



Artículo de revisión

Ensayos clínicos: conceptos generales e interpretación

Pablo Castaño González^a y José A. Gómez-Puerta^{b,*}

^a Artmédica, Medellín, Antioquia, Colombia

^b Servicio de Reumatología, Hospital Clínic, Barcelona, España

INFORMACIÓN DEL ARTÍCULO

Historia del artículo:

Recibido el 16 de noviembre de 2018

Aceptado el 9 de noviembre de 2021

On-line el 25 de enero de 2022

Palabras clave:

Ensayo clínico

Aleatorización

Estudios controlados

Keywords:

Clinical trial

Randomisation

Controlled studies

R E S U M E N

Existen diferentes tipos de diseños de ensayos clínicos que sirven a propósitos específicos y dependerán tanto de la naturaleza de la enfermedad reumatológica como de las características de las intervenciones terapéuticas investigadas y de la pregunta que se desea responder con el estudio. En la presente revisión se describen los principales tipos de ensayos clínicos que se realizan en el área de la reumatología, incluyendo sus aplicaciones y limitaciones.

© 2021 Asociación Colombiana de Reumatología. Publicado por Elsevier España, S.L.U.

Todos los derechos reservados.

Clinical trials: General concepts and interpretation

A B S T R A C T

There are different types of clinical trial designs that serve specific purposes, which will depend on the nature of the rheumatological disease, the characteristics of the therapeutic interventions being investigated and the question the study is intended to answer. This review describes the main types of clinical trials conducted in the field of rheumatology, including their applications and limitations.

© 2021 Asociación Colombiana de Reumatología. Published by Elsevier España, S.L.U. All rights reserved.

Introducción

Con la aparición de las terapias biológicas hace ya dos décadas, en la literatura de la especialidad se ha incrementado la

presencia de ensayos clínicos que evalúan la eficacia o la seguridad de estas y otras estrategias terapéuticas; por tanto, es importante para los reumatólogos conocer conceptos básicos sobre los tipos de ensayos clínicos al momento de interpretarlos, diseñarlos o participar en su ejecución.

Dentro de las diferentes modalidades de ensayos clínicos se incluyen: los de un solo brazo, los controlados con placebo, los ensayos con grupos cruzados (*crossover trials*), los ensayos factoriales, los ensayos de no inferioridad y los estudios

* Autor para correspondencia.

Correo electrónico: jagomez@clinic.cat (J.A. Gómez-Puerta).

<https://doi.org/10.1016/j.rcreu.2021.11.004>

0121-8123/© 2021 Asociación Colombiana de Reumatología. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

diseñados para evaluar nuevos dispositivos de diagnóstico, entre otros.

La selección del tipo de estudio depende de la pregunta clínica que se pretenda contestar, de las características de la enfermedad y de las intervenciones terapéuticas que se evalúen, de los desenlaces elegidos, de la disponibilidad de un grupo control y del tiempo y los recursos económicos con los que se cuente para su desarrollo. A continuación se describen las diferentes fases y las características de los principales diseños de ensayos clínicos usados en reumatología.

Fases de los ensayos clínicos

Cuando los ensayos clínicos evalúan el efecto de una nueva intervención terapéutica que pretende recibir aprobación de las agencias reguladoras de medicamentos y dispositivos médicos, se usa una clasificación por fases de la investigación en humanos. Los estudios fase I suelen involucrar menos de 100 participantes (generalmente voluntarios sanos) y su objetivo es investigar la seguridad, la tolerabilidad, la farmacocinética y la farmacodinámica. Los estudios fase II generalmente se hacen con menos de 300 individuos con una enfermedad específica y buscan evaluar el efecto biológico o clínico de la dosis segura encontrada en la fase I y continuar investigando su perfil de seguridad. Los estudios fase III suelen ser ensayos clínicos aleatorizados multicéntricos, con un gran número de pacientes con una patología determinada (entre 300 y miles), cuyo objetivo es confirmar la eficacia y la seguridad con relación a la terapia estándar para esa condición (o el placebo, si no existe tal terapia) y así conseguir aprobación para su comercialización. Los estudios fase IV o de vigilancia poscomercialización pretenden continuar evaluando la seguridad del medicamento o del dispositivo durante un periodo más largo y en un mayor número de individuos que en los estudios fase III¹.

Ensayos de un solo brazo

En este tipo de ensayo, una muestra de individuos con una enfermedad específica se expone a una terapia determinada y se sigue durante un periodo para evaluar la respuesta a esta, es decir, todos los participantes reciben el tratamiento experimental.

Esta clase de estudio permite obtener datos preliminares sobre la eficacia de una intervención experimental y, a pesar de que es el más simple de diseñar, la interpretación de los resultados debe ser cautelosa, porque es difícil saber si existe un *efecto placebo de la terapia utilizada*, si la respuesta obtenida se debe a los cambios propios de la historia natural de la enfermedad, así como establecer la magnitud de la respuesta sin un grupo control para comparar. Debido a estas limitaciones, los ensayos con una sola rama deben llevarse a cabo en enfermedades bien seleccionadas, en las cuales el efecto placebo sea nulo o mínimo, la historia natural sea conocida y que no sea posible o sea difícil o poco ético utilizar un brazo control. Su principal aplicación es en enfermedades raras o huérfanas, en las que el número de pacientes es escaso; en cambio, dichos ensayos no son útiles

en pacientes con osteoartritis, fibromialgia y otras condiciones de dolor crónico en las que existen altas tasas de efecto placebo².

Por ejemplo, en un ensayo clínico de un solo brazo se investigó la seguridad y la eficacia de canakinumab (anticuerpo monoclonal humano anti-IL-1 β) en el síndrome de Schnitzler, una enfermedad autoinflamatoria rara caracterizada por urticaria crónica, gammapatía monoclonal y manifestaciones clínicas de inflamación sistémica como fiebre intermitente, artralgias o artritis, dolor óseo, adenopatías y hepatoesplenomegalia, y en la que existe mayor riesgo de desarrollar linfomas y amiloidosis AA. En dicho ensayo se incluyeron 8 pacientes que recibieron canakinumab a una dosis de 150 mg subcutáneos cada mes, por 6 dosis, y fueron seguidos por 9 meses. El desenlace primario fue la remisión completa al día 14, lo que alcanzaron todos los pacientes; 3 meses después de suspender la terapia con canakinumab, 4 pacientes presentaron recaídas, mientras que 2 persistieron en remisión tras 7 meses de la última dosis³.

Ensayos controlados con placebo

En esta clase de estudios se selecciona una muestra de individuos con una enfermedad específica que son aleatorizados a 2 o más grupos: uno o más que reciben un tratamiento experimental (puede ser el mismo a diferentes dosis o diferentes tratamientos) y otro que recibe placebo. El individuo aleatorizado recibe solo un tipo de tratamiento durante toda la duración del ensayo. Posteriormente, los individuos se siguen durante el estudio y se comparan los resultados⁴.

Muchos ensayos clínicos en reumatología utilizan este diseño. Un ejemplo es el ensayo clínico fase II aleatorizado, controlado con placebo, multicéntrico, doble ciego, en el que se investigó la eficacia y la seguridad de apremilast (inhibidor oral de la fosfodiesterasa 4) en 111 pacientes con enfermedad de Behçet con 2 o más úlceras orales activas que fueron aleatorizados a 2 grupos: apremilast 30 mg cada 12 horas y placebo oral cada 12 horas; la duración fue de 12 semanas, y después de este periodo hubo una fase de extensión del tratamiento activo (fase en la cual los pacientes que recibieron placebo pasaron a recibir tratamiento con apremilast) durante 12 semanas. El desenlace primario fue el número de úlceras tras las primeras 12 semanas, y los resultados favorecieron al apremilast, lo que motivó la realización de un ensayo clínico fase III: el estudio Relief, que incluyó un mayor número de pacientes (207) y una fase de extensión de tratamiento activo más larga (52 semanas), en la que se confirmaron los beneficios de este medicamento en el tratamiento de las úlceras orales y genitales.

Los ensayos controlados con placebo en los cuales las diversas intervenciones hayan sido asignadas de forma aleatoria permiten controlar tanto por factores de confusión conocidos como por desconocidos, tras un análisis por intención de tratar (ITT).

En el análisis ITT los pacientes son analizados de acuerdo con la intervención a la que han sido asignados inicialmente (paciente aleatorizado, paciente analizado, con independencia o no de recibir la intervención). En el caso de que las muestras queden bien balanceadas tras la aleatorización (suelen quedar así en la mayoría de los ensayos clínicos), las comparaciones

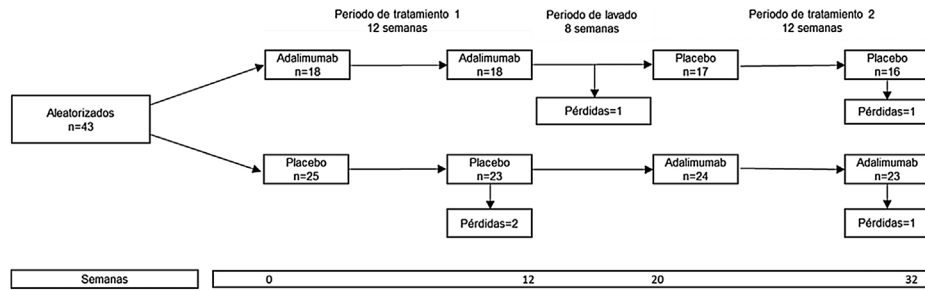


Figura 1 – El estudio HUMOR ilustra las características de un ensayo clínico cruzado: la asignación aleatoria a ADA y placebo en el periodo 1, un periodo de lavado, y el cruzamiento de los grupos en el periodo 2.

Fuente: modificado de Aitken et al.⁸.

realizadas entre ambas intervenciones (terapia experimental vs. placebo) reflejarán de una forma más precisa la magnitud del efecto de la medicación o la intervención sobre la enfermedad seleccionada para el ensayo clínico⁵. Es preciso tener en cuenta que en algunos estudios se hacen aleatorizaciones estratificadas, basadas generalmente en conocimientos de estudios previos, para balancear las muestras o para seleccionar el grupo que se va a estudiar (p. ej., incluir 3 pacientes con artritis reumatoide seropositiva por cada paciente con artritis reumatoide seronegativa, un paciente hombre con lupus eritematoso sistémico [LES] por cada 9 mujeres, etc.)⁶.

Por otra parte, en el análisis por protocolo solo se incluyen aquellos pacientes que recibieron al menos una dosis de la intervención o tratamiento y suelen ser utilizados para valorar la seguridad de un medicamento o dispositivo; por esta razón pueden verse afectados por las pérdidas de pacientes durante el seguimiento.

La desventaja de estos ensayos en paralelo (ensayos en los que cada paciente recibe durante todo el estudio la intervención a la que fue asignado inicialmente) consiste en que requieren un tamaño muestral grande, debido a la variabilidad en la respuesta en cada paciente y de cada tratamiento; si se requiere detectar pequeñas diferencias, la muestra debe ser mayor.

Ensayos con grupos cruzados (crossover)

En este tipo de diseño cada participante es aleatorizado a una secuencia determinada de tratamientos que serán administrados durante periodos de tratamiento específicos, separados por un periodo de lavado del efecto de la intervención antesora (*wash out*), es decir, cada individuo actúa como su propio control porque recibe de manera consecutiva cada uno de los tratamientos del estudio en momentos diferentes. Por ejemplo, en un ensayo cruzado 2 × 2 cada participante es aleatorizado a una de dos secuencias de tratamientos: A y luego B, o B y luego A. El periodo de lavado, también conocido como de estabilización, tiene por objetivo eliminar el probable efecto residual después de la suspensión del primer tratamiento administrado y, en caso de que las estrategias en comparación sean fármacos, se recomienda que sea al menos 5 veces la vida media del medicamento del estudio que tenga la vida media más larga (no es útil en terapias con efectos terapéuticos residuales largos como el rituximab o el ácido zoledrónico).

Puesto que cada sujeto sirve como su propio control, estos estudios tienen diferentes ventajas en comparación con los diseños en paralelo (aquellos en los cuales cada participante recibe solamente una de las intervenciones en estudio): la reducción del tamaño muestral, una menor variabilidad interindividual y una mayor probabilidad de reclutar pacientes, ya que cada participante recibirá en algún momento al menos un tratamiento activo. Las principales limitaciones que tienen son: un mayor riesgo de pérdidas durante el seguimiento debido a la exposición de cada individuo a 2 o más terapias, lo que implica una mayor duración del estudio y una mayor probabilidad de efectos secundarios y muerte; el efecto residual: persistencia en un periodo del efecto de la intervención del periodo anterior; el efecto secuencia: el orden de las intervenciones afecta el resultado final y solo el efecto de la primera puede ser evaluado de forma fiable; el efecto periodo: las características basales de un paciente cambian durante el estudio, por lo que durante el segundo o los subsecuentes periodos no son similares a las del inicio; la imposibilidad de evaluar la eficacia y la seguridad a largo plazo. Por tales razones, este diseño resulta útil cuando se investigan terapias con un efecto rápidamente reversible tras su suspensión o una vida media corta, con un inicio de efecto rápido que permita periodos de tratamiento cortos y en enfermedades crónicas con manifestaciones clínicas relativamente estables en el tiempo y bajo riesgo de pérdidas (p. ej., sería poco útil en el LES)⁷.

Un ejemplo de ensayo clínico con grupos cruzados es el estudio HUMOR (*HUMira for erosive hand Osteoarthritis*), en el que se evaluó la eficacia de adalimumab (ADA) en 51 pacientes mayores de 50 años que cumplían criterios del Colegio Americano de Reumatología para osteoartritis de manos y que tenían al menos una erosión por radiografía convencional y sinovitis por resonancia magnética⁸. Cada participante fue aleatorizado a recibir ADA 40 mg subcutáneos cada 2 semanas o placebo subcutáneo por 12 semanas, seguido por un periodo de lavado de 8 semanas y luego cruzado al otro tratamiento por otras 12 semanas. Expresado de otra manera, cada paciente fue aleatorizado a dos grupos o secuencias: el grupo 1, que recibió el tratamiento activo y luego placebo, y el grupo 2, que recibió placebo y luego el tratamiento activo (fig. 1). El desenlace primario fue el cambio en la escala visual análoga de dolor a las 12 semanas; los resultados fueron negativos porque no hubo diferencia significativa entre los grupos.

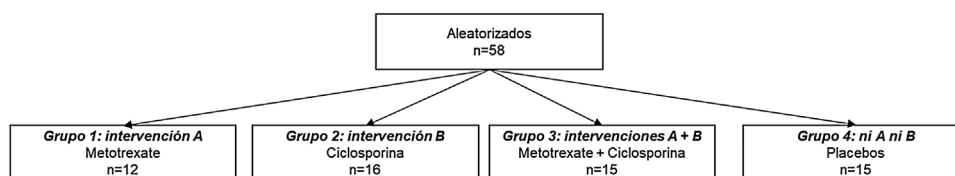


Figura 2 – El estudio SELAM es un ejemplo de los elementos de un ensayo clínico factorial. En el grupo 1 de la figura se usaron las dos intervenciones (MTX y ciclosporina), en los grupos 2 y 3 una sola de las intervenciones (MTX y ciclosporina, respectivamente) y en el grupo 4 ninguna de ellas (placebos orales).

Fuente: tomado de Ibrahim et al.⁹.

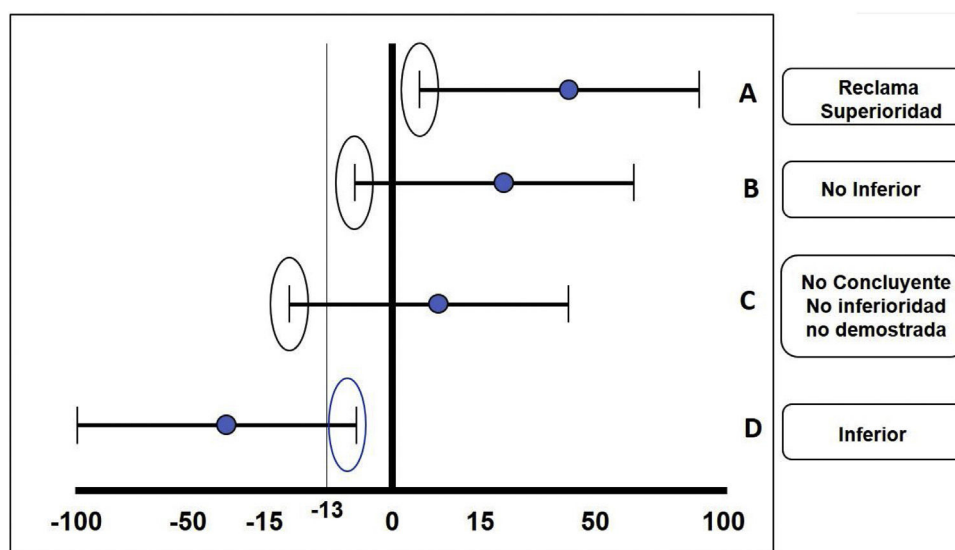


Figura 3 – Se ilustran los diferentes resultados posibles en el caso de un ensayo de inferioridad: A) Reclama superioridad. B) No inferioridad. C) No concluyente. D) Inferioridad. En este caso se toma como ejemplo un margen inferior de -13 .

Fuente: basado en el diseño del estudio ORAL Strategy¹¹.

Ensayos clínicos factoriales

En este tipo de ensayos clínico se evalúan simultáneamente 2 o más intervenciones experimentales, cada una por separado y en combinación. En el diseño más sencillo (2×2) existirían 2 tratamientos o factores: A y B, y las ramas del ensayo clínico serían 4: A solo, B solo, A + B y ninguno de ellos. La mayor eficiencia de estos estudios se presenta cuando se asume que los tratamientos en estudio no tienen interacción significativa, es decir, el efecto del tratamiento A no depende de que se administre el tratamiento B, y viceversa; no obstante, estos estudios también pueden emplearse cuando se desea estudiar específicamente si existe interacción entre los tratamientos.

El estudio SELAM⁹ fue un ensayo clínico factorial controlado con placebo en el que se investigó la eficacia de dos tratamientos: ciclosporina (A) y metotrexato (MTX) (B), durante 56 semanas, en 58 adultos con dermatomiositis o polimiositis con respuesta incompleta a glucocorticoides. Había 4 ramas en el estudio: MTX 15-25 mg/semana + glucocorticoides (A), ciclosporina 1-5 mg/kg/día + glucocorticoides (B), MTX + ciclosporina + glucocorticoides (A+B) y placebo de MTX + placebo de ciclosporina + glucocorticoides (ni A ni B); nótese que los esteroides

en este caso son una cointervención presente en todos los grupos (fig. 2). El desenlace primario fue la puntuación en el examen manual muscular o *manual muscle test* a los 12 meses. En todos los grupos hubo una mejoría en el desenlace, pero no se encontraron diferencias significativas al comparar cada uno de los brazos de tratamiento activo (solo o en combinación) con el brazo de placebo.

Ensayos de no inferioridad

Los ensayos de no inferioridad se usan cuando se desea estudiar si el efecto de una intervención experimental es no inferior comparado con el de una intervención estándar que previamente haya probado ser superior al placebo. La racionalidad para emplear este diseño reside en que el tratamiento experimental que pretende demostrar ser no inferior tiene una potencial ventaja sobre el tratamiento estándar en términos de seguridad, costo, administración, posología, tolerancia o alguna otra característica que resulte conveniente o ventajosa¹⁰. Para ejemplificar esto puede mencionarse el caso de los medicamentos biosimilares, los cuales se comparan con el biológico innovador para demostrar no inferioridad, mas no superioridad, y cuya principal ventaja debería ser un menor costo.

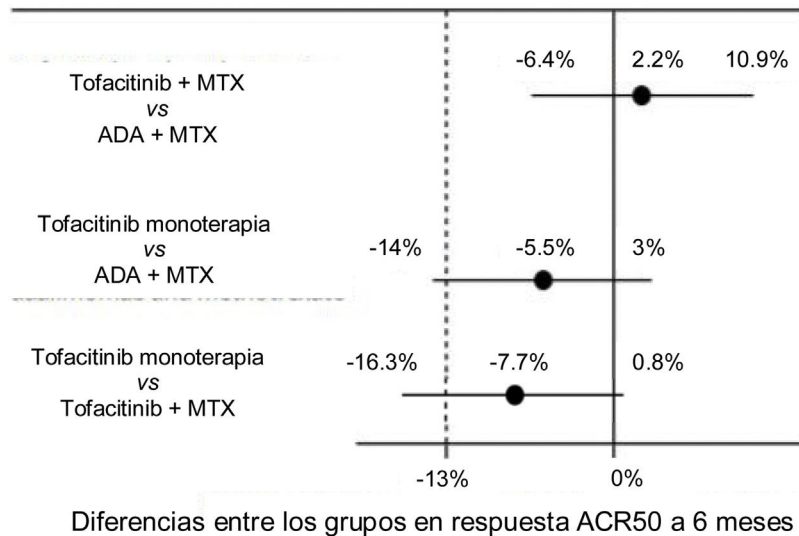


Figura 4 – Resultados del ORAL Strategy de acuerdo a los tres brazos del estudio.
Fuente: modificado de Fleischmann et al.¹¹.

La estrategia estadística fundamental en este tipo de estudios es el margen de no inferioridad o criterio de no inferioridad, el cual se deriva de los resultados obtenidos en un ensayo clínico de superioridad del tratamiento estándar realizado previamente o un metaanálisis de varios de estos. Los resultados de un ensayo clínico de no inferioridad pueden ser: 1) no inferioridad demostrada: el tratamiento experimental es no inferior al comparador; 2) no inferioridad no demostrada: el tratamiento experimental no se puede considerar no inferior al comparador (dicho en otras palabras, es inferior); 3) superioridad: el tratamiento experimental es superior al comparador, y 4) no concluyente: el efecto del tratamiento comparador supera los márgenes de inferioridad predeterminados por el estudio, pero el intervalo de confianza incluye valores dentro del efecto deseado del medicamento en relación con el comparador; esto suele pasar cuando el tamaño de la muestra no fue suficiente o se calculó un margen de no inferioridad muy ambicioso para dicha molécula o intervención (fig. 3).

Por ejemplo, el estudio ORAL Strategy¹¹ fue un ensayo clínico aleatorizado, controlado, doble ciego, fase 3b/4, cabeza a cabeza, con un diseño de no inferioridad en el que se evaluó la eficacia de tres estrategias terapéuticas en pacientes con artritis reumatoide activa con respuesta inadecuada a MTX: tofacitinib en monoterapia, tofacitinib más MTX y ADA más MTX. Las intervenciones experimentales son los grupos de tofacitinib, en tanto que el tratamiento estándar comparador activo es el grupo de ADA más MTX. El desenlace primario fue la proporción de pacientes que alcanzaron al menos una respuesta ACR50 a los 6 meses. El margen de no inferioridad fue -13%, porcentaje que representaba la mitad de la diferencia absoluta observada entre ADA más MTX en relación con el placebo —la cual se obtuvo a partir de un metaanálisis de ensayos clínicos de ADA¹²—, y la línea media era 0. Se declaraba no inferioridad si el límite inferior del intervalo de confianza para la diferencia entre los grupos era mayor de -13% y superioridad si el límite inferior era

mayor de 0. En la figura 4 se ilustran los resultados del estudio.

Otro diseño de no inferioridad es el estudio de nefritis lúpica, el cual fue un ensayo clínico aleatorizado de etiqueta abierta en el que se compararon ciclofosfamida (CFM) endovenosa versus micofenolato mofetilo (MMF) durante 24 semanas para el tratamiento de inducción de remisión en 140 adultos con nefritis lúpica proliferativa (clases II y IV)¹³. La motivación principal argumentada para realizar el estudio era la suposición de que el MMF tendría menores efectos adversos y mejor tolerancia que la CFM, la terapia estándar de aquel entonces. El desenlace primario fue la tasa de remisión completa a las 24 semanas. El margen de no inferioridad fue fijado por los investigadores en -10% (una sola cola). El criterio para establecer no inferioridad era que el límite inferior del intervalo de confianza de la diferencia absoluta de las tasas de remisión entre el grupo de MMF y el grupo de CFM fuera mayor de -10%, mientras que el criterio para superioridad era que fuera mayor de 0. Dicha diferencia fue del 16,7%, con un intervalo de confianza del 95% (IC 95%) de 5,6-27,9%; el límite inferior de este intervalo no solo era mayor de -10%, sino que cruzaba el 0, por lo que se asumió superioridad del MMF, lo que constituye un ejemplo de un ensayo de no inferioridad en el que se reclama superioridad.

Conclusiones

Existen diferentes diseños de ensayos clínicos que sirven a propósitos específicos, los cuales dependen de la naturaleza de la enfermedad reumatológica, de las características de las intervenciones terapéuticas investigadas y de la pregunta que se desea responder con el estudio. Ante el gran número de publicaciones de ensayos clínicos, es responsabilidad del clínico saber interpretar dichos resultados y entender la metodología del estudio y su validez, para finalmente darle la aplicabilidad que considere apropiada con un paciente determinado.

Conflicto de intereses

El doctor Gómez Puerta ha presentado ponencias para laboratorios que realizan ensayos clínicos, incluyendo Abbvie, Bristol-Myers Squibb, Galápagos, Janssen, Lilly, MSD, Pfizer y Roche. Ninguno de los laboratorios mencionados realizó ninguna contribución para la elaboración del presente manuscrito.

BIBLIOGRAFÍA

- Friedman LM, Furberg CD, DeMets DL, Reboussin DM, Granger CB. Introduction to clinical trials. En: *Fundamentals of Clinical Trials*. New York, NY: Springer New York; 2010. p. 1-23.
- Evans SR. Clinical trial structures. *J Exp Stroke Transl Med*. 2010;3:8-18, <http://dx.doi.org/10.6030/1939-067x-3.1.8>.
- De Koning HD, Schalkwijk J, van der Ven-Jongekrijg J, Stoffels M, van der Meer JWM, Simon A. Sustained efficacy of the monoclonal anti-interleukin-1 beta antibody canakinumab in a 9-month trial in Schnitzler's syndrome. *Ann Rheum Dis*. 2013;72:1634-8, <http://dx.doi.org/10.1136/annrheumdis-2012-202192>.
- Katz DL. Study design. En: *Clinical Epidemiology & Evidence-Based Medicine. Fundamental Principles of Clinical Reasoning & Research*. Sage Publications; 2001. p. 147-70.
- Capurro ND, Gabrielli NL, Letelier SLM. Importancia de la intención de tratar y el seguimiento en la validez interna de un estudio clínico randomizado. *Rev Med Chil*. 2004;132:1557-60.
- Molina Arias M, Ochoa Sangrador C. Ensayo clínico (III). Aleatorización. Enmascaramiento. *Evid Pediatr*. 2015;11:15.
- Rothman KJ. Epidemiology in clinical settings. En: *Epidemiology: An Introduction*. 2nd edition Oxford University Press; 2012. p. 235-53.
- Aitken D, Laslett LL, Pan F, Haugen IK, Otahal P, Bellamy N, et al. A randomised double-blind placebo-controlled crossover trial of HUMira (adalimumab) for erosive hand Osteoarthritis — the HUMOR trial. *Osteoarthr Cartil*. 2018;26:880-7, <http://dx.doi.org/10.1016/j.joca.2018.02.899>.
- Ibrahim F, Choy E, Gordon P, Doré CJ, Hakim A, Kitaz G, et al. Second-line agents in myositis: 1-year factorial trial of additional immunosuppression in patients who have partially responded to steroids. *Rheumatology (Oxford)*. 2015;54:1050-5, <http://dx.doi.org/10.1093/rheumatology/keu442>.
- Ferreira-González I. Bases para la interpretación de los estudios de no inferioridad: a propósito de los estudios ROCKET-AF RE-LY y ARISTOTLE. *Rev Esp Cardiol*. 2014;67:432-5, <http://dx.doi.org/10.1016/j.recesp.2013.10.020>.
- Fleischmann R, Mysler E, Hall S, Kivitz AJ, Moots RJ, Luo Z, et al. Efficacy and safety of tofacitinib monotherapy, tofacitinib with methotrexate, and adalimumab with methotrexate in patients with rheumatoid arthritis (ORAL Strategy): A phase 3b/4, double-blind, head-to-head, randomised controlled trial. *Lancet*. 2017;390:457-68, [http://dx.doi.org/10.1016/S0140-6736\(17\)31618-5](http://dx.doi.org/10.1016/S0140-6736(17)31618-5).
- Machado MA, Maciel AA, de Lemos LLP, Costa JO, Kakehasi AM, Andrade EIG, et al. Adalimumab in rheumatoid arthritis treatment: A systematic review and meta-analysis of randomized clinical trials. *Rev Bras Reumatol*. 2013;53:419-30.
- Ginzler EM, Dooley MA, Aranow C, Kim MY, Buyon J, Merrill JT, et al. Mycophenolate mofetil or intravenous cyclophosphamide for lupus nephritis. *N Engl J Med*. 2005;353:2219-28, <http://dx.doi.org/10.1056/NEJMoa043731>.