

V. Semnificație statistică; p și intervale de încredere; semnificație clinică

P și intervalul de încredere sunt rezultatele invariabile ale testelor statistice, și drept urmare le găsim în toate articolele ce descriu o cercetare originală (încă din abstract). Așadar, în finalul unei cercetări, se efectuează analiza datelor, iar rezultatele acestora sunt p și intervalul de încredere (CI), care ne arată binecunoscuta “semnificație statistică”.

De ce avem nevoie de statistică? Pentru că vrem să tragem concluzii cât mai valide din cantități limitate de date și diferențe importante sunt deseori mascate de variabilitatea biologică și/sau imprecizia experimentală. Pe de altă parte, mintea umană excelează în găsirea de tipare și relații și tinde să generalizeze în exces.

Se presupune că populația este infinită, iar noi ne facem întotdeauna cercetările pe un eșantion finit, fie că este vorba de câteva zeci de subiecți, fie că sunt câteva zeci de mii (ca în marile studii cardiologice, de ex. ALLHATⁱ). Statistica (în particular p și CI) ne folosește tocmai pentru a vedea, la sfârșit, dacă rezultatele obținute pe eșantionul nostru sunt valabile în general, pe întreaga populație și pot fi extrapolate la aceasta, sau sunt rezultatul întâmplării survenite în eșantionul nostru.

Să presupunem că vrem să vedem dacă fumatul este factor de risc pentru infarctul de miocard. Pentru aceasta, alegem un eșantion de n pacienți (numărul se calculează în funcție de 1) semnificația clinică a fumatului = riscul relativ și/sau riscul atribuibil care consider că merită osteneala a fi evidențiate, și de 2) semnificația statistică pe care vreau să o obțin). Îi urmărim și numărăm câți fac infarct dintre fumători și câți dintre nefumători, și calculăm riscul relativ (RR)=2; în urma aplicării unui test statistic (în acest caz tip X^2), obținem un $p=0,01$, iar calculând intervalul de încredere al RR, obținem $CI \in [1,3, 4]$.

În privința p -ului, acest lucru nu înseamnă altceva decât că, dacă în realitate (la nivelul populației) riscul de a face infarct al fumătorilor ar fi egal cu riscul de a face infarct al nefumătorilor, probabilitatea ca noi să fi obținut un $RR \geq 2$ pe un eșantion de n subiecți este de 1%. Cu cât p este mai mic, această probabilitate (ca rezultatul obținut de noi în eșantionul nostru să nu fie cel real) este mai mică.

Intervalul de încredere (de obicei se calculează cel 95%) ne dă mai multe informații: în exemplul nostru, ne spune că în realitate (adică la nivelul populației), suntem 95% siguri că riscul relativ este între 1,3 și 4, adică în realitate, dacă fumezi ai un risc de a face infarct miocardic de la de 1,3 până la de 4 ori mai mare decât dacă nu ai fuma. (Aceasta este interpretarea intervalului de încredere. De fapt, intervalul de încredere 95% înseamnă că dacă noi am repeta acest studiu pe n pacienți de 100 de ori, în 95 dintre cazuri am găsi RR între 1,3 și 4).

A nu se confunda semnificația statistică cu cea clinică! Nu înseamnă că dacă p ar fi fost 0,0001 în exemplul de mai sus, fumatul ar fi fost un factor de risc mai mare! Ar fi însemnat doar că sunt mult mai sigur că $RR=2$ obținut în studiul meu e real, iar intervalul de încredere corespondent ar fi fost mult mai îngust!

Dacă, de exemplu, vrem să demonstrăm că medicamentul m prelungeste viața în insuficiența cardiacă, iar rezultatul este că cei tratați cu placebo au trăit în medie 10 ani, iar cei tratați cu m au trăit în medie 10 ani și o zi, $p=0,000001$ arată doar că sunt foarte, foarte sigur că acea zi în plus se datorează medicamentului m , și nu este rezultatul întâmplării. Cu cât p este mai mic, intervalul de încredere este mai îngust, iar gradul de certitudine mai mare.

Un anumit nivel al lui p care indică faptul că o asociere (cum au fost asocierile dintre fumat și infarctul miocardic, sau între medicamentul m și supraviețuirea mai lungă, în exemplele

noastre) este semnificativă statistic este determinat arbitrar; în cercetarea medicală, acest nivel este stabilit, prin convenție, la 0,05.

Aplicând intervalele de încredere, asocierea (sau diferența) nu este semnificativă statistic atunci când intervalul de încredere îl cuprine pe 1, în cazul rapoartelor (de ex. risc relativ, odds ratio), sau pe 0, în cazul diferențelor (de ex. risc atribuibil, reducerea relativă a riscului, reducerea absolută a riscului etc.) (sau ∞ în cazul NNT).

Drept dovadă că p și intervalele de încredere ne indică același lucru (sunt în același timp semnificative sau nu statistic), atunci când $p=0,05$, una dintre extremele intervalului de încredere 95% este 1, în cazul rapoartelor și 0, în cazul diferențelor.

Intervalul de încredere ne dă – spre deosebire de p – și informații asupra puterii statistice a studiului, atunci când acesta este negativ (vezi și capitolele XI, XII, XV). De exemplu, într-un studiuⁱⁱ care compara două dispozitive pentru cicatrizarea ulcerelor piciorului diabetic, unul clasic și scump cu unul nou și mult mai ieftin, autorii ajung la concluzia că cel nou, mai ieftin, este la fel de bun ca și primul, deoarece nu s-a evidențiat o diferență statistică între ele ($p=0,21$). Cum avem în fiecare grup numai câte douăzeci de pacienți, ne punem pe bună dreptate întrebarea dacă într-adevăr nu există nici o diferență, sau puterea statistică a fost prea mică pentru a evidenția una. Dacă revista în care a fost publicat studiul ar fi cerut de la autori și intervalele de încredere, sau recenzorii le-ar fi calculat ei înșiși, ar fi văzut că de studiul nu a avut suficientă putere statistică: vindecarea este cu 10% mai mică în grupul cu dispozitivul mai ieftin (reducerea absolută a riscului = RAR), intervalul de încredere 95% fiind [-8,3%, 28,3%] - într-adevăr, nesemnificativ statistic (conține valoarea 0), dar vedem că în realitate dispozitivul mai ieftin poate duce de la creșterea cu 8% a vindecării, la scăderea cu 28,3% a vindecării, ultima cifră fiind mai mult decât semnificativă clinic în favoarea dispozitivului mai scump! În concluzie, studiul nu are nici o valoare și nu reușește să arate nici pe departe ce își propune (concluzia abstractului a fost: “Dispozitivul Optima Diab este la fel de sigur și eficient ca și dispozitivul TCC în tratamentul ulcerului piciorului diabetic, dar costul său mai scăzut îl fac să fie preferat în centrele care se ocupă de piciorul diabetic...”)!

Bibliografie

ⁱ. ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial. Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic: The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). *JAMA*. 2002; 288:2981-97.

ⁱⁱ. Piaggese A, Macchiarini S, Rizzo L et al. An Off-the-Shelf Instant Contact Casting Device for the Management of Diabetic Foot Ulcers. A randomized prospective trial versus traditional fiberglass cast. *Diabetes Care* 2007; 30:586–590