

CLASIFICACIÓN DE LOS ESTUDIOS CLÍNICOS Y ANALISIS DE LA LITERATURA MÉDICA PUBLICADA

Orellana B.* Moreira J.P.
Médico Internista Endocrinología
Hospital Roosevelt

RESUMEN

La aproximación del médico práctico a la literatura científica no es sencilla. No hemos tenido una formación sistemática en tal sentido, y la disponibilidad de la literatura, que en una época pasada era muy dificultosa, ha evolucionado a una abundancia de publicaciones muy difícil de abarcar, llegando en muchas ocasiones a saturarnos de información, la cual no siempre nos resulta útil en su totalidad. Este trabajo está orientado a compartir una forma de leer los trabajos científicos para extraer información que pueda ser útil en la práctica clínica, a través de una serie de preguntas que debemos de formularnos al tener de frente una publicación científica, logrando analizar de forma secuencial la validez interna del estudio, evaluar los resultados, y determinar la aplicabilidad del mismo a la práctica clínica diaria.

INTRODUCCIÓN

Enfrentado a la inmensa cantidad de información publicada en la literatura médica respecto a preguntas clínicas específicas, con resultados muchas veces contradictorios, es necesario disponer de herramientas que permitan analizar en forma crítica esa información. Dichas herramientas habitualmente consisten en una serie de criterios que intentan responder preguntas específicas de acuerdo al tipo de estudio que se está analizando. A su vez, el tipo de estudio está relacionado con la pregunta que generó el inicio del proceso de búsqueda de información.

La presente revisión tiene como objetivo la identificación y la interpretación de los principales estudios clínicos en investigación. La guía clasifica los diferentes estudios de investigación y desarrolla con mayor detalle los principales, como el ensayo clínico, los estudios de cohortes, los estudios de casos y controles y los metaanálisis,

elaborando cuadros comparativos con el propósito de que no solo la lectura, sino el entendimiento de los mismos sea más sencillo y organizado.

De la misma forma, se plantea una secuencia de pasos útiles a seguir en el momento que nos enfrentemos a literatura médica, de una manera secuencial y ejemplificada, para de alguna forma no solo distinguir si la lectura del mismo valga la pena, sino aprovechar de mejor manera la información que nos proporciona, y que los resultados se traduzcan en mejores decisiones a la hora de tratar a nuestros pacientes.

La Investigación Biomédica:

La investigación biomédica puede subclasificarse en investigación básica/preclínica e investigación clínica. La investigación biomédica preclínica es importante para aumentar el conocimiento de los mecanismos biológicos básicos. Los estudios comúnmente se realizan en instituciones o departamentos

preclínicos en campos tales como anatomía, bioquímica, biología celular, inmunología, microbiología, biología molecular, neurociencia, farmacología y fisiología. La investigación preclínica puede contribuir al descubrimiento de nuevos tratamientos médicos y la investigación clínica abarca desde los estudios de investigación o de laboratorio clínico hasta las pruebas de nuevos procedimientos clínicos, nuevas herramientas de diagnóstico

clínico y nuevos productos en seres humanos.

La clasificación de los estudios clínicos se basa habitualmente en la asignación del factor de estudio y en su finalidad. No obstante, en el cuadro que se presenta a continuación se presentan los criterios de clasificación de una forma más sencilla, hasta cierto punto arbitrario y otros criterios que pueden tomarse en cuenta.

<i>Criterios de Clasificación</i>	<i>Categorías</i>	
<i>Asignación del factor de estudio</i>	Experimental (manipulan la causa)	Observacional (no manipula la causa)
<i>Finalidad</i>	Análítica (evalúan causalidad)	Descriptiva (no evalúan causalidad)
<i>Direccionalidad</i>	Longitudinal (mediciones en distintos momentos del tiempo)	Transversal (mediciones en mismo momento en el tiempo)
<i>Existencia del grupo control</i>	Controlado	No controlado
<i>Dirección temporal</i>	Prospectivo (datos recogidos mientras se producen)	Retrospectivo (datos recogidos después de producirse)
<i>Tipo de asignación del factor del estudio</i>	Aleatorizada (interviene el azar)	No aleatorizada
<i>Tipo de datos</i>	Primarios	Secundarios (o de registro)
<i>Unidad en estudio</i>	Sujetos	Poblaciones
<i>Sujetos del grupo control</i>	Paralelo (los controles son sujetos)	Cruzado (el mismo sujeto es control)
<i>Temporalidad del grupo control</i>	Concurrente	Histórico

TIPOS DE ESTUDIOS

Los estudios descriptivos sugieren hipótesis pero no permiten encontrar asociaciones causales. Se consideran adecuados para describir la frecuencia y las características más importantes de un problema de salud en una población, y para proporcionar datos

en los que basar la hipótesis. Para evaluar una relación causa/efecto se requiere de estudios analíticos observacionales o experimentales. Los estudios clínicos no se acostumbran realizar a partir de toda la población. La mayoría de las veces no sería abordable. Por eso se definen unos criterios de la población en estudio, se selecciona una muestra y

a partir de su estudio se extrapolan posteriormente los resultados a la población diana. Un buen proceso de muestreo, junto con unos adecuados criterios de selección, permitirá a los estudios observacionales que la muestra en estudio sea representativa. En el caso de los estudios experimentales, se utilizará la asignación aleatorizada a las distintas estrategias en estudio, lo cual permitirá una distribución equitativa de los distintos factores que puedan influir en los resultados. (1)

Tipos de estudios clínicos más habituales

- Descriptivos
- *En poblaciones*
- *Estudios ecológicos*
- *En individuos*
- *A propósito de un caso*
- *Series de casos*
- *Transversales/prevalencia*
- Analíticos
- *Observacionales*
- *Estudios de casos y controles*
- *Estudios de cohortes (retrospectivos y prospectivos)*
- *Intervención*
- *Ensayo clínico*
- *Ensayo de campo*
- *Ensayo comunitario*

Se diferencian 2 tipos principales de errores en la investigación clínica. El primero, denominado error aleatorio o debido exclusivamente a las fluctuaciones del azar, se puede corregir con relativa facilidad, por ejemplo aumentando el tamaño de la

muestra para aumentar la precisión de la medida. El segundo, más problemático y difícil de corregir, se denomina error sistemático o sesgo, y se debe a la desviación sistemática de los valores verdaderos que deseamos estimar. De acuerdo con estos 2 tipos de error quedan definidas la validez o exactitud, o grado en que realmente medimos lo que queremos medir, y la precisión o grado de ausencia del error aleatorio.

Los factores de confusión son un caso especial de sesgos en los que un factor de confusión, que no es causa ni consecuencia del factor en estudio, se relaciona con éste y también con el resultado del estudio. Como medidas para luchar contra el sesgo, en la fase de diseño disponemos de la asignación aleatorizada a la intervención, la restricción (limitar las características de inclusión) o apareamiento, (inclusión de controles de características similares). La solución en la fase de análisis es más compleja y requiere de técnicas estadísticas de ajuste o estratificación. Los sesgos afectan a la validez interna; el grado en que los resultados están libres de error. La validez externa se refiere al grado en que los resultados obtenidos en la muestra son aplicables a la población global. (1,2)

ENSAYO CLÍNICO

Se trata de un estudio experimental, prospectivo, controlado, en el que se asigna de manera aleatorizada la estrategia de intervención. El ensayo

clínico se considera el estándar de oro para evaluar la eficacia de una intervención determinada.

Las características básicas para controlar los posibles sesgos en el ensayo clínico prototipo son la asignación aleatorizada, el enmascaramiento, la presencia del grupo control y el control de las pérdidas y los abandonos.(3)

La asignación aleatorizada se considera el principal pilar. Como consecuencia, dado que en promedio los grupos se pueden considerar idénticos, aparte de la intervención, cualquier diferencia en los resultados es, en teoría, debida a ésta.

En la técnica del enmascaramiento a simple ciego es el paciente el único que no conoce la asignación. En la del doble ciego, además tampoco la conoce el investigador. Actualmente, cuando se utiliza la denominación doble ciego, se suele incluir en el grupo que desconoce la asignación a cualquier persona que esté en relación con el desarrollo y la evaluación del estudio.

La presencia del grupo control se fundamenta en la necesidad de sustraer de la observación del efecto de la intervención test posibles interferencias que puedan sesgar los resultados. Los factores más conocidos son: 1) la regresión a la media o tendencia hacia valores promedio tras presentar valores extremos; 2) el efecto Hawthorne o la tendencia a un comportamiento distinto del habitual durante el transcurso de un estudio (p. ej., mejor respuesta al control de la presión arterial); 3) el efecto placebo por el que se observan respuestas ante la instauración de un teórico tratamiento (sustancia inactiva en realidad), y 4) la evolución natural de la enfermedad en estudio.

Aunque las pérdidas y los abandonos son difícilmente evitables, se aconseja adecuar el diseño para minimizarlos, preverlos para el cálculo adecuado del tamaño de la muestra, describirlos de manera desglosada y, finalmente, realizar un buen análisis evaluando el posible impacto de las pérdidas en los resultados.

Limitaciones

Restricciones éticas

Muestras muy seleccionadas (validez externa cuestionada) Las intervenciones pueden diferir de la práctica habitual. Evalúa la relación entre una única intervención y su efecto Costo elevado.

Los estudios deben diseñarse para responder a una pregunta clínicamente relevante o variable principal. La predeterminación del tamaño de la muestra debe fijarse en

ésta, y su resultado al final del estudio será el que permitirá concluir o no la eficacia. El resto de las variables se considerarán como secundarias y tienen únicamente un papel de

soporte de la principal. Esta estrategia permite conservar el nivel de falsos positivos. Otras estrategias son posibles, pero se requieren ajustes estadísticos especificados. En los ensayos convencionales se predetermina el tamaño de la muestra y se analizan los datos después del reclutamiento total del estudio. Hay diseños secuenciales con análisis intermedios que permiten el análisis gradual de los datos. Se distinguen habitualmente 3 poblaciones de análisis en el ensayo clínico. La población de seguridad incluye a todos los sujetos expuestos a la intervención. La población de intención de tratar —ITT— intenta minimizar sesgos de selección de la población, maximiza la similitud con la práctica general y tiende a no sobreestimar la eficacia; idealmente incluye a todos los sujetos aleatorizados o permite muy pocas exclusiones, como no tomar la medicación en estudio, los diagnósticos de entrada incorrectos o la falta de datos postaleatorización. La población por protocolo —PP excluye los casos con desviaciones importantes de protocolo y tiene como objetivo homogeneizar las características del estudio, y maximiza la posibilidad de discriminar entre intervenciones. Generalmente, en estudios de superioridad se considera más adecuada la aproximación ITT para estudios confirmatorios de eficacia y la de PP para estudios exploratorios. Para la no inferioridad, puesto que la ITT dificulta encontrar diferencias y no sería una postura

conservadora, se prefiere la aproximación PP.(2,3).

DISEÑOS DE SUPERIORIDAD, EQUIVALENCIA Y NO INFERIORIDAD

Aunque la mayor parte de estudios están diseñados desde una hipótesis de superioridad, también es interesante comprobar si 2 tratamientos son similares (diseño de equivalencia), o bien evaluar si un tratamiento test no es peor que un control activo de referencia (diseño de no inferioridad). Las pruebas de hipótesis clásicas no son útiles y deben ser redefinidas incluyendo unos márgenes de igualdad para delimitar la zona de aceptación de equivalencia, o bien aquel valor de no inferioridad, que determinará el valor máximo admisible para la inferior eficacia. Más allá de esos valores predefinidos, las diferencias se consideran demasiado importantes y no son aceptables. Los estudios de equivalencia no se suelen usar como tales en investigación clínica, a excepción de los ensayos de bioequivalencia. En los estudios de evaluación de eficacia lo que preocupa es vigilar que no se supere el valor de no-inferioridad (por falta de eficacia), pero no el hipotético escenario opuesto (superar el margen contrario por “exceso” de eficacia). Para facilitar la interpretación, estos estudios se suelen evaluar mediante intervalos de confianza (IC). Para la no-inferioridad se fija la estimación de la comparación entre grupos y su límite inferior del intervalo de confianza del 95% bilateral (IC del

95%) (o 97,5% unilateral hacia el lado que implique el peor escenario)

ESTUDIOS DE COHORTES

El estudio de cohortes es un estudio analítico, observacional, habitualmente prospectivo, en el que se compara una cohorte expuesta con una cohorte no expuesta a algún factor, y la aparición en el tiempo de eventos o enfermedades. Se les ha denominado también con los términos que escriben sus características (prospectivo, de seguimiento, concurrente, de incidencia, longitudinal), pero al no ser éstas exclusivas, de estos estudios, pueden llevar a confusiones.

El diseño es similar al ensayo clínico, pero la exposición no está intervenida por lo que el investigador actúa únicamente como observador. Al no estar los grupos generados por azar es más difícil asumir la comparabilidad entre las 2 cohortes, y están sujetos a más sesgos. Los más conocidos son el de selección (cohortes no comparables), de migración (cambios de expuesto a no expuesto, y viceversa), y de observación (distintas intensidades de observancia entre ambas cohortes). Por ser observacional, es especialmente útil cuando se sospecha de un factor de

incluye o no el valor δ de no inferioridad. (3)

riesgo y no es ética o práctica la asignación aleatorizada a la exposición. Dado que se parte de sujetos sanos (sin los eventos o enfermedades a estudiar), permiten estimar la incidencia de una determinada afección.

Es muy importante el estudio profundo de las características basales para: 1) descartar la enfermedad al inicio; 2) clasificar a los sujetos por el grado de exposición, y 3) evaluar el posible papel de potenciales factores de confusión. En algunos casos se han propuesto estudios de cohortes retrospectivos. Aunque se disminuye la larga duración habitual de los prospectivos, tiene más problemas: menor exhaustividad y calidad de la recogida de información, posible heterogeneidad en cuanto a criterios diagnósticos y de evaluación. Por tanto, son menos precisos que los prospectivos y resulta más complicado establecer la relación causa-efecto. El análisis de resultados se acostumbra presentar en forma de riesgo relativo entre cohortes, o riesgo de padecer la enfermedad en relación con la exposición.

Cacterísticas de los estudios de cohortes frente a otros estudios observacionales

Limitaciones

Costo elevado y larga duración.
Dificultad en la ejecución.
No son útiles en enfermedades raras.
Requieren generalmente un tamaño de la muestra elevado
El paso del tiempo puede introducir cambios en los métodos y criterios diagnósticos
Posibilidad de sesgos por pérdida en el seguimiento

ESTUDIOS DE CASOS Y CONTROLES

El estudio de casos y controles es analítico, observacional y retrospectivo. Se compara la exposición previa a uno o varios factores en un grupo de individuos que presentan una condición médica (casos) respecto de los que no la presentan (controles). Es crucial una buena selección de los casos, a poder ser nuevos o casos incidentes, o en su defecto prevalentes (útiles en enfermedad crónica, de baja letalidad, si el diagnóstico ha sido reciente o cuando se dispone de pocos recursos), pero todavía lo es más la de los controles. En el caso ideal, ambos están aparejados por características demográficas y potenciales factores de riesgo.

En estos estudios, al recogerse la información de manera retrospectiva,

ésta es de menor calidad y está más sujeta a sesgos. Los más conocidos son los sesgos de memoria y el del entrevistador. Útiles en las primeras fases de la investigación de una determinada afección, se pueden formular hipótesis de trabajo tras éste e incluso plantear hipótesis etiológicas. Son más efectivos en tiempo y costo que los estudios de cohortes, pero tienen menor evidencia de causalidad. El resultado que nos facilita este tipo de estudios es, básicamente, una estimación del riesgo de padecer una determinada afección asociado a un factor de exposición. La medida se denomina comúnmente con el anglicismo odds ratio (OR), probabilidades relativas y desigualdad relativa). En casos de enfermedades de baja prevalencia, la OR y el riesgo relativo dan resultados bastante similares. (3)

Características de los estudios de casos y controles frente a los estudios observacionales

Limitaciones

Recogida retrospectiva de la información

Dificultad en establecer la secuencia temporal entre la enfermedad y la exposición

Conseguir controles representativos de la población sin la afección y que sean lo más parecido posible a los casos

No se puede estudiar más de una enfermedad.

No sirven para el cálculo de incidencia o prevalencia

ESTUDIOS TRANSVERSALES

Un estudio transversal o de prevalencia estudia simultáneamente la exposición y la enfermedad en una población bien definida en un momento determinado.

Aunque permite examinar asociaciones entre una afección y múltiples variables, la medición simultánea no permite conocer la secuencia temporal y no es, por tanto, posible determinar causalidad en las relaciones. Se puede incluir diversas afecciones en

un mismo estudio. Además, resultan útiles en el caso de afecciones o factores de riesgo de larga duración.

Son de los estudios más utilizados por su sencillez y bajo costo. Prácticamente, muchos de los estudios experimentales o analíticos han tenido su origen en estos diseños. En ocasiones su utilidad radica en poder valorar necesidades preventivas, curativas y rehabilitadoras de la comunidad.

También han resultado útiles para la evaluación de programas sanitarios. Pueden encontrarse con otras denominaciones: estudio de prevalencia, estudio de morbilidad o encuesta de salud.

Limitaciones

Imposibilidad de establecer la secuencia temporal entre la enfermedad y el factor de riesgo

La ambigüedad temporal hace que la definición y medición de la exposición sean dificultosas

No permite concluir causalidad

Muy susceptibles a sesgos de información y selección.

No son útiles en enfermedades raras ni de corta duración

ESTUDIOS DE PRUEBAS DIAGNOSTICAS

Estudio descriptivo que tiene como objetivo estimar la capacidad de una prueba diagnóstica para discriminar entre las personas que presentan una condición y las que no. Se requiere la existencia de una prueba de referencia reconocida.

Tras la confrontación de la prueba de referencia con la diagnóstica en estudio, se calculan las siguientes probabilidades:

Sensibilidad: proporción de individuos con la característica que presentan un resultado positivo, indica la utilidad para identificar a los pacientes con la

característica (ideal para pruebas de tamizaje)

Especificidad: proporción de individuos sin la característica que presentan un resultado negativo; indica la utilidad para identificar a los pacientes sin la característica (pruebas confirmatorias)

Especificidad: proporción de individuos sin la característica que presentan un resultado negativo; indica la utilidad para identificar a los pacientes sin la característica (pruebas confirmatorias)

Valor predictivo de un resultado positivo (VPP): probabilidad de que un paciente tenga la característica si la prueba resulta positiva.

Valor predictivo de un resultado negativo (VPN): probabilidad de que un paciente no tenga la característica si la prueba resulta negativa.

Sensibilidad y especificidad son características intrínsecas de la prueba que no se afectan por la prevalencia. Dado que tanto sensibilidad como especificidad varían en función del criterio elegido como punto de corte, una forma más global consiste en el uso de curvas de características operativas del receptor. La evolución del área bajo dicha curva es el mejor indicador de la capacidad predictiva del test, independientemente de la prevalencia de la enfermedad en la población de referencia y en base al cual se podrán establecer comparaciones entre diferentes pruebas diagnósticas. (1)

VPP y VPN se afectan mucho por la prevalencia de la enfermedad en estudio (si disminuye la prevalencia, disminuye el VPP y aumenta el VPN, y viceversa)

REVISIÓN SISTEMÁTICA/METAANÁLISIS

Una revisión sistemática es un estudio que, basado en métodos explícitos, exhaustivos y reproducibles, sintetiza los resultados de otros estudios primarios. Aunque se utiliza como sinónimo de meta análisis, ciertos autores prefieren limitar este último

término al procedimiento estadístico y no al proceso global.

Algunos autores lo sitúan en la cúspide de una hipotética pirámide de la evidencia, por encima del ensayo clínico aleatorizado y controlado. No obstante, dada su absoluta dependencia en la calidad de los estudios en los que se basa, sus potenciales sesgos, y también la proliferación incontrolada y la generalización de resultados a partir de metaanálisis de integridad dudosa, su generalización como herramienta de referencia es peligrosa y en muchos casos no está justificada.

Entre sus ventajas cuentan la ganancia en precisión y la potencia estadística (útil para explorar subgrupos y generar nuevas hipótesis), permitir concluir ante aparentes divergencias de resultados, y todo ello con un método trazable y objetivo. Por la parte negativa, tenemos los sesgos y el tratamiento de la heterogeneidad de los resultados.

El sesgo más importante radica en la identificación e inclusión de los estudios, fundamentado en los sesgos de publicación (se tiende a no publicar resultados negativos), de idioma (se publican en inglés los resultados más favorables y en idiomas locales los menos), de citación (se citan más los resultados positivos, por tanto son más fácilmente identificables), de publicación múltiple, de calidad deficiente (faltan datos para poder metaanalizar en publicaciones de peor calidad), y de los criterios de

selección de los estudios, entre otros.
(10)

CRITERIOS DE CALIDAD EN UN ENSAYO CLÍNICO

Existen al menos 25 escalas y nueve listas o “checklists” que se han utilizado para evaluar la validez y la calidad de los ensayos clínicos. Estas escalas y listas incluyen entre 3 y 57 elementos. Casi todos los elementos se basan en criterios sugeridos o generalmente aceptados que se mencionan en los libros de texto.

Una de las escalas más utilizadas, debido a su sencillez, para evaluar la calidad de los ensayos clínicos es la de Jadad. (4) La escala de Jadad, puntúa los ensayos clínicos de 0 a 5 en función de que cumplan cada uno de los siguientes criterios:

- Si el estudio fue aleatorizado.
- Si el estudio fue doble ciego
- Si están descritas las pérdidas y las retiradas

Si se describió y es adecuado el método para generar la secuencia de aleatorización

Si las condiciones de enmascaramiento son adecuadas

Se considera que el ensayo es de pobre calidad si su puntuación es inferior a 3.

Muchos de los instrumentos de medida de la calidad tienden a confundir la calidad de los informes con la validez del diseño y la realización de un ensayo clínico. No es infrecuente que durante la lectura del informe o publicación de un

ensayo clínico se encuentre falta de información relevante para valorar tanto la calidad metodológica como la importancia de los resultados.

La declaración CONSORT (6) es un conjunto de recomendaciones que pretende uniformar a la información que deben incluir todas las publicaciones de ensayos clínicos. Las Recomendaciones de las Normas Consolidadas para los Informes de los Ensayos Clínicos (CONSORT), incluyen una lista de comprobación y un diagrama de flujo. La lista de comprobación incluye 21 puntos que se refieren principalmente a los métodos, resultados y discusión del informe de un ensayo clínico, identificando las claves de la información necesaria para poder evaluar la validez interna y externa del informe. El diagrama de flujo del ensayo permite extraer información sobre el progreso de los pacientes que participaron en el mismo.

El grupo de trabajo en Medicina Basada en la Evidencia de la universidad canadiense de McMaster, promotor de la metodología de la Medicina Basada en la Evidencia, publicó desde 1993 en la revista JAMA una serie de artículos que bajo el título genérico de “Guías para usuarios de la literatura médica” presenta un conjunto de guiones para ayudar a los clínicos a leer distintos tipos de artículos científicos.(7)

Siguiendo estas recomendaciones, otros grupos como la iniciativa CASP (Critical Appraisal Skills Programme, programa de habilidades en lectura

crítica por sus siglas en inglés) han creado herramientas para facilitar la lectura de artículos científicos. (5) La lectura crítica de un ensayo clínico debe incluir:

Valoración de la validez ¿son válidos los resultados del ensayo clínico?

Clara definición de los objetivos. ¿Se orienta el ensayo a una pregunta claramente definida?

Asignación aleatoria ¿Fue aleatoria la asignación de los pacientes al tratamiento? ¿Se ocultó la asignación aleatoria?

Seguimiento adecuado y análisis por intención a tratar. ¿Se tuvieron en cuenta tanto los médicos como los pacientes, durante el estudio?

Diseño ciego. ¿Se mantuvieron ciegos respecto al tratamiento, tanto los médicos como los pacientes, durante el estudio?

Similitud de las características basales de los grupos ¿Eran similares los grupos al inicio del ensayo?

Igualdad de tratamiento de los distintos grupos (salvo en la intervención estudiada) ¿Fueron tratados todos los grupos de forma similar, aparte de la intervención experimental?

Valoración de la importancia. ¿Cuáles son los resultados?

Magnitud del efecto ¿Cuál es la magnitud del efecto del tratamiento?

Precisión de los resultados ¿Cuan precisos son los resultados?

Valoración de la aplicabilidad ¿Son aplicables los resultados del ensayo?

Reproducibilidad de los resultados ¿Son aplicables los resultados a mi paciente?

Análisis de todos los resultados clínicamente importantes ¿Se han considerado todos los resultados clínicamente importantes?

Relación entre beneficios, perjuicios y costos. ¿Merecen la pena los beneficios frente a los perjuicios y costos?

VALORACIÓN DE LA VALIDEZ ¿SON VÁLIDOS LOS RESULTADOS DEL ENSAYO CLÍNICO?

Por valoración de la validez se entiende la comprobación de la validez interna del ensayo clínico que puede definirse como la adecuación metodológica de su diseño y desarrollo que garantice que los resultados no estén sesgados.

Los tres aspectos esenciales para analizar la validez interna de un ensayo clínico son:

- Clara definición de los objetivos
- Asignación aleatoria
- Seguimiento adecuado y análisis por intención de tratar

Otros aspectos secundarios o adicionales incluyen:

- El diseño ciego
- La similitud de las características basales de los grupos
- Igualdad de tratamiento de los distintos grupos

Clara definición de los objetivos. ¿Se orienta el ensayo a una

pregunta claramente definida?: El ensayo clínico debe definir claramente sus objetivos definiendo explícitamente

Las características de los pacientes incluidos

- La intervención estudiada
- La intervención o intervenciones frente a las que se compara
- Las variables mediante las que se evalúa su eficacia

Así por ejemplo en el ensayo CAPPP (8) el objetivo fue determinar si el tratamiento con captopril como tratamiento de base era superior a la terapia convencional: diuréticos o betabloqueadores. Se definen:

Pacientes: entre 25 y 65 años, HTA esencial conocida y tratada previamente o no. PAM superior a 100 mmHg en dos ocasiones diferentes.

Intervención: captopril

Intervención frente a la que se compara: betabloqueadores o diuréticos.

Variable resultado: Variable combinada IAM mortal o no, ECV y otras causas de muerte cardiovascular

Asignación aleatoria. ¿Fue aleatoria la asignación de los pacientes al tratamiento? ¿Se ocultó la asignación aleatoria?: Un aspecto fundamental cuando se valoran ensayos clínicos es comprobar que se trata de un ensayo clínico aleatorio. En un ensayo clínico aleatorio la asignación de los pacientes a los distintos grupos de tratamiento se realiza al azar, con

algún sistema similar al de lanzar una moneda.

La aleatorización se define como un procedimiento sistemático y reproducible por el que los sujetos participantes en un ensayo clínico son distribuidos en los distintos grupos de tratamiento. La asignación aleatoria implica que todos los participantes en el ensayo tienen la misma probabilidad de entrar en cualquiera de los brazos del estudio. No son sistemas aleatorios la asignación de los pacientes por su fecha de nacimiento, por el centro en el que son atendidos, etc.

La importancia de la asignación aleatoria radica en que este diseño de investigación es el que más se aproxima a la creación de grupos de pacientes similares. No sólo respecto de los factores que se sabe que pueden condicionar el resultado, tales como la gravedad de la enfermedad, sino también respecto de otros factores cuya influencia sea desconocida. Los estudios en los que la asignación de los tratamientos no es aleatoria tiende a mostrar mayores efectos del tratamiento en ocasiones falsamente positivos.

La ocultación o cegamiento de la aleatorización no debe confundirse con el cegamiento del tratamiento (diseño abierto, simple o doble ciego). La ocultación o cegamiento de la aleatorización consiste en que el investigador desconoce la secuencia de aleatorización, en otras palabras desconoce a qué grupo será asignado el próximo paciente que incluya en el ensayo clínico. Esto resulta especialmente

importante en los estudios abiertos o simple ciego en los que el investigador conoce cuál es el tratamiento que corresponde a cada grupo. Ya que de forma consciente o inconsciente puede influir en la inclusión o exclusión del paciente en el ensayo en función del tratamiento que debería de recibir. Para ello el investigador debe conocer el grupo al que es asignado el paciente sólo cuando éste ya ha aceptado su participación y firmado el consentimiento informado.

Volvamos a tomar como ejemplo es estudio CAPPP: “La secuencia de aleatorización fue generada por ordenador y comunicada a los investigadores por medio de sobres numerados sellados, uno por cada paciente, con instrucciones de usarlos por orden numérico”.

Seguimiento adecuado y análisis por intención de tratar. ¿Se tuvieron en cuenta todos los pacientes en el momento del análisis?: La verificación del correcto análisis de todos los pacientes participantes en el estudio incluye la comprobación de un seguimiento completo y un análisis de los resultados por intención de tratar.

¿Fue completo el seguimiento?: Durante el desarrollo de los ensayos clínicos se producen siempre pérdidas de pacientes. Sin embargo todos los pacientes incluidos en un ensayo clínico deben ser tenidos en cuenta a la conclusión del mismo. Se conocen como pérdidas post-aleatorización a aquellos pacientes que abandonan o son retirados del estudio una vez que se les ha

asignado el tratamiento. Las pérdidas post-aleatorización incluyen:

Los abandonos (drop out) son los pacientes incluidos en un ensayo clínico que no quieren o no pueden continuar en el estudio ni en el seguimiento.

Las retiradas (withdrawal) son los pacientes que son obligados a abandonar el estudio por presentar determinadas circunstancias especificadas en el protocolo (como la aparición de nuevos criterios de exclusión, reacciones adversas graves, etc.) pero en los que si es factible su seguimiento.

Pérdida de seguimiento (loss to follow-up) son los pacientes incluidos en el estudio, que no realizan los controles indicados y de los que no es posible obtener datos de su seguimiento.

Desviación del protocolo: son los pacientes que por distintas circunstancias no siguen el protocolo asignado (incumplimientos terapéuticos, cambio de brazo por error).

Todas las pérdidas post-aleatorización (abandonos, retiradas, pérdidas de seguimiento) deben aparecer reflejadas en el informe del ensayo clínico.

¿Se analizaron en el grupo en que fueron aleatoriamente distribuidos?: En los ensayos clínicos, como en la práctica diaria, se producen con cierta frecuencia desviaciones del protocolo, cuando un paciente que fue asignado aleatoriamente a recibir un tratamiento, deja de tomarlo, lo sigue parcialmente o recibe en realidad, por distintos motivos, el

tratamiento de otro grupo. Ante estas circunstancias cabrían teóricamente tres posibilidades:

Excluir al paciente del análisis final,

Analizarlo en el grupo del tratamiento que realmente realizó, aunque fuera asignado a otro (análisis por tratamiento)

Analizarlo en el grupo de tratamiento al que fue asignado aleatoriamente, aunque realmente no lo recibiera (análisis por intención a tratar).

Aunque pueda resultar paradójico, el enfoque correcto es analizar al paciente en el grupo al que fue aleatoriamente asignado.

El análisis por intención de tratar es la forma de análisis de los resultados que incluye a todos los pacientes que han sido inicialmente asignados a cada grupo de tratamiento independientemente de que completaran o no el período de tratamiento y/o seguimiento. El análisis por intención a tratar tiene dos ventajas:

Se conservan las propiedades conseguidas con la asignación aleatoria, la homogeneidad de los grupos.

Se aproxima a la realidad de la práctica clínica (efectividad), donde los cambios o incumplimiento de los tratamientos son habituales.

Diseño ciego. ¿Se mantuvieron “ciegos” respecto al tratamiento, tanto los médicos como los pacientes, durante el estudio?: El enmascaramiento o cegamiento del tratamiento consiste en una serie de medidas o precauciones que se

toman con el fin de que a lo largo del estudio, ya sea el paciente, el investigador o ambos, desconozcan el tratamiento que se está aplicando. Los tipos de enmascaramiento pueden ser:

Simple ciego: son aquellos ensayos en los que el sujeto desconoce el grupo de tratamiento al que pertenece. (Aunque en ocasiones puede ser el investigador el que desconoce la asignación de tratamientos)

Doble ciego: son aquellos ensayos en los que tanto el sujeto como el investigador desconocen la asignación a los grupos de tratamiento.

Igualdad de tratamiento de los distintos grupos. ¿Fueron tratados todos los grupos de forma similar?: Se debe comprobar que todos los pacientes incluidos en el ensayo han sido tratados de forma similar, salvo en la intervención a estudiar, en cuanto a estudios, controles y coin-tervenciones. Si los pacientes de alguno de los grupos recibieron mayor atención, fueron visitados con mayor frecuencia, o recibieron distintos tratamientos complementarios, los resultados pueden verse sesgado.

Por ejemplo, en el estudio CAPP para conseguir los valores adecuados de presión arterial (diastólica menos o igual a 90 mmHg) además de incrementar la dosis de captopril, betabloqueadores o diuréticos, “se podría añadir al régimen de tratamiento un

antagonista del calcio en ambos grupos”.

VALORACIÓN DE LA IMPORTANCIA ¿CUÁLES SON LOS RESULTADOS?

- La importancia clínica de un tratamiento se valora mediante:
- La magnitud del efecto hallado
- La precisión de los resultados

Magnitud del efecto ¿Cuál es la magnitud del efecto del tratamiento?: La magnitud del efecto del tratamiento se puede expresar según los casos mediante:

El riesgo relativo, la reducción relativa del riesgo, la reducción absoluta del riesgo, el número de pacientes necesario a tratar o la odds ratio para el caso de variables binarias (presencia o ausencia de un evento).

La diferencia de medias de medianas, etc., en el caso de variables continuas.

Como ejemplo en el estudio CAPPP: En la variable resultado principal definida como IAM fatal o no fatal, ECV y otras causas de muerte cardiovascular, el riesgo relativo fue de 1.05. Esto significa que el riesgo de presentar cualquiera de los eventos de la variable principal fue de 1.05 veces mayor en el grupo captopril que en el grupo control de tratamiento convencional (diuréticos o betabloqueadores). Un riesgo relativo de 1, indica la igualdad de riesgo en ambos grupos.

Precisión de los resultados ¿Cuán precisos son los resultados?: Los ensayos clínicos al trabajar con

muestras limitadas de población obtienen estimaciones sobre el efecto de los tratamientos. Un ensayo clínico con una muestra de pacientes determinada obtiene una estimación puntual de cuál es el efecto del tratamiento en esos pacientes concretos, por lo que interesa conocer cuál será la estimación del efecto del tratamiento en la población general. Por ello, además de conocer la magnitud del efecto, es necesario conocer la precisión de esa estimación a través de los intervalos de confianza.

Intervalo de confianza es el intervalo dentro del que se encuentra la verdadera magnitud del efecto con un grado prefijado de seguridad. El grado de seguridad que se emplea habitualmente es el 95%. Un intervalo de confianza del 95% es aquel en el que el 95% de los casos, o con una seguridad del 95%, el verdadero valor de la magnitud del efecto se encuentra dentro de ese intervalo. Cuanto más estrecho sea el intervalo de confianza, mayor es la precisión con la que se estima el efecto. Cuando el intervalo de confianza del riesgo relativo incluye al 1, indica que no se han demostrado diferencias entre el grupo tratamiento y el grupo control

Tomemos nuevamente como ejemplo el estudio CAPPP. El riesgo relativo de la variable principal fue de 1.05; con un intervalo de confianza del 95% entre 0.9 y 1.22. El verdadero valor del riesgo estimado se sitúa con una probabilidad del 95% entre un ligero beneficio para el tratamiento con captopril y un ligero beneficio para el tratamiento

convencional (betabloqueadores o diuréticos).

VALORACIÓN DE LA APLICABILIDAD ¿SON APLICABLES LOS RESULTADOS DEL ENSAYO?

La valoración de la aplicabilidad de los resultados de un ensayo clínico al paciente concreto que motivó nuestra pregunta debe incluir:

- La reproducibilidad de los resultados,
- El análisis de todas las variables clínicamente importantes,
- La relación entre los beneficios y los perjuicios y costos.

Reproducibilidad de los resultados: El ensayo clínico controlado aleatorio es el mejor diseño para demostrar y comparar la eficacia de los tratamientos. Tienen una gran validez interna, pero las características ideales en las que se desarrollan limitan su aplicabilidad o validez externa.

Una forma de valorar la aplicabilidad de los resultados de un ensayo al paciente individual es comprobar si sus características se ajustan a las de los pacientes incluidos en el estudio. Si nuestro paciente hubiera podido formar parte del ensayo clínico, si cumple todos los criterios de inclusión y ninguno de exclusión, los resultados podrían ser directamente aplicables. De no ser así, se debe evaluar si las diferencias influirían de forma decisiva en los resultados.

Análisis de todos los resultados clínicamente importantes: Se debe considerar si en ensayo analiza todos los resultados clínicamente importantes incluidos los efectos secundarios. El ensayo clínico debe medir los resultados utilizando variables clínicamente importantes. Por ejemplo, en el caso de ensayos sobre tratamiento antihipertensivo son variables clínicamente importantes: la mortalidad cardiovascular o los eventos cardiovasculares. Variables como la tensión arterial o la hipertrofia ventricular son medidas intermedias y tienen una escasa significación para el paciente.

Relación entre beneficios, perjuicios y costos: para decidir la aplicabilidad de los resultados de un ensayo clínico debe realizarse un análisis final de los beneficios que aporta el tratamiento frente a los perjuicios y costos.

REFERENCIAS BIBLIOGRÁFICAS:

1. Laporte JR. Principios básicos de investigación clínica. Madrid: Ergon, 1993.
2. Guyatt G, Rennie D. Guías para usuarios de literatura médica. Manual para la práctica clínica basada en la evidencia. Barcelona: Ars Medica, 2004
3. Pita Fernández, S. Estudios experimentales en la práctica clínica. Investigación terapéutica. Ensayos clínicos. En: Gómez de la Cámara, A. ed. Manual de Medicina Basada en la Evidencia. Elementos para su desarrollo y aplicación en Atención Primaria. Madrid: Jarpyo Editores; 1998. p. 147-163

4. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996;17:1-12
5. Programa de lectura crítica CASP-españa. 11 preguntas para dar sentido a un ensayo clínico.
6. Consort Statement
Fisterra.com. Pilar Gayoso Diz. Lectura crítica de un artículo sobre tratamiento.
7. Hansson L, Lindholm L, Niskanen L, Lanke J, Hedner T, Nikiason A et al. Effect of angiotensin-converting-enzyme inhibition compared with conventional therapy on cardiovascular morbidity and mortality in hypertension: the Captopril Prevention Project (CAPPP) randomised trial. *Lancet* 1999; 353: 611-616.
8. Doval H. Tajer C. Como leer un ensayo clínico para incorporarlo a nuestro conocimiento médico y aplicarlo en la práctica.
9. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. Evidence based Medicine: How to practice and teach EBM. 2nd ed. Edinburgh: Churchill Livingstone; 2000.