

Chapter 14 / Capítulo 14



Critical reading of scientific articles on
health: A practical guide (Spanish Edition)
ISBN: 978-9915-9928-3-9
DOI: 10.62486/978-9915-9928-3-9.Ch14
Pages: 209-242

© 2026 The authors. This is an open access
book chapter distributed under the terms of
the Creative Commons Attribution (CC BY)
4.0 License.

Randomized Controlled Trials

Ensayos Clínicos Aleatorizados

Weymar Lehi Poma Luna  

Ariel Sosa-Remón  

Guiselle Carol Cabrera Morales  

Helen Fernández Burgoa  

1 Hospital Municipal Boliviano Holandes. La Paz, Bolivia.

2 Instituto Nacional de Oncología y Radiobiología. Habana, Cuba.

3 Hospital Oncológico, Caja Nacional de Salud. La Paz, Bolivia.

4 Ministerio de Salud y Deportes. Comité Regional de Integración Docente Asistencial e Investigación. Tarija Bolivia.

ABSTRACT

Introduction: Randomized controlled trials (RCTs) are the most methodologically rigorous experimental design for evaluating the efficacy and safety of healthcare interventions. Thanks to randomization, allocation concealment, and blinding, they minimize bias and establish causal relationships with high internal validity, making them the gold standard of Evidence-Based Medicine. Development: This chapter addresses the fundamental methodological principles of RCTs, including parallel, crossover, factorial, and cluster designs, as well as superiority, non-inferiority, and equivalence trials. Essential aspects such as participant selection criteria, randomization, blinding, outcome definition, follow-up, and statistical analysis using intention-to-treat, relative risk, absolute risk reduction, number needed to treat, and survival analysis are analyzed. Emphasis is placed on evaluating internal and external validity to ensure the reliability and applicability of the results. Conclusions: Randomized controlled trials (RCTs) represent the most robust tool for generating reliable scientific evidence to support clinical decisions and health policies. Their value depends not only on demonstrating therapeutic efficacy but also on methodological quality, transparency in execution, and the clinical relevance of the findings. Rigorous critical appraisal allows us to distinguish between statistically significant results and truly relevant benefits for patients.

Keywords: Randomized controlled trials; Randomization; Internal validity; Intention-to-treat analysis; Evidence-based medicine

RESUMEN

Introducción. Los ensayos clínicos aleatorizados (ECA) constituyen el diseño experimental de mayor rigor metodológico para evaluar la eficacia y seguridad de intervenciones sanitarias. Gracias a la aleatorización, el ocultamiento de la asignación y el enmascaramiento, permiten minimizar sesgos y establecer relaciones causales con elevada validez interna, siendo considerados el estándar de oro de la Medicina Basada en la Evidencia. Desarrollo. El capítulo aborda los principios metodológicos fundamentales de los ECA, incluyendo los diseños paralelos, cruzados, factoriales y por conglomerados, así como los ensayos de superioridad, no inferioridad y equivalencia. Se analizan aspectos esenciales como los criterios de selección de participantes, la aleatorización, el cegamiento, la definición de desenlaces, el seguimiento y el análisis estadístico mediante intención de tratar, riesgo relativo, reducción absoluta del riesgo, número necesario a tratar y análisis de supervivencia. De esta manera, se enfatiza la evaluación de la validez interna y externa para garantizar la confiabilidad y aplicabilidad de los resultados. Conclusiones. Los ECA representan la herramienta más robusta para generar evidencia científica confiable que sustente decisiones clínicas y políticas sanitarias. Su valor no depende únicamente de demostrar eficacia terapéutica, sino también de la calidad metodológica, la transparencia en la ejecución y la pertinencia clínica de los hallazgos. Una lectura crítica rigurosa permite distinguir entre resultados estadísticamente significativos y beneficios verdaderamente relevantes para los pacientes.

Palabras clave: Ensayos Clínicos Aleatorizados; Aleatorización; Validez Interna; Análisis por Intención de Tratar; Medicina Basada en la Evidencia

1. CASO CLÍNICO

La Sra. Julia, mujer de 60 años de edad, con diagnóstico de diabetes mellitus tipo 2 desde hace 12 años, acude a consulta en el Servicio de Medicina Interna para evaluación de su control metabólico. A pesar de recibir tratamiento farmacológico con metformina y sulfonilureas, presenta un control glucémico subóptimo, evidenciado por una hemoglobina glucosilada (HbA1c) de 8,6%. Como antecedentes relevantes, presenta hipertensión arterial adecuadamente controlada y una importante carga de riesgo cardiovascular familiar, dado que su hermano falleció por infarto agudo de miocardio a los 62 años. Durante la consulta, la paciente expresa interés por alternativas terapéuticas que, además de optimizar el control glucémico, contribuyan a reducir el riesgo de eventos cardiovasculares mayores. Esta inquietud resulta particularmente relevante considerando que la diabetes mellitus tipo 2 constituye un factor de riesgo independiente para enfermedad cardiovascular y que la reducción del riesgo cardiometabólico representa actualmente uno de los objetivos prioritarios del tratamiento integral de estos pacientes. En este contexto, el médico considera el posible uso de los inhibidores del cotransportador sodio-glucosa tipo 2 (iSGLT2), una clase farmacológica que ha demostrado beneficios clínicos más allá del control glucémico, incluyendo reducción de hospitalizaciones por insuficiencia cardíaca, enlentecimiento de la progresión de la enfermedad renal crónica y disminución de eventos cardiovasculares en poblaciones seleccionadas. Sin embargo, también reconoce que estos fármacos presentan limitaciones relacionadas con su costo, potenciales efectos adversos como infecciones genitourinarias y cetoacidosis diabética y la necesidad de identificar con precisión los subgrupos de pacientes que podrían obtener el mayor beneficio clínico en comparación con otras estrategias terapéuticas, como los agonistas del receptor del péptido similar al glucagón tipo 1 (GLP-1).

A partir de este escenario clínico surge la siguiente pregunta de investigación:

¿La adición de un inhibidor del cotransportador sodio-glucosa tipo 2 al tratamiento estándar es más eficaz y segura que el tratamiento convencional para reducir eventos cardiovasculares y mejorar los desenlaces clínicos en pacientes con diabetes mellitus tipo 2 y alto riesgo cardiovascular?

Responder esta interrogante requiere recurrir al diseño de investigación con mayor capacidad para establecer relaciones causales entre una intervención y sus desenlaces: los ensayos clínicos aleatorizados (ECA), considerados el estándar de referencia para evaluar la eficacia y seguridad de las intervenciones terapéuticas en salud.

2. OBJETIVOS

Los ECA son considerados el estándar de oro en investigación clínica porque permiten responder a una pregunta clave: ¿Funciona realmente una intervención y es segura en comparación con la práctica habitual o un placebo? Este estudio bien diseñado tiene varios objetivos:

- Evaluar la eficacia terapéutica: Busca determinar si la intervención produce el efecto deseado en la población de estudio.
- Evaluar la seguridad de la intervención: Además del beneficio, un ECA permite identificar los efectos adversos asociados al tratamiento.
- Establecer una relación causal: Gracias a la aleatorización y el enmascaramiento, los ECA son el diseño más confiable para decir que el efecto observado es consecuencia del tratamiento, y no de otros factores.
- Generar evidencia para guías y protocolos: Los resultados de los ECAs son la base para la medicina basada en la evidencia y permiten actualizar las guías clínicas internacionales. Es por ello que gracias a grandes ECAs, hoy sabemos que los iSGLT2 reducen hospitalizaciones por insuficiencia cardíaca y esta información ya se incorporó en guías de diabetes y cardiología.
- Optimizar la toma de decisiones clínicas: Un ECA no solo responde si un fármaco funciona, sino también qué magnitud de beneficio ofrece, cuántos pacientes hay que tratar (NNT) y si los beneficios superan los riesgos.

3. INTRODUCCIÓN

Los ECA representan el pilar metodológico de la Medicina Basada en la Evidencia (MBE). Desde mediados del siglo XX, este diseño se ha consolidado como el estándar de oro para

evaluar la eficacia y seguridad de intervenciones sanitarias, ya que, gracias a la aleatorización y el enmascaramiento, logra minimizar sesgos y establecer relaciones causales con un alto grado de validez interna.

Un ECA se caracteriza por asignar de manera aleatoria a los participantes a diferentes grupos de comparación (intervención, placebo o tratamiento estándar). Esta asignación al azar distribuye equitativamente los factores de confusión conocidos y desconocidos, lo que permite atribuir los resultados observados principalmente a la intervención evaluada. Por ello, cuando se realiza con calidad metodológica y ética, un ECA ofrece la evidencia más confiable para responder preguntas clínicas relevantes. En los últimos años, la importancia de los ECAs se ha reafirmado en múltiples contextos, desde los grandes ensayos de terapias cardiovasculares y oncológicas hasta los estudios pivotaes que guiaron el uso de vacunas y tratamientos durante la pandemia de COVID-19. Sin embargo, también se han identificado desafíos y limitaciones que requieren atención:

- La necesidad de mejorar la calidad del reporte. Un metaanálisis reciente de más de 21,000 ensayos (1966–2024) mostró que, aunque la adherencia a la guía CONSORT ha aumentado, persisten deficiencias importantes en la descripción de la aleatorización, el ocultamiento de la asignación y los protocolos publicados.
- El impulso hacia los “large simple RCTs” o ensayos grandes y pragmáticos, apoyados en registros electrónicos de salud, que permiten evaluar de manera más eficiente y con costos reducidos intervenciones en poblaciones amplias.
- La búsqueda de mayor equidad y representatividad. La OMS publicó en 2024 una guía global que insta a incluir más países de ingresos bajos y medianos en los ensayos, para que la evidencia no se limite a contextos de alto recurso. Asimismo, la ICH E21, en desarrollo, busca revertir la exclusión histórica de mujeres embarazadas y lactantes, promoviendo una investigación más inclusiva.
- La necesidad de fortalecer la integridad científica. Algunos casos recientes de retractaciones han puesto de relieve la importancia de garantizar transparencia, ética e independencia en la conducción de los ECAs.

En este contexto, los ECAs no solo son fundamentales para establecer si un tratamiento funciona, sino también para determinar su magnitud de beneficio, su seguridad, su aplicabilidad en diferentes poblaciones y su relevancia clínica frente a costos y alternativas existentes. Para el médico, esto se traduce en un puente directo entre la investigación científica y la toma de decisiones en el consultorio o en la sala hospitalaria. Por ejemplo, gracias a los grandes ECAs multicéntricos sabemos hoy que los inhibidores del SGLT2 no solo reducen la glucosa, sino también las hospitalizaciones por insuficiencia cardíaca y la progresión de la enfermedad renal crónica, lo que ha transformado la práctica clínica y las guías internacionales de diabetes y cardiología.

Los ECA constituyen la herramienta más poderosa para evaluar intervenciones en salud. Su evolución reciente apunta hacia ensayos más simples, globales, inclusivos y transparentes, lo cual garantiza que la evidencia que generan sea más aplicable, confiable y justa para la diversidad de pacientes que atendemos en la práctica clínica.

4. METODOLOGÍA

Cuando un profesional de la salud se enfrenta a un artículo sobre un ensayo clínico aleatorizado, la sección de Metodología es el corazón del estudio. Allí se describe cómo fue diseñado, cómo se seleccionaron los participantes, de qué manera se aplicó la aleatorización, qué tipo de comparador se utilizó, cómo se definieron los desenlaces y qué estrategias se emplearon para garantizar la validez de los resultados. Un lector crítico debe detenerse con calma en este apartado, ya que la solidez del ensayo depende más del diseño que del hallazgo final.

4.1 TIPO DE DISEÑO

El tipo de diseño de un ensayo clínico determina la manera en que los participantes son distribuidos y cómo se comparan los resultados entre los grupos. Comprender esta característica es esencial para interpretar críticamente un artículo, ya que de ella dependen la validez de los hallazgos y la aplicabilidad clínica.

El diseño más común es el **ensayo paralelo**, en el cual cada participante es asignado a un grupo y permanece en él durante todo el estudio: uno recibe la intervención y el otro recibe

placebo o tratamiento estándar. Este modelo es fácil de interpretar y ofrece comparaciones claras entre grupos. En nuestro caso, el diseño más apropiado sería un ensayo paralelo donde un grupo reciba un iSGLT2 añadido al tratamiento habitual y el otro grupo reciba únicamente el tratamiento estándar. Así, al final del seguimiento se comparan los eventos cardiovasculares, la glucosa y los efectos adversos.

Existen también los **ensayos cruzados** (*crossover*), en los que los mismos participantes reciben las dos intervenciones en distintos momentos, separados por un “período de lavado” para evitar que los efectos de la primera intervención influyan en la segunda. Este diseño reduce la variabilidad, pues cada persona actúa como su propio control, pero no es adecuado cuando la enfermedad progresa con el tiempo o el efecto de la intervención es irreversible. En nuestro caso si la Sra. Julia participara en un crossover para evaluar un antihipertensivo, podría recibir primero el fármaco A, luego pasar un período de lavado y después recibir el fármaco B. No obstante, este diseño no sería apropiado para un iSGLT2, ya que sus efectos cardiovasculares y renales persisten en el tiempo y podrían alterar la comparación.

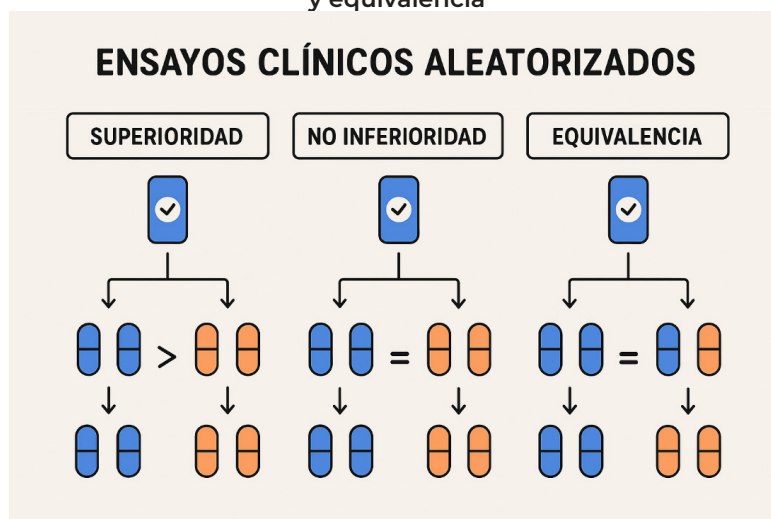
Otro diseño es la factorial, que permite estudiar más de una intervención en el mismo ensayo y en la misma población. Por ejemplo, un ensayo podría evaluar al mismo tiempo un iSGLT2 y una dieta hipocalórica, con cuatro grupos: solo fármaco, solo dieta, ambos o ninguno. Este diseño permite ahorrar recursos y responder dos preguntas en un solo estudio, pero puede complicar la interpretación si existe interacción entre las intervenciones.

Están los ensayos por conglomerados (*clusterrandomizedtrials*), en los que la aleatorización no se hace en individuos, sino en grupos completos, como hospitales, centros de salud o comunidades. Estos diseños son útiles cuando existe riesgo de “contaminación” entre los participantes o cuando la intervención se aplica a nivel colectivo. Por ejemplo, un ensayo que evalúe un programa de educación para mejorar el control de la diabetes podría aleatorizar centros de salud enteros en lugar de pacientes individuales. Para el lector crítico, identificar el tipo de diseño en la sección de metodología es el primer paso para valorar un ECA. Si el diseño está bien elegido y justificado, los resultados tendrán mayor validez y pertinencia clínica.

4.2 ENSAYOS DE SUPERIORIDAD, NO INFERIORIDAD Y EQUIVALENCIA

Un aspecto central en la metodología de un ensayo clínico es el marco de comparación que los investigadores adoptan para responder a su pregunta de investigación. En la lectura crítica, identificar si se trata de un ensayo de superioridad, no inferioridad o equivalencia es clave, porque cambia completamente la manera en que interpretamos los resultados.

Figura 1. Tipos de ensayos clínicos aleatorizados según su marco de diseño: superioridad, no inferioridad y equivalencia



Fuente: Elaboración propia

Ensayo de superioridad: Un ensayo de superioridad busca demostrar que la nueva intervención es más eficaz que el tratamiento de comparación. En este tipo de estudios, el resultado esperado es una diferencia significativa a favor del nuevo fármaco o procedimiento.

Ensayo de no inferioridad: En ocasiones, la meta no es demostrar que el nuevo tratamiento

sea mejor, sino que no sea clínicamente peor que el estándar. Esto es útil cuando la nueva opción ofrece otras ventajas: menos efectos adversos, administración más sencilla, menor costo o mejor adherencia. Aquí los investigadores establecen un margen (Δ), es decir, la diferencia máxima que aceptarían para considerar que el nuevo tratamiento no es peor que el comparador.

Ensayo de equivalencia: En los ensayos de equivalencia se busca demostrar que dos tratamientos tienen el mismo efecto clínico, dentro de un rango predefinido. Este marco suele aplicarse cuando lo que interesa es sustituir un medicamento por otro (por ejemplo, un genérico por el original) y confirmar que ambos son igualmente eficaces y seguros.

En el acápite de interpretación estadística se verá esto con más detalle

4.3 PARTICIPANTES Y CRITERIOS DE INCLUSIÓN/EXCLUSIÓN

Un elemento fundamental de la metodología en un ensayo clínico aleatorizado es la definición clara de quiénes fueron los participantes. Al leer críticamente un artículo, el primer paso es preguntarse: *¿los pacientes estudiados se parecen a los que yo atiendo en la práctica clínica?*. Para responder, los investigadores especifican dos aspectos clave: los criterios de inclusión y los criterios de exclusión.

Los criterios de inclusión definen las características necesarias para participar. Generalmente incluyen la edad, el diagnóstico confirmado de la enfermedad y, en algunos casos, parámetros de severidad o riesgo. Estos criterios aseguran que los participantes realmente pertenecen a la población objetivo del fármaco. Por ejemplo, en un ECA sobre iSGLT2, podrían incluir a adultos entre 40 y 75 años con diabetes mellitus tipo 2, HbA1c mayor a 7% y alto riesgo cardiovascular.

En cambio, los criterios de exclusión buscan proteger la seguridad de los pacientes y evitar factores que dificulten la interpretación de los resultados. Estos criterios limitan el riesgo de complicaciones graves y reducen variables que podrían sesgar los resultados. Por ejemplo, se podría excluir a pacientes con insuficiencia renal avanzada, mujeres embarazadas o aquellos con mala adherencia a controles médicos.

Para el lector crítico, la clave está en entender que cada criterio tiene un propósito, pero también un efecto sobre la validez externa del estudio, es decir, sobre su aplicabilidad. Si los criterios son demasiado estrictos, el estudio puede terminar incluyendo solo a un subgrupo muy seleccionado de pacientes, y sus resultados no necesariamente serán extrapolables a la práctica clínica habitual.

Aplicado al caso, un ensayo con criterios de inclusión que consideren diabetes tipo 2 mal controlada y alto riesgo cardiovascular sí la representaría. Pero si el ensayo excluyera a todos los pacientes mayores de 60 años o con hipertensión arterial, dejaría fuera a Julia y a muchos de nuestros pacientes habituales, limitando la utilidad clínica de los resultados.

4.4 ALEATORIZACIÓN Y OCULTAMIENTO DE LA ASIGNACIÓN

La **aleatorización** es el corazón de todo ensayo clínico aleatorizado. Su propósito es distribuir a los participantes de forma imparcial entre los grupos de intervención y control, de manera que cualquier diferencia observada en los resultados pueda atribuirse al tratamiento y no a otras características de los pacientes.

En la lectura crítica, lo primero que debe identificar el lector es cómo se realizó la aleatorización. Existen varios métodos de aleatorización que buscan garantizar la comparabilidad entre los grupos de estudio. La aleatorización simple asigna a los participantes de manera completamente aleatoria, similar a lanzar una moneda al aire o usar un generador de números aleatorios. Es fácil de implementar, pero en estudios con muestras pequeñas puede generar desequilibrios en el tamaño de los grupos o en características basales relevantes.

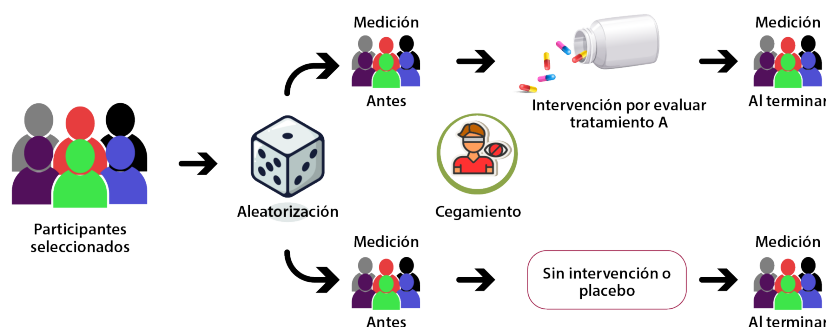
La aleatorización en bloques se utiliza para mantener un número similar de participantes en cada grupo a lo largo del estudio. Los pacientes se agrupan en bloques de tamaño fijo o variable (por ejemplo, 4 o 6), dentro de los cuales se distribuyen de forma equilibrada entre las intervenciones. Este método es especialmente útil en ensayos con reclutamiento progresivo, ya que evita diferencias importantes en el tamaño de los grupos en distintos momentos del estudio.

Por su parte, la aleatorización estratificada se emplea cuando existen variables clínicas relevantes que podrían influir en los resultados, como la edad, el sexo o la presencia de comorbilidades. En este caso, los participantes se clasifican previamente en subgrupos (estratos) según estas características, y luego se aleatorizan dentro de cada uno de ellos. De esta manera,

se asegura una distribución equilibrada de factores pronósticos clave entre los grupos de intervención.

En la práctica, es frecuente que los ensayos clínicos combinen estos métodos, por ejemplo, utilizando aleatorización estratificada con bloques, lo que permite mantener tanto el equilibrio en el tamaño de los grupos como en variables clínicas relevantes. Para el lector crítico, es fundamental verificar que el método de aleatorización esté claramente descrito y que sea adecuado para el diseño y tamaño del estudio.

Figura 2. Diseño clásico de un ensayo clínico aleatorizado



Fuente: Adaptado de Zurita-Cruz et al., 2018.

Ahora bien, la aleatorización por sí sola no es suficiente. Es igualmente importante el **ocultamiento de la asignación**, que consiste en impedir que los investigadores sepan de antemano a qué grupo se asignará un paciente. Este paso evita que, consciente o inconscientemente, los médicos incluyan a ciertos pacientes en un grupo con la esperanza de que se beneficien más o de que los resultados favorezcan al nuevo tratamiento. Un artículo de calidad debe explicar con claridad qué mecanismos se usaron para mantener oculta la secuencia de asignación, como sistemas computarizados centralizados o sobres sellados y opacos.

Volviendo a la Sra. Julia, imaginemos que participa en un ECA sobre iSGLT2. Si el investigador supiera antes de tiempo que ella será asignada al grupo control, podría —por buena intención— decidir no incluirla, pensando que su riesgo cardiovascular es alto y que debería recibir el nuevo medicamento. Esto introduciría un sesgo que invalidaría la comparación. Solo con una asignación estrictamente aleatoria y oculta podemos estar seguros de que los grupos son comparables y que las diferencias encontradas se deben al fármaco y no a la selección de pacientes.

Por ello, al leer un ensayo clínico, el médico debe detenerse a verificar dos puntos esenciales: ¿la aleatorización fue adecuada? ¿La asignación estuvo realmente oculta? Si la respuesta es afirmativa, entonces el estudio tiene un pilar sólido para sostener la validez de sus conclusiones.

4.5 ENMASCARAMIENTO

El enmascaramiento, también conocido como cegamiento, es otra piedra angular en los ECA. Consiste en mantener en secreto la asignación de los participantes a los grupos, con el fin de evitar que las expectativas o creencias de quienes participan en el estudio influyan en los resultados.

En la práctica, existen distintos niveles de enmascaramiento. El más simple es el **ciego simple**, en el que los pacientes no saben qué tratamiento reciben. Esto reduce el riesgo de que reporten mejoría solo por creer que están tomando un medicamento “nuevo”. Más sólido es el **doble ciego**, donde ni los pacientes ni los investigadores que administran el tratamiento conocen la asignación. Así se evita que el médico, de manera consciente o inconsciente, trate a los pacientes de forma distinta según el grupo al que pertenecen. El nivel más riguroso es el **triple ciego**, en el cual incluso los analistas de datos desconocen la asignación hasta que el estudio finaliza.

¿Por qué es importante este detalle metodológico para el lector crítico? Porque la ausencia de cegamiento puede **generar sesgo de observación y efecto placebo**. Si un investigador sabe que su paciente está en el grupo experimental, puede evaluar con más optimismo sus síntomas o ser más riguroso en la búsqueda de efectos adversos en el grupo control. Del mismo modo, un paciente que sospecha estar en el grupo placebo puede reportar menos mejoría, aunque su evolución sea similar.

En nuestro caso, si la paciente se encuentra en un ensayo sobre iSGLT2, lo ideal sería que ella

no supiera si recibe el nuevo fármaco o un placebo y que su médico tratante tampoco lo supiera. De esa manera, cualquier diferencia en la reducción de eventos cardiovasculares o en la aparición de efectos adversos se debería realmente al medicamento y no a las percepciones o conductas de los involucrados.

Figura 3. Niveles de cegamiento en ensayos clínicos aleatorizados y su impacto en el riesgo de sesgo



Fuente: Elaboración propia

Para el lector crítico, debe preguntarse: ¿quién estuvo cegado en este ensayo? Si los pacientes, los médicos y los evaluadores de resultados estuvieron adecuadamente enmascarados, la confianza en los hallazgos aumenta considerablemente. Cuando el cegamiento no es posible —por ejemplo, en estudios quirúrgicos o en intervenciones conductuales—, el lector crítico debe verificar si se usaron desenlaces objetivos (como la mortalidad) o evaluadores independientes para minimizar el sesgo.

4.6 INTERVENCIÓN Y SEGUIMIENTO

En todo ensayo clínico aleatorizado, la **intervención** debe estar descrita con precisión, de modo que cualquier lector pueda entender exactamente qué tratamiento se administró, en qué dosis, por cuánto tiempo y bajo qué condiciones. Este nivel de detalle es esencial tanto para la reproducibilidad científica como para que el lector pueda valorar si la intervención es aplicable a su propia práctica clínica.

Un buen artículo especifica el nombre del fármaco o la estrategia terapéutica, la vía de administración, la frecuencia de las dosis y la duración del tratamiento. También debe indicar si hubo ajustes de dosis, protocolos de rescate o combinaciones con otros medicamentos de uso estándar. En nuestro caso, un ECA de iSGLT2 debería indicar, por ejemplo, que los participantes recibieron *dapagliflozina 10 mg por vía oral una vez al día*, añadida al tratamiento habitual con metformina y sulfonilurea, durante un período de tres años. Sin esta claridad, el lector crítico no podría interpretar con certeza los resultados.

El segundo aspecto es el seguimiento, es decir, la forma en que los investigadores acompañaron a los pacientes a lo largo del tiempo para medir tanto los beneficios como los riesgos del tratamiento. Un seguimiento adecuado incluye visitas periódicas para evaluar el desenlace principal (como la aparición de eventos cardiovasculares en nuestro ejemplo), la medición de variables intermedias (HbA1c, presión arterial, peso corporal), y el registro sistemático de eventos adversos. Cuanto más largo y completo sea el seguimiento, más confiables serán las conclusiones.

El lector crítico debe prestar especial atención a las **pérdidas de seguimiento**. Ningún ensayo logra que todos los pacientes permanezcan hasta el final. Si el ensayo ya está en marcha, y las pérdidas son tan masivas que comprometen el poder estadístico (1-B), lo correcto no es reemplazar sujetos “uno a uno”, es mejor:

- Reportar honestamente las pérdidas en el diagrama de flujo (CONSORT).
- Discutir las limitaciones que esto genera en los resultados.
- Si el estudio es preliminar, podrías considerar ampliar el reclutamiento de forma global y transparente, pero nunca como una “sustitución” de individuos específicos, y siempre

previa aprobación del Comité de Ética e Investigación.

Es importante que el estudio reporte cuantos abandonaron, en que grupo y motivo. Una pérdida elevada o asimétrica puede sesgar los resultados. Si, por ejemplo, en un ensayo de iSGLT2 se pierde al 20% de los pacientes del grupo experimental por efectos adversos, y solo al 5% del grupo control, la validez del hallazgo se vería comprometida.

Finalmente, los artículos de calidad describen cómo evaluaron la **adherencia** al tratamiento. Esto puede hacerse con diarios de medicación, conteo de pastillas, medición de metabolitos en sangre o dispositivos electrónicos. En la práctica clínica sabemos que la adherencia es una de las principales barreras para obtener resultados; un ensayo que no la evalúe deja abierta la duda de si los beneficios observados se deben al fármaco o a que solo un grupo de pacientes cumplió estrictamente la indicación.

En conclusión, al leer la sección de intervención y seguimiento de un ensayo clínico, el lector crítico debe preguntarse: ¿Está claramente descrita la intervención, incluyendo dosis, vía y duración?, ¿El seguimiento fue suficiente para detectar beneficios y riesgos relevantes?, ¿Se reportaron las pérdidas y la adherencia de manera transparente?

4.7 DESENLACES

Un ensayo clínico solo tiene sentido si responde a una pregunta concreta y medible. Esa respuesta se obtiene a través de los **desenlaces** o **outcomes**, que son los eventos o variables que los investigadores deciden medir para evaluar la eficacia y la seguridad de la intervención.

En la lectura crítica de un artículo, el primer paso es identificar el **desenlace primario**. Este es el más importante, porque el estudio fue diseñado y calculado en función de él. El desenlace primario debe ser clínicamente relevante, claramente definido y medible de manera objetiva. En un ECA sobre diabetes y riesgo cardiovascular, un desenlace primario apropiado podría ser la aparición de un evento cardiovascular mayor (infarto, accidente cerebrovascular o muerte cardiovascular). De esta manera, el estudio no solo mide el efecto sobre la glucosa, sino también sobre aquello que más preocupa al paciente y al médico: la reducción de complicaciones graves.

Además del desenlace primario, los ensayos suelen reportar **desenlaces secundarios**, que aportan información complementaria pero no deben interpretarse con el mismo peso. Estos desenlaces ayudan a entender mejor el perfil del tratamiento, aunque no fueron el foco principal del cálculo del tamaño muestral. En el caso, los desenlaces secundarios podrían incluir el cambio en la HbA1c, el peso corporal, la presión arterial o la función renal.

Otro aspecto importante son los **desenlaces de seguridad**, que deben estar descritos con tanto detalle como los de eficacia. Un buen ensayo clínico informa no solo los beneficios, sino también los riesgos: infecciones urinarias, deshidratación o hipotensión en el caso de los iSGLT2. Al leer un artículo, se debe preguntar: *¿se reportaron de manera completa los efectos adversos? ¿Se clasificaron por severidad y frecuencia?*

Un punto crítico en la lectura crítica es reconocer si los desenlaces son **duros (clínicos)** o **sustitutos (intermedios)**. Los desenlaces duros, como la mortalidad o los eventos cardiovasculares, tienen un impacto directo en la vida del paciente. Los desenlaces sustitutos, como la HbA1c, la presión arterial o el colesterol, predicen un beneficio potencial pero no siempre garantizan que se traduzca en una mejoría real en la salud. En la práctica clínica, ambos son útiles, pero los desenlaces duros tienen más peso para cambiar conductas y guías de práctica.

Finalmente, es importante que los desenlaces estén previamente definidos y registrados en un protocolo público (como ClinicalTrials.gov). Si un artículo cambia los desenlaces primarios después de iniciado el ensayo, o resalta solo los resultados positivos entre muchos desenlaces explorados, el lector debe ser cauteloso: puede tratarse de un sesgo de reporte.

En nuestro caso, un ensayo sólido debería tener como desenlace primario la reducción de eventos cardiovasculares mayores y, como secundarios, la mejoría en el control glucémico y la reducción de hospitalizaciones por insuficiencia cardíaca. Solo así podremos responder con certeza a la pregunta que motivó su consulta: *¿este medicamento no solo baja el azúcar, sino que además la protege de complicaciones graves?*

4.8 DURACIÓN DEL SEGUIMIENTO Y MANEJO DE PERDIDAS

La duración del seguimiento es uno de los aspectos que más influyen en la validez de los resultados. Un seguimiento demasiado corto puede mostrar una reducción transitoria en una

variable intermedia, pero no alcanzar a registrar los eventos clínicos relevantes ni los efectos adversos a largo plazo. Por el contrario, un seguimiento suficiente permite observar si el beneficio inicial se mantiene en el tiempo y si aparecen complicaciones inesperadas.

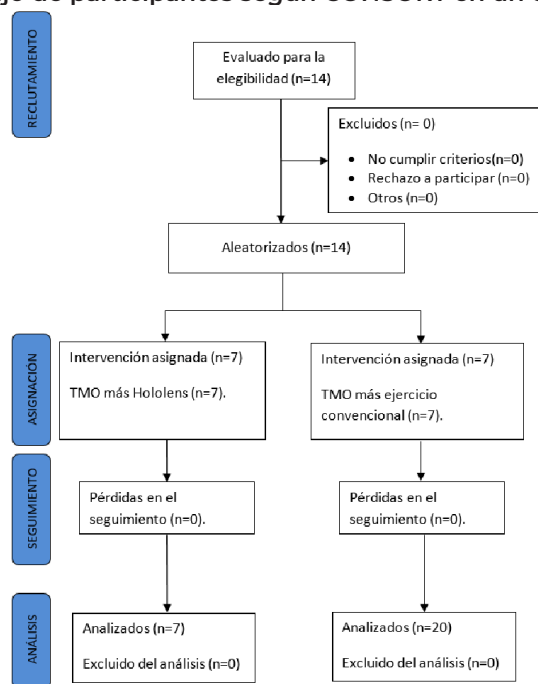
En la lectura crítica, el lector debe preguntarse: *¿el tiempo de seguimiento fue adecuado para detectar los desenlaces que se querían medir?* En un ensayo sobre fármacos para la diabetes, como en el caso de la Sra. Julia, un seguimiento de solo tres o seis meses podría mostrar una reducción en la HbA1c, pero no sería suficiente para evaluar si el medicamento reduce infartos, hospitalizaciones por insuficiencia cardíaca o la progresión de la enfermedad renal. Para este tipo de desenlaces cardiovasculares y renales, lo recomendable son seguimientos de al menos dos o tres años.

El otro aspecto crucial es el **manejo de las pérdidas de seguimiento**. En todo ensayo habrá pacientes que abandonen por efectos adversos, falta de motivación o traslado a otro centro de salud. Lo importante no es solo cuántos se pierden, sino también si la tasa de abandono es diferente entre los grupos. Una pérdida mayor en el grupo experimental, por ejemplo, podría ocultar problemas de seguridad o tolerancia.

Un buen artículo debe reportar con transparencia el número de pacientes que iniciaron el estudio, cuántos completaron el seguimiento y las razones de los abandonos. Lo ideal es que incluya un **diagrama de flujo CONSORT**, que muestre claramente el recorrido de los participantes desde la elegibilidad hasta el análisis final, como previamente habíamos mencionado.

En conclusión, al revisar la sección de metodología, el lector debe verificar que el tiempo de seguimiento sea suficiente para responder a la pregunta clínica y que las pérdidas hayan sido reportadas y analizadas con transparencia. Solo así podrá confiar en que los resultados reflejan lo que ocurriría en sus propios pacientes a largo plazo.

Figura 4. Diagrama de flujo de participantes según CONSORT en un ensayo clínico aleatorizado



Fuente: Elaboración propia

4.9 ANÁLISIS ESTADÍSTICO

Un ensayo clínico no solo se define por su diseño, su población y su seguimiento; también por la manera en que sus resultados serán analizados. Por eso, en la lectura crítica es fundamental revisar qué tipo de análisis estadístico fue planeado y si corresponde al diseño del estudio.

Existen dos enfoques principales: el análisis por intención de tratar (Intention-to-Treat, ITT) y el análisis por protocolo (Per Protocol, PP). El análisis por intención de tratar incluye a todos los pacientes en el grupo al que fueron asignados originalmente, aunque no hayan seguido el tratamiento hasta el final. Este enfoque preserva la aleatorización y refleja mejor lo que ocurre en

la práctica clínica real, donde no todos los pacientes cumplen al cien por ciento con la indicación. En cambio, el **análisis por protocolo** considera únicamente a aquellos que completaron el tratamiento según lo planeado, mostrando así el efecto “puro” de la intervención, aunque con mayor riesgo de sesgo porque excluye a quienes abandonaron o cambiaron de tratamiento.

Hoy en día, las guías metodológicas internacionales recomiendan que el análisis principal de un ECA sea por intención de tratar, ya que garantiza mayor validez interna y es más representativo de la práctica clínica. No obstante, muchos artículos también reportan el análisis por protocolo como complemento. El lector crítico debe comprobar si el estudio justifica adecuadamente qué población analizó y si existen diferencias importantes entre ambos enfoques. Además del tipo de análisis, es importante que el artículo especifique cómo se manejarán los **datos faltantes**. Estrategias como la imputación de valores o los análisis de sensibilidad permiten evaluar si los resultados se mantienen robustos pese a las pérdidas. Si el estudio omite este detalle, la interpretación de sus hallazgos puede ser poco confiable.

Al revisar la sección de metodología de un ECA, el lector debe fijarse en que el artículo describa: Qué población será analizada (ITT, PP o ambas). Cómo se tratarán los datos faltantes. Si el análisis planeado es coherente con los objetivos y el diseño del estudio.

Tabla 1. Elementos metodológicos esenciales en los Ensayos Clínicos Aleatorizados: descripción estandarizada y aplicación al caso

Elemento metodológico	Cómo se describe en un artículo científico	Aplicación al caso
Tipo de diseño	Se especifica si es paralelo, cruzado, factorial o por conglomerados, y si el objetivo es de superioridad, no inferioridad o equivalencia.	Ensayo paralelo, multicéntrico, de superioridad: iSGLT2 añadido al tratamiento estándar vs placebo.
Participantes	Se detallan los criterios de inclusión (diagnóstico, edad, factores de riesgo) y de exclusión (condiciones que pueden alterar seguridad o validez).	Adultos de 40–75 años con DM2, HbA1c $\geq$ 7%, con riesgo cardiovascular. Se excluyen pacientes con insuficiencia renal avanzada o embarazo.
Aleatorización y ocultamiento	Se describe el método de asignación (simple, bloques, estratificada), así como el sistema de ocultamiento (centralizado, sobres opacos, software).	Aleatorización en bloques y estratificada por edad y presencia de enfermedad cardiovascular; asignación centralizada en línea para evitar sesgo de selección.
Enmascaramiento (cegamiento)	Se especifica quién desconocía la asignación: pacientes, médicos, evaluadores o analistas (simple, doble o triple ciego).	Doble ciego: ni la Sra. Julia ni sus médicos saben si recibe iSGLT2 o placebo; los eventos cardiovasculares son evaluados por comité independiente.
Intervención	Se define el fármaco o tratamiento, dosis, vía, frecuencia, duración y protocolos de rescate.	Dapagliflozina 10 mg/día por vía oral, añadida al tratamiento estándar, durante 3 años. Ajustes según tolerancia.
Seguimiento	Se describe la frecuencia de visitas, qué variables se registraron (clínicas, bioquímicas, de seguridad) y cómo se evaluó la adherencia.	Consultas cada 3 meses, medición de HbA1c, presión arterial, función renal; adherencia evaluada con conteo de comprimidos y registros electrónicos.
Desenlaces	Se define un desenlace primario y los secundarios, junto con desenlaces de seguridad.	Primario: eventos cardiovasculares mayores (infarto, ACV, muerte cardiovascular). Secundarios: HbA1c, peso, presión arterial, progresión de ERC. Seguridad: infecciones urinarias y episodios de hipotensión.
Duración del seguimiento	Se justifica si el tiempo fue suficiente para evaluar beneficios y riesgos relevantes.	Tres años, adecuado para detectar diferencias en desenlaces cardiovasculares.
Pérdidas de seguimiento	Se reporta cuántos pacientes se retiraron y por qué, idealmente con diagrama de flujo CONSORT.	10% de abandonos en ambos grupos; causas reportadas: efectos adversos leves, falta de adherencia, traslado a otros centros.
Tipo de análisis estadístico planeado	Se aclara si el análisis será por intención de tratar (ITT), por protocolo (PP) o ambos, y cómo se tratarán los datos faltantes.	Análisis principal por ITT (todos los asignados se analizan en su grupo original). Secundario por PP para verificar consistencia. Datos faltantes manejados con imputación múltiple.

Nota: La tabla resume los componentes metodológicos que deben estar presentes en la publicación de un ensayo clínico aleatorizado según las guías internacionales (CONSORT 2010 y sus extensiones, SPIRIT 2013/2025, ICH E9[R1]). La primera columna identifica el elemento metodológico, la segunda describe cómo se debe reportar en un artículo científico y la tercera ejemplifica su aplicación en el caso ficticio de iSGLT2.

5. VALIDEZ

En la lectura crítica, se debe preguntar si los resultados son **válidos**. En otras palabras,

¿podemos confiar en lo que este estudio encontró y aplicarlo a nuestros pacientes? La validez de un ensayo clínico se analiza en dos dimensiones: **interna** y **externa**.

5.1 VALIDEZ INTERNA

La validez interna se refiere a la confianza de que las diferencias observadas entre los grupos se deben realmente a la intervención y no a errores sistemáticos o sesgos. Es la base científica del estudio: si un ensayo carece de validez interna, sus conclusiones pierden valor. Para evaluar esta validez, el lector crítico debe revisar varios aspectos:

- Aleatorización adecuada: asegura que los grupos sean comparables desde el inicio.
- Ocultamiento de la asignación: evita el sesgo de selección.
- Cegamiento: reduce la influencia de las expectativas de pacientes e investigadores.
- Seguimiento completo y balanceado: minimiza el sesgo por pérdidas.
- Análisis coherente con el diseño: un ITT preserva la aleatorización y refleja mejor la práctica clínica.

5.2 VALIDEZ EXTERNA

La validez externa, también llamada generalización o aplicabilidad, se refiere a qué tanto los resultados de un ensayo pueden extrapolarse a la práctica clínica cotidiana. Un estudio puede tener una validez interna impecable pero ser poco útil si la población estudiada no se parece a la que vemos a diario en nuestros hospitales y consultorios. Para analizar esta validez, el lector debe preguntarse:

- ¿Los criterios de inclusión y exclusión fueron tan estrictos que dejaron fuera a pacientes comunes (ancianos, embarazadas, personas con múltiples comorbilidades)?
- ¿Las condiciones del estudio se asemejan a la práctica real o son demasiado controladas?
- ¿El seguimiento fue lo bastante largo para captar los desenlaces clínicos relevantes?

5.3 EL EQUILIBRIO ENTRE VALIDEZ INTERNA Y EXTERNA

Un buen ensayo clínico debe aspirar a ambas: alta validez interna y adecuada validez externa. A veces, sin embargo, existe una tensión entre ambas. Los estudios muy estrictos logran gran control y reducen sesgos, pero sacrifican representatividad. Los ensayos pragmáticos, en cambio, buscan parecerse a la práctica real, pero corren mayor riesgo de variabilidad y pérdidas de seguimiento.

La lectura crítica consiste en identificar este equilibrio y decidir si los hallazgos del ensayo son confiables y aplicables al paciente que tenemos frente a nosotros.

Para el lector crítico

- La validez interna responde a la pregunta: *¿puedo confiar en que los resultados son ciertos para esta población estudiada?*
- La validez externa responde a la pregunta: *¿puedo aplicar estos resultados a mis pacientes en la práctica clínica?*

6. INTERPRETACIÓN ESTADÍSTICA

Leer un ensayo clínico aleatorizado no significa solo verificar que fue bien diseñado: también es imprescindible entender sus resultados. La sección de *Resultados* en un artículo suele presentar tablas, gráficos y medidas estadísticas que pueden parecer complejas. Sin embargo, con un enfoque sistemático, cualquier lector puede aprender a interpretar correctamente los hallazgos y aplicarlos a su práctica clínica.

Para la correcta interpretación, se utilizarán estos datos:

Tabla 2. Resultados principales de un ensayo clínico aleatorizado (Control vs. iSGLT2) en pacientes con diabetes tipo 2 y alto riesgo cardiovascular (N=2000, seguimiento 36 meses)

Grupo	N	MACE (n, %)	Hosp. IC (n, %)	Mortalidad (n, %)	Inf. genital (n, %)	CAD (n, %)	Δ HbA1c (media ± DE)	Δ Peso (kg, media ± DE)	Δ PAS (mmHg, media ± DE)
Control	1000	136 (13.6%)	125 (12.5%)	71 (7.1%)	32 (3.2%)	2 (0.2%)	-0.59 ± 0.96	-0.37 ± 2.48	-1.1 ± 7.8
iSGLT2	1000	81 (8.1%)	68 (6.8%)	44 (4.4%)	64 (6.4%)	4 (0.4%)	-0.87 ± 1.00	-2.94 ± 2.99	-2.7 ± 8.0

Fuente: Elaboración propia

6.1 PRINCIPIOS GENERALES

El primer principio que un lector crítico debe recordar es que un resultado numérico nunca debe interpretarse de manera aislada. Una diferencia estadísticamente significativa no siempre significa que sea clínicamente relevante, y un resultado no significativo no necesariamente implica que el tratamiento no funcione.

Por ello, la lectura crítica exige responder dos preguntas distintas:

- ¿El resultado es estadísticamente significativo? (es decir, ¿podemos descartar que sea producto del azar?).
- ¿El resultado es clínicamente importante? (es decir, ¿el beneficio es suficiente para cambiar la práctica en beneficio del paciente?).

Un segundo principio es distinguir entre las dos herramientas estadísticas más frecuentes: el valor p y el intervalo de confianza (IC). El valor p indica la probabilidad de obtener los resultados observados si en realidad no existiera diferencia entre los grupos. Aunque tradicionalmente se considera "significativo" un $p < 0,05$, este umbral es arbitrario y no refleja la magnitud del efecto.

El intervalo de confianza, en cambio, ofrece más información: indica un rango dentro del cual se encuentra el valor real del efecto con un 95% de seguridad estadística. Un IC estrecho refleja mayor precisión; un IC amplio indica mayor incertidumbre. Además, el IC permite valorar tanto la significancia estadística como la relevancia clínica, porque muestra el rango posible de beneficios o riesgos.

6.2 MEDIDAS DE EFECTO DE VARIABLES DICOTÓMICAS (CRUDAS)

RIESGO RELATIVO

El Riesgo Relativo (RR) compara la probabilidad de que ocurra un evento en el grupo intervención frente al grupo control y se interpreta de esta forma:

- Si $RR = 1$, no hay diferencia.
- Si $RR < 1$, la intervención protege (reduce el riesgo).
- Si $RR > 1$, la intervención aumenta el riesgo.

Tabla 3. Riesgo relativo de eventos cardiovasculares mayores (MACE) a 36 meses en pacientes con diabetes tipo 2 y alto riesgo cardiovascular (N=2000)

Grupo	Eventos / N	Riesgo (%)	
Control	136 / 1000	13,6	
iSGLT2	81 / 1000	8,1	
Medida de efecto	Valor puntual	IC95%	p-valor
Riesgo Relativo (RR)	0,6	0,46 – 0,77	<0,001

Fuente:Elaboración propia

Para su interpretación crítica:

- Riesgo Relativo (RR): El RR de 0,60 significa que los pacientes tratados con iSGLT2 tuvieron un 40% menos de riesgo relativo de presentar un evento cardiovascular mayor en comparación con los pacientes del grupo control.
- Intervalo de confianza (IC95%): El IC va de 0,46 a 0,77, lo que significa que, con un 95% de confianza estadística, el efecto verdadero del fármaco se encuentra en ese rango. Como

el IC no cruza el valor nulo ($RR=1$), la diferencia es estadísticamente significativa. Además, el rango es clínicamente relevante: incluso en el peor escenario ($RR=0,77$), el iSGLT2 aún ofrece un beneficio del 23%.

- Valor p: El $p < 0,001$ confirma que la probabilidad de que esta diferencia se deba al azar es extremadamente baja.

En la práctica clínica, esto significa que si la Sra. Julia recibe iSGLT2, tiene entre un 23% y un 54% menos riesgo de sufrir un infarto, ACV o muerte cardiovascular que si continúa solo con el tratamiento convencional.

REDUCCIÓN ABSOLUTA DE RIESGO (RAR)

La RAR mide la diferencia en el riesgo de presentar un evento entre el grupo control y el grupo intervención. A diferencia del Riesgo Relativo, que expresa una proporción, la RAR muestra la diferencia concreta en porcentaje de pacientes afectados, lo cual facilita la traducción a la práctica clínica.

Tabla 4. Reducción absoluta del riesgo (RAR) de eventos cardiovasculares mayores (MACE) a 36 meses en pacientes con diabetes tipo 2 y alto riesgo cardiovascular (N=2000)

Grupo	Eventos / N	Riesgo (%)	
Control	136 / 1000	13,6	
iSGLT2	81 / 1000	8,1	
Medida de efecto	Valor puntual	IC95%	p-valor
RAR	5,50%	2,8% – 8,2%	<0,001

Fuente:Elaboración propia

Para su Interpretación crítica

- RAR = 5,5%: Esto significa que, por cada 100 pacientes tratados con iSGLT2 durante 3 años, se evitan 5,5 eventos cardiovasculares mayores que hubieran ocurrido con el tratamiento estándar.
- IC95% (2,8% – 8,2%): Estamos 95% seguros de que el beneficio real se encuentra dentro de ese rango. En el peor escenario, el iSGLT2 prevendría 2 a 3 eventos por cada 100 pacientes; en el mejor, hasta 8.
- $p < 0,001$: Confirma que la diferencia es estadísticamente significativa y muy poco probable que se deba al azar.

Si la Sra. Julia y 99 pacientes similares fueran tratados con iSGLT2 en lugar del tratamiento convencional, entre 3 y 8 de ellos evitarían un infarto, ACV o muerte cardiovascular en tres años gracias al nuevo fármaco.

REDUCCIÓN DE RIESGO RELATIVO (RRR)

La RRR expresa el beneficio de la intervención en términos relativos: indica qué proporción del riesgo en el grupo control fue reducida gracias al tratamiento. Es una medida muy frecuente en artículos porque suele mostrar efectos más llamativos que la RAR. Sin embargo, debe interpretarse con cautela: un porcentaje alto puede impresionar, pero sin la RAR y el NNT pierde contexto clínico.

Tabla 5. Reducción relativa del riesgo (RRR) de eventos cardiovasculares mayores (MACE) a 36 meses en pacientes con diabetes tipo 2 y alto riesgo cardiovascular (N=2000)

Grupo	Eventos / N	Riesgo (%)	
Control	136 / 1000	13,6	
iSGLT2	81 / 1000	8,1	
Medida de efecto	Valor puntual	IC95%	p-valor
RRR	40%	23% – 54%	<0,001

Fuente: Elaboración propia

Para su interpretación crítica

- RRR = 40%: El tratamiento con iSGLT2 reduce en un 40% el riesgo relativo de presentar un evento cardiovascular mayor comparado con el tratamiento convencional.
- IC95% (23% – 54%): Con 95% de certeza, el beneficio real está entre un 23% y un 54% de reducción relativa del riesgo.
- $p < 0,001$: La diferencia es estadísticamente significativa.

En el caso de la Sra. Julia, la RRR muestra de manera impactante el beneficio: “el iSGLT2 reduce un 40% el riesgo de infarto, ACV o muerte cardiovascular”. Sin embargo, el médico crítico no debe quedarse solo con este número. Para tomar decisiones, es necesario complementarlo con la **RAR** y el **NNT**, que muestran cuántos pacientes se benefician en términos absolutos.

NÚMERO NECESARIO A TRATAR (NNT)

El NNT indica a cuántos pacientes hay que tratar con la intervención, durante un período específico, para evitar un evento adicional respecto al control. Importante que el NNT siempre debe ir acompañado del horizonte temporal (p. ej., a 36 meses) y, cuando sea posible, de su IC95%, que se obtiene invirtiendo los límites del IC de la RAR.

Tabla 6. Número necesario a tratar (NNT) para prevenir un evento cardiovascular mayor (MACE) a 36 meses en pacientes con diabetes tipo 2 y alto riesgo cardiovascular (N=2000)

Grupo	Eventos / N	Riesgo (%)		
Control	136 / 1000	13,6		
iSGLT2	81 / 1000	8,1		
Medida de efecto	Valor puntual	IC95% del NNT*	Horizonte	p-valor**
NNT	18	13 – 36	36 meses	<0,001

Fuente: Elaboración propia

Para su interpretación crítica

- Debemos tratar a 18 pacientes como la Sra. Julia con iSGLT2 durante 36 meses para prevenir 1 evento cardiovascular mayor adicional respecto al tratamiento estándar.
- Precisión: el IC95% 13–36 indica que, con la incertidumbre estadística del estudio, el “verdadero” NNT podría estar entre 13 (más favorable) y 36 (más conservador).
- Decisión clínica: el NNT traduce el beneficio a un lenguaje clínico tangible. Junto con el NNH (para infecciones genitales ≈ 31 en nuestro ejemplo) permite ponderar beneficio vs. riesgo y conversar con la paciente de forma clara.

Tip: el NNT debe estar reportado siempre con el tiempo (“a 36 meses”) y no redondear al número más cercano, sino hacia arriba (regla estándar) para mantener una interpretación conservadora.

NÚMERO NECESARIO A DAÑAR (NNH)

El NNH indica a cuántos pacientes hay que tratar con la intervención, durante un periodo específico, para causar 1 evento adverso adicional respecto al control. Igual que con el NNT, el NNH siempre debe reportarse con su horizonte temporal (p. ej., a 36 meses) y, cuando sea posible, con IC95%, obtenidos invirtiendo los límites del IC de la diferencia absoluta de riesgos.

Tabla 7. Aumento absoluto del riesgo (AAR) y número necesario para dañar (NNH) en infecciones genitales a 36 meses en pacientes con diabetes tipo 2 y alto riesgo cardiovascular (N=2000)

Grupo	Eventos / N	Riesgo (%)		
Control	32 / 1000	3,2		
iSGLT2	64 / 1000	6,4		
Medida de efecto	Valor puntual	IC95%	Horizonte	p-valor
AAR	3,20%	1,3% – 5,1%	36 meses	<0,001
NNH	31	20 – 75*	36 meses	—

Fuente: Elaboración propia

Para su interpretación crítica

- Si tratamos 31 pacientes como la Sra. Julia con iSGLT2 durante 36 meses, 1 paciente adicional presentará infección genital en comparación con el tratamiento estándar.
- Precisión: el IC95% 20–75 indica la incertidumbre: en el escenario menos favorable, 1 evento adicional por cada 20 tratados; en el más favorable, por cada 75.

Decisión clínica: el NNH debe evaluarse junto con el NNT. En nuestro ejemplo: NNT (MACE) $\approx$ 18 a 36 meses, NNH (infección genital) $\approx$ 31 a 36 meses. Esto sugiere que, en promedio, el beneficio cardiovascular ocurre con mayor frecuencia que el daño infeccioso, aunque ambos deben discutirse con la paciente y prevenirse (educación, higiene, vigilancia de síntomas)

Traducir la estadística a la práctica clínica es esencial para la toma de decisiones compartidas. Cuando leas un ensayo clínico, utiliza el Número Necesario a Tratar (NNT) para visualizar el impacto real.

6.3 MEDIDAS DE EFECTO DE VARIABLES DICOTÓMICAS (AJUSTADAS)

Cuando el interés está en el tiempo al evento (no solo si ocurrió, sino *cuándo*), se utiliza el análisis de supervivencia: Se construyen curvas de Kaplan-Meier para visualizar la probabilidad acumulada de permanecer libre de evento, se aplica un modelo de regresión de Cox que ajusta por covariables (edad, sexo, factores de riesgo), el resultado es el Hazard Ratio (HR), que veremos continuación:

Concepto de Kaplan-Meier: Cuando un ensayo clínico estudia un desenlace dicotómico (por ejemplo, muerte, infarto, hospitalización), no siempre basta con saber cuántos pacientes tuvieron el evento al final del seguimiento. También importa cuándo ocurrió: ¿los eventos se acumularon rápidamente en los primeros meses o de forma más tardía?

Para responderlo, se utilizan las curvas de Kaplan-Meier, que muestran la probabilidad acumulada de estar libre del evento a lo largo del tiempo (figura 5).

- Cada caída en la curva representa que un paciente (o un grupo de pacientes) presentó el evento.
- La distancia entre las curvas indica la diferencia entre los grupos a lo largo del seguimiento.

Concepto de Hazard Ratio (HR): En los ensayos clínicos con desenlaces de tiempo hasta evento (muerte, infarto, hospitalización), no basta con saber cuántos pacientes tuvieron el evento. Es igualmente importante entender cuándo ocurrió. El HR es la medida estadística más usada en este contexto, puede entenderse como el riesgo instantáneo de que ocurra un evento en un momento dado, en pacientes que aún no lo han presentado y se lo calcula con la regresión de Cox que permite ajustar con covariables relevantes

Para su interpretación del HR:

- HR = 1: no hay diferencia.
- HR < 1: la intervención reduce el riesgo (ej. HR=0,70 $\rightarrow$ un 30% menos riesgo instantáneo)
- HR > 1: la intervención aumenta el riesgo.

Concepto de regresión de Cox: El HR se obtiene a través de un modelo de regresión de Cox, que permite analizar el tiempo hasta el evento, comparar grupos (intervención vs control) y ajustar por covariables (edad, sexo, comorbilidades, etc.), aumentando la validez del resultado.

Para entender todo esto presentaremos la siguiente tabla:

Tabla 8. Eventos cardiovasculares mayores (MACE) a 36 meses: análisis crudo y ajustado mediante regresión de Cox en pacientes con diabetes tipo 2 y alto riesgo cardiovascular (N=2000)

Grupo	Eventos / N	Riesgo acumulado (%)	RR (IC95%)	HR ajustado (IC95%)*	p-valor
Control	136 / 1000	13,6	Ref.	Ref.	
iSGLT2	81 / 1000	8,1	0,60 (0,46–0,77)	0,72 (0,58–0,89)	0,002

Fuente: Elaboración propia con apoyo de Inteligencia artificial

Para su interpretación crítica de la tabla 8:

- El RR (0,60) muestra la diferencia acumulada al final del seguimiento: iSGLT2 reduce los eventos de 13,6% a 8,1%.
- El HR (0,72) muestra la reducción del riesgo en cualquier momento del seguimiento, tomando en cuenta cuándo ocurrió cada evento y ajustando por covariables.
- Ambos resultados van en la misma dirección, pero el HR ofrece una visión más dinámica y ajustada.

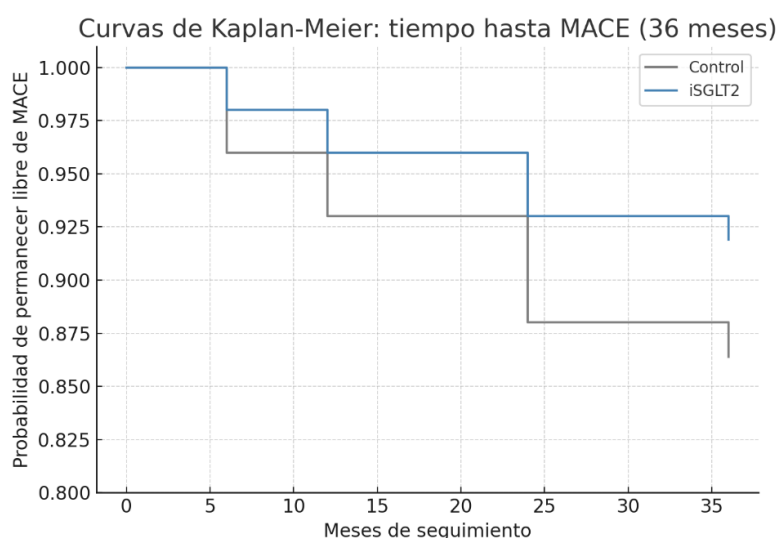
El lector crítico debe revisar siempre: Las curvas de Kaplan-Meier: ¿se separan temprano y se mantienen separadas?, El HR ajustado: ¿qué variables se incluyeron en el ajuste y están clínicamente justificadas?

El RR menciona *qué pasó al final del estudio*; el HR menciona *cómo se fueron acumulando los eventos a lo largo del tiempo*.

Para la interpretación crítica de la figura 5:

- Se observa cómo las curvas comienzan a separarse desde los primeros meses y se mantienen diferentes hasta el final.
- La curva del grupo iSGLT2 se mantiene más alta → mayor probabilidad de permanecer libre de eventos.
- La diferencia acumulada concuerda con lo que vimos en la tabla (13,6% vs 8,1% de eventos).

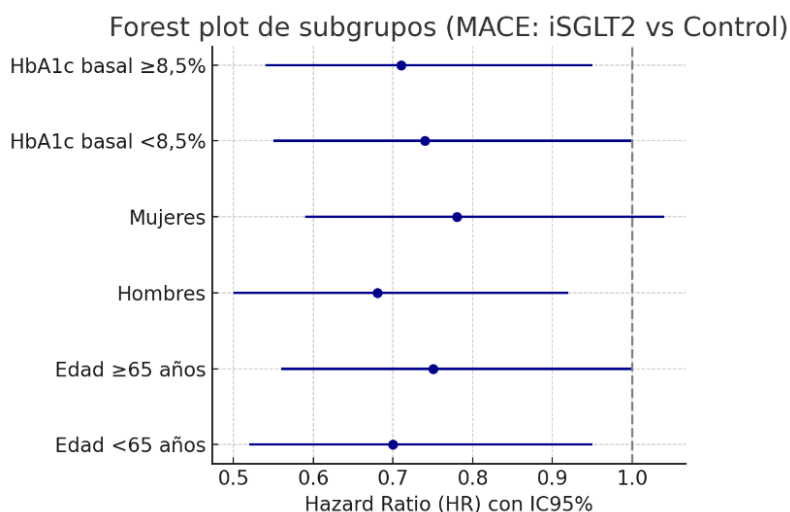
Figura 5. Curvas de Kaplan-Meier: tiempo hasta un evento cardiovascular mayor (MACE) en pacientes con diabetes tipo 2 y alto riesgo cardiovascular (seguimiento 36 meses, N=2000)



Nota: El análisis de supervivencia se realizó mediante el método de Kaplan-Meier; comparación entre grupos con prueba de log-rank.

Así mismo en algunos artículos prefieren presentar estos resultados por medio de una gráfica de Forest plot como se muestra a continuación:

Figura 6. Forest plot de subgrupos: hazard ratio (HR) para eventos cardiovasculares mayores (MACE) a 36 meses en pacientes tratados con iSGLT2 vs. control



Fuente: Elaboración propia

En este tipo de gráficos, los cuadrados representan el efecto de cada subgrupo, mientras que el diamante (o el punto central en este caso) representa el resultado global del estudio.

Regla de oro: Si el diamante o sus brazos horizontales (intervalos de confianza) no tocan ni cruzan la línea vertical de la unidad (valor nulo = 1), el resultado es estadísticamente significativo. En este capítulo, al estar a la izquierda de la línea, demuestra que el iSGLT2 es superior al control en todos los análisis

Para su interpretación:

- Muestra que el beneficio de iSGLT2 ($HR < 1$) es consistente en todos los subgrupos (edad, sexo, HbA1c basal).
- Algunos IC95% se acercan a 1, lo que refleja mayor incertidumbre en subgrupos más pequeños.
- En la práctica, estos gráficos se usan para reforzar la robustez y generalización de los resultados.

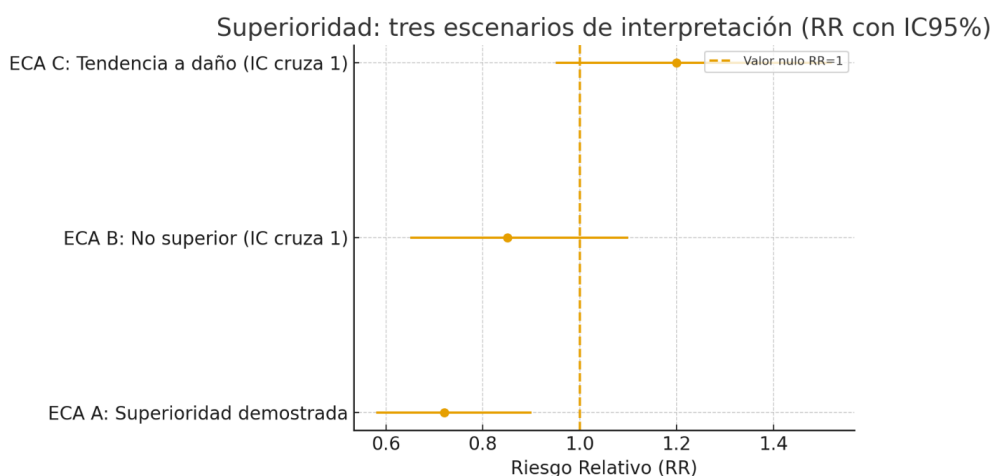
En un artículo, debajo del gráfico suele haber una nota: "Los análisis de subgrupos son exploratorios y deben interpretarse con cautela."

ENSAYO DE SUPERIORIDAD

Un ensayo de superioridad se diseña para demostrar que una intervención es mejor que el tratamiento comparador (placebo o tratamiento estándar). La hipótesis nula es que no hay diferencia entre grupos, y el objetivo del ensayo es rechazarla. Este tipo de ensayo es el más frecuente en medicina, sobre todo cuando se evalúan nuevos fármacos que se espera superen al placebo o a la terapia tradicional, utiliza el RR y suele hacerse por ITT, son los ensayos que cambian guías cuando muestran que una terapia supera a la que ya existe.

Así mismo en su interpretación se pueden encontrar tres casos

Figura 7. Ensayos clínicos de superioridad: tres escenarios de interpretación según riesgo relativo (RR) con intervalo de confianza al 95%



Nota: RR = riesgo relativo; IC95% = intervalo de confianza al 95%.

Su interpretación es de la siguiente forma:

- ECA A: Superioridad demostrada ($RR=0,72$; IC95% 0,58–0,90): el IC no cruza 1 → hay evidencia de beneficio.
- ECA B: No superior ($RR=0,85$; IC95% 0,65–1,10): el IC cruza 1 → no se puede afirmar superioridad (resultado "negativo/inconcluso").
- ECA C: Tendencia a daño ($RR=1,20$; IC95% 0,95–1,52): el punto sugiere perjuicio, pero el IC cruza 1 → estadísticamente no concluyente.

En nuestro ejemplo la pregunta de investigación sería: ¿El iSGLT2 añadido al tratamiento estándar reduce más los eventos cardiovasculares mayores (MACE) que el tratamiento convencional en pacientes con diabetes tipo 2 y alto riesgo cardiovascular?

Tabla 9. Ensayo de superioridad: incidencia de eventos cardiovasculares mayores (MACE) a 36 meses en pacientes con diabetes tipo 2 y alto riesgo cardiovascular (N=2000)

Grupo	Eventos / N	Riesgo (%)	
Control	136 / 1000	13,6	
iSGLT2	81 / 1000	8,1	
Medida de efecto	Valor puntual	IC95%	p-valor
Riesgo Relativo (RR)	0,6	0,46 – 0,77	<0,001
Reducción absoluta (RAR)	5,50%	2,8% – 8,2%	<0,001
NNT	18	13 – 36	—

Fuente: Elaboración propia

Su interpretación crítica

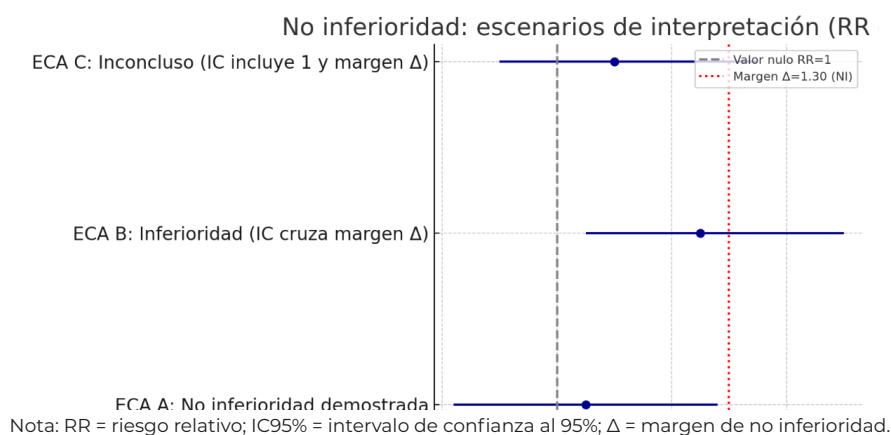
- El RR de 0,60 indica que el iSGLT2 reduce en un 40% el riesgo relativo de MACE frente al control.
- El IC95% (0,46–0,77) no cruza el 1, lo que confirma que el beneficio es estadísticamente significativo. Incluso en el peor escenario (RR=0,77), el tratamiento sigue mostrando un beneficio clínico relevante.
- La RAR de 5,5% significa que, por cada 100 pacientes tratados, se previenen entre 3 y 8 eventos cardiovasculares.
- El NNT de 18 traduce el hallazgo a la práctica: se necesita tratar a 18 pacientes como la Sra. Julia durante 3 años para evitar un evento adicional.
- El $p < 0,001$ respalda la significancia estadística, pero el verdadero valor interpretativo lo aporta el IC, que muestra la magnitud y la precisión del efecto.

Este ensayo de superioridad demuestra que el iSGLT2 es más eficaz que el tratamiento estándar para prevenir eventos cardiovasculares en pacientes con diabetes tipo 2 y alto riesgo, lo cual lo convierte en una opción sólida para su manejo.

ENSAYO DE NO INFERIORIDAD (NI)

Un ensayo de no inferioridad (NI) busca demostrar que una nueva intervención no es clínicamente peor que el tratamiento estándar, dentro de un margen predefinido que debe estar clínicamente justificada (Δ). Este tipo de ensayo es útil cuando el nuevo tratamiento puede ofrecer ventajas adicionales (menos efectos adversos, menor costo, vía de administración más sencilla, mejor adherencia), aunque no necesariamente supere en eficacia al comparador.

En vez de demostrar “ser mejor”, se intenta probar que el tratamiento es “suficientemente bueno”, es decir, que no pierde eficacia más allá de lo clínicamente aceptable. Su hipótesis nula nos dice que la intervención es inferior al estándar, su IC no debe cruzar el margen y debe reportarse tanto en ITT como PP

Figura 8. Ensayos clínicos de no inferioridad: escenarios de interpretación según riesgo relativo (RR) con intervalo de confianza al 95% y margen preespecificado (Δ)

Para su interpretación:

- ECA A: No inferioridad demostrada (RR=1,05; IC95% 0,82–1,28): el IC está por debajo del margen $\Delta=1,30 \rightarrow$ el nuevo tratamiento no es clínicamente peor.
- ECA B: Inferioridad (RR=1,25; IC95% 1,05–1,50): el IC cruza el margen $\Delta=1,30 \rightarrow$ el tratamiento es peor que el estándar.
- ECA C: Resultado inconcluso (RR=1,10; IC95% 0,90–1,35): el IC incluye tanto el valor nulo (1) como el margen $\Delta=1,30 \rightarrow$ no puede concluirse ni no inferioridad ni superioridad.

En nuestro ejemplo: ¿El iSGLT2 es no inferior a un agonista de GLP-1 en la prevención de eventos cardiovasculares en pacientes con diabetes tipo 2 y alto riesgo cardiovascular?

Se define un margen de no inferioridad $\Delta = 1,3$ para el Riesgo Relativo (RR). Esto significa que se acepta hasta un 30% más de riesgo con iSGLT2 respecto al GLP-1, siempre que ofrezca otras ventajas (por ejemplo, menor costo y mayor facilidad de administración).

Tabla 10. Ensayo de no inferioridad: incidencia de eventos cardiovasculares mayores (MACE) a 36 meses en pacientes tratados con iSGLT2 frente a GLP-1 (N=2000)

Grupo	Eventos / N	Riesgo (%)		
GLP-1	110 / 1000	11		
iSGLT2	115 / 1000	11,5		
Medida de efecto	Valor puntual	IC95%	Margen Δ	Conclusión
Riesgo Relativo (RR)	1,05	0,82 – 1,28	1,3	No inferioridad demostrada

Fuente: Elaboración propia

Para su interpretación crítica:

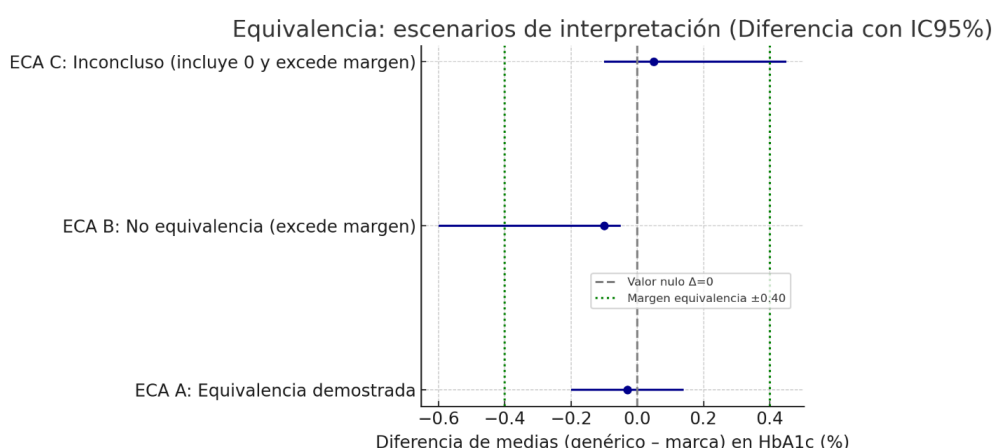
- El RR de 1,05 indica que la incidencia de MACE fue similar en ambos grupos, con solo un 5% más de riesgo en el grupo iSGLT2.
- El IC95% (0,82–1,28) se mantiene dentro del margen de no inferioridad ($\Delta=1,30$), lo que significa que no podemos descartar una ligera pérdida de eficacia, pero esta se encuentra dentro del rango clínicamente aceptable.
- Como resultado, el iSGLT2 puede considerarse no inferior al GLP-1 para prevención de eventos cardiovasculares.
- En este contexto, la decisión clínica podría inclinarse hacia el iSGLT2 si sus ventajas adicionales (menor costo, facilidad de administración, efectos renales positivos) resultan relevantes para la paciente.
- Si bien los agonistas de GLP-1 son una opción probada, un ensayo de no inferioridad podría justificar el uso de iSGLT2 en su lugar, especialmente si se busca un tratamiento oral, accesible y con beneficios adicionales, sin que ello signifique un sacrificio clínicamente importante en eficacia cardiovascular.

ENSAYO DE EQUIVALENCIA

Un ensayo de equivalencia busca demostrar que dos intervenciones producen efectos clínicos iguales dentro de un rango predefinido de tolerancia. A diferencia de los ensayos de no inferioridad, que aceptan que un tratamiento puede ser “no peor” que otro, los ensayos de equivalencia intentan probar que no hay diferencias significativas en ninguna dirección.

Este tipo de diseño se usa con frecuencia para comparar medicamentos genéricos o biosimilares con sus versiones de referencia, con el objetivo de demostrar que ofrecen la misma eficacia y seguridad. Su hipótesis nula nos dice que la diferencia entre ambos medicamentos es mayor al margen en alguna dirección, su margen de equivalencia ($\pm\Delta$) debe establecerse antes y justificar clínicamente qué diferencia se considera irrelevante, su IC debe encontrarse dentro del margen.

En nuestro ejemplo su pregunta de investigación: ¿Un iSGLT2 genérico es equivalente a la versión de marca en la reducción de HbA1c en pacientes con diabetes tipo 2 y alto riesgo cardiovascular? Se define un margen de equivalencia de $\pm 0,4\%$ para la HbA1c. Esto significa que diferencias menores a $\pm 0,4\%$ no se consideran clínicamente relevantes.

Figura 9. Ensayos clínicos de equivalencia: escenarios de interpretación según diferencia de medias en HbA1c con intervalo de confianza al 95% y márgenes preespecificados ($\pm\Delta$)

Nota: IC95% = intervalo de confianza al 95%; Δ = margen de equivalencia; la diferencia de medias se expresa como cambio en HbA1c (%).

Para su interpretación:

- ECA A: Equivalencia demostrada: Diferencia = $-0,03\%$ (IC95% $-0,20\%$ a $+0,14\%$), el IC queda completamente dentro de $\pm 0,40 \rightarrow$ equivalentes
- ECA B: No equivalencia (excede margen): Diferencia = $-0,10\%$ (IC95% $-0,60\%$ a $-0,05\%$). El IC rebasa el margen inferior ($-0,40$) $\rightarrow$ no se puede concluir equivalencia (es más “potente” que el margen permite).
- ECA C: Inconcluso (incluye 0 y excede margen): Diferencia = $+0,05\%$ (IC95% $-0,10\%$ a $+0,45\%$). El IC incluye 0 (no hay diferencia clara) pero supera el margen superior ($+0,40$) $\rightarrow$ no concluye equivalencia.

Tabla 11. Ensayo de equivalencia: cambio en HbA1c a 36 meses en pacientes tratados con iSGLT2 marca versus genérico

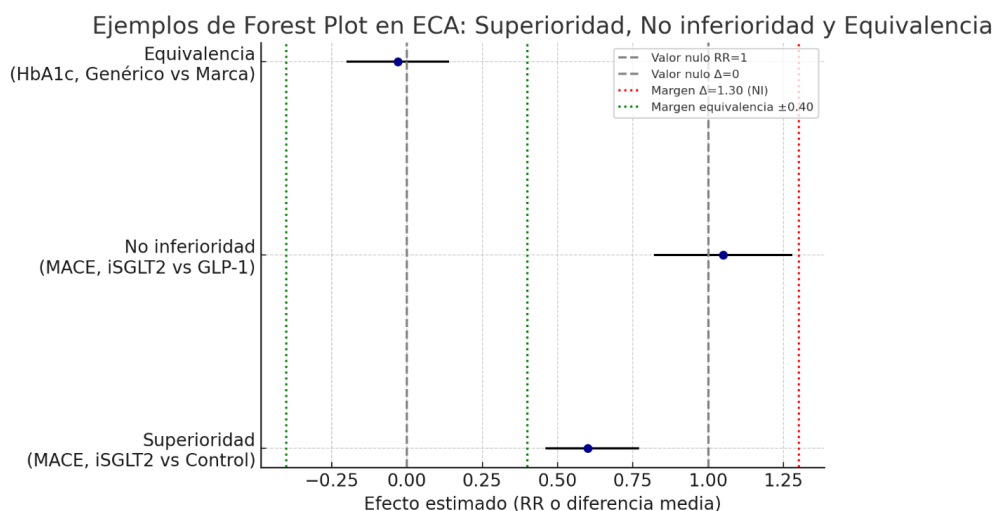
Grupo	Δ HbA1c media ($\pm$ DE)			
iSGLT2 marca	$-0,87 \pm 1,00$			
iSGLT2 genérico	$-0,90 \pm 1,05$			
Medida de efecto	Diferencia media (genérico – marca)	IC95%	Margen de equivalencia	Conclusión
Cambio en HbA1c	$-0,03\%$	$-0,20\%$ a $+0,14\%$	$\pm 0,40\%$	Equivalentes

Fuente: Elaboración propia

Para su interpretación crítica

- La diferencia media en reducción de ambos en HbA1c fue $-0,03\%$, es decir, prácticamente nula.
- El IC95% ($-0,20\%$ a $+0,14\%$) se encuentra completamente dentro del margen de equivalencia ($\pm 0,40\%$), lo que respalda la conclusión de equivalencia.
- Esto significa que ambos tratamientos, marca y genérico, ofrecen la misma eficacia para reducir HbA1c en pacientes con diabetes tipo 2 y riesgo cardiovascular.
- En un ensayo de equivalencia, si el genérico de iSGLT2 muestra eficacia y seguridad similares al de marca, su médico podría recetarle con confianza la versión más accesible, sabiendo que obtendrá el mismo beneficio clínico.

Muchas veces todos estos márgenes se presentan en un Forest plot y queda de la siguiente forma:

Figura 10. Ejemplos de forestplot en ensayos clínicos aleatorizados: superioridad, no inferioridad y equivalencia

Nota: RR = riesgo relativo; IC95% = intervalo de confianza al 95%; Δ = margen de no inferioridad; margen de equivalencia = $\pm 0.40\%$

Para la interpretación del forestplot comparativo con los tres tipos de ensayos clínicos:

- Superioridad (RR=0,60; IC95% 0,46–0,77): el IC no cruza 1 → demuestra que iSGLT2 es superior al control.
 - No inferioridad (RR=1,05; IC95% 0,82–1,28): el IC queda dentro del margen $\Delta=1,30$ → concluye no inferioridad frente a GLP-1.
 - Equivalencia (Δ HbA1c –0,03; IC95% –0,20 a +0,14): el IC está completamente dentro de $\pm 0,40$ → demuestra equivalencia entre genérico y marca
- Perlas clínicas y metodológicas para interpretar ECA según su marco

EN ENSAYOS DE SUPERIORIDAD

- Lo que importa es que el IC95% no cruce 1 (para RR) o 0 (para diferencias).
- Si el IC incluye el valor nulo → el ensayo no puede concluir superioridad, aunque el valor puntual sugiera beneficio.
- El p-valor acompaña, pero lo esencial es el IC: muestra tanto la magnitud del efecto como su precisión.
- Pregunta clave para el lector: *¿el efecto es estadísticamente significativo y clínicamente relevante?*

EN ENSAYOS DE NO INFERIORIDAD (NI)

- Aquí lo central es el margen Δ de no inferioridad, que debe estar justificado clínicamente.
- Un IC que incluye 1 no invalida la NI, siempre que se mantenga dentro del Δ .
- Los ensayos de NI deben reportar tanto análisis por intención de tratar (ITT) como por protocolo (PP): si coinciden, la conclusión es robusta.
- Pregunta clave para el lector: *¿el IC se mantiene dentro del margen predefinido de NI? ¿Ese margen es razonable?*

EN ENSAYOS DE EQUIVALENCIA

- Requieren un margen bilateral ($\pm\Delta$) definido antes de iniciar el ensayo.
- Para concluir equivalencia, el IC95% debe quedar completamente dentro de ambos márgenes.
- Si el IC toca o sobrepasa cualquiera de los límites, no se puede concluir equivalencia, aunque el valor puntual esté cerca de 0.
- Pregunta clave para el lector: *¿el IC cabe íntegramente dentro de los márgenes de equivalencia?*

ERRORES FRECUENTES

- Concluir superioridad en un ensayo de NI porque el IC muestra un valor favorable (¡no es lo que estaba diseñado para probar!).
- Confiar solo en el valor p sin revisar el intervalo de confianza ni los márgenes.
- No verificar si el margen Δ fue definido antes y con criterio clínico (un margen amplio puede “maquillar” la inferioridad).
- Olvidar que la relevancia clínica es tan importante como la estadística: un resultado significativo puede no ser relevante en la práctica.

6.4 MEDIDAS DE EFECTO DE VARIABLES CONTINUAS

En muchos ensayos clínicos los desenlaces no se expresan como “evento sí/no”, sino como valores numéricos continuos, en estos casos, la medida de efecto más frecuente es la diferencia de medias entre grupos, calculada al final del estudio o como cambio desde el valor basal, son ejemplos: niveles de HbA1c, presión arterial, colesterol LDL, peso corporal, eGFR, etc. Sus características son:

Hipótesis nula (H_0): no hay diferencia entre la media de ambos grupos (diferencia = 0).

Hipótesis alternativa (H_1): existe diferencia entre las medias (diferencia $\neq 0$, o en un sentido específico si el ensayo es unilateral).

Intervalo de confianza (IC95%): si el IC incluye el 0, no hay significancia estadística.

Importancia clínica: incluso cuando hay diferencia significativa, el lector crítico debe preguntarse si esa magnitud de cambio es realmente útil para los pacientes.

DIFERENCIA DE MEDIAS

La diferencia de medias es la medida de efecto más común cuando los desenlaces son variables cuantitativas (p. ej., HbA1c, peso, presión arterial).

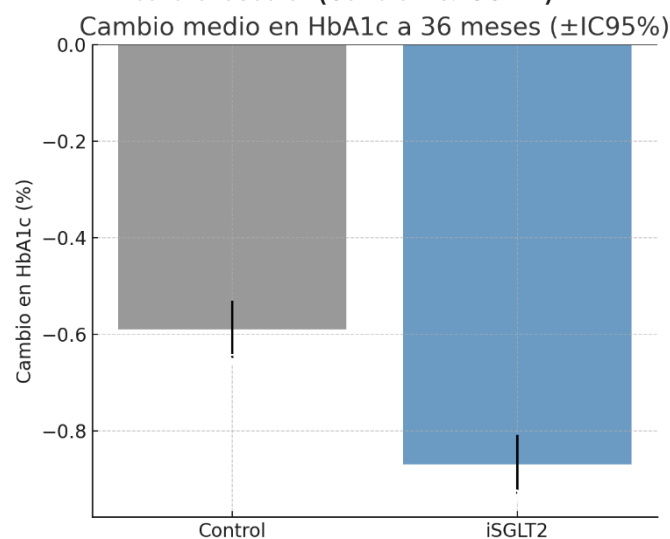
Eso se presenta de la siguiente forma:

Tabla 12. Cambio en HbA1c a 36 meses en pacientes con diabetes tipo 2 y alto riesgo cardiovascular (Control vs. iSGLT2)

Grupo	N	Δ HbA1c media ($\pm$ DE)	Diferencia media (IC95%)	p-valor
Control	1000	$-0,59 \pm 0,96$		
iSGLT2	1000	$-0,87 \pm 1,00$	$-0,28 (-0,39 \text{ a } -0,17)$	$<0,001$

Fuente: Elaboración propia

Figura 11. Cambio medio en HbA1c (%) a 36 meses en pacientes con diabetes tipo 2 y alto riesgo cardiovascular (Control vs. iSGLT2)



Nota: HbA1c = hemoglobina glicosilada; los valores representan la media de cambio con intervalos de confianza al 95% (IC95%).

Para su interpretación crítica

- La diferencia de $-0,28\%$ indica que el iSGLT2 logró una reducción de HbA1c ligeramente mayor que el control.
- El IC95% ($-0,39$ a $-0,17$) no incluye 0, lo que confirma que la diferencia es estadísticamente significativa.
- Sin embargo, la magnitud del efecto es modesta: una reducción de $\sim 0,3\%$ en HbA1c rara vez cambia la práctica por sí sola.
- El lector crítico debe distinguir entre significancia estadística y relevancia clínica: este beneficio glucémico puede no justificar el uso del fármaco, pero cobra sentido si se combina con sus efectos sobre eventos cardiovasculares y renales.

DIFERENCIA DE MEDIAS AJUSTADAS

En muchos ensayos clínicos, la comparación de medias entre grupos no se hace de manera “cruda” (simple resta de medias), sino ajustada por variables basales u otros factores relevantes. Esto se logra mediante técnicas estadísticas como:

- ANCOVA (Análisis de covarianza): ajusta la diferencia final tomando en cuenta los valores basales de la variable (p. ej., HbA1c inicial).
- Modelos lineales mixtos: permiten incorporar medidas repetidas en el tiempo (longitudinales), ajustando por la evolución de cada paciente.

Esto se realiza porque reduce el efecto de posibles desequilibrios basales (aunque haya aleatorización, siempre puede haber pequeñas diferencias), además que incrementa la precisión de la estimación y permite interpretar la diferencia como “el efecto neto del tratamiento”, aislando otras influencias.

Para entenderlo mejor imaginemos que en nuestro ensayo ficticio, al inicio los pacientes del grupo iSGLT2 tenían una HbA1c ligeramente mayor ($8,7\%$ vs $8,5\%$). Si solo comparamos la reducción final ($-0,87\%$ vs $-0,59\%$), la diferencia es $-0,28\%$, pero, al usar un modelo ANCOVA que ajusta por HbA1c basal, edad y sexo, la diferencia ajustada puede ser más precisa:

Tabla 13. Cambio en HbA1c a 36 meses: comparación cruda y ajustada en pacientes con diabetes tipo 2 y alto riesgo cardiovascular (Control vs. iSGLT2)

Grupo	Δ HbA1c media ($\pm$ DE)	Diferencia cruda (IC95%)	Diferencia ajustada (IC95%)*	p-valor
Control	$-0,59 \pm 0,96$			
iSGLT2	$-0,87 \pm 1,00$	$-0,28$ ($-0,39$ a $-0,17$)	$-0,25$ ($-0,36$ a $-0,14$)	$<0,001$

Nota: HbA1c = hemoglobina glicosilada; DE = desviación estándar; IC95% = intervalo de confianza al 95%.

Para su interpretación crítica:

- El ajuste no cambia radicalmente la conclusión (sigue siendo significativo), pero la estimación es más precisa.
- Para el lector crítico, es clave verificar qué variables se usaron en el ajuste: Si son pertinentes (ej., HbA1c basal, edad), el análisis gana validez, si son excesivas o irrelevantes, el ajuste puede introducir sesgo.
- La diferencia ajustada suele ser la que se reporta en el artículo principal.
- Cuando lea un ECA con desenlaces continuos, se debe revisar siempre si la diferencia reportada es cruda o ajustada. Los ensayos de mayor calidad suelen usar ANCOVA o modelos mixtos para dar resultados más sólidos.

MODELOS MIXTOS PARA MEDIDAS REPETIDAS (MMRM)

En muchos ensayos, los pacientes son evaluados en múltiples momentos de seguimiento (por ejemplo, HbA1c a los 6, 12, 24 y 36 meses). Comparar solo la media al final puede desaprovechar información, además de que algunos pacientes pueden tener datos faltantes en ciertos puntos.

Para manejar esto, se utilizan los MMRM:

- Incorporan todas las mediciones de cada paciente.
- Permiten estimar el efecto del tratamiento a lo largo del tiempo, no solo al final.
- Tienen en cuenta la correlación intrapaciente (las mediciones de un mismo individuo están relacionadas).

- Son más robustos frente a datos faltantes (si faltan algunas visitas, aún pueden incluirse las demás del mismo paciente).

Para entenderlo utilizaremos el siguiente ejemplo:

Tabla 14. Cambio en HbA1c a lo largo del estudio (36 meses): análisis crudo y ajustado en pacientes con diabetes tipo 2 y alto riesgo cardiovascular (Control vs. iSGLT2)

Grupo	N	Δ HbA1c media (±DE)	Diferencia cruda (IC95%)	Diferencia ajustada (IC95%)*	p-valor
Control	1000	-0,59 ± 0,96			
iSGLT2	1000	-0,87 ± 1,00	-0,28 (-0,39 a -0,17)	-0,25 (-0,36 a -0,14)	<0,001

Fuente: Elaboración propia con apoyo de Inteligencia artificial

Para su interpretación crítica

- El análisis crudo muestra una diferencia de -0,28% a favor de iSGLT2, estadísticamente significativa (no cuenta el 0 entre los intervalos)
- El análisis ajustado corrige por posibles diferencias iniciales (HbA1c basal ligeramente distinta, edad, sexo), y la estimación final es -0,25%, también significativa.
- Aunque las cifras varían un poco, la conclusión es la misma: el iSGLT2 logra una reducción mayor de HbA1c.

El lector crítico debe fijarse en qué variables fueron incluidas en el ajuste: Si son clínicamente relevantes, aumentan la validez del resultado, si son excesivas o arbitrarias, pueden introducir sesgo.

6.5 ANÁLISIS DE SUPERVIVENCIA

En los ensayos clínicos aleatorizados, cuando el desenlace principal es un evento (infarto, muerte, hospitalización), no basta con conocer qué proporción de pacientes lo presentó al final del estudio. También es necesario saber en qué momento ocurrió y cómo fue la acumulación de eventos en el tiempo.

Para responder estas preguntas, se utilizan técnicas de análisis de supervivencia, los mismos que ya se mencionaron en base a HR, curvas de Kaplan-Meier y regresión de Cox

Para aclarar:

- La curva de Kaplan-Meier muestra cómo se van acumulando los eventos en el tiempo.
- El log-rank test compara estadísticamente ambas curvas. ($p < 0,05$ las curvas son diferentes)
- El HR (Cox) resume el efecto del tratamiento durante todo el seguimiento y puede ajustarse por covariables.

6.6 ANÁLISIS DE SUBGRUPOS

En los ensayos clínicos, una pregunta frecuente es si el efecto del tratamiento es igual para todos los pacientes o si varía según ciertas características. Para responder esto, los investigadores realizan análisis de subgrupos, dividiendo a los participantes en categorías como:

- Edad (ej.: <65 vs ≥65 años)
- Sexo (hombres vs mujeres).
- Severidad de la enfermedad al inicio (HbA1c baja vs alta, estadio de insuficiencia cardíaca, etc.).
- Comorbilidades (con vs sin enfermedad renal).

El objetivo es explorar si el efecto del tratamiento es consistente en toda la población del estudio.

Riesgos de sobreinterpretación: Aunque son atractivos, los análisis de subgrupos presentan limitaciones importantes:

- Menor poder estadístico: al dividir la muestra, cada subgrupo tiene menos pacientes y aumenta la probabilidad de resultados falsos.
- Multiplicidad de comparaciones: cuantos más subgrupos se analicen, mayor es la probabilidad de encontrar diferencias “significativas” solo por azar.
- Hallazgos espurios: pueden aparecer diferencias que no tienen base biológica ni clínica. Un ejemplo clásico fue un ensayo de aspirina en el que, en análisis secundarios, parecía que el efecto solo existía en ciertos signos zodiacales: un resultado absurdo, pero “estadísticamente significativo” por azar.
- Influencia en la práctica clínica: algunos lectores cometen el error de recomendar un

tratamiento solo a un subgrupo específico cuando el efecto global del ensayo lo respalda para toda la población.

Cómo usarlos con cautela:

- Los análisis de subgrupos deben estar preespecificados en el protocolo del estudio, no inventados después de analizar los datos.
- Lo más importante en un ensayo es siempre el resultado global. Los subgrupos sirven para explorar consistencia, no para generar conclusiones definitivas.
- Tienen valor si: existe un fundamento fisiopatológico o clínico que justifique la diferencia, los resultados son consistentes en la mayoría de los subgrupos, el ensayo fue diseñado con suficiente poder para analizar el subgrupo (lo cual rara vez ocurre).

6.7 ANÁLISIS DE ROBUSTEZ

Cuando un ensayo clínico muestra resultados estadísticamente significativos, el lector crítico no debe conformarse con el valor *p* ni con el intervalo de confianza. Una pregunta fundamental es si esos resultados son **robustos**, es decir, si se mantienen consistentes cuando se aplican distintos métodos de análisis, escenarios alternativos o ajustes estadísticos.

El análisis de robustez busca responder a esta inquietud. En la práctica, los investigadores suelen contrastar los hallazgos principales con diferentes aproximaciones para demostrar que no dependen de un único modelo o supuesto. Por ejemplo, uno de los contrastes más frecuentes es comparar los resultados obtenidos bajo el principio de intención de tratar (ITT) con los del análisis por protocolo (PP).

- **En el análisis ITT** se incluyen todos los pacientes según el grupo al que fueron aleatorizados, aunque no hayan cumplido el tratamiento o se hayan perdido durante el seguimiento. Esto preserva la fuerza de la aleatorización y es considerado el análisis principal.
- **El análisis PP**, en cambio, restringe la muestra a los participantes que siguieron el tratamiento según lo planificado, lo que en teoría refleja mejor la eficacia del fármaco, pero introduce más riesgos de sesgo. Cuando ambos análisis coinciden en la dirección y magnitud del efecto, el lector puede tener mayor seguridad en que el hallazgo es confiable.

Otra forma habitual de evaluar la robustez es a través de **análisis de sensibilidad**. Estos se utilizan especialmente para explorar el impacto de las pérdidas de seguimiento o de los datos faltantes. El investigador puede probar distintos escenarios, por ejemplo, asumir que todos los pacientes perdidos en el grupo intervención presentaron el evento, o que ninguno lo presentó. Si el resultado principal se mantiene significativo en ambos extremos, puede considerarse más sólido.

También es frecuente repetir los análisis utilizando **modelos estadísticos alternativos**. Un ensayo puede presentar un HR ajustado por edad y sexo, pero luego mostrar que al añadir otras covariables como HbA1c basal, hipertensión o dislipidemia, los resultados se mantienen en la misma dirección. Esta consistencia refuerza la confianza en que el efecto del tratamiento no depende de un único modelo.

En conjunto, el análisis de robustez no busca generar resultados nuevos, sino **poner a prueba la estabilidad de las conclusiones**. Para el lector crítico, representa una herramienta esencial: un ensayo cuyo beneficio se mantiene bajo diferentes enfoques transmite mayor seguridad que aquel cuyos resultados cambian de forma sustancial al modificar las condiciones de análisis.

Para poder ser entendido se presentan los siguientes datos:

Tabla 15. Análisis de robustez para eventos cardiovasculares mayores (MACE) a 36 meses en pacientes con diabetes tipo 2 y alto riesgo cardiovascular

Tipo de análisis	HR (IC95%)	p-valor
Intención de tratar	0,72 (0,58–0,89)	0,002
Por protocolo	0,70 (0,55–0,88)	0,001
Sensibilidad*	0,75 (0,61–0,92)	0,004

Fuente: Elaboración propia

Para su interpretación, la tabla muestra que, independientemente del método de análisis, los resultados favorecen al iSGLT2. El HR se mantiene entre 0,70 y 0,75, lo que indica una reducción robusta del riesgo de eventos cardiovasculares mayores. El análisis de sensibilidad, diseñado para tensionar los datos al máximo, confirma que el efecto no depende de un único escenario, sino

que se mantiene clínicamente relevante en diferentes supuestos.

Un excelente lector critico debe conocer los siguientes puntos:

- La fortaleza de un ensayo no se mide solo por el valor p. La robustez implica que los resultados se mantengan estables bajo diferentes formas de análisis.
- ITT primero, PP después. El análisis por ITT es el estándar porque conserva la aleatorización. El PP puede mostrar un efecto más pronunciado, pero si ambos coinciden en dirección y magnitud, la confianza en el hallazgo aumenta.
- Los datos perdidos importan. Un buen ensayo explora escenarios distintos para las pérdidas de seguimiento mediante análisis de sensibilidad. Si el beneficio persiste incluso en el “peor escenario posible”, el resultado es sólido.
- Los modelos de ajuste deben ser coherentes. Añadir o quitar covariables relevantes (edad, sexo, comorbilidades) no debería cambiar radicalmente la estimación del efecto. Si ocurre, el lector debe sospechar fragilidad en los resultados.
- La consistencia es la clave. Cuando las conclusiones se mantienen a través de diferentes métodos (RR, HR, ITT, PP, sensibilidad), el hallazgo gana fuerza y se convierte en evidencia más confiable.

6.8 ERRORES FRECUENTES EN LA INTERPRETACIÓN

La lectura crítica de ensayos clínicos no consiste únicamente en identificar resultados estadísticamente significativos. Muchos lectores caen en interpretaciones equivocadas que pueden llevar a decisiones clínicas poco fundamentadas. A continuación, se resumen los errores más comunes y cómo evitarlos.

Confundir RRR con RAR: Uno de los errores más extendidos es quedarse solo con la Reducción Relativa del Riesgo (RRR). Esta medida suele producir cifras llamativas: “el nuevo fármaco reduce el riesgo un 40%”. Sin embargo, si el riesgo absoluto es bajo, la ganancia real para el paciente puede ser mínima. La Reducción Absoluta del Riesgo (RAR) y el Número Necesario a Tratar (NNT) traducen el beneficio a un lenguaje clínico comprensible: “por cada 18 pacientes tratados se evita un evento”. El lector crítico debe priorizar siempre estas medidas absolutas.

Enfocarse solo en el p-valor: El p-valor responde a la pregunta de si la diferencia observada podría explicarse por azar, pero no informa sobre la magnitud ni sobre la importancia clínica del efecto. Un $p < 0,05$ no convierte automáticamente un hallazgo en relevante, ni un $p > 0,05$ descarta utilidad clínica. El intervalo de confianza (IC95%) aporta mucha más información, ya que muestra el rango posible de efecto y su precisión.

Ignorar pérdidas de seguimiento: Las pérdidas durante el seguimiento pueden distorsionar gravemente los resultados. Si un porcentaje importante de pacientes se pierde y no se sabe si presentaron el evento, el efecto puede estar sobreestimado o infraestimado. El lector crítico debe revisar siempre si el porcentaje de pérdidas (si es alto, la validez interna está comprometida), cómo se manejaron estadísticamente (análisis por intención de tratar, análisis de sensibilidad).

No diferenciar significancia estadística de relevancia clínica: Un resultado puede ser estadísticamente significativo, pero clínicamente irrelevante. Por ejemplo, una reducción de 1 mmHg en la presión arterial puede alcanzar $p < 0,05$ en un ensayo grande, pero no cambia la práctica clínica. A la inversa, un resultado no significativo podría seguir siendo clínicamente valioso si muestra una tendencia consistente en estudios previos o en metaanálisis. El lector debe recordar que significancia estadística $\neq$ relevancia clínica.

Tabla 16. Errores frecuentes en la interpretación de ensayos clínicos aleatorizados y cómo evitarlos

Error frecuente	Cómo evitarlo
Confundir RRR con RAR	Revisar siempre la Reducción Absoluta del Riesgo (RAR) y el NNT para entender el beneficio real.
Enfocarse solo en el p-valor	Priorizar el intervalo de confianza (IC95%), que muestra magnitud y precisión del efecto.
Ignorar pérdidas de seguimiento	Verificar el porcentaje de pérdidas y cómo fueron manejadas (ITT, sensibilidad).
No diferenciar significancia de relevancia clínica	Preguntar si la magnitud del efecto es clínicamente importante, más allá de la significancia estadística.

Fuente: Elaboración propia con apoyo de Inteligencia artificial

7. SESGOS Y LIMITACIONES

Aunque los ensayos clínicos aleatorizados son considerados el “estándar de oro” para evaluar la eficacia de las intervenciones, no están exentos de limitaciones. La lectura crítica exige que el médico pueda reconocer los sesgos más frecuentes y entender cómo afectan la validez de los resultados.

7.1 SESGOS

Sesgo de selección: Ocurre cuando los grupos no son comparables desde el inicio. La aleatorización busca evitarlo, pero si la secuencia de asignación no se genera o no se oculta adecuadamente, los investigadores pueden influir de manera consciente o inconsciente en qué pacientes ingresan a cada brazo.

Sesgo de desempeño: Aparece cuando los pacientes o investigadores saben qué tratamiento reciben y esto modifica su conducta. El cegamiento (simple, doble o triple) reduce este riesgo. Por ejemplo, un paciente que sabe que está en el grupo “activo” puede mejorar su adherencia, mientras que uno en el grupo “placebo” podría desanimarse.

Sesgo de detección: Se produce cuando los evaluadores conocen la asignación de los pacientes y su juicio clínico o interpretación de pruebas se ve influenciada. El cegamiento de quienes evalúan los desenlaces es clave para minimizarlo.

Sesgo de desgaste (attritionbias): Resulta de las pérdidas durante el seguimiento. Si muchos pacientes abandonan y estos no se distribuyen de forma equilibrada entre los grupos, el efecto observado puede estar distorsionado. El análisis por intención de tratar ayuda a mitigar este sesgo.

Sesgo de reporte: Ocurre cuando los investigadores seleccionan qué desenlaces publicar (generalmente los favorables), omitiendo los negativos o neutros. La transparencia y el registro previo del protocolo en plataformas como ClinicalTrials.gov son esenciales para evitarlo.

7.2 LIMITACIONES

Generalización restringida: los pacientes incluidos suelen ser seleccionados con criterios estrictos. Por ello, los resultados pueden no reflejar lo que ocurre en la práctica clínica real con poblaciones más heterogéneas.

Duración insuficiente: algunos ensayos no siguen a los pacientes el tiempo suficiente para detectar desenlaces relevantes (ejemplo: mortalidad a largo plazo).

Uso de desenlaces sustitutos: en lugar de medir directamente la mortalidad o la calidad de vida, se utilizan variables intermedias (p. ej., niveles de laboratorio), que no siempre predicen de manera confiable el beneficio clínico.

Conflictos de interés: los estudios financiados por la industria pueden tener sesgos de diseño, análisis o reporte a favor de la intervención.

Costo y complejidad: los ECA requieren grandes recursos, lo que limita su viabilidad para todas las preguntas de investigación.

El lector crítico debe recordar que ningún ensayo es perfecto. La clave está en reconocer qué sesgos pudieron intervenir, cómo los autores intentaron controlarlos y qué impacto real tienen en la validez interna y en la aplicabilidad externa de los resultados; todo esto permitirá tener un análisis crítico para decidir si es aplicable o no a la práctica médica que se enfrenta.

8. CHECKLIST

Este checklist está pensado para que, al leer un artículo, se pueda marcar si cada componente está adecuadamente reportado (Sí/No/Parcial). Se basa en las guías CONSORT 2010 (25 ítems) y su explicación–elaboración; lo complementa con SPIRIT 2013 (para cotejar el protocolo), ICH-E9(R1) (estimands y sensibilidad) y requisitos de registro OMS/ICTRP (transparencia).

Tabla 17. Checklist internacional para lectura crítica de ECA

Sección	Subsección	Criterios / Contenido (sin modificar el texto)
A. Título y resumen	—	Identifica explícitamente el estudio como ensayo aleatorizado.
		Resumen estructurado: diseño, población, intervención, comparador, desenlaces, resultados principales con medidas de efecto e intervalos de confianza, y conclusiones.
B. Introducción	—	Fundamento: justificación científica clara.
		Objetivos/hipótesis: precisos y con el marco declarado (superioridad, no inferioridad, equivalencia).
C. Métodos	1) Diseño del ensayo	Tipo: paralelo, factorial, crossover, por conglomerados (clúster), adaptativo, pragmático, multi-brazo; razón de asignación.
		Si es no inferioridad/equivalencia: margen (Δ) definido y justificado clínicamente, estrategia analítica preespecificada (IC una/dos colas, poblaciones ITT/PP).
		Si es adaptativo: reglas de adaptación, control del error tipo I, análisis interinos y plan estadístico previo.
		Si es factorial: definición de efectos principales e interacciones y cómo se analizarán.
		Si es crossover: periodos, efecto de arrastre y periodo de lavado.
		Si es clúster: número y tamaño de clústeres, ICC en cálculo muestral y análisis.
	2) Participantes y entorno	Criterios de inclusión/exclusión claros; escenarios (hospital/APS/país/centros); período de reclutamiento.
	3) Intervenciones	Detalle suficiente para replicar (dosis, pauta, vía, duración); cointervenciones permitidas/prohibidas; adherencia. Idealmente, seguir una guía de descripción estandarizada (p. ej., TIDieR).
	4) Desenlaces	Primarios y secundarios preespecificados (definición operativa, momento de medición, dirección clínica del efecto).
		Si hay sustitutos: declarar validez y relevancia.
	5) Tamaño muestral y monitoreo	Supuestos usados (tamaño de efecto/varianza/tasa de eventos/Δ si aplica), pérdidas esperadas, planes de interinos y DSMB (comité de monitoreo de datos).
	6) Aleatorización y ocultamiento	Generación de secuencia (tipo y restricciones).
		Ocultamiento de la asignación (mecanismo).
		Responsabilidades: quién generó/asignó/inscribió a los participantes.
	7) Cegamiento	Quién estuvo ciego (pacientes, personal, evaluadores) y cómo se garantizó; similitud de comparadores.
	8) Plan estadístico (SAP) y estimand	Métodos para el desenlace primario y secundarios; poblaciones de análisis (ITT/PP).
		Manejo de datos faltantes y análisis de sensibilidad; control de multiplicidad; subgrupos preespecificados.
		Estimand definido: población objetivo, variable, eventos intercurrentes y medida-resumen; análisis principal y sensibilidades alineadas con ese estimand.
	9) Ética, registro y protocolo	Aprobación ética; registro prospectivo con número visible; acceso a protocolo/versión y a enmiendas.
D. Resultados	—	Flujograma (diagrama CONSORT): aleatorizados, tratados, analizados, pérdidas/exclusiones (con causa).
		Cronograma de reclutamiento/seguimiento; razones de detención si hubo.
		Características basales por grupo (tabla 1) con comparabilidad.
		N analizado por desenlace y población (ITT/PP).
		Resultados del desenlace primario con medidas de efecto (absolutas y relativas) y IC; claridad del horizonte temporal.
		Para no inferioridad/equivalencia: figura con IC y margen Δ (límites visibles).
		Análisis adicionales: subgrupos/ajustes; distinguir prespecificados de exploratorios.
		Daños/seguridad: definiciones, gravedad/seriedad, denominadores por grupo, poblaciones de seguridad y método de captura de eventos.
E. Discusión		Limitaciones: sesgos potenciales, imprecisión, multiplicidad; generalizabilidad (aplicabilidad).
		Balance beneficios-daños y contexto de evidencia (comparación con estudios previos y guías).

Sección	Subsección	Criterios / Contenido (sin modificar el texto)
F. Otra información		Financiación y rol del patrocinador (diseño, análisis, escritura, publicación).
		Conflictos de interés y acceso a datos/código si corresponde.
Extensiones útiles para diseños especiales (marcar cuando apliquen)		No inferioridad/Equivalencia: margen Δ, justificación clínica, coherencia ITT/PP, gráfica con Δ; interpretación centrada en márgenes.
		Clúster: ICC y su IC, número/tamaño de clústeres, flujo de clústeres e individuos, ajuste por clustering.
		Pragmáticos: flexibilidad de intervención y comparador, cointervenciones, adherencia, representatividad del entorno.
		Adaptativos: reglas de adaptación, simulaciones, control de error, enmascaramiento operacional, independencia del DSMB.
		Multi-brazo: control de multiplicidad y jerarquía de comparaciones.
		Factoriales: definición y análisis de interacción.
		Crossover: periodos, efectos de secuencia/arrastre, periodo de lavado.
		Sustitutos: validez del marcador proxy y coherencia con desenlaces duros.
		Resultados reportados por pacientes (PRO/QoL): instrumentos validados, mínimos cambios clínicamente importantes.
Doble verificación con protocolo y registro		Antes de cerrar la lectura crítica:
		1. Contrasta el artículo con su protocolo (objetivos, desenlaces, SAP, manejo de intercurrentes, adherencia).
		2. Verifica el registro prospectivo (fechas, población, desenlaces declarados) y su consistencia con lo publicado.
		3. Comprueba que el estimand esté definido y que existan análisis de sensibilidad coherentes.

Fuente: Elaboración propia

9. UTILIDAD DE LOS ECA

Los ECA son considerados el estándar de oro para evaluar la eficacia y seguridad de las intervenciones en salud. Su utilidad radica en que, cuando están bien diseñados y ejecutados, ofrecen la mejor evidencia posible para guiar la práctica clínica. Sin embargo, también tienen limitaciones que el lector crítico debe conocer para aplicarlos con criterio.

9.1 UTILIDAD

Las aportaciones principales son:

- **Causalidad confiable:** Gracias a la aleatorización, los ECA reducen el sesgo de selección y equilibran variables conocidas y desconocidas entre grupos. Esto permite atribuir con mayor certeza las diferencias de resultados a la intervención estudiada.
- **Medidas directas de eficacia y seguridad:** Un ECA bien diseñado permite estimar el beneficio absoluto y relativo de un tratamiento, así como los riesgos de eventos adversos. Esto se traduce en indicadores clínicos como el NNT y el NNH, que ayudan a decidir si una intervención merece ser utilizada en la práctica.
- **Base para guías clínicas:** La mayoría de las recomendaciones terapéuticas de sociedades científicas y guías internacionales se apoyan en ECA de alta calidad. Sin ellos, la medicina se basaría en observaciones menos confiables o en experiencia personal.
- **Protección del paciente:** Los ECA incluyen mecanismos éticos (consentimiento informado, comités de ética, monitoreo de seguridad) que buscan equilibrar la innovación con la seguridad del participante.

9.2 LIMITACIONES A CONSIDERAR

A pesar de su relevancia, los ECA no son perfectos:

- Suelen incluir poblaciones muy seleccionadas, lo que limita la generalización a la práctica diaria.
- Tienen costos elevados y duración prolongada, por lo que muchas preguntas clínicas importantes quedan sin respuesta.
- Algunos ensayos priorizan desenlaces sustitutos (biomarcadores, cifras de laboratorio) en lugar de desenlaces clínicamente relevantes como mortalidad o calidad de vida.
- Existe el riesgo de conflictos de interés cuando son financiados por la industria.

9.3 APLICABILIDAD EN LA PRÁCTICA CLÍNICA

El lector crítico debe recordar que la verdadera utilidad de un ECA no se mide solo en términos de significancia estadística, sino en su capacidad de responder:

- ¿Este tratamiento ofrece un beneficio clínico relevante para mis pacientes?
- ¿El riesgo de efectos adversos es aceptable?
- ¿El contexto de mi práctica permite aplicar lo estudiado (disponibilidad, costos, adherencia, sistema de salud)?

10. UTILIDAD

La aplicación de los hallazgos del ensayo clínico en la Sra. Julia trasciende el control de la hemoglobina glicosilada. La utilidad clínica de los iSGLT2 en este escenario se traduce en una nefroprotección activa y una reducción significativa del riesgo cardiovascular. Al trasladar la evidencia a la práctica, el médico no solo trata una cifra laboratorial, sino que interviene en el pronóstico vital de la paciente, disminuyendo la probabilidad de progresión a enfermedad renal crónica terminal y hospitalizaciones por insuficiencia cardíaca. Esta sección permite al clínico decidir, con base en la magnitud del beneficio (RRA y NNT), si la intervención justifica el costo y los posibles efectos adversos en su contexto local.

Capítulo 13. Ensayos clínicos aleatorizados (ECA) 

Resumen práctico para la lectura crítica

1 ¿Qué es un ECA?  Es el diseño de referencia para evaluar la eficacia y seguridad de una intervención. La asignación aleatoria reduce sesgos y permite establecer relaciones causales con alta validez interna.	2 Objetivos principales  • Evaluar eficacia terapéutica • Evaluar seguridad • Establecer causalidad • Generar evidencia para guías clínicas • Optimizar la toma de decisiones	3 Elementos metodológicos clave  • Tipo de diseño: paralelo, crossover, factorial o por conglomerados • Marco del ensayo: superioridad, no inferioridad o equivalencia • Participantes: criterios de inclusión y exclusión • Aleatorización y ocultamiento de la asignación • Cegamiento: simple, doble o triple • Intervención, seguimiento y adherencia • Desenlace primario, secundarios y seguridad • Análisis estadístico: ITT, PP y manejo de datos faltantes	
4 Validez Validez interna  Depende de una aleatorización adecuada, ocultamiento, cegamiento, seguimiento completo y análisis coherente. Validez externa  Indica qué tanto los resultados pueden aplicarse a pacientes y contextos reales de práctica clínica.	5 Interpretación estadística  • En variables dicotómicas: RR, RAR, RRR, NNT y NNH • En variables continuas: diferencia de medias y análisis ajustados • En tiempo al evento: curvas de Kaplan-Meier y hazard ratio (HR) • El IC95% muestra magnitud y precisión; el valor p no debe interpretarse aisladamente • La significancia estadística no siempre implica relevancia clínica	6 Tipos de ECA  Superioridad: demuestra que una intervención es mejor que el comparador  No inferioridad: demuestra que no es clínicamente peor dentro de un margen Δ  Equivalencia: demuestra efectos clínicamente similares dentro de $\pm\Delta$	
7 Subgrupos y robustez  Los subgrupos deben ser prespecificados y se interpretan con cautela. La robustez se valora comparando ITT, PP y análisis de sensibilidad.	8 Sesgos y limitaciones  • Sesgo de selección • Sesgo de desempeño • Sesgo de detección • Sesgo de diagnóstico • Sesgo de reporte Limitaciones: poblaciones seleccionadas, desenlaces sustitutos, costos altos y conflictos de interés.	9 Checklist para lectura crítica  Verificar: título y resumen, introducción, métodos, resultados, discusión, financiación, registro, protocolo, CONSORT/SPRIT y coherencia entre artículo y protocolo.	10 Utilidad clínica  Un ECA útil no solo debe mostrar significancia estadística, sino un beneficio clínico relevante, un perfil de seguridad aceptable y aplicabilidad al contexto del paciente.

 **Perlas clave**
 No confundir RRR con RAR
  Revisar siempre el IC95%
  Valorar NNT y NNH
  Vigilar pérdidas de seguimiento
  Diferenciar significancia estadística de relevancia clínica

 Resumen didáctico del capítulo para personal de salud

11. PREGUNTAS DE AUTOEVALUACIÓN

PREGUNTAS DE SELECCIÓN MÚLTIPLE

1. ¿Cuál es la principal ventaja de la aleatorización en un ECA? a) Reducir el tamaño muestral necesario b) Equilibrar variables conocidas y desconocidas entre grupos c) Eliminar la necesidad de cegamiento d) Evitar el reporte de pérdidas

2. En un ensayo de no inferioridad, ¿qué condición debe cumplirse para concluir no inferioridad? a) Que el IC no cruce el valor nulo (1 o 0) b) Que el IC se mantenga por debajo del margen Δ preespecificado c) Que el valor p sea $<0,05$ d) Que las pérdidas de seguimiento sean menores al 5%

3. ¿Qué medida expresa el número de pacientes que deben ser tratados para evitar un evento adverso? a) RRR b) RAR c) NNT d) NNH

4. En un análisis de supervivencia, ¿qué representa el Hazard Ratio (HR)? a) El riesgo acumulado al final del estudio b) El riesgo instantáneo de evento en cualquier momento del seguimiento c) El porcentaje de pacientes que completaron el protocolo d) El número necesario para causar un evento adverso

5. ¿Cuál de los siguientes es un sesgo que puede ocurrir en los ECA? a) Sesgo de selección b) Sesgo de desempeño c) Sesgo de detección d) Todas las anteriores

PREGUNTAS ABIERTAS

6. Explique con sus propias palabras la diferencia entre un ensayo de superioridad y uno de no inferioridad.

7. ¿Por qué se considera el análisis por intención de tratar (ITT) como el estándar en los ECA?

8. ¿Cómo se interpreta clínicamente el NNT y qué limitaciones tiene como medida de efecto?

9. ¿Qué información aporta una curva de Kaplan-Meier que no puede obtenerse a partir de un RR acumulado al final del seguimiento?

10. Mencione tres errores frecuentes en la interpretación de un ECA y explique cómo evitarlos en la práctica clínica.

RESPUESTAS:

1. b) Equilibrar variables conocidas y desconocidas entre grupos. La aleatorización distribuye de forma aleatoria tanto los factores medibles como los no medibles entre los grupos, reduciendo el sesgo de selección y permitiendo que las diferencias en los resultados se atribuyan con mayor validez a la intervención.

2. b) Que el IC se mantenga por debajo del margen Δ preespecificado. En los ensayos de no inferioridad, el objetivo no es demostrar diferencia, sino que la intervención no sea peor que el comparador más allá de un margen clínicamente aceptable (Δ). Para ello, todo el intervalo de confianza debe permanecer dentro de ese margen.

3. c) NNT. El número necesario a tratar (NNT) indica cuántos pacientes deben recibir una intervención para prevenir un evento adicional. Es una medida directa de aplicabilidad clínica, derivada de la reducción absoluta del riesgo.

4. b) El riesgo instantáneo de evento en cualquier momento del seguimiento. El hazard ratio (HR) representa la razón de tasas de ocurrencia del evento a lo largo del tiempo, considerando el momento en que ocurren los eventos, no solo su frecuencia acumulada.

5. d) Todas las anteriores. Los ensayos clínicos pueden presentar diferentes tipos de sesgo: de selección (si la aleatorización falla), de desempeño (si no hay cegamiento adecuado) y de detección (si la evaluación del desenlace no es objetiva).

6. Diferencia entre ensayo de superioridad y no inferioridad. Un ensayo de superioridad busca demostrar que una intervención es mejor que otra en términos de un desenlace clínico. En cambio, un ensayo de no inferioridad pretende demostrar que una intervención no es significativamente peor que el tratamiento estándar dentro de un margen clínicamente aceptable. Mientras el primero busca diferencia, el segundo busca equivalencia práctica en eficacia.

7. Importancia del análisis por intención de tratar (ITT). El análisis por intención de tratar incluye a todos los pacientes según el grupo al que fueron originalmente asignados, independientemente de si completaron o no la intervención. Esto preserva los beneficios de la aleatorización, evita sesgos y refleja mejor la efectividad en condiciones reales de práctica clínica.

8. Interpretación clínica del NNT y sus limitaciones. El NNT indica cuántos pacientes deben tratarse para prevenir un evento adicional, lo que permite dimensionar el beneficio de una intervención en términos prácticos. Sin embargo, tiene limitaciones: depende del riesgo basal de la población, puede variar según el tiempo de seguimiento y no refleja directamente los posibles efectos adversos.

9. Aporte de la curva de Kaplan-Meier. La curva de Kaplan-Meier permite visualizar la

probabilidad de permanecer libre de un evento a lo largo del tiempo, mostrando cuándo ocurren los eventos y cómo se separan los grupos durante el seguimiento. Esto no puede obtenerse únicamente con un riesgo relativo acumulado al final del estudio.

10. Errores frecuentes en la interpretación de un ECA y cómo evitarlos. Un error frecuente es confundir la reducción relativa del riesgo con la absoluta, lo que puede sobreestimar el beneficio; esto se evita revisando siempre el RAR y el NNT. Otro error es enfocarse únicamente en el valor p , ignorando el intervalo de confianza, por lo que se debe priorizar la magnitud y precisión del efecto. Finalmente, ignorar las pérdidas de seguimiento puede sesgar los resultados, por lo que es fundamental evaluar su magnitud y cómo fueron manejadas (por ejemplo, mediante análisis ITT o de sensibilidad).

REFERENCIAS

- CONSORT Group. CONSORT 2025 statement: Updated guideline for reporting randomised trials. *BMJ*. 2025. Available at: <https://pmc.ncbi.nlm.nih.gov/articles/PMC11996237/>.
- Chan A-W, Mayo-Wilson E, Li T, et al. SPIRIT 2025 statement: Harmonized protocol guidance for randomized trials. *JAMA*. 2025. Available at: <https://doi.org/10.1001/jama.2025.8411>.
- Butcher NJ, Monsour A, Mew EJ, et al. Guidelines for reporting outcomes in trial reports: The CONSORT-Outcomes 2022 extension. *JAMA*. 2022;328(22):2252–64. Available from: <https://jamanetwork.com/journals/jama/fullarticle/2799401/>.
- Junqueira DR, Zorzela L, Golder S, et al. CONSORT Harms 2022 statement, explanation, and elaboration: Updated guideline for the reporting of harms in randomized trials. *BMJ*. 2023;381:e073725. Available at: <https://doi.org/10.1136/bmj-2023-073725>.
- Kahan BC, Elbourne D, Hall SS, et al. Reporting of factorial randomized trials: Extension of the CONSORT 2010 statement. *BMJ*. 2025;388:e081123. Available at: <https://www.equator-network.org/reporting-guidelines/reporting-of-factorial-randomized-trials-extension-of-the-consort-2010-statement/>.
- Sterne JAC, Savović J, Page MJ, et al. RoB 2: A revised tool for assessing risk of bias in randomised trials. *BMJ*. 2019;366:l4898. Available at: <https://doi.org/10.1136/bmj.l4898>.
- Higgins JPT, Thomas J, Chandler J, et al., editors. Cochrane handbook for systematic reviews of interventions. Version 6.5 (updated 2024). Chapter 8, Assessing risk of bias in a randomized trial (RoB 2). Available at: <https://www.cochrane.org/es/authors/handbooks-and-manuals/handbook/current/chapter-08>.
- International Council for Harmonisation (ICH). E9(R1) addendum on estimands and sensitivity analysis in clinical trials (Step 5). 2020. Available from: <https://www.ema.europa.eu/en/ich-e9-statistical-principles-clinical-trials-scientific-guideline>.
- Mauri L, D'Agostino RB Sr. Challenges in the design and interpretation of noninferiority trials. *N Engl J Med*. 2017;377(14):1357–67. Available at: <https://doi.org/10.1056/NEJMr1510063>.
- Schandelmaier S, Briel M, Varadhan R, et al. Development of the instrument to assess the credibility of effect modification analyses (ICEMAN) in randomized controlled trials and meta-analyses. *CMAJ*. 2020;192(32):E901–6. Available at: <https://doi.org/10.1503/cmaj.200077>.
- Kahan BC, Hindley J, Edwards M, et al. The estimands framework: A primer on the ICH E9(R1) addendum. *BMJ*. 2024;384:e076316. Available at: <https://doi.org/10.1136/bmj-2023-076316>.
- Dumas E, Stensrud MJ. How hazard ratios can mislead and why it matters in practice. *Eur J Epidemiol*. 2025;40(6):603–9. Available at: <https://doi.org/10.1007/s10654-025-01250-9>.