

## Capítol III

### Punts bàsics de l'avaluació: Metodologia

#### Tema A

Quan podem dir que els resultats d'un estudi són vàlids i aplicables?

#### Tema B

Quina decisió cal prendre si el nou medicament és més eficaç o segur que el de referència, però també és més car?

#### Tema C

Quan podem dir que el nou medicament és un equivalent terapèutic?

#### Tema D

Com hem de definir les condicions d'ús i el lloc en terapèutica?



## Capítol III

### TEMA A

### QUAN PODEM DIR QUE ELS RESULTATS D'UN ESTUDI SÓN VÀLIDS I APLICABLES?

1- La qualitat dels assaigs clínics publicats: les evidències sota sospita

2- Aprendre a llegir entre línies

2.1 Validesa interna

A- Aspectes bàsics

B- Aspectes secundaris

2.2. Eficàcia i validesa externa o aplicabilitat: cap a l'efectivitat

2.3 Rellevància dels resultats

A- Concepte i càlcul: RAR, NNT, RRR, RR, OR

B- Precisió i IC

3. Decidir en la pràctica

4. Annex

- Llista CONSORT



## 1- LA QUALITAT DELS ASSAIGS CLÍNICS PUBLICATS: LES EVIDÈNCIES SOTA SOSPITA

La validesa interna i l'aplicabilitat dels resultats dels estudis és el punt crucial per a la selecció del fàrmac.

La lectura superficial de molts assaigs clínics pivot ens aporta evidències substancials dels avantatges d'un nou fàrmac. Però una anàlisi en profunditat ens posa de manifest aspectes crítics que matisen i, de vegades, invaliden aquests avantatges. El fet que l'autorització per comercialitzar un nou fàrmac es basi en un únic o en pocs assaigs clínics fase III, fa que l'anàlisi detallada de dits assaigs sigui especialment decisiva per a la presa de decisions.

La presència i la detecció de biaixos en els assaigs clínics publicats és un tema cada vegada més present en la literatura mèdica.

Pel que fa a la necessitat d'aquesta lectura crítica i rigorosa en els últims mesos tenim que la comunitat científica, per motius diferents, ha fet revisions profundes d'alguns assaigs ( Vegeu exemples recents a la taula 1):

- Estudi CLASS, celecoxib en osteoartritis i artritis reumatoide
- Estudi PROWESS, Drotrecogina alfa activada en sepsis
- Estudi WHI, teràpia hormonal substitutiva en la dona postmenopàusica
- Rasburicasa en hiperuricèmia postquimioteràpia

La comunitat científica està molt sensibilitzada per aquest tema i, com a exemple d'alguns dels esforços i de les propostes encaminades a millorar la qualitat dels assaigs clínics i de la publicació hi ha **la llista-guia CONSORT**, acceptada pels editors de les revistes mèdiques de major impacte (Annals Intern Med, JAMA, NEJM, Lancet...). La llista ha estat actualitzada recentment i constitueix una eina excel·lent com a guia per a la revisió crítica d'un article. (*Altman DG et al: The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med 2001; 1134:663-694*). Es pot accedir al protocol d'avaluació original, a la informació ampliada i a exemples a internet, accés lliure en <http://www.consort-statement.org/revisedstatement.htm>. Vegeu també punt l'annex d'aquest tema).

**Finançament, autoria i responsabilitat.** També és un motiu de preocupació el tema de l'autoria de les publicacions i de la relació dels autors amb les companyies promotores de l'assaig. Ha generat una normativa i uns requisits que s'han de tenir en compte per a la presentació de manuscrits per a la publicació a les revistes biomèdiques: *Frank Davidoff, Catherine D DeAngelis, Jeffrey M. Drazen, et al Finançament, autoria i responsabilitat. Versió en espanyol de la Rev Esp Cardiol. Accés lliure en Rev Esp Cardiol 2001; 54: 1247 - 1250. <http://www.secardiologia.es/>.*

## **Taula 1** **Exemples recents d'assaigs clínics i "evidències" que s'han posat en dubte**

### **Publicació de les dades parcials**

Com a exemple de les dades publicades de forma parcial i que poden modificar l'estimació d'una relació risc/benefici, hi ha l'estudi CLASS (Silverstein Fe et al JAMA 2000; 284: 1246-5512) que comparava el Celecoxib, el Diclofenac i l'Ibuprofè a la indicació d'osteoartritis i artritis reumatoide. Concloïa que el Celecoxib s'associava a una menor incidència d'úlcers simptomàtiques i complicacions ulceroses que els AINE clàssics. En una editorial recent s'analitzen les dades de l'estudi complet i s'explica que l'estudi original tenia defectes significatius de disseny: va combinar la troballa de dos estudis diferents i va incloure els resultats de només els sis primers mesos de l'assaig. Quan es van analitzar les dades de tot el període de 12 mesos, el Celecoxib no va produir menys complicacions ulceroses que els dos AINE clàssics. (Jüni P et al BMJ 2002; 324:1287-8)3.

### **Modificació del protocol de l'assaig clínic**

Un altre exemple és el de l'estudi PROWESS sobre la Proteïna C activada en la sepsis. L'estudi va sofrir una modificació en els criteris d'inclusió durant l'assaig. Dels 720 pacients que ja s'havien inclòs es va decidir excloure'n els que presentaven riscos majors de mort, no lligats a sepsis, i es va procedir a canviar la línia cel·lular productora de la Proteïna C activada, així com d'incloure nous criteris d'anàlisi. La diferència en mortalitat no era estadísticament significativa en el subgrup de pacients que responien als criteris inicials (el 28% versus el 30%) mentre que sí que era significativa en els pacients inclosos després del canvi de protocol (el 22% versus el 31%) (Siegel JP NEJM Sounding Board; 2002; 347: 1030-4). Un comitè d'experts (Anti infective Drugs advisory Committee) de la FDA, dels quals la meitat dels membres van votar a favor i l'altra meitat en contra de la seva aprovació, va procedir a reavaluar l'assaig clínic en funció de l'escala de valoració APACHE II i es va obtenir informació addicional. L'anàlisi "post hoc" ha evidenciat subgrups de malalts que podrien obtenir un benefici de la Drotrecogina alfa i d'altres que estarien exposats a riscos de sagnia greu.

### **Noves evidències que canvien les coses**

El tercer exemple és el de la Teràpia Hormonal Substitutiva (THS) amb nous assaigs clínics i que demostren uns riscos que no es coneixien. Recentment s'han publicat els resultats de l'assaig clínic WHI ( Writing Group for the Women's Health Initiative Investigators JAMA 2002; 288: 321-333), realitzat a 16.608 dones amb l'objectiu de determinar els riscos i beneficis a llarg termini de la THS en la dona postmenopàusica. Els resultats mostren un major risc de càncer de mama i de tipus cardiovascular, que els que es coneixien fins al moment. El risc que s'havia estimat prèviament a partir d'estudis d'observació mostrava avantatges de la THS en la disminució de risc cardiovascular. Els resultats d'aquest assaig clínic han modificat la relació benefici/risc de la THS en prevenció primària (Fletcher WS et al JAMA 2002; 298, 3: 366-), i han generat les advertències oportunes per part de les agències reguladores per limitar els usos i advertir dels riscos cardiovasculars a curt termini i de neoplàsia de mama a llarg termini.

### **Assaigs petits que "justifiquen" un registre**

La Rasburicasa va ser autoritzada per l'EMEA a partir de només un assaig de fase III. (Informe EMEA 2001). Aquest estudi (Goldman SC Blood 2001, 97:2998-3003) és un assaig no cec amb una grandària de mostra molt petita (52 pacients pediàtrics) que compara l'Alopurinol i la Rasburicasa. El resultat s'expressa en una variable intermèdia (àrea sota la corba d'àcid úric durant 96 hores). L'estudi té una grandària de la mostra massa petita per trobar diferències d'incidència de fallada renal o de requisits de suport renal. El fet és que amb aquestes evidències el producte va rebre l'autorització per ser comercialitzat

## 2- APRENDRE A LLEGIR ENTRE LÍNIES

La validesa i l'aplicabilitat dels resultats d'un assaig s'ha de revisar i s'ha de fer tenint en compte els conceptes i els aspectes metodològic següents:

- La **validesa interna** es planteja si els resultats obtinguts proporcionen una resposta correcta a la pregunta plantejada en l'objectiu de l'estudi, amb absència de biaixos en els procés de disseny, d'execució i d'anàlisi.

- La **validesa externa** d'un estudi ens informa sobre l'aplicabilitat de les conclusions a altres pacients, diferents dels que hi van participar. Per determinar la validesa externa d'un estudi s'han de valorar aspectes com la definició de la malaltia, els criteris de selecció dels subjectes, la intervenció que s'utilitza com a referència, la pauta i la dosi de les intervencions que es comparen, etc. Es tracta de veure si aquests aspectes són els adequats al problema de salut d'interès.

### 2.1 VALIDESA INTERNA D'UN ASSAIG CLÍNIC

Podem dir que un assaig clínic és vàlid si permet assegurar que l'efecte que s'ha observat és a causa de la intervenció diferent que el grup d'intervenció i el grup control han rebut i no per altres factors que puguin haver conduït a una falsa conclusió.

#### A) ASPECTES BÀSICS

Perquè un assaig clínic tingui un bon nivell de validesa s'han de considerar tres aspectes bàsics:

- 1- Assignació aleatòria
- 2- Seguiment exhaustiu de tots els subjectes
- 3- Anàlisi per intenció de tractar.

- **Assignació aleatòria.** La característica metodològica clau d'un assaig clínic és l'assignació aleatòria dels subjectes als grups d'estudi: el grup control i el grup intervenció. En intervenir només l'atzar a l'assignació, hi ha la tendència a assegurar que es produeix una distribució equilibrada de totes les variables, tant de les conegudes com de les desconegudes. En els estudis amb una grandària de la mostra reduïda, es poden emprar les tècniques d'assignació estratificada o per blocs que permeten augmentar la probabilitat de formar grups equilibrats.

- **Seguiment exhaustiu de tots els subjectes.** Es tracta de respondre a dues qüestions: s'han tingut en compte tots els pacients en els resultats?, s'ha realitzat un seguiment complet dels pacients? Com que els grups formats per assignació aleatòria tendeixen a ser comparables, s'ha de fer un seguiment de tots els pacients que els componen i s'han de tenir en compte en finalitzar l'estudi. Si no es fa així, els subjectes que finalitzen l'estudi poden formar grups que no siguin similars, de manera que la comparació de les respostes observades no sigui imparcial.

De tota manera, a la pràctica sempre existeix un nombre de pacients als quals no es pot fer el seguiment. Per poder avaluar aquest aspecte, és important que a la publicació de l'assaig clínic es proporcioni **informació sobre els abandonaments (dropouts)** o pacients que interrompen el seguiment, i sobre les **retirades (withdrawals)** o subjectes que han d'abandonar per l'aparició de reaccions adverses o altres esdeveniments que obliguen la seva exclusió. Si les seves característiques són diferents de la dels subjectes que finalitzen l'estudi, o si s'han produït per motius relacionats amb les intervencions de l'estudi, la seva exclusió de l'anàlisi podria invalidar les conclusions.

Actualment les normes de publicació a les revistes científiques segueixen les recomanacions del grup CONSORT, i solen presentar un diagrama de flux on consten les pèrdues de pacients de cada fase de l'assaig: 1-Reclutament, 2-Assignació, 3-Seguiment, 4-Anàlisi.

- **Anàlisi per intenció de tractar.** Els investigadors poden estar temptats d'excloure determinats subjectes de l'anàlisi perquè no han rebut el tractament previst, perquè han tingut un mal compliment o perquè no han seguit la pauta de visites o altres aspectes del protocol de l'estudi. Tanmateix, com que els motius pels quals s'han produït aquests fets poden estar relacionats amb el pronòstic dels subjectes, la seva exclusió podria introduir un biaix de comparació. Per evitar-ho s'ha d'utilitzar l'estratègia de **l'anàlisi per intenció de tractar**, segons la qual els subjectes s'inclouen al grup al qual han estat assignats independentment de si han rebut o no correctament la intervenció, el seguiment i l'avaluació.

Conceptes:

- **Anàlisi per intenció de tractar "Intention-To-Treat analyses" (ITT).** Segons la definició més acceptada és l'anàlisi que inclou "tots els pacients als grups als quals s'han assignat aleatòriament, independentment de si complien o no els criteris d'inclusió, independentment del tractament que realment rebien i independentment que abandonin el tractament o es desviïn del protocol". Alguns assaigs utilitzen un concepte una mica més restrictiu, per exemple: "Anàlisi que es realitza sobre els pacients reclutats i randomitzats que han rebut, almenys, una dosi del medicament correcte segons lo adjudicat a l'aleatorització".

- **Anàlisi per intenció de tractar modificada "Modified Intention-To-Treat analyses" (MITT):** Molts assaigs defineixen en la metodologia que es realitza una MITT, per exemple, a l'estudi (*Herbrecht R et al. N Eng J Med 2002; 347:408-15*) s'assigna tractament antifúngic (Voriconazole o Anfotericina B) quan hi ha sospita d'infecció per fongs. En la MITT s'analitzen els pacients que compleixen els criteris de l'anàlisi "per intenció de tractar" i que, a més, presenten el diagnòstic confirmat, definitiu o probable, d'infecció fúngica realitzat. Un grup independent realitza la MITT. Aquest tipus d'anàlisi ha d'estar previst i descrit en el mètode. Les anàlisis "post hoc" tenen un valor molt limitat.



- **Anàlisi per protocol (PP).** Aquestes anàlisis estudien només els pacients dels dos grups que han seguit totalment el protocol, per exemple, només es tenen en compte els pacients sense violacions del protocol, que compleixen tots els criteris d'inclusió i d'exclusió, que han rebut el tractament corresponent, que hi ha una adequada avaluació al final del tractament amb el compliment del tractament correcte). En ocasions es presenten anàlisis de tipus PP adaptats com, per exemple, els anomenats "anàlisi de pacients amb criteris d'avaluació" definits a la secció de mètode de l'assaig.

En general els resultats de l'anàlisi ITT han de ser més similars a aquells que podem esperar quan s'aplica el fàrmac en clínica que amb els resultats de les anàlisis MITT i PP. L'ITT reflecteix el que ocorre realment a la pràctica clínica, permet obtenir de l'assaig clínic la informació més pròxima a l'efectivitat del tractament (efecte més proper a les condicions reals), en què es donen circumstàncies com la de diagnòstic incorrecte, falta d'adhesió al tractament, etc. Això és important en els assaigs clínics amb una orientació pragmàtica, i és important des del nostre punt de vista, ja que ens interessa estimar l'efectivitat del fàrmac quan l'emparam en els nostres pacients.

En molts d'assaigs es presenten taules amb ambdós tipus de resultats (ITT i PP), ja que poden aportar informació complementària per facilitar la interpretació, per exemple: en els assaigs d'antiretrovirals és molt rellevant el grau de compliment terapèutic. A les anàlisis ITT s'estudien tots els pacients d'ambdós grups, amb independència de l'adherència. A l'anàlisi PP s'estudien només els que han seguit de forma correcta el tractament. L'anàlisi ITT ens dóna una estimació global de l'eficàcia, l'anàlisi PP ens dóna una estimació de l'eficàcia dels bons complidors. Si el metge prescriu el tractament a un pacient concret que coneix que és bon adherent, el que s'espera és obtenir un resultat d'efectivitat similar a l'anàlisi PP.

Les anàlisis que avaluen la seguretat d'un fàrmac s'han de fer sobre la ITT, ja que el motiu que un pacient abandoni el tractament pot estar relacionat amb l'aparició d'una reacció adversa.

En els assaigs clínics d'equivalència, l'anàlisi per ITT deixa de ser conservadora, perquè permet concloure amb més facilitat que els resultats no són estadísticament significatius i, per tant, inferir que les alternatives en estudi són terapèuticament equivalents.

## **Taula 2**

### **Exemple: Definició i nombre de pacients avaluats a cada tipus d'anàlisi**

*Voriconazole en aspergil·losi invasora versus Anfotericina B deoxicolat. Herbrecht R, et al N Engl J Med 2002;347:408-15*

**Anàlisi per intenció de tractar (ITT):** són els reclutats aleatòriament N=391.

**Anàlisi de seguretat:** són els reclutats i randomitzats, amb almenys una dosi del medicament correcte segons la randomització. (S'exclouen 12 pacients perquè no van rebre cap tractament). N= 379.

**Anàlisi per intenció de tractar modificada (MITT):** ITT + Diagnòstic confirmat definitiu probable per un grup independent. S'exclouen 102 pacients per manca de diagnòstic confirmat d'aspergil·losi invasora. N=277.

**Anàlisi CCE (Anàlisi a pacients amb criteris d'avaluació, similar a anàlisi PP):** els pacients inclosos a l'anàlisi MITT + no teràpia antifúngica concomitant + no violacions protocol + adequada avaluació al final del tractament + haver rebut almenys 5 dies del tractament de l'estudi. N=242.

**Taula 3.**

**Exemples: Resultats expressats per anàlisi ITT i per anàlisi PP**

**ESTUDI M 98-863 Antirretrovirals en pacients naïves.**

Estudi fase III, randomitzat doble cec compara Lopinavir/Ritonavir amb el Nelfinavir (gold estàndard) en combinació amb d4T/3TC, en pacients naïves. S'analiza l'eficàcia en termes de resposta virològica (proporció de pacients amb <400 còpies la setmana 34) i durada fins a la setmana 48. Núm. de pacients: 653.

	Lopinavir/ Ritonavir	Nelfinavir	p	NNT
PCR-HIV <400 per intenció de tractar (ITT)	75%	63%	P<0,001	8,3
PCR-HIV <400 tractats (PP)	93%	82%	P<0,001	9,1

A les 48 setmanes el 75% dels pacients amb Lopinavir/Ritonavir versus el 63% amb Nelfinavir tenien menys de 400 còpies. P<0,001. (anàlisi per intenció de tractar). Els resultats si es considera els pacients en tractament són del 93% i del 82% (anàlisi per protocol) respectivament p <0,001.

**B) ASPECTES SECUNDARIS**

A més dels tres punts esmentats, existeixen una sèrie de característiques secundàries que orienten sobre la qualitat de l'assaig clínic:

**Emmascarament.** Les expectatives d'èxit o de fracàs de la intervenció, tant dels pacients com dels investigadors, poden influir en l'avaluació de la resposta. Aquest problema es pot prevenir amb l'ús de tècniques d'emmascarament o cec, que impliquen que els pacients i/o els observadors desconeguin el tractament que rep cada un, de manera que no puguin interferir en els resultats. La tècnica més bona és la de **doble cec** o **cegament doble**, segons la qual ni els pacients ni els investigadors coneixen el tractament assignat. No sempre és possible aplicar aquestes tècniques. Com a mínim és recomanable que les persones que valoren la resposta desconeguin el tractament assignat, de manera que es redueixi el possible biaix.

**Comparabilitat inicial dels grups.** Els grups han de ser similars en tot, excepte en el tractament que reben. S'ha d'haver avaluat la comparabilitat inicial dels grups pels factors pronòstics ja que l'assignació aleatòria no assegura que s'hagi produït una distribució equilibrada de totes les variables rellevants, especialment quan el nombre de subjectes no és elevat. Si els grups presenten algunes diferències importants per algun factor fortament associat amb el pronòstic, encara que no sigui estadísticament significativa, la validesa de l'estudi es pot veure compromesa.

Si els grups no són similars per alguna variable, l'anàlisi estadística permet controlar la influència de l'esmentada diferència i obtenir una estimació

ajustada de l'efecte del tractament mitjançant l'anàlisi estratificada o una tècnica multivariant.

**La comparabilitat dels grups s'ha de mantenir durant tot el seguiment.** Si un dels grups s'ha seguit amb més atenció que l'altre, existeix una major probabilitat de detectar qualsevol esdeveniment que ocorri durant el seguiment i, fins i tot, de modificar la intervenció o d'administrar altres tractaments, que interfereixen amb l'avaluació de la resposta.

**Variables intermèdies o vicariants.** Molts assaigs clínics quantifiquen els resultats d'eficàcia mitjançant una variable intermèdia. Les variables intermèdies han de reflectir el resultat final de forma convincent, sinó la fortalesa de les conclusions es debilita. La variable intermèdia ha de presentar una associació consistent, forta i independent amb el desenllaç clínic. Vegeu taula.

<b>Taula 4</b> <b>Característiques d'una variable intermèdia ideal.</b>	
<ul style="list-style-type: none"> <li>- Fiable, reproducible, clínicament disponible, fàcil de quantificar i capaç de mostrar un efecte "dosi/resposta".</li> <li>- Vertadera predictora de la malaltia (o del risc d'emmalaltir).</li> <li>- La seva relació amb la malaltia ha de ser biològicament plausible.</li> <li>- Sensible.</li> <li>- Específica.</li> <li>- Punt de tall precís entre els valors normals i els anormals.</li> <li>- Valor predictiu positiu acceptable.</li> <li>- Valor predictiu negatiu acceptable.</li> <li>- Susceptible de control de qualitat.</li> <li>- Els canvis de la variable intermèdia han de reflectir ràpidament i amb precisió la resposta al tractament, així com normalitzar els seus valors en les fases de remissió o de guariment.</li> </ul>	

<b>Taula 5</b> <b>Resum. Passos bàsics que s'han de seguir per valorar la validesa interna d'un assaig clínic</b> <i>(Adaptat de Jimenez 1997, Guyatt 1995)</i>	
1. Validesa interna de l'estudi (metodologia correcta)	
Preguntes principals:	<ul style="list-style-type: none"> <li>L'assignació dels subjectes ha estat aleatòria?</li> <li>S'han seguit tots els subjectes?</li> <li>S'ha utilitzat l'estratègia d'anàlisi per intenció de tractar?</li> </ul>
Preguntes secundàries	<ul style="list-style-type: none"> <li>S'ha utilitzat alguna tècnica d'emascarament?</li> <li>Els grups eren comparables a l'inici de l'estudi?</li> <li>Els grups s'han seguit i tractat de la mateixa forma?, excepte el tractament d'estudi?</li> <li>S'empra una variable intermèdia o una variable final? En el primer cas està validada?</li> </ul>

## 2.2 EFICÀCIA I VALIDESA EXTERNA O APLICABILITAT: CAP A L'EFECTIVITAT

Els assaigs clínics donen resultats d'eficàcia; però el que ens interessa, quan aplicam les conclusions als nostres pacients, és conèixer quina serà la seva efectivitat real, tenint en compte tant les característiques de base del grup de candidats a tractar, com els aspectes d'atenció assistencial vigents en el nostre medi.

Els estudis amb criteris **restrictius** tenen més limitacions per a l'extrapolació. Els estudis **pragmàtics** es poden generalitzar més fàcilment. Una major validesa externa s'associa al caràcter **multicèntric** de l'estudi ja que suposa la participació de diferents institucions sanitàries, de pacients de diverses procedències i de diferents estils de pràctica clínica.

S'ha de valorar si la definició de la malaltia, els criteris de selecció dels subjectes, la intervenció que s'utilitza com a referència, la pauta i la dosi de les intervencions que es comparen, etc. són adequades al problema de salut que ens interessa. Si els pacients tenen les mateixes característiques que els pacients inclosos a l'estudi, els resultats han de ser perfectament aplicables. Per això els estudis han d'indicar clarament els criteris de selecció, la procedència i les característiques dels subjectes estudiats.

És convenient tenir en compte les següents qüestions sobre la utilitat pràctica dels resultats de l'assaig. Respondre a unes simples preguntes ens ajuden a centrar el tema:

El fàrmac amb què es compara és l'estàndard adequat i el que s'empra a la pràctica habitual del meu hospital? Sí, No

En el cas que el fàrmac de comparació no sigui l'estàndard del meu hospital, l'assaig pivot aporta evidències indirectes d'utilitat? Sí, No.

Les característiques basals dels meus pacients són similars a les de l'assaig? Sí, No. Vegeu criteris d'inclusió i d'exclusió.

Els resultats es poden aplicar a la meua pràctica? Sí, No.

El seguiment clínic i assistencial dels meus pacients és similar a les de l'assaig? Sí, No.

Als assaigs en els quals s'utilitza una variable intermèdia, la variable és consistent amb el resultat final?

S'han tengut en compte tots els desenllaços clínicament importants?

## Les evidències de superioritat

Per considerar un fàrmac superior a un altre i justificar la seva inclusió a l'hospital, hem de tenir en compte dos punts bàsics més:

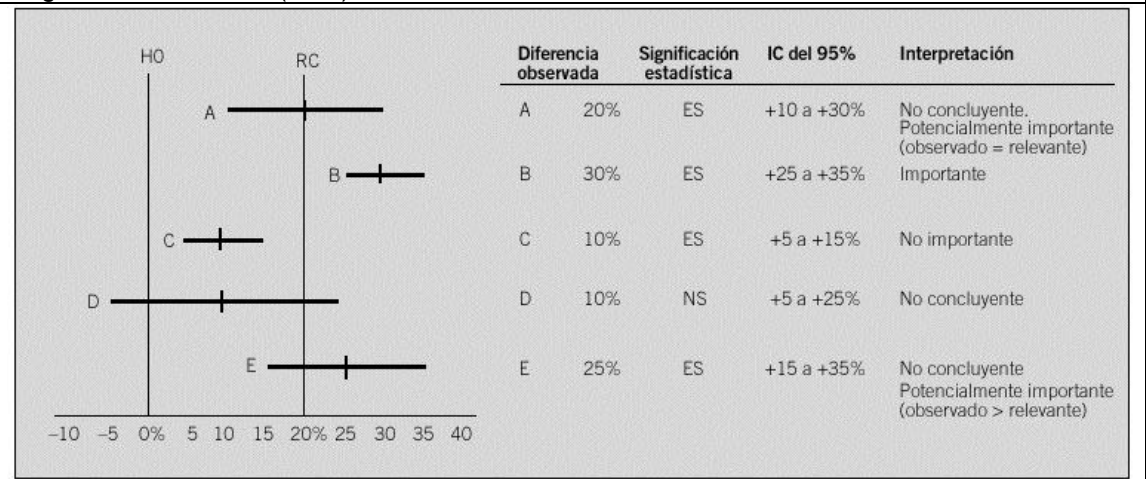
- Valorar si la **magnitud de l'efecte del tractament** és de rellevància clínica. Per això hem d'establir quin és el valor mínim que consideram clínicament rellevant. Si el valor que s'ha trobat a l'assaig és superior i el seu IC95% de confiança no inclou l'esmentat valor mínim, podem dir que el tractament en estudi és superior, (exemple B de la gràfica). Si el valor és inferior, estadísticament significatiu i el seu IC95% no inclou el valor mínim clínicament rellevant, podem considerar que el medicament és un poc superior però sense importància clínica (cas C de la gràfica). Els casos A i E indiquen superioritat i d'una possible importància clínica, encara que no són concloents s'han de tenir en compte com a indicatius de superioritat.

### Gràfica 1

**Resultats hipotètics d'un estudi que compara un tractament actiu davant un placebo i utilitza com a mesures la diferència d'eficàcia (RAR). Es considera que la mínima RAR rellevant clínicament és del 20%.**

ES: Diferència significativa estadísticament. NS: No significativa estadísticament.

Argimon JM Med Clin (Barc) 2002, 118:382-4.



- Les dades dels subgrups i l'anàlisi detallada de l'assaig són de gran interès per determinar al final les **condicions d'ús** del nou fàrmac. Això ajuda a delimitar els subgrups de pacients que es poden beneficiar més dels efectes del medicament i establir les condicions d'ús a l'hospital. De fet, la major part dels nous medicaments s'introdueixen per la seva superioritat en petits grups de pacients i condicions d'ús estrictes. Quan un estudi presenta anàlisi de subgrups s'ha de ser molt cautelós i avaluar si el subgrup és gran, si respon a una hipòtesi establerta prèviament a l'inici de l'estudi, si s'han analitzat pocs subgrups i si s'han

observat els mateixos resultats en altres estudis. Si no es compleixen aquestes circumstàncies caldrà ser escèptic.

### Les evidències d'equivalència terapèutica

La decisió d'incloure un nou medicament com a equivalent terapèutic es realitza segons les evidències. Diferents tipus d'estudis ens aporten informació per a la presa de decisions:

- Estudis pragmàtics, sobretot de tipus observacional.
- Assaigs directes entre dues opcions amb l'objectiu de demostrar l'equivalència o la no-inferioritat.
- Estudis directes entre dues opcions amb l'objectiu de demostrar la superioritat però no apareixen diferències estadístiques significatives entre les opcions.
- Evidències indirectes per l'existència d'assaigs amb un tercer element comparatiu.

**Vegeu tema C del capítol III** informació ampliada

<b>Taula 6</b>
<b>Resum. Aspectes bàsics de la validesa externa o de l'aplicabilitat</b>
Rellevància dels resultats
Es presenten diferències significatives?
Es presenten els IC dels resultats principals i dels subgrups?
Quina és la magnitud de l'efecte del tractament i és de rellevància clínica?
Si s'utilitza una variable intermèdia, és consistent amb el resultat final?
Utilitat pràctica dels resultats
El fàrmac amb què es compara és l'estàndard adequat i el que s'empra a la pràctica habitual del meu hospital?
Les característiques basals dels meus pacients són similars a les de l'assaig? Vegeu criteris d'inclusió i d'exclusió.
Els resultats es poden aplicar a la meva pràctica?
El seguiment clínic i assistencial dels meus pacients és similar a les de l'assaig?
S'han tengut en compte tots els desenllaços clínicament importants?

## 2.3 RELLEVÀNCIA DELS RESULTATS

La decisió de si un tractament és o no més eficaç que el de referència no s'ha de basar exclusivament en el resultat d'una prova de significació estadística. Si el lector només coneix el valor de p, sabrà si l'estudi ha demostrat o no que existeix un efecte però no si és important en clínica.

És fonamental que l'article proporcioni una estimació de la **magnitud** de l'efecte del tractament perquè el lector pugui avaluar si és un efecte d'importància clínica, i una estimació de la **precisió** amb els seus intervals de confiança.

### A- FORMA DE PRESENTAR LA MAGNITUD DE L'EFECTE

A la majoria dels estudis pivot s'empren variables binàries del tipus "sí" o "no", per a l'obtenció dels resultats principals (exemples: % de mort de pacients, % curació de pacients, % de pacients amb determinada resposta clínica). A la publicació original és habitual transformar i presentar moltes variables contínues de forma binària.

La forma d'expressar els resultats pot ser diferent i és necessari familiaritzar-se amb els diferents índexs i les seves relacions. Moltes vegades els resultats s'expressen en mesures relatives, sobretot la Reducció del Risc Relatiu (RRR), ja que, d'aquesta forma, s'obtenen valors més espectaculars i aparents. També, amb freqüència, s'usen el Risc Relatiu (RR) i l'Odds Ratio (OR) que s'han incorporat com a elements d'expressió dels resultats no només als estudis observacionals, sinó també en els assaigs clínics i en les metanàlisis.

Les mesures absolutes es basen en calcular la diferència entre les freqüències del resultat observat en cada grup de subjectes, mentre que les mesures relatives es basen en el quocient entre les esmentades freqüències, quantificant quantes vegades és més probable que ocorri un esdeveniment en un grup que en un altre.

Les formes habituals de presentació dels resultats són:

Mesures absolutes:

- Reducció Absoluta del Risc (RAR)
- Número Necessari a Tractar (NNT)

Mesures relatives:

- Reducció Relativa del Risc (RRR)
- Risc relatiu del grup tractament respecte al grup control (RR)
- Odds Ratio (OR)

Vegeu els exemples de concepte i de càlcul a la taula

<b>Taula 7</b>		
<b>EXEMPLE : Forma de presentar les dades d'eficàcia d'un assaig clínic</b>		
<b>Exemple sobre els resultats de mortalitat d'un assaig clínic en el qual el Grup A correspon al grup control i el Grup B al grup intervenció</b>		
<i>Fórmula</i>	<i>Terminologia</i>	<i>Càlcul i resultat</i>
A	Risc (incidència, proporció) grup control o risc	Exemple 20/100 =20%

	<i>base</i>	=0,2
<i>B</i>	<i>Risc (incidència, proporció) grup intervenció :</i>	<i>Exemple 15/100 =15% =0,15</i>
<i>A – B</i>	<i>RAR (Reducció Absoluta del Risc o Diferències de Risc o Risc atribuïble )</i>	<i>0,2-0,15 = 0,05</i>
<i>1 / (A-B)</i>	<i>NNT (Número Necessari a Tractar). És a dir que és igual a 1/RAR</i>	<i>1/(0,2-0,15) = 20</i>
<i>B / A</i>	<i>RR (Risc Relatiu o Raó de Riscs)</i>	<i>0,15 / 0,20 = 0,75</i>
<i>B´/A´</i>	<i>OR (Odds Ratio)</i>	<i>15/85 : 20/80 = 0,176/0,25 = 0,706</i>
<i>[ (A-B) / A ] x 100</i>	<i>RRR (Reducció Relativa del Risc). És a dir que és igual a 1-RR</i>	<i>[ (0,2-0,15) / 0,2 ] x 100 = 25% = 0,25</i>
<i>B´ = casos/no casos grup intervenció: 15/85 = 0,176</i>		
<i>A´ = casos/no casos grup control: 20/80 = 0,25</i>		

Per al nostre propòsit convé usar les mesures absolutes.

### Mesures absolutes:

#### - Concepte i càlcul de la RAR

La RAR (Reducció Absoluta del Risc o Diferències de Risc o Risc atribuïble ) és la forma més simple d'expressar la diferència d'eficàcia entre els grups estudiats. És la diferència de proporcions d'un esdeveniment entre el grup control i el grup intervenció.

Les dades bàsiques de proporció o incidència de l'esdeveniment de cada grup estudiat i la seva diferència o RAR ens informen d'un forma simple de la magnitud de la diferència d'eficàcia.

La RAR s'ha de conèixer abans de prendre cap decisió sobre l'avantatge d'un fàrmac i quins pacients s'han de tractar.

#### - Concepte i càlcul de l'NNT

L'NNT és el número necessari de pacients a tractar per produir una unitat addicional d'eficàcia. L'NNT és una dada fàcil d'obtenir dels resultats dels assaigs clínics, sempre que la mesura d'eficàcia s'expressi en forma d'una variable binària.

L'NNT ens és útil per dimensionar l'eficàcia del fàrmac mitjançant l'expressió del grau d'“esforç” que s'ha de realitzar per aconseguir una unitat addicional d'eficàcia. També ens serà molt útil per als càlculs farmacoeconòmics

L'NNT=1/RAR. Si expressam la RAR en percentatges, llavors l'NNT =100/RAR

#### **Taula 8**

#### **Càlcul del Numero necessari de pacients a tractar (NNT)**

*L'NNT és l'invers de la reducció del risc absolut (1/RAR) i es pot calcular a partir dels resultats d'eficàcia d'un assaig clínic amb l'aplicació d'una fórmula senzilla*



<b>Número Necessari a Tractar</b>	<i>Resultats d'un assaig clínic:</i> <i>a: percentatge de pacients del grup tractat amb el nou fàrmac amb resultats d'eficàcia (braç experimental).</i> <i>b: percentatge de pacients del grup tractat amb placebo o amb el fàrmac de referència amb resultats d'eficàcia (braç control).</i> <i>a-b: reducció del risc absolut (RAR) expressat en percentatge.</i>
100 $NNT = \frac{100}{a-b}$	

### Mesures relatives

L'expressió dels resultats exclusivament com mesures relatives, fet que ocorre sovint en difondre les conclusions d'un assaig clínic, pot resultar tendenciosa i fer que se sobrevalori l'efecte beneficiós real d'un tractament. El problema és que les mesures relatives no tenen en compte el risc basal o de partida dels pacients i, per això, no permeten diferenciar els beneficis grans dels petits. Vegeu taules 9 i 10.

#### - Concepte d'RRR. Alerta amb la forma de presentar les dades d'eficàcia:

L'RRR és la diferència en les proporcions d'esdeveniments entre el grup control i el de tractament, dividit per la proporció d'esdeveniments en el grup control.

$$[(A-B) / A] \times 100$$

**Taula 9**  
**Exemple de RAR i d'RRR**

*Si un fàrmac disminueix la mortalitat d'un 6% a un 3%, podem dir realment que :*

- *Ha disminuït la mortalitat un 3% en termes absoluts (6% -> 3%) (RAR).*
- *Ha disminuït la mortalitat un 50% en termes relatius (6% -> 3%) (RRR).*
- *Ha augmentat la supervivència del 94% al 97%, és a dir un 3,2% en termes relatius (RRR).*

*En tots els casos deim la veritat.*  
*Per això, és millor parlar en termes de RAR (Reducció absoluta del risc), que tots entenem de forma més clara i precisa.*

**Taula 10**  
**Mesures de l'efecte d'un tractament en funció del risc basal dels subjectes del grup control, amb la suposició dels riscos basals del 40%, del 20%, del 10% i del 5 %.**

	<b>Reg basal del grup control</b>			
	40 %	20 %	10 %	5 %
<b>Incidència en el grup tractat</b>	30 %	15 %	7,5 %	3,75 %
<b>RRR (Reducció Relativa del Risc)</b>	25 %	25 %	25 %	25 %
<b>RAR (Reducció Absoluta del Risc)</b>	10 %	5 %	2,5 %	1,25 %
<b>NNT (Número Necessari a Tractar)</b>	10	20	40	80

### - Concepte i càlcul de l'RR

L'RR (Risc Relatiu) és una relació entre la proporció d'esdeveniments del grup control i del grup de referència.

És una mesura relativa dels resultats relacionada directament amb l'RRR (Reducció del Risc Relatiu), de fet  $RRR=(1-RR)\times 100$ .

Per exemple, la proporció de pacients que sofreixen una reacció adversa amb el tractament estàndard és del 0,10 (10%), mentre que la proporció de pacients que presenten la mateixa reacció adversa amb el tractament en estudi és del 0,125 (12,5%). L'RR és igual a 1,25 (és a dir el quocient de 0,125 / 0,1). També podem dir que amb el tractament estàndard un 25% menys dels pacients sofreixen la reacció adversa [ $RRR=(1-1,25) \times 100 = 25\%$ ].

Un altre exemple: La mortalitat al braç placebo és del 15%. Aquest percentatge és del 5% al braç intervenció. L'RR és del 0,33 ( quocient de 0,5/0,15), és a dir, hi ha un 67% més de mortalitat al grup placebo [ $RRR=(1-0,33) \times 100 = 67\%$ ].

L'RR pren valors entre 0 i infinit. El valor 1 és el valor neutre i significa que no hi ha diferències entre els grups comparats. Quan ens apropam al valor 0 o a un valor molt superior a 1, significa que la diferència entre grups és molt gran.

Si pren un valor  $>1$  significa que el grup el resultat del qual posam en el numerador té una major proporció que el que posam en el denominador, al contrari si el RR és  $<1$ .

(Nota: En general a la fórmula de càlcul del RR, el risc base ( "control event rate") se situa al denominador i el del grup intervenció al numerador. Si intercanviam el lloc on posam les dues proporcions l'RR serà l'invers (1/RR)).

### - Concepte i càlcul de l'OR ( Odds Ratio o Momio Ratio)

La mesura de l'efecte més emprada en la metanàlisi, tant per expressar les dades de cada assaig com el resultat global comparat, és l'OR (Exemple Biblioteca Cochrane). Estadísticament es prefereix l'OR a l'RR a causa, fonamentalment, de les seves propietats matemàtiques per al càlcul del resultat global. Permet, mitjançant la regressió logística, ajustar la relació en estudi per l'efecte d'altres variables. Actualment en molts assaigs clínics els resultats s'expressen en OR.

L'Odds. És la raó en la qual el numerador és la probabilitat que ocorri un esdeveniment i el denominador és la probabilitat que l'esdeveniment no ocorri. En les apostes de joc de pilota dels frontons s'anomena "momio".

Habitualment s'estima mitjançant el nombre de pacients que compleixen els criteris d'un objectiu dividit pel nombre de pacients que no compleixen l'objectiu.

*L'OR (Odds Ratio)*. La divisió d'una odds per una altra odds és una raó d'odds.

La fórmula de càlcul és:

$$OR = p_1 / (1-p_1) : p_2 / (1-p_2).$$

Exemple, si després d'un tractament antibiòtic en un grup de deu pacients, quatre presenten diarrea ( $4/6=0,66$ ) i al grup control de 10 pacients 1 presenta diarrea ( $1/10=0,11$ ), l'OR serà de  $=0,66/0,11 = 6$ .

En una taula 2x2 l'OR també es pot calcular segons la fórmula:

$$OR = AD/BC$$

	Èxit	Fracàs
Grup 1 (intervenció)	A	B
Grup 2 (control)	C	D

### - Relació entre l'RR i l'OR

En els esdeveniments que es produeixin amb baixa incidència (<10%), el valor numèric de l'OR i el de l'RR seran semblants. Quan l'esdeveniment és més freqüent del 10% però la diferència entre grup control i grup intervenció és petita també resulten similars les estimacions de l'RR i de l'OR.

#### **Taula 11**

#### **Exemples: Expressió de resultats com a RR i OR**

*Exemple 1:*

*Risc en grup control 0,2. Risc en grup intervenció 0,05*

$$RR = 0,05/0,2 = 0,25$$

$$OR = 0,05/0,95 : 0,2/0,8 = 0,052 : 0,250 = 0,21$$

*Exemple 2:*

*Risc en grup control 0,9. Risc en grup intervenció 0,4*

$$RR = 0,4/0,9 = 0,44$$

$$OR = 0,4/0,6 : 0,9/0,1 = 0,67 : 9 = 0,07$$

Els RR és un quocient o relació de *proporcions* i l'OR és un quocient o relació de dos *odds*. Tant les proporcions com les odds expressen el mateix, però usen dues escales numèriques diferents, les proporcions oscil·len entre 0 i 1 i les odds entre 0 i infinit. Per passar d'una escala a una altra s'utilitzen les expressions :  $odds = p/(1-p)$ ; i  $p = odds / (1+odds)$ .

En general els resultats expressats en OR seran molt més aparents, és a dir, amb valors més extrems (més a prop de 0 o més a prop d'infinít) que els expressats en RR, especialment quan la incidència d'esdeveniment d'algun dels dos grups és superior al 10% i les diferències entre ells són grans. Aquesta circumstància és freqüent en els assaigs clínics. Els IC dels OR també solen ser més amplis que els dels RR. Als estudis epidemiològics se solen manejar taxes d'incidència molt baixes, i per això els valors d'OR i d'RR són similars.

#### - Relació entre mesures absolutes i relatives

Un RR o un OR poden ser difícils de traduir si no es tradueix en mesures de la diferència absoluta de riscos. La RAR i l'NNT es poden calcular a partir de l'RR si es coneix la incidència de l'esdeveniment al grup control. També existeixen fórmules per calcular l'NNT derivades dels resultats d'un assaig clínic a partir de l'OR.

A internet hi ha diferents programes d'accés lliure que faciliten el seu càlcul ràpid. Vegeu taula 12.

**Taula 12.**  
**Càlcul d'NNT, RR, OR i RRA amb intervals de confiança 95%, programes en internet**

**"Risk Reduction Calculator"** <http://araw.mede.uic.edu/cgi-bin/nntcalc.pl>  
(gratuït)

**"EBM calculator"** <http://www.cebm.utoronto.ca/palm/ebmcalc/ebm.zip>  
(gratuït). És un programa per a agendes PDA tipus PALM

**"UBC Clinical Significance Calculator"**  
<http://www.healthcare.ubc.ca/calc/clinsig.html>  
<http://www.healthcare.ubc.ca/calc/clinsig.html>

**"Calculador cisa"** <http://home.clara.net/sisa/two2hlp.htm>

**Altres programes estadístics d'accés lliure en internet:**

[Http://www.fisterra.com/material/investiga/program\\_internet/program\\_internet.htm](http://www.fisterra.com/material/investiga/program_internet/program_internet.htm)

*Nota: Per calcular l'IC, s'ha de conèixer: - el n de cada grup i els resultats d'esdeveniments (+/-) de cada grup per fer la taula de 2 x 2. Si es disposa del n i del %, es poden obtenir amb una simple calculadora.*

També hi ha taules

<b>Taula 13.</b>					
<b>Càlcul de l'NNT a partir d'una OR per a un assaig clínic (Mc Lister JAMA 2000)</b>					
Incidència de l'esdeveniment al grup control	Intervenció terapèutica (OR)				
	0,50	0,60	0,70	0,80	0,90
	NNT				
0,05	41	52	69	104	209
0,1	21	27	36	54	110
0,2	11	14	20	30	61
0,3	8	10	14	22	46
0,5	5	8	11	18	38
0,7	6	9	13	20	44
0,9	12	18	27	46	101

Hi ha evidència empírica que suggereix que els índexs relatius OR i RR tendeixen a ser més semblants entre diferents assaigs del mateix tractament que els índexs absoluts RAR i NNT.

Els índexs absoluts últims depenen del tractament i també del risc basal. Així en determinats assaigs pot passar que el subgrup de pacients amb un risc base de major gravetat es beneficiïn més del tractament (exemple: Drotrecogina alfa en sepsis), mentre que en altres casos ocorre a l'inrevés, el subgrup de pacients amb una millor situació de base es "curaran" més (exemple: medicament per al tractament d'una infecció).

Un aspecte important, per decidir si el resultat d'un assaig expressat en RAR o NNT és aplicable al nostre medi, és avaluar si el risc basal dels pacients de l'assaig o d'alguns dels subgrups és similar al del nostre medi.

L'aplicació dels índexs relatius a situacions basals diferents al de l'assaig constitueix una aproximació que, malgrat les evidències empíriques esmentades, s'ha de realitzar amb cautela.

## B) PRECISIÓ AMB QUÈ S'ESTIMA L'EFECTE

Cada expressió de la magnitud de l'efecte s'ha d'acompanyar dels intervals de confiança. El vertader efecte d'un tractament és desconegut. La mesura observada no és més que una estimació puntual. Si s'hagués estudiat una mostra de subjectes diferents, s'hauria obtingut una estimació probablement similar, però no exactament igual. Se sol realitzar el càlcul de l'interval de confiança del 95%. Obviament la precisió expressada per l'amplitud de l'esmentat interval està molt relacionada amb la grandària de la mostra. Si el número de subjectes és major, també és major la precisió i més estret l'interval de confiança.

La precisió s'estima amb l'interval de confiança (IC) del 95%, que és l'interval entre els límits del qual es té el 95% de confiança que es trobi la vertadera magnitud de l'efecte. L'IC presenta grans avantatges respecte al grau de significació estadística.

Quan s'empra com a mesura de l'efecte una diferència com la RAR o l'RRR, si l'IC del 95% inclou el valor 0, el resultat no és estadísticament significatiu. Si, al contrari, exclou el valor 0, la diferència és estadísticament significativa.

Quan l'IC s'aplica a l'estimació d'un RR o d'una OR, no es presenten diferències estadísticament significatives quan l'IC 95% inclou el valor 1.

A més de saber si el resultat és o no estadísticament significatiu, l'IC permet conèixer entre quins límits és probable que es trobi la vertadera magnitud, i veure si inclou o exclou el mínim valor considerat de rellevància clínica.

La majoria de les publicacions incorporen actualment els IC 95%, però no sempre. Per al càlcul dels intervals de confiança de l'NNT i les altres mesures, hi ha programes que ens ajuden. Vegeu la taula anterior amb els programes de càlcul disponibles a internet.

### 3. DECIDIR EN LA PRACTICA

#### Són vàlids i aplicables els resultats?

En resum, en revisar un assaig hem d'incidir en els aspectes bàsics de validesa interna, sobretot si el tipus d'anàlisi és per intenció de tractar o per protocol i, en aquest cas, la proporció de pacients no avaluats.

També és important la variable emprada en els resultats, veure la seva rellevància clínica i, si es tracta d'una variable intermèdia, que sigui realment predictiva de resultats finals de pes clínic.

Convé avaluar els resultats en variables absolutes i veure si la magnitud de la superioritat és rellevant, tant en el resultat principal com en els subgrups. Els IC 95% s'han de conèixer per definir aquesta superioritat.

Els resultats principals i els secundaris han d'estar definits prèviament com a objectius de l'estudi i estar exposats en el mètode.

Quant a la validesa externa, el més important és que el medicament de referència sigui l'estàndard adequat, juntament amb les altres qüestions de l'apartat 2.2.

La qüestió és que moltes vegades la informació i l'evidència aportada per l'assaig és incompleta i que genera incertesa sobre el seu interès pràctic en els nostres pacients.

Per facilitar, de forma sistemàtica i estructurada, l'avaluació de la qualitat d'un assaig, s'han confeccionat llistes-guia o check-list, que poden ser de tipus qualitatiu o amb escales quantitatives. S'han publicat multitud de llistes-guia, més o menys complexes (vegeu a l'annex la llista CONSORT), que serveixen per repassar, de forma sistemàtica, cada un dels punts importants.

Des d'un punt de vista pràctic pot ser preferible emprar alguns conceptes de l'MBE, com les escales de classificació de l'evidència científica que diferencien, de forma jeràrquica, els diferents nivells de l'evidència en funció del rigor científic del disseny de l'estudi. D'aquí es deriven la magnitud i el rigor de les recomanacions que es tradueixen en una definició del grau d'evidència i de recomanació.

Hi ha diferents escales de gradació de la qualitat, bastant similars entre si, encara que difereixen en alguns matisos. En una revisió recent (*Martínez-López I et al Informes normalizados de evaluación de medicamentos: resultados de dos años de aplicación. 2003 Aten Farm (en premsa)*), nosaltres hi hem aplicat una escala d'evidència científica modificada de *Jovell et al (Jovell AJ, Navarrès-Rubio MD: Evaluación de la evidencia científica. Med Clin (Barc) 1995; 105:740-3)*, que ens ha permès classificar els assaigs clínics pivot dels darrers medicaments estudiats. És només una aproximació, que es basa

en establir un nivell i restar categories en funció de les condicions de validesa interna o externa no adequades. D'aquesta forma podem resumir en una frase el nivell d'evidència de cada estudi.

<b>Taula 14</b>				
<b>Nivells d'evidència.</b>				
<i>Modificat de Jovell AJ , Navarrès-Rubio MD: Evaluación de la evidencia científica. Med Clin (Barc) 1995; 105:740-3)</i>				
	<i>Nivell d'evidència</i>	<i>Disseny</i>	<i>Condicions rigor científic</i>	<i>Validesa externa o aplicabilitat</i>
I.	<i>Excel·lent</i>	<i>Metanàlisis d'assaigs controlades i randomitzades</i>	<i>Qualitat dels estudis primaris</i>	<i>Compara amb l'estàndard correcte</i>
II	<i>Adequada</i>	<i>Assaig controlat i randomitzat de mostra gran (n &gt; 500 per a aquesta "chec-list")</i>	<i>Poder estadístic. Qualitat de l'estudi, pèrdues seguiment, no anàlisi "post hoc", variable principal</i>	<i>Compara amb l'estàndard correcte</i>
III	<i>Regular</i>	<i>Assaig controlat i randomitzat de mostra petita (n &lt;500 per a aquesta "chec-list")</i>	<i>Poder estadístic. Qualitat de l'estudi, pèrdues seguiment, no anàlisi "post hoc" Variable principal</i>	<i>Compara amb l'estàndard correcte</i>
IV	<i>Baixa</i>	<i>Assaig prospectiu randomitzat no cec Assaig prospectiu controlat no randomitzat</i>	<i>Controls coincidents en temps. Qualitat de l'estudi, pèrdues seguiment, no anàlisi "post hoc", variable principal</i>	<i>Compara amb l'estàndard correcte</i>
V	<i>Baixa</i>	<i>Assaig retrospectiu no randomitzat</i>		
VI	<i>Baixa</i>	<i>Estudi de cohorts Estudi de sèries observacionals</i>		
VII	<i>Pobre</i>	<i>Casos i controls</i>		
VIII	<i>Pobre</i>	<i>Sèries clíniques, estudis descriptius</i>		
<i>Si l'estàndard amb el qual es compara no és adequat (-2 categoria), estudi no publicat (-1 categoria). No qualitat. (-1 categoria)</i>				

En el nostre cas, durant el període d'estudi de dos anys, s'han realitzat 28 informes d'avaluació. Els nivells d'evidència que els estudis han aportat van ser els següents: Nivell I (excel·lent): 0 casos; Nivell II (adequada): 10 casos (50%); Nivell III (regular): 4 casos (20%); Nivells IV a VI (baixa): 6 casos (30%).

Tanmateix, al marge de classificacions que faciliten col·locar una "etiqueta", el més important és la lectura crítica i detallada de l'assaig, que és l'aspecte clau per decidir sobre la selecció i definir les condicions d'ús d'un nou medicament.



#### 4. Annex tema A del capítol III

##### La llista-guia Consort

La proposta CONSORT, realitzada per un grup d'experts, exposa les recomanacions bàsiques que s'han de seguir per publicar assaigs clínics a revistes biomèdiques. Conté un diagrama de flux, que serveix de model perquè els autors d'un assaig exposin, de forma ordenada, el nombre de pacients inicialment inclosos a l'estudi i les pèrdues ocorregudes en cada fase al costat dels motius. (Ref: *Altman DECAGRAM et al: Ann Intern Med 2001; 134:663-694*).

Al congrés de la SEFH de l'any 2002 de Barcelona, el nostre servei va presentar una comunicació en la qual es revisava la qualitat dels assaigs clínics "pivot" que han tingut impacte en la introducció de nous medicaments a l'hospital en els últims anys. Es van avaluar 8 assaigs clínics dels fàrmacs: Darbepoetina, EPO 40.000 UI, Levonorgestrel DIU, Peg-interferón alfa-2b, Tenecteplasa, Tobramicina TNS, Verteporfino i Zoledronat, i es va valorar el percentatge de compliment dels criteris CONSORT. Els criteris es compleixen en un 68% de mitjana (rang 52-96%-96). ( Ref. *F. Puigventós et al Comunicació al Congrés de la SEFH. Barcelona 2002*)

Resum

##### **Taula 15**

##### **ASSAIGS CLÍNICS AMB IMPACTE EN LA INTRODUCCIÓ DE NOUS FÀRMACS A L'HOSPITAL: APLICACIÓ DE LA LLISTA-GUIA CONSORT**

*F. Puigventós Latorre, P Ventayol Bosch, I Martínez López, M Crespí Monjo, A Escrivá Torralva, F. Comes Gallard, O. Prim Sánchez.*

**Objectiu:** avaluar la qualitat dels principals assaigs clínics que han justificat la inclusió dels últims fàrmacs a l'hospital mitjançant l'aplicació de la llista-guia CONSORT.

**Mètodes:** s'identifiquen els assaigs clínics publicats dels últims fàrmacs inclosos a l'hospital. Dos revisors, de forma independent, van avaluar els articles, varen valorar el grau de compliment dels criteris i els apartats de la llista-guia CONSORT; en cas de discrepància va ser revisat i consensuat per un tercer revisor.

**Resultats:** s'han avaluat 8 assaigs clínics dels fàrmacs: Darbepoetina, EPO 40.000 UI, Levonorgestrel DIU, Peg-interferó alfa-2b, Tenecteplasa, Tobramicina TNS, Verteporfino i Zoledronat, i s'ha valorat el percentatge de compliment dels criteris CONSORT:

Darbepoetina	EPO-40000	Levonorgestrel	Peg-INF	Tobramicina	TNK	Verteporfino	Zoledrònic
Kidney Intern 2001,60:741-7.	Ann Inter Med 2000; 133:845-54.	BMJ 1998; 316:1122-6.	Lancet 2001; 358:958-65.	N Eng J Med 1999; 340:23- 30.	Lancet 1999; 354:716-22.	Arch Ophtalmol 1999; 117: 139- 45.	J Clin Oncol 2001, 19: 6558-67.
52 %	83 %	61 %	70 %	54 %	71 %	96 %	59 %

Els criteris es compleixen en una mitjana del 68% (52-96%-96), hi ha homogeneïtat entre els revisors en un 88% (només el 12% va requerir una tercera revisió). Els criteris que menys es compleixen són: mètodes usats per garantir qualitat de la mesura dels resultats (12%), descripció detallada de com s'aplica la randomització (12%), avaluació del procés i èxit d'encegar (20%), criteris de suspensió de l'assaig (33%), i presentació de resultats amb intervals de confiança (50%).

**Conclusions:** la majoria d'assaigs avaluats no compleixen una elevada proporció dels criteris. La guia CONSORT és fàcil d'aplicar i ajuda a sistematitzar les revisions dels assaigs clínics. Per a la seva aplicació en l'avaluació de nous medicaments és necessari valorar altres aspectes addicionals, com l'adequació del grup control, la rellevància clínica dels resultats i realitzar una ponderació dels criteris.

**Comissió de Farmàcia: selecció de medicaments**  
**XIII Jornada Serveis de Farmàcia d'Hospitals Comarcals de Catalunya**  
**Olot, octubre 2003**

<b>Taula 16.</b>					
<b>Llista-Guia CONSORT 2001</b>					
Protocol d'avaluació original, a la informació ampliada i a exemples a internet, accés lliure en <a href="http://www.consort-statement.org/revisedstatement.htm">http://www.consort-statement.org/revisedstatement.htm</a>					
<b>Secció</b>	<b>Ítem</b>	<b>Descripció.</b> <b>Ha de constar:</b>	<b>Si</b>	<b>No</b>	<b>No apl</b>
<b>TÍTOL I RESUM</b>					
	1	Com van ser assignats els participants a la intervenció? (ex: "adjudicació aleatòria", "randomització", "assignats aleatòriament")			
<b>INTRODUCCIÓ</b>					
Antecedents	2	Antecedents científics i explicació racional			
<b>MÈTODES</b>					
Participants	3a 3b	Criteris d'elecció dels participants i lloc de recollida de les dades			
Intervencions	4	Detalls precisos de la intervenció que s'han realitzat a cada grup i com i quan serà aplicada			
Objectius	5	Objectius específics i hipòtesi			
Resultats	6a 6b	Definició clara de les mesures dels resultats primaris i de les mesures dels resultats secundaris i, quan sigui aplicable, mètodes usats per garantir la qualitat de les mesures (Ex: observacions múltiples, formació dels assessors).			
Grandària de la mostra	7a 7b	Com va ser calculada la grandària de la mostra i, quan siguin aplicables, exposició d'anàlisis intermèdies i criteris de suspensió de l'assaig			
Randomització: generació de la seqüència	8a 8b	Mètode usat per generar la seqüència de randomització, incloent-hi detalls de qualsevol restricció ( ex: blocs, estratificació, ...)			
Randomització: assignació oculta	9	Mètode usat per ocultar l'assignació aleatòria. (Ex: sobres numerats, telèfon central), que especifiqui si la seqüència està oculta fins que la intervenció és assignada.			
Randomització: aplicació	10	Qui genera la seqüència de l'assignació, qui incorpora als participants i qui assigna els participants als seus grups.			
Cec (emascarament)	11a 11b	Si els que administren les intervencions i avaluen els resultats són cecs respecte a l'assignació de grups. Si això és així, com es va avaluar el procés i èxit de cegar.			
Els mètodes estadístics	12a 12b	Mètodes estadístics usats per comparar grups en els resultats primaris. Mètodes per a les anàlisis addicionals, com ara anàlisi del subgrup i anàlisi ajustades.			
<b>RESULTATS</b>					
Flux dels participants	13a 13b	Flux dels participants a través de cada etapa (es recomana un <b>diagrama</b> ). Específicament, per a cada grup estudiat, comunicar el nombre de participants assignats aleatòriament, els que han rebut el tractament previst, els que han acabat el protocol de l'estudi, i els analitzats per al resultat primari. Descripció de les desviacions del protocol de l'estudi previst, juntament amb les raons de les esmentades desviacions.			
El reclutament	14	Dates que defineixen els períodes de reclutament i seguiment.			
Dades base	15	Característiques demogràfiques i clíniques cada grup.			
Nombres analitzats	16	Els nombres de participants (denominador) a cada grup, inclosos a cada anàlisi i especificar si l'anàlisi s'ha realitzat "per intenció de tractar". Indica els resultats en números absoluts quan és factible (ex: 10/20, no 50%).			
Resultats i estimació	17	Per a cada resultat primari i secundari, un resum dels resultats per a cada grup, i de la grandària estimada de l'efecte i la seva precisió (ex: interval de la confiança del 95%).			
Anàlisis auxiliars	18	Comunicar qualsevol altra anàlisi realitzada, incloent-hi anàlisi de subgrup i anàlisis ajustades, indicant aquelles que s'han preespecificat i aquelles que són exploratòries.			
Efectes adversos	19	Tots els esdeveniments adversos importants o efectes secundaris en cada grup de la intervenció.			
<b>DISCUSSIÓ</b>					
Interpretació	20	Interpretació dels resultats, considerant la hipòtesi de l'estudi, fonts de biaixos potencials o d'imprecisió, així com els perills associats			

		amb la multiplicitat d'anàlisi i resultats.			
Generalització	21	Generalització (validesa externa) dels resultats d'assaig.			
Evidència general	22	Interpretació general dels resultats en el context de l'evidència actual.			

## DIAGRAMA CONSORT 2001

Flux de les diferents fases d'un AC

- 1-Reclutament
- 2-Assignació
- 3-Seguiment
- 4-Anàlisi

**1**

**NOMBRE INICIAL DE PACIENTS AVALUATS PER SER INCLOSOS EN L'ASSAIG, (n =...)**



**NOMBRE DE PÈRDUES ANTERIORS A L'ASSIGNACIÓ ALEATÒRIA, (n =...)**

**MOTIUS D'EXCLUSIÓ**  
 - No compleixen els criteris d'inclusió (n =...)  
 - No donen el consentiment per participar-hi, (n =...)  
 - Altres motius (n = ...)



**2**

**ASSIGNACIÓ ALEATÒRIA**  
**NOMBRE DE PACIENTS ASSIGNATS ALEATÒRIAMENT, (n =...)**



**GRUP EXPERIMENTAL (n =...)**  
 - Reben la intervenció assignada (n =...)  
 - No reben la intervenció assignada (n =...)  
 (Explicau motius, n = )

**GRUP DE REFERÈNCIA (n =...)**  
 - Reben la intervenció assignada (n =...)  
 - No reben la intervenció assignada (n =...)  
 (Explicau motius, n = )



**3**

**PÈRDUES DE SEGUIMENT (n =...)**  
 (Explicau motius)  
 -ABANDONAMENT DE LA INTERVENCIÓ (n =...)  
 (Explicau motius)

**-PÈRDUES DE SEGUIMENT (n =...)**  
 (Explicar motius)  
 -ABANDONAMENT DE LA INTERVENCIÓ (n =...)  
 (Explicau motius)



**4**

**PACIENTS QUE S'ANALITZEN (n =...)**  
 Pacients exclosos de l'anàlisi (Explicau motius, (n =...))

**PACIENTS QUE S'ANALITZEN (n =...)**  
 Pacients exclosos de l'anàlisi (Explicau motius) (n =...))

## Referències:

Guyatt G, Rennie D, Haywart RSA. *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. AMA ed 2002 <http://pubs.ama-assn.org/misc/usersguides.dtl>

Argimon JM. La ausencia de significación estadística en un ensayo clínico no significa equivalencia terapéutica. *Med Clin (Barc)* 2002; 118 (18): 701-3

Argimon JM. El intervalo de confianza, algo más que un valor de significación estadística. *Med Clin (Barc)* 2002; 118(10): 382-4

Altimiras J et al: *Farmacoepidemiología y estudios de utilización de medicamentos*. EN J. Bonal et al ed *Farmacia Hospitalaria*, 3 ed 2002

Guyatt G et al. *Therapy and understanding results*. in *Users Guides to the medical literature*. ESTIMA ed 2002

Bucher H et al. *Therapy and applying the results*. In *Users Guides to the medical literature*. ESTIMA ed 2002

Oxman A et al. *Summarizing the evidence*. In *Users Guides to the medical literature*. ESTIMA ed 2002

Abraira V. *Medidas del efecto de un tratamiento: Reducción absoluta del riesgo, reducción relativa del riesgo y riesgo relativo*. *SEMERGEN* 2000, 26: 535-6

Abraira V. *Medidas del efecto de un tratamiento: odds ratio y número necesario para tratar*. *SEMERGEN* 2001, 27:418-20

Abraira V. *Estimación: intervalos de confianza*. *SEMERGEN* 2002; 28(2): 84-5

Abraira V. *¿Qué es el análisis por intención de tratar?* *SEMERGEN* 2000; 26: 393-4

Rodríguez JL et al. *Doble ciego. El control de los sesgos en la realización de los ensayos clínicos, contradicciones, insuficiencias e implicacions*. *Med Clin (Barc)* 2002; 118: 192-5

Martínez-gonzález MA et al. *¿Que es una odds ratio?* *Med Clin (Barc)* 1999; 112:416-22

Jiménez Villa J. *Lectura crítica de la literatura científica (I) validez del estudio*. *FMC Formación Médica Continuada en Atención Primaria*. 2000, 7, 4: 209-220

Jimenez Villa J. *Lectura crítica de la literatura científica II. Evaluación de los resultados*. *FMC Formación Médica Continuada en Atención Primaria*. 2000, 7, 5: 283-296

Jovell AJ, Navarro-Rubio MD. *Evaluación de la evidencia científica*. *Med Clin (Barc)* 1995; 105: 740-43

Granados A. *La evaluación de las tecnologías médicas*. *Med Clin(Barc)* 1995; 104: 581-585.

Ioannidis JPA, Lau J. *Can quality of clinical trials and meta-anayses be quantified ?* *Lancet* 1998; 352: 590-91

Berlin JA, Rennie D. *Measuring the quality of trials. The quality of quality scales*. *JAMA* 1999, 282, 11: 1083-84.

Juni P et al. *The hazards of scoring the quality of clinical trials for meta-analysis*. *JAMA* 1999; 282: 1054-60.

*Barton S. Which clinical studies provide the best evidence. BMJ 2000; 321: 255-56*

*Jiménez Villa J. Lectura crítica de un ensayo clínico. Cuadernos de gestión: 1997, 3, 2: 76-89*

*García Lopez FJ, Gutiérrez S, Galende I, Avendaño C. Evaluación de la calidad de los ensayos clínicos: justificación, utilidad e inconvenientes. Med Clin (Barc ) 1999; 112 (supl 1): 35-42*

*The Asilomar Working Group on Recommendations for Reporting of Clinical Trials in the Biomedical Literature: Checklist of information for inclusion in reports of clinical trials; Ann Intern Med 1996, 124, 8: 741-743*

*Guyatt GH, Sackett DL, Cook DJ: User's guides to the medical literature. II How to use an article about therapy and prevention . A . How valid are the results of the study ? JAMA 1993, 270,21: 2598-2601*

*Guyatt GH, Sackett DL, Cook DJ: User's guides to the medical literature. II How to use an article about therapy and prevention . B. What were the results and will they help in caring for my patients ? JAMA 1994, 271, 1: 59-63.*

*Mora R. Como mejorar la calidad estadística de los artículos presentados a revistas biomédicas: lista de comprobación para los autores. Med Clin (Barc) 1999; 113: 138-149*