

5. Midiendo la enfermedad y la exposición

Estadística descriptiva, medición de la ocurrencia y frecuencia de una enfermedad, prevalencia, incidencia (como proporción y como tasa) y sobrevivida, promedios ponderados, exponentes, y logaritmos.

“A menudo digo que cuando uno puede medir aquello del cual habla, y expresarlo en números, sabe algo sobre el tema; pero cuando no lo puede expresar en números, el conocimiento es de naturaleza escueta y no satisfactoria; puede ser el comienzo del conocimiento, pero uno ha apenas empezado la etapa de Ciencia en sus propios pensamientos, cualquiera sea el tema.”

Lord Kelvin (citado en Kenneth Rothman, *Modern Perspectives in Epidemiology*, 1 ed. Boston, Little Brown, 1986, pág. 23)

Al principio de este texto resaltamos cuatro aspectos claves de la epidemiología: su naturaleza multidisciplinaria, y su preocupación con poblaciones, la medición y la comparación. Como todos los científicos empíricos, los epidemiólogos dedican mucha atención a temas de medición – y la aplicación de números a los fenómenos. Cada objeto de estudio – una enfermedad, una exposición, un evento, una condición – debe ser definido y medido. Dado que la epidemiología trata de poblaciones, los epidemiólogos necesitan métodos para describir y resumir los fenómenos en poblaciones. Este capítulo discute varios aspectos de la medición, incluyendo la definición, el cálculo y la interpretación de las medidas claves de los eventos y estados de salud en las poblaciones. El siguiente capítulo trata de las comparaciones entre estas mediciones.

Los números: aplicándolos a los fenómenos

La “numerología” es el concepto de resumir los fenómenos cuantitativamente. Enfrentados a una realidad infinitamente compleja y detallada, el investigador tratar de identificar y cuantificar los aspectos que son importantes. Dos de los innumerables ejemplos de este proceso en epidemiología son:

Puntaje de Aterosclerosis: un epidemiólogo, David Freedman, que recibió su doctorado de la Universidad de Carolina del Norte (UNC), realizó la investigación para su tesis sobre la relación de la aterosclerosis en pacientes que se someten a una angiografía coronaria con los niveles plasmáticos de homocisteína. Un problema básico que tuvo que enfrentar fue cómo medir la aterosclerosis en las angiografías coronarias. ¿Debía clasificar los pacientes como teniendo una obstrucción clínicamente significativa, contar el número de obstrucciones, o tratar de asignarle un puntaje a la extensión de la aterosclerosis? Un puntaje de aterosclerosis captaría la mayor información y representaría mejor la naturaleza del fenómeno de aterosclerosis, tal cual sería afectado por los niveles de homocisteína. ¿Pero este puntaje de aterosclerosis, debe medir la superficie involucrada o el grado del estrechamiento? ¿Cómo debe considerar las lesiones distales a la oclusión, que no afectan para nada el flujo de sangre? Estas y otras decisiones necesariamente dependerían de su modelo conceptual de

cómo la homocisteína afecta el endotelio. Por ejemplo, ¿la homocisteína estaría involucrada principalmente en causar el daño inicial, en cuyo caso sería relevante la superficie total involucrada, o estaría involucrada en la progresión de la aterosclerosis, en cuyo caso sería el grado de estrechamiento que sería relevante? Se pueden llegar a compromisos por las limitaciones en cuanto a las mediciones que pueden ser realizadas en los angiogramas.

Medición del abandono del hábito de fumar: a primera vista, el abandono del hábito de fumar, en un estudio de los efectos del abandono del hábito o de la efectividad de un programa de abandono del hábito, parecería sencillo de definir y medir. Pero, aún en este caso, surgen varias interrogantes. Los beneficios del abandono pueden exigir una abstinencia durante un período largo (p.ej., años). Sin embargo, las técnicas de validación bioquímicas, consideradas necesarias cuando los participantes tendrían una razón para exagerar su éxito para abandonar, pueden detectar al que continúa fumando durante un período limitado de tiempo (p.ej., aproximadamente siete días para la cotinina salival). ¿Debe el abandono ser definido como el no uso de tabaco durante 7 días, para facilitar la validación, o por lo menos por un año, en que la tasa de recaída se hace mucho más baja?

Los modelos conceptuales subyacen las mediciones

En general, la forma en que usamos los números y el tipo de medidas que construimos depende de:

1. el objetivo de la medición
2. la naturaleza de los datos que tenemos disponibles
3. nuestros conceptos sobre el fenómeno

Estos tres factores determinarán el tipo de medidas que se usarán.

De ser posible, nos gustaría observar el fenómeno desarrollarse a través del tiempo. En la práctica a menudo debemos tomar unas medidas e inferir el resto del proceso. Los modelos conceptuales impregnan tanto el proceso de la aplicación de los números a los fenómenos como el proceso de analizar estadísticamente los datos resultantes para identificar patrones y asociaciones. Como no podemos registrar todos los aspectos del fenómeno de interés, debemos identificar aquellos aspectos que son biológica, psicológica, o de alguna manera epidemiológicamente importantes. Estos aspectos son plasmados en definiciones operacionales y clasificaciones. El método por el cual aplicamos los números y los analizamos debe mantener las características importantes sin agobiarnos con información superflua. Este concepto básico se mantiene para los datos sobre individuos (la unidad de observación habitual en epidemiología) y sobre poblaciones. Aunque empleamos modelos matemáticos y estadísticos como marcos para la organización de los números resultantes, para la estimación de las medidas y parámetros claves, y para examinar las relaciones, los modelos conceptuales guían todas estas acciones.

Niveles de medición

El punto en el cual ejercen su influencia los objetivos, la disponibilidad de los datos y los modelos conceptuales es **el nivel de medición** para un fenómeno o constructo específico. Consideremos el constructo de logro educativo, una variable que es ubicua en la investigación epidemiológica.

Podemos (1) clasificar a las personas como que son egresados o no de la secundaria; (2) clasificarlas en múltiples categorías (secundaria incompleta, secundaria completa, escuela de oficios, escuela técnica, universitario, profesional, postgrado); (3) registrar el nivel más alto alcanzado en la educación; o (4) registrar sus puntajes en pruebas estandarizadas, que tendríamos que administrar.

La primer opción presentada muestra la “medición” más básica que se puede realizar: una clasificación **dicotómica** (de dos categorías). Los sujetos se clasificarán como “casos” o “sanos”, “expuestos” o “no expuestos”, hombre o mujer, etc. Las comunidades pueden ser clasificadas como que tienen o no leyes para el uso obligatorio de cinturón de seguridad, como que tienen o no un programa de intercambio de agujas, etc.

Potencialmente la clasificación **policotómica** (más de dos categorías) daría más información, como país de origen, preferencia religiosa, grupo sanguínea ABO, o histología de un tumor (p.ej., células escamosas, de células pequeñas (en avena), adenocarcinoma). Una clasificación policotómica puede ser **nominal** – nombrando las categorías pero no ordenándolas en un rango, como ocurre con los ejemplos recién mencionados – u **ordinal**, en que los valores o categorías pueden ser ordenados en un rango de acuerdo a alguna dimensión. Por ejemplo, podemos clasificar pacientes como “no casos”, “casos posibles” “casos seguros”, o lesiones como mínima, moderada, severa, y fatal.

Los valores de los distintos niveles de una variable nominal no dan más información que la identificación del nivel, por lo cual pueden ser intercambiadas sin restricciones. Podemos codificar células escamosas como “1”, oat cell “2”, y adenocarcinoma “3”; o podemos ponerle “2” a células escamosas y “1” a oat cell “1” o aún “5”. Los números simplemente sirven de nombres. Los valores de los distintos niveles de una variable ordinal significan el orden de rango de los niveles. Los valores pueden ser cambiados, pero generalmente no intercambiados. Podemos usar “1”, “2”, “3” respectivamente para no caso, caso posible y caso seguro, o podemos usar “1” “3” “8”, pero no podemos usar “1” “3” “2”, dado que este código no mantendría el orden.

Cuando los propios valores, o por lo menos el tamaño de los intervalos entre ellos, dan información, entonces el fenómeno ha sido medido a nivel de **intervalo**. La temperatura medida en escala Fahrenheit es un escala de intervalo, dado que, aunque podemos decir que 80° F es el doble de 40° F, la proporción no tiene sentido en términos del fenómeno subyacente. Las escalas psicológicas a menudo se consideran escalas de intervalo. Lo que diferencia una escala de intervalo de la mayor parte de las medidas que usamos en las ciencias físicas es la ausencia de un **punto cero** fijo. Dado que sólo los intervalos tienen sentido, la escala puede ser llevada para arriba o para abajo sin que se cambie su significado. Una escala de intervalo con valores “1”, “1,5”, “2”, “3”, “4” podrían bien ser codificados “24”, “24,5”, “25”, “26”, “27”.

Una escala de **proporción**, sin embargo, tiene un punto cero no arbitrario, de manera que tanto los intervalos como las razones tienen significado. La mayor parte de las mediciones físicas (altura, presión arterial) son escalas de proporción. Los valores de una escala de proporción pueden ser multiplicados o divididos por una constante, como para un cambio de unidades, dado que las comparaciones de los intervalos y las razones no se alterarán. Si el valor B es el doble del valor A antes de la multiplicación, será el doble del valor A después de la multiplicación. Una escala de proporción con valores “1”, “1,5”, “2”, “3”, “4” puede ser transformada en “2”, “3”, “4”, “6”, “8”

(con la sustitución apropiada de las unidades), pero no en “2”, “2.5”, “3”, “4”, “5”, dado que solo se conservarían los intervalos y no las razones.

Un tipo de escala de proporción es un **conteo** como el orden de nacimiento o paridad. Un conteo es una variable **discreta** porque sus posibles valores pueden ser enumerados. Una variable **continua**, como la define la matemática, puede tomar cualquier valor dentro de un rango posible, y una infinidad de valores entre dos valores cualesquiera. Las mediciones en epidemiología no son ni remotamente tan precisas como en las ciencias físicas, pero muchas mediciones usadas en epidemiología tienen un número suficiente de valores posibles como para ser tratadas como si fueran continuas (p.ej., altura, peso, o presión arterial).

Sin embargo, sean continuas o discretas, tanto las escalas de intervalo como las de proporción implican generalmente una relación lineal entre los valores numéricos y el constructo que se quiere medir. Así, si medimos logros educacionales por el número de años de escolaridad completados, estamos implicando que el aumento de 10° grado al 11° grado es lo mismo que el aumento de 11° a 12° grado, aunque el completar este último grado permite la obtención del diploma de secundaria. También implicamos que completar el 12° grado con tres cursos avanzados en una escuela de alto rendimiento es lo mismo que completar 12° grado con cursos de recuperación en una escuela de bajo rendimiento, o completar 12° grado pero leyendo a nivel de 9° grado, o completar 12° grado pero no habiendo hecho más matemáticas que álgebra elemental, etc., sin mencionar los aspectos educacionales de viajar, hablar varios idiomas o haber aprendido un oficio. Aún la edad cronológica puede no ser considerada una escala de intervalo o proporción cuando ciertas edades tienen un significado especial (p.ej., 15 años, 18 años, 21 años, 40 años, 65 años). Muchas medidas que parecen estar en escalas de intervalo o de proporción en realidad no se comportan como tales, debido a un **efecto de umbral** (las diferencias entre los valores bajos no tienen un verdadero significado), **efecto de saturación** (diferencias entre los valores altos no tienen un verdadero significado) y otras **relaciones no lineales**.

Medidas absolutas y relativas — la importancia del denominador

Mientras que los valores absolutos de edad, logro educacional, presión arterial, y cigarrillos por día tienen significado, otras medidas se expresan como concentraciones (p.ej., 20 µg de plomo por decilitro de sangre, 500 células T por centímetro cúbico de sangre, 1.3 personas por habitación, 393 personas por kilómetro cuadrado) o **relativas** a alguna otra dimensión (p.ej., índice de masa corporal [peso/altura²], porcentaje de calorías dado por grasas, razón de colesterol total a colesterol HDL). La mayor parte de las medidas poblacionales no tienen mayor significado si no están relacionadas con el tamaño y las características de una población y/o los valores esperados, aunque sólo sea implícitamente. ¿Cómo podemos evaluar si un número de casos constituye un brote o aún una epidemia? (con excepción de casos de viruela que ha sido erradicada en el mundo entero). Por eso los epidemiólogos a menudo se refieren con desdén al número absoluto de casos o muertes como “datos del numerador”. Las excepciones ayudan a demostrar el principio general. Un pequeño número de casos de angiosarcoma de hígado en una fábrica llevó a una investigación que descubrió el riesgo que plantea el cloruro de vinilo. Un pequeño número de casos de adenocarcinoma de la vagina en mujeres adolescentes en un hospital llevó a la identificación del efecto de dietilbestrol (DES) para dicha patología. Un pequeño número de casos de Síndrome de Inmunodeficiencia Adquirida (SIDA) alertó a la salud pública del comienzo de la pandemia. Dado que estos casos eran

condiciones muy raras o nunca observadas previamente, ya había una expectativa definida sobre su frecuencia.

Tipos de razones

Como se demostró con varios de los ejemplos citados anteriormente, expresamos una cantidad en relación a otra formando una razón, que es simplemente el cociente entre dos números, un numerador dividido por el denominador. Las razones son generalizadas en epidemiología dado que permiten que el número de casos se exprese en relación a la población de la cual surgen.

Dos tipos especiales de razones en epidemiología son las proporciones y las tasas. Las proporciones son razones en las cuales el numerador “esta incluido en” o es “parte del denominador”. La aseveración de que 12% de la población tiene 65 o más años expresa una proporción, dado que las personas de 65 y más años son una fracción componente de la población. Dado que el numerador es una fracción componente del denominador, una proporción solo puede tener valores entre 0 y 1 inclusive. Las proporciones a menudo se expresan como porcentaje, pero se puede utilizar cualquier factor de multiplicación para obtener un número más fácil de expresar. Por ejemplo, la proporción 0.00055 a menudo se expresaría como 5.5 por 10,000 o 55 por 100,000. Señalemos que la razón de abortos con respecto a los nacimientos vivos, aunque de la misma magnitud, *no* es una proporción, dado que el numerador no está incluido en el denominador.

Aunque muchos tipos de razones (incluyendo las proporciones) son llamadas frecuentemente “tasas”, en su uso correcto una **tasa** es la razón de un cambio en una cantidad con respecto al cambio en otra cantidad, siendo la cantidad en el denominador a menudo tiempo (Elandt-Johnson, 1975). Un ejemplo clásico de una tasa es la velocidad que es un cambio de localización dividido por un cambio en el tiempo. Tasa de nacimiento, tasa de mortalidad y tasas de enfermedad son ejemplos si consideramos los eventos — nacimientos, muertes, casos recién diagnosticados — como la representación de un “cambio” en “la cantidad”. Las tasas mismas pueden ser absolutas o relativas, según si el numerador mismo es una razón que expresa el cambio en relación con algún denominador. La mayor parte de las tasas en epidemiología son tasas relativas, dado que como se ha descrito más arriba, el número de casos o eventos, en general, debe ser relacionado a la población de la cual surgen.

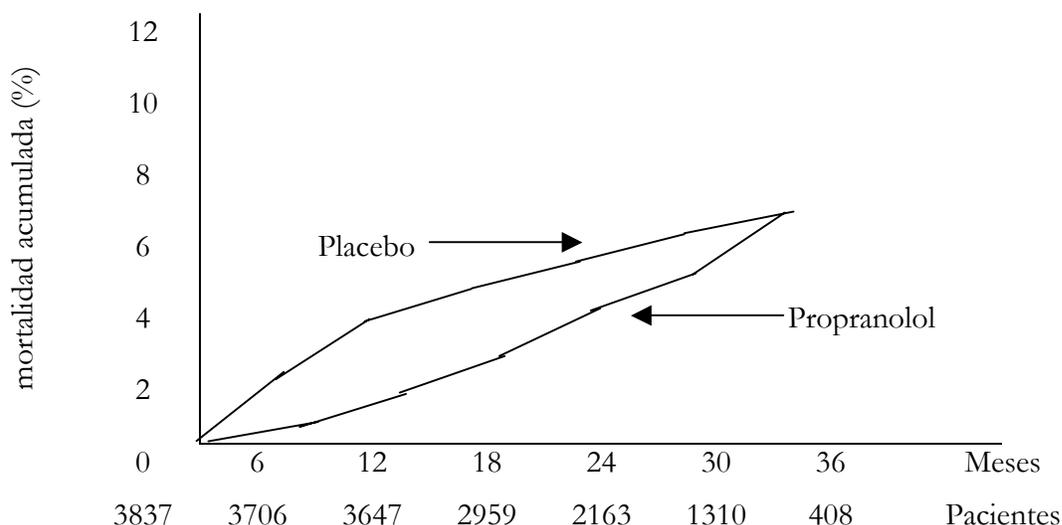
“Capturando el fenómeno”

Todas las medidas, por supuesto, son resúmenes o indicadores de una realidad compleja. La pregunta siempre es, “¿Captura la medida lo que es importante sobre el fenómeno, dado nuestro objetivo?”. Este principio se aplica tanto a nivel individual (por ejemplo, ¿cuándo pueden ser representados por un único valor representativo la presión sanguínea o los latidos cardiacos, que varían constantemente?) como a nivel poblacional.

Por ejemplo, aunque la proporción de pacientes que sobreviven 5 años es una medida de la efectividad de un tratamiento, si la proporción es baja, es particularmente importante cuando ocurren las muertes. La aseveración de que “la tasa de sobrevivencia a los cinco años de la cirugía de bypass coronario fue de 60%” no nos dice si el 40% que murieron lo hicieron durante el

procedimiento, poco después, gradualmente durante el período, o sólo después de 3 años de pasada la cirugía. Cuando el tiempo hasta la ocurrencia de un evento es importante, se usa el análisis de supervivencia, como en la siguiente figura, similar a la presentada por el Ensayo de Beta Bloqueantes en el Infarto Agudo de Miocardio [Beta-blocker Heart Attack Trial (BHAT)], un ensayo doble ciego, aleatorizado del uso de propranolol para tratar pacientes que sufren infarto agudo de miocardio.

Tabla de vida de la mortalidad acumulada en en el Ensayo de Beta Bloqueantes en el Infarto Agudo de Miocardio Beta [Blocker Heart Attack Trial (BHAT)]

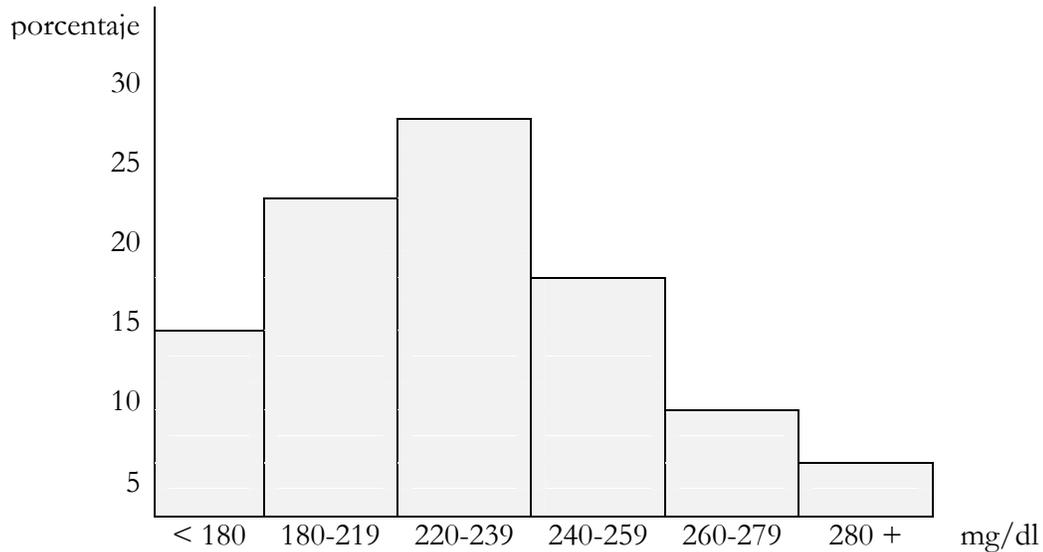


[Fuente: *JAMA*, Marzo 26, 1982; 247:1707]

Distribuciones –el panorama más completo

En general, cuando el objeto de estudio tiene que ver no sólo con la “presencia” u “ocurrencia” de un fenómeno sino con una variable *policotómica* o una *medida*, debemos examinar toda su distribución, p.ej.,

Niveles de colesterol sérico - Distribución



Aunque las distribuciones proveen mucha información, son engorrosas para trabajar y presentar. Por lo tanto tratamos de “captar” la información esencial sobre la distribución usando medidas de resumen, como el promedio, la mediana, o los cuartiles, y el desvío estándar o el rango intercuartílico (ver más adelante). Mientras que a veces es fundamental comprimir una distribución, una curva, o un fenómeno más complejo en una cifra o dos, hay que cuidar que la simplificación necesaria no distorsione los cálculos, la presentación y la interpretación resultantes. De hecho pueden ser las personas en un extremo de la distribución que son más importantes o dan más información sobre las consecuencias sobre la salud.

Si los datos se distribuyen en forma conocida, podemos caracterizar la distribución entera por sus parámetros (p.ej., el promedio y el desvío estándar para una distribución “normal” [Gaussiana]). Pero puede ser riesgoso suponer que los datos se ajustan a una distribución particular, sin verificar este supuesto examinando un histograma (p.ej., ver Estadística para Clínicos [*Statistics for Clinicians*] figura 7-7, varias distribuciones con el mismo promedio y desvío estándar pero apariencias notoriamente diferentes):

Medidas de resumen usadas comúnmente para descripción y comparación

Media – El valor “promedio” de la variable

Mediana – El centro de la distribución de la variable – la mitad de los valores quedan por debajo y la mitad por encima

Cuartiles – Los valores que delimitan el 1er, 2o. y 3er. cuarto de la distribución de la variable [la mediana es el 2o. cuartil].

Percentiles – Los valores que delimitan un porcentaje de la distribución, p.ej., el 20avo. percentil (también conocido como el segundo decil) es el valor por debajo del cual se encuentran el 20% inferior de las observaciones.

Desvío estándar – Groseramente es la distancia de una observación típica con respecto a la media o promedio (más precisamente, es la raíz cuadrada del promedio de las distancias de las observaciones desde la media elevadas al cuadrado). [No debe ser confundido con el **error estándar**, que es una medida de imprecisión de una estimación].

Rango intercuartílico – La distancia entre el 1er y el 3er cuartil.

Sesgo – El grado de asimetría alrededor del valor promedio de una distribución. Ser sesgado a la derecha o tener un sesgo a la derecha significa que la distribución se extiende a la derecha; en una distribución sesgada en forma positiva, la media (promedio global) se encuentra a la derecha de la mediana, debido a la influencia de los valores extremos.

Kurtosis – Es lo puntiagudo de la distribución con relación a la longitud y tamaño de las colas de la misma. Una distribución muy puntiaguda se dice que es “leptocúrtica”, una chata es “platicúrtica”.

Cuando se interpretan medidas de resumen, es importante considerar si las medidas representan los elementos más característicos de las distribuciones de las cuales provienen. Algunos ejemplos:

Promoción de la salud comunitaria:

Supongamos que encuestas antes y después de un programa de control del consumo de alcohol en la comunidad muestran que hay una disminución en el consumo promedio de alcohol de 1 bebida /día en la población blanco. Esta disminución puede reflejar :

- una disminución de 5 bebidas/día para cada persona en el 20 por ciento de la población que consume más

o

- una disminución de 1.25 bebidas/día para todas las personas menos aquellos en el 20% de mayor consumo,

con implicancias muy diferentes para la salud.

Diferencias blancos-negros en el peso al nacer:

La distribución del peso al nacer tiene una forma aproximadamente gaussiana (“normal”) con un rango que va desde aproximadamente 500 gramos (el límite inferior de la viabilidad) a aproximadamente 5,000 gramos con un promedio de alrededor de 3,000 gramos. Estadísticamente la distribución es suave y razonablemente simétrica. Sin embargo, las implicancias biológicas son muy variadas a lo largo de la distribución, dado que la mayoría de las muertes infantiles ocurren en los recién nacidos que pesan menos de 2,500 gramos. Para los recién nacidos que pesan entre 1,000 y 2,000 gramos, la tasa de mortalidad es de aproximadamente 33%; para los bebés de menos de 1,000 gramos la tasa de mortalidad es de 75%.

Las distribuciones de los pesos al nacer para los Norteamericanos de raza blanca y los de raza negra son en general similares, pero la de los negros tiene un pequeño desplazamiento hacia la izquierda. Ese pequeño desplazamiento hacia la izquierda se traduce en una

proporción sustancialmente mayor de recién nacidos con un peso menor a 2,500 gramos en que las tasas de mortalidad son mucho mayores.

Ingresos per capita

¿Deben los recursos para la atención de salud de la gente pobre ser adjudicadas en base al ingreso per cápita de las regiones? Por lo menos una investigación ha encontrado que las barreras a la atención de salud a las que se enfrentan los pobres en EEUU parecen ser las mismas en las regiones ricas y en las otras, de manera que el ingreso per cápita (i.e., el ingreso promedio por persona) no sería tan buen criterio para determinar las necesidades de los programas de salud pública como sería el número de personas pobres en el área (Berk M, Cunningham P, Beauregard K. The health care of poor persons living in wealthy areas. *Social Science in Medicine* 1991;32(10):1097-1103).

Moraleja: para interpretar un cambio o una diferencia en una medida de resumen es necesario conocer algo sobre la forma de la distribución y la relación entre la variable y el resultado de salud relevante.

Heterogeneidad y distribuciones de los factores desconocidos – cualquier medida de resumen es un promedio ponderado

Dado que las poblaciones difieren en características que afectan la salud, un número global, como una proporción o un promedio, a menudo esconden subgrupos que difieren en forma significativa de la visión general del fenómeno. Aún cuando no podamos identificar estos subgrupos, debemos ser conscientes de su probable existencia. Dado que la mayor parte de las enfermedades varían entre subgrupos, las medidas epidemiológicas son más interpretables conociendo la composición del grupo al cual están referidas, por lo menos en términos de las características demográficas básicas (en especial edad, sexo, área geográfica, nivel socioeconómico, empleo, estado civil, etnia) y las exposiciones importantes (p.ej., hábito de fumar).

P.ej., un grupo de trabajadores sufre 90 muertes por cáncer de pulmón por 100,000 por año. Para saber que opinar sobre esto es imprescindible conocer la distribución por edad del grupo y si es posible la distribución de las tasas de fumadores.

Casi cualquier medida en epidemiología puede ser considerada un promedio ponderado de la medida para los subgrupos componentes. Podemos usar medidas “específicas” (p.ej., “tasas específicas por edad” “tasas específicas por edad y sexo”) cuando la medida global (“cruda”) no da suficiente información. También podemos producir medidas “ajustadas” o “estandarizadas” en que una ponderación estándar se usa para facilitar la comparación entre grupos. Las medidas ajustadas son típicos promedios ponderados – los pesos son claves. El concepto de promedios ponderados es fundamental y volverá a aparecer en varios temas en epidemiología. (¿Dificultades con los promedios ponderados? Mira el apéndice sobre promedios ponderados.)

Tipos de medidas epidemiológicas

Objetivo de la medida:

Hay tres clases principales de medidas epidemiológicas según el propósito u objetivo. Usamos **medidas de frecuencia o distribución** para contestar preguntas como “¿Cuánto?”, “¿Cuántos?” “¿Qué tan frecuentemente?”, “¿Qué tan probable?”, o “¿Qué tan riesgoso?”. Usamos **medidas de asociación** para las preguntas que tienen que ver con la fuerza de asociación entre distintos factores. Usamos **medidas de impacto** para responder a preguntas como “¿Qué tan importante?”.

Disponibilidad de los datos:

También podemos clasificar las medidas epidemiológicas según el tipo de datos que son necesarios para obtenerlas:

1. Medidas derivadas de los sistemas rutinarios de recolección de datos, p.ej., registro de eventos vitales, registros de cáncer, notificación de enfermedades transmisibles.
2. Medidas derivadas de datos recolectados en estudios epidemiológicos o con objetivos relacionados (p.ej., estudios clínicos, registros de seguros de salud).
3. Medidas derivadas de estudios teóricos en biometría - no se necesitan datos! P.ej., riesgo de padecer la enfermedad en los expuestos = $\Pr[\text{Enf} | \text{Ex}]$

$$\text{Densidad de Incidencia} = \frac{d(N_t)}{N_t dt}$$

La utilidad de la tercer clase de medidas está en refinar los conceptos de las mediciones y en aumentar la comprensión de las mismas. Las medidas en las dos primeras clases generalmente involucran compromisos entre el ideal teórico y la realidad práctica. La epidemiología es un campo fundamentalmente práctico. En el resto del capítulo veremos brevemente el primer tipo de medida y algo más prolongadamente el segundo tipo.

Medidas derivadas de datos recolectados en forma rutinaria

En este sector encontramos los datos de estadísticas vitales recogidos por las autoridades de salud y las agencias de estadísticas como la Organización Mundial de la Salud, El Centro Nacional Norteamericano para Estadísticas de Salud, los departamentos de salud de los estados, y sus equivalentes en otros países. Ejemplos de medidas publicadas de dichos datos son:

- Tasa de mortalidad total
- tasa de mortalidad específica por causa

- tasa de natalidad (nacimientos por 1,000 habitantes)
- tasa de mortalidad infantil
- razón aborto/nacimiento vivo
- tasa de mortalidad materna

[Ver Mausner and Kramer, cap. 5; Remington y Schork, cap. 13.]

El denominador para las estadísticas vitales y otras tasas basadas en la población (p.ej., tasas de mortalidad, tasas de natalidad, tasas de matrimonio) generalmente se toma del censo nacional o de otros datos de eventos vitales, como en el caso de la tasa de mortalidad infantil:

$$\text{Tasa de mortalidad infantil} = \frac{\text{Muertes de niños menores de un año de edad en un año}}{\text{Total de nacimientos vivos en un año}}$$

Los resultados generalmente se llevan a una escala en que puedan ser expresados sin decimales (p.ej., 40 muertes por 1000 o 4,000 muertes por 100,000).

Apartado opcional – Evaluación de la precisión de una tasa estimada, de la diferencia de tasas, o razón de tasas de estadísticas vitales

Si r es una tasa (p.ej., una tasa de mortalidad infantil) y n es el denominador de esa tasa (p.ej., número de nacimientos vivos), se puede entonces construir un intervalo de confianza de 95% para r usando la fórmula:

$$r \pm 1.96 \times \sqrt{(r/n)}$$

P.ej., en un área con 30 muertes infantiles y 1,000 nacimientos vivos, $r = 30/1,000 = 30$ por 1,000 o 0.03. El intervalo de confianza de 95% para r es:

$$0.03 \pm 1.96 \times \sqrt{(0.03/1,000)} = 0.03 \pm 0.0107 = (0.0193, 0.0407),$$

o entre 19.3 y 40.7 por mil.

El intervalo de confianza de 95% para la diferencia, D , entre dos tasas, r_1 y r_2 , basadas respectivamente, en el número de muertes d_1 y d_2 , y los denominadores n_1 y n_2 , es:

$$(r_1 - r_2) \pm 1.96 \times \sqrt{(r_1/n_1 + r_2/n_2)}$$

El intervalo de confianza del 95% para la razón, R , de r_1 y r_2 es:

$$R \pm R \times 1.96 \times \sqrt{(1/d_1 + 1/d_2)}$$

Donde d_2 (el número de muertes de la tasa del denominador) es de por lo menos 100.

Fuente: Joel C. Kleinman. Infant mortality. Centers for Disease Control. National Center for Health Statistics *Statistical Notes*, Winter 1991;1(2):1-11.

La fundamentación de lo explicado más arriba es como se explica a continuación. El número de eventos raros en una población grande puede ser descrita por la distribución de Poisson, que tiene como característica notable que su media es igual a su varianza. Para una distribución de Poisson con una media d (y varianza d), si el número de eventos es suficientemente grande (p.ej., 30), el 95% de la distribución caerá en el intervalo $d \pm 1.96\sqrt{d}$. Si dividimos esta expresión por el tamaño de la población (n), podemos obtener el intervalo de confianza de la tasa con la siguiente fórmula:

$$d/n \pm (\sqrt{d})/n = r \pm \sqrt{r/n}$$

Los sistemas de notificación y registro de enfermedades específicas, ingresos hospitalarios, y consultas ambulatorias dan información sobre incidencia o utilización de los servicios de salud para algunas condiciones. Las enfermedades transmisibles hace mucho tiempo que son notificables, aunque cuan completo se hace la notificación, es bastante variable. Inversiones importantes en los registros de cáncer de los estados de EEUU están creando la base para un sistema de registro nacional de cáncer en EEUU. Varios estados tienen sistemas de notificación para los accidentes automovilísticos. En general, sin embargo, los datos sobre enfermedades no fatales están menos disponibles y son menos completos que los datos de mortalidad.

Recordemos: todas las tasas, razones, y otras medidas pueden ser:

Específicas para un grupo definido por edad, sexo, y/u otros factores.

Ajustadas por edad, sexo, u otra(s) variable(s) relevante(s);

Crudas (i.e., ni específica ni ajustada).

Estos términos se aplican con respecto a una variable particular por lo tanto no son mutuamente excluyentes. Por ejemplo, una tasa puede ser ajustada con respecto a edad, específica con respecto a género, y cruda con respecto a etnias, región geográfica, etc. (p.ej., una tasa ajustada por edad para mujeres de todas las etnias y todas las regiones geográficas).

El concepto básico subyacente a los procedimientos de ajuste es el del promedio ponderado. Las limitaciones de las medidas ajustadas derivan de este aspecto – la validez de la comparación depende de la similitud de los pesos componentes; la validez de la interpretación depende de la homogeneidad numérica y conceptual de las medidas componentes específicas.

Medidas obtenidas de datos recolectados en estudios epidemiológicos

Para la mayor parte de los estudios epidemiológicos, los datos recolectados en forma rutinaria no se adecuan a lo que se necesita, de manera que los datos deben ser recolectados específicamente para los propósitos del estudio. La recompensa por el tiempo, el esfuerzo, y el gasto es una mayor oportunidad de estimar medidas que son más precisas para las inferencias etiológicas y de otro tipo. Tres medidas principales de este tipo son la prevalencia, la incidencia, y la letalidad de los casos.

Prevalencia – la proporción de casos dentro de la población	
Prevalencia =	$\frac{\text{Casos}}{\text{Población-en-riesgo}}$

Prevalencia – un tipo de imagen “instantánea” – es la más básica de las medidas epidemiológicas. Se define como el número de casos dividido por la población-en-riesgo. Hagamos notar que:

- La prevalencia es una proporción, por lo tanto sus valores deben estar entre 0 y 1, inclusive.
- Población en riesgo (PER) significa “pasible de tener la condición”.
- La prevalencia puede ser usada para estimar la probabilidad de que una persona seleccionada al azar de la PER tenga la enfermedad [Pr(E)].

Ejemplo:

$$\begin{aligned} \text{Prevalencia} &= \frac{\text{No. de personas con demencia senil en un momento dado}}{\text{No. de personas en la población de estudio en riesgo de sufrir demencia senil}} \\ &= \frac{175}{1,750} = 0.10 = 10\% \end{aligned}$$

Apartado opcional – Evaluando la precisión de una prevalencia estimada.

Dado que la prevalencia es una proporción, se puede obtener un intervalo de confianza usando la distribución binomial, o cuando hay por lo menos 5 casos, la aproximación normal a la distribución binomial. La varianza de una variable aleatoria binomial es pq (donde p es la probabilidad de un “éxito” y $q=(1-p)$), de manera que el error estándar para la probabilidad estimada es $\sqrt{pq/n}$. Así, el intervalo de confianza del 95% para una prevalencia estimada, p , es $p \pm 1.96\sqrt{[p(1-p)/n]}$. Para el ejemplo anterior, los límites de confianza del 95% son $0.10 \pm 1.96\sqrt{[(0.10)(0.90)/1750]} = (0.086, 0.114)$. Cuando hay menos de 5 casos se requiere la utilización de un procedimiento exacto.

La prevalencia tiene tres componentes:

1. Casos existentes
2. Población “en riesgo” de sufrir la condición
3. Punto (o a veces un período) en el tiempo al que corresponde la prevalencia

Incidencia – la ocurrencia de casos nuevos	
Incidencia =	$\frac{\text{casos nuevos}}{\text{Población-en-riesgo durante el período de tiempo}}$

Incidencia – una “película” – describe lo que está pasando en la población. La incidencia se define como el número de casos nuevos dividido por la población en riesgo durante ese período. Por lo tanto la incidencia incluye tres componentes:

1. Casos nuevos
2. Población en riesgo.
3. Intervalo de tiempo.

Hagamos notar que:

- La incidencia involucra el pasaje de tiempo.
- La incidencia puede ser expresada como una proporción o como una tasa.
- La incidencia puede ser usada para estimar el riesgo de un evento durante un período de tiempo especificado.

Ejemplo:

$$\text{p.ej., Incidencia acumulada} = \frac{\text{Casos nuevos de demencia senil en 5 años}}{\text{No. de personas en riesgo}}$$

En la epidemiología de las enfermedades infecciosas, esta medida a menudo se denomina **tasa de ataque** o **tasa de ataque secundaria**, sobretodo cuando se refiere a la proporción de casos nuevos entre los contactos de un caso primario.

Letalidad de los casos es una medida de la severidad de la enfermedad. Aunque a menudo se llama la “tasa” de letalidad, la medida es generalmente calculada como una proporción:

Letalidad – proporción de los casos que mueren	
Letalidad a 5 años =	$\frac{\text{Muertes por la condición}}{\text{Número de personas con la condición}}$

Si el período de tiempo en discusión no engloba el período entero de riesgo de morir por la condición, entonces debe explicitarse el período de tiempo o no es posible interpretar la medida. La

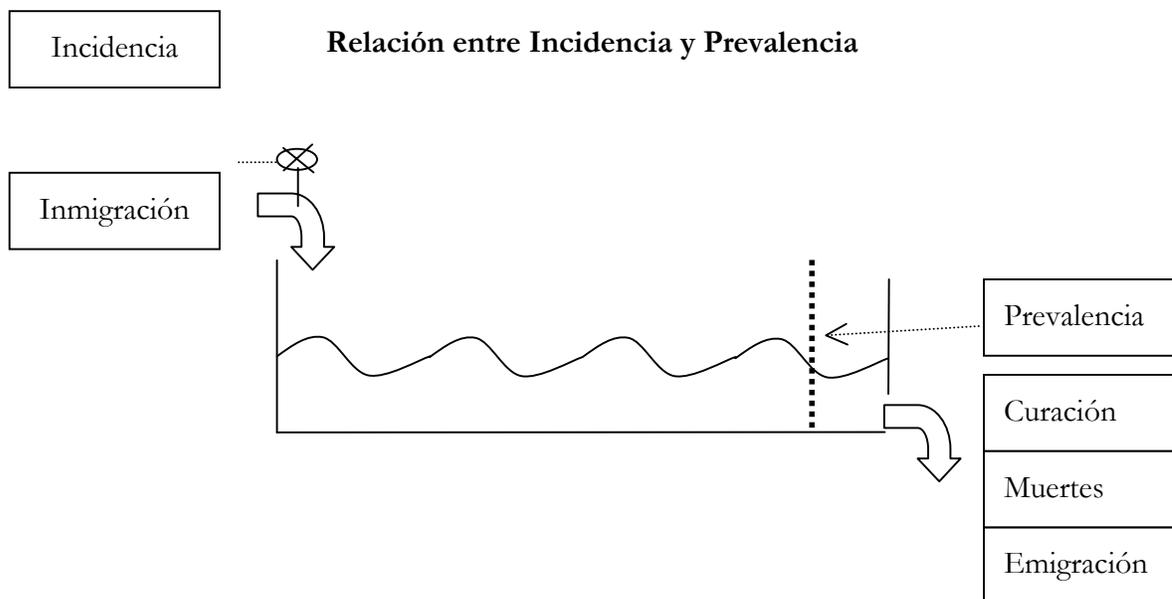
letalidad del SIDA aumenta cada año después del diagnóstico, pero el de un episodio de influenza o un procedimiento quirúrgico no cambia después del primer mes.

Ejemplo:

$$\text{Tasa de letalidad} = \frac{\text{Muertes por demencia senil en 5 años}}{\text{Número de personas diagnosticadas con demencia senil}}$$

Relación entre incidencia y prevalencia

Incidencia, mortalidad, y prevalencia están íntimamente relacionadas, por supuesto, igual que nacimientos, muertes y tamaño poblacional. Los demógrafos estudian estos últimos fenómenos, y sus técnicas son utilizadas en epidemiología (con otras denominaciones, naturalmente, para “proteger a los inocentes!”).



En una población estacionaria, en que no hay migración de casos o de personas sanos, si la incidencia, la prevalencia, y la duración de la condición permanecen constantes, el número de casos nuevos que ocurren debe ser balanceado por el número de casos existentes que dejan la población a través de la curación o la muerte. En esta situación, la prevalencia es función de la incidencia y la duración promedio de ser un caso. Para una enfermedad rara, la prevalencia \approx incidencia \times duración (ver “Incidencia y Prevalencia en una población”, más adelante).

Influencias sobre la relación entre incidencia y prevalencia

Las relaciones entre incidencia, mortalidad y prevalencia son afectadas por factores como:

Virulencia de la enfermedad - ¿Es rápidamente fatal?

Atención médica - ¿Cuando son detectados los casos por el sistema de atención?

¿Pueden ser curados los casos?

¿La detección temprana cambia el pronóstico?

Comportamiento - ¿La gente reconoce y actúa rápidamente frente a los síntomas?

¿Los pacientes cumplen el tratamiento?

¿Causas de muerte que compiten? - ¿La gente con la enfermedad tienen probabilidades de morir por otras causas?

Migración - ¿Es probable que las personas con la enfermedad se vayan de la zona?

¿Es probable que las personas con la enfermedad migren a la zona?

Dado que la prevalencia es afectada por factores (p.ej., duración y migración) que no afectan el desarrollo o la detección de la enfermedad o la condición, generalmente se prefieren las medidas de incidencia más que las medidas de prevalencia para estudiar la etiología y/o la prevención. Tanto la incidencia como la prevalencia son útiles para varios otros propósitos (vigilancia y control de enfermedades, planificación de la atención médica). Además la prevalencia es más fácil de estimar que la incidencia y puede servir para inferencias etiológicas a pesar de sus limitaciones.

Es importante señalar, sin embargo, que aunque la incidencia misma no es afectada por factores no relacionados con la etiología, la incidencia observada refleja la influencia de una variedad de factores no etiológicos (la rapidez con que la enfermedad produce síntomas que llevan a una consulta, acceso a atención de salud, si el proveedor de salud selecciona los procedimientos diagnósticos correctos, precisión de los resultados de los exámenes y su interpretación, y precisión y demora en el informe de los resultados). Hay, por lo tanto, grandes dificultades para la interpretación de la incidencia reportada de muchas enfermedades y condiciones (p.ej., enfermedad de Alzheimer, SIDA, VIH, otras enfermedades de transmisión sexual, enfermedad de Lyme, y cáncer de próstata, para nombrar solo algunos).

Un ejemplo de como la historia natural de la enfermedad distorsionó las tendencias en la incidencia observada es el de los primeros años de la epidemia del SIDA, cuando la notificación de los casos de SIDA era el medio principal de rastreo de la epidemia de HIV. Debido a la considerable variabilidad en el tiempo entre la infección por HIV y el desarrollo de las infecciones oportunistas que señalaban el inicio del SIDA, la tendencia al aumento en los casos de SIDA exageraba la tendencia en aumento de las infecciones por VIH. El mecanismo de este efecto puede ser ilustrado como sigue. Supongamos que la cantidad de infecciones nuevas por VIH durante los cuatro primeros años fueron 500, 1,000, 1,500, 2,000, respectivamente, indicando un aumento lineal de 500/año. Supongamos que 5% de las infecciones por VIH progresan a SIDA en el año siguiente a la infección, para un mediana de tiempo hasta el SIDA de 10 años. En el primer año ocurrirán 25 casos de SIDA (5% de 500 infecciones). Durante el segundo año se producirán 75 casos de SIDA (5% de

500 más 5% de 1,000). En el tercer año se producirán 150 casos de SIDA (5% de 500 más 5% de 1,000 más 5% de 1,500). En el cuarto año se producirán 250 casos de SIDA, de manera que la tendencia en SIDA (25, 75, 150, 250) inicialmente aparentaría aumentar en forma más pronunciada que la tendencia en VIH (las infecciones por VIH se duplican en el año 2, pero los casos de SIDA se triplican) y luego aparentan estabilizarse a pesar de que no hay cambios en la tendencia de la incidencia del VIH. También habrá un cambio en la razón de SIDA a VIH, como ocurrió en los primeros años de la epidemia. (El fenómeno fue descrito en un artículo del *American Journal of Epidemiology* alrededor de 1987; estoy buscando la cita).

Prevalencia versus incidencia		
	<u>Prevalencia</u>	<u>Incidencia</u>
Casos	Entidades	Eventos
Población de origen (PER)	En riesgo de <u>ser</u> un caso	En riesgo de <u>convertirse</u> en un caso
Tiempo	Estático (puntual)	Dinámico (de intervalo)
Usos	Planificación	Investigación etiológica

Consideraciones relevantes tanto para prevalencia como incidencia

Casos

1. **Definición de caso** – ¿Qué es un caso?

Ejemplos: artritis, colestiasis, enfermedad cardiovascular, diabetes, trastornos siquiátricos, tratamiento epidemiológico de sífilis o gonorrea, cáncer de próstata

2. **Desarrollo del caso** – ¿Cuándo es un caso?

Temas: inducción, latencia, progresión, reversibilidad

Ejemplos: aterosclerosis, cáncer, colestiasis, diabetes, hipertensión, SIDA

3. **Detección de un caso** – ¿Cuándo un caso es un “caso”?

Temas: La detectibilidad es función de la tecnología y la factibilidad. Lo que se puede detectar no es lo mismo que lo que se detecta.

Ejemplos: Aterosclerosis, cáncer de mama, colestiasis, osteoporosis, infecciones asintomáticas, cáncer de próstata

Población de origen [Población en riesgo (PER)]

1. ¿Cuál es la población relevante? — ¿quién está realmente “en riesgo”?

P.ej., edad (la mayor parte de las enfermedades), sexo (cáncer de mama), ETS y actividad sexual, cáncer de útero e histerectomía, cáncer de vesícula biliar y colecistectomía, genotipos?

2. ¿Qué pasa con las manifestaciones previas?
 ¿de la misma enfermedad? (influenza, tumores, lesiones)
 de una enfermedad relacionada (accidente vascular después de enfermedad cardíaca coronaria, cáncer en un lugar diferente)
4. ¿Y la muerte por otras causas? (riesgos en competencia)
 P.ej., las muertes por diabetes disminuyen la tasa de mortalidad por enfermedad coronaria, las muertes por enfermedad cardíaca disminuyen la tasa de mortalidad por cáncer de pulmón al punto que los fumadores tienen un riesgo excesivo para ambos

Seleccionando el denominador correcto

La selección del denominador más apropiado puede ser muy complejo. Por ejemplo, cuál es el denominador más apropiado para las lesiones o muertes por accidentes de tránsito?

- ¿Población total?
- ¿Población de 18 años y más?
- ¿Los que tienen licencia de conducir?
- ¿Vehículos registrados?
- ¿Kilómetros recorridos por los vehículos?
- ¿Kilómetros recorridos por los pasajeros?

Cual elegir depende si el tema de interés tiene que ver con:

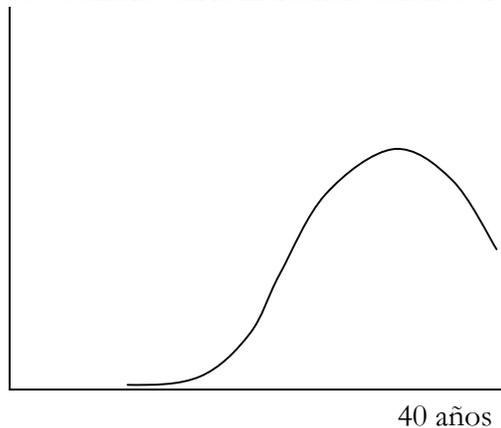
- Riesgo de lesión por edad y/o sexo (¿denominador poblacional?)
- Efecto sobre el riesgo del uso del cinturón de seguridad (¿kilómetros pasajero?)
- Efecto sobre las muertes del límite de velocidad especificado (¿kilómetros pasajero?)
- Papel del alcohol en las fatalidades por accidentes de tránsito
- Evaluación de políticas alternativas de transporte

Por ejemplo, los conductores mayores tienen una tasa más alta de choques por 100 millones de kilómetros recorridos que los adolescentes. Pero la tasa de choques por conductor con licencia no es más alta para los conductores mayores, porque los conductores mayores limitan su tiempo manejando.

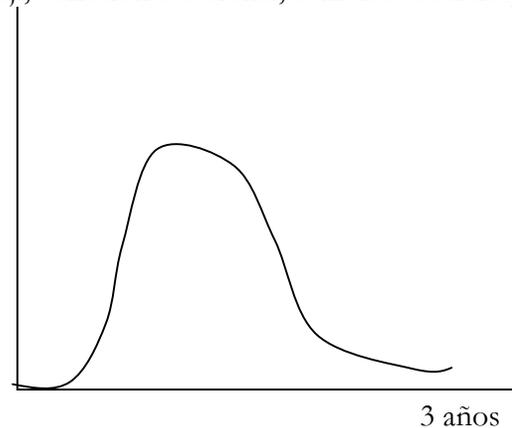
Pasaje de tiempo [sólo incidencia] – ¿cuál período de observación?

1. Historia natural de la enfermedad - período de riesgo versus período de observación
 P.ej., sobrevivientes de la bomba atómica y los tumores sólidos, lesiones por accidentes de vehículos de motor, malformaciones congénitas
2. Diferentes períodos de observación para distintos sujetos (¿1 persona observada durante 2 años = 2 personas observadas 1 año?)

3. Cambios en la incidencia durante el período (p.ej., variación estacional, cambios seculares)



Cáncer en los sobrevivientes de la bomba atómica



Malformaciones congénitas

Tipos de población de origen para incidencia

Las poblaciones de origen pueden ser definidas de varias maneras, incluyendo la residencia en un área geográfica determinada, empleo en una compañía o industria, asistencia a una escuela o universidad, pertenencia a una organización, tributario de atención médica de un proveedor particular, o reclutamiento explícito para el estudio. La incidencia involucra el pasaje de tiempo y por lo tanto implica algún tipo de seguimiento de la población. Una característica clave de la población de origen es la manera en que sus integrantes pueden cambiar a través del tiempo. Rothman y Greenland (1998) presentan una discusión detallada de los tipos de poblaciones y la terminología que ha sido usada para describirlas. La distinción principal que haremos es entre una **cohorte fija**, cuyos integrantes cambian sólo por atricción, y una **población dinámica** (Rothman y Greenland llaman esto una **cohorte abierta**), cuyos integrantes pueden cambiar de varias maneras. (La terminología de cohorte fija versus población dinámica viene de Ollie Miettinen a través de Kleinbaum, Kupper, y Morgenstern.)

Cohorte– el ingreso a la población se define en base a algún aspecto o evento en la vida de los integrantes de la población de estudio (p.ej., vivir en el área geográfica cuando ocurre un evento ambiental importante, comenzar a trabajar en un empleo o industria, recibir un tratamiento médico o quirúrgico, comienzo de una afección, comienzo de una exposición, o simplemente apuntarse para el estudio). Los abandonos de la cohorte (por muerte, emigración, pérdida) son problemáticos; los ingresos a la cohorte sólo se permiten en relación con el evento que define el comienzo del seguimiento para esa persona. Señalemos que una vez que se ha completado el reclutamiento la cohorte se hará más pequeña al pasar el tiempo por las pérdidas, y la distribución por edad de toda la cohorte se hará mayor.

Variantes:

Cohorte Retrospectiva o histórica - la población es definida en algún momento del pasado (p.ej., basado en los registros de empleo) y luego seguida hacia delante en el tiempo hasta el presente a través del uso de los registros disponibles.

“Cohorte dinámica” – el tiempo de seguimiento se cuenta desde el momento del ingreso al estudio o en relación a algún evento que ocurre en distintos momentos para diferentes personas (p.ej., un procedimiento médico), de manera que el ingreso a una cohorte ocurre durante un período de tiempo. En un estudio clásico de cohortes, el tiempo de seguimiento para cada sujeto y el tiempo calendario son idénticos; en una cohorte dinámica, el seguimiento de cada participante puede ocurrir durante un intervalo de tiempo calendario diferente.

Población dinámica – una población es definida durante un período de tiempo y su experiencia es monitorizada durante ese período. La población de estudio puede ser definida de la misma manera (p.ej., residencia geográfica, empleo, afiliación, etc.). En una población dinámica, sin embargo, tanto el ingreso como los egresos son esperados y tenidos en cuenta. Por ejemplo, la población de un área geográfica experimentará nacimientos, muertes, y posiblemente una migración importante. A través del tiempo, una población dinámica puede aumentar o disminuir en tamaño, y su distribución por edad puede cambiar o mantenerse igual.

Caso especial:

Una población dinámica se dice que es **estable** o **estacionaria** cuando su tamaño y distribución por edad no cambian a través del tiempo. A menudo se supone que es estacionaria, dado que simplifica enormemente el análisis (Ver Rothman y Greenland, 1998, para más sobre este tema.)

Tipos de medidas de incidencia: incidencia acumulada (proporción de incidencia) y densidad de incidencia (tasa de incidencia)

Hay dos tipos principales de medidas de incidencia, que difieren fundamentalmente en la forma en que construyen el denominador: **incidencia acumulada** y **densidad de incidencia** (de nuevo, esta es la terminología de Olli Miettinen, adoptada por Kleinbaum, Kupper, y Morgenstern; Rothman y Greenland usan **proporción de incidencia** y **tasa de incidencia**, respectivamente). La incidencia acumulada (IA) es simplemente la proporción de una población que sufre el evento o desarrolla la condición durante el período de tiempo establecido. La densidad de incidencia (DI) es la velocidad a la cual se desarrollan los casos nuevos en la población, con relación al tamaño de la población.

Incidencia acumulada (proporción de incidencia)	
IA	= $\frac{\text{Casos nuevos en el período especificado}}{\text{Número de personas en riesgo}}$

Densidad de incidencia (tasa de incidencia)	
	Casos nuevos durante el período especificado
DI =	<hr style="width: 50%; margin: auto;"/>
	Población-tiempo

Incidencia acumulada (IA), también conocida como proporción de incidencia (PI)

La definición de IA se basa en el siguiente escenario “ideal”:

1. Se identifica una población libre del resultado en algún momento del tiempo (una cohorte);
2. Todos los integrantes de la cohorte están en riesgo de sufrir el evento o el resultado (por lo menos una vez) durante todo el período de tiempo;
3. Todos los primeros eventos o resultados para cada persona son detectados.

Por ejemplo, consideremos una investigación del riesgo que un oficial de policía novato sufrirá una lesión por arma de fuego durante sus primeros seis meses de patrullaje. Los datos son recolectados para una cohorte de 1,000 oficiales de policía recién entrenados que comienzan sus tareas de patrullaje con el Departamento de Policía de San Francisco (DPSEF). En sus primeros seis meses en el departamento, 33 de los policías sufren una herida de arma de fuego. Los otros 967 policías han llevado a cabo sus patrullajes durante los 6 meses sin sufrir heridas de arma de fuego. La IA a seis meses de una lesión por arma de fuego es de $33/1,000 = 0.033$. Usamos esta IA observada para estimar el riesgo a seis meses de sufrir una lesión por arma de fuego en los nuevos policías de patrullaje en San Francisco

Este ejemplo se ajusta al escenario ideal para IA: hay una población “en riesgo” y “a la vista” durante el período entero, y todos los primeros eventos fueron conocidos. Por el momento ignoraremos todas las razones que pueden resultar en que un integrante de la cohorte no permanezca “en riesgo” (p.ej, transferencia a un trabajo de escritorio, licencia médica prolongada, abandono del trabajo) y “a la vista” (p.ej., tomado por otro departamento de policía).

Algunas cosas a señalar sobre la IA:

1. El período de tiempo debe ser especificado (p.ej., “IA de 5 años) o evidente a partir del contexto (p.ej., enfermedad aguda después de la exposición a una fuente de alimentos contaminados);
2. Dado que la IA es una proporción, lógicamente cada persona puede ser contada como un caso una sola vez, aún si él o ella sufre más de un evento;
3. Como toda proporción, los valores de la IA solo pueden estar entre 0 y 1 (inclusive), que es una razón por la cual puede ser usada para estimar el riesgo en forma directa (la probabilidad de un evento).

Cálculo de muestra:

200 personas libres de la enfermedad crónica X observados durante 3 años
se desarrollan 10 casos de X
IA de 3 años = 10 casos / 200 personas = 10/200 = .05

Así, el riesgo a los 3 años para que una de las 200 personas desarrolle la enfermedad X, condicionado al hecho de no morir de otra causa, se estima como 0.05 o 5%.

Apartado opcional – Evaluando la precisión de una incidencia acumulada estimada

Dado que la incidencia acumulada es una proporción, el intervalo de confianza se obtiene de la misma manera que para la prevalencia (ver apartados previos).

Riesgo y “odds”

En epidemiología, el término “riesgo” habitualmente se entiende como la probabilidad de que un evento tendrá lugar en un intervalo de tiempo establecido o implícito (igual estén atentos a otros usos). En su uso epidemiológico, un riesgo es una probabilidad condicional, porque es la probabilidad de sufrir un evento o convertirse en un caso condicionado al hecho de mantenerse “en riesgo” (elegible para ser un caso) y “a la vista” (disponible para que el evento sea detectado).

Cualquier probabilidad puede ser transformada en una medida relacionada, los “odds”. El *odds* se define como la razón entre la probabilidad de un resultado y la probabilidad de otro resultado. Cuando los únicos resultados son (caso, no caso), el odds es la razón de la probabilidad de convertirse en un caso a la probabilidad de no convertirse en un caso. Si el riesgo o probabilidad de convertirse en un caso [Pr(E)] es p , entonces el odds de convertirse en un caso son $p/(1-p)$. Si el riesgo, o probabilidad, de desarrollar la enfermedad X es 0.05 (5%), el odds de desarrollar la enfermedad X es $.05/.95 = 0.0526$ (el odds es siempre mayor que el riesgo, sobretodo cuando el riesgo es alto).

Las propiedades matemáticas de los odds los hacen muy ventajosos para múltiples usos. Mientras que las probabilidades están restringidos al intervalo 0 – 1, el odds puede ser cualquier número no negativo. Un odds = 1.0 (“cincuenta-cincuenta”) corresponde a una probabilidad = 0.5, el centro del conjunto de posibles valores. El logaritmo del odds puede, por lo tanto ser cualquier número real, siendo el $\log(\text{odds}) = 0$ correspondiendo al centro del conjunto de posibles valores. El logaritmo natural (neperiano) del odds (llamado “logit”, por “transformación logarítmica”) se usa ampliamente en bioestadística y epidemiología. Para continuar el ejemplo anterior, con un riesgo de 5%, el odds = 0.0526, el $\ln(\text{odds})$, o logit = -2.944; dado que el $\ln(\text{odds})$ es cero cuando el riesgo es .5, un riesgo menor a 0.5 da un logit negativo. [¿se te han olvidado los logaritmos? Mira el apéndice sobre logaritmos y exponentes.]

Incidencia acumulada cuando hay pérdidas al seguimiento

En el ejemplo anterior, las 200 personas que estaban originalmente libres de la enfermedad X fueron observadas durante los 3 años. ¿Qué pasaría si 20 personas hubieran fallecido por otras causas antes

de desarrollar la enfermedad X? Entonces, no todos los 200 hubieran estado “en riesgo” los 3 años enteros.

Hay cuatro alternativas principales para estimar la IA de tres años:

1. Ignorar las muertes:

$$IA \text{ de 3 años} = 10/200 = .05$$

2. Ignorar los que murieron (analizamos sólo las personas que fueron seguidas durante los 3 años enteros):

$$IA \text{ de 3 años} = 10/(200-20) = .056$$

3. Llegar a un compromiso contando las 20 personas que murieron como si fueran 10 personas en riesgo los 3 años enteros:

$$IA \text{ de 3 años} = 10/(200-20/2) = .053$$

4. Usar una tabla de vida, en que (a) IA se calcula para cada parte del período (p.ej., anual) para estimar el riesgo durante esa parte del período; (b) los riesgos son convertidos en probabilidades de sobrevivida (1-riesgo); y (c) los riesgos se multiplican para obtener la probabilidad de sobrevivida a los 3 años y por lo tanto el riesgo a tres años (1 - probabilidad de sobrevivida).
5. Tomar la inversa de la proporción de sobrevivida estimada de Kaplan-Meier. Este método es igual al anterior salvo que los períodos se hacen tan cortos que ocurre un único caso en un período cualquiera. Los períodos en que no hay casos tienen un 100% de sobrevivida, de manera que la sobrevivida estimada de Kaplan-Meier es el producto de la proporción de sobrevivientes en cada intervalo en que ocurre un caso

Cada uno de estos métodos incluye ciertos supuestos sobre el momento en que ocurre la enfermedad en el período de tres años, sobre si será detectada cuando ocurre, y si las personas que murieron por otras causas tenían mayor o menor probabilidad de desarrollar X si hubieran vivido.

Densidad de incidencia (DI)

$$DI = \frac{\text{Casos nuevos durante el período especificado}}{\text{Número de personas-año de observación}} \quad (\text{personas-meses, etc.})$$

Señalemos que:

- DI es una tasa relativa, no una proporción.

- Las unidades de tiempo deben ser especificadas, dado que de otra manera el valor numérico es ambiguo (p.ej., 15 casos/100,000 personas-año = 15 casos/1,200,000 personas-meses).*
- En forma ideal, la densidad de incidencia es la tasa instantánea de ocurrencia de la enfermedad en cada momento en el tiempo. En la práctica, los epidemiólogos generalmente calculan una DI promedio durante uno o más períodos.

Interpretación:

DI contesta la pregunta “¿Cuán rápido está ocurriendo la enfermedad en la población, con relación a su tamaño?”, o “¿Cuál es la intensidad con que está ocurriendo la enfermedad?”. Se ha discutido que la DI no tiene interpretación a nivel individual (ver Morgenstern H, Kleinbaum, DG, Kupper LL, 1980). Sin embargo, es posible que la DI puede ser considerada como respuesta, aunque sea indirectamente a la pregunta, “¿Qué tan pronto podrá esto ocurrirme a mi?”.

Cálculo de muestra:

En nuestro ejemplo original para IA, teníamos 10 casos de la enfermedad crónica X que se desarrollaron entre 200 personas que inicialmente eran libres de la enfermedad X y que fueron observados durante 3 años sin pérdidas al seguimiento. A continuación están los valores de IA y DI para este ejemplo:

$$CI \text{ de 3 años} = 10 \text{ casos} / 200 \text{ personas} = 10/200 = .05$$

$$DI \approx 10 \text{ casos} / (200 \text{ personas} \times 3 \text{ años}) = 10 / 600 \text{ personas-años}$$

$$\approx 0.167 \text{ casos por persona-año (p-a)} = 0.167 / \text{p-a} = 167 / 1000\text{p-a}$$

La razón de la aproximación es que, como veremos, las personas dejan de contribuir con tiempo-persona cuando desarrollan la enfermedad de manera que el denominador debe disminuirse de acuerdo con esto. El cálculo más correcto es $10 / (200 \times 3 \text{ años} - 10 \times 1.5 \text{ años}) = 10/585 = 0.17/\text{p-a}$, suponiendo que los casos ocurrieron en forma uniforme durante los 3 años.

* The importance of stating units can perhaps be appreciated from the following: “On Sept. 23, 1999, NASA fired rockets intended to nudge its Mars Climate Orbiter into a stable low-altitude orbit. But after the rockets fired, NASA never heard from its expensive spacecraft again, and scientists later concluded that it had either crashed on the Martian surface or had bounded away, escaping the planet completely. “The reason for the debacle, scientists concluded months later, was that the manufacturer, the Lockheed Martin Corporation, had specified the rocket thrust in pounds, while NASA assumed that the thrust had been specified in metric-system newtons.” Browne, Malcom W. Refining the art of measurement. Science Times, New York Times, 3/20/2001, page D6¶.

Calculando DI

Al calcular la DI, usamos los mismos casos que para la IA salvo que podemos querer permitir múltiples eventos por persona. Si consideramos que las recurrencias son independientes unas de otras, podemos simplemente contarlos como casos nuevos. Si no, podemos definir la enfermedad como la primer ocurrencia. Hay otras consideraciones que también afectan la elección.

Hay varios métodos usados para calcular la población-tiempo.

- 1) Si los individuos son seguidos a través del tiempo, de manera que se conoce el tiempo de observación libre de enfermedad para cada persona, simplemente sumamos todo el tiempo libre de enfermedad para todas las personas:

$$\text{tiempo-población} = \Sigma (\text{tiempo libre de enfermedad de cada persona})$$

- 2) Si se está siguiendo una cohorte fija, pero no con suficiente detalle para conocer el período de tiempo libre de enfermedad para cada individuo, podemos estimar el tiempo-población como sigue:

$$\begin{aligned} \text{tiempo-población} &= \text{tamaño poblacional promedio durante el período} \\ &\times \text{duración del período de observación} \end{aligned}$$

Si hay N_0 personas libres de enfermedad al principio del período, y durante el período hay “C” casos, “M” muertes por otras causas que no son la enfermedad de interés, y “A” personas cuya situación de enfermedad es desconocida (“abandonos”), entonces el número de personas libres de enfermedad al final del período es: $(N_0 - C - M - A)$. El promedio de personas libres de enfermedad, suponiendo que los casos, las muertes y los abandonos ocurren en forma uniforme durante el período, es:

$$\frac{N_0 + (N_0 - C - M - A)}{2} = (N_0 - C/2 - M/2 - A/2)$$

Y el tiempo- población en riesgo puede ser estimado como:

$$(N_0 - C/2 - M/2 - A/2) \times (\text{intervalo de tiempo})$$

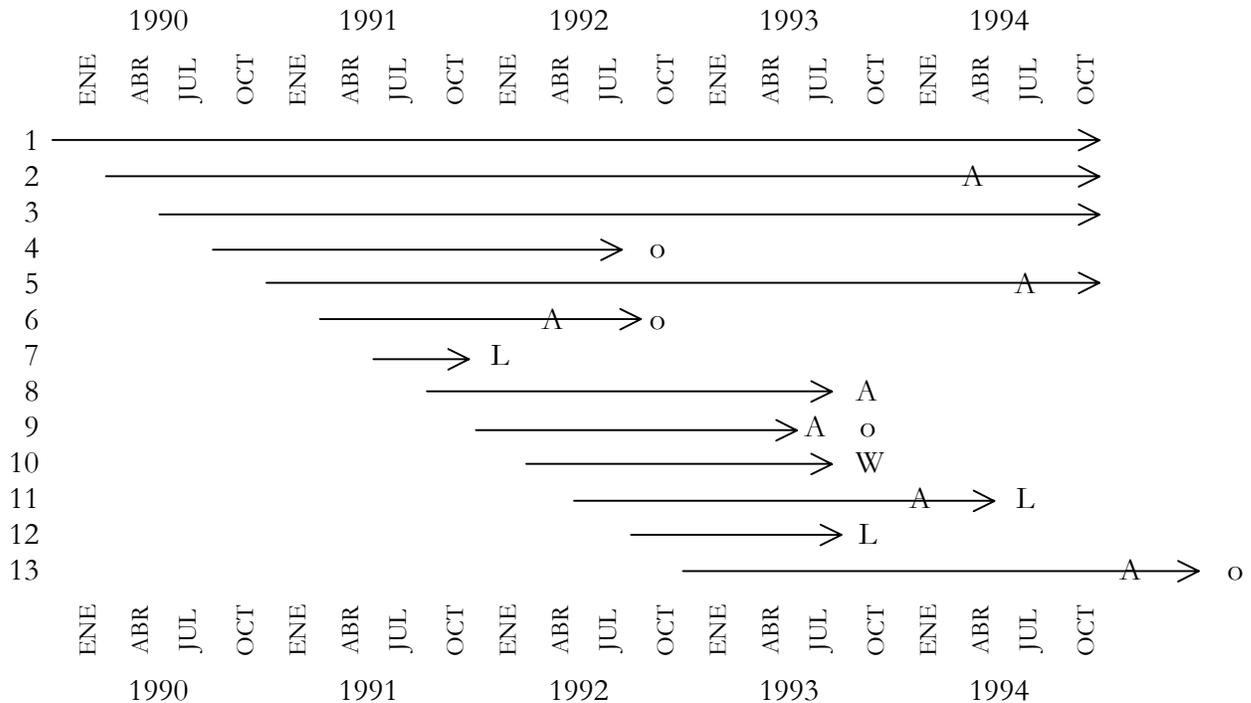
- 3) Si estamos siguiendo una población dinámica (también conocida como una “cohorte abierta”) en vez de una cohorte fija, podemos usar la misma estrategia de multiplicar el tamaño de la población libre de enfermedad por el período de tiempo. Puede ser posible estimar el número promedio de personas libres de enfermedad tomando el promedio del número de personas libres de enfermedad al principio y al final del período. Si podemos suponer que la población es “estable” (el número de personas libres de enfermedad que se pierden de la población a través de la emigración, muerte y el desarrollo de la enfermedad de interés es balanceada por la inmigración), entonces el número de personas libres de enfermedad es aproximadamente constante. Si tenemos cualquier estimación utilizable del

número promedio de personas libres de enfermedad (N_0), podemos estimar el tiempo-población como $N_0 \times$ (intervalo de tiempo).

Si la enfermedad es rara, el número de personas libres de enfermedad (N_0) será aproximadamente igual al número total de personas (N), que es más probable que sea conocido. En este caso, podemos estimar el tiempo-población como $N \times$ (intervalo de tiempo), donde N es el tamaño poblacional promedio sin tomar en cuenta el estado de enfermedad. Las tasas de mortalidad anuales y otras tasas de estadísticas vitales típicamente se calculan usando la estimación de la población a mitad de año (1 de Julio) como denominador, que es aproximadamente el tamaño promedio de la población cualquier día del año si la población es aproximadamente constante o cambia en forma monótona.

Cálculo de tiempo-persona en una cohorte cuando se conocen los tiempos individuales de seguimiento

Gráfico de la experiencia de seguimiento hipotética de 13 pacientes con Alzheimer avanzado que están siendo tratados en domicilio en el período de Enero 1990 - Diciembre 1993 y seguidos hasta Diciembre 31, 1994 para ingreso a un hogar de ancianos, en orden, por fecha de ingreso al estudio (de Kleinbaum, Kupper, and Morgenstern, 1982).



Claves :

A = admitido a hogar

L = perdido al seguimiento

W = retirado

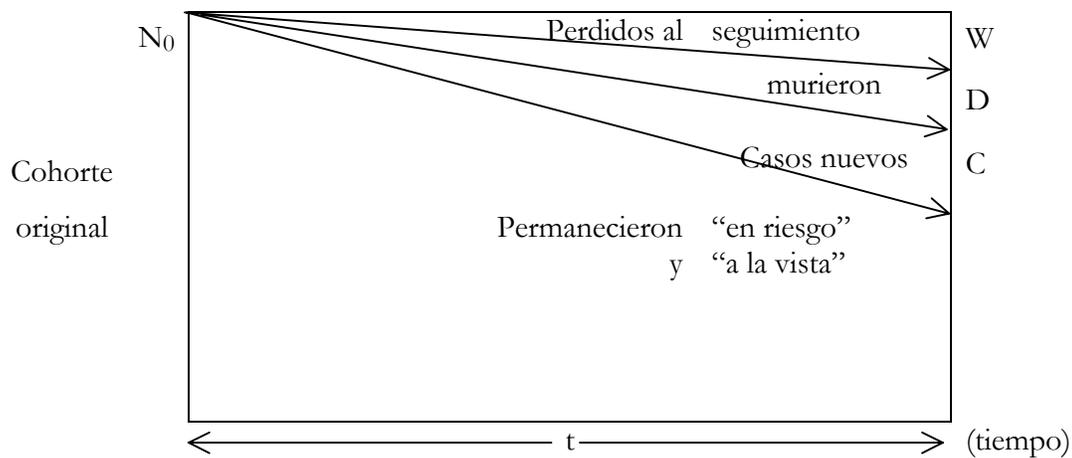
o = fallecido

$$DI = \frac{\text{Casos}}{\text{Suma del seguimiento libre de enfermedad de todos los individuos}}$$

Sujeto	Casos	Seguimiento
1		5.0
2	1	4.0
3		4.5
4		2.0
5	1	3.5
6	1	1.0
7		0.5
8	1	2.0
9	1	1.5
10		1.5
11	1	1.5
12		1.0
13		2.0
Total	6	30.0

$$DI = \frac{6}{30 \text{ años persona}} = 0.20 \text{ pacientes admitidos por año}$$

**Cálculo de tiempo-persona en una cohorte
cuando los tiempos de seguimiento individuales no son conocidos**

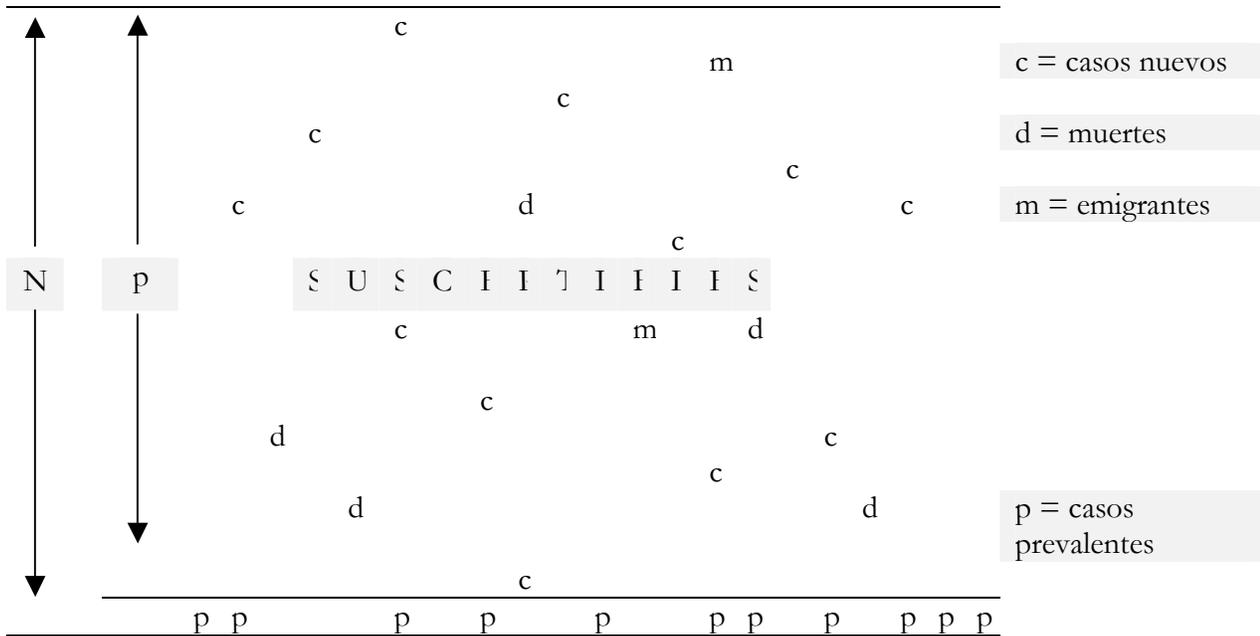


$$DI = \frac{C}{(N_0 - W/2 - D/2 - C/2) t}$$

(t = intervalo de tiempo)

(Dado que el área de un triángulo = base \times altura/2, el tiempo-persona perdido al seguimiento puede ser estimado por una vez y media el número de retirados [la base del triángulo] por la duración del intervalo de tiempo [la altura]. El procedimiento es igual para el tiempo de seguimiento perdido por muertes y los casos incidentes. Estas estimaciones suponen que los casos son detectados a medida que ocurren y que sólo se cuenta el primer caso por sujeto.)

Cálculo de tiempo-persona en una población, estable, dinámica



- Procesos existentes:
- Inmigración de casos, no casos
 - Emigración de casos, no casos
 - Muerte de casos, no casos
 - Desarrollo de nuevos casos

$$DI = \frac{\text{casos}}{N_{0t}} \quad \text{o} \quad DI = \frac{\text{casos}}{N_t}$$

(t = intervalo de tiempo)

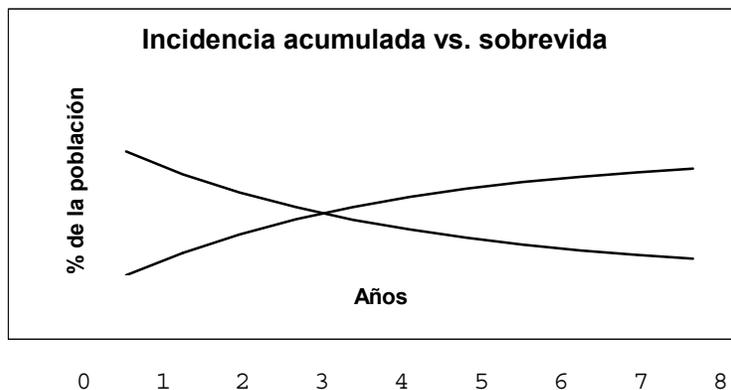
Relación entre IA y DI

Tanto la IA como la DI son en realidad viejos conocidos que han cambiado sus aspectos. Cuando calculamos la esperanza de vida en el primer tópico, usamos los términos tasa de mortalidad, riesgo instantáneo, mortalidad acumulada, sobrevida acumulada. La DI es esencialmente el riesgo instantáneo, ahora aplicado a eventos distintos a la muerte. La IA es esencialmente la proporción de mortalidad acumulada, ahora aplicada a eventos de cualquier tipo. Ambos representan distintas medidas de resumen del análisis de sobrevida (conocido en ingeniería como análisis de tiempo-falla).

La DI es la velocidad a la cual está cambiando el tamaño de la población no afectada, en relación al tamaño de la población no afectada; IA es la proporción de la población original que ha sido

afectada por el tiempo t . IA es una medida acumulada desde el momento inicial de referencia hasta un punto específico posterior en el tiempo. La IA estima el riesgo promedio para un integrante de la cohorte. En principio la DI puede referirse a un instante en el tiempo, aunque sólo puede ser calculado como promedio de un intervalo. La DI se denomina a veces la “fuerza de morbilidad”, en analogía con la función de riesgo instantáneo (la “fuerza de mortalidad”).

La siguiente figura muestra la relación entre IA y su inverso, la proporción no afectada (sobrevivida). La DI es la velocidad relativa de la disminución en la curva de sobrevivida.

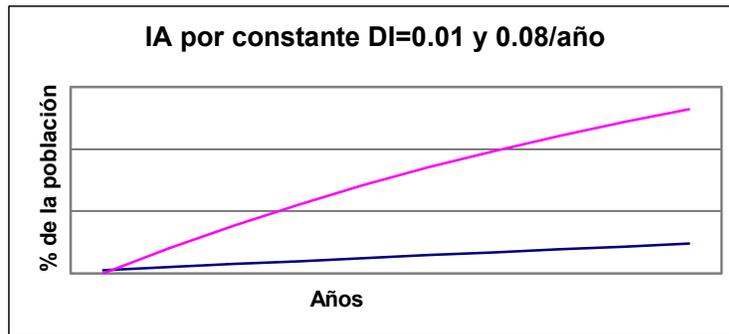


La incidencia de SIDA en hombres homosexuales de San Francisco puede parecerse a la mitad izquierda de esta gráfica.

Se puede apreciar la relación matemática entre IA y DI a través del tiempo considerando una enfermedad incurable en una cohorte hipotética fija definida en un momento en el tiempo y sin forma de ingreso o salida de ella más que por la enfermedad en cuestión. Suponiendo que la DI_t (la fuerza de morbilidad) es constante a través del tiempo, los casos se desarrollarán durante el período de seguimiento siguiente. Sin embargo, dado que el número de integrantes de la cohorte no afectados (en riesgo) está disminuyendo, el número de casos nuevos será menor en cada intervalo de tiempo sucesivo. Dado que el número de casos es menor en cada intervalo, la pendiente de la curva para la IA tenderá a aplanarse a medida que se aproxima a 1.0 (su valor máximo), cuando la cohorte entera ha desarrollado la enfermedad. La proporción de no afectados (lo inverso de IA: $1 - IA =$ también se hace menos pronunciado en su pendiente. Hemos supuesto, obviamente, que DI_t es constante. En esta situación, la relación matemática entre IA y DI es:

$$IA = 1 - \exp[-\int (DI_t dt)] = 1 - \exp(-DI \Delta t)$$

Para una enfermedad rara con una DI constante (o durante un intervalo de tiempo suficientemente corto): $IA \approx DI \times \Delta t$ (donde Δt es el intervalo de tiempo), porque como la cohorte no se agota, el número de casos nuevos en cada intervalo de tiempo permanece aproximadamente igual.



Ejemplo:

- $DI = 0.01/\text{año}$ (1 caso por 100 años persona)
- En 5 años, IA será 0.049, o aproximadamente lo mismo que $DI \times 5$ ($=0.05$); 95% de la cohorte permanece libre de enfermedad y por supuesto expuesta a la DI de 0.01/año.
- En 10 años, la IA será .096, apenas por debajo de la $DI \times t$ ($=0.10$); 90% de la cohorte permanece libre de enfermedad.
- $DI = 0.05/\text{año}$ (5 casos por 100 años persona)
- En 5 años, IA será 0.226, algo menor que $DI \times 5$ ($=0.25$); 77% de la cohorte permanece libre de enfermedad.
- En 10 años, IA será 0.40, mientras que $DI \times t$ ($=0.50$); sólo 60% de la cohorte permanece libre de enfermedad.

AI vs. DI - un ejemplo de la vida real

(cortesía de Savitz DA, Greenland S, Stolley PD, Kelsey JL. Scientific standards of criticism: a reaction to “Scientific standards in epidemiologic studies of the menace of daily life”, [Estándares científicos de crítica: una reacción a “ Estándares científicos en los estudios epidemiológicos de los peligros de la vida diaria”] por A.R. Feinstein. *Epidemiology* 1990;1:78-83; Fue en realidad Charles Poole que se dio cuenta de este *faux pas* [Poole C, Lanes SF, Davis F, *et al.* “Occurrence rates” for disease (letter). *Am J Public Health* 1990; 80:662]; el tema específico en discusión aquí es el efecto del alcohol sobre el riesgo de cáncer de mama.

“. . . tasas de ocurrencia de cáncer de mama sustancialmente diferentes: aproximadamente **6.7 por mil** (601/89,538) en la cohorte de enfermeras y aproximadamente **18.2 por mil** (131/7,188) en la cohorte NHANES.” (Feinstein AR. Scientific standards in epidemiologic studies of the menace of daily life. *Science* 1988;242:1259 citada en Savitz DA y col., p.79, énfasis agregado.)

Implicancias:

- (1) Distintas tasas sugieren errores en el diagnóstico de cáncer de mama.

(2) Si hay sobre o sub-detección, puede haber una detección sesgada.

(3) El sesgo puede producir un diagnóstico más completo o sobre-diagnóstico entre los bebedores

Sin embargo:

Enfermeras: 601 casos/89,538 mujeres en 4 años

IA = 6.7 por mil (4 años)

DI = 1.68 por 1,000 mujeres - años

NHANES: 121 casos/7,188 mujeres en 10 años (10 casos tendrían que haber sido excluidos por Feinstein)

IA = 16.8 por mil (10 años)

DI = 1.68 por 1,000 mujeres-años

Este ejemplo ilustra la importancia de señalar el período de seguimiento para una IA y el problema que puede surgir al comparar las IA para diferentes períodos de seguimiento.

Dos medidas complementarias de incidencia: IA y DI

Incidencia acumulada (IA)

1. aumenta con el período de observación (i.e., es “acumulada”)
2. tiene problemas con:
 - eventos múltiples en un único sujeto
 - distintos tiempos de seguimiento para los sujetos

Pero

3. no es necesario conocer el momento exacto de la aparición de la enfermedad
4. estima directamente el riesgo

Densidad de incidencia (DI)

1. sugiere la posibilidad se extrapolar hacia el futuro - “libre de duración”;
2. puede absorber:
 - múltiples eventos en un único sujeto
 - distintos tiempos de seguimiento de los sujetos
3. no requiere una cohorte para su estimación o interpretación
4. puede ser más apropiada para la inferencia etiológica

Eligiendo entre IA y DI

A. Objetivo

Estimación de tasa o riesgo

B. Historia Natural

¿El período de interés está incluido en el período de observación? (período de riesgo restringido versus extendido)?

P.ej., si uno quisiera analizar la longevidad relativa de hombres y mujeres, el riesgo de muerte en la vida (IA) sería inútil.

C. Disponibilidad de datos, p.ej.

Cohorte fija, cohorte dinámica, población dinámica

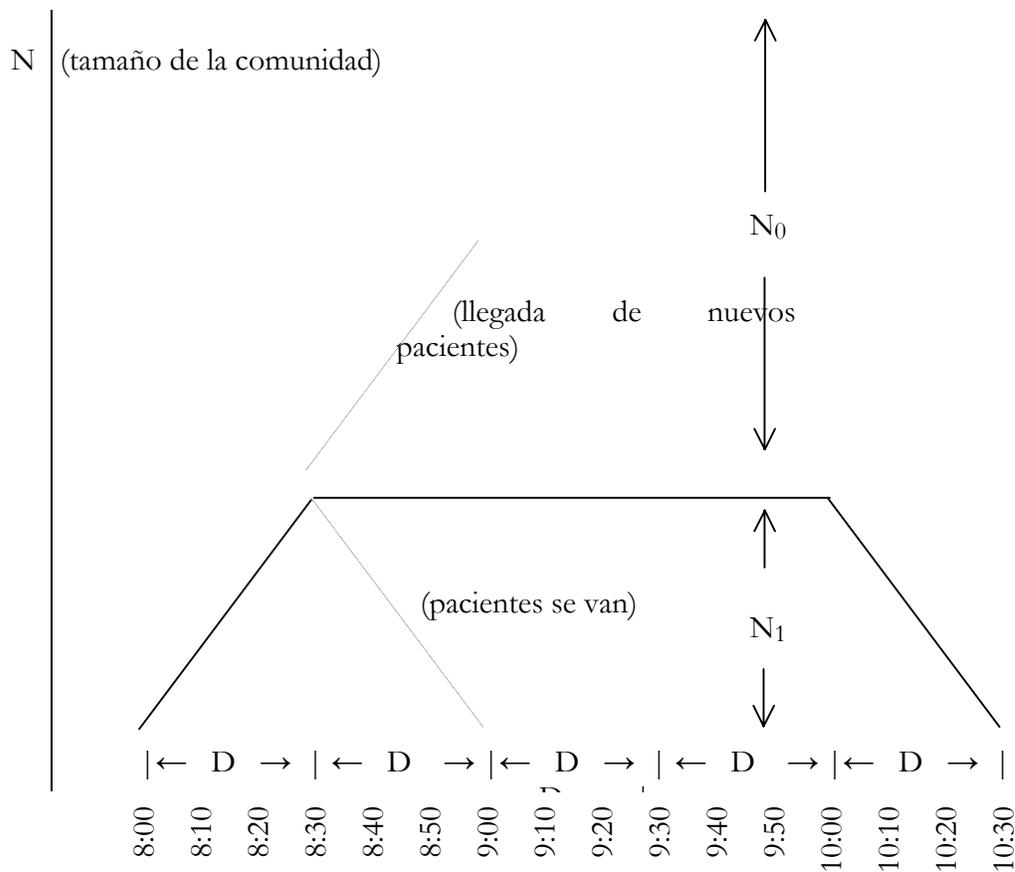
Distintos tiempos de seguimiento

El conocimiento de cuando ocurrirán los eventos puede favorecer un método u otro.

Incidencia y prevalencia en una población

Una variedad de situaciones familiares – por ejemplo, personas en la cola en el cajero del almacén, pacientes en la sala de espera o el hospital, accesos simultáneos a un servicio de internet – son analogías útiles para la relación entre incidencia y prevalencia en una población. Consideremos una clínica comunitaria.

Incidencia, prevalencia, y duración: flujo de pacientes en una clínica comunitaria



Si la clínica abre a las 8:00am, un paciente llega cada 10 minutos (6/hora), y lleva 30 minutos para que el paciente sea atendido y tratado, el número de pacientes en la clínica aumentará los primeros

30 minutos y luego permanecerá constante en 3 pacientes hasta que la clínica cierre y los 3 últimos pacientes son tratados. Si la velocidad con la que llegan los pacientes aumentara a 10/hora, en la primer media hora que se necesita para tratar el primer paciente, 5 más llegarán, de manera que el número de pacientes se estabilizará en 5, en vez de 3. De igual manera, el prolongar el tiempo de tratamiento de 30 a 60 minutos haría aumentar el número en la clínica en la primera hora, para un total de 6 pacientes en la clínica en cualquier momento dado hasta la hora del cierre.

Con los supuestos originales, 6 pacientes llegan a la clínica cada hora entre las 8:00am y las 10:00am, y 6 pacientes se van cada hora entre 8:30am y 10:30am. Entre las 8:30am y 10:00am la clínica está en equilibrio, con 3 pacientes allí en cualquier momento dado. Este número de equilibrio, N_1 , es igual a la velocidad de llegada (6/hora) por el tiempo promedio que el paciente permanece en la clínica (0.5 horas):

$$N_1 = \text{Tasa de llegada} \times D$$

Donde D es el promedio de duración de una visita clínica.

Si la clínica es la única en una comunidad de tamaño N (o es la asignada para la atención de N personas), podemos expresar la tasa de llegada como una función del tamaño de la comunidad:

$$\begin{aligned} \text{Tasa de llegada} &= I \times N_0 \\ \text{(pacientes/hora)} & \end{aligned}$$

Donde I es la incidencia de visitas a la clínica y N_0 es el número de personas disponibles para ir a la clínica (N menos N_1 personas ya dentro de la clínica, lo cual supone que las personas pueden volver a la clínica en cuanto se van o que inmediatamente dejan la comunidad y son reemplazados por otras personas que son elegibles para ir a la clínica). También podemos expresar el número de pacientes en la clínica, N_1 , como función del tamaño de la comunidad, usando P como la “prevalencia” poblacional de concurrencia a la clínica.

$$N_1 = P \times N$$

Usando estas tres ecuaciones, podemos escribir:

$$\begin{aligned} N_1 &= \text{Tasa de llegada} \times D \\ &= (I \times N_0) \times D \\ P \times N &= (I \times N_0) \times D \end{aligned}$$

$$P = \frac{N_0}{N} I \times D$$

Odds de prevalencia \approx incidencia \times duración promedio

Si el número de visitantes a la clínica es pequeño en relación al tamaño de la comunidad, entonces $N_0/N \approx 1$, y tenemos la aproximación **prevalencia \approx incidencia \times duración promedio**. Si no es así, la relación puede ser expresada como **odds de prevalencia = incidencia \times duración promedio**, dado que:

$$P = \frac{N_0}{N} I \times D = \frac{N - N_1}{N} I \times D$$

$$P = (1 - P) \times I \times D \quad \text{y} \quad \frac{P}{(1 - P)} = I \times D$$

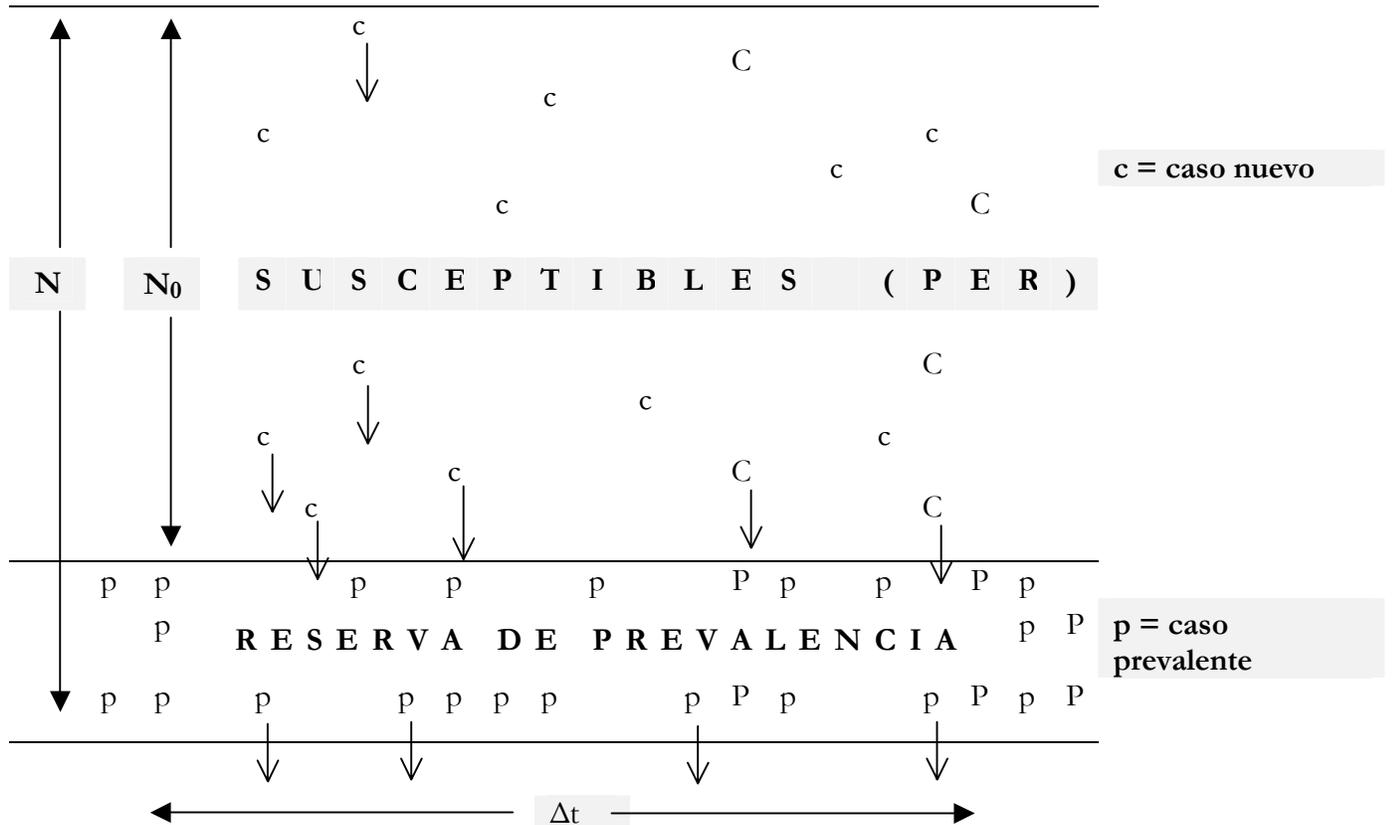
Odds es definido como la razón de dos probabilidades, habitualmente la razón de una probabilidad dividida por su inverso (probabilidad de/probabilidad de no). La prevalencia de una condición es una estimación de la probabilidad de que un integrante seleccionado al azar de la población sea un caso [Pr(caso)]. Si la prevalencia es **p**, entonces el odds de prevalencia es **p/(1-p)**. De manera que el odds de prevalencia, i.e., el odds de que una persona seleccionada al azar de la población tenga la enfermedad (i.e., es un caso prevalente) es:

$$\begin{aligned} \text{Odds de prevalencia} &= \text{prevalencia} / (1 - \text{prevalencia}) \\ &= (N \times \text{prevalencia}) / (N - N \times \text{prevalencia}) \\ &= (N \times \text{prevalencia}) / N_0 = (N/N_0) \times \text{prevalencia} \end{aligned}$$

Incidencia, prevalencia, y duración en una población estacionaria

El siguiente diagrama muestra el proceso descrito como podría darse para los casos de una enfermedad que ocurre en una población seguida durante un período de tiempo, en equilibrio con respecto a la incidencia de la enfermedad, duración, y los ingresos y egresos hacia y desde la población. Una derivación de la relación odds de prevalencia = incidencia \times duración se encuentra a continuación. (Ver Rothman y Greenland, 1998 para más detalles sobre este tema).

Incidencia y prevalencia en una población de tamaño N observada durante un intervalo de tiempo Δt



c son los casos incidentes (nuevos)

p son los casos prevalentes (existentes)

Δt indica el intervalo de tiempo

↓ indica salida desde la población no afectada o de la reserva de prevalencia

Tamaño de población = N = personas libres de enfermedad + casos existentes = N_0 + reserva de prevalencia

El supuesto de que la incidencia y la prevalencia son constantes significa que:

$$\text{Casos nuevos} = \text{Finalizaciones}$$

$$(\text{Incidencia} \times N_0) \times \Delta t = (\text{Prevalencia} \times N \times \text{tasa de finalización rate}) \times \Delta t$$

$$\text{Prevalencia} \times \frac{N}{N_0} = \frac{\text{Incidencia}}{\text{Tasa de finalización}}$$

Dado que la tasa de finalización es la velocidad a la cual los casos existentes dejan la reserva de prevalencia, esta tasa es la recíproca de la duración promedio de un caso. Para demostrar esto veamos la tasa de finalización de un caso único:

$$\text{Tasa finalización} = \frac{\text{Finalizaciones}}{\text{No. de casos} \times \Delta t}$$

Para un caso único,

$$\text{Tasa finalización} = \frac{1}{1 \times \Delta t} = \frac{1}{\Delta t}$$

$$\text{Duración promedio (i.e., } \Delta t) = 1 / \text{Tasa finalización}$$

Así, en la relación descrita entre incidencia y prevalencia, podemos sustituir Duración (D) por $1 / \text{Tasa de finalización}$:

$$\text{Prevalencia} \times \frac{N}{N_0} = \frac{\text{Incidencia} \times \text{Duración}}{\text{Duración}}$$

De esta manera, en una población que está en estado estable con respecto a una condición dada, el odds de prevalencia de esa condición es igual a la incidencia por la duración promedio (también la prevalencia, si es suficientemente pequeña). A la inversa, si observamos que el odds de prevalencia de una condición permanece constante (y podemos suponer que la población es estable sin migración neto de casos), entonces la incidencia debe balancear la pérdida de casos por muerte o curación. Dado que la prevalencia es a menudo más fácil de determinar que la incidencia, podemos usar esta relación para realizar inferencias sobre la incidencia.

Estimando la incidencia a partir de los datos de prevalencia

Esta relación ha sido usada como base para estimar la seroincidencia de VIH a partir de los datos de seroprevalencia, usando un procedimiento de detección en suero diseñado para identificar las personas recientemente infectadas (Janssen y cols., 1998). Esta técnica usa el hecho de que las pruebas de ELISA para el anticuerpo VIH se han hecho mucho más sensibles desde que se desarrollaron por primera vez. Las personas que fueron VIH positivas con la prueba de los anticuerpos VIH actuales (muy sensibles) son nuevamente testadas con una versión anterior, menos sensible, de la prueba. Dado que lleva tiempo para que los títulos de anticuerpos anti-VIH aumenten a un nivel detectable con la prueba menos sensible, hay un período de tiempo (como de cuatro meses) durante el cual la prueba menos sensible será negativa. Los resultados discordantes de las dos pruebas de anticuerpos

anti - VIH define una “condición” de poca sobrevivida (duración promedio 129 días) cuya prevalencia puede ser usada para estimar la ocurrencia de nuevas infecciones por VIH. Solucionando la ecuación $\text{odds de prevalencia} = I \times D$ nos da $I = \text{odds de prevalencia} / D \approx P/D$ para prevalencias bajas. De manera que si la seroprevalencia de la infección reciente en una población estable es de 2%, la incidencia de nuevas infecciones por VIH es aproximadamente $0.02/129 \text{ días} = 0.057 / \text{año} = 5.7/100 \text{ personas año}$.

Sin embargo, el supuesto de la población estable no se ve frecuentemente en la práctica, y el modelo descrito también está groseramente simplificado en el sentido de que trata la población entera como una entidad homogénea, ignorando la influencia de la edad (ver Rothman y Greenland, 1998). Cuando examinamos la relación entre incidencia y prevalencia dentro de un rango de edad de interés específico, necesitamos considerar el efecto de ingresos y egresos al mismo por envejecimiento desde y hacia la banda de edad de interés. Por ejemplo, las fuerzas armadas de EEUU han llevado a cabo las pruebas serológicas para el anticuerpo VIH de todos los reclutas, personal militar activo, y de los reservistas desde que está disponible la prueba. Dentro de cada grupo de edad, la seroprevalencia de anticuerpos anti VIH ha sido aproximadamente constante a través de los años. Si pudiéramos ignorar los efectos de la edad, usando la relación $\text{prevalencia de odds} = \text{incidencia} \times \text{duración promedio}$, podemos concluir que la incidencia de VIH debe ser igual a la (pequeña) proporción de sujetos que dejan la reserva de prevalencia cada año debido a la baja o la muerte. Sin embargo, otra forma de salir de la reserva de prevalencia de un grupo de edad particular es de llegar a edades que ya no son de este grupo y pasar al siguiente. Dado que la seroprevalencia de VIH aumenta con la edad (hasta la edad de aproximadamente 35 años), puede ser inferido que las infecciones (casos incidentes) ocurren más rápidamente que lo necesario para balancear las muertes y las bajas entre los casos. La razón es que cada año, algunas de las personas en el grupo de edad son reemplazadas por personas del grupo de edad inmediatamente inferior, un grupo con una seroprevalencia menor. Si las infecciones no ocurrieran a una velocidad suficiente para balancear este flujo hacia fuera de casos prevalentes, la prevalencia en cada grupo de edad disminuiría con el tiempo a medida que los grupos de menor prevalencia van envejeciendo (ver David Sokol y John Brundage, *Métodos de vigilancia para la epidemia de SIDA [Surveillance methods for the AIDS epidemic]*, *NYS J Medicine* Mayo 1988).

Es más, una medida de incidencia con significado aún requiere la identificación de una cohorte o población de origen. Aunque la prueba serológica antigua para la detección de la infección reciente por VIH ha sido utilizada para estimar la “incidencia” por VIH entre los pacientes de la clínica, la interpretación de esas estimaciones es muy problemática (Schoenbach, Poole, and Miller, 2001).

Bibliografía

Bailar, John C., III; Elaine M. Smith. Progress against cancer? *N Engl J Med* 1986; 314:1226-32. (included in an assignment)

Elandt-Johnson, Regina C. Definition of rates: some remarks on their use and misuse. *Am J Epidemiol* 1975;102:267-271.

Gable, Carol Brignoli. A compendium of public health data sources. *Am J Epidemiol* 1990; 131Z:381-394.

Gaffey, WR. A critique of the standardized mortality ratio. *J Occupational Medicine* 1976;18:157-160.

Glantz, Stanton H. *Primer of biostatistics*. NY, McGraw-Hill, 1981.

Hook, Ernest B. Incidence and prevalence as measures of the frequency of birth defects. *Am J Epidemiol* 1983;116:743-7

Janssen RS, Satten GA, Stramer SL, et al. New testing strategy to detect early HIV-1 infection for use in incidence estimates and for clinical and prevention purposes. *JAMA* 1998; 280:42-48.

Morgenstern H, Kleinbaum, DG, Kupper LL. Measures of disease incidence used in epidemiologic research. *Int J Epidemiology* 9:97-104, 1980.

Remington, Richard D. and M. Anthony Schork. *Statistics with applications to the biological and health sciences*. Englewood Cliffs, NJ, Prentice-Hall, 1970.

Rothman, Kenneth J. Clustering of disease. Editorial. *Am J Public Health* 1987; 77:13-15.

Victor J. Schoenbach, Charles Poole, William C. Miller. Should we estimate incidence in undefined populations? *Am J Epidemiol* 2001;153(10):935-937.

Smouse, Evan Paul; Martin Alva Hamilton. Estimating proportionate changes in rates. *Am J Epidemiol* 1983; 117:235-43.

Zeighami EA, Morris MD. The measurement and interpretation of proportionate mortality. *Am J Epidemiol* 1983; 117:90-7.

Apéndice sobre promedios ponderados

Como la epidemiología estudia poblaciones, y las poblaciones contienen subgrupos, los promedios ponderados figuran en forma importante en epidemiología. Casi cualquier medida basada en la población puede ser considerada un promedio ponderado del valor de esa medida obtenida de los subgrupos comprendidos en la población. Los promedios ponderados se usan para estandarizar o ajustar las medidas crudas para hacerlas más comparables entre poblaciones con diferente distribución proporcional de los subgrupos. Tanto el concepto como las matemáticas son fundamentales.

Un promedio ponderado es como un promedio común salvo que los componentes que están siendo promediados pueden tener más o menos influencia (peso) sobre el promedio resultante. Por ejemplo, supongamos que medimos la presión arterial sistólica en 10 ocasiones y obtenemos los siguientes valores (mmHg): 95, 100, 100, 105, 105, 105, 110, 110, 115, 120. Si queremos la presión arterial sistólica promedio, simplemente sumamos las mediciones individuales y dividimos por el número de lecturas: $1,065/10 = 106.5$ mmHg. Dado que algunos valores ocurren más de una vez, deberíamos obtener el mismo resultado usando un promedio ponderado:

Número de lecturas	Valor	Suma ponderada
1	95	95
2	100	200
3	105	315
2	110	220
1	115	115
1	120	120
10		1,065

$$\text{Promedio} = 1,065 / 10 = 106.5 \text{ mmHg.}$$

Un negocio pequeño puede usar una disposición como esta para calcular el precio promedio pagado pro algún bien durante algún período de tiempo. En esa situación, la primer columna podría mostrar el número de bolsas adquiridas, la segunda columna el precio por bolsa, y la tercer columna el monto total pago en dólares.

Con un poco de generalización (para permitir que el “número de lecturas” sea un número fraccionado), tenemos un procedimiento para crear un promedio ponderado. Ejemplos conocidos son los puntajes de cursos promedios ponderados por las horas de curso, costo promedio por acción de un conjunto de acciones compradas en múltiples oportunidades, y precio promedio por litro de nafta comprado durante las vacaciones.

Matemáticamente, un promedio ponderado es una combinación lineal donde los coeficientes (p_i) son proporciones cuya suma es igual a 1.0. Varias fórmulas equivalentes son:

$$\frac{w_1a_1 + w_2a_2 + \dots + w_na_n}{w_1 + w_2 + \dots + w_n} = \frac{w_1a_1 + w_2a_2 + \dots + w_na_n}{W}$$

$$= \frac{w_1a_1}{W} + \frac{w_2a_2}{W} + \dots + \frac{w_na_n}{W} = \sum \left(\frac{w_ia_i}{w_i} \right)$$

$$= p_1a_1 + p_2a_2 + \dots + p_na_n = \sum(p_ia_i)$$

donde $W = w_1 + w_2 + \dots + w_n$ and $p_1 + p_2 + \dots + p_n = 1$

Para el ejemplo del precio de la gasolina, los w_i representan las cantidades adquiridas en cada parada y las a_i representan el precio de cada compra.

Apéndice sobre exponentes y logaritmos

(Adaptado de Defares JG y Sneddon IN. *An introduction to the mathematics of medicine and biology*. The Netherlands, North-Holland, 1960[Una Introducción a las matemáticas de la medicina y la biología])

Algunos hechos simples:

$$2^2 = 2 \times 2 = 4$$

$$2^3 = 2 \times 2 \times 2 = 8$$

$$\text{raíz cuadrada de } 4 = 2$$

$$\text{raíz cúbica de } 8 = 2$$

Exponentes:

b^x significa b elevado a la $x^{\text{ésima}}$ potencia; x es denominado el exponente.

Si x es 2, entonces $b^x = b^2 = b \times b$. Si x es 3, entonces $b^x = b^3 = b \times b \times b$. A partir de esto podemos razonar que:

$$1) \quad b^m \times b^n \text{ debe ser igual a } b^{(m+n)}$$

(El producto de un número elevado a la $n^{\text{ésima}}$ potencia multiplicado por el mismo número a la $m^{\text{ésima}}$ potencia es igual a ese número elevado a la suma de potencias.)

$$2) \quad b^m/b^n \text{ debe ser igual a } b^{(m-n)}$$

(El cociente de un número elevado a la $n^{\text{ésima}}$ potencia dividido por el mismo número elevado a la $m^{\text{ésima}}$ potencia es igual a ese número elevado a la diferencia de potencias (la potencia del numerador menos la potencia del denominador).

$$3) \quad (b^m)^n \text{ debe ser igual a } b^{(m \times n)}$$

(La $n^{\text{ésima}}$ potencia del número elevado a la $m^{\text{ésima}}$ potencia es igual al número elevado a la $(m \times n)^{\text{ésima}}$ potencia.

Para exponentes que no son números enteros, definimos b^x para conservar las tres reglas antes mencionadas. De manera que $b^0=1$ and $b^{-x} = 1 / b^x$.

Cuando el número de base (b en los ejemplos anteriores) es e, un número trascendente que es aproximadamente 2.7183, podemos escribir e^x o (para la conveniencia tipográfica) $\exp(x)$. e y los logaritmos neperianos tienen propiedades especiales que los hacen recomendables para uso en matemáticas y estadísticas.

Logaritmos:

Si para un número b (mayor de 1.0), es posible encontrar un número x tal que:

$$y = b^x$$

podemos decir que x es el logaritmo de y en base b:

$$x = \log_b y$$

El logaritmo es la inversa de la exponenciación, de manera que si $y=b^x$:

$$\log_b(y) = \log_b(b^x) = x \quad y$$

$$b^x = b^{(\log_b y)} = y$$

Para mantener consistencia con las reglas para los exponentes, antes mencionados, vemos que:

$$1) \quad \log_b(xy) = \log_b x + \log_b y$$

(el logaritmo de un producto es la suma de los logaritmos)

$$2) \quad \log_b(x/y) = \log_b x - \log_b y$$

(el logaritmo de un cociente es igual al logaritmo del numerador menos el logaritmo del denominador), y

$$3) \quad \log_b(x^n) = n \log_b x$$

Los logaritmos son definidos de manera que las reglas se generalizan al caso de los exponentes negativos y fracciones. El logaritmo de un número negativo, sin embargo, aún no ha sido definido.

La base b debe ser un número positivo mayor de 1.0, y es habitualmente 10 (para los “logaritmos comunes”) o e (para los logaritmos naturales o neperianos). Estos últimos son los que más se ven en

matemáticas, estadísticas, y epidemiología. La expresión $\ln(x)$ o simplemente $\log(x)$ se usan a menudo cuando se comprenden los logaritmos neperianos.

Señalemos que para la base e ($= 2.7183$), $\exp(x)$ (a) debe ser mayor de cero, (b) será igual a 1 cuando $x=0$, and (c) aumentará muy rápidamente para una x grande. En contraste, $\ln(x)$ (a) será negativo para $x<1$, (b) será igual 0 cuando $x=1$, y (c) será positivo para $x>1$. De manera que si x es una razón positiva cuyo valor nulo es 1.0, $\ln(x)$ representará una transformación de x con el valor nulo en 0 y los otros valores se distribuirán en forma simétrica alrededor de él. Estas propiedades de los logaritmos son útiles para transformar las distribuciones de las variables y para el análisis de razones.