

## La interpretació de les metaanàlisis: a propòsit de l'eficàcia dels antidepressius

El terme metaanàlisi va ser descrit en el camp de la recerca educativa<sup>1</sup> i s'utilitza en medicina des de mitjan dècada de 1980. Tot i que hi ha diversos sinònims de la mateixa metodologia (per exemple, overview, systematic review, pooled data, etc.) ha acabat imposant-se el terme meta-analysis des que així va ser inclòs com a Medical Subject Heading a la base de dades bibliogràfica Medline.

Una metaanàlisi és útil perquè dona una resposta quantitativa per interpretar l'enorme volum de recerca clínica que es fa actualment en ciències de la salut. Ha esdevingut una metodologia d'elecció per avaluar, classificar i combinar l'evidència científica disponible, la qual cosa s'ha posat de manifest en les revisions quantitatives realitzades per la Col·laboració Cochrane.

Val a dir que la metaanàlisi no és només l'aplicació de mètodes estadístics formals que permeten combinar dades de diferents estudis clínics amb qüestions de recerca similars, sinó que sobretot implica el disseny i l'aplicació d'un protocol rigorós amb un procés de cerca bibliogràfica, selecció i avaluació dels estudis que s'hi volen incloure<sup>2</sup>. Aquesta avaluació necessita, com a mínim, i de manera independent per part de dos investigadors, l'extracció de les dades i la valoració de la qualitat metodològica de cada estudi, així com l'anàlisi de l'heterogeneïtat entre estudis.

Tot això és el que van portar a terme els autors de la metaanàlisi sobre antidepressius que es va publicar el mes de febrer d'enguany<sup>3</sup> i que va tenir un cert ressò als mitjans de comunicació. La metaanàlisi va incloure tots els assaigs clínics que es van lliurar a la Food and Drug Administration (FDA) dels EUA per a l'aprovació de la comercialització de quatre antidepressius de nova generació: fluoxetina, venlafaxina, nefazodona i paroxetina. Els autors volien conèixer el grau de millora respecte el nivell de depressió inicial (mesurat segons l'escala Hamilton de depressió) dels pacients aleatoritzats a placebo o a antidepressiu. De tots els assaigs clínics identificats, finalment se'n van incloure 35 (5 de fluoxetina, 6 de venlafaxina, 8 de nefazodona i 16 de paroxetina) que implicaven un total de 5.133 pacients, 3.292 aleatoritzats a un antidepressiu i 1.841 a placebo.

Els resultats globals van mostrar molt poca diferència entre els que prenién antidepressiu i els que prenién placebo: la mitjana de millora, abans i després del tractament, va ser de

9,60 punts en l'escala de Hamilton per al grup d'antidepressiu i de 7,80 punts en el grup placebo. Per tant, considerant tots els pacients inclosos en els 35 estudis, la diferència entre els que van prendre antidepressiu (qualsevol dels quatre) i placebo va ser d'1,80 punts. Aquesta diferència és estadísticament significativa; tanmateix, els autors de l'estudi puntualitzen que no és clínicament significativa. Per afirmar-ho es van basar en el criteri que el National Institute for Health and Clinical Excellence (NICE) del Regne Unit fa servir, és a dir, 3 punts.

Aquest criteri de significació clínica és el que consta a la guia de pràctica clínica sobre depressió comissionada pel NICE<sup>4</sup>. En aquest punt cal aclarir que el NICE no desenvolupa guies; el que fa és proveir el marc i finançar-les. En aquest cas va ser desenvolupada pel National Collaborating Centre for Mental Health del Regne Unit, que és en realitat una xarxa d'organitzacions professionals implicades en la salut mental, organitzacions d'atenció sanitària, institucions acadèmiques i el mateix NICE. El lideratge d'aquest centre corre a càrrec, de manera conjunta, del Royal College of Psychiatrists' research unit (College Research Unit - CRU) i el Centre for Outcomes Research and Effectiveness (CORE) de la British Psychological Society.

Probablement és per aquest pes professional en el desenvolupament de la guia de pràctica clínica que els autors de la metaanàlisi consideren important aquest criteri de significació clínica. Amb tot, val a dir que a la guia no s'explicita d'on prové aquesta diferència de 3 punts en l'escala Hamilton de depressió per considerar la significació clínica. Hom esperaria que es basés en un estudi on diversos clínics valoressin un canvi en un nombre suficient de pacients (mostra prèviament calculada) i ho compararessin amb els punts obtinguts per aquests pacients en l'escala de depressió. O potser és que aquest valor està molt consensuat entre els clínics i personalment se m'escapa tal consens.

Davant d'això, el criteri de significació estadística sembla més fiable, però també és cert que amb mostres molt grans, com passa en el cas de les metaanàlisis, la significació estadística és relativament senzilla d'atènyer. Finalment, tot i que la metaanàlisi s'ha dut a terme de manera curiosa, la interpretació dels resultats, és a dir, si realment els antidepressius tenen o no efecte en la millora de la depressió, és dubtosa. Tanmateix, anem a pams i desgranem com s'han tractat els potencials biaixos d'una metaanàlisi (vegeu taula 1) per veure si els resultats són o no fiables.

Taula 1. Principals limitacions d'una metaanàlisi

Del conjunt de limitacions que aquest tipus d'anàlisi té, en destaquen:

1. El biaix de selecció, a causa tant d'una cerca bibliogràfica inapropiada o, dins d'aquest, el biaix de publicació, pel qual només es poden seleccionar els estudis que han estat publicats, i que solen ser els que demostren l'eficàcia de la intervenció a estudiar, també anomenats estudis positius.
2. El biaix en els criteris d'inclusió introduït pels investigadors, que pot conduir a l'exclusió d'estudis rellevants o, a la inversa, la inclusió d'estudis no adequats, per exemple, repetits (a vegades es publiquen 2 articles diferents amb anàlisis sobre la mateixa mostra de pacients).
3. Els biaixos existents en els estudis individuals seleccionats, que es poden agreujar quan estan combinats de forma quantitativa.
4. La generalització dels resultats de la metaanàlisi, que estan restringits al conjunt d'estudis combinats i poden no ser generalitzables al total d'estudis que tracten el tema de recerca seleccionat.
5. L'heterogeneïtat dels estudis seleccionats així com la variabilitat en llur qualitat metodològica.

Pel que fa al biaix de selecció o publicació, aquesta metaanàlisi sobre antidepressius es cura en salut tenint en compte que els estudis que inclou no són els publicats sinó tots els que han arribat a l'FDA, s'hagin o no publicat amb posterioritat. A més, els autors van dur a terme una cerca específica a les pàgines web de les companyies farmacèutiques desenvolupadores dels fàrmacs inclosos en la metaanàlisi. D'alguna manera, es podria dir que aquest possible biaix s'ha minimitzat al màxim.

En relació amb els criteris de selecció dels estudis, els investigadors poden haver introduït un biaix en no haver inclòs a la metaanàlisi els estudis de sertralina i citalopram. Tanmateix, aquesta no va ser una decisió deliberada sinó basada en criteris de rigor metodològic, ja que els estudis disponibles amb aquests fàrmacs van ser exclosos perquè no especificaven les puntuacions mitjanes de millora i en anar a cercar informació complementària als webs tampoc no la van trobar. A més, per comprovar si aquests estudis podien fer variar els resultats, els autors van portar a terme una anàlisi de sensibilitat observant què passava amb el resultat global si aquests estudis eren dins o fora de la metaanàlisi. Es comprovava que el resultat global no variava i, per tant, segons això, sembla que la possibilitat de biaix, en el sentit de deixar fora de l'anàlisi un estudi rellevant, és mínima.

Amb tot, el biaix en l'altre sentit, és a dir, com a resultat d'incloure estudis no adequats, pot haver-se donat. No tant perquè hi pugui haver dades repetides (cosa que hauria controlat l'FDA, se suposa) sinó perquè s'hagin inclòs estudis l'objectiu dels quals no fos valorar l'eficàcia sinó altres aspectes previs, com ara el càlcul de la dosi adequada o la toxicitat. En aquests tipus d'estudis són freqüents els resultats negatius atès que les dosis són temptatives i gairebé sempre s'han de fer dos o tres d'aquests assaigs<sup>5</sup>.

Curiosament una altra metaanàlisi d'antidepressius i també amb els estudis de l'FDA, publicada el mateix mes però en

una revista més clàssica<sup>6</sup> va posar de relleu que els estudis de l'FDA eren de fase 2 i fase 3. Els estudis de fase 2 es porten a terme només per obtenir dades preliminars d'eficàcia i dades sobre efectes adversos<sup>7</sup>. Des del meu punt de vista, aquests estudis no s'haurien d'incloure en les metaanàlisis d'eficàcia, atès que els estudis que serveixen realment per valorar-la i que són vàlids per extrapolar els resultats a la clínica són els de fase 3 (vegeu-ne la definició de l'FDA<sup>8</sup>).

Pel que fa a la validesa dels estudis individuals seleccionats, cal dir que les metaanàlisis són habitualment síntesis de dades quantitatives procedents de la combinació d'assaigs clínics, tot i que també es poden combinar dades d'estudis observacionals longitudinals o transversals (per exemple, en el cas de metaanàlisis d'estudis de proves diagnòstiques). Per tant, la validesa d'una metaanàlisi sempre depèn dels estudis que combina. En tractar-se d'assaigs clínics, s'estan combinant els dissenys de major validesa interna possible, però no només cal tenir present el rigor metodològic sinó també altres característiques inherents als objectius dels estudis originals. Per exemple, en aquest cas el seguiment de tots els estudis és de 6 setmanes i, tot i que pot ser acceptable per l'FDA per aprovar la comercialització d'un antidepressiu, el cert és que clínicament seria desitjable que els seguiments fossin més llargs atès que aquests fàrmacs tarden setmanes a produir l'efecte. És cert que se sol parlar d'una a tres setmanes, però també s'ha observat clínicament que poden passar cinc setmanes.

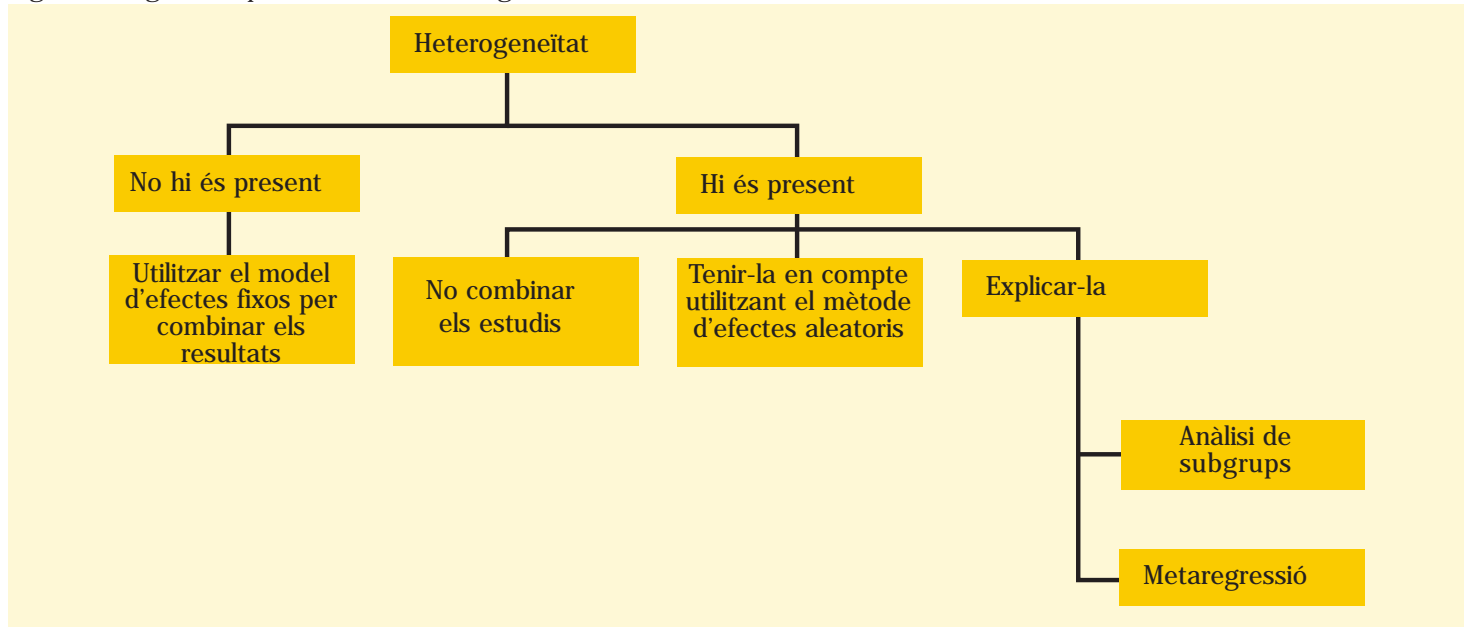
També cal considerar la dificultat de generalització que tenen els assaigs clínics i, per extensió, les metaanàlisis d'assaigs clínics.

Precisament perquè es tracta del disseny de major validesa interna, els criteris d'inclusió tenen tendència a ser al màxim restrictius possible i, per exemple, no s'hi inclouen pacients amb comorbiditat ni tampoc pacients molt greus quan, de fet, la realitat és força la contrària (els pacients amb depressió solen també tenir ansietat i una bona part d'ells, sobre un 15%, se suïciden). Per tant, és difícil extrapolar els resultats obtinguts en aquesta població de pacients relativament lleus i sense comorbiditats (que, d'altra banda, són resultats molt vàlids per a aquesta població) a la població de pacients habituals de depressió.

Si bé és cert que una metaanàlisi pot augmentar la capacitat de generalització perquè inclou diferents estudis amb mostres de pacients amb característiques clíniques que poden diferir; aquestes mateixes diferències poden comportar heterogeneïtat respecte als efectes del tractament (cinquena limitació esmentada més amunt), aspecte que podria limitar la possibilitat de fer la síntesi quantitativa dels estudis. Existeixen proves estadístiques per comprovar si l'heterogeneïtat hi és o no present; en aquesta metaanàlisi, els investigadors utilitzen tant el test Q com l'índex I<sup>2</sup> i es comprova que realment n'hi ha. Amb tot, hi ha diferents maneres de treballar-la (vegeu figura 1). En aquest aspecte, l'article és pràcticament impecable perquè porta a terme totes les possibilitats, fins i tot combina els resultats mitjançant el model d'efectes fixos comparant-lo amb el model d'efectes aleatoris per veure com es comporta.

En relació amb la variabilitat de la qualitat metodològica, els autors ho van tenir en compte i després d'avaluar-la van concloure que la qualitat estava ben controlada i que les característiques metodològiques no afectaven els resultats finals.

Figura 1. Algorisme per a tractar l'heterogeneïtat dels estudis a incloure en una metanàlisi



## Corol·lari

L'article de Kirsch i col·laboradors presenta una metaanàlisi portada a terme de manera rigorosa, però el possible biaix de la inclusió d'estudis no adequats (estudis de dosi-resposta i no d'eficàcia) no es pot descartar i, per tant, pot fer qüestionar els resultats. A més, la significació clínica considerada (3 punts en l'escala Hamilton de depressió) necessita més explicacions. I, finalment, com en qualsevol metaanàlisi d'assaigs clínics, però especialment en el cas de la depressió, cal tenir en compte la dificultat de generalització dels resultats obtinguts. Es tracta de mostres poc semblants a les de la pràctica clínica habitual (amb comorbiditat i amb una proporció relativament alta de casos greus). Així doncs, una recomanació final seria que cal dedicar més neurones i recursos a incloure en els assaigs clínics pacients amb comorbiditat, diferents graus de gravetat o edats extremes, per exemple, i a pensar en aspectes metodològics que ens permetin una millor translació de la recerca a la pràctica clínica. Els recursos probablement estan a l'abast de pocs lectors d'aquest butlletí, però no és pas gens ni mica el cas de les neurones.

Marta Aymerich  
 Metgessa. Directora de l'Agència d'Avaluació de Tecnologia i Recerca Mèdiques (fins al maig del 2008).  
 Actualment és directora executiva del projecte de Facultat de Medicina a la Universitat de Girona.

## Bibliografia

1. Glass CV. Primary, secondary and meta-analysis of research. *Educ Res.* 1976;5:3-8.
2. Jovell AJ, Aymerich M, Serra-Prat M, Navarro-Rubio MD. Meta-anàlisi de ensayos controlados y aleatorizados. En: Jovell AJ, Aymerich M. Evidencia científica y toma de decisiones en sanidad. Monografies Mèdiques de l'Acadèmia de Ciències Mèdiques de Catalunya i de Balears. Barcelona, 1999.
3. Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, et al. Initial severity and antidepressant benefits: A metaanalysis of data submitted to the Food and Drug Administration. *PLoS Med.* 2008;5(2):e45.
4. National Institute for Clinical Excellence (NICE). Depression: management of depression in primary and secondary care. Clinical practice guideline No 23. London: National Institute for Clinical Excellence, 2004.
5. Comunicació personal amb Enric Álvarez, Cap de Servei de Psiquiatria de l'Hospital de Sant Pau i Professor titular de Psiquiatria de la Universitat Autònoma de Barcelona.
6. Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Eng J Med.* 2008;358:252-60
7. Food and Drug Administration (FDA). The CDER Handbook. [Consultat el 2 de juny de 2008]. Disponible a: <http://www.fda.gov/CDER/HANDBOOK/phase2.htm>
8. Food and Drug Administration (FDA). The CDER Handbook. [Consultat el 2 de juny de 2008]. Disponible a: <http://www.fda.gov/CDER/HANDBOOK/phase3.htm>