

## XI. Erorile de tip I și II, puterea și mărimea eșantionului

### $\alpha$ și testarea ipotezei

Facem un experiment fiindcă vrem să știm dacă tratamentul modifică prognosticul. Pentru aceasta avem nevoie de analiză statistică deoarece răspunsul poate fi (sau mai bine zis de cele mai multe ori este) ambiguu.

Cele mai multe programe statistice prezintă rezultatele sub forma unei valori  $p$  care răspunde la întrebarea: dacă tratamentul nu are nici un efect în realitate, care este probabilitatea ca numai variabilitatea întâmplătoare să ducă la o diferență egală sau mai mare decât cea care a rezultat în studiul nostru? Valoarea  $p$  este o fracție care ia valori de la 0 la 1. Dacă  $p$  este mic, suntem înclinați să credem că diferența observată în studiul respectiv se datorează tratamentului, și nu întâmplării.

Întotdeauna vrem să luăm o hotărâre în urma studiului, și anume dacă diferența observată este sau nu semnificativă statistic, iar acest lucru se face foarte simplu: înainte de a începe studiul, se decide o valoare de prag pentru  $p$ , valoare numită  $\alpha$ , care este de obicei egală cu 0,05. După efectuarea experimentului în urma căruia a fost calculată valoarea lui  $p$ , dacă  $p < \alpha$  tratamentul a avut un efect semnificativ statistic (și respingem ipoteza nulă că tratamentul nu a fost eficient), iar dacă  $p > \alpha$  experimentul nu ne convinge că tratamentul a avut vreun efect.

### Erorile de tip I și II

Concluzia că un rezultat este sau nu semnificativ statistic este mai puțin solidă în realitate, și asta pentru că putem face două tipuri de greșeli:

- Eroarea de tip I ( $\alpha$ ) apare atunci când tragem concluzia că un efect este semnificativ statistic, însă diferența a fost datorată întâmplării. Tratamentul nu a avut nici un efect, iar variabilitatea întâmplătoare prezentă în cele două grupuri ne-a făcut să tragem concluzia greșită că între ele există o diferență. Probabilitatea de a efectua o astfel de greșală este  $p$ .
- Eroarea de tip II ( $\beta$ ) apare atunci când tragem concluzia că nu există un efect semnificativ statistic, dar tratamentul este eficace. Tratamentul a avut efect, însă variabilitatea întâmplătoare prezentă în cele două grupuri a ascuns diferența. Această eroare apare atunci când studiul nu a avut destulă **putere** statistică.

### $\beta$ și puterea

Chiar dacă tratamentul are efect, este posibil să nu obținem o diferență semnificativă statistic. Pur și simplu din întâmplare, datele noastre pot duce la un  $p > \alpha$ , cu alte cuvinte facem o eroare de tip II ( $\beta$ ).

Probabilitatea ca, atunci când există o diferență în realitate, să obținem și noi semnificație statistică în studiul nostru se numește **putere**.

$\beta + \text{puterea} = 100\%$ , de unde se vede că ele sunt complementare (crescând puterea, scădem probabilitatea de a face o eroare de tip II și invers).

Putem privi experimentul (studiul) nostru ca pe un test diagnostic prin care vrem să vedem dacă un tratament este sau nu eficient. Astfel,  $\alpha$  este rata fals pozitivilor,  $(100\% - \alpha)$  este specificitatea, puterea este sensibilitatea, iar  $\beta$  este rata fals negativilor.

Testul nostru va fi cu atât mai sensibil cu cât puterea este mai mare și cu atât mai specific cu cât  $\alpha$  ( $p$ ) este mai mic.

### Mărimea eșantionului

De câți subiecți (date) avem nevoie pentru un studiu?

Răspunsul depinde de:

1. Cât de mare este diferența pe care vrem să o evidențiem (semnificația clinică)?

Cu cât efectul clinic este mai mare, cu atât avem nevoie de mai puțini subiecți, și invers. Cum în zilele noastre efectele clinice ale tratamentelor sunt foarte mici (reduceri absolute ale riscului = RAR uneori de sub 1%), studiile sunt efectuate pe efective din ce în ce mai mari. Reprezentanții firmelor farmaceutice se laudă de multe ori cu mărimea eșantionului din studiul lor (“studiul a fost făcut pe x mii sau zeci de mii de pacienți”); acest lucru nu este neapărat de bine, ci arată că efectul clinic al tratamentului a fost atât de mic, încât a fost nevoie de eșantioane uriașe pentru a se atinge semnificația statistică.

Am văzut la exprimarea riscului, că RAR depinde de riscul bazal, așadar cu cât riscul bazal este mai mare, cu atât vom avea nevoie de mai puțini pacienți și invers. Studiile de prevenție terțiară (statine la pacienți care deja au făcut un infarct) au nevoie de mai puțini pacienți decât cele de prevenție secundară (statine la pacienți cu boală coronariană), de eșantionul cel mai mare având nevoie studiile de prevenție primară (statine la sănătoși); chiar și studiile de prevenție primară se fac la pacienți aflați la un risc rezonabil de mare pentru a nu ne trezi că studiul nu a avut destulă putere (cum spunea un prieten francez, profesor de medicină internă și epidemiologie clinică, de ce s-au făcut studiile HPS<sup>i</sup> și ASCOT<sup>ii</sup> în Marea Britanie și Scandinavia, și nu în țări mediteraneene?).

Se poate întâmpla ca estimarea frecvenței evenimentelor din ramura martor să nu fie adecvată [estimarea se face din studii de prevalență sau alte studii terapeutice; este posibil ca eșantionul nostru să nu se potrivească cu estimarea din întâmplare, sau pentru că populația noastră a fost diferită de cea din alte studii, sau pentru că estimarea a fost prea vagă deoarece nu am găsit alte studii în domeniul respectiv sau care să urmărească efectul (*end-point*-ul) respectiv]. O astfel de surpriză au avut coordonatorii studiului EUROPA<sup>iii</sup>, care a trebuit, pentru a avea mai multe evenimente încât să obțină semnificație statistică, să prelungească durata studiului cu un an (soluție preferabilă includerii de noi pacienți).

Alegând pacienți la risc mare pentru a crește puterea statistică poate avea și efect invers, ca în studiul ISOLDE<sup>iv</sup>, unde din cauza gravității bolii pacienții au abandonat studiul în număr foarte mare și atunci retragerea din studiu a devenit efect de studiat (când se simte foarte rău, pacientul nu mai simte nici un efect placebo, iar când nici medicamentul nu are efect, normal că abandonează pacienții din ambele ramuri ale studiului!).

2. Cât de mare a fost variabilitatea datelor?

Cu cât variabilitatea datelor este mai mare (deviația standard a valorilor comparate înainte și după tratament, de exemplu), cu atât este nevoie de un eșantion mai mare.

3. Ce eroare de tip I ne asumăm (cât suntem de dispuși să găsim o diferență care nu există în realitate). De obicei  $\alpha=0,05$ , dar uneori poate vrem să obținem un  $p$  mai mic (atunci când investim mulți bani într-un studiu) sau suntem dispuși să acceptăm un  $p$  mai mare (într-un studiu pilot, de exemplu). Cu cât vrem să obținem un  $p$  mai mic, cu atât vom avea nevoie de un eșantion mai mare.

4. Ce eroare de tip II ne asumăm (cât suntem de dispuși să ratăm evidențierea unei diferențe care există în realitate). Prin convenție,  $\beta$  trebuie să fie cel mult 20% (deci puterea măcar 80%).

Majoritatea studiilor publicate în NEJM (cea mai bună revistă medicală din lume!) în anii 1970, și lucrurile nu s-au îmbunătățit în deceniul următor, nu aveau puterea statistică de a evidenția efectul propus<sup>v</sup>. În zilele noastre, dat fiind că majoritatea studiilor terapeutice sunt sponsorizate de industria farmaceutică și aceasta nu-și permite să cheltuiască sume uriașe fără a reuși să evidențieze un efect cât de mic pentru a-și vinde medicamentul, studiile sunt

proiectate de la început ca să aibă o putere destul de mare (se vede calculul în secțiunea Metode a articolelor).

5. De cât timp și bani dispune cel care face (sponsorizează) studiul. Modificări relativ mici de  $\alpha$  sau putere duc la modificări mari ale eșantionului (pentru că efectul medicamentului este dat, nu putem umbla decât la erorile de tip I și II atunci când calculăm mărimea eșantionului) și atunci, în funcție de banii și timpul disponibile, îți asumi riscuri mai mari sau mai mici.

### Calculul puterii unui studiu

Atunci când întâlnim un studiu în care nu s-a ajuns la o semnificație statistică, vrem să știm dacă studiul este într-adevăr negativ sau tratamentul ar fi putut avea efect, însă studiul nu a avut putere să îl evidențieze. Pentru asta, putem introduce datele din studiu (numărul de pacienți incluși în fiecare grup, numărul de pacienți care au suferit efectul urmărit și RR sau RAR) și obținem puterea pentru  $\alpha=0,05$  (care a fost puterea studiului pentru a obține măcar o semnificație statistică de  $p=0,05$ ?).

Autorii studiului PEACE<sup>vi</sup> au dorit să demonstreze, după studiile HOPE<sup>vii</sup> cu ramipril și EUROPA<sup>iii</sup> cu perindopril, că și trandolaprilul ar scădea morbi-mortalitatea la pacienții coronarieni fără hipertensiune sau insuficiență cardiacă. Studiul PEACE nu a fost, însă, pozitiv și s-a vehiculat ideea că, pacienții de aici fiind la risc mai mic decât cei din HOPE și EUROPA, nu s-a atins semnificația statistică pentru că studiul nu a avut puterea necesară. Și într-adevăr, pentru efectul principal (un efect compozit, altul decât cel din EUROPA) puterea studiului a fost de 16%, pentru mortalitatea cardiovasculară puterea a fost de 10%, iar pentru mortalitatea generală puterea a fost de 32% (calculare efectuate cu Epi Info 6<sup>1</sup>). Dacă merita să mărim talia eșantionului pentru a obține semnificație statistică vom vedea la discutarea intervalelor de încredere (Capitolul XII).

---

### Bibliografie

- i. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20,536 high risk individuals: a randomized placebo-controlled trial. *Lancet* 2002 360: 7-22.
- ii. Sever PS et al. Prevention of coronary and stroke events with atorvastatin in hypertensive patients who have average or lower-than-average cholesterol concentrations, in the Anglo-Scandinavian Cardiac Outcomes Trial-Lipid-Lowering Arm (ASCOT-LLA): a multicentre randomised trial. *Lancet* 2003; 361: 1149-1158.
- iii. *The EUROpean trial On reduction of cardiac events with Perindopril in stable coronary Artery disease Investigators*. Efficacy of perindopril in reduction of cardiovascular events among patients with stable coronary artery disease: randomised, double-blind, placebo-controlled, multicentre trial (the EUROPA study). *Lancet*. 2003; 362: 782–88.
- iv. P S Burge PS, Calverley PMA, Jones PW, Spencer S, Anderson JA, Maslen TK on behalf of the ISOLDE study investigators. Randomised, double blind, placebo controlled study of fluticasone propionate in patients with moderate to severe chronic obstructive pulmonary disease: the ISOLDE trial. *BMJ* 2000;320:1297–303
- v. Freiman JA, Chalmers TC, Smith H, Kuebler RR. The importance of beta, the type II error, and sample size in the design and interpretation of the randomized controlled trial. Survey of

---

<sup>1</sup> se poate lua gratuit de pe site-ul Center for Disease Control: <http://www.cdc.gov/epiinfo/Epi6/EI6dnjp.htm>

two sets of “negative” trials. În JC Baillar III, F Mosteller (editors), Medical uses of statistics, NEJM Books, Boston, Massachusetts, 2d Edition, 1992, p.357-73.

vi. The PEACE Trial Investigators. Angiotensin-Converting-Enzyme Inhibition in Stable Coronary Artery Disease. *N Engl J Med* 2004;351:2058-68.

vii. The Heart Outcomes Prevention Evaluation Study Investigators. Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high-risk patients. *N Engl J Med* 2000;342:145-153.