

Neurología Argentina

www.elsevier.es/neurolarg



Revisión

Validez externa de los ensayos clínicos en Neurología

Pablo Ioli

Neurólogo de planta, Coordinador de la Unidad de Investigación Clínica, Servicio de Neurología, Hospital Privado de Comunidad, Mar del Plata, Argentina

INFORMACIÓN DEL ARTÍCULO

Historia del artículo:

Recibido el 21 de septiembre de 2010

Aceptado el 30 de octubre de 2010

Palabras clave:

Ensayos clínicos aleatorizados

Validez externa

Generalizabilidad de los ensayos clínicos

Desenlaces

Escala neurológicas

Tratamiento

Enfermedades neurológicas

R E S U M E N

Introducción: Aun cuando un ensayo clínico aleatorizado (ECA) sea metodológicamente correcto, queda el desafío de cómo trasladar y valorar los resultados en nuestro paciente. Esta capacidad de generalizar y extender a la práctica diaria dichos resultados es lo que se llama validez externa de los ECA.

Objetivos: En esta revisión analizaremos los aspectos salientes que deben ser tenidos en cuenta para ponderar la distancia y la aplicabilidad de tratamientos provenientes de ECA a nuestros pacientes respecto de los participantes en tales estudios.

Desarrollo: Abordaremos el proceso de reclutamiento, cómo son evaluados y tratados los pacientes de los estudios clínicos, y finalmente cómo se analizan sus desenlaces y cómo éstos pueden ser traducidos en medidas de efecto con mayor significancia clínica.

Conclusiones: Los factores que son relevantes en la evaluación de la validez externa de un ECA son muchos y variados. Valorar el conocimiento de estos factores y la lectura crítica pueden ayudar a interpretar los ECA para aplicarlos en nuestros pacientes.

© 2010 Sociedad Neurológica Argentina. Publicado por Elsevier España, S.L.

Todos los derechos reservados.

External validity of clinical trials in Neurology

A B S T R A C T

Keywords:

Randomized controlled trials

External validity

Clinical trials generalizability

Outcomes

Neurological scales

Treatment

Neurological diseases

Introduction: Even when a randomized controlled trial (RCT) is methodologically correct, the challenge of extending and weighing the results in our patient's treatment remains. The capability of generalization and extension of these results to clinical practice is the so called external validity of RCT.

Objective: In this narrative review, we will analyze the cornerstone topics that must be taken into account in order to measure the distance and applicability of treatments coming from RCTs to our patients in relation to the participants of the trials.

Development: We will assess the recruitment process, how patients are evaluated and treated in clinical trials. We will subsequently study how the outcomes of such trials are analyzed and how they can be translated in measures of effects with more clinical meaning.

Correo electrónico: pabloioli@yahoo.com.ar (P. Ioli).

Conclusions: The relevant factors in the evaluation of external validity in an RCT are multiple and diverse. Knowing these factors and making a critical interpretation of clinical trials can help to understand the RCTs in order to apply them to our patients.

© 2010 Sociedad Neurológica Argentina. Published by Elsevier España, S.L.
All rights reserved.

Introducción

Cuando habitualmente analizamos los aspectos metodológicos de los ensayos clínicos estudiamos la presencia de sesgos, es decir, errores sistemáticos en el diseño de un estudio. Evaluamos si la muestra es representativa de la patología que se estudia, si el proceso de aleatorización es correcto, si los pacientes en cada grupo están bien balanceados, si el seguimiento y medición de los desenlaces es adecuado, o avanzamos en la utilidad de tal o cual prueba estadística. Aun cuando vayamos hasta los detalles más sutiles de un ensayo clínico aleatorizado (ECA) o de un estudio de pronóstico, nos queda el paso más importante por delante: cómo aplicar, y valorar hasta donde esto es posible, los resultados a nuestro paciente. Esta capacidad de generalizar y extender a la práctica diaria dichos resultados es lo que se llama validez externa de los ECA.

Es bien conocida la relación inversa entre validez interna de un ECA y su validez externa¹ (fig. 1).

A mayor calidad metodológica del estudio, más “puros”, selectos y virtuales serán los pacientes, y más probablemente nuestros pacientes de todos los días diferirán de los participantes y de las condiciones en las que fueron estudiados. Esta distancia puede en ocasiones ser considerable². Imagine, por ejemplo, que un joven quiere conquistar una señorita y, buscando la evidencia con la más alta eficacia demostrada, se dedica a mirar películas de James Bond. Bajo esas condiciones ideales de conquista, la eficacia es muy elevada, pero muy probablemente él no sea tan apuesto como el agente, su auto no sea deportivo, descapotable, ni siquiera último modelo, su trabajo no va a salvar al mundo y probablemente tenga un objetivo mucho menos contorneado. Un buen ejemplo de alta validez interna y poca generalizabilidad. Frecuentemente esta decepción, la de la generalización de los ECA, nos invade cuando vemos algunos estudios.

Valorar la validez externa es entonces preguntarse hasta dónde los resultados de un estudio son trasladables a individuos de una población diferente de la que participó en el mismo. Por otra parte, y desde un punto de vista estadístico, podríamos decir que un estudio tiene validez externa si sus

resultados difieren de los resultados que se hubieran obtenido idealmente en toda la potencial población a la que la intervención está dirigida, sólo por error atribuido al azar³.

En este artículo revisaremos algunos aspectos que nos pueden facilitar la extensión de la evidencia médica a nuestra práctica diaria.

Lugares y ámbitos del estudio y de fuera del estudio

Dentro de un mismo estudio multicéntrico y multinacional suele haber diferencias en cómo y cuándo se aplica una intervención. En el European Carotid Surgery Trial (ECST), el estudio europeo que comparó endarterectomía carotídea versus tratamiento médico en pacientes con estenosis carotídea sintomática⁴, existieron, por ejemplo, centros “rápidos” en los que la endarterectomía se realizó antes de los 51 días de la aleatorización del paciente y que presentaron una reducción del riesgo absoluto de stroke isquémico ipsilateral a 5 años mucho mayor, para estenosis tanto moderadas como para severas, que en los llamados centros “lentos”. Esto no fue reflejado en la publicación original, de manera que, amparado en un resultado global favorable, la evolución real podía depender de en qué país y en qué sistema de salud (determinante de la demora hasta la endarterectomía) se iba a realizar la intervención².

En el IST, por ejemplo, el 50% de los pacientes italianos recibieron glicerol, contra el 3% en el resto de los países, y otras diferencias importantes también involucraron otros tratamientos, como corticoides o hemodilución⁵.

Así, los ECA realizados en un país son habitualmente generalizables a otros, pero tal generalizabilidad no debería darse siempre por garantizada².

Incluso los mismos centros que participaron en un estudio pueden tener resultados diferentes en una misma intervención y tipo de pacientes dentro del contexto de un ECA y fuera de él. Es conocido caso del estudio ACAS, en el que los mismos equipos quirúrgicos tuvieron mortalidades y complicaciones diferentes durante el estudio y fuera del mismo^{6,7}.



Figura 1 – Relación inversa entre la validez interna de un ECA y su validez externa.

Pacientes elegidos y aleatorizados versus pacientes del día a día

Debido a que no podría existir un ECA para cada situación y para cada paciente (tampoco sería exigible), un primer ejercicio podría ser comparar las características de nuestro paciente con las de las tablas del/de los estudio/s que determinaron la eficacia del tratamiento. No obstante, no necesariamente la

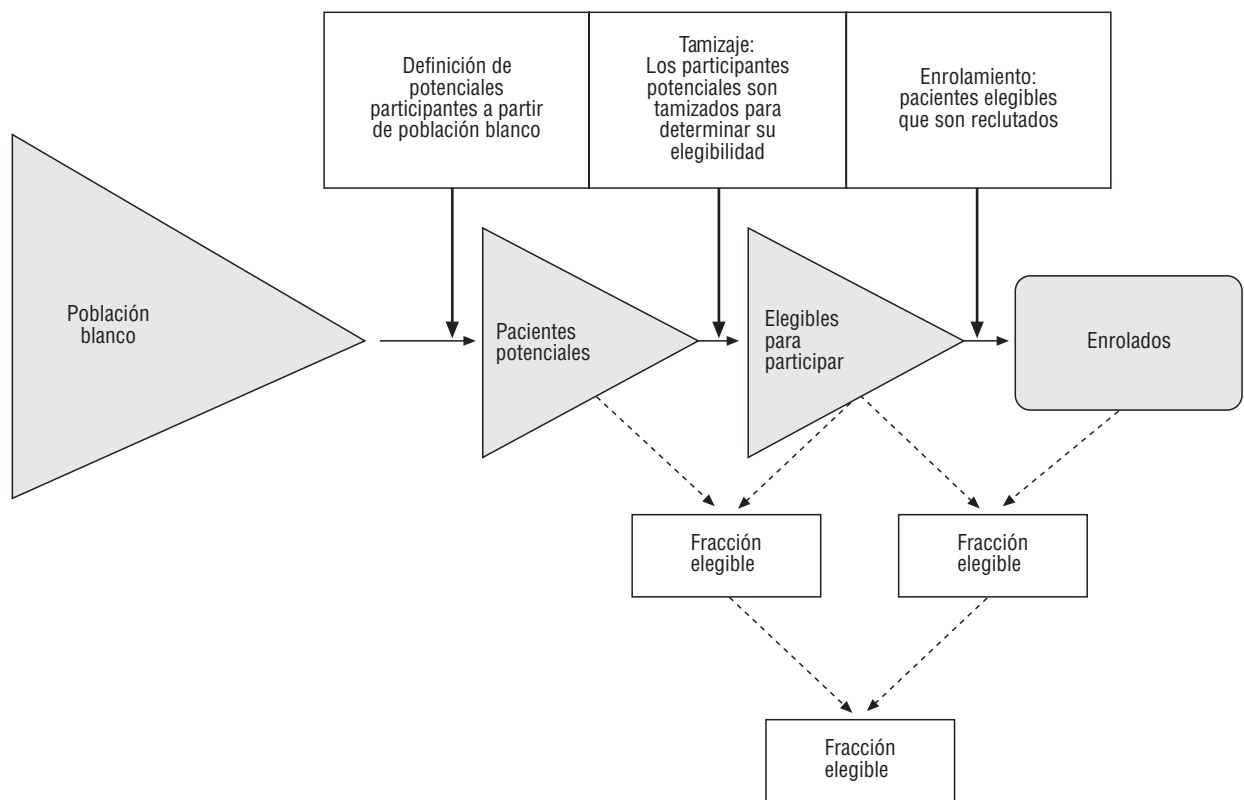


Figura 2 – Cadena de filtros por la que pasan los pacientes participantes en un estudio.

potencial semejanza de nuestro paciente al paciente “promedio” del estudio asegura que se comporte de la misma manera, ya sea porque el tratamiento no es eficaz (ni ineficaz) en todos los pacientes, o porque puede diferir en alguna característica no tenida en cuenta pero médicamente relevante.

La mayoría de los ECA no dan información acerca del proceso de reclutamiento, al menos en su publicación⁸, a pesar de que éste es un factor determinante en el establecimiento de la validez externa de un estudio⁹. De hecho, esto ha sido una crítica al CONSORT original, que posteriormente fue revisado^{10,11}.

La cadena de filtros que terminan en el paciente participante en un estudio comienza con la definición de la *población blanco* (p. ej.: todos los pacientes con diagnóstico de enfermedad de Parkinson [EP] de una región y tiempo dado), luego se definen los *participantes potenciales* (pacientes con EP de tal región, de novo, vírgenes de tratamiento, que son los que van a ser tamizados). El cociente entre estos dos primeros determina la *fracción elegible*. Posteriormente, los realmente *elegibles para participar* (los que cumplían todos los criterios de inclusión/exclusión), y finalmente los *enrolados*. El cociente entre estos dos últimos determina la *fracción enrolada*, y el producto de estas dos fracciones determina la *fracción reclutada (FR)*. Por definición, la FR representa la proporción de pacientes participantes potenciales que son finalmente enrolados y permite tener una estimación de la distancia entre el espectro de pacientes con una enfermedad respecto de los reclutados en el estudio. No se ha establecido la FR adecuada para estudios en Neurología, y probablemente ésta varíe según la patología,

pero por ejemplo en oncología se estima que oscila entre el 18 y el 22%¹² (fig 2).

Pero aun en los pacientes enrolados puede haber otros filtros, como las llamadas “estrategias de enriquecimiento”. Se trata de pacientes que fueron enrolados (firman el consentimiento informado) y, luego de cumplir criterios de selección, son descartados debido a que tras un período basal (*run-in*) de observación no terminan de cumplir con requerimientos específicos del estudio (criterios de inclusión), como, por ejemplo, una complementación no adecuada del diario de fluctuaciones motoras en pacientes con EP (una circunstancia en la que no sólo se requiere del paciente sino de un acompañante lúcido). Esta situación es particularmente frecuente en epilepsia, demencia y EP. En un estudio de tacrina en la enfermedad de Alzheimer se reclutaron 632 pacientes durante un período de 6 semanas a diferentes dosis de tacrina o placebo y, tras un período de lavado, sólo 215 (34%), que habían tenido mejoría con tacrina, fueron aleatorizados a ese tratamiento, es decir, que los finalmente ingresados ya tenían una primera buena respuesta¹³.

En el estudio PROACT II se ilustra la importancia de conocer cómo fue el reclutamiento¹⁴. Los autores describen con detalle este proceso en el que se tamizaron 12.323 pacientes con stroke isquémico agudo, de los cuales 10.893 (83%) fueron excluidos por no cumplir ciertos criterios de inclusión/exclusión (4.053 tenían más de 6 h de evolución, 2.410 tenían déficits leves o que mejoraron rápidamente, 696 eran mayores de 85 años, entre otros factores), más otros criterios de exclusión tomográficos. Quedaron 474 pacientes elegibles (fracción elegible: 3,8%) que fueron sometidos a angiografía, y de ellos 180

fueron enrolados (fracción enrolada: 38%), es decir, una FR de 1,4%, o se necesitaron tamizar 68 pacientes para que uno recibiera la intervención. Esto da una buena medida de la aplicabilidad de una terapia y sobre qué aspectos se deben tomar las decisiones en políticas de salud (priorizar políticas de acceso rápido a emergencias antes que comprar angiografía).

Esta distancia entre pacientes reclutados y potenciales participantes ha sido también establecida para antiagregación en prevención secundaria de stroke isquémico y accidente isquémico transitorio. Aplicando los criterios de inclusión/exclusión de 7 grandes estudios de prevención secundaria en stroke en una serie consecutiva de pacientes, se halló que del 25 al 67% hubieran cumplido con dichos criterios y que la mortalidad fue superior entre los pacientes inelégibles que entre los participantes¹⁵.

Se han diseñado algunas herramientas, como el GATE (Graphic Appraisal Tool for Epidemiology), para evaluar cuán generalizables pueden ser los resultados de un estudio particular a un paciente^{16,17}.

Otro aspecto relacionado con el reclutamiento es el tipo de pacientes incluidos, teniendo en cuenta su riesgo basal. En un interesante estudio se analizó cómo podrían afectarse el poder estadístico y la validez externa de grandes ECA en prevención secundaria de stroke isquémico, haciendo el ejercicio de estratificar el riesgo basal (de recurrencia de stroke) de los pacientes incluidos en otros estudios. Concretamente se estratificaron los pacientes de los estudios UK-TIA Aspirin Trial¹⁸, Dutch TIA Trial¹⁹, ECST²⁰, Oxford TIA²¹ Cohort, en función del riesgo de los pacientes incluidos en los estudios PROFESS²², MATCH²³ y SPORTIF^{24,25}. El análisis sugiere que aunque la inclusión de pacientes con mayor riesgo puede hacer que el estudio requiera menos pacientes y menos tiempo de seguimiento, el efecto sobre estos dos aspectos es mucho respecto de la importante pérdida de generalizabilidad de los resultados. El estudio afirma que, por ejemplo, reclutar individuos con más de un factor de riesgo vascular no justifica, desde un punto de vista estadístico, la exclusión de pacientes con un riesgo basal menor que haría que el espectro de la población incluida aumentase la validez externa.

Pacientes tratados en el estudio y tratados en la práctica diaria

El problema de la adherencia al tratamiento en la práctica diaria y la distancia con el habitual correcto compromiso con el fármaco en estudio por parte de los pacientes en los ECA es probablemente uno de los aspectos más difíciles de salvar. Ya sea por el interés del paciente en participar, como por los regímenes de visitas, controles y análisis (que pueden incluir dosificaciones del fármaco de investigación, como suele ocurrir en la epilepsia), los exhaustivos diarios acerca del estado del paciente (esto es especialmente común en la EP, en la epilepsia, en la esclerosis múltiple y en la demencia) y el apoyo familiar indispensable en patologías con compromiso funcional, el ámbito en el que un paciente participante recibe su medicación es muy diferente del de la consulta diaria.

Aunque la cantidad de pastillas por día no es la única barrera para la adherencia a la medicación^{26,27}, ésta puede caer un 51% ($\pm 20\%$) cuando se toman cuatro comprimidos por día²⁸.

En un estudio canadiense, la adherencia a anticolinesterásicos en pacientes con Alzheimer fue del 64% al primer año, y la discontinuación subió hasta el 84% a los 40 meses²⁹. En pacientes que reciben inmunomoduladores para el tratamiento de la esclerosis múltiple, casi tres cuartos de ellos pierden al menos una dosis por cada 6 meses de tratamiento, y 1 de cada 10 pierden más de 10 dosis en 6 meses³⁰.

Desenlaces en los estudios y desenlaces que interesan a los pacientes

Frecuentemente los neurólogos utilizamos escalas para cuantificar el compromiso que tiene el paciente. De esa manera estamos midiendo la enfermedad (o al menos algunos aspectos de ella), pero no necesariamente al enfermo. Los pacientes suelen darle más peso a aspectos relacionados con la calidad de vida que a la intensidad o frecuencia de un síntoma. La FDA se ha ocupado de este asunto y desarrolló el documento PROMs (Guidance for Industry Patient-Reported Outcome Measures), donde especifica que un PRO (*patient reported outcome*) es una "medida del estado de salud del paciente que proviene directamente del paciente"³¹. Esta diferencia entre lo que medimos los médicos y lo que a los pacientes les interesa que se mida ha sido demostrada en la esclerosis múltiple, en la que los pacientes aparecen más preocupados por aspectos menos tangibles del compromiso relacionado con su enfermedad, como el estado mental o la vitalidad, en contraposición a las manifestaciones físicas directas³².

Por eso, los desenlaces que se miden en un ECA no necesariamente reflejan un beneficio relevante, aunque medible, para el paciente.

Otro aspecto complejo relacionado se refiere a desenlaces en los que se utilizan escalas y que muestran diferencias estadísticamente significativas pero con relevancia clínica controvertida. Si bien es fundamental que una escala permita detectar, en pequeños cambios, cambios con relevancia clínica, cuanto mayor es la capacidad de una escala para detectar esas pequeñas variaciones (su *responsiveness*), mayor será la probabilidad de que en un ECA se detecten cambios que no son clínicamente relevantes³³. Asimismo, en las escalas ordinales, diferencias de puntuación entre el grupo tratamiento y el grupo control pueden tener menor, mayor o nula importancia médica, dado que según qué se evalúe, cada punto de la escala tiene significados clínicos diferentes. La aplicación de ciertas escalas en Neurología ha sido particularmente criticada por esta tendencia, especialmente UPDRS (Unified Parkinson's Disease Rating Scale)³⁴, EDSS (Kurtzke's expanded Disability Status Scale), RMI (Rivermead Mobility Index) y Ashworth³³.

Otros desenlaces sin escalas ordinales también requieren atención. En epilepsia, la eficacia global de un antiepiléptico respecto del placebo, demostrada en un estudio, puede esconder diferencias pequeñas de efecto en pacientes a ambos extremos del espectro de probabilidad de remisión de crisis,

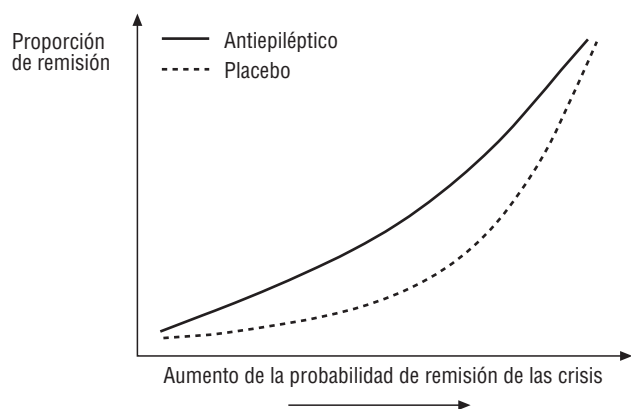


Figura 3 – Respuesta a los antiepilépticos (línea continua) y placebo (línea discontinua) en términos de reducción de crisis.

como pacientes con alta probabilidad de remisión (p. ej., con dos crisis en dos años, EEG y RM normales) y con baja probabilidad (p. ej., un paciente con 10 crisis, examen neurológico, EEG y RM anormales)³⁵. El comportamiento de las crisis en nuestro paciente puede ser más útil para conocer la efectividad del antiepiléptico que el propio tamaño de efecto global establecido para el antiepiléptico en el estudio original (fig. 3).

Asumiendo que los desenlaces y los métodos con los que éstos se miden son los adecuados, cuando leemos un ECA los médicos nos enfrentamos con el obstáculo de interpretar en términos clínicos las medidas de efecto utilizadas y trasladarlas a nuestro paciente individual. El primer paso en este sentido sería analizar el comportamiento de subgrupos con características similares a nuestro paciente. El abordaje debe ser cuidadoso, ya que cuando un ECA muestra que la intervención no es beneficiosa, se pueden encontrar por azar un 7-21% de subgrupos con efecto. Y por el contrario, cuando un ECA muestra que la intervención es beneficiosa, se pueden encontrar por azar un 41-66% de subgrupos sin efecto. Existen muchas reglas que deben ser utilizadas para tal análisis y escapan al alcance de esta revisión, pero constituye una aproximación que puede acercar los ECA al paciente real^{36,37}.

Una segunda estrategia consiste en traducir al lenguaje clínico medidas de efecto, como el odds ratio, el riesgo relativo o la reducción del riesgo absoluto o relativo, entre otras. Quizá la herramienta más útil para ese fin sea el número necesario a tratar (NNT), que es igual a $1/\text{DRA}$ (diferencia de riesgo absoluto) $\times 100$. Una herramienta asociada consiste en ajustar el NNT para diferentes tiempos de seguimiento, como: $\text{NNTs} = \text{NNTT} \times \text{T/S}$, donde S = duración de seguimiento deseada (S) y $\text{NNTT} = \text{NNT}$ para la duración de seguimiento del estudio (T). Por ejemplo, un fármaco con un NNT de 10 para 6 meses para un desenlace dado en el estudio original, si queremos evaluar el potencial beneficio a 2 años en mi paciente: $10 \times 0,5 / 2 = 2,5$, “por cada 2,5 pacientes tratados se beneficia 1 en 2 años”; como regla, cuanto mayor es el tiempo, menor es el NNT. O también, NNT ajustado para diferentes riesgos basales, como: $\text{NNTr} = \text{NNTT} / \text{R}$, donde R = riesgo

basal de mi paciente y $\text{NNTT} = \text{NNT}$ del estudio. Por ejemplo, un fármaco con un NNT de 10 para 6 meses para un desenlace dado en el estudio original. Si se desea evaluar el potencial beneficio en nuestro paciente, que tiene el doble de riesgo que el de los pacientes del ECA original, entonces: $10 / 2 = 5$, “por cada 5 pacientes tratados se beneficia 1 en 6 meses”; cuanto mayor es el riesgo, menor es el NNT³⁸.

Deben considerarse algunas limitaciones de los NNT: los NNTs y NNTr asumen que el riesgo de un desenlace es constante, son población, trastorno, desenlace e intervención específicos, y los NNT derivados de metaanálisis parten de riesgos basales heterogéneos. Además, no hay un valor umbral, y no nos dicen nada acerca de efectos colaterales, costos, valores y preferencias. Una medida relacionada es el llamado costo derivado del NNT (CDNNT), que agrega una aproximación económica como parte del efecto, estableciendo cuál sería el costo de tratar a tal cantidad de individuos para obtener un desenlace favorable o desfavorable. Se calcula como: $\text{NNT} \times \text{costo del tratamiento con el fármaco evaluado}$ (durante el tiempo determinado en el estudio original)³⁹.

Otro abordaje es traducir los resultados de los ECA a los llamados “temas abordados críticamente” (CAT en su sigla en inglés, por *critically appraised topics*); aquí, a partir de una situación clínica real se formula una pregunta contestable con su estructura en la que se definen población (características del individuo), intervención (tratamiento que se evalúa), control (grupo comparador) y desenlace (conocido como paradigma PICO: P: población, I: intervención, C: control, O: outcome-desenlace. En pacientes con deterioro cognitivo leve [P], ¿cuán superior es el donepezilo [I] respecto del placebo [C] para reducir el riesgo de evolución a enfermedad de Alzheimer [O]?). Luego se realiza una revisión sistemática, seguida de una síntesis de la información aplicada al problema planteado. Sus cinco pasos incluyen:

- 1) Generar una pregunta.
- 2) Buscar evidencia adecuada.
- 3) Abordar críticamente la evidencia (validez, intervalos de confianza, impacto).
- 4) Determinar la aplicabilidad de la evidencia.
- 5) Contestar la pregunta original.

Existen varios sitios en donde hay publicados CAT en Neurología y con herramientas para desarrollarlos^{16,40,41}.

Una última opción es realizar un ensayo clínico aleatorizado en un paciente (conocido como ECA n=1), que constituye el máximo nivel de evidencia para el paciente individual. Son estudios experimentales diseñados para evaluar la efectividad de una intervención en un paciente particular. Sus conclusiones sólo son aplicables al paciente en estudio y surgen como una alternativa científica a la tradicional “prueba terapéutica”. Permiten además hacer investigación por uno mismo en lugar de valorar la realizada por otros^{1,42-44}. En nuestro servicio de Neurología hemos llevado adelante algunos ECA n=1; por ejemplo, un estudio en el que a un adolescente con diagnóstico de trastorno de hiperactividad con déficit atencional (ADHD) se le administró en forma doble ciego (ni el médico tratante y los evaluadores, ni el paciente y su familia, sabían

qué recibía cada día) metilfenidato o placebo por día, con asignación aleatorizada de uno u otro, midiendo la respuesta con escalas propias de la enfermedad (escala de Conners), pero también otras ajustadas a las dificultades del paciente, que incluían adaptaciones de la escala para sus docentes y padres. Aquí no se trataba de ver si el metilfenidato era o no eficaz en ADHD, sino si el metilfenidato era o no efectivo en este paciente y para los desenlaces elegidos⁴⁵.

Seguridad en un ensayo clínico y seguridad en la vida real

Lo descrito anteriormente como “estrategias de enriquecimiento” tiene otra consecuencia: los pacientes finalmente incluidos pueden tener una mejor tolerabilidad y menos efectos adversos.

Por otra parte, los ECA están diseñados para demostrar superioridad de un tratamiento sobre otro, y la población de estudio está calculada para mostrar eficacia y no seguridad. Esto se hace más evidente en estudios de no-inferioridad, donde se requiere un menor tamaño de muestra y, por tanto, es menos probable detectar efectos adversos menos frecuentes⁴⁶.

Conclusión

Cada vez que leemos un nuevo ECA nos enfrentamos a la necesidad de valorar la generalizabilidad de sus resultados y de cómo se llegó a ellos. Aun cuando nuestra aproximación a la evidencia de una nueva intervención sea a través de una guía de práctica clínica de una sociedad científica respetable, éstas nos enfrentan a la paradoja de dar recomendaciones de aplicación en poblaciones grandes (de pacientes individuales) que se basan, cada vez más, en “la mejor evidencia disponible”: entiéndase la evidencia con mayor validez interna, por lo que este tipo de guías no reemplazan la necesidad de considerar la validez externa de sus recomendaciones.

Los factores que son relevantes en la evaluación de la validez de un ECA son muchos y variados y con pesos diferentes, y si bien se han desarrollado herramientas para valorarlos^{16,17}, el conocimiento de estos factores y la lectura crítica ayudarán a interpretar los ECA para aplicarlos en nuestros pacientes.

Conflicto de intereses

El autor ha participado como investigador principal en estudios de fase III para el laboratorio Boehringer Ingelheim y Servier, y actualmente para Schering-Plough.

BIBLIOGRAFÍA

- Sackett DL, Richardson WS, Rosenberg W, Haynes RD. Evidence-based Medicine: How to practice and teach EBM. London: Churchill Livingstone; 1997. p. 167.
- Rothwell PM. External validity of randomised controlled trials: “to whom do the results of this trial apply?”. *Lancet*. 2005; 365:82-93.
- Dekkers OM, von Elm E, Algra A, Romijn JA, Vandembroucke JP. How to assess the external validity of therapeutic trials: a conceptual approach. *Int J Epidemiol*. 2010;39:89-94.
- European Carotid Surgery Trialists' Collaborative Group (1998). Randomised trial of endarterectomy for recently symptomatic carotid stenosis: Final results of the MRC European Carotid Surgery Trial (ECST). *Lancet*. 1998;351:1379-87.
- Ricci S, Celani MG, Righetti E, Cantisani AT, for the International Stroke Trial Collaborative Group. Between country variations in the use of medical treatments for acute stroke: An update. *Cerebrovasc Dis*. 1996; 6 Suppl 2:133.
- Asymptomatic Carotid Atherosclerosis Study Group. Carotid endarterectomy for patients with asymptomatic internal carotid artery stenosis. *JAMA*. 1995;273:1421-8.
- Wennberg DE, Lucas FL, Birkmeyer JD, Bredenberg CE, Fisher ES. Variation in carotid endarterectomy mortality in the Medicare population: trial hospitals, volume, and patient characteristics. *JAMA*. 1998;279:1278-81.
- Gross CP, Mallory R, Heiat A, Krumholz HM. Reporting the recruitment process in clinical trials: Who are these patients and how did they get there? *Ann Intern Med*. 2002;137: 10-6.
- Charlson ME, Horwitz RI. Applying results of randomised trials to clinical practice: impact of losses before randomisation. *BMJ (Clin Res Ed)*. 1984;289:1281-4.
- Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials: The CONSORT statement. *JAMA*. 1996;276:637-9.
- Altman DG, Schulz K, Moher D, et al. The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. *Ann Intern Med*. 2001;134:663-94.
- Wright JR, Bouma S, Dayes I, et al. The importance of reporting patient recruitment details in phase III trials. *J Clin Oncol*. 2006;24:843-5.
- Davis KL, Thal LJ, Gamzu ER, et al, and the Tacrine Collaborative Study Group. A double-blind, placebo-controlled multicenter study of tacrine for Alzheimer's disease. *N Engl J Med*. 1992; 327:1253-9.
- Furlan A, Higashida R, Wechsler L, et al. Intra-arterial prourokinase for acute ischemic stroke. The PROACT II study: A randomized trial. *JAMA*. 1999;282:2003-11.
- Maasland L, van Oostenbrugge RJ, Franke CF, Scholte Op Reimer WJ, Koudstaal PJ, Dippel DW. Netherlands Stroke Survey Investigators. Patients enrolled in large randomized clinical trials of antiplatelet treatment for prevention after transient ischemic attack or ischemic stroke are not representative of patients in clinical practice: the Netherlands Stroke Survey. *Stroke*. 2009;40:2662-8.
- <http://www.fmhs.auckland.ac.nz/soph/depts/epi/epiq/ebp.aspx> (último acceso, agosto/2010).
- Gartlehner G, Hansen RA, Nissman D, Lohr KN, Carey TS. A simple and valid tool distinguished efficacy from effectiveness studies. *J Clin Epidemiol*. 2006;59:1040-8.
- UK-TIA study group. The United Kingdom transient ischaemic attack (UK-TIA) aspirin trial: final results. *J Neurol Neurosurg Psychiatry*. 1991;54:1044-54.
- Dutch TIA Trial Study Group. A comparison of two doses of aspirin (30 mg vs. 283 mg a day) in patients after a transient ischemic attack or minor ischemic stroke. *N Engl J Med*. 1991; 325:1261-6.
- European Carotid Surgery Trialists' Collaborative Group. Randomised trial of endarterectomy for recently symptomatic

- carotid stenosis: final results of the MRC European Carotid Surgery Trial (ECST). *Lancet*. 1998;351:1379-87.
21. Hankey GJ, Slattery JM, Warlow CP. The prognosis of hospital-referred transient ischaemic attacks. *J Neurol Neurosurg Psychiatry*. 1991;54:793-802.
 22. Sacco RL, Diener HC, Yusuf S, Cotton D, Ounpuu S, Lawton WA, et al. PROFESS Study Group. Aspirin and extended-release dipyridamole versus clopidogrel for recurrent stroke. *N Engl J Med*. 2008;359:1238-51.
 23. Diener HC, Bogousslavsky J, Brass LM, Cimminiello C, Csiba L, Kaste M, et al. MATCH Investigators. Aspirin 214 Stroke January 2006 and clopidogrel compared with clopidogrel alone after recent ischaemic stroke or transient ischaemic attack in high-risk patients (MATCH): randomised, double-blind, placebo-controlled trial. *Lancet*. 2004;364:331-7.
 24. Olsson SB, Executive Steering Committee on behalf of the SPORTIF III Investigators. Stroke prevention with the oral direct thrombin inhibitor ximelagatran compared with warfarin in patients with non-valvular atrial fibrillation (SPORTIF III): randomised controlled trial. *Lancet*. 2003;362:1691-8.
 25. Halperin JL, Executive Steering Committee, SPORTIF III, and V Study Investigators. Ximelagatran compared with warfarin for prevention of thromboembolism in patients with nonvalvular atrial fibrillation: Rationale, objectives, and design of a pair of clinical studies and baseline patient characteristics (SPORTIF III and V). *Am Heart J*. 2003;146:431-8.
 26. Osterberg L, Blaschke T. Adherence to medication. *N Engl J Med*. 2005;353:487-97.
 27. Arlt S, Lindner R, Rösler A, von Renteln-Kruse W. Adherence to medication in patients with dementia. *Drugs Aging*. 2008;25:1033-47.
 28. Claxton AJ, Cramer J, Pierce C. A systematic review of the associations between dose regimens and medication compliance. *Clin Ther*. 2001;23:1296-310.
 29. Amuah JE, Hogan DB, Eliasziw M, Supina A, Beck P, Downey W, et al. Persistence with cholinesterase inhibitor therapy in a population-based cohort of patients with Alzheimer's disease. *Pharmacoepidemiol Drug Saf*. 2010;19:670-9.
 30. Tremlett H, Van der Mei I, Pittas F, Blizzard L, Paley G, Dwyer T, et al. Adherence to the immunomodulatory drugs for multiple sclerosis: contrasting factors affect stopping drug and missing doses. *Pharmacoepidemiol Drug Saf*. 2008;17:565-76.
 31. Food and Drug Administration. U.S. Department of Health and Human Services. Guidance for industry. Patient-reported outcome measures: use in medical product development to support labeling claims. Disponible en: <http://www.fda.gov/ohrms/dockets/dockets/06d0044/06d-0044-EC13-Attach-1.pdf> (último acceso, agosto 2010).
 32. Rothwell PM, McDowell Z, Wong CK, Dorman PJ. Doctors and patients don't agree: cross sectional study of patients' and doctors' perceptions and assessments of disability in multiple sclerosis. *BMJ*. 1997;314(7094):1580-3.
 33. Hobart JC, Cano SJ, Zajicek JP, Thompson AJ. Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. *Lancet Neurol*. 2007;6:1094-105.
 34. Marras C, Lang AE. Outcome measures for clinical trials in Parkinson's disease: achievements and shortcomings. *Expert Rev Neurother*. 2004;4:985-93.
 35. Marson AG, Williamson PR. Interpreting regulatory trials in epilepsy. *Curr Opin Neurol*. 2009;22:167-73.
 36. Rothwell PM. Treating individuals 2. Subgroup analysis in randomized controlled trials: importance, indications, and interpretation. *Lancet*. 2005;365(9454):176-86.
 37. Rothwell PM, Mehta Z, Howard SC, Gutnikov SA, Warlow CP. Treating individuals 3: from subgroups to individuals: general principles and the example of carotid endarterectomy. *Lancet*. 2005;365(9455):256-65.
 38. Bussière M, Wiebe S. Progress in clinical neurosciences: measuring the benefit of therapies for neurological disorders. *Can J Neurol Sci*. 2005;32:419-24.
 39. De Robles P, Recchia L, Gonorazky S, González Aguilar P. El costo derivado del número necesario a tratar. Una aproximación accesible y preliminar a la dimensión económica de los tratamientos. Ejemplos en Neurología. *Revista Neurológica Argentina*. 2006;31:59-64.
 40. <http://www.hpc.org.ar/pages/vdato.asp?id=472> (último acceso, agosto 2010).
 41. <http://www.cnsuwo.ca/ebn/cats/> (último acceso, agosto 2010).
 42. Guyatt G, Sackett D, Taylor DW, et al. Determining optimal therapy-randomized trials in individual patients. *N Engl J Med*. 1986;314:889-92.
 43. Guyatt GH, Sackett D, Adachi J, et al. A clinician's guide for conducting randomized trials in individual patients. *Can Med Ass J*. 1988;139:497-503.
 44. Guyatt GH, Keller JL, Jaeschke R, et al. Clinical usefulness of N of 1 randomized control trials: three year experience. *Ann Int Med*. 1990;112:293-2.
 45. Rauek S, Ioli P, Gonorazky SE. Ensayo clínico aleatorizado en un paciente (ECA-n=1) con déficit atencional e hiperactividad (DAH) para determinar la efectividad del tratamiento con metilfenidato. *Revista Neurológica Argentina*. 2007;74 Supl poster 254.
 46. Garattini S, Bertele V. Non-inferiority trials are unethical because they disregard patients' interests. *Lancet*. 2007;370:1875-7.