

III Congreso de la A.V.P.A.
Epidemiología para veterinarios

Dr. Ignacio de Blas
24 de noviembre de 2005

Asociación de Veterinarios de Porcino de Aragón (AVPA)

Introducción

A lo largo del taller ilustraremos con ejemplos prácticos cuestiones relativas a la evaluación de procesos patológicos que pudieran aparecer en las explotaciones porcinas entre los que se incluirán:

1. Medición de la enfermedad en explotaciones de porcino: detección de enfermedad versus estimación de prevalencias, cálculo de tamaño de muestra, influencia de la fiabilidad diagnóstica, precisión de la medición y comparación de tasas de enfermedad
2. Factores de riesgo: diseño de estudios observacionales, cálculo e interpretación del Odds Ratio y del Riesgo Relativo, evaluación económica de su impacto en la producción porcina.

La metodología de este curso se basará en una pequeña introducción teórica que permita a los asistentes tener un marco conceptual que permita una mejor comprensión de los ejemplos planteados y a continuación se expondrán algunos ejemplos prácticos y aplicados al trabajo de campo de la clínica porcina. Posteriormente se plantearán otros casos para que los asistentes trabajen con ellos e incluso si desean plantear algún supuesto o duda extraído de su actividad profesional.

Debido a las limitaciones de tiempo se ofrece en forma de anexos el desarrollo completo de los conceptos teóricos utilizados durante el seminario.

Para la realización de los ejemplos se utilizarán las aplicaciones informáticas Win Episcopo 2.0 y WinEpi Tasas 2.0 (disponibles gratuitamente en la web <http://winepi.unizar.es>), y de forma excepcional y en función de las cuestiones planteadas por los participantes se podría utilizar Microsoft Excel.

Anexo 1:

Recogida de muestras

Conceptos básicos.

Cuando investigamos epidemiológicamente una enfermedad en el contexto de una población, podemos trabajar sobre el total de la misma, lo que encarece y dificulta el estudio, o sobre una parte de la misma que denominamos **muestra**, que simplifica y abarata el proceso, a la par que reduce su duración temporal. Sin embargo para garantizar los resultados obtenidos la muestra debe cumplir una condición indispensable: debe ser representativa de la población para que los resultados obtenidos se puedan extrapolar a la misma.

La investigación de una variable a partir de una muestra tiene la desventaja de perder parte de la información común a la población y que por tanto el valor obtenido no sea el real, sino una **estimación** de ese valor real de la población.

Para comprender plenamente los conceptos descritos en el presente capítulo es preciso describir una serie de conceptos fundamentales:

- **Población diana:** aquella población de la que se desea obtener información
- **Población a estudiar u origen:** aquella población de la que realmente se obtiene la información. En ocasiones coincide con la anterior, pero si no es representativa de la primera los resultados obtenidos no será extrapolables a la población diana.
- **Unidad de muestreo:** es el elemento básico del procedimiento de muestreo. Puede ser el animal, el rebaño, la explotación,... dependiendo del planteamiento del estudio.
- **Muestra:** es el conjunto de unidades seleccionadas a partir de la población a estudiar y sobre las que realmente se realiza la investigación epidemiológica.
- **Estrato:** es la colección de unidades de muestreo agrupadas por una característica común, por ejemplo el sexo, la raza o la especie.
- **Fracción de muestreo:** es la relación entre el tamaño de la población a estudiar y el tamaño de la muestra

Para hacer **representativa** una muestra tomada de una población requiere, por un lado, que posea un *tamaño adecuado* y por otro, que el *método de selección* de la misma sea el apropiado, de manera que todas las subpoblaciones existentes en la misma estén representadas de forma correcta. A pesar de ello existen otra serie de condicionantes que hay que considerar como son el coste de la obtención de datos, el tiempo necesario y los riesgos intrínsecos que implican la recogida de los datos tanto para el observador como para el observado.

Por esa razón es preciso que la representatividad de la muestra esté marcada por dos condiciones básicas:

- **Aleatoriedad:** todas las unidades de la población tienen la misma probabilidad de formar parte de la muestra
- **Homogeneidad:** la muestra debe conservar las proporciones que se guardan en la población en lo referente a las características estudiadas.

Métodos de muestreo.

A la hora de realizar la selección de las unidades que formarán parte de la muestra podemos recurrir a distintos métodos, que se pueden clasificar en dos grandes grupos: probabilísticos y no probabilísticos.

- **Probabilísticos o aleatorios:** Todas las unidades de la población tienen la misma probabilidad de formar parte de la muestra, siendo el azar el que determina los que forman parte de la muestra, de tal forma que la realización de estos métodos precisan de la utilización de sistemas de *lotería* o de *números aleatorios*. Según las características de la población investigada y de los objetivos de la investigación pueden ser de diferentes:
 - **Simple:** Es necesario tener identificados todas las unidades, de forma que se seleccionan individualmente.
 - **Sistemático:** Está indicado cuando no se conoce la identidad de las unidades, de forma que se establece entre ellos un orden (por ejemplo, orden en el que irán saliendo los animales al campo), y se seleccionan dos números aleatorios. El primer número definirá el orden establecido del primer individuo que integrará la muestra, y posteriormente se selecciona un segundo número correspondiente al intervalo de muestreo, que habitualmente se utiliza el valor obtenido al dividir el tamaño de la población por el tamaño de muestra necesario.
 - **Estratificado:** A veces interesa analizar la población en función de la existencia de diferencias en un determinado carácter, por ejemplo el sexo. En estos casos dividiremos la población total en grupos y de cada grupo se toma una muestra por métodos *simples* o *sistemáticos*. Habitualmente el número de unidades a muestrear en cada grupo será proporcional al tamaño de ese grupo respecto al total de la población (si un sexo supone en la población el 75% de los animales, en la muestra ese sexo debe estar representado en un 75%), aunque la forma más correcta de proceder es la determinación del tamaño de muestra necesario para cada estrato.
 - **Conglomerados (Cluster):** En otros casos se conoce el número de grupos que hay dentro de la población pero no el número de unidades de cada uno de esos grupos, de forma que se seleccionará una muestra por los métodos *simple* o *sistemático* de los grupos y se tomarán la totalidad de las unidades de cada grupo seleccionado. Es el caso en que se conoce el número de granjas de una región, pero no el número de animales de cada una, la forma correcta de proceder consistiría en la selección de parte de las granjas y la investigación de la totalidad de los animales de cada una de esas granjas.
 - **Por etapas (Multiestadio):** Consistente en un sistema mixto de los métodos *conglomerados* y *estratificado*. Se trata de realizar el muestreo en dos fases: una primera fase entre grupos y una segunda fase entre unidades dentro de esos grupos seleccionados en la primera fase previamente. En el ejemplo anterior se trata de seleccionar parte de las granjas, y posteriormente en cada una de ellas se seleccionan parte de los animales: en ambos casos mediante los métodos *simple* o *sistemático*.
- **No probabilístico:** Agrupa todos aquellos métodos en que no todas las unidades tienen la misma probabilidad de formar parte de la muestra, siendo el investigador (veterinario) o el investigado (ganadero) el que decide las que forman parte de la muestra. En estos casos, corremos el riesgo de que la muestra no sea representativa de la población, y por tanto los datos obtenidos no son susceptibles de ser analizados por los métodos estadísticos habituales.

En función de quien decide las unidades que integran la muestra, los tipos de muestreo no probabilístico serán:

- **Con voluntarios:** El investigado decide voluntariamente qué unidades forman parte de la muestra según sus intereses.
- **De conveniencia:** El investigador decide qué unidades forman parte de la muestra, en función de que posean o no algún carácter que desea analizar.

Selección del tamaño de muestra.

A la hora de fijar el tamaño muestral es importante valorar que el número de unidades muestreadas va a condicionar el tiempo, el coste y validez del trabajo. Por lo tanto la utilización de un método de selección adecuado no garantiza que la muestra sea representativa de la población, ya que la muestra también debe cumplir una segunda condición: que el tamaño sea el adecuado para el tipo de investigación que se desea realizar. Esto supone que el tamaño de la muestra depende de los objetivos de la investigación, ya que no será necesario el mismo número de animales para detectar la presencia de una enfermedad en una población (se trata de detectar al menos un animal positivo), que para determinar la prevalencia (debemos detectar una proporción de animales enfermos en la muestra igual a la que existe en el total de la población).

Para poblaciones en el que el número de unidades es superior a 1.000 se puede generalizar a efectos de cálculo previo y orientativo, y utilizar la siguiente fórmula:

$$n = 2 \cdot \sqrt{N}$$

Sin embargo en condiciones normales para estimar el tamaño de la muestra a seleccionar en una población con unos objetivos determinados es necesario establecer una medida de la dispersión de los datos, manifestada por la varianza (y por extensión la desviación estándar), así como fijar el error aceptado y el nivel de confianza deseado (probabilidad de obtener una respuesta correcta).

El error aceptado y el nivel de confianza son establecidos arbitrariamente por el investigador, mientras que la desviación estándar se obtiene a través de la literatura.

- **Estimar una media**

Para conocer la media de una variable continua en una población (el caso del peso o de la temperatura corporal) el tamaño de la muestra a seleccionar dependerá de la **desviación estándar de la media esperada (SD)**.

En este caso, la fórmula a utilizar es:

$$n = \left(\frac{Z_{\alpha} \cdot SD}{L} \right)^2$$

donde:

n :tamaño de la muestra.

Z α : valor de la distribución normal (t de Student) para un nivel de confianza deseado. Generalmente se utiliza un nivel de confianza del 95%, cuyo valor de t_{α} es 1.96.

L: error aceptado o precisión, margen admitido a ambos lados de la media esperada.

SD: desviación estándar correspondiente a la media esperada de enfermedad. Este valor suele ser desconocido por lo que debe utilizarse un valor esperado o aproximado. El cálculo de ese valor aproximado se realiza definiendo inicialmente los valores máximo y mínimo que se espera encontrar, entre esos dos valores límites se encuentra un rango que con un 95% de probabilidad incluye todos los valores de $X \pm 2 SD$, lo que significa que dividiendo ese rango estimado por 4 obtendremos un valor aproximado de la SD.

- **Estimar una proporción**

Cuando interesa determinar una proporción en una población (por ejemplo la prevalencia de la enfermedad) el tamaño de la muestra a tomar depende, como en el caso anterior, del error aceptado, del nivel de confianza y de la desviación estándar.

La fórmula de la desviación estándar (SD) para variables dicotómicas (presencia o ausencia de enfermedad) es la siguiente, siendo P la proporción (en tanto por uno) del carácter estudiado:

$$SD = \sqrt{P \cdot (1 - P)}$$

En el peor de los casos se puede escoger la situación que nos obligue a seleccionar un tamaño mayor de muestra, que es que corresponde a una probabilidad del 50%.

Para estos casos se utiliza la fórmula:

$$n = \left(\frac{Z_{\alpha} \cdot SD}{L} \right)^2$$

donde:

n: tamaño de la muestra.

Z α : valor de la distribución normal (t de Student) para el nivel de confianza dado.

L: error aceptado o precisión, generalmente se utiliza el 5%

SD: desviación estándar correspondiente a la proporción esperada (**P**)

Así, trabajar con una prevalencia esperada del 50%, un error aceptado del 5% y un nivel de confianza del 95%, indica que esperamos que, de cada 100 mediciones que se realicen, en 95 la prevalencia tenga un valor comprendido en el intervalo $50 \pm 5\%$ (45% - 55%).

- **Ajuste del tamaño muestral**

En los dos casos anteriormente descritos el tamaño de la muestra obtenido mediante esa fórmula sólo es válido cuando la fracción muestral es inferior al 10%, lo que generalmente ocurre con poblaciones mayores de 1.000 unidades. Sin embargo, cuando el tamaño de la muestra a tomar es mayor del 10% de la población, lo que suele ocurrir con poblaciones de menos de 1.000 unidades, es necesario realizar una corrección de la desviación estándar que permitirá estimar el tamaño real de la muestra que se debe tomar, de forma que el tamaño final de la muestra será:

$$n' = \frac{n}{1 + \frac{n}{N}}$$

donde:

n': tamaño de muestra ajustado

n: tamaño de la muestra (n)

N: tamaño de la población

- **Cálculo del error aceptado**

A veces el investigador tiene limitado el número de muestras, por ejemplo porque la capacidad del laboratorio es limitada. Entonces el tamaño de la muestra está definido a priori y lo que interesa calcular es el error aceptado para establecer un criterio objetivo que permita tomar la decisión de seguir adelante con la investigación o no, de tal forma que si el error aceptado (L) es demasiado grande en opinión del investigador, es inútil continuar con el control.

En este caso las fórmulas son:

$$L = t \cdot SE = t \cdot \frac{SD}{\sqrt{n}} \quad \text{o} \quad L = t \cdot SD \cdot \sqrt{\frac{N-n}{N \cdot n}}$$

La primera de ellas para poblaciones consideradas infinitas (mayores de 1.000 unidades) y la segunda para poblaciones finitas (menores de 1.000 unidades).

• Detección de enfermedad

A veces es necesario determinar si una población está infectada, de tal forma que consideramos una población como positivas cuando al menos **d** animales están enfermos. Este parámetro **d** lo fija el investigador y es el número de animales que espera encontrar enfermos. En enfermedades graves o cuando interesa obtener una máxima seguridad en el resultado, **d** se lleva al mínimo valor posible, es decir, se establece en 1.

El cálculo se realiza mediante la siguiente fórmula:

$$n = \left(1 - (1 - NC)^{\frac{1}{d}} \right) \cdot \left(N - \frac{(d-1)}{2} \right)$$

donde:

n : tamaño de la muestra requerido

N: tamaño de la población

d : número de animales enfermos que esperamos que existan en la población

NC : nivel de confianza expresado en tanto por uno (0.95)

A partir de esta fórmula y conocido un tamaño de muestra determinado (n) con el que se desea trabajar, es posible estimar el **máximo número de animales positivos (D)** que pueden aparecer en una población donde todas las **n** muestras recogidas han resultado negativas.

$$D = \left(1 - (1 - NC)^{\frac{1}{n}} \right) \cdot \left(N - \frac{(n-1)}{2} \right)$$

donde: **D**: máximo número de animales enfermos que pueden quedar en la población

Es importante destacar que asumimos que la prueba diagnóstica tiene una validez del 100%

Errores de muestreo.

La selección de una muestra a partir de una población que interesa estudiar y de la que se desea obtener una determinada información puede generar errores en el proceso del muestreo y que pueden suponer una importante desviación de los resultados obtenidos respecto de la realidad.

En el estudio de una población a partir de una muestra pueden cometerse tres tipos de errores:

- **Error sistemático:** Está relacionado con el método de selección de la muestra al seleccionar deliberadamente algunas unidades o sustituir otras de forma incorrecta. *Este tipo de errores no disminuyen al aumentar el tamaño de la muestra pero sí al utilizar un método aleatorio*
- **Error aleatorio:** Error debido generalmente al azar que se relaciona directamente con el tamaño de la muestra tomada. *Este error no disminuye utilizando métodos aleatorios pero sí al incrementar el tamaño de la muestra*
- **Error de información:** En ocasiones existe una tercera posibilidad de error que no está relacionada directamente con la muestra, sino que depende de la forma en que se recoge la información una vez que la muestra ya ha sido seleccionada.

Anexo 2:

Medición de la enfermedad

Conceptos previos.

Frecuentemente se busca la comparación entre poblaciones animales, lo que significa que debe realizarse la medición de la enfermedad en términos relativos, es decir, expresarla como una relación entre un numerador y un denominador, donde el numerador corresponde a los individuos que sufren el evento y el denominador a los que son susceptibles de adquirir dicho evento, y que generalmente denominamos como **población en riesgo**, que se define como el número de individuos que biológicamente son capaces de experimentar la enfermedad.

Las relaciones entre numerador y denominador pueden ser de dos tipos:

- **Proporción:** cuando el numerador está incluido en el denominador. Su valor se acota entre 0 y 1 (100 si se expresa en porcentaje). Se expresa matemáticamente como $a/(a + b)$
- **Tasa (ratio):** cuando el numerador no está incluido en el denominador, en este caso sus valores posibles van de 0 a infinito. Se expresa matemáticamente como a/c .

Los eventos que afectan a una población pueden medirse a nivel estático (generalmente referido a un momento concreto del tiempo) o a nivel dinámico (durante un periodo de tiempo definido). En ambos casos existe un componente externo del tiempo que no afecta directamente sobre la medición.

En los estudios de enfermedad, la medición de la frecuencia con que esta se presenta se realiza mediante la **prevalencia** que corresponde al término estático y la **incidencia** que corresponde al término dinámico. Además de estos parámetros de medición de la enfermedad, existen otros que indican básicamente la importancia de la enfermedad en una población y que son la **morbilidad**, la **mortalidad** y la **letalidad**.

Morbilidad, mortalidad y letalidad.

- **Morbilidad**

Es la probabilidad de que los individuos de una población adquieran la enfermedad en un momento o periodo del tiempo (según se estudie a nivel dinámico o estático, en el primer caso será equivalente a la incidencia y en el segundo a la prevalencia de enfermedad como veremos en el siguiente apartado).

Se calcula como el cociente entre el número de animales con la enfermedad en un periodo o momento del tiempo definidos y el número de animales en riesgo de adquirir la enfermedad durante ese mismo periodo o momento del tiempo.

- **Mortalidad**

Corresponde a la probabilidad de que los individuos de la población mueran como consecuencia de una determinada enfermedad en un periodo de tiempo.

Se calcula como la relación entre el número de muertes debidas a la enfermedad en el periodo de tiempo en estudio y la población en riesgo de morir por esa enfermedad en ese mismo periodo de tiempo.

En el caso de la mortalidad es posible estimar otro parámetro que se refiere al impacto de la enfermedad, denominado **índice de mortalidad proporcional**, que indica la importancia de la muerte como consecuencia de la enfermedad con relación al total de muertes acaecidas en la población. Se calcula como el número de muertes debidas a la enfermedad en un periodo de tiempo dado dividido por el número de muertes totales en la población en ese mismo periodo.

- **Letalidad**

Corresponde a la probabilidad de morir entre aquellos individuos que presentan la enfermedad en un periodo determinado de tiempo. Si bien no se trata de una tasa sino de una proporción, en la bibliografía anglosajona aparece denominada *Case Fatality Rate*.

Se calcula como la relación entre el número de muertos debido a la enfermedad en un periodo de tiempo y el número de animales que presentan la enfermedad en ese periodo de tiempo.

Prevalencia, incidencia acumulada y tasa de incidencia.

Aunque los parámetros de medida de la enfermedad previamente definidos son importantes, especialmente cuando interesa conocer el impacto de la misma en la población, los parámetros de medición más frecuentemente utilizados son la incidencia y la prevalencia, ya que permiten conocer el impacto de la enfermedad en la población y estimar su frecuencia y distribución en el tiempo.

- **Prevalencia**

Corresponde a la probabilidad de que un individuo seleccionado al azar en la población sufra la enfermedad, y por lo tanto, es la proporción de enfermos presentes en una población durante un periodo o momento del tiempo, en el primer caso se habla de periodo de prevalencia y en el segundo de prevalencia puntual, si bien esta última es la que más frecuentemente se utiliza y supone generalmente hablar de una situación momentánea en el tiempo.

La **prevalencia puntual** se caracteriza por medir todos los casos de enfermedad existentes sin diferenciar los casos antiguos (comenzaron antes del momento del estudio) de los nuevos (los que comenzaron en ese momento). Se expresa con relación al número de animales que existen en la población y que están en riesgo de presentar la enfermedad en el momento del estudio (incluidos los que presentan la enfermedad).

Su cálculo equivale al cociente entre el número de animales enfermos (nuevos en ese momento o periodo de tiempo más los que ya existentes) y el número total de animales en riesgo de presentarla, incluidos los enfermos. Los valores de la prevalencia varían entre 0 y 1.

El **periodo de prevalencia** corresponde a los casos que ya existían en el momento en que comienza el periodo de estudio más los nuevos casos de enfermedad que aparecen durante el periodo de estudio. Se trata de un parámetro que se utiliza en escasas ocasiones, correspondiendo su cálculo al número de casos de enfermedad al comienzo del periodo de estudio (antiguos) más todos los nuevos casos que se presentan en ese periodo, dividido por el total de individuos en riesgo de sufrir la enfermedad (incluidos los que ya la tienen al comienzo del periodo de estudio).

- **Incidencia**

Este parámetro describe la velocidad con que la enfermedad se distribuye en la población, es decir, indica el flujo de individuos sanos a enfermos, y equivale a la medida de los nuevos casos de enfermedad que se presentan en una población durante un periodo determinado de tiempo. Determinar la incidencia implica realizar siempre dos mediciones en la población a estudiar, una medición al principio del estudio, en la que se determinarán los individuos sanos, y una segunda durante el periodo de estudio o al final del mismo, que determinará cuantos animales han adquirido la enfermedad (nuevos casos).

En función de las características de la población a estudiar y de la enfermedad, existen dos parámetros diferentes de incidencia: la incidencia acumulada (*Cummulative Incidence*) y la tasa de incidencia (*Incidence Rate*). La primera de ellas es recomendable en estudios de cortos periodos de tiempo y/o cuando las poblaciones son estables, mientras que la segunda

se aplica a estudios de largos periodos de tiempo y/o con poblaciones poco estables (variables en número a lo largo del periodo de estudio).

- ***Incidencia acumulada (CI)***

Corresponde a la proporción de animales sanos al comienzo del periodo de estudio y que enferman a lo largo de dicho periodo. Se calcula como la relación existente entre el número de animales que adquieren la enfermedad en un periodo de tiempo y el número de animales sanos en riesgo de adquirir la enfermedad al comienzo de ese periodo de estudio. Es una medida dinámica de la enfermedad y su valor varía entre 0 y 1.

En ocasiones se puede producir la pérdida de parte de la población en riesgo durante ese periodo de estudio, generalmente como consecuencia de la muerte de parte de la población, en tal caso, la población sana en riesgo que constituye el denominador, se estima como la suma de la población sana en riesgo al principio y la población sana en riesgo al final, dividido por dos.

- ***Tasa de incidencia (IR)***

La tasa de incidencia es una relación entre los nuevos casos de enfermedad en un periodo de tiempo dado y la población susceptible de adquirirla en ese periodo de tiempo pero con un componente de temporal que implica considerar que no todos los individuos están en riesgo de adquirir la enfermedad durante el mismo tiempo.

El estudio de una enfermedad a lo largo de un periodo de tiempo supone considerar que no todos los animales enfermarán a la vez dentro de ese periodo, por lo que no todos los animales contribuirán de la misma forma como individuos en riesgo (al denominador). Además la población en riesgo puede variar constantemente a lo largo del periodo de estudio, de forma que la solución indicada consistirá en dividir el periodo de tiempo en subperiodos considerando para cada individuo de la población el número de subperiodos dentro del periodo global en que está en riesgo, de forma que la población en riesgo total será la suma de los subperiodos en riesgo con que contribuya cada individuo de la población (es decir, multiplicar cada individuo por el número de subperiodos en riesgo). Esos subperiodos del tiempo en los que los animales están en riesgo constituyen lo que se denomina ***unidades-tiempo en riesgo***.

Por tanto, la tasa de incidencia equivale a número de nuevos casos de enfermedad que se presentan en una población dividido por la suma de los subperiodos en riesgo de los individuos de una población durante un determinado periodo de tiempo. Se trata de una tasa, es decir, el numerador no está incluido en el denominador, por lo que su valor varía entre 0 e infinito.

Relaciones entre los distintos parámetros.

- ***Prevalencia y tasa de incidencia***

La prevalencia de enfermedad está condicionada por la tasa de incidencia y por la duración de la enfermedad, de forma que cuando la enfermedad tiene duración constante, se puede establecer la siguiente relación:

$$P / (1-P) = IR \cdot D$$

donde:

P: prevalencia

1-P: proporción de animales sanos

IR: tasa de incidencia

D: duración de la enfermedad

Esta relación permite estimar la duración de una enfermedad cuando conocemos su prevalencia y su tasa de incidencia. Si analizamos la fórmula se pone de manifiesto que un

descenso en la tasa de incidencia o en la duración de la enfermedad implica un descenso en la prevalencia de la misma y por el contrario, un aumento en alguna de las dos implica un incremento de la prevalencia de enfermedad.

- **Incidencia acumulada y tasa de incidencia**

Como en el caso anterior existe una relación entre ambos parámetros, relación que viene definida por la ecuación:

$$CI = 1 - e^{-IR \cdot t}$$

donde:

CI: incidencia acumulada

IR: tasa de incidencia

t: número de periodos de tiempo

- **Mortalidad, letalidad y tasa de incidencia**

La mortalidad puede estimarse en función de la tasa de incidencia y de la letalidad cuando el riesgo de muerte se produce durante largos periodos de tiempo, utilizando la fórmula:

$$Mt = 1 / (1/IR + 1/L)$$

donde:

Mt: mortalidad

IR: tasa de incidencia

L: letalidad

Importancia de los intervalos de confianza.

La medición de la enfermedad en una población suele realizarse sobre una muestra representativa de la misma, de manera que el resultado obtenido sobre ésta última se pueda extrapolar a la población origen. Teniendo en cuenta que el tamaño de la muestra se determina mediante cálculos de probabilidad, hay que introducir unos márgenes de error que pueden hacer que el resultado obtenido sobre la muestra no corresponda con el real.

Para evitar estas situaciones, y para conocer la calidad del valor de la medida realizada sobre la muestra se calculará su **Intervalo de Confianza** que indica que el valor real del parámetro en la población de la que se extrajo la muestra se encuentra acotado entre los límites superior e inferior de dicho intervalo con una probabilidad definida por el nivel de confianza establecido (generalmente el 95%).

Así, a modo de ejemplo, para la prevalencia de enfermedad obtenida en una muestra obtenida a partir de una población a estudiar, los límites superior e inferior del intervalo de confianza de dicha medición se calculan como:

$$\left(P - Z_{\alpha} \cdot \sqrt{\frac{P \cdot (1-P)}{n}}, \quad P + Z_{\alpha} \cdot \sqrt{\frac{P \cdot (1-P)}{n}} \right)$$

donde:

P: prevalencia obtenida en la muestra

n: tamaño de la muestra

Z α : valor de la distribución normal (t de Student) para un nivel de confianza determinado (para el 95% de nivel de confianza el valor de Z es 1.96).

Estandarización de tasas.

En ocasiones al estudiar los resultados obtenidos en dos poblaciones se observa que de forma global parece que existen diferencias en la presentación de la enfermedad en ambas poblaciones y sería preciso probar estadísticamente la existencia de diferencias significativas.

Sin embargo al estratificar dichos resultados en función de alguna característica de la población (edad, sexo, especie,...) y los comparamos podemos encontrarnos con resultados sorprendentes y contradictorios. Así pues esta circunstancia nos hace plantearnos la necesidad de estandarizar las tasas calculadas para poder comparar el nivel de enfermedad en dos poblaciones que presenten distinta distribución.

Para explicar con mayor claridad este aspecto lo ilustraremos con un sencillo ejemplo. Supongamos dos poblaciones A y B, en las que los animales están distribuidos por edades según el siguiente esquema, y a partir de las que hemos recogido unas muestras para estimar los nuevos casos de enfermedad (incidencia)

Población	Estrato	Tamaño	Muestra	Casos	Incidencia (%)
A	Machos	258	144	41	28,47
	Hembras	760	227	8	3,52
	Total	1.018	371	49	13,21
B	Machos	48	42	12	28,57
	Hembras	1.462	265	9	3,40
	Total	1.510	307	21	6,84

Como se puede observar a priori existe una diferencia en las tasas globales de enfermedad entre ambas poblaciones (13,21 y 6,84% respectivamente) lo que parece indicar que el proceso patológico se presenta con distinta intensidad en ambas poblaciones lo que podría dar lugar a la elaboración de una hipótesis de trabajo para la búsqueda de factores de riesgo o protección entre ambas poblaciones, así como para probar estadísticamente la existencia de diferencias significativas.

Sin embargo al comparar los resultados por estratos vemos como en ambos casos la incidencia es similar en los machos de ambas poblaciones (alrededor del 28%) y ocurre algo parecido en las hembras, pero en unas magnitudes muy inferiores (cerca del 3%).

Así pues esta circunstancia nos hace plantearnos la necesidad de estandarizar las tasas calculadas para poder comparar el nivel de enfermedad en ambas poblaciones ya que como observamos la distribución de los animales en ambas poblaciones en función del criterio seguido, es muy diferente lo que influye de forma clave en el cálculo de la incidencia global en cada población y nos impide comparar adecuadamente los resultados obtenidos.

En poblaciones animales se recomienda la estandarización indirecta en la que se estandarizan las tasas de cada estrato utilizando como referencia una de las poblaciones. Para ello seguiremos los siguientes pasos:

- **Cálculo del número de casos esperados (CE) para cada estrato:** multiplicaremos la incidencia media estimada en cada estrato por el número total de animales del estrato y obtendremos los siguientes resultados:

Población	Estrato	Tamaño	Incidencia (%)	CE	Inc. Estimada (%)
A	Machos	258	28,47	72	28,47
	Hembras	760	3,52	27	3,52
	Total	1.018		99	9,72
B	Machos	48	28,57	14	28,57
	Hembras	1.462	3,40	50	3,40
	Total	1.510		64	4,24

Hay que destacar que tras realizar esta operación y ponderar la contribución de los casos de cada estrato al global de la población, teniendo en cuenta la estructura de la población nos encontramos con que en ambos casos estábamos sobrevalorando la incidencia.

- **Estandarización de tasas de la población B con respecto a la distribución de la población A:** multiplicaremos las incidencias estimadas en cada estrato de la población B por el número de individuos del estrato correspondiente en la población A, y luego se recalculan las incidencias estandarizadas utilizando como denominador la distribución de la población A obteniéndose los siguientes resultados:

Población	Estrato	Tamaño	Inc. Estimada (%)	CE	Inc. Estand. (%)
A	Machos	258	28,47	72	28,47
	Hembras	760	3,52	27	3,52
	Total	1.018		99	9,72
B	Machos	48	28,57	74	28,57
	Hembras	1.462	3,40	26	3,40
	Total	1.510		100	9,82

Así pues al comparar las incidencias globales estandarizadas en ambas poblaciones observamos que la diferencia aparente que encontrábamos al principio del estudio (13,21% y 6,84%) equivale en realidad a unas incidencias muy similares (9,72% y 9,82% respectivamente) lo que pone de manifiesto la gran influencia de la estructura de las poblaciones sobre la medición de la enfermedad.

Así pues podemos concluir que:

1. el cálculo de la tasa global de enfermedad en una población está directamente condicionada por la existencia de diferencias en la presentación de la misma en función de diversas características de la población.
2. la comparación de las tasas de una enfermedad en distintas poblaciones precisa de la estandarización de las mismas ya que las diferencias en la distribución de la población en estratos influye de forma determinante en los resultados obtenidos.

Anexo 3:

Estudios observacionales

Tipos de estudios.

La evaluación de la asociación entre los determinantes (factores de riesgo) y la presentación de enfermedad en una población supone estimar el **grado de riesgo de presentación de la enfermedad que supone el estar expuesto a esos determinantes**, es decir la variación absoluta o relativa, en el riesgo de presentar la enfermedad debido a la exposición al factor.

Los estudios epidemiológicos experimentales representan el mecanismo ideal de evaluación del papel de un determinante en una enfermedad o de la eficacia de un tratamiento. El problema reside en la necesidad de unas condiciones estrictamente controladas que escapan a las posibilidades del veterinario clínico. Sin embargo, el veterinario clínico dispone de otro tipo de estudios epidemiológicos, los estudios observacionales, que realizan esa misma evaluación en condiciones naturales (en el propio campo), lo que los hace más accesibles y le permiten estudiar esas asociaciones en condiciones naturales y no bajo condiciones controladas como ocurre en el caso de los experimentales.

Los estudios epidemiológicos observacionales pueden ser de tres tipos dependiendo del momento en que se llevan a cabo y de la información disponible:

- **Transversales:** practicados en un *momento* puntual del tiempo. Se trabaja con un bloque de información único acerca de la población animal, la obtenida en ese momento del tiempo sin considerar circunstancias previas.
- **Longitudinales:** se desarrollan durante un *período* definido de tiempo y suponen recoger con dos series de datos relativos a la misma población, al comienzo y al final del periodo de estudio.

Estos pueden ser a su vez:

- **Caso-Control:** son *retrospectivos*, es decir, se conoce el estado de enfermedad en los animales en la actualidad y se analiza la exposición a los factores en el pasado. Los integrantes del estudio se seleccionan según su estado sanitario.
- **Cohorte:** son *prospectivos*, se conoce el grado de exposición a los factores y se analiza la presentación posterior de la enfermedad. Los integrantes se seleccionan según la exposición al factor.

En cualquier estudio observacional la información puede presentarse en una tabla de contingencia de 2x2, con la siguiente estructura:

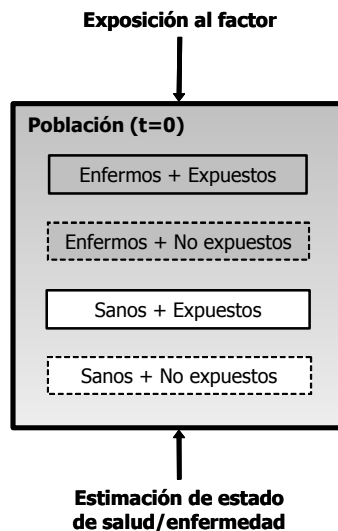
		Expuestos al factor		
		Sí	No	
Enfermedad	Enfermos	a	b	a + b
	Sanos	c	d	c + d
		a + c	b + d	N=a+b+c+d

donde:

- a:** enfermos expuestos al factor
- b:** enfermos no expuestos al factor
- c:** sanos expuestos al factor
- d:** sanos no expuestos al factor

Estudios transversales: Razón de Prevalencias

En este tipo de estudio se analiza la presentación o no de la enfermedad y la exposición o no al factor asociado en un **momento concreto del tiempo**. Para ello, partimos de una población inicial (o de una muestra representativa) y se recoge la información en ese momento concreto del tiempo sobre la enfermedad y la exposición al factor. Al inicio de este tipo de estudios, sólo conocemos el tamaño total de la población (N) o de la muestra de la misma (n), ya que no se han usado como criterios de selección ni el estado de salud ni del estado de exposición al factor asociado.



Para aclarar la mecánica seguida por este tipo de estudio plantearemos un caso en el que se estudia una enfermedad en una población donde la prevalencia es del 30% y la exposición al factor es del 60%. Así pues para una muestra representativa de 100 animales se distribuirían según la siguiente tabla de contingencia:

		Expuestos al factor		
		Sí	No	
Enfermedad	Enfermos	12	18	30
	Sanos	48	22	70
		60	40	100

La principal desventaja de este tipo de estudios es que dependen de la *prevalencia* de enfermedad (no sabemos si se trata de casos nuevos o antiguos) en un momento del tiempo y con la exposición al determinante en ese mismo momento, lo que supone ignorar cuantos casos de enfermedad son debidos a la exposición al factor en ese momento o a situaciones previas.

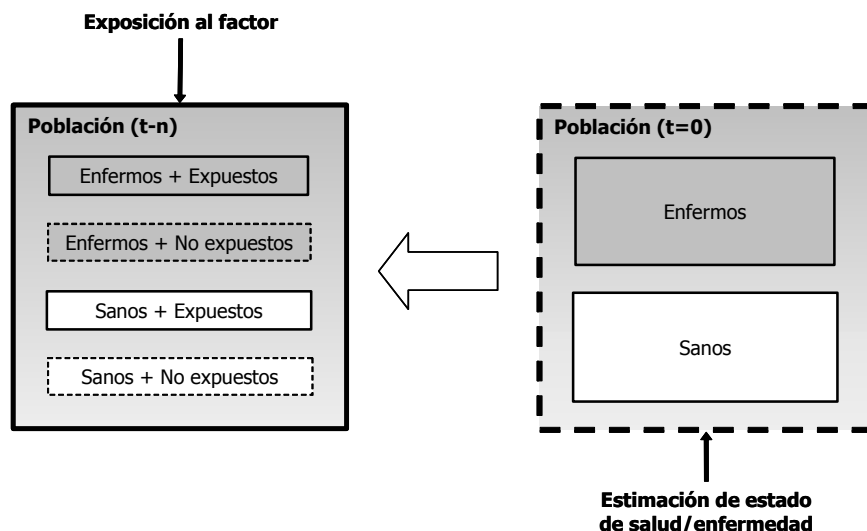
En estos estudios, el parámetro de elección para evaluar la relación factor/enfermedad es **la Razón de Prevalencias (RP)**, que es igual al cociente entre la prevalencia entre los expuestos al factor en un momento del tiempo y la prevalencia entre los no expuestos al factor en ese mismo momento. Este parámetro sólo es aplicable en estudios transversales.

$$RP = \frac{\text{Prev. expuestos}}{\text{Prev. no expuestos}} = \frac{\frac{a}{a+c}}{\frac{b}{b+d}}$$

A veces también se utiliza el Odds Ratio (OR) que equivale a un valor aproximado del riesgo que suponer estar expuesto o no al factor, y que explicaremos posteriormente.

Estudios caso-control: Odds Ratio

En este tipo de estudios se analiza la enfermedad y la exposición al factor asociado a lo largo de un **período de tiempo de forma retrospectiva**. Para ello partimos de una población (o muestra representativa) y formamos dos grupos iguales de animales, uno de enfermos y otro de sanos. A continuación buscamos que animales estuvieron expuestos al factor en un momento dado del pasado, así pues este tipo de estudios se organizan en función de la presentación de la enfermedad.



En este tipo de estudios conocemos N, A+B y C+D al inicio del mismo, tal y como podemos ver en nuestro ejemplo:

		Expuestos al factor		
		Sí	No	
Enfermedad	Enfermos	20	30	50
	Sanos	35	15	50
		55	45	100

Los principales inconvenientes de este tipo de estudios son que se trabaja con *prevalencias* y que la información retrospectiva de la exposición al factor a veces está ausente o es incorrecta.

En estos estudios, el parámetro de elección para medir el riesgo de exposición al factor es el **Odds Ratio (OR)**, que representa un valor aproximado del riesgo de enfermedad que supone estar expuesto al factor.

En su cálculo es preciso establecer una razón entre la tasa de enfermos/sanos entre los expuestos partido por la misma tasa entre los no expuestos, obteniéndose la siguiente fórmula para el cálculo de este parámetro.

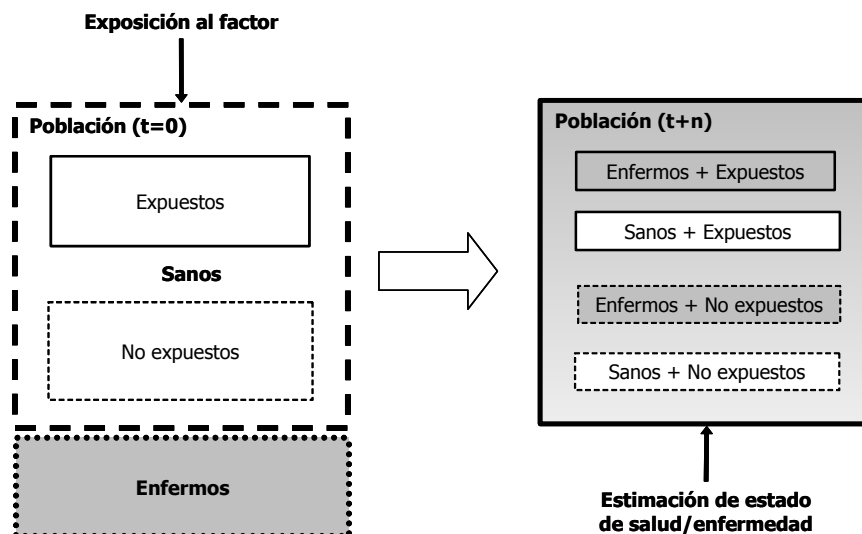
$$OR = \frac{a/c}{b/d} = \frac{a \cdot d}{b \cdot c}$$

El valor del OR obtenido tiene un valor generalmente próximo al Riesgo Relativo.

Hay que poner de manifiesto que este parámetro también se puede utilizar en los otros dos tipos de estudios observacionales (transversal y de cohorte), aunque son de elección los específicos de cada uno de ellos.

Estudios cohorte: Riesgo Relativo.

Como en el caso anterior la investigación también se realiza en un **período de tiempo** pero en este caso **con carácter prospectivo**. Para ello, partiendo de una población o muestra representativa se forman dos grupos de animales: sanos y enfermos. Desecharemos este último grupo y entre los animales sanos seleccionaremos dos grupos preferentemente de igual tamaño: animales expuestos al factor y animales no expuestos. A continuación recogeremos la información correspondiente a la aparición de la enfermedad en un momento posterior al inicio del estudio.



En este tipo de estudios conocemos N, A+C y B+D al inicio del mismo, tal y como podemos ver en nuestro ejemplo:

		Expuestos al factor		
		Sí	No	
Enfermedad	Enfermos	10	22	32
	Sanos	40	28	68
		50	50	100

Los principales inconvenientes de este tipo de estudios son que se trabaja con *prevalencias* y que

Este tipo de estudios observacionales es el que más parecido a los estudios experimentales, y su gran *ventaja* radica en que trabaja con incidencias, de forma que se conoce el efecto real de la exposición al factor. Por contra su coste económico es muy alto y la duración suele ser prolongada, ya que precisa el seguimiento de los participantes en el estudio durante bastante tiempo.

El hecho de que los estudios de cohorte trabajen con incidencias, hace que nos encontremos con dos situaciones diferentes, cálculos con incidencia acumulada o con tasa de incidencia. De forma que en este último caso la tabla de contingencia será ligeramente diferente a la ya expuesta:

		Expuestos al factor		
		Sí	No	
Enfermedad	Enfermos	a	b	a+b
	UTR	N1	N2	N1+N2

donde **UTR** corresponde a las Unidades de Tiempo en Riesgo

Esto supone que el cálculo de la relación entre la exposición al factor y la enfermedad variará según el tipo de incidencia utilizada.

El parámetro de elección para medir la relación entre el factor y la enfermedad es el **Riesgo Relativo (RR)** que proporciona el valor exacto del riesgo de presentar la enfermedad que supone la exposición al factor. En estos casos se puede utilizar el OR como aproximación del riesgo, de forma que RR y OR serán muy similares en enfermedades con prevalencias bajas, incrementándose la diferencia al aumentar la prevalencia.

Así pues el Riesgo Relativo (RR) equivale a la razón entre la incidencia en los expuestos y la incidencia en los no expuestos, y en caso de trabajar con Incidencia Acumulada, la fórmula será:

$$RR = \frac{\text{Inc. expuestos}}{\text{Inc. no expuestos}} = \frac{\frac{a}{a+c}}{\frac{b}{b+d}}$$

mientras que si se trabaja con *Tasa de Incidencia*, se aplicará la siguiente fórmula:

$$RR = \frac{\text{Inc. expuestos}}{\text{Inc. no expuestos}} = \frac{\frac{a}{N1}}{\frac{b}{N2}}$$

Este parámetro sólo puede utilizarse en estudios observacionales de cohorte, puesto que las incidencias entre expuestos y no expuestos sólo son calculables en este tipo de estudios.

Interpretación de resultados

Para todos los parámetros anteriormente descritos encontramos que valores de **1** indican que no existe asociación mensurable entre la exposición al factor y la enfermedad, de forma que valores diferentes a **1** indican que el factor influye de forma cuantificable sobre la presentación de enfermedad, presentándose en este caso dos posibles situaciones:

- valores **mayores de 1** indican que un incremento del riesgo de presentación de la enfermedad entre los expuestos al factor, equivalente a tantas veces dicho factor; en este caso se denomina **factor de riesgo**
- valores **menores de 1** indican una disminución del riesgo de enfermedad entre los expuestos, equivalente a tantas veces el inverso del factor; en este caso el determinante se denomina **factor de protección**

Sin embargo, en la mayoría de los casos no se trabaja con toda la población en el estudio y se realizan estudios a partir de muestras representativas lo que supone la introducción de un margen de error que se representa por un intervalo de confianza, de tal forma, que cuando el valor **1** está incluido en el intervalo, se considerará que no existen elementos de juicio suficientes para afirmar que el factor influye en la enfermedad. En caso contrario se aplicarán las reglas anteriormente descritas.

Además hay que tener en cuenta que tanto el Odds Ratio como el Riesgo Relativo anteriormente descritos son los denominados **crudos**, es decir, que se han calculado mediante un análisis univariable, en el que no se considera la posible interferencia de otros factores:

- **interacción**: un factor que afecta a la presentación de la enfermedad puede estar condicionado por otros
- **confusión (confounding)**: dos factores parecen estar asociados pero realmente se debe a que existe un tercer factor común a ambos.

El análisis realizado hasta ahora no contemplaba la intervención de varios factores a la vez en la presentación de la enfermedad, de forma que para analizar estas situaciones se calcula el Odds Ratio y/o el Riesgo Relativo **ajustados** que permiten conocer el grado de confusión e interacción, utilizando técnicas epidemiológicas como la estratificación y/o el matching.

Riesgo atribuible y Número Necesario para Perjudicar/Tratar (NNP/NTT)

Un parámetro de interés en la evaluación económica de los factores de riesgo/protección estudiados en un estudio observacional de cohorte es el **Riesgo Atribuible** que se define como la diferencia entre la incidencia acumulada en los expuestos y en los no expuestos.

Utilizando este valor se puede calcular el **Número Necesario para Perjudicar** (NNP) en el caso de los factores de riesgo y el **Número Necesario para Tratar** (NTT) en el caso de los factores de protección, que equivalen al inverso del Riesgo Atribuible en valor absoluto.

Así pues entendemos como Número Necesario para Perjudicar el número de individuos que se deben exponer al factor para aumentar en un caso la incidencia de la enfermedad, mientras que el Número Necesario para Tratar es el número de individuos que es preciso tratar para evitar la aparición de un nuevo caso en la población.

Por consiguiente un valor de NNP de 3,5 indica que cada 3,5 individuos que exponemos a un determinado factor de riesgo, supondrá la aparición de un nuevo caso de enfermedad, y de forma análoga un NTT de 5 indica que cada 5 individuos tratados (expuestos a un factor de protección) se evitará la aparición de un nuevo caso.

Esto es importante ya que desde el punto de vista económico nos permite estimar la oportunidad establecer una serie de medidas preventivas orientadas a la supresión del factor de riesgo o de aplicar un determinado tratamiento para evitar una enfermedad. Por ejemplo en el caso de un NTT igual a 5 correspondiente a la aplicación de una vacuna de un coste individual de 0,2 céntimos frente a una enfermedad cuya aparición supone unas pérdidas de 0,5 céntimos, vemos como la aplicación de dicha vacuna puede no tener ningún interés desde el punto de vista económico ya que el coste de aplicación para evitar un caso es de 1 céntimo ($5 \times 0,2$ céntimos) frente a unas pérdidas de 0,5 céntimos.

Sin embargo en el caso de un NNP de 3,5 asociado a un factor de riesgo cuya supresión costaría 0,1 céntimos y donde el coste de la enfermedad es de 0,5 céntimos, vemos que en este caso los costes derivados de la supresión del factor (0,35 céntimos) son claramente inferiores al coste derivado de la aparición de un nuevo caso, y por tanto en este caso sería interesante introducir las medidas adecuadas para la supresión del factor.