

ACTA SESION FORMACION II - Madrid 21 noviembre 2003

**INTRODUCCION A LA INTERPRETACION DE
LOS ENSAYOS CLINICOS.**

Asistentes:

Esther Inés
Juanse Hernández
JC Pueyo
Kike Rodríguez
Marion Zibelli
Udiarraga García
William Mejías
Xavi Franquet
Mauro Guarinieri

Facilita:

Joan Tallada

**Introducción a la interpretación de
los ensayos clínicos**

Joan Tallada

**Foro Español de Activistas de Tratamientos
FEAT
Madrid, 21/11/03**

Referencia: *Caroline Sabin*
Reader in Medical Statistics and Epidemiology
Department of Primary Care and Population Sciences, RF&UCMS

Joan comienza explicando los distintos tipos de estudios: Experimental frente a Observacional, Transversal frente a Longitudinal y Prospectivo frente a Retrospectivo, etc.

Tipos de estudios

Experimental frente a **Observacional**

Transversal frente a **Longitudinal**

Prospectivo frente a **Retrospectivo**

Experimental frente a Observacional

Estudio Experimental

El investigador interviene en el cuidado del paciente de modo pre-planificado, experimental y registra los resultados.

Estudio Observacional

El investigador no interviene en el cuidado del paciente de ningún modo que no sea la atención clínica rutinaria. El investigador sólo registra lo que sucede.

P: Un estudio observacional, ¿siempre es objetivo?

R: Si, pues el investigador siempre trabaja con datos.

Transversal frente a Longitudinal

Estudio Transversal (Cross-sectional)

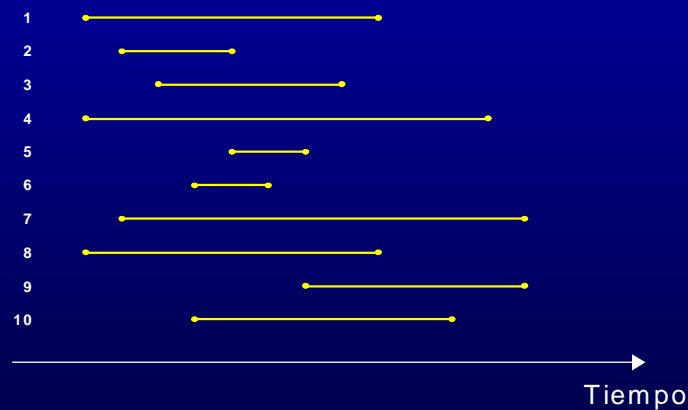
Se estudia a los pacientes en un único corte temporal.

Estudio Longitudinal

Se estudia a los pacientes al seguirlo a lo largo de un periodo de tiempo.

Transversal frente a Longitudinal

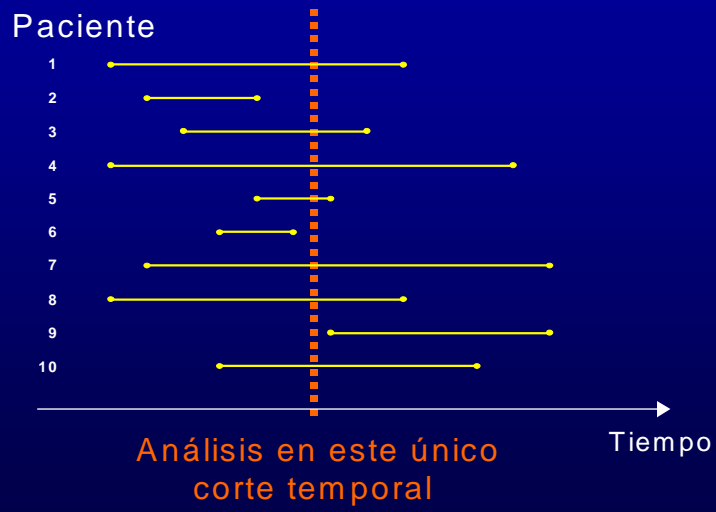
Paciente



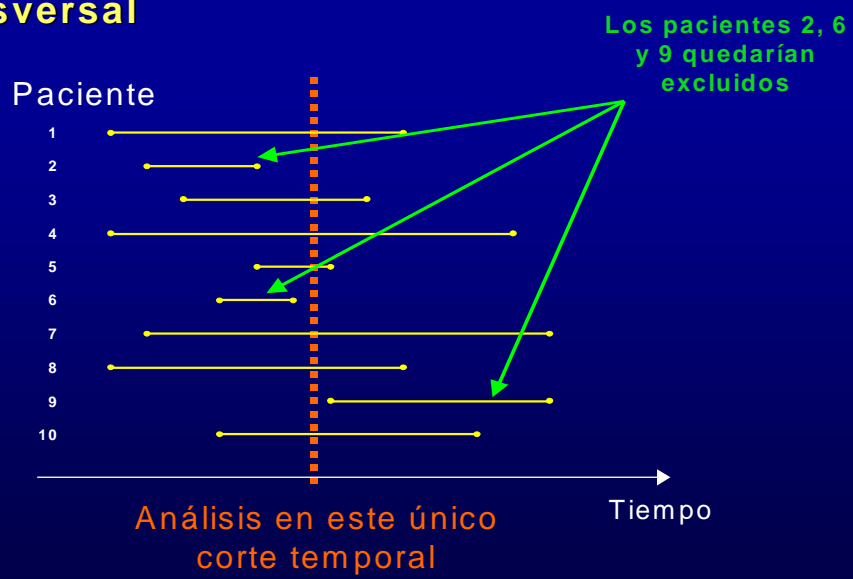
P: Explica la diferencia entre estudio transversal y longitudinal con un ejemplo.

R: El estudio transversal sería como una fotografía, mientras que el longitudinal sería como un video.

Transversal

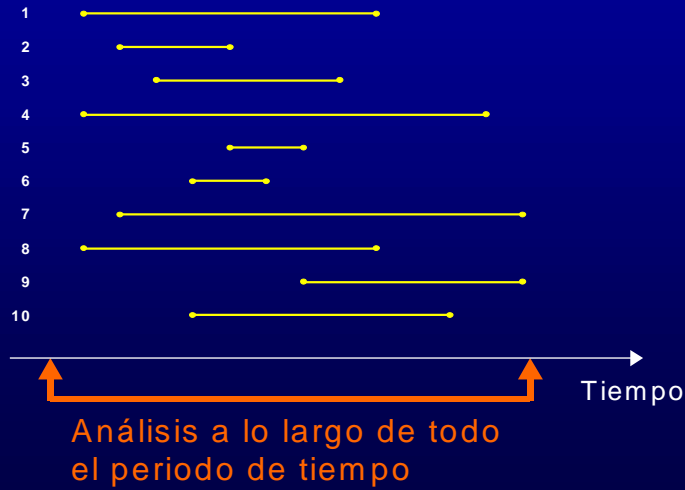


Transversal



Longitudinal

Paciente



Longitudinal

Paciente



P: En los estudios longitudinales, ¿todas las personas han de participar durante el mismo periodo de tiempo?

R: No siempre, unos pueden empezar antes y otros después. Un ejemplo: Estudio longitudinal de 48 semanas realizado en un periodo de 3 años.

Prospectivo frente a Retrospectivo

Estudio Prospectivo

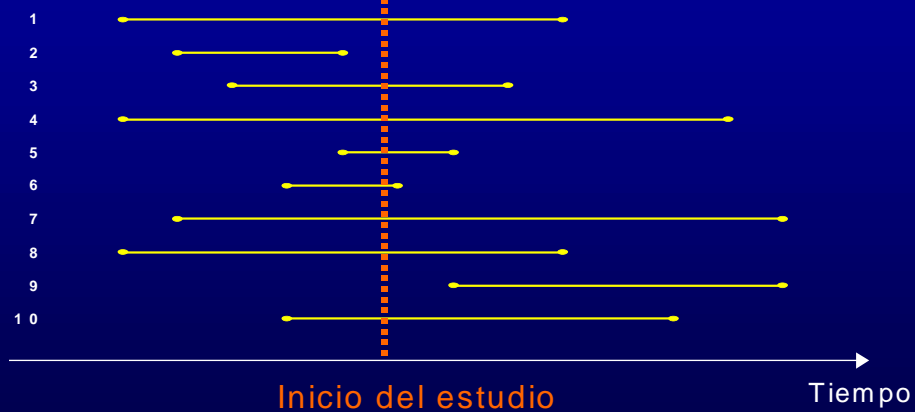
Se sigue a los pacientes hacia adelante en el tiempo.

Estudio Retrospectivo

“Se sigue” a los pacientes atrás en el tiempo (se trata habitualmente de analizar retrospectivamente los registros de los pacientes).

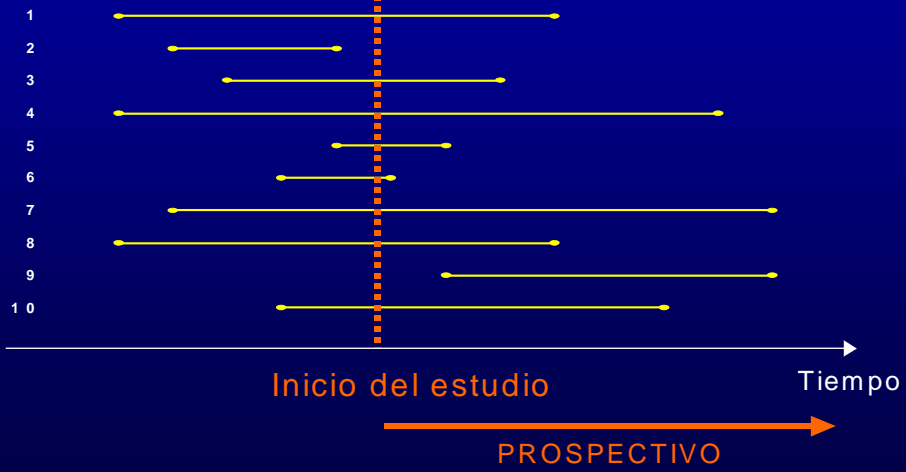
Prospectivo frente a Retrospectivo

Paciente



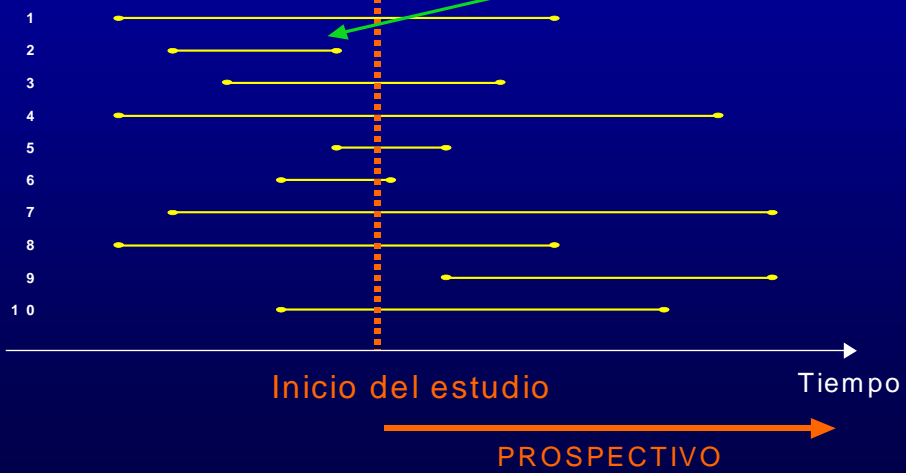
Prospectivo frente a Retrospectivo

Paciente

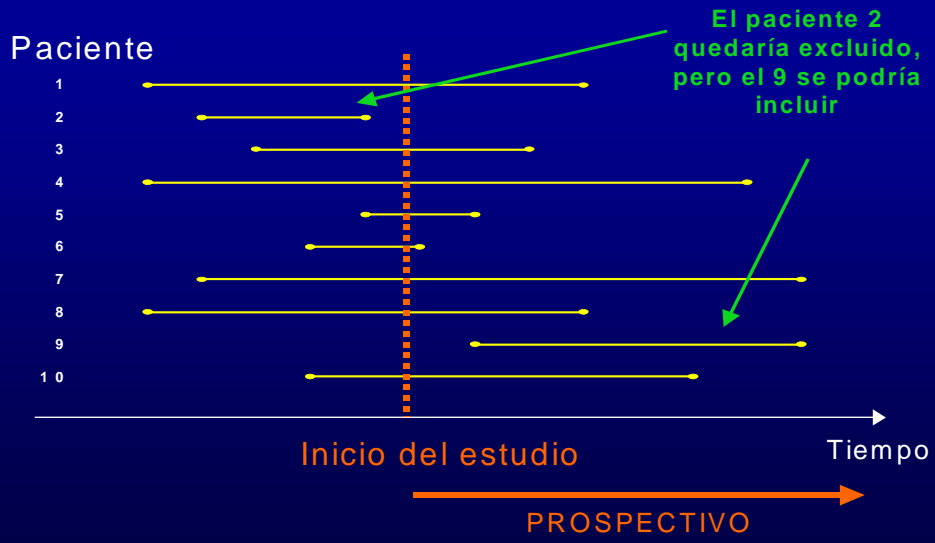


Prospectivo frente a Retrospectivo

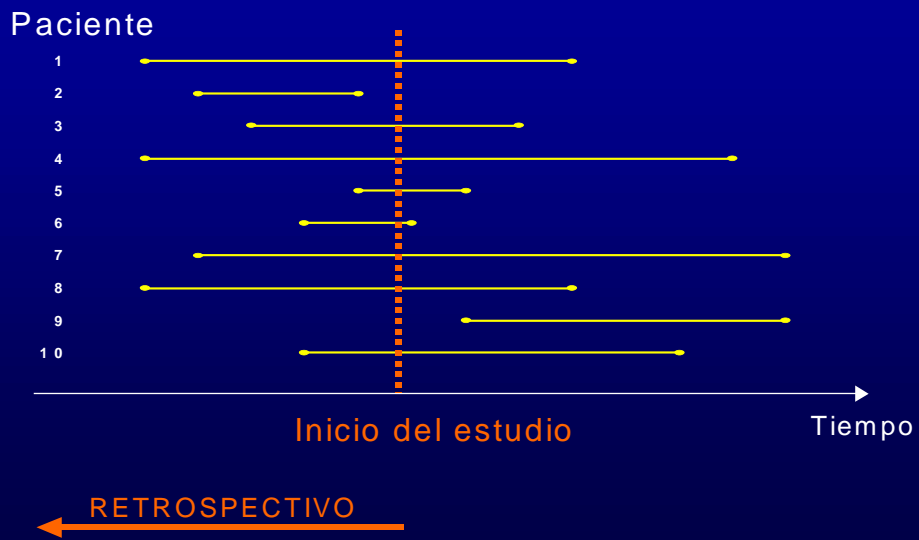
Paciente



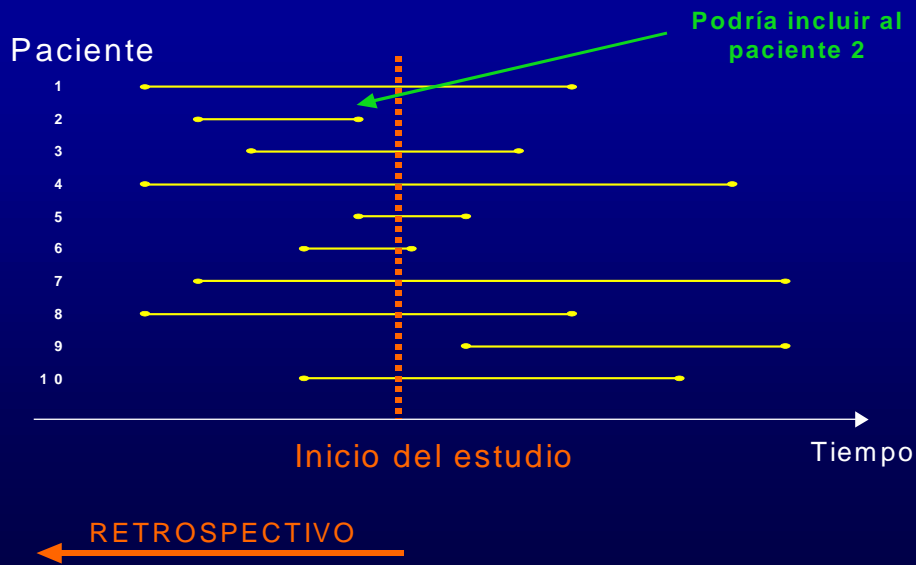
Prospectivo frente a Retrospectivo



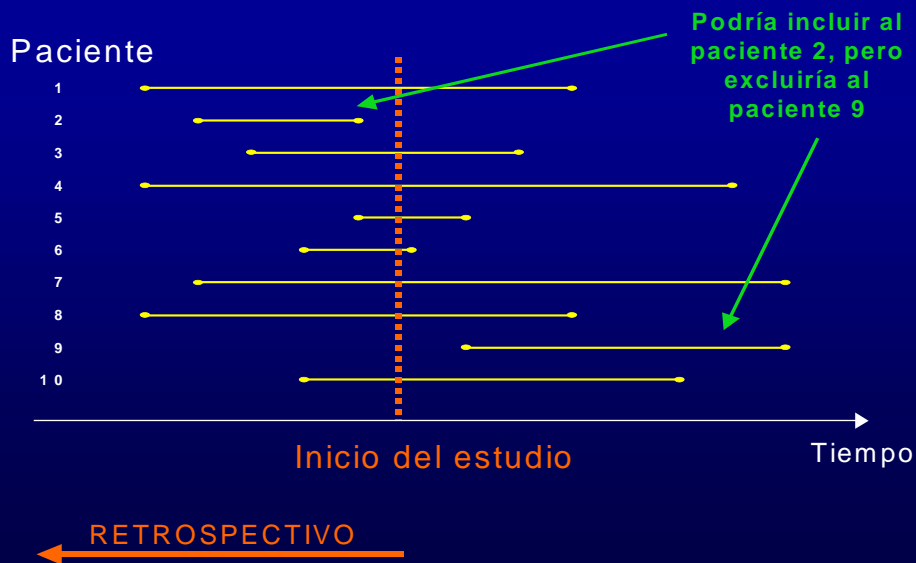
Prospectivo frente a Retrospectivo



Prospectivo frente a Retrospectivo



Prospectivo frente a Retrospectivo



P: ¿Es importante el tiempo empleado en un estudio?

R: Si lo es. Para que el tiempo en el que se va a desarrollar el experimento sea el adecuado, éste dependerá de lo que se quiera estudiar, pero generalmente a más tiempo, más fiabilidad.

Por ejemplo en caso de un estudio de toxicidad a largo plazo (lipodistrofia) si que es necesario que se realice durante un gran espacio de tiempo.

P: ¿Y el número de participantes?

R: También es un factor a tener en cuenta. Según lo que se quiera estudiar se necesitará un grupo más o menos amplio de personas, por ejemplo para comprobar la eficacia y seguridad de un fármaco si que es necesario que el número de participantes sea grande.

P: ¿De lo anteriormente expuesto, que tipos de estudios serían los más completos?

R: La mejor combinación de tipos para el mejor estudio es que sea experimental, prospectivo y longitudinal.

Diseños de estudios.-

Ensayos Controlados Aleatorizados, ECA – Estudios experimentales

Principales tipos de diseños de estudio

- **Ensayo controlado aleatorizado (ECA)**
- **Estudio de cohorte**
- **Estudio de control de casos (case control)**
- **Estudio transversal**
- **Series de casos/evaluación de notificación de casos**
- **Revisión de la literatura/revisión sistemática**
- **Meta-análisis**

Con frecuencia se utilizan estos términos de manera equivocada (sobre todo los de cohorte y los de control de casos).

En los estudios experimentales hay dos grupos, el grupo que recibe el fármaco y el grupo control o que recibe un placebo y esto se hace de forma aleatoria y controlada. El grupo control recibe dosis de placebo internacionales recomendadas, nunca por debajo o por encima. Los marcadores subrogados que se utilizan son los CD4 y la carga viral y se hacen evaluaciones de resultados o *endpoints*.

Ensayos controlados aleatorizados (ECA)

- Experimental y (habitualmente) longitudinal.

Ensayos **controlados** aleatorizados (ECA)

- Los ECA son **comparativos**: se comparan dos o más estrategias de tratamiento (con uno o más **grupo(s) control**).
- El grupo control nos permite concluir que cualquier mejora en el resultado se debe al tratamiento puesto a prueba, y no a cualquier otro factor de carácter externo.
- La comparación se suele establecer entre un **nuevo régimen** (o una nueva estrategia) y un **régimen existente**.
- Si no tenemos un régimen existente, el grupo control puede consistir en personas no tratadas que toman normalmente **placebo**.

Ensayos controlados **aleatorizados** (ECA)

- Los individuos se asignan a grupos de tratamiento por un proceso denominado **aleatorización** (como lanzar una moneda al aire cuando se trata de dos grupos).
- Esto garantiza que los grupos de tratamiento son **parecidos** al inicio del ensayo; cualquier diferencia sólo es imputable al "boleo".
- La aleatorización es la **característica más importante** de un ECA. El ECA es interpretado como el **modelo ideal** para obtener las pruebas de la eficacia del tratamiento.

P: ¿Como se define al grupo que se quiere experimentar?

R: Grupo o brazo de tratamiento.

P: ¿Que régimen se aplica al grupo control?

R: El estandarizado o aconsejado.

P: ¿Qué es basal?

R: Es el dato o datos de inicio del estudio.

Ciego

- Ensayo de **doble ciego**: ni el paciente ni el médico conocen qué tratamiento toma el primero.
- Garantiza que no **se puede introducir un sesgo** por diferencias en el seguimiento o la atención a determinados grupos.
- También se puede "cegar" al resto de personal del estudio (por ejemplo, quien recoge los datos) de manera que la información sobre los resultados se registra sin sesgos.

P: Los estudios aleatorizados ¿a veces pueden resultar poco equitativos?, es decir que los grupos no sean homogéneos o estén mal repartidos...

R: No, al contrario, son los que nos dan mayor garantía de que la distribución de los participantes de ambos grupos es lo mas parecida posible, en la aleatorización estadísticamente las probabilidades de que ocurra lo contrario son muy lejanas.

P: ¿Se pueden hacer reajustes en ciertos estudios aleatorizados, por ejemplo si son de género?

R: Si.

Joan explica que la palabra randomizado se suele ver mucho en la descripción de los estudios, significa lo mismo que aleatorizado, (en inglés randomized) pero está mal traducido o expresado, ésta proviene de RAM, -acrónimo inglés utilizado en informática para la aleatorización.

Tipos de ECA: Ensayos transversales

- **En general, los ensayos transversales necesitan de menos participantes ya que utilizan a cada uno de ellos como su propio control para reducir la variabilidad interindividual.**
- **No obstante, los ensayos transversales sólo son adecuados para tratamientos dirigidos a los síntomas antes que los que ofrecen una “cura”.**
- **Los ensayos transversales son difíciles de analizar si acaecen cambios temporales o estacionales durante el periodo de tratamiento.**

Ejemplo de ensayo transversal en VIH

Seguridad y aceptabilidad del condón Reality™ para el sexo anal entre hombres que tienen sexo con hombres.
Renzi C et al. AIDS (2003); 17: 727-31.

Se aleatoriza a parejas serodiscordantes para recibir o bien:

- a) Condones Reality durante 6 semanas seguido de condones de latex por 6 semanas, O BIEN
- b) Condones de latex durante 6 semanas seguido de condones Reality por 6 semanas

Evaluación de resultados (*Endpoints*): Frecuencia de resbalo con retirada, dolor o incomodidad en el uso, hemorragia rectal, voluntad de empleo futuro.

Limitaciones de ECA en la infección por VIH

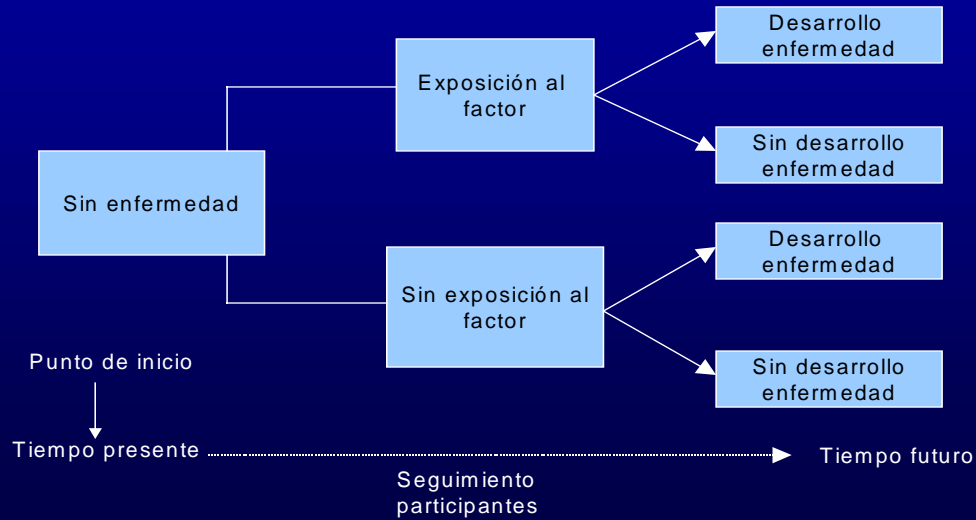
- Sólo es posible en el caso de que haya una intervención respecto a la cual se pueda aleatorizar a la gente.
- El seguimiento de participantes en un ensayo puede ser diferente en cada hospital; también pueden diferir los tipos de pruebas realizadas.
- Los participantes en el ECA puede que no sean representativos de la población hospitalaria: en concreto, puede que algunos grupos (mujeres, UDVP) estén excluidos.
- El ensayo puede ser corto (48 semanas) y sólo centrarse en la comparación de los dos o tres tratamientos principales.
- Se concentra en evaluaciones de resultado (*endpoints*) de marcadores surrogados antes que en eventos clínicos a largo plazo.

Estudios de cohorte.

P: Define cohorte.

R: Es un grupo definido con iguales caracteres.

Un estudio de cohorte tradicional



El ejemplo de la diapositiva de un estudio de cohorte, suscita confusión a algunos de los asistentes...

P: ¿Qué se entiende por factor? ¿El factor puede ser una patología en concreto?

R: El factor puede ser una patología o puede ser cierto fármaco, por ejemplo un IP.

P: ¿La exposición al factor si es una enfermedad es siempre intencionada?

R: No.

Estudios de cohorte en la infección por VIH

- Distinguen las cohortes “epidemiológicas” tradicionales, en la que la visita de cohorte es independiente de la atención clínica (por ejemplo, MACS, ALIVE, HERS), y **bases de datos observacionales de cohortes clínicas**.
- Esto último suele incluir a todos los pacientes de una clínica y los datos registrados son los que se obtienen de forma rutinaria.
- Las cohortes son útiles para describir la historia natural del VIH, los cambios en la progresión de la enfermedad, y en la identificación de los marcadores de pronóstico.

Ejemplo de estudio de cohorte

Cohorte de Hemofilia del Hospital Royal Free (HRF)

- 111 hombres con hemofilia registrados en el Centro de Hemofilia del HRF que se infectaron con VIH entre 1979 y 1985.
- Estos hombres han sido seguidos por más de 20 años para describir la historia natural de la infección por VIH.
- La información se obtiene de modo rutinario: datos demográficos, de sucesos clínicos, de laboratorio y sobre los tratamientos.
- Actualmente, 39 de estos hombres siguen vivos y 28 de ellos están bajo seguimiento en el hospital.

Estudios de control de casos

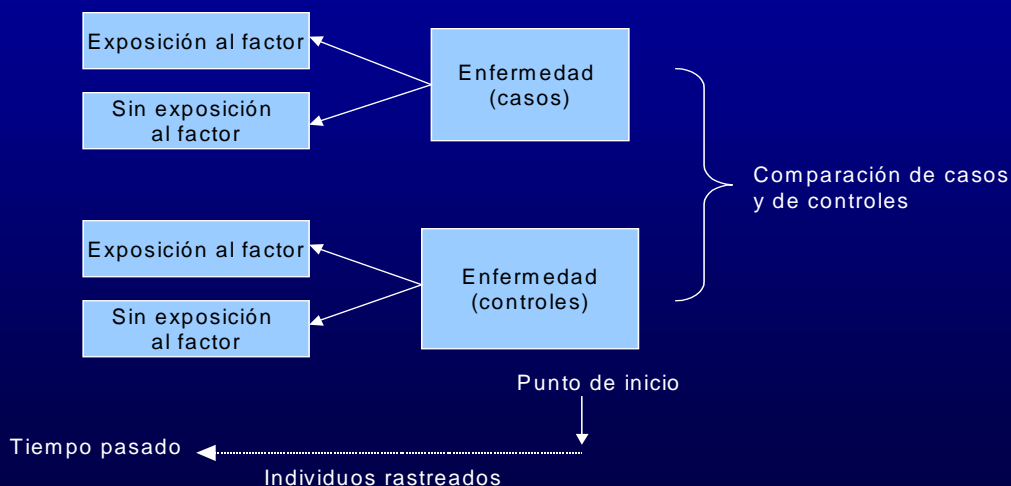
Estudios de control de casos

- Observacional y longitudinal (tradicionalmente retrospectivo).
- Se compara a un grupo de pacientes con una enfermedad (**casos**) con un grupo de pacientes sin la enfermedad (**controles**).
- El objetivo es observar si la **exposición** a algún factor ha sucedido con más o menos frecuencia *en el pasado* en los casos que en los controles.
- Los casos y los controles se pueden **emparejar** a partir de la información demográfica básica (p.e., sexo y edad) para hacer que los dos grupos sean lo más parecidos posible.

P: ¿Podría ser un estudio de control de casos la comparación de un grupo con otro que realice ciertas prácticas de riesgo?

R: Si.

Diagrama de un estudio de control de casos tradicional



Ejemplo de estudio de control de casos

Redes sexuales de mujeres embarazadas con y sin infección por VIH. Johnson KM et al. AIDS (2003); 17: 605-12.

Casos: Mujeres VIH+ (n=75) que se visitan un hospital en Perú.

Controles: Mujeres VIH- (n=137) que se visitan en el mismo hospital.

VARIABLES DE INTERÉS: historia sexual, tamaño de la red sexual (a partir de la historia sexual propia y de las parejas, más las relaciones de segunda y tercera generación).

Resultados: el tamaño de la red sexual predijo el estado serológico, con independencia del número de mujeres de las parejas.

Conclusiones: "El riesgo de infección por VIH entre las mujeres embarazadas en Lima depende en buena medida del riesgo de sus parejas masculinas".

P: ¿Cuál sería el diseño que aportaría mayor rigor científico a un estudio?

R: Que sea un estudio experimental, prospectivo, aleatorizado y doble ciego.

Estudios transversales (*cross-sectional*)

- Se realizan en un **único** corte temporal: no hay seguimiento.
- Habitualmente incluye un cuestionario o encuesta y se utiliza para determinar la prevalencia de una condición, para describir la situación actual o para determinar las actitudes y las creencias.
- **Ventajas:** relativamente barato y rápido
- **Inconvenientes:** no es posible calcular la incidencia de una enfermedad, pero sí puede determinar la **prevalencia**.

Ejemplo de estudio transversal

Prevalencia del VIH y comportamiento de riesgo entre clientes de trabajadoras sexuales en Abidjan, Costa de Marfil. Vuylstene BL et al. AIDS (2003); 17: 1691-4.

- Entrevistas a 526 clientes al abandonar la habitación de la trabajadora sexual (se pidió a todos una muestra de saliva).
- 423 (80,4%) proporcionaron una muestra de saliva; la prevalencia en este grupo fue del 13,4%.
- El uso declarado de condón fue muy alto (el 92,7% lo utilizaba siempre, el 95.4% lo utilizó en esa visita).
- Una edad más avanzada y el estar casado/cohabitar se asociaba de manera significativa con la infección por VIH.

Series de casos / revisión de notificación de casos

- Una forma de evidencia bastante pobre pero puede proporcionar datos preliminares útiles.
- Útil como herramienta descriptiva: p.e., para definir la historia natural de una enfermedad o describir prácticas actuales.
- No hay elemento de comparación: por ello no es posible mostrar una conexión entre la exposición y la enfermedad.
- Por definición, estos estudios son habitualmente retrospectivos, por lo que puede que haya problemas con los datos históricos.

Ejemplo de una revisión de notificación de casos

Uso de nevirapina en niños infectados con VIH-1.
Verweel G et al. AIDS (2003); 17: 1639-47.

- Revisión de las notificaciones de casos de todos los niños con VIH 96 semanas después de empezar nevirapina entre agosto de 1997 y marzo de 1999.
- Se inscribió a 74 niños.
- El fármaco se toleró bien; exantema (*rash*) en el 20% pero sólo fue severo en 4 niños.

Repaso de la literatura/repaso sistemático

- Es útil cuando ya hay muchos datos publicados sobre un tema.
- Repaso a la literatura: cualquier repaso de la literatura, que no se hace necesariamente de manera sistemática, y no pretende identificar *todos* los estudios.
- Revisión sistemática: se presta atención a garantizar que se incluyan *todos* los estudios disponibles.
- El objetivo de ambos es describir los hallazgos y las variaciones entre estudios, antes que intentar establecer un cálculo común de las características del efecto.

Meta-análisis

- Es útil cuando los estudios publicados por separado son en general demasiado pequeños como para alcanzar conclusiones sólidas.
- El objetivo es conjuntar datos existentes sobre un tema particular con el objetivo de ofrecer un cálculo combinado del efecto.
- Puede llevarse a cabo utilizando datos publicados (extensión de un repaso sistemático) u obteniendo datos no elaborados proporcionados directamente por los investigadores.
- Susceptible a **un sesgo de publicación**, por lo que tiene que garantizar o intentar que incluya además estudios no publicados (p.e., *abstracts* de Conferencias, estudios en marcha).

Ejemplo de un meta-análisis

Efectos de los alelos CCR5-Δ32 y CCR2-64I en la progresión de la enfermedad en niños con VIH-1 infectados por vía perinatal. Ioannidis JPA et al. AIDS (2003); 17: 1631-8.

- Datos de genotipo extraídos de 10 estudios (n=1317)
- Estudios identificados a través del buscador de MEDLINE, *abstracts* de los principales encuentros, comunicación con expertos, anuncio en *Nature Medicine*, y presentaciones en reuniones de SIDA.
- En relación a la progresión a SIDA clínico, ni el CCR5-Δ32 ni el CCR2-64I mostraron en conjunto una tendencia no significativa de protección.

Descripción de datos

Descripción de los datos cuantitativos

Los datos se muestran agrupados por valores estadísticos.

Esto permite a quien lee el estudio obtener una impresión sobre nuestro conjunto de datos y permite su comparación entre ellos.

Habitualmente se presentan dos tipos de medidas:

- Una medición de *valor medio*.
- Una medición de la *variabilidad*.

Mediciones del 'promedio'

Tres mediciones del promedio comúnmente descritas:

- La media (aritmética)
- La mediana
- La moda

La media

$$\text{Media} = \frac{\text{Suma de todos los valores individuales}}{\text{Número de valores}}$$

Ejemplo – Recuentos de CD4 (células/mm³) después de la seroconversión en 111 hombres con hemofilia

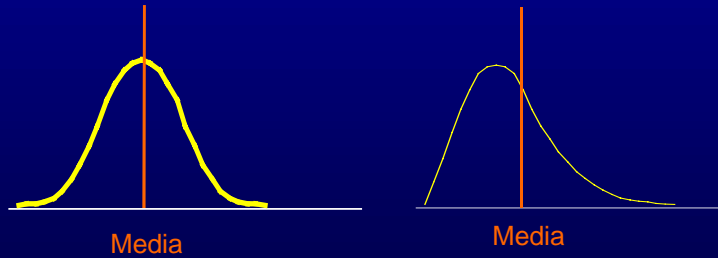
530, 490, 1100, 580, 410, 370, 220, 700,,
510, 790, 1380, 1920, 40, 330, 720, 580, 440

$$\begin{aligned}\text{Media} &= \frac{530+490+1100+\dots+580+440}{111} \\ &= \frac{76840}{111}\end{aligned}$$

$$= 692,25 \text{ células/mm}^3$$

¿Cuándo la media no es indicativa?

La media funciona bien cuando los datos son simétricos pero no cuando están descompensados.



La mediana

La **mediana** es el valor del “medio” al poner en orden todos los valores

- En el caso de que el número de medidas sea impar, entonces la mediana es el valor del medio.
- La mediana es particularmente adecuada para datos de distribución descompensada.

P: ¿Qué ocurre cuando la media no es compensada?

R: Que los datos no son simétricos, y algunos elementos son dispares al conjunto, aquí la media da una falsa percepción y esto lo solucionaría la mediana.

Ejemplo (cont.) – Recuentos de CD4 (en orden numérico)

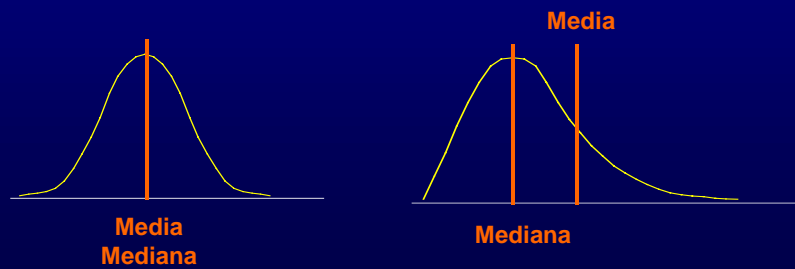
40, 40, 150, 200, 210, 220, 240, 270, 280, 290, 290, 310,, 1380, 1470, 1490, 1700, 1920, 2090, 2500

Mediana = Valor número 56 en el orden (n=111)
= 600 células/mm³

Ejemplo (cont.)

Media = 692 células/mm³
Mediana = 600 células/mm³

En general :



¿Cuándo es inadecuada la media/mediana?

No se puede calcular en el caso de valores nominativos.

Ejemplo – *Grupos por vía transmisión por 100 pacientes*

Homosexual	n=54
IDU	n=23
Heterosexual	n=11
Receipt of blood products	n=5
Other	n=5

Medición de colocación - la moda

La **moda** es el valor que se repite más veces.

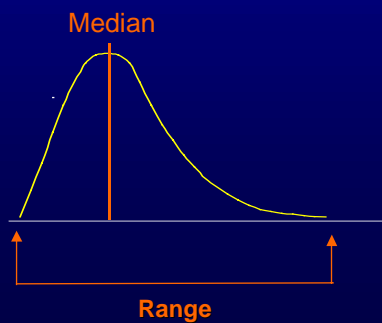
- La moda se menciona muy raramente, pero puede ser apropiada en mediciones de categorías.
- En el ejemplo anterior, la categoría que era la moda era *homosexual* (ya que fue la vía más común en el grupo).

¿Media, mediana o moda?

Tipo de datos	
Numéricos, distribución normal	Media
Numéricos, descompensados	Mediana
Categorica, nominal	Moda
Categorica, ordinal, pocas categorías	Moda
Categorica, ordinal, un número considerable de categorías	Mediana

El rango

El **rango** es el valor máximo y el valor mínimo

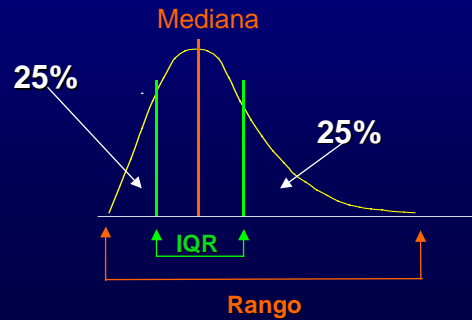


- La mediana, que es el valor medio de todos los valores
- El modo, el valor que se repite más veces
- El rango, es el valor máximo y el valor mínimo

El rango y los rangos intercuartiles (IQR)

El **rango** es el valor máximo y el valor mínimo

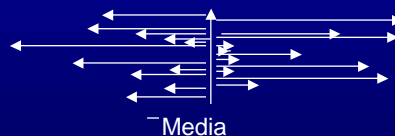
El **rango intercuartil (IQR)** es aquel en el que entran los datos centrales que componen el 50% del total.



Surgen comentarios durante la explicación del rango intercuartil.

La varianza y la desviación estándar

La **varianza** ofrece una medición de la desviación promedio de cada punto desde la media conjunta



- La **varianza** no se expresa en el mismo tipo de unidades que la media, por lo que no se pueden comparar.
- Por ello, prefiere utilizarse la **desviación estándar (SD)** (raíz cuadrada de la varianza).

Varias preguntas sobre el cálculo de la varianza y la desviación estándar, y de cómo se han de utilizar, con amplias respuestas de Joan.

Ejemplo (cont.)

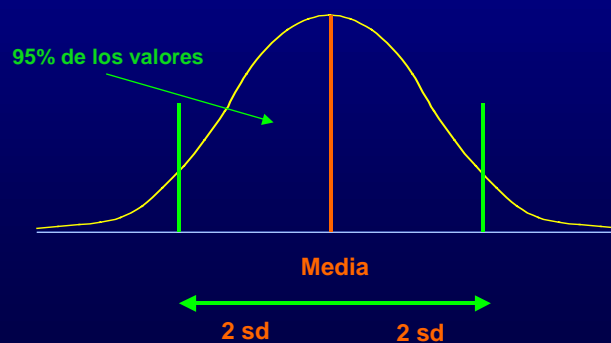
$$\begin{aligned}\text{Varianza} &= \frac{1}{111} \times [(530-692.3)^2 + (490-692.3)^2 + \dots + (440-692.3)^2] \\ &= 161304,9\end{aligned}$$

la media

$$\begin{aligned}\text{Desviación estándar} &= \sqrt{161304,9} \\ &= 401,63 \text{ células/mm}^3\end{aligned}$$

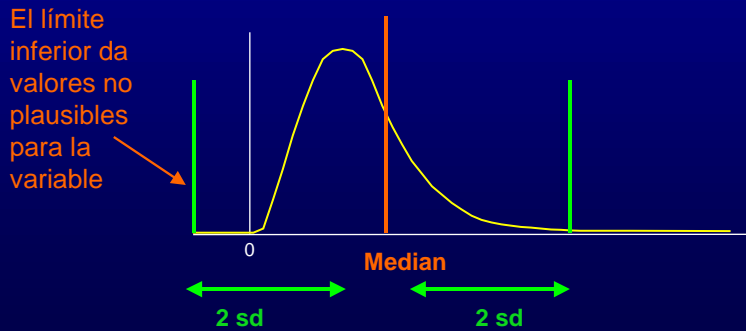
¿Cómo se utiliza la SD?

Si los datos están distribuidos normalmente, aproximadamente el 95% de los valores de los datos estarán dentro del rango: media \pm 2 SD



¿Cómo se utiliza la SD?

Si los datos no tienen una distribución normal, el rango no te ofrece valores razonables, por lo que la SD no tiene utilidad.



Muestras y poblaciones

Población: el grupo entero que nos gustaría estudiar idealmente.

Muestra: el subgrupo de de la población realmente estudiado.

Estamos interesado en el valor de un **parametro** en la población (p.e., proporción de personas que fuman).

Tenemos que **calcular** esto utilizando los datos recorgidos en nuestra muestra.

¿Qué es la incertidumbre?

Utilizamos la *muestra* para sacar conclusiones sobre la *población*.

Pero dado que no hemos muestreado la población entera tenemos que introducir algunas incertidumbres en nuestros cálculos.

Necesitamos tener esto en cuenta al realizar pruebas sobre los datos.

Cuantificar la incertidumbre

Queremos saber cuánta incertidumbre hemos introducido.

Podemos cuantificar esto calculando el *intervalo de confianza* de nuestro cálculo.

Los intervalos de confianza proporcionan un rango de valores plausibles respecto al parámetro de la verdadera población que estarán en consonancia con los datos obtenidos en la muestra.

Ejemplo

Un nuevo fármaco (A) ha sido probado en 10 pacientes. De estos 8 (80%) han experimentado respuesta virológica a las 24 semanas.

- ¿Crees que el índice de respuesta 'verdadero' es del 80%?
- Si los pacientes tratados, ¿también esperarías que los pacientes que responden son 80?
- Si el índice de respuesta 'verdadero' fuese del 70%, ¿cuál sería la probabilidad de que, por casualidad, hubieran respondido 8 de 10?
- Si el índice de respuesta 'verdadero' fuese del 50%, ¿cuál sería la probabilidad de que hubieran respondido 8 de 10?

Intervalos de confianza

Podemos generar **intervalos de confianza** para la proporción que responde.

Éste es el rango de los valores 'verdaderos' que podrían haber resultado en 8 de 10 de los pacientes que responden.

En este caso, el intervalo de confianza del 95% está entre el **44% al 98%**: tenemos la confianza en un 95% de que el índice de respuesta verdadero se encuentra en ese rango.

Así que, existe una posibilidad de que 8 sobre 10 de que estos pacientes hubiesen respondido incluso si la respuesta verdadera hubiese sido cualquier valor en un rango entre el 44% y el 98%.

Interpretación de los intervalos de confianza

“El porcentaje estimado en conjunto de pacientes con ≤ 400 copias/ml de ARN de VIH plasmático a las 24 semanas fue del 64% (intervalo de confianza [IC] del 95%: 60 al 67%).”

Bartlett JA et al. Overview of the effectiveness of triple combination therapy in antiretroviral-naïve HIV-1 infected adults. *AIDS* (2001); 15: 1369-77.

Interpretación de los intervalos de confianza

Si el intervalo de confianza es amplio (esto es, el rango incluye un amplio número de valores posibles para el parámetro verdadero) no podemos alcanzar una conclusión fiable.

Interpretación de los intervalos de confianza

Si el intervalo de confianza es amplio (esto es, el rango incluye un amplio número de valores posibles para el parámetro verdadero) no podemos alcanzar una conclusión fiable.

→ el cálculo es **impreciso**

Interpretación de los intervalos de confianza

Si el intervalo de confianza es amplio (esto es, el rango incluye un amplio número de valores posibles para el parámetro verdadero) no podemos alcanzar una conclusión fiable.

→ el cálculo es **impreciso**

Si el intervalo es estrecho, hemos sido capaces de alcanzar una conclusión fiable sobre el valor del parámetro.

Interpretación de los intervalos de confianza

Si el intervalo de confianza es amplio (esto es, el rango incluye un amplio número de valores posibles para el parámetro verdadero) no podemos alcanzar una conclusión fiable.

→ el cálculo es **impreciso**

Si el intervalo es estrecho, hemos sido capaces de alcanzar una conclusión fiable sobre el valor del parámetro.

→ el cálculo es **preciso**

Interpretación de los intervalos de confianza

- Una buena manera de juzgar si disponemos de un cálculo preciso es fijarse en los límites inferior y superior del IC.
- Si el valor real fuera el de cualquiera de esos dos límites, ¿cambiarías tus conclusiones?

Interpretación de los intervalos de confianza

Se ha probado un nuevo fármaco en 20 pacientes naïve a los antirretrovirales, de los 15 (75%) han respondido.

Intervalo de confianza del 95%: del 56% al 94%.

Interpretación de los intervalos de confianza

Se ha probado un nuevo fármaco en 20 pacientes naïve a los antirretrovirales, de los 15 (75%) han respondido.

Intervalo de confianza del 95%: del 56% al 94%.

- Si la respuesta verdadera fuese el 56%, ¿qué pensarías de este tratamiento?

Interpretación de los intervalos de confianza

Se ha probado un nuevo fármaco en 20 pacientes naïve a los antirretrovirales, de los 15 (75%) han respondido.

Intervalo de confianza del 95%: del 56% al 94%.

- Si la respuesta verdadera fuese el 56%, ¿qué pensarías de este tratamiento?
- Si la respuesta verdadera fuese el 94%, ¿qué pensarías de este tratamiento?

Intervalos de confianza: el efecto de aumentar el tamaño de la muestra

Si asumimos que responde el 80% de los pacientes, ¿cuál sería el intervalo de confianza?

Tamaño muestra	No (%) responden	IC 95%
10	8 (80,0%)	(44% al 98%)

Intervalos de confianza: el efecto de aumentar el tamaño de la muestra

Si asumimos que responde el 80% de los pacientes, ¿cuál sería el intervalo de confianza?

Tamaño muestra	No (%) responden	IC 95%
10	8 (80,0%)	(44% al 98)
50	40 (80,0%)	(66% al 90%)
100	80 (80,0%)	(71% al 87%)
500	400 (80,0%)	(77% al 84%)
1000	800 (80,0%)	(78% al 83%)

Intervalos de confianza: el efecto de aumentar el tamaño de la muestra

Si asumimos que responde el 80% de los pacientes, ¿cuál sería el intervalo de confianza?

Tamaño muestra	No (%) responden	IC 95%
10	8 (80,0%)	(44% al 98)
50	40 (80,0%)	(66% al 90%)
100	80 (80,0%)	(71% al 87%)
500	400 (80,0%)	(77% al 84%)
1000	800 (80,0%)	(78% al 83%)

Conforme aumenta el tamaño de la muestra, en general el intervalos de confianza se hace más estrecho, indicando un cálculo más preciso.

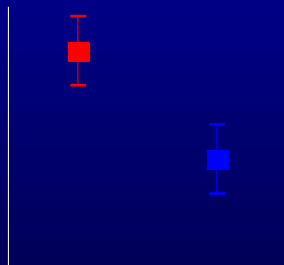
Interpretación de intervalos de confianza (2)

“Se calcula que un 0% (95% IC: 0%-8,6%) de los niños que reciben AZT neonatal inmediato con o sin AZT intraparto se han infectado, en comparación con el 22,1% (95% IC: 14,2% - 31,8%) de los que no han tomado AZT o sólo AZT antenatal tardío.”

Sirinavin SS et al. Effect of immediate neonatal zidovudine on prevention of vertical transmission of human immunodeficiency virus type 1. *Int J Infect Dis* (2000);4:148-52.

Interpretación de los intervalos de confianza

Los intervalos de confianza no se yuxtaponen



Podemos concluir que existe una diferencia real entre los dos grupos

Interpretación de intervalos de confianza (3)

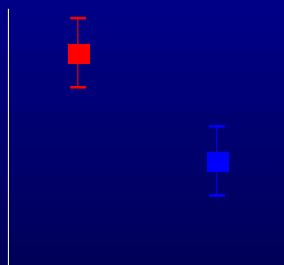
“El porcentaje estimado en conjunto de pacientes con ≤ 400 copias/ml de ARN de VIH plasmático a las 24 semanas fue del 64% (intervalo de confianza [IC] del 95%: 60 al 67%).

Los porcentajes de pacientes con < 50 copias/ml de ARN de VIH plasmático a las 48 semanas por clase de fármaco fueron: triple IP, 46% (95% IC: 41 al 52%); triple ITINN, 51% (43 al 59%); triple ITIN, 45% (36 al 54%).”

Bartlett JA et al. Overview of the effectiveness of triple combination therapy in antiretroviral-naïve HIV-1 infected adults. *AIDS* (2001); 15: 1369-77.

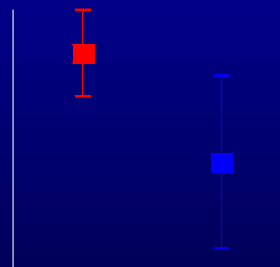
Interpretación de los intervalos de confianza

Los intervalos de confianza no se yuxtaponen



Podemos concluir que existe una diferencia real entre los dos grupos

Los intervalos de confianza se yuxtaponen



No podemos sacar ninguna conclusión sobre la diferencia si no tenemos más información

Interpretación de los intervalos de confianza

- Si los intervalos de confianza se yuxtaponen, no podemos extraer ninguna conclusión sobre si existe o no una diferencia real entre los grupos.
- Incluso si pareciera que existe una diferencia entre los dos grupos, necesitamos determinar si esta diferencia podría haberse dado por casualidad o si es probable de que se trate de un hallazgo real.
- Para evaluar esto hace falta llevar a cabo un test de **significación (o de hipótesis)**.

Sobre el valor p

Interpretación de valor p

El valor p expresa la probabilidad de que los resultados obtenidos se puedan extrapolar a la población a la que hace referencia la muestra. Es decir, que la diferencia obtenida sea o no estadísticamente significativa.

Es un valor expresado entre 0 (es imposible que los resultados se obtuvieran por casualidad) y 1 (tenemos la certeza de que los resultados se dieron por casualidad).

Se ha establecido arbitrariamente $p < 0,05$ como el umbral de lo estadísticamente significativo.

P: ¿Qué interpretación se le da al valor p, si éste es por ejemplo 0.07, ó 0.08?

R: Cuanto mayor sea de 0.05 (que es el umbral de lo estadísticamente significativo), menos probabilidades hay de que los datos sean fiables, y por lo tanto menos se podrán extrapolar a la citada población.

¿Qué es un ensayo clínico?

Cualquier forma de **experimento planificado** que incluya a **pacientes** y que esté diseñado para determinar el **tratamiento más adecuado** para una dolencia médica concreta

Selección de pacientes de un ensayo

- Un ensayo debe tener criterios de inclusión y de exclusión explícitos: definiciones precisas sobre quién puede ser incluido en el estudio.
- Los participantes deben ser ampliamente representativos del futuro grupo de pacientes a quien se pueda aplicar los resultados del estudio.
- Sin embargo, rara vez los participantes en ensayos son una selección aleatoria de todas las personas que viven con VIH.

Selección de participantes para un ensayo

Discutir el ensayo con el paciente y determinar la elegibilidad



Consentimiento informado



El paciente entra en el ensayo



Aleatorización

Evaluación de la respuesta –el *endpoint* primario

- En cualquier ensayo hay que definir al menos una medición de resultados o *endpoint* primario que nos indique los efectos clave del tratamiento sobre el participante.
- La medición primaria normalmente se refiere a la **eficacia**.
- Si no hay coherencia entre los resultados de diferentes mediciones, la primaria será aquella a partir de la cual se establecerá o no el valor del fármaco.

Evaluación de la respuesta - *endpoint* secundarios

- Además de la medición primaria, también podemos necesitar definir un número determinado de mediciones secundarias.
- Se suelen relacionar con la toxicidad o la calidad de vida, o pueden ser otras mediciones de eficacia que no refleja el *endpoint* primario.

Definiciones de *endpoints* – ejemplo

Sustitución de análogos de nucleósido por abacavir en pacientes con lipoatrofia asociada al VIH. Carr A et al. JAMA (2002); 288: 207-215.

Endpoint primario:

Cambios medios en la masa grasa medidos por DXA a la semana 24

Endpoints secundarios:

Eventos adversos
Antropometría
Masa grasa total y central
Mediciones bioquímicas, de lípidos, y glucémicas.
Carga viral
Recuentos de CD4
Calidad de vida

Definiciones de *endpoints* – ejemplo

Sustitución de análogos de nucleósido por abacavir en pacientes con lipoatrofía asociada al VIH. Carr A et al. JAMA (2002); 288: 207-215.

Endpoint primario:

Cambios medios en la masa grasa medidos por DXA a la semana 24

Endpoints secundarios:

Eventos adversos
Antropometría
Masa grasa total y central
Mediciones bioquímicas, de lípidos, y glucémicas.
Carga viral
Recuentos de CD4
Calidad de vida

Comentamos el ejemplo de la diapositiva expuesta, y se llega a la conclusión que pueden haber endpoints “no muy fiables” por ejemplo si el estudio no se ha realizado en un periodo de tiempo suficientemente amplio si así se requiere.

También se apunta que los Endpoint secundarios se han de preveer antes del inicio del ensayo, para evitar errores de apreciación.

Endpoints clínicos frente a surrogados

- En general estamos más interesados en el efecto de un nuevo tratamiento en un resultado clínico (p.e., nuevos sucesos de SIDA o defunciones)
- Ahora bien, hoy día, los ensayos de TARGA que utilizan *endpoints* clínicos tiene que ser extremadamente amplios y seguir a los pacientes por periodos de tiempo muy largos para conseguir suficiente poder como para detectar diferencias ente regímenes de tratamiento.
- Por ello, con frecuencia se considera el efecto del régimen de tratamiento a partir de una medición surrogada (p.e., cambios en los CD4 o el ARN del VIH)

P: Define marcador subrogado.

R: Es un marcador de sustitución, suelen ser datos indirectos, y pueden ser por ejemplo los recuentos de CD4, la carga viral etc.

Análisis por intención de tratamiento (ITT)

Todos los pacientes aleatorizados para tomar el tratamiento deben incluirse en el análisis en los grupos a los que fueron aleatorizados

Análisis por intención de tratamiento (ITT)

- Ofrece una medida del efecto del tratamiento en la vida real
- Es el único **cálculo no sesgado** del efecto del tratamiento.
- La mayoría de las grandes revistas científicas exigen análisis por ITT
- Las presentaciones deben incluir análisis por ITT como el análisis principal a no ser que se proporcione una justificación bien fundamentada para no hacerlo así.

Análisis en tratamiento

Sólo incluye a los pacientes que completan un curso entero del tratamiento al que fueron aleatorizados.

Análisis en tratamiento

- Se sugiere que muestra el efecto óptimo del tratamiento cuando se toma siguiendo las recomendaciones.
- Sin embargo, este enfoque tiene el potencial de proporcionar cálculos **muy sesgados** del efecto del tratamiento ya que es probable que los participantes con las peores respuestas sean los que abandonan o cambian de tratamiento.
- Es probable que este enfoque ofrezca un cálculo demasiado optimista sobre el efecto del nuevo tratamiento.
- **NO DEBE UTILIZARSE EN ANÁLISIS PRIMARIOS**

Métodos alternativos de análisis de ITT

Cuando faltan datos completos sobre marcadores surrogados, hay varias alternativas:

- **ITT Abandono=Fracaso (ITT M=F)**

Todos los valores faltantes se tratan como si fueran fracasos en el análisis, con independencia de su valor más reciente: todos los pacientes entran en el denominador. Esto ofrece el valor más conservador o pesimista sobre el nuevo tratamiento

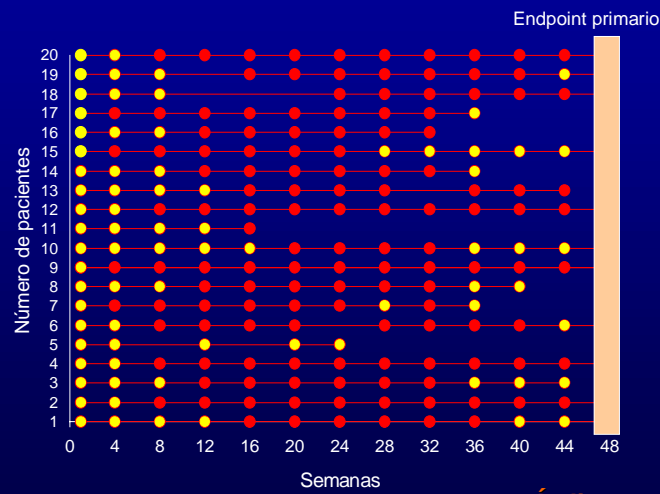
Métodos alternativos de análisis de ITT

Cuando faltan datos completos sobre marcadores surrogados, hay varias alternativas:

- **ITT última observación arrastrada (LOCF)**

Se considera la última medición disponible de cada persona como la válida hasta el final del tratamiento. Es un ITT ya que incluye a todos los pacientes en el denominador, pero las autoridades regulatorias no lo valoran positivamente (FDA o EMEA).

Ejemplos de diferentes enfoques

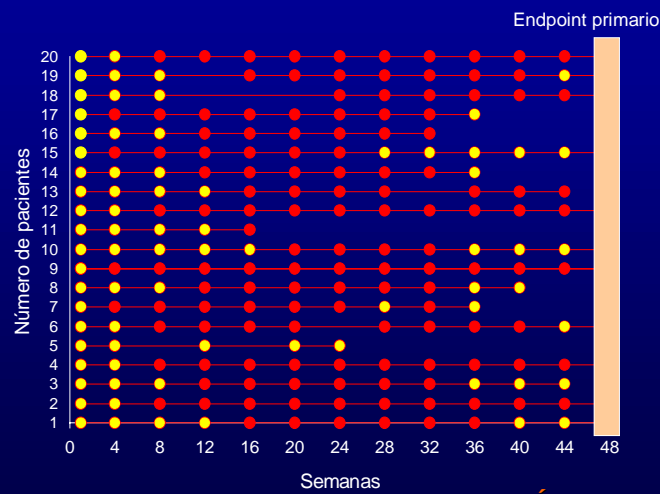


Respondedor:
ET/
ITT abandono
= exclusión

1
1
1
-
-
0
-
-
1
-
0
1
-
-
1
-
1
-
1
0

Índice respuesta = 8/11
= 73%

Ejemplos de diferentes enfoques

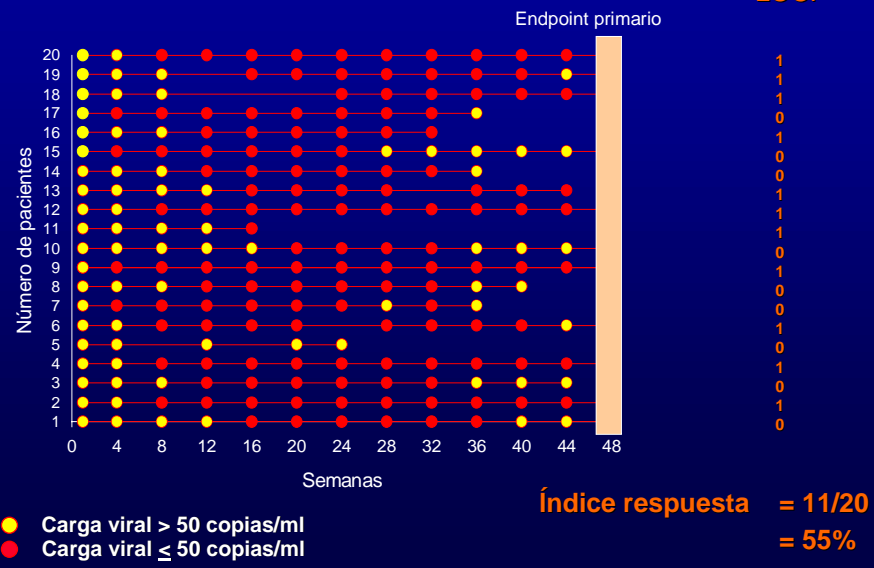


Respondedor:
ITT abandono
= fracaso

1
1
1
0
0
0
0
0
0
1
0
0
1
0
0
1
0
0
1
0
1

Índice respuesta = 8/20
= 40%

Ejemplos de diferentes enfoques



Ejemplos de diferentes enfoques: resumen

<i>Enfoque</i>	<i>Índice respuesta</i>
En Tratamiento/ITT Aband.=Exclusión	73%
ITT Abandono = Fracaso	40%
ITT Abandono = LOCF	55%

Tras varias preguntas se aclara la diferencia entre análisis en tratamiento y análisis por intención de tratamiento.

Amplio debate sobre el concepto, ya que hay diferencias de opiniones sobre si algún participante abandona el estudio, será por culpa del fármaco o no.

Se comenta que el ITT es el estudio que nos va a dar una visión más global de lo que ha ocurrido durante el ensayo para poder extrapolarlo después a la realidad.

Y que nos va a reflejar lo que ocurrirá en la vida real al iniciar un tratamiento, en este punto todos coincidimos que para considerar que un tratamiento sea aceptable se miden múltiples factores, tales como eficacia, seguridad, toxicidad, adherencia etc...

Por tanto si el tratamiento no se cumple por el motivo que sea, podremos decir que éste no se ajusta a nuestras necesidades.

Debemos pues enfocarlo como el Abandono = Fracaso.

Análisis de subgrupos

- **Con frecuencia existe la tentación de considerar el efecto del régimen de tratamiento en una serie de subgrupos en los análisis.**
- **Por ejemplo, considera el efecto del régimen en grupos como:**
 - **Hombres/mujeres**
 - **Baja/alta carga viral a nivel basal**
 - **Bajo/alto recuento de CD4 a nivel basal**
 - **Naïve/Experimentados al inicio del tratamiento**

Análisis de subgrupos

- Hay varios riesgos de hacer demasiados análisis de subgrupos.
- El aumento del número de pruebas que se lleven a cabo significa que algunas de las comparaciones pueden ser significativas probablemente por casualidad.
- Pese a que el estudio tenga poder suficiente como para detectar diferencias de conjunto, los subgrupos pueden estar basados en tamaños de muestras mucho más pequeños y no tendrán suficiente poder estadístico.

Análisis ínterin

- En cualquier ensayo existe una preocupación de que uno de los brazos de tratamiento puede ser inferior por alguna razón a los otros (p.e., un régimen puede ser mucho más eficaz o puede estar asociado con un índice superior de toxicidad grave respecto a otros).
- Si esto es así, puede considerarse que no es ético continuar poniendo a los pacientes en riesgo de una toxicidad grave o de fracaso de tratamiento.
- A resultados de esto, se planifican uno o más análisis ínterin en un momento temporal preplanificado para seguir el progreso del ensayo.

Análisis ínterin

- Sin embargo, siempre queda la posibilidad de que los hallazgos iniciales, en especial si el número de pacientes es pequeño, puedan haber surgido por casualidad.
- Si se permite que el ensayo siga adelante hasta el final, estas tendencias pueden desaparecer.
- Hay que ser muy cuidadoso respecto a interrumpir el ensayo precozmente a partir de los resultados del análisis ínterin.
- Si se va a llevar a cabo un análisis ínterin habitualmente se recomienda que el ensayo sólo se interrumpa si la evidencia de una diferencia entre los brazos es muy sólida (p.e., $p < 0,0001$).

Análisis ínterin: el papel del DSMB

- Normalmente se nombra un Comité de Seguimiento y Seguridad de Datos (en inglés, *Data Safety and Monitoring Board*, DSMB).
- Incluye un cierto número de 'expertos' independientes en el tema, y suele incorporar a un estadístico.
- El DSMB evalúa los datos de seguridad en forma regular (esta información no suele estar cegada) e informa de ello al Comité Científico (en inglés, *Steering Committee*) del ensayo.
- El DSMB puede recomendar interrumpir precozmente un ensayo si lo cree necesario.

Análisis ínterin (cont.)

- Si los resultados ínterin sugieren superioridad de uno de los brazos, pero el DSMB no recomienda interrumpir el ensayo, la presentación de los resultados puede entorpecer la finalización exitosa del ensayo:
 - ❖ Los pacientes que ya han sido aleatorizados podría cambiar al brazo superior, lo que llevaría a altas tasas de abandono en el inferior.
 - ❖ Los nuevos pacientes podrían negarse a ser aleatorizados al brazo inferior.

Análisis ínterin (cont.)

- Si los resultados ínterin sugieren superioridad de uno de los brazos, pero el DSMB no recomienda interrumpir el ensayo, la presentación de los resultados puede entorpecer la finalización exitosa del ensayo:
 - ❖ Los pacientes que ya han sido aleatorizados podría cambiar al brazo superior, lo que llevaría a altas tasas de abandono en el inferior.
 - ❖ Los nuevos pacientes podrían negarse a ser aleatorizados al brazo inferior.
- Si se van a publicar los datos del análisis ínterin, es importante que o bien se mantenga el ciego o que los resultados se presenten seprados por grupos.