

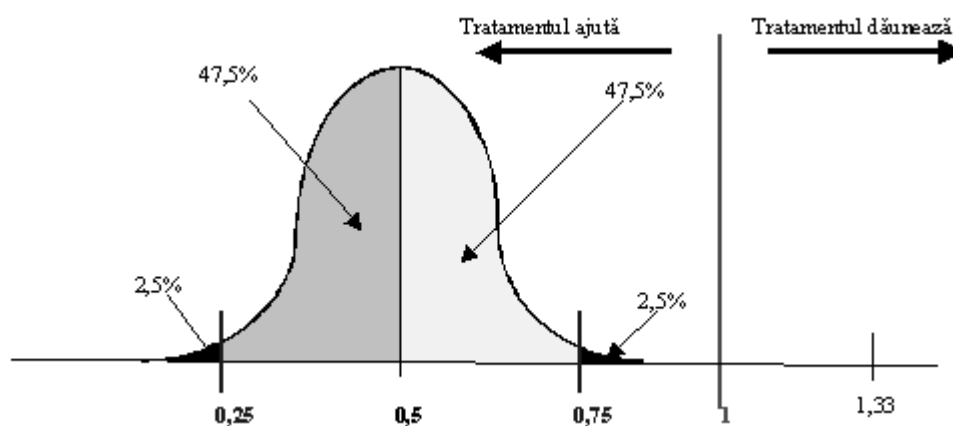
XII. Cât de precisă este estimarea efectului tratamentului?

În timp ce în capitolele IX și X am vorbit despre semnificația clinică a rezultatelor unui RCT (mărimea efectului), de această dată ne vom ocupa de semnificația statistică, și în special de intervalele de încredere.

Nu putem niciodată cunoaște care este reducerea riscului în realitate (la nivelul populației), ci avem la îndemână estimarea punctuală (din studiul respectiv), pe care încercăm statistic să o extrapolăm la nivelul populației - aceasta se face prin intermediul intervalului de încredere, adică intervalul în care putem fi 95% siguri că se găsește parametrul respectiv la nivelul populației.

Așadar, parametrul despre care vorbim (RR, RRR, RAR, OR etc.) se poate găsi în realitate oriunde în acest interval, este adevărat cu o probabilitate din ce în ce mai mică pe măsură ce ne îndreptăm spre capetele intervalului, și cu o probabilitate de 5% să se găsească înafara lui (2,5% sub limita inferioară și tot 2,5% deasupra limitei superioare a intervalului - Figura XII.1).

Figura XII.1. Rezultatul unui studiu clinic randomizat, exprimat sub forma riscului relativ (RR cu CI 95%).

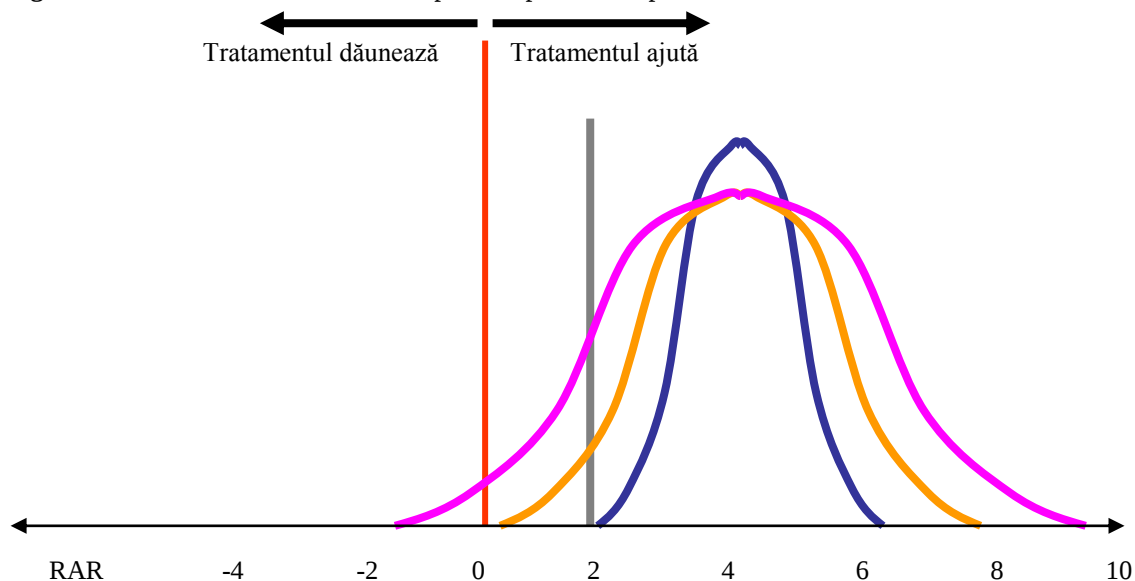


RR este de 0,5 (estimarea punctuală), intervalul de încredere 95% fiind [0,25 - 0,75], marcat prin liniile verticale scurte. Linia verticală lungă este RR de 1 (tratamentul nu reduce riscul, nici nu-l crește). Aria neagră = probabilitatea ca RR să fie mai mic de 0,25 (stânga) sau mai mare de 0,75 (dreapta) (fiecare dintre aceste probabilități este de 2,5%). Aria gri închis = probabilitatea de 47,5% ca RR să fie între 0,25 și 0,5 (probabilitatea scade pe măsură ce ne îndreptăm spre marginea 0,25 a CI). Aria gri deschis = probabilitatea de 47,5% ca RR să fie între 0,5 și 0,75 (probabilitatea scade pe măsură ce ne îndreptăm spre marginea 0,75 a CI). Interpretare: tratamentul a scăzut riscul la jumătate în studiul nostru, iar în realitate, cu o probabilitate de 95%, tratamentul scade riscul la între 1/4 și 3/4 din riscul bazal.

Pentru ca două tratamente să fie diferite (să existe semnificație statistică), trebuie ca intervalele de încredere ale rapoartelor (RR, OR) să nu-l cuprindă pe 1, ale diferențelor (RRR și RAR, sau diferența dintre scorurile de calitate a vieții, scoruri de durere sau pur și simplu diferențele dintre valorile TA, colesterolului etc.) să nu-l cuprindă pe 0, iar al NNT să nu tindă către infinit.

În Figura XII.2 sunt prezentate rezultatele a trei studii terapeutice pozitive. Toate trei arată aceeași reducere absolută a riscului (RAR), de 4%, însă gradul de precizie al estimării RAR este diferit, de la mic (studiul roz, cu CI95% cel mai larg, care cuprinde și valoarea 0, deci rezultatul nu este semnificativ statistic), la mare (studiul bleumarin, cu CI95% cel mai îngust). Din punct de vedere statistic, studiile bleumarin și portocaliu sunt semnificative (CI95% nu cuprind valoarea 0), deci suntem 95% siguri că tratamentul scade riscul. Din punct de vedere clinic (sau al sănătății publice), însă, se poate considera că, pentru boala și efectul (*end-point-ul*) respective, tratamentul nu merită făcut dacă nu produce o RAR de cel puțin 2%, și atunci

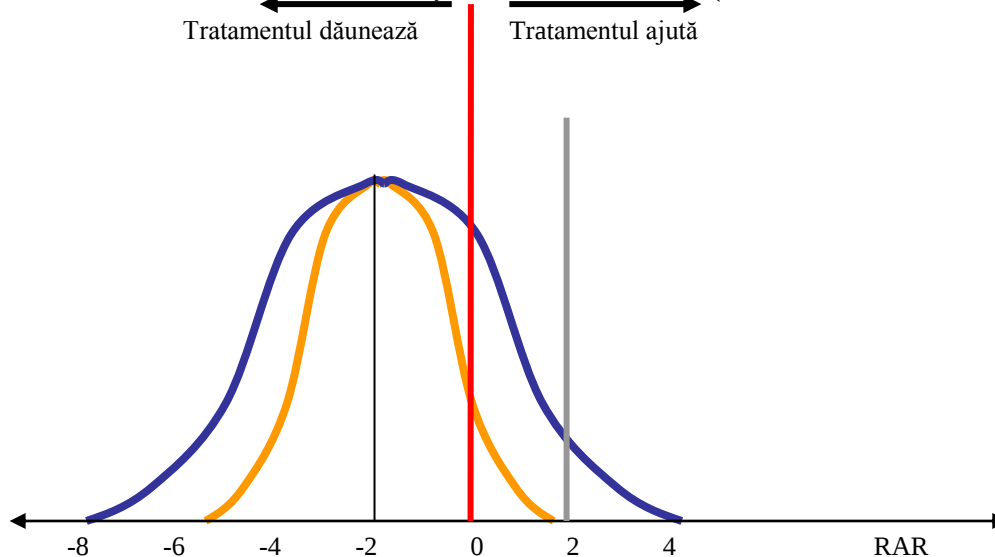
Figura XII.2. Rezultatele a 3 studii terapeutice ipotetice, exprimate sub forma RAR cu CI95%.



Pentru efectul (*end-point*-ul) respectiv, considerăm că o reducere absolută a riscului (RAR) de 2% (linia verticală gri) este cel mai mic beneficiu pe care pacienții îl consideră destul de important ca să merite să urmeze tratamentul. În fiecare caz, maximul înălțimii “clopotului” reprezintă estimarea punctuală a RAR, iar extremitățile sunt CI95%. Putem considera fie că este vorba de 3 RCT diferite din care din întâmplare a rezultat aceeași RAR de 4%, fie că este vorba despre același RCT, dar cu eșantioane diferite, determinând puteri statistice diferite ale studiului, de la eșantionul cel mai mic (clopotul roz), cu CI95% cel mai mare, la studiul cu eșantionul cel mai mare (clopotul bleumarin), cu CI95% cel mai îngust.

Figura XII.3. Rezultatele a 2 studii terapeutice ipotetice “negative”, exprimate sub forma RAR cu CI95%.

Ambele studii au arătat că **tratamentul crește** riscul absolut cu 2% (rezultat nesemnificativ statistic, deoarece



intervalele de încredere cuprind valoarea 0). Dacă ambele extremități ale CI95% ar fi fost mai mici decât 0, studiul ar fi arătat clar că tratamentul dăunează. Considerând, ca și în Figura 2, RAR=2% ca valoare de prag a semnificației clinice și/sau economice (linia verticală gri), studiul portocaliu poate fi considerat clar negativ, nu însă și studiul bleumarin.

nici studiul portocaliu nu este concludent, pentru că intervalul de încredere cuprinde valoarea 2. Dintre toate, numai studiul bleumarin este clar pozitiv.

În Figura XII.3 sunt prezentate rezultatele a două studii terapeutice “negative”, în care estimarea punctuală arată că tratamentul ar crește riscul absolut cu 2%. Dacă ambele

extremități ale CI95% ar fi fost în stânga liniei roșii, am fi putut spune că sigur tratamentul nu ajută (ba chiar mai mult, că sigur tratamentul face rău). Studiul portocaliu este negativ – chiar în eventualitatea, destul de improbabilă că, în realitate, RAR s-ar găsi între linia roșie = RAR de 0 și extremitatea superioară a intervalului de încredere, rezultatul tot n-ar fi depășit limia semnificației clinice sau economice pe care ne-am fixat-o la 2% pentru acest tratament și efect, așadar nu ar fi meritat. Celălalt studiu, însă, cel bleumarin nu este clar negativ, deoarece nu este exclusă posibilitatea ca tratamentul să scadă, totuși, riscul cu mai mult de 2% (extremitatea superioară a intervalului de încredere depășește valoarea de 2%). Așadar, deoarece intervalul de încredere este prea larg, cuprinzând nu numai valoarea 0, ci și valori importante din punct de vedere clinic, putem spune că studiile portocaliu și roz din figura XII.2 și bleumarin din figura XII.3 nu sunt concludente, și trebuie făcute alte studii cu eșantioane mai mari pentru a îngusta intervalul de încredere și a căpăta o siguranță mai mare. Altă soluție este, atunci când dispunem de mai multe astfel de studii neconcludente, cu intervale de încredere largi, să le cumulăm și să facem o metaanaliză (care nu este decât o altă metodă de mărire a eșantionului, și deci a puterii statistice).

În Tabelul XII.1 sunt prezentate rezultatele a patru studii ASCOT¹, trei imaginare, efectuate pe eșantioane de 50, 10 și 5 ori mai mici și cel real, efectuat pe 19.257 pacienți. Rezultatele se referă la unul dintre obiectivele secundare – total evenimente și proceduri cardiovasculare. Se vede cum intervalul de încredere devine din ce în ce mai îngust pe măsură ce se crește eșantionul. Dacă studiul ar fi fost efectuat pe 385 pacienți (de 50 ori mai puțin decât în realitate), intervalul de încredere ar fi fost extrem de larg și nu s-ar mai fi obținut semnificație statistică: RRR variază de la cu 34,5% mai mare în ramura amlodipină față de atenolol până la cu 47,6% mai mic, iar NNT de la trebuie să tratăm 10 pacienți cu amlodipină/peridopril în loc de atenolol/tiazidic petru a salva 1 pacient, până la trebuie să tratăm 22 de pacienți cu atenolol/tiazidic în loc de amlodipină/perindopril pentru a salva 1 pacient.

Tabelul XII.1. Variațiile intervalelor de încredere ale riscurilor și NNT în funcție de mărimea eșantionului (studiul ASCOT cu eșantion real și eșantioane imaginare mai mici) – *end point*: total evenimente și proceduri cardiovasculare.

Mărimea eșantionului	Incidența în grupul atenolol	Incidența în grupul amlodipină	RR (CI95%)	RRR (CI95%)	RAR (CI95%)	NNT (CI95%)
: 50 (385)	27/193	32/192	0,85 (0,53-1,35)	15% [(-34,5)-47,6]	2,5% [(-4,5)-9,9]	40 [10-(-22)]
: 10 (1926)	136/964	160/962	0,85 (0,69-1,05)	15% [(-4,7) – 31,3]	2,5% [(-0,7) – 5,7]	40 [18-(-142)]
: 5 (3.850)	272/1927	320/1923	0,85 (0,73-0,98)	15% (1,6-27)	2,5% (0,2-4,8)	40 (21-500)
Real (19.257)	1362/9639	1602/9618	0,85 (0,78-0,90)	15% (10-22)	2,5% (1,5-3,5)	40 (29-67)

Dacă studiul ar fi fost efectuat pe 3850 pacienți (de 5 ori mai puțin decât în realitate), s-ar fi obținut semnificația statistică, însă intervalul de încredere s-ar fi întins până la limita acesteia (0,98 pentru RR, 1,6% pentru RRR, 0,2% pentru RAR).

Și acum trebuie să discutăm despre o chestiune foarte delicată, care este semnificația clinică (vezi și Capitolul X.Din nou despre măsura efectului, semnificație clinică și semnificație statistică). Delicată pentru că este subiectivă și arbitrară.

Recunosc că eu, care sunt obsedat de semnificația clinică, am căutat în multe locuri să vedem dacă a stabilit, vreodată, cineva niște praguri peste care să spunem că un rezultat este semnificativ clinic. În studiile de cost-eficiență, acest prag este stabilit la 50.000 USD pentru un an de viață câștigat (stabilit în țările dezvoltate, pentru cele mai sărace s-ar putea să fie prea mult; unii nu concep să-și ia o mașină care nu are cinci stele NCAP la siguranță, 6 air-

baguri și ABS, EBD etc., alții nu au bani să-și cumpere nici una cu mult mai ieftină! Desigur, pentru idealisti o viață este inestimabilă și merită să plătești oricât pentru ea, însă avem de plătit pentru mai multe miliarde de vieți pe Pământ și nu avem cu ce).

După unii², RRR devine destul de importantă dacă este de măcar 25%, însă am văzut în capitolul IX că RR și RRR nu sunt măsuri ale semnificației clinice, pentru că sunt aceleași indiferent de riscul bazal, semnificația clinică fiind dată de RAR și NNT.

De fapt nu s-a stabilit un prag universal, pentru că nu există așa ceva, acesta depinde de multe lucruri. Pragul este stabilit, de la caz la caz de responsabilii cu politica sanitară și de pacienți (clinicienii sunt ultimii care să participe la așa ceva). Responsabilii cu sănătatea publică îl stabilesc în primul rând în funcție de costuri (una este să reduci riscul absolut cu 1% cu aspirină, sau un diuretic tiazidic, medicamente foarte ieftine, și alta este să realizezi același lucru cu clopidogrel, sau un trombolitic de ultimă generație, ori amlodipină sau perindopril; desigur, în costuri intră multe alte cheltuieli înafara prețului tratamentului). Pentru pacient, pragul dincolo de care merită mai degrabă să facă un tratament decât altul este stabilit luând în considerare prețul pe care îl plătește el, calitatea vieții (scăzută de un tratament prin reacțiile adverse, sau crescută de un medicament care este, poate, mai scump), elemente pe care fiecare pacient le evaluează în felul său, în funcție de venituri și filozofia lui de viață. De aceea, în timp ce responsabilii cu politica sanitară trebuie să găsească ei informația și să evalueze dacă un tratament merită sau nu mai mult decât altul să fie compensat sau gratuit, pacientul trebuie să ia o hotărâre în funcție de preferințele lui, după ce medicul îi explică ce obține de fapt în raport cu un alt tratament sau cu nici un tratament (RAR, NNT) – vezi Capitolul XVI. Comunicarea riscului.

Dacă stabilim pragul de semnificație clinică al RAR la 2% (pacientul consideră că nu merită să cumpere, sau responsabilii cu politica sanitară consideră că nu merită să compenseze amlodipină+perindopril în loc de atenolol+un tiazidic decât dacă acest din urmă regim scade riscul absolut cu cel puțin 2%), studiul ASCOT imaginar pe un eșantion de 3.850 de pacienți (de 5 ori mai mic decât în realitate) (Tabelul XII.1) nu este concludent, deoarece marginea inferioară a intervalului de încredere trece destul de bine sub pragul de 2%.

Cum studiul ASCOT a fost efectuat în realitate pe un eșantion mult mai mare, iar intervalul de încredere este mai îngust (1,5%-3,5%), cu o probabilitate foarte mare RRA real se află mai sus de 2%. Rămâne doar de judecat dacă această diferență (între 1,5 și 3,5% după 5,5 ani de tratament, cu un NNT corespunzător de 29-67, și am ales din articol *end-point*-ul cu semnificația statistică cea mai mare) este destul de tentantă, ținând cont de prețul cu mult mai mare al combinației amlodipină+perindopril. Chiar dacă luăm varianta optimistă a intervalului de încredere, din 29 de pacienți 28 dau diferența de preț degeaba, și poate că dacă ar fi întrebați ar prefera să o dea lunar pe o carte sau, de ce nu, pe o sticlă de vodcă. E viața lor și au dreptul să aleagă.

Dintre pacienții pe care îi vedem atât de des la jurnalele de știri în fața farmaciilor, descumpăniți că nu li se mai compensează medicamentele scumpe pe care le luaseră până atunci, niciunul nu știe care sunt RAR sau NNT. Normal, din moment ce nici medicii care le-au prescris habar nu au! Aceste lucruri sunt foarte sensibile din punct de vedere filozofic și politic – acum puține luni a fost mare scandal în Marea Britanie când National Institute for Clinical Excellence (NICE) a hotărât să sisteze compensarea medicamentelor folosite în demență în stadiile incipiente, pentru că efectul lor nu merita cheltuiala (2,5 lire sterline pe zi)³.

Bibliografie

1. Dahlof B, Sever PS, Poulter NR et al, for ASCOT investigators. Prevention of cardiovascular events with an antihypertensive regimen of amlodipine adding perindopril as required versus atenolol adding bendroflumethiazide as required, in the Anglo-Scandinavian

Cardiac Outcomes Trial – Blood Pressure Lowering Arm (ASCOT-BPLA): a multicentre randomised controlled trial. *Lancet*. 2005; 366:895-906.

2. Sackett D, Haynes B, Guyatt G, Tugwell T. Clinical epidemiology. A basic science for clinical medicine. 2nd Edition. London: Little, Brown 1991. p. 203.

3. <http://news.bbc.co.uk/2/hi/health/6036519.stm> (accesat ultima dată pe 18/01/2007)