidad de medidas es simplemente el tener una medida para cada una de las diferentes maneras de hacer una pregunta sobre impacto. Es decir, la pregunta puede ser hecha en términos absolutos (“Cuánto” riesgo) o en términos relativos (“Qué proporción” de riesgo). Se puede hacer la pregunta específicamente en referencia a las personas expuestas al factor o en referencia a toda la población. Además, el factor en consideración puede causar o prevenir el daño. Las diferentes combinaciones de estas alternativas requieren distintas medidas. La razón para tener más nombres que medidas (y para el uso del mismo nombre para distintas medidas) no es muy clara.

### **Perspectiva absoluta**

La **perspectiva absoluta** para el riesgo atribuible es expresada por las preguntas, “¿Cuánto del riesgo es atribuible al factor?” y “¿Cuántos casos pueden ser evitados si el factor estuviera ausente?” La respuesta se obtiene estimando la **diferencia de riesgo** o la diferencia en el número de casos entre las personas expuestas y no expuestas. La diferencia de riesgo, por ejemplo, nos da una estimación de la **cantidad de riesgo en las personas expuestas** que es “atribuible” al factor (suponiendo causalidad). Si nos interesa la cantidad de riesgo que es atribuible a la exposición **en la población total** (suponiendo causalidad), multiplicamos la diferencia de riesgo por la **prevalencia de exposición** en la población. Si nos interesa el **número real de casos** que son atribuibles, i.e., que podrían haber sido evitados por la eliminación completa de la exposición (antes de que hubieran ocurrido efectos irreversibles), podemos multiplicar la diferencia de riesgo por el tamaño de la población.

### **Perspectiva relativa**

La **perspectiva relativa** del riesgo atribuible es expresada por la pregunta, “¿Qué proporción del riesgo es atribuible al factor?” y “¿Qué proporción de los casos de la enfermedad podría ser evitada si el factor estuviera ausente?”. Aquí, necesitamos expresar la cantidad de riesgo atribuible al factor en relación con el riesgo total en las personas expuestas o en la población total. La medida para la

población expuesta también es llamada la “**proporción de riesgo atribuible**” (PRA) o la “**fracción de exceso**” (ver Rothman y Greenland). La medida para la población total a veces es llamada “la proporción de riesgo atribuible poblacional” (PRAP).

### **Proporción de riesgo atribuible**

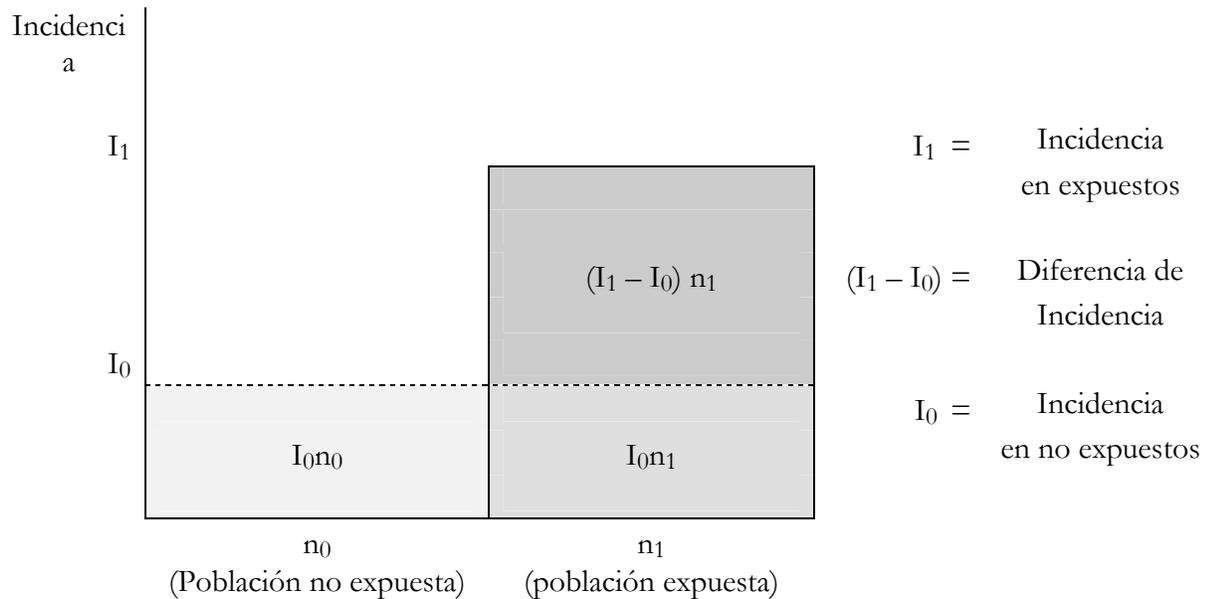
La proporción de riesgo atribuible se relaciona directamente con la fuerza de asociación entre la exposición y la enfermedad – si la exposición lleva el riesgo al doble, entonces la mitad del riesgo es atribuible a la exposición; si la exposición multiplica el riesgo por tres, entonces dos tercios del riesgo es atribuible a la exposición; si la exposición multiplica el riesgo por cuatro, entonces la proporción de riesgo atribuible es tres cuartos, etc.

### **Proporción de riesgo atribuible poblacional**

La PRAP refleja no sólo la fuerza de asociación sino también la prevalencia de la exposición en la población. Obviamente una exposición puede hacer más daño (tener más impacto) si es presentada por muchas personas en vez de unas pocas. La PRAP agrega este aspecto a la proporción de riesgo atribuible. Señalemos que muchos textos y artículos antiguos se refieren a la PARP simplemente como “riesgo atribuible”.

El siguiente diagrama muestra los fundamentos de las diversas medidas de riesgo atribuible. La idea básica es, simplemente que, si observamos la incidencia  $I_1$  en una población expuesta y una incidencia menor  $I_0$  en una población comparable no expuesta, y hacemos el supuesto de que la exposición está causando la mayor incidencia en la población expuesta, es lógico suponer que la diferencia,  $I_1 - I_0$ , es la cantidad de incidencia que se debe a la exposición. Por lo tanto, dependiendo de la manera en que preguntamos, esta “incidencia atribuible” se expresa como una diferencia absoluta o como una diferencia relativa, y en relación sólo a las personas expuestas o a la población total.

## Representación esquemática del riesgo atribuible en una población



$P_1 =$  Proporción expuesta

En el esquema anterior:

$n_0$  y  $n_1$  representan, respectivamente, el número de personas no expuestas y expuestas o las cantidades de tiempo persona no expuesta y expuesta;  $n = n_0 + n_1$

$P_0$  y  $P_1$  representan, respectivamente, las proporciones de personas no expuestas y expuestas o población tiempo (i.e.,  $P_1 = n_1/n$ )

$I_0$  es la proporción de incidencia de enfermedad (incidencia acumulada) en personas no expuestas, de manera que  $I_0 n_0$  es el número esperado de casos entre las personas no expuestas, i.e., el área no sombreada del triángulo.

$I_1 n_1$  es, de manera similar, el número esperado de casos entre las personas expuestas, i.e., el área combinada de los dos rectángulos sombreados.

$(I_1 - I_0)$  es la diferencia de incidencias o riesgo atribuible. Nos da la cantidad de incidencia “atribuible” a la exposición, i.e., la cantidad de incidencia en las personas expuestas en exceso, por encima de lo que se esperaría que hubieran tenido en ausencia de la exposición ( $I_0$ ).

$(I_1 - I_0) n_1$  es el número esperado de casos entre las personas expuestas más allá de lo que se esperaría de su incidencia de fondo ( $I_0$ ), i.e., casos atribuibles (el área del rectángulo cuadrículado). Los casos atribuibles son simplemente el riesgo atribuible multiplicado por el número de personas expuestas.

RR es el riesgo relativo (razón de riesgos, RIA),  $I_1/I_0$

La proporción de riesgo atribuible (PRA) en las personas expuestas es la proporción de casos expuestos que son “atribuibles” a la exposición. La proporción es:

$$PRA = \frac{\text{“Casos atribuibles”}}{\text{Todos los casos expuestos}} = \frac{(I_1 - I_0)n_1}{I_1 n_1} = \frac{I_1 - I_0}{I_1} = \frac{RR - 1}{RR}$$

(los RR se obtienen dividiendo el numerador y el denominador por  $I_0$ ).

De manera similar, la proporción de riesgo atribuible poblacional [PRAP], la proporción de todos los casos que es atribuible a la exposición, es:

$$PRAP = \frac{\text{“Casos atribuibles”}}{\text{Todos los casos}} = \frac{(I_1 - I_0)n_1}{I_1 n_1 + I_0 n_0} = \frac{I_1 n_1 - I_0 n_1}{I_1 n_1 + I_0 n_0} = \frac{P_1(RR-1)}{1 + P_1(RR-1)}$$

La fórmula de la derecha muestra la relación de la PRAP con la prevalencia de la exposición y el “riesgo excesivo” ( $RR-1$ ). El denominador no puede ser menor a 1, de manera que si el numerador es muy pequeño (p.ej., una prevalencia de exposición muy baja y/o una asociación débil), la PRAP también será muy pequeña. Al contrario, para una exposición muy prevalente (p.ej.,  $P_1=0.80$ ) y una asociación muy fuerte (p.ej.,  $RR=9$ ), el numerador [ $0.80 \times (9-1)$ ] será grande (6.4). El denominador se acercará a este valor, dado que el 1 tendrá poca influencia. Así la PRAP mostrará que una gran proporción (p.ej. cerca de 1.0) de los casos es atribuible. A medida que aumenta la prevalencia, la PRAP se acerca más a la PRA (cuando  $P_1=1$ , como ocurre en la población expuesta, la fórmula de la PRAP se reduce a eso para la PRA).

La influencia conjunta de la fuerza de asociación y la prevalencia de exposición sobre la PRAP puede ser más fácil de apreciar en la siguiente reformulación algebraica:

$$PRAP = \frac{1}{1 + 1/[P_1(RR-1)]}$$

## Definiciones y fórmulas

**Riesgo atribuible** [absoluto]: la cantidad de riesgo en el grupo expuesto que está relacionada con su exposición. El riesgo atribuible es estimado por la diferencia de incidencias acumuladas o la diferencia de densidades de incidencia:

$$RA = I_1 - I_0$$

**Riesgo atribuible poblacional** [absoluto]: la cantidad de riesgo en la población (i.e., en personas expuestas y no expuestas en conjunto) que se relaciona con la exposición. El riesgo atribuible poblacional es igual al riesgo atribuible multiplicado por la prevalencia de la exposición:

$$RAP = RA \times P_1 = (I_1 - I_0)P_1 = I - I_0$$

[Esta medida no se usa muy frecuentemente, pero es útil aquí para completar el patrón. La “I” sin subíndice se refiere al total, o incidencia cruda. La equivalencia de los términos del medio y de la derecha en la expresión anterior pueden apreciarse sustituyendo  $(I_1P_1 + I_0P_0)$  por I y  $(I_0P_0 + I_0P_1)$  por  $I_0$ .]

**Proporción (o porcentaje) de riesgo atribuible** [PRA]: proporción (porcentaje) del riesgo en el grupo expuesto que se relaciona con su exposición.

$$PRA = \frac{I_1 - I_0}{I_1} = \frac{RR - 1}{RR} = \frac{RA}{I_1}$$

**Proporción de riesgo atribuible poblacional** (o porcentaje) [PRAP]: La proporción (porcentaje) del riesgo en la población que está relacionada con la exposición.

$$PRAP = \frac{P_1 (RR - 1)}{1 + P_1 (RR - 1)} = \frac{I - I_0}{I} = \frac{PRA}{I}$$

## Estudios caso control

Las medidas absolutas (RA y PRA) requieren estimaciones de incidencia, de manera que no pueden ser estimadas de los resultados de un estudio caso control sin información adicional sobre la incidencia. Si la enfermedad es rara, la PRA y la PRAP pueden ser estimadas a partir de un estudio caso control usando el OR como estimación del RR. La PRA es entonces simplemente  $(OR - 1)/OR$ . Se puede entonces derivar una fórmula para el PRAP usando álgebra y el Teorema de Bayes :

$$PRAP = \frac{P_{E|D} (RR - 1)}{RR} = (P_{E|D}) \times PRA$$

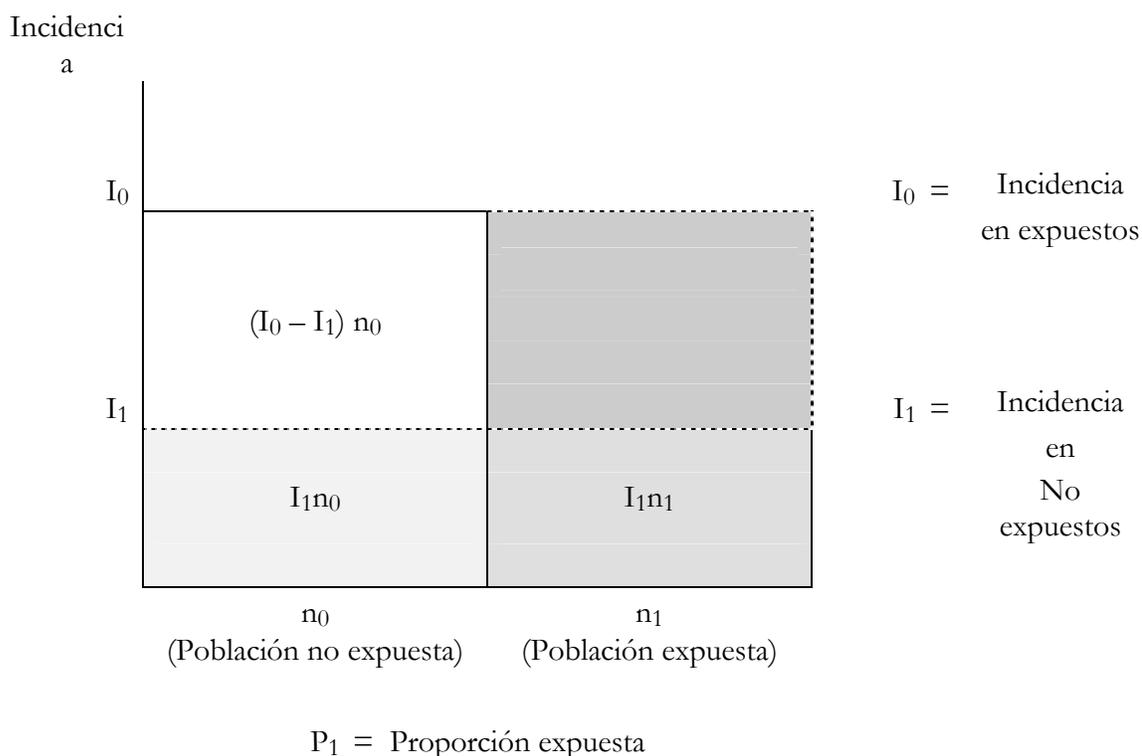
donde  $P_{E|D}$  es la proporción de casos que son expuestos. Dado que el lado derecho de la fórmula no requiere conocer la prevalencia de exposición en la población ni la verdadera incidencia de la enfermedad, esta fórmula puede ser usada en los estudios caso control que brindan una estimación del RR o de la RDI.

## Fracción prevenible

Cuando  $I_1 < I_0$  (p.ej., para uso de vacunas, equipo protector, o artefacto para control de polución), se necesita una fracción “prevenible”. Dado que una exposición de prevención (suponiendo causalidad) disminuye el riesgo del daño, no podemos pensar en términos de “casos atribuibles” dado que los casos no han ocurrido! En vez, definimos la fracción prevenible como la proporción de “casos potenciales” que fueron prevenidos, i.e., que **no** ocurrieron por la exposición protectora. Para las vacunas esta proporción se denomina **eficacia de la vacuna** o efectividad.

Igual que con el riesgo atribuible, hay dos variantes, una para los expuestos a la intervención de prevención y una para la población total (ambas basadas en la perspectiva “relativa”; la perspectiva absoluta no parece ser utilizada). El siguiente diagrama, similar al de las fracciones atribuibles, será usado.

**Representación esquemática de la fracción prevenible en una población**



Donde  $n_1$ ,  $n_0$ ,  $I_1$ ,  $I_0$  son los mismos que antes y  $(I_0 - I_1)n_1$  constituyen los “casos prevenibles”, i.e., el número de casos potenciales que hubieran ocurrido si la exposición no se asociara con una incidencia menor (recuerden que  $I_0$  es mayor que  $I_1$ ) o no hubiera estado presente.  $I_1n_1$  son los casos que ocurrieron a pesar de la intervención.

Por lo tanto, la fracción prevenible en los expuestos (FP<sub>1</sub>) cuantifica los casos prevenidos como proporción de todos los casos potenciales en las personas expuestas. La fracción prevenible en la población (FP) representa los casos prevenidos como una proporción de todos los casos potenciales en la población entera. En cada caso, los “casos prevenidos” son casos que hubieran ocurrido sin no fuera por la exposición de prevención; los “casos potenciales” son los casos prevenidos más los casos reales.

Del diagrama:

### Fracción prevenible en los expuestos

(FP<sub>1</sub> - para los expuestos a la intervención preventiva)

$$FP_1 = \frac{\text{“casos potenciales prevenidos”}}{\text{Todos los casos potenciales expuestos}} = \frac{(I_0 - I_1) n_1}{I_0 n_1} = \frac{(I_0 - I_1)}{I_0} = 1 - RR$$

(dado que  $I_1 < I_0$ ,  $RR < 1.0$ ).

### Fracción prevenible en la población (FP)

$$FP = \frac{\text{“casos potenciales prevenidos”}}{\text{Todos los casos potenciales}} = \frac{(I_0 - I_1) n_1}{I_0 n} = \frac{(I_0 - I_1) P_1}{I_0} = P_1 FP_1$$

(recuerde que  $n_1/n$  es la proporción expuesta,  $P_1$ ).

La fracción prevenible representa la proporción (o porcentaje) de la potencial carga de enfermedad que es prevenida por el factor protector. La siguiente fórmula muestra esto claramente:

$$FP = \frac{(I_0 - I_1) n_1}{I_0 n} = \frac{(I_0 n_1 - I_1 n_1) + (-I_0 n_0 + I_0 n_0)}{I_0 n} = \frac{(I_0 n_1 + I_0 n_0)}{I_0 n} - \frac{(I_1 n_1 + I_0 n_0)}{I_0 n} = \frac{I_0 - I_1}{I_0}$$

$I_0$  es el riesgo en las personas no expuestas a la intervención protectora. Si nadie recibe sus beneficios, entonces el riesgo en la población entera también sería  $I_0$ . El riesgo global real,  $I$ , representa el promedio de riesgos de los expuestos a la intervención protectora y de los no expuestos, ponderados por sus tamaños respectivos ( $I_1 n_1 + I_0 n_0$ ). Así  $I_0 - I$  es la diferencia entre el riesgo que *se podría haber* observado y el riesgo que *fue* observado, diferencia que se supone atribuible

a la efectividad de la intervención protectora y su utilización. La última fórmula muestra esta diferencia como una proporción del riesgo en ausencia de la intervención preventiva.

Para todas estas medidas, por supuesto, se acepta el supuesto, por lo menos a los efectos de la discusión, de que la relación es causal, y en algunos casos, que el quitar la causa (o introducir la intervención protectora) es completa e inmediatamente efectivo. En cualquier ejemplo específico, por supuesto, este último supuesto puede variar.

### **Enfoque único del riesgo atribuible y la fracción prevenible**

Aunque hay muchas sutilezas, la idea básica del riesgo atribuible y la fracción prevenible es sencilla. Esta simplicidad se ve eclipsado por una serie de fórmulas. Los siguientes conceptos ayudan a comprender la sencillez subyacente y puede ser la manera más fácil de derivar las fórmulas cuando se necesitan.

El objetivo básico es cuantificar el impacto de una exposición o intervención preventiva en términos de carga de enfermedad. Los impactos importantes ocurren en los casos de:

1. alta frecuencia de enfermedad
2. riesgo o factor de prevención potente
3. proporción importante de las personas expuestas al factor

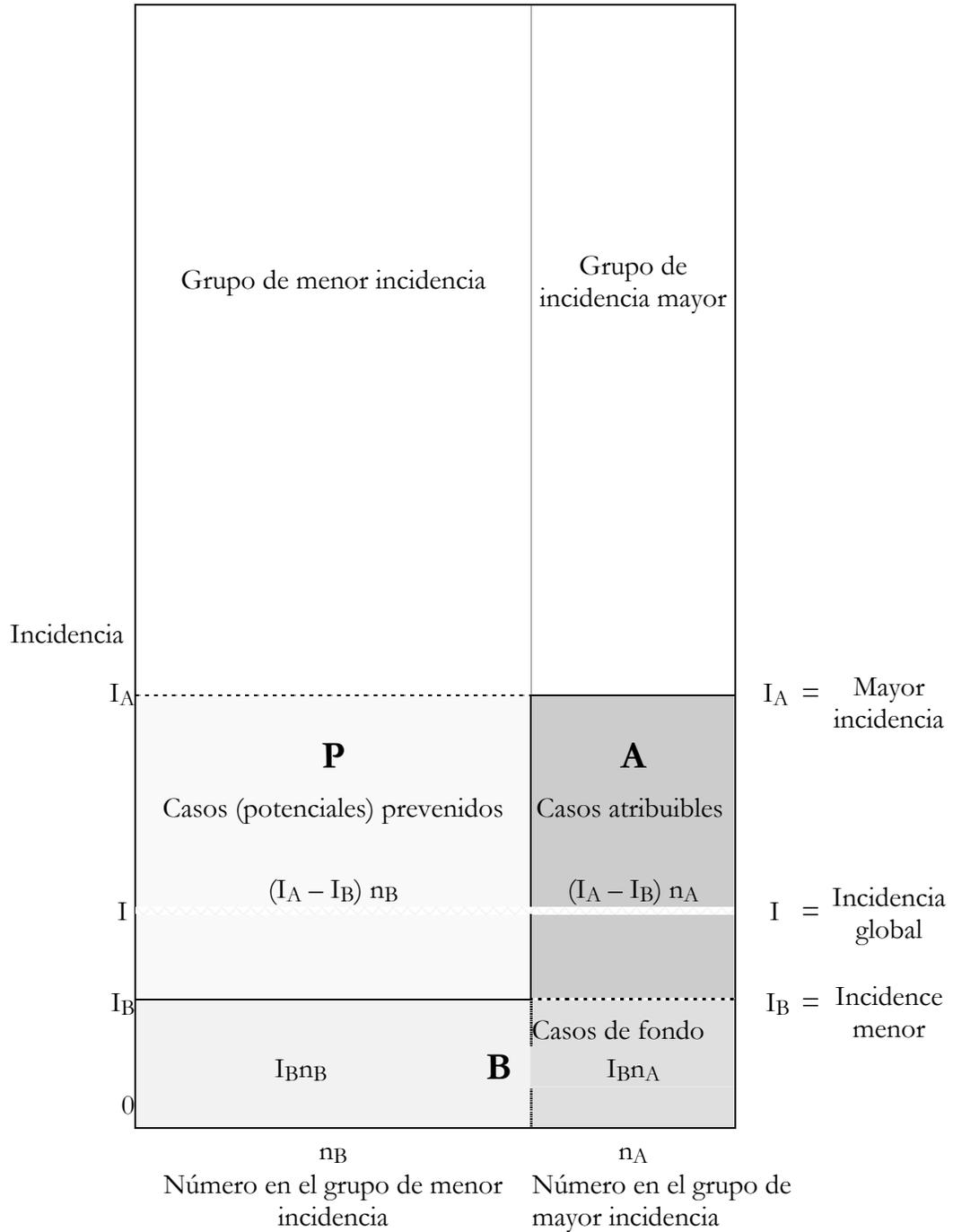
Un aspecto que complica las fórmulas es el hecho de que la incidencia en las personas expuestas al factor de riesgo es mayor que en los no expuestos, pero la incidencia en las personas expuestas al factor de prevención es menor que en los no expuestos. Podemos obviar estas diferencias pensando en términos de incidencia más alta e incidencia más baja.

El diagrama siguiente representa una población en riesgo en que las personas pueden ser clasificadas en dos grupos de exposición, uno con una menor incidencia (p.ej., físicamente activos) y el otro con una mayor incidencia (p.ej., sedentarios). El ancho de cada rectángulo alto indica el número de personas en el grupo correspondiente de exposición. La actividad física y el estilo de vida sedentario son un buen ejemplo porque sirven para la perspectiva de factor de riesgo (riesgo atribuible) y para la de factor preventivo. Usemos  $I_B$  e  $I_A$  para representar las incidencias baja y alta respectivamente, y  $N_B$  y  $N_A$  para representar el número de personas o la cantidad de tiempo poblacional en las categorías de baja y alta incidencia respectivamente.

En este diagrama, el rectángulo **A** [ $N_A (I_A - I_B)$ ] representa los casos **atribuibles**. Estos son los casos que no hubieran ocurrido si no fuera por el factor de riesgo (o la ausencia del factor de prevención). El rectángulo **P** [ $N_B (I_A - I_B)$ ] representa los casos **prevenidos**. Estos casos son sólo potenciales, dado que no han ocurrido. Son los casos que hubieran ocurrido si el factor de prevención no hubiera estado presente (o si el factor de riesgo se extendiera al grupo de baja incidencia). El rectángulo **B** [ $N_B I_B + N_A I_B$ ] representa los casos inevitables (de fondo). Ellos ocurren a pesar del factor de prevención o la ausencia del factor de riesgo. El número total de casos

es representado por la suma de los rectángulos para los dos grupos de exposición  $[ N_B I_B + N_A I_A ]$ . Si  $I$  es la incidencia global (cruda) el número total de casos también puede ser escrito como  $[ (N_B + N_A) I ]$ . El total de casos potenciales (i.e., los observados más los prevenidos) corresponde a  $[ (N_B + N_A) I_A ]$ , el resultado de someter a la población total a una incidencia mayor.

## Representación esquemática de los casos atribuibles y prevenibles



$n_1$  = Número con riesgo o factor de prevención.

$P_1$  = Proporción con el riesgo o factor de prevención.

Con el esquema y notación podemos expresar las fracciones atribuibles y prevenibles de una manera más paralela. El riesgo atribuible poblacional (PRAP) es simplemente  $\mathbf{A}/(\mathbf{A} + \mathbf{B})$  y la fracción prevenible (FP) es simplemente  $\mathbf{A}/(\mathbf{A} + \mathbf{B} + \mathbf{P})$ . Podemos entonces escribir las fórmulas que derivamos antes como:

$$\text{PRAP} = \frac{\text{“Casos atribuibles”}}{\text{Todos los casos}} = \frac{\mathbf{A}}{\text{Todos los casos}} = \frac{N_A (I_A - I_B)}{N_B I_B + N_A I_A} = \frac{P_A (I_A - I_B)}{I}$$

El último paso utiliza el hecho de que  $N_A/(N_B + N_A)$  es la prevalencia de la exposición y que la incidencia global  $I$  es igual al total de casos dividido por la población total. Dado que la proporción de riesgo atribuible (PRA) tiene que ver solo con el grupo expuesto,  $P_A = 1$  y  $PRA = (I_A - I_B)/I_A$ , que también es igual a  $(RR - 1)/RR$ .

De igual manera, la fracción prevenibles es:

$$\text{FP} = \frac{\text{“Casos prevenidos”}}{\text{Todos los casos potenciales}} = \frac{\mathbf{P}}{(N_B + N_A) I_A} = \frac{N_B (I_A - I_B)}{(N_B + N_A) I_A} = \frac{P_B (I_A - I_B)}{I_A}$$

Si dividimos el numerador y el denominador por  $I_A$ , obtenemos  $P_B (1 - RR)$ . La fracción prevenible en los expuestos a la intervención de prevención sólo tiene influencia en el grupo de menor incidencia, de manera que  $P_B = 1$  y  $FP_1 = (1 - RR)$ .

Con esta notación podemos ver la equivalencia esencial entre PRAP y la FP. Ambas involucran la diferencia de incidencia por el número de personas en el grupo “expuesto”. Ambas son expresadas como proporción de todos los casos, salvo que en el caso de la FP debemos incluir los casos potenciales en el denominador – si no, no sería una proporción.

### PRAP en un estudio caso control

Los estudios caso control, como veremos en el próximo capítulo, no dan una estimación de la incidencia salvo que se obtenga información adicional. Sin embargo, podemos usar razones de incidencia (RR) en lugar de incidencia, y entonces usamos el OR para estimar el RR. Si el grupo control se basa en la población, puede darnos una estimación directa de la prevalencia de la exposición. Si no, podemos modificar la fórmula y usar la prevalencia de exposición en los casos.

Esta prevalencia es simplemente los casos expuestos ( $N_A I_A$ ) dividido por todos los casos ( $N_B I_B + N_A I_A$ ), lo cual es muy similar a la fórmula de la PRAP.

El resultado,  $(N_A I_A) / (N_B I_B + N_A I_A)$ , es muy similar a la fórmula de la PRAP. La única diferencia está en los numeradores:  $N_A I_A$  versus  $N_A (I_A - I_B)$ . Podemos ir desde la expresión de la prevalencia de exposición a la PRAP multiplicando la primera por  $(I_A - I_B)/I_A$  o, en forma equivalente, por  $(RR - 1)/RR$ , que podemos estimar con  $(OR - 1)/OR$  si la enfermedad es rara. De manera que podemos estimar PRAP para una enfermedad rara a partir de un estudio caso control que no puede medir ni incidencias ni la prevalencia de exposición usando el OR para estimar RR en:

$$\text{PRAP} = \frac{N_A(I_A - I_B)}{N_B I_B + N_A I_A} = \frac{P_{A|D}(RR - 1)}{RR} = P_{A|D}(\text{PRA})$$

Sería bueno familiarizarse con este diagrama y estas fórmulas, ya que eso nos permitirá desarrollar una comprensión profunda de incidencia, prevalencia, riesgo relativo, impacto, y los promedios ponderados y también derivar cualquiera de las fórmulas básicas de riesgo atribuible o de fracción prevenible. Imaginemos como cambiará el diagrama a medida que aumenta o disminuye la prevalencia de la exposición. ¿Cómo se moverá la línea ondulante (la incidencia global, I) a medida que cambian las otras variables? ¿Cómo se relaciona con  $I_B$ ,  $I_A$ ,  $N_B$ ,  $N_A$ ? (Esta es definitivamente una relación clave para conocer). ¿Qué pasa cuando todos son expuestos? ¿Cuándo nadie está expuesto? ¿Cuál es la prevalencia de exposición en los casos? ¿Cómo cambiará a medida que cambia la incidencia y/o la prevalencia de la exposición global?

### **Interpretación de las fracciones atribuibles**

Aunque el concepto básico de riesgo atribuible tiene un sentido intuitivo, resulta que tiene muchas sutilezas y matices. Es sólo recientemente que se han comenzado a apreciar las sutilezas, de manera que mucho de lo que se ha escrito (aún por el suscrito) y algo de lo que se está escribiendo (con suerte, no por el suscrito) no es totalmente preciso. La confusión se agrava por la cantidad de términos que han sido introducidos, con usos que varían de un autor a otro. Además de las múltiples terminologías para el mismo concepto (un problema común en epidemiología), también hay situaciones en que un mismo término se aplica a distintos conceptos. Por lo tanto, si te encuentras confundido por algo que estás leyendo en esta área, siempre considera la posibilidad de que lo que estás leyendo puede ser confuso también.

Las complicaciones surgen cuando tratamos de interpretar las fracciones atribuibles (p.ej., PRA, PRAP) en términos de etiología (causales), que es, por supuesto, en lo que estábamos interesados desde el inicio. Consideremos las dos siguientes preguntas, que figuran en forma destacada en los litigios por responsabilidad de producto en que la justicia mantiene que la enfermedad del demandante es “más probablemente que no” una consecuencia de la exposición al producto (p.ej., asbestos, drogas de indicación médica, implantes mamarios de siliconas, tabaco).

- Entre los no fumadores expuestos a X, *¿qué proporción de Y fueron causados por X?*
- *¿Cuál es la probabilidad de que el caso de Y de una persona Z es el resultado de X?*

Lo que diferencia estas dos interrogantes de las preguntas ilustrativas planteadas al principio de la sección es el uso de la terminología causal (“causado por”, “resultado de”) en vez de la forma más general (y más vaga) “atribuido a”. La incidencia y las medidas derivadas de la incidencia muestran sólo los efectos netos o en exceso, no los procesos causales que los producen. Por ejemplo, aunque un estilo de vida sedentario aumenta el riesgo de la enfermedad cardíaca coronaria, el ejercicio físico puede aumentar notoriamente el riesgo de un evento cardíaco. Cuando comparamos la tasa de eventos cardíacos en el grupo sedentario con la tasa en un grupo de personas que realizan ejercicios en forma regular, la diferencia en tasas de incidencia mide el aumento en la tasas de eventos cardíacos *asociado con* el estilo de vida sedentario. Pero si algo de la incidencia de los eventos cardíacos en los que se ejercitan, en realidad *resulta de la práctica del ejercicio*, entonces la diferencia en la incidencia entre los dos grupos mide en realidad el daño *neto* de un estilo de vida sedentario, más que el efecto total. Comparando las tasas de incidencia estamos permitiendo que algunos de los eventos cardíacos en los que realizan ejercicios contrarresten algunos de los eventos en el grupo sedentario, siendo el tamaño relativo del daño y beneficio dependiente del tipo de personas (p.ej., características genéticas o distribución de alguna otra exposición) que se ejercita o no. En general, los datos epidemiológicos no revelarán qué contribuye a la diferencia de incidencias neta.

De igual manera, si la acción de un factor causal puede facilitar la oportunidad para que otro factor cause la enfermedad (porque la enfermedad ya ha ocurrido), entonces no hay manera de saber, a partir de los datos epidemiológicos, cual de los factores causó la enfermedad en una persona o la población expuesta a ambos factores causales. Por esta razón, es problemático interpretar las medidas de riesgo atribuible como **fracciones etiológicas**, aunque muchos autores han usado la terminología en forma intercambiable. Según Greenland (1999: 1167), la “falacia clave en la mayor parte de la literatura y los testimonios con respecto a la probabilidad de causalidad es el uso de las siguientes ecuaciones generalmente incorrectas: Fracción etiológica = Fracción de Tasa y Probabilidad de Causalidad = Fracción de Tasa ...” donde la fracción etiológica (FE) es “la fracción de estos individuos para quienes la exposición fue una causa contribuyente de la enfermedad” (Greenland, 1999: 1166) y la fracción de tasa (FT) es la diferencia de las tasas de incidencia dividida por la tasa de incidencia en los expuestos (análogo a la PRA, salvo por el hecho de que se deriva de las tasas de incidencia y no de las proporciones de incidencia) (p1167). En términos algebraicos  $FE = (A_1 + A_2) / A_T$ , donde  $A_1$  serían las personas expuestas que hubieran desarrollado la enfermedad en algún momento pero cuyo desencadenamiento fue precipitado debido a la exposición,  $A_2$  son las personas expuestas cuya enfermedad nunca hubiera ocurrido sin la exposición y  $A_T$  es  $A_1 + A_2$  más las personas expuestas que desarrollaron la enfermedad totalmente independiente de la exposición. La FE estima la **probabilidad de causalidad**, dado que  $(A_1 + A_2) / A_T$  es la probabilidad de que una persona seleccionada al azar de  $A_T$  tuviera su enfermedad acelerada por ( $A_1$ ) o causada totalmente por ( $A_2$ ) la exposición. La proporción  $A_2 / A_T$  es la fracción de exceso, dado que da la proporción del total de casos que no hubieran ocurrido sin la exposición (Greenland, 1999), no importando el tiempo hasta la ocurrencia. (Greenland observa que el no distinguir la fracción de exceso de la fracción etiológica es “un problema grave en la mayor parte de la literatura” y considera el término “riesgo atribuible” como particularmente engañoso aunque “domina la literatura norteamericana”, tanto en bioestadística como en epidemiología [p.1168].)

---

Respuesta para la pregunta del comienzo del capítulo sobre la asociación entre enfermedad cardiaca coronaria (ECC) y los anticonceptivos orales (ACO):

La proporción de casos de ECC en la muestra de 40 debe estar entre aproximadamente  $30/60 = 0.5$  (la proporción de casos entre las usuarias de ACO) y  $20/90 = 0.2222$  (la proporción entre las no usuarias) Si la muestra consiste de 22 usuarias y 18 no usuarias, entonces la mejor estimación de la proporción muestral de casos de ECC es:

$$\begin{array}{l} \text{Proporción} \\ \text{con} \\ \text{ECC} \end{array} = 0.5 \left( \frac{22}{40} \right) + 0.2222 \left( \frac{18}{40} \right) = 0.5(0.55) + 0.2222(0.45) = 0.375$$

Por lo tanto la mejor estimación de la proporción global con ECC es aproximadamente 0.375 o 15 mujeres en la muestra de 40.

## Resumen

Hay tres categorías de medidas: frecuencia/alcance, asociación, impacto

### **(1) Medidas de frecuencia o alcance (en especial prevalencia e incidencia)**

En epidemiología, la incidencia es la ocurrencia de cualquier evento nuevo relacionado con la salud (p.ej., enfermedad, muerte, recuperación). La incidencia se cuantifica como:

**PROPORCIÓN:** la proporción de la población que sufre el evento; también se denomina como “RIESGO” dado que estima el riesgo promedio por persona para el período. El ODDS (de riesgo) es simplemente una transformación del riesgo [riesgo/(1-riesgo)].

**TASA:** el número de eventos relacionados con la salud por persona por unidad de tiempo; corresponde al riesgo promedio por persona por unidad de tiempo.

MEDIDA	ESTIMADOR EPIDEMIOLÓGICO	UNIDADES	LÍMITES
Riesgo	Incidencia acumulada (IA)	Sin dimensiones	0 a 1
Tasa	Densidad de incidencia (DI)	1/tiempo	0 to “infinito”
Odds <sub>r</sub>	IA / (1-IA)	Sin dimensiones	0 to “infinito”

IA (una proporción) se usa para estimar el riesgo de un individuo de desarrollar una enfermedad. DI (una tasa) se usa para estimar la fuerza de intensidad de las ocurrencias. El riesgo y la tasa están relacionadas, dado que cuanto mayor es la intensidad de las ocurrencias en una población, mayor es el riesgo de un evento para cualquier integrante de la población. Cuando la IA es pequeña (i.e., por baja intensidad de una enfermedad o un intervalo de tiempo corto), la DI es aproximadamente igual a la IA dividida por el número de años de seguimiento. Cuando la IA no es pequeña, la relación es matemáticamente más compleja.

### **Aplicación**

La elección de una medida de incidencia (IA o DI) depende de:

#### **a. OBJETIVOS DEL ESTUDIO**

La IA da una estimación directa del riesgo de un individuo, por lo que puede ser útil para la toma de decisiones clínicas o personales;

La DI es a menudo preferida para la evaluación del impacto poblacional de un evento de salud o para la prueba de hipótesis etiológicas.

## ***b. CONSIDERACIONES PRÁCTICAS***

*La LA puede ser preferida:*

- si el evento de salud tiene un período de riesgo
- si es difícil de asegurar el momento de cambio en el estado de salud
- para facilidad de comprensión.

*La DI puede ser preferida:*

- si el evento de salud tiene un período de riesgo extendido
- si la duración de los seguimientos es variable
- si hay grandes pérdidas al seguimiento
- si el evento de salud puede recurrir (p.ej., fracturas de huesos).

Una razón de dos estimaciones de riesgo es una “razón de riesgos” (RR). Una razón de dos estimaciones de tasas es una “razón de tasas” (RR). Una razón de dos odds es una “razón de odds” (OR). Todas estas medidas son denominadas a veces “riesgo relativo” (RR), aunque en forma estricta sólo el primero tiene que ver con riesgos.

### **(2) Medidas de asociación**

p.ej. Razón de proporciones (RIA), razón de tasas (RDI), razón de odds (OR) .

$$RIA = \frac{a / (a + c)}{b / (b + d)} = \frac{IA \text{ en expuestos}}{IA \text{ en no expuestos}} = \begin{matrix} \text{“razón de riesgos”,} \\ \text{“riesgo relativo”} \end{matrix}$$

donde a= casos expuestos, b= casos no expuestos, c= sanos no expuestos, d= etc.

En términos de probabilidad,  $RIA = Pr(D|E) / Pr(D|\bar{E})$

¿Cómo interpretamos la RIA?

- a. Si  $RIA = 1$ , entonces no hay asociación entre la exposición y la enfermedad.
- b. Si  $RIA > 1$  entonces la exposición parece estar asociada con un aumento en el riesgo de la enfermedad i.e., la exposición puede ser dañina.
- c. Si  $RIA < 1$  entonces la exposición parece estar asociada con una disminución del riesgo de la enfermedad, i.e., la exposición puede ser protectora.

(RIA menores a 1.0 son incómodas de entender desde el punto de vista conceptual, así que en muchos casos es más útil invertir la categoría de enfermedad o exposición para obtener el recíproco de la RIA. Una RIA de 0.4 se convierte entonces en una RIA de 2.5)

La RIA puede ser estimada directamente si el estado de exposición se conoce antes de la ocurrencia de la enfermedad, como en los estudios de seguimiento prospectivos o en los estudios de cohortes retrospectivos.

Cuando una enfermedad es rara, el  $OR_r$  se aproxima a la RIA – una cosa útil de conocer porque se pueden usar los modelos de regresión logística para estimar los odds ratio:

$$OR_r = \frac{\text{Odds de la enfermedad en los expuestos}}{\text{Odds de la enfermedad en los no expuestos}} = RIA$$

El OR (sea el “OR de riesgo” o el “OR de exposición”) es fácil de calcular como la razón de productos cruzados:  $(a \times d) / (b \times c)$ .

Los OR de riesgo y de exposición se calculan en forma idéntica a partir de la tabla  $2 \times 2$ , pero eso no significa que sean iguales “epidemiológicamente”. Recuerde que los números de la tabla  $2 \times 2$  son sólo una abstracción de la verdadera experiencia del estudio y deben ser usados teniendo en cuenta el diseño (i.e., un diseño caso control no es equivalente a un diseño longitudinal). En un estudio de cohortes, típicamente calculamos una RIA o una RDI promedio. En un estudio de seguimiento sin cohorte fija típicamente calculamos una RDI. En un estudio caso control típicamente calculamos un OR. En un estudio de corte transversal, típicamente calculamos una razón de prevalencia o un OR de prevalencia. Si se conoce una de las incidencias acumuladas, la estimación del OR (p.ej., de un modelo de regresión logística – ver capítulo sobre Análisis e interpretación de datos) puede ser convertido en una estimación de razón de riesgo con la siguiente fórmula (Zhang and Yu, 1998; la nomenclatura se cambió para adecuarla a la utilizada en este capítulo):

$$RR = \frac{OR}{(1 - IA_0) + (IA_0 \times OR)}$$

Se puede estimar un odds ratio de prevalencia de la misma manera, si se conoce la prevalencia en los no expuestos.

### 3) **Medidas de impacto**

“Cuánto” de una enfermedad puede ser atribuida a una exposición puede ser considerado como:

- la cantidad de riesgo o incidencia en los expuestos (diferencia de incidencias acumuladas) o en la población total (habitualmente presentado como un número de casos)
- una proporción del riesgo o la incidencia en los expuestos (PRA) o en la población total (PRAP).

Los contribuyentes a las medidas de impacto son:

1. La fuerza de asociación – afecta todas las medidas de impacto.
2. Nivel de la incidencia de fondo – afecta sólo la cantidad de incidencia (diferencia de incidencias acumuladas, diferencias de densidades de incidencia)
3. Prevalencia población al de la exposición – afecta sólo el impacto en la población (p.ej., PRAP).

## Apéndice — Relacionando los factores de riesgo a los resultados en salud

Advertencia: este capítulo ha sido convertido a MS Word recientemente, de manera que pueden haberse introducido algunos errores algebraicos.

### ***Estimando la incidencia específica por exposición y el riesgo atribuible de un estudio caso control cuando se conoce la incidencia cruda***

Este procedimiento utiliza el hecho de que la incidencia cruda puede ser expresada como un promedio ponderado de las incidencias específicas por exposición:

$$I = P_1 I_1 + P_0 I_0$$

donde:

$I$  = incidencia cruda

$I_1$  = incidencia en la población expuesta

$I_0$  = incidencia en la población no expuesta

$P_1$  = proporción de la población que está expuesta

$P_0$  = proporción de la población que no está expuesta

Dado que el RR (riesgo relativo, RIA, RDI) =  $I_1/I_0$ , es posible sustituir  $RR \times I_0$  por  $I_1$  en la expresión anterior:

$$I = P_1 \times RR \times I_0 + P_0 I_0$$

De igual manera, dado que  $P_1 + P_0 = 1$ , podemos sustituir  $1 - P_1$  por  $P_0$ :

$$I = P_1 \times RR \times I_0 + (1 - P_1) \times I_0$$

Resolviendo para  $I_0$  nos da:

$$I_0 = \frac{I}{P_1 \times RR + (1 - P_1)} = \frac{I}{1 + P_1 (RR - 1)}$$

Dado que para una enfermedad rara podemos estimar el RR usando el OR, la fórmula final es:

$$I_0 = \frac{I}{1 + P_1 (OR - 1)}$$

Esta fórmula puede ser usada para estudios caso control en que:

1. El grupo control ha sido seleccionado de tal manera que la proporción expuesta estima  $P_1$  en la población;
2. Hay información disponible para estimar la incidencia cruda de la enfermedad;
3. La enfermedad es suficientemente rara (p.ej., incidencia menor de 10% al final del seguimiento) de manera que el OR estima el RR bastante bien.

Una vez que hemos estimado  $I_0$ , estimamos  $I_1$  multiplicado por el OR. A partir de  $I_1$  e  $I_0$  podemos estimar el riesgo atribuible.

### ***Demostración de que el OR estima la Razón de Incidencia Acumulada cuando las Incidencias Acumuladas son pequeñas***

Para esta demostración y la que sigue, una nomenclatura diferente simplificará la presentación. Usaremos E para enfermedad y X para exposición, de manera que podemos usar mayúsculas para su presencia y minúsculas para su ausencia. Así el subíndice X se referirá a la presencia de la exposición, y el subíndice x se referirá a la ausencia de exposición. De igual manera, el subíndice E se refiere a los casos, el subíndice e a los sanos. P es la probabilidad, que también puede ser interpretada como la prevalencia (cuando se refiere a la exposición) o al riesgo (cuando se refiere a la enfermedad). La barra vertical significa “dado que” o “condicionado a”. Así:

$P_X$  = Probabilidad de estar expuesto (i.e., prevalencia de la exposición)

$P_x$  = Probabilidad de no estar expuesto ( $1 - P_X$ )

$P_E$  = Probabilidad de enfermar (riesgo)

$P_e$  = Probabilidad de no tener la enfermedad ( $1 - P_E$ )

$P_{X|E}$  = Probabilidad de estar expuesto condicionado a tener la enfermedad (i.e., prevalencia de la exposición en los casos)

$P_{x|E}$  = Probabilidad de no estar expuesto teniendo la enfermedad (i.e.,  $1 - P_{X|E}$ )

$P_{X|e}$  = Probabilidad de estar expuesto no teniendo la enfermedad (i.e., prevalencia de la exposición en los sanos)

$P_{x|e}$  = Probabilidad de no estar expuesto no teniendo la enfermedad (i.e.,  $1 - P_{X|e}$ )

$P_{E|X}$  = Probabilidad de enfermedad (riesgo) condicionado a estar expuesto (i.e., riesgo de enfermedad en los expuestos)

$P_{e|X}$  = Probabilidad de no tener la enfermedad estando expuesto (i.e.,  $1 - P_{E|X}$ )

$P_{E|x}$  = Probabilidad de tener la enfermedad no estando expuesto (i.e., riesgo de enfermedad en los no expuestos)

$P_{e|x}$  = Probabilidad de no estar enfermo no estando expuesto (i.e.,  $1 - P_{E|x}$ )

Por definición,  $odds_r = riesgo / (1 - riesgo)$ . El odds ratio de riesgo ( $OR_r$ ) es la razón del odds para las personas expuestas al odds para las personas no expuestas. En nomenclatura de probabilidad:

$$OR_r = \frac{P_{E|X} / (1 - P_{E|X})}{P_{E|x} / (1 - P_{E|x})} = \frac{P_{E|X}}{P_{E|x}} \times \frac{(1 - P_{E|x})}{(1 - P_{E|X})} = RR \times \frac{(1 - P_{E|x})}{(1 - P_{E|X})}$$

Dado que  $P_{E|X} / P_{E|x} = RR$ . Cuando la enfermedad es rara,  $P_{E|X}$  y  $P_{E|x}$  son ambos pequeños, de manera que  $OR_r \approx RR$ . Usando la IA para estimar el riesgo y la RIA para estimar RR, obtenemos  $OR_r \approx RIA$  cuando la IA es pequeña tanto en el grupo expuesto como no expuesto. Para ilustrar esto, crea algunas tablas  $2 \times 2$  que reflejen los riesgos de varias enfermedades y razones de incidencias acumuladas y calcula el OR. Podrás verificar que el OR siempre se aleja de 1.0 más que la razón de incidencias acumuladas pero cuando la incidencia es menor de aproximadamente 10%, el OR se desvía poco de la razón de incidencias acumuladas. El OR también puede ser expresado como  $OR_r = RIA + (OR_r - )P_{E|X}$  (cuando  $OR > 1$ ; ver Hogue, Gaylor, y Schultz, 1983), lo cual demuestra que la diferencia absoluta entre OR y RIA se relaciona con el tamaño del OR y el riesgo de enfermedad en las personas expuestas.

### ***Estimando la PRAP de un estudio caso control sin información adicional***

En un estudio caso control en que el OR estima la RIA, podemos usar el OR y la proporción de exposición en los casos para estimar la PRAP.

Comencemos con la fórmula presentada anteriormente:

$$PRAP = \frac{P_1 (RR - 1)}{1 + P_1 (RR - 1)}$$

Traduciendo con la nueva nomenclatura ( $P_X$  for  $P_1$ ):

$$PRAP = \frac{P_X (RR - 1)}{1 + P_X (RR - 1)}$$

Quitando los paréntesis en el denominador y sustituyendo ( $P_{E|X}/P_{E|x}$ ) por RR en el denominador (únicamente):

$$PRAP = \frac{P_X (RR - 1)}{1 + P_X (RR - 1)} = \frac{P_X (RR - 1)}{1 + P_X RR - P_X} = \frac{P_X (RR - 1)}{1 + P_X (P_{E|X} / P_{E|x}) - P_X}$$

Multiplicando el numerador y el denominador por  $P_{E|x}$ :

$$PRAP = \frac{P_{E|x} P_X (RR - 1)}{P_{E|x} (1 + P_X (P_{E|X} / P_{D|e}) - P_E)} = \frac{P_{E|x} P_X (RR - 1)}{P_{E|x} + P_X P_{E|X} - P_{E|x} P_X}$$

Sustituyendo  $(1-P_X)$  por la segunda  $P_X$  en el denominador (únicamente) y re-organizando los términos:

$$PRAP = \frac{P_{E|x} P_X (RR - 1)}{P_{E|x} + P_X P_{E|X} - P_{E|x} (1-P_X)} = \frac{P_{E|x} P_X (RR - 1)}{P_{E|x} + P_X P_{E|X} - P_{E|x} + P_X P_{E|x}}$$

Quitando  $+ P_{E|x}$  y  $- P_{E|x}$  del denominador, tenemos entonces  $P_X P_{E|X} + P_X P_{E|x}$ .

Dado que el promedio ponderado de riesgo en los expuestos (ponderado por la probabilidad de exposición) más el riesgo en los no expuestos (ponderado por la probabilidad de no exposición), el denominador se simplifica a  $P_E$  (esto se presentó una página atrás como:  $I = P_1 I_1 + P_0 I_0$ ). Por lo tanto tenemos:

$$PRAP = \frac{P_{E|x} P_X (RR - 1)}{P_X P_{E|X} + P_X P_{E|x}} = \frac{P_{E|x} P_X (RR - 1)}{P_E} = \frac{P_X P_{E|x} (RR - 1)}{P_E}$$

Usando el Teorema de Bayes podemos mostrar que:

$$\frac{P_X}{P_E} = \frac{P_{X|E}}{P_{E|X}}$$

de manera que la fórmula anterior puede ser escrita como:

$$PRAP = \frac{P_{X|E} P_{E|x} (RR-1)}{P_{E|X}} = \frac{P_{X|E} (RR-1)}{P_{E|X} / P_{E|x}} = \frac{P_{X|E} (RR-1)}{RR}$$

Que también puede ser escrito como  $P_{X|E} PRA$ .

En un estudio caso control, sabemos que  $P_{X|E}$ , la probabilidad (proporción) de exposición entre los casos y usar el OR para estimar el RR (suponiendo que la enfermedad es rara).

## Bibliografía

Capítulos de textos (ver la lista anterior en Midiendo la Enfermedad).

Davies, Huw Talfryn Oakley; Iain Kinloch Crombie, Manouche Tavakoli. When can odds ratios mislead? *BMJ* 1998;316:989-991.

Deubner, David C., Herman A. Tyroler, John C. Cassel, Curtis G. Hames, and Caroline Becker. Attributable risk, population attribution risk, and population attributable fraction of death associated with hypertension in a biracial population. *Circulation* 1975;52:901-908

Freeman, Jonathan; George B. Hutchison. Duration of disease, duration indicators, and estimation of the risk ratio. *Am J Epidemiol* 1986; 124:134-49. (Advanced)

Gladden, Beth C. On graphing rate ratios. *Am J Epidemiol* 1983; 118:905-908.

Greenland, Sander. Relation of probability of causation to relative risk and doubling dose: a methodologic error that has become a social problem. *Am J Public Health* 1999 (August)

Greenland, Sander. Interpretation and choice of effect measures in epidemiologic analyses. *Am J Epidemiol* 1987; 125:761-768 and correspondence in *Am J Epidemiol* 1988; 128:1181-1184.

Greenland, Sander; James M. Robins. Conceptual problems in the definition and interpretation of attributable fractions. *Am J Epidemiol* 1988; 128:1185-1197. (Intermediate)

Greenland, Sander; Schlesselman JJ, Criqui MH. The fallacy of employing standardized regression coefficients and correlations as measures of effect. *Am J Epidemiol* 1986; 123:203-208. (Advanced)

Hebert, James R.; Donald R. Miller. Plotting and discussion of rate ratios and relative risk estimates. Letter. *J Clin Epidemiol* 1989; 42(3); 289-290.

Hogue, Carol J.R.; David W. Gaylor, and Kenneth F. Schulz. Estimators of relative risk for case-control studies. *Am J Epidemiol* 1983;118:396-407.

Koopman JS, Longini IM, Jacquez JA, Simon CP, et al. Assessing risk factors for transmission of infection. *Am J Epidemiol* 1991 (June); 133(12):1199-1209.

Lee, James. Odds ratio or relative risk for cross-sectional data? *Intl J Epidemiol* 1994;23:201-203.

Lee, James; KS Chia. Estimation of prevalence rate ratios for cross sectional data an example in occupational epidemiology. *Br J Ind Med* 1993;50:861-2.

Thompson, Mary Lou; J.E. Myers, D. Kriebel. Prevalence odds ratio or prevalence ratio in the analysis of cross-sectional data: what is to be done. *Occup Environ Med* 1998;55:272-277.

Walter, Stephen D. Calculation of attributable risks from epidemiological data. *Intl J of Epidemiol* 7:175-182, 1978.

Walter, Stephen D. Choice of effect measure for epidemiological data. *J Clinical Epidemiology* 2000;53:931-939.

Zhang, Jun; Kai F. Yu. What's the relative risk: a method of correcting the odds ratio in cohort studies of common outcomes. *JAMA* 1998;280(19):1690-1691.