

EPIDEMIOLOGIE CLINICĂ ȘI BIOSTATISTICĂ

Cristian Băicuș, Paul Bălănescu (UMF Carol Davila, Medicină internă și Metodologia cercetării, Spitalul Clinic Colentina)

1. Tipuri de studii

Toate cunoștințele medicale provin din studii științifice, pornind de la cercetările fundamentale și sfârșind cu cele pe oameni - de la cazuri (practic studiile pe care le fac medicii zilnic) la studiile clinice randomizate. Din punctul de vedere al medicului practician și al celui de sănătate publică, rezultatele diferitelor tipuri de studii sunt mai mult sau mai puțin credibile în funcție de cât de corecte (valide) sunt respectivele studii. Bineînțeles că un studiu poate fi făcut mai bine sau mai prost (există studii clinice randomizate proaste, și există studii caz-martor bune), însă oricât de bine este făcut un studiu, acesta nu-și poate depăși condiția, adică un studiu caz-martor perfect rămâne un studiu caz-martor, deoarece prin design acesta este predispus la mai multe potențiale erori sistematice (*bias*) decât unul de categorie superioară (de cohortă, sau randomizat).

Epidemiologia este de două feluri, descriptivă și analitică, ca și statistica ce o însoțește, epidemiologia și statistica descriptivă ocupându-se cu studii care duc la simple descrieri sau tendințe, pe când în epidemiologia analitică intră studiile în care comparăm, căutând diferențe sau asocieri între factori de risc și efecte, deci în studiile analitice există întotdeauna un grup martor, iar rezultatul statistic va fi acel celebru *p* mic.

Epidemiologia descriptivă conține clasic **cazurile** (până la 5) și **seriile de cazuri** (peste 5), studiile ecologice (de corelație) și **studiile transversale** (*cross-sectional*).

Din cazuri sau serii de cazuri, pe care le întâlnim zilnic, învățăm permanent, numai că de multe ori concluziile pe care le tragem nu sunt corecte, pe de o parte din cauza variabilității biologice, iar pe de alta pentru că mintea omenească excelează în găsirea de modele și relații care în realitate nu există. De exemplu, o fetiță de 4 ani i-a spus băiețelului cu care se juca că el nu va putea fi medic, pentru că singurii trei medici pe care ea îi cunoștea erau femei, și din această experiență limitată generalizase în mod greșit (1).

Studiile transversale se mai numesc și studii de prevalență, deoarece utilitatea lor principală este în stabilirea prevalenței (frecvența unei boli sau factor de risc în populație la un moment dat, numărul de cazuri/populație). Clasic fac parte din epidemiologia descriptivă (descriu prevalențe, repetate la intervale de timp pot arăta tendințe), ele pot fi folosite și ca studii analitice: într-o

populație de indivizi despre care nu știm nimic (nici dacă sunt expuși, nici dacă au boala), măsurăm în același timp prezența unui factor de risc = expunerea (prezența unei proteine C reactive=PCR crescute) și a unui efect (prezența bolii coronariene), și căutăm să vedem dacă există o asociere între cele două. **Mărimea asocierii** care rezultă din studiul transversal este dată de **raportul prevalențelor** (prevalența bolii coronariene la cei cu PCR crescută / prevalența bolii coronariene la cei cu PCR scăzută; dacă acesta este 2, spunem că cei cu PCR crescută au un risc de două ori mai mare să aibă boală coronariană decât cei cu PCR normală). După cum vă dați seama, **factorul de risc** și **efectul** fiind măsurate în același timp, lipsește argumentul temporal de cauzalitate (nu știm dacă întâi a fost PCR crescută și apoi boala coronariană, condiție esențială a cauzalității, sau invers).

Și astfel am ajuns la **studiile analitice (epidemiologia analitică)**, care sunt de două feluri, **observaționale** și **intervenționale**. Cele observaționale, în care cercetătorul doar observă/măsoară și nu intervine asupra expunerii, sunt **studiile de cohortă** și cele **caz-martor**. În studiile de cohortă direcția cercetării este de la expunere către efect (boală), de exemplu luăm populația din orașul Framingham și dozăm colesterolul fiecărui individ, după care îi urmărim pe toți pentru a vedea incidențele (definite ca numărul de cazuri noi de boală raportate la întreaga populație la risc) bolii coronariene, precum și dacă acestea sunt diferite la cei cu colesterolul mare față de cei cu colesterolul normal. De aici va rezulta raportul incidențelor, sau **riscul relativ**; dacă acesta este 2 (așadar supraunitar, >1), putem spune că cei cu colesterolul mare au un risc de două ori mai mare să facă boală coronariană decât cei cu colesterolul normal. Dacă acesta este 1, înseamnă că nu există nici o diferență, iar colesterolul nu este **factor de risc** pentru boala coronariană. Dacă, dimpotrivă, riscul relativ este 0,5 (așadar subunitar, <1), putem spune că expunerea la valori mari de colesterol este un **factor de protecție** (cei cu colesterol mare au un risc la jumătate față de cei cu colesterol normal să facă boală coronariană). Se vede și cum în studiile de cohortă, care sunt prin designul lor prospective, se poate ști întotdeauna că în succesiunea temporală a fost mai întâi factorul de risc prezent (expunerea), și apoi a apărut boala, deci acest important criteriu de cauzalitate este satisfăcut.

În **studiile caz-martor** este invers, ele prin design sunt retrospective pentru că pornim de la cazuri (cei care deja au boala, cazurile, pacienți cu boală coronariană), cărora le asociem niște martori care ar trebui să fie similari cazurilor cu excepția faptului că nu au boala (aici intervin cele mai

multe erori sistematice, în alegerea cazurilor); apoi măsurăm atât la cazuri cât și la martori prevalența factorului de risc (colesterol crescut), iar **măsura asocierii** care se calculează din aceste studii se numește **raportul cotelor** (*odds ratio*), care estimează **riscul relativ** pe care nu îl putem calcula direct aici, nefiind un studiu prospectiv din care să poată fi calculate incidențe. Din motive strict matematice, cota (*odds*) supraestimează întotdeauna incidența (riscul), iar *odds ratio* supraestimează întotdeauna riscul relativ.

În **studiile intervenționale** cercetătorul intervine, în sensul că include un grup de pacienți cu boală coronariană care nu iau statine, și prin tragere la sorți (randomizare) alocă, la întâmplare jumătate dintre pacienți grupului care va lua statine, și jumătate dintre pacienți grupului care va lua placebo. Prin acest proces de randomizare se realizează distribuția egală în cele două grupuri a celorlalți factori de risc pentru infarctul miocardic (adică în ambele grupuri va fi o distribuție egală în ceea ce privește vârsta, sexul, prevalența diabetului, hipertensiunii etc), astfel încât dacă la sfârșit va fi o diferență în ceea ce privește incidențele infarctului în cele două grupuri, vom putea spune că diferența se datorează tratamentului cu statină și nu a intervenit vreun **factor de confuzie**. Mărimea asocierii este măsurată tot prin **riscul relativ**, care este raportul dintre incidența infarctului în grupul expus (statină) și cel martor (placebo), iar dacă acesta este 0,7 putem spune că statina scade riscul de infarct la pacienții cu boală coronariană. Echivalentul observațional al acestui studiu ar fi unul de cohortă în care includem pacienți cu boală coronariană dintre care unii iau statină, alții nu iau, îi urmărim și vedem dacă există o diferență între incidențele infarctului între grupul statină și grupul martor. Dacă **riscul relativ** iese 0,7, înseamnă că incidența infarctului în grupul statină a fost mai mică decât în grupul martor, însă aici nu mai putem fi siguri că s-a întâmplat strict **din cauza** statinei, ci poate că în grupul statină s-au selectat pacienți mai bogați/educați, care au avut grijă de sănătatea lor din mai multe puncte de vedere (au făcut mișcare, și-au tratat și hipertensiunea și diabetul mai bine, au mâncat mai sănătos, nu au fumat etc., toți aceștia fiind posibili **factori de confuzie** care să fi fost adevărata cauză a scăderii incidenței infarctului în grupul statină); înseamnă că în acest caz a intervenit o **eroare sistematică de selecție** (*selection bias*) care a distorsionat rezultatele studiului în sensul scăderii incidenței infarctului în grupul statină. Pentru a vedea rolul independent al statinei, putem ajusta în analiza statistică (**analiză multivariabilă**, aici **regresie logistică**) pentru toți acești factori de confuzie, și să vedem dacă după ajustare statina mai pare să scadă în continuare riscul de infarct. Problema este că și după ajustare, nu putem

niciodată și dacă nu mai există și alți factori de risc pentru infarct la care nu ne-am gândit sau și mai probabil care nu au fost descoperiți încă, și atunci nu putem ajusta pentru ei, pe când într-un studiu clinic randomizat, prin procesul de randomizare grupurile vor fi egale atât în privința factorilor de risc cunoscuți, cât și a celor necunoscuți, deci suntem mai siguri de validitatea rezultatelor unui studiu clinic randomizat decât de ale unui studiu de cohortă. În studiile caz-martor pot interveni în plus alte erori sistematice, introduse fie de alegerea defectuoasă a martorilor, fie de imposibilitatea demonstrării succesiunii temporale factor de risc – efect. Din acest motiv, studiul clinic randomizat este considerat cel mai valid tip de studiu în ierarhia dovezii.

2. Mărimea asocierii

Mărimea asocierii este măsurată prin diferiți parametri, în funcție de tipul de studiu. În studiile cu design prospectiv, în care se pornește de la expunere și indivizii sunt urmăriți în timp pentru a vedea câți suferă efectul (studiile observaționale de cohortă și studiile clinice randomizate), mărimea asocierii este măsurată prin **riscul relativ** (*relative risk, risk ratio*), care este raportul dintre riscul la expuși și riscul la neexpuși, sau altfel spus raportul dintre incidența bolii în cohorta expusă (tratată) și incidența bolii în cohorta martor (placebo) [$\text{RiscExpuși}/\text{RiscMartor} = \text{IncidențaExpuși}/\text{IncidențaMartori(Placebo)}$]. Dacă raportul = 1, numitorul=numărătorul, deci riscul este același și la expuși, și la martori, așadar expunerea nu modifică riscul, expunerea respectivă nu are nici un efect. Dacă raportul >1 înseamnă că riscul este mai mare la expuși, așadar expunerea este factor de risc. Dacă raportul <1 înseamnă că riscul este mai mic printre expuși (tratați), deci expunerea este factor de protecție.

În celelalte tipuri de studii, neputându-se calcula incidența bolii, nu putem calcula riscul relativ, și atunci îl estimăm pe acesta prin alte măsuri: după cum am spus, în studiile transversale avem raportul prevalențelor, iar în studiile caz-martor avem raportul cotelor (*odds ratio*). Toate se interpretează la fel: dacă avem un risc relativ, sau raport al prevalențelor, sau *odds ratio* de 2 înseamnă că cei expuși au un risc de 2 ori mai mare de a suferi efectul decât cei neexpuși. De asemenea, din motive matematice, rezultatele analizei multivariabile vor fi sub formă de *odds ratio*, chiar dacă rezultatele provin dintr-un studiu prospectiv (cohortă sau randomizat).

Mărimea asocierii ne arată cât de mare este legătura etiologică (relația de cauzalitate) dintre un factor de risc și un efect (boală). Cu cât aceasta este mai mare, cu atât reprezintă un argument mai puternic pentru cauzalitate, iar atunci când este vorba despre studii observaționale (de cohortă,

transversale sau caz-martor) se recomandă a se lua în considerare numai rezultatele (risc relativ, *odds ratio*, raport al prevalențelor) mai mari decât 2, sau și mai bine mai mari decât 5 când este vorba despre factori de risc, sau mai mici decât 0,5 sau și mai bine mai mici decât 0,2 atunci când este vorba despre factori de protecție. Un risc relativ (sau *odds ratio*) de 2-3 rezultat dintr-un studiu observațional crește calitatea dovezii de la slabă la moderată, iar un risc relativ (sau *odds ratio*) de peste 5 crește calitatea dovezii la înaltă, adică aceeași cu a unui studiu clinic randomizat (2). Motivul este acela că, dacă s-au strecurat factori de confuzie care nu au putut fi înlăturați, și care au dus în mod fals la acest raport, se presupune că și dacă i-am descoperi pe aceștia, i-am măsura și am ajusta pentru ei, tot ar mai rămâne ceva din riscul relativ diferit de 1.

3. Ierarhia dovezii și gradele de recomandare

Există multe sisteme de ierarhizare a dovezii, din care decurge gradul de recomandare din ghiduri, aproape fiecare societate savantă având sistemul ei, însă principiile sunt comune. Ierarhizarea dovezii se face în funcție de validitatea ei, iar validitatea ați văzut că provine mai ales din designul studiului, astfel încât la vârf se află studiile clinice randomizate, după care urmează studiile observaționale (puse de multe ori la grămadă, deși validitatea studiilor de cohortă este mult mai mare decât a celor caz-martor), iar la sfârșit celelalte tipuri de studii (serii de cazuri, cercetare preclinică și pe animale, părerea expertului). În principiu, o sinteză sistematică din studii randomizate are o validitate mai mare decât unul sau câteva studii randomizate, ca și o sinteză sistematică de studii observaționale în raport cu studiile observaționale, deci o astfel de dovadă duce la recomandări mai puternice decât studiile unice.

Sistemul GRADE (2), adoptat de din ce în ce mai multe societăți pentru ghidurile lor clasifică dovezile ca fiind de calitate bună dacă provin din studii randomizate, slabă dacă provin din studii observaționale și foarte slabă dacă provin din orice alte studii. Calitatea poate fi scăzută dacă studiile au un risc mare de erori sistematice (*bias*), sau rezultatele sunt imprecise (interval de încredere larg, care poate include rezultate favorabile sau adverse importante; număr mic de evenimente, rezultatele putându-se schimba dacă unul sau două dintre ele nu ar fi existat). Calitatea este crescută dacă mărimea asocierii este importantă ($RR > 2$ sau $< 0,5$ crește calitatea studiilor observaționale la moderată, > 5 sau $< 0,2$ crește calitatea la înaltă).

4. Mărimea efectului

Mărimea efectului este parametrul la care trebuie să ne uităm în primul rând atunci când evaluăm rezultatele unui studiu, după ce am hotărât că acesta este bine făcut. Mărimea efectului ne arată importanța clinică a efectului, și este dată de mai mulți parametri. Înafară de evaluarea subiectivă a mărimii efectului, există studii care determină efectul minim important (*minimally important difference*), adică efectul care trebuie să apară pentru ca majoritatea pacienților chiar să își dea seama că le este mai bine, pentru diverse scale; pentru scala St.George de calitate a vieții în BPOC, efectul minim important este de 4, un efect mediu este de 8, iar un efect mare este de 12 unități. Combinația salmeterol + fluticazonă îmbunătățește calitatea vieții cu aproape 4 unități, iar placebo o îmbunătățește cu 2 (3).

Dacă efectul este o variabilă dihotomică, atunci vorbim despre riscuri și reduceri ale riscului. Astfel, când este vorba despre riscuri relative sau *odds ratio*, una este ca acesta să fie 0.9 (foarte aproape de 1), și alta de 0,2, când este vorba despre factori de protecție, și una este ca acesta să fie de 2 (factorul respectiv crește riscul de 2 ori), și alta de 12 (cum crește fumatul riscul de cancer pulmonar).

După cum am spus, mărimea riscului relativ (sau *odds ratio*, sau raportul prevalențelor) este o măsură a cauzalității. Măsura impactului unui factor este dat de **reducerea relativă a riscului**, **reducerea absolută a riscului**, numărul de pacienți care trebuie tratat pentru a salva unul (*number needed to treat*), și **riscul atribuibil**.

Să zicem că avem un studiu terapeutic, în care incidența deceselor în grupul placebo este de 10%, iar cea în grupul tratament este de 8%. Riscul relativ (RR) este în acest caz $8\%/10\%=0,8$. Reducerea relativă a riscului (RRR) este de $1-RR = 1-0,8 = 0,2$ (sau 20%), așadar putem spune că tratamentul reduce riscul de deces cu 20%. Reducerea absolută a riscului (RAA) se obține prin scădere (se mai numește și diferența riscului), adică $10\%-8\%=2\%$, adică la fiecare 100 de pacienți tratați, tratamentul respectiv salvează 2 (90 oricum nu mureau, iar 8 oricum mor, tratați sau netratați). Pentru a afla câți pacienți trebuie să tratăm pentru a salva unu (*number needed to treat*) împărțim $100/RAR = 100/2 = 50$, adică trebuie să tratăm 50 de pacienți pentru a salva 1. Trebuie să-i tratăm atâta timp cât a durat studiul, dacă acesta a durat 4 ani, înseamnă că trebuie să tratăm 50 de pacienți timp de 4 ani pentru a salva 1 pacient de la moarte.

Dacă avem un studiu etiologic de cohortă în care în cohorta expusă antiinflamatoarelor incidența infarctului miocardic este de 12%, iar în cohorta martor este de 7%, diferența riscului se numește

în acest caz **riscul atribuibil** = $12\% - 7\% = 5\%$, adică la 100 de pacienți expuși la antiinflamatoare, 7% oricum fac infarct (din cauza altor factori de risc), iar 5% fac infarct din cauza antiinflamatoarelor.

De multe ori, semnificația clinică trebuie judecată în context, iar evaluarea ei este mai degrabă un demers filozofic, variind de la individ la individ, în funcție de valoarea pe care o dă acesta efectului respectiv în contextul prețului medicamentului, al efectelor adverse sau pur și simplu al necesității de a face un tratament. Din acest motiv trebuie ca hotărârea că un efect este sau nu important clinic trebuie să o ia pacientul, pe baza informațiilor furnizate de medic, printr-un proces care se numește decizie împărtășită (*shared decision making*). Acest lucru arată, de altfel, și absurditatea ghidurilor care hotărăsc de la ce nivel al riscului (cardiovascular, de fractură de șold etc) să începem tratamentul cu statină sau antiosteoporoză, când acest prag nu poate fi hotărât decât de fiecare pacient pentru el (și de plătitor, în funcție de cât este dispus și poate să dea pentru un an de viață câștigat, de exemplu).

5. P și intervale de încredere

În statistica analitică se fac comparații și se caută diferențe sau asociații, de obicei între un factor de risc (sau prognostic) și un efect. De exemplu, dintr-un studiu făcut pe n pacienți rezultă un risc relativ de 2. Pe noi însă ne interesează întotdeauna care este adevărul – care este riscul relativ în realitate? Este el într-adevăr 2? Este într-adevăr acesta diferit de 1 (care înseamnă că, indiferent dacă indivizii sunt expuși sau nu, incidența bolii este aceeași, deci factorul respectiv nu este factor de risc)? Așadar, semnificația statistică ne ajută la extrapolarea rezultatelor de la eșantionul pe care s-a efectuat studiul, la realitate.

P ne arată care este probabilitatea ca, dacă în realitate riscul relativ este de 1 (deci factorul respectiv este indiferent), noi să fi obținut, în mod fals, din întâmplare, în studiul nostru pe n pacienți un risc relativ ≥ 2 . Dacă în urma testelor statistice aplicate am obținut $p=0,01$, înseamnă că probabilitatea ca în realitate să nu existe nici o legătură între factorul testat și efect este de 1%. Prin convenție s-a hotărât că riscul asumat de a săvârși o eroare de tip I (alfa) trebuie să fie mai mic decât 5%, și atunci valoarea de prag pentru semnificația statistică în medicină este $p=0,05$.

Intervalul de încredere ne dă mai multe informații decât p : pe lângă semnificația statistică, ne arată și valorile pe care le poate lua în realitate parametrul respectiv. De exemplu, dacă pentru studiul nostru de mai sus pe n pacienți avem riscul relativ (RR) de 2 cu intervalul de încredere 95%

(CI95%) între 1,2 și 4, înseamnă că în realitate RR se află, cu o probabilitate de 95%, între 1,2 și 4, adică cei expuși pot avea un risc de a face boala de între de 1,2 ori și 4 ori mai mare decât cei neexpuși (în realitate, CI95% ne spune că, dacă am repeta studiul nostru pe 100 de eșantioane de n pacienți, de 95 de ori RR s-ar afla între 1,2 și 4). Vă aduc aminte că, fiind un raport, $RR > 1$ înseamnă factor de risc (expunerea crește riscul), $RR = 1$ înseamnă factor indiferent și $RR < 1$ înseamnă factor de protecție (expunerea scade riscul). Din acest motiv, dacă intervalul de încredere conține valoarea 1, atunci nu avem semnificație statistică: $RR = 2$, CI95% [0,5-4] arată că în studiul nostru am obținut $RR = 2$, însă în realitate acesta poate fi de la 0,5, adică expunerea scade riscul la jumătate, și până la 4, adică expunerea crește riscul de 4 ori, așadar în urma studiului nostru nu am demonstrat că factorul respectiv ar putea fi factor de risc.

Atunci când este vorba despre diferențe, intervalul de încredere nu trebuie să conțină valoarea 0. De exemplu, dacă în urma unui studiu, un medicament scade tensiunea arterială în medie cu 4 mmHg, CI95% [-2; 10] înseamnă că efectul medicamentului este de la creșterea tensiunii cu 2 mmHg, până la scăderea ei cu 10 mmHg, deci în urma studiului nu putem ști dacă medicamentul are vreun efect. Dacă CI95% era [1; 10] am fi spus că medicamentul scade tensiunea arterială, în cel mai rău caz cu doar 1mmHg, în cel mai bun caz cu 10 mmHg. La fel, reducerea relativă a riscului, reducerea absolută a riscului sunt diferențe și intervalul lor de încredere nu trebuie să conțină valoarea 0.

Intervalul de încredere ne dă și informații despre puterea statistică a studiului nostru: cu cât este mai larg, cu atât gradul de incertitudine este mai mare și este evident că puterea statistică a fost mai mică.

Așadar, deoarece intervalul de încredere ne dă informații mai multe decât p , se recomandă să îl furnizăm în orice studiu pe care îl facem, iar atunci când citim un studiu să ne uităm mai degrabă după el decât după p .

Intervalul de încredere poate fi calculat și în studiile descriptive, nu numai analitice. Iată de exemplu intervalul de încredere al unei proporții dintr-un studiu de prevalență – intenția de vot atunci când avem 2 candidați, X și Y: pentru X zic că ar vota 61% dintre cei chestionați, CI95% [53; 72%]; în aceste caz, intervalul de încredere nu trebuie să conțină valoarea 50%, cu care nu câștigă și nu pierde niciunul.

Atunci când nu avem semnificație statistică ($p > 0,05$, iar intervalul de încredere conține valoarea 1 pentru rapoarte sau 0 pentru diferențe) înseamnă fie că factorul respectiv este indiferent (nu este nici factor de risc, nici de protecție), fie că nu a avut studiul nostru puterea statistică să detecteze asociația dintre factorul respectiv și efect (s-a produs o eroare de tip II, beta).

6. Tipurile de variabile și testele statistice

Fie când faci un studiu, fie când îl evaluezi, cel mai greu este să-ți dai seama dacă acesta este corect (valid), sau există factori de confuzie și/sau erori sistematice (*bias*). Statistica este partea cea mai ușoară dintr-un studiu, pentru că nu ai nevoie decât să stabilești tipurile de variabile, să știi pentru fiecare tip de variabilă ce test statistic se potrivește, și să ai un soft statistic prietenos. Mai mult, dacă nu te pricepi la statistică poți oricând să rogi pe altcineva să te ajute după ce ai făcut studiul (deși în mod normal trebuie vorbit cu statisticianul încă din faza de proiect), pe când dacă nu ai făcut un studiu valid, sau nu te-ai gândit să măsoară o variabilă care ar putea fi factor de confuzie, după ce s-a terminat studiul nu se mai poate face nimic, iar studiul poate fi aruncat la gunoi.

În plus trebuie subliniat că statistica nu este decât o unealtă, o prelungire a gândirii biologice/medicale, și nu un scop în sine. Pentru a hotărî ce statistică trebuie făcută, trebuie să ai niște argumente și obiective clare, medicale, motiv pentru care încă din protocol și/sau aplicația pentru grant trebuie să arăți de la început care sunt obiectivele, care sunt variabilele măsurate și ce asociații epidemiologice vei testa și pentru ce factori de confuzie vei ajusta. Mai mult, încă din această fază trebuie să estimezi mărimea eșantionului (pe câți indivizi vei face studiul), pe baza mărimii efectului pe care o estimezi din alte studii din literatură sau dintr-un studiu pilot. Cu cât efectul pe care dorim să îl demonstrăm este mai mic, cu atât este nevoie de un eșantion mai mare, și din acest motiv studiile sunt pe eșantioane din ce în ce mai mari (mii și zeci de mii de indivizi, uneori chiar sute de mii, sau metaanalize pe sute de mii), pentru că efectele tratamentelor sunt foarte mici. Alt motiv pentru care analiza statistică trebuie hotărâtă înainte de începerea studiului, pe baza obiectivelor este că, dacă facem multe teste statistice ajungem la rezultate semnificative statistic din întâmplare (la fiecare 100 de testări, vom obține din întâmplare 5 $p=0,05$ și 1 $p=0,01$) (problema comparațiilor multiple), motiv pentru care analizele care nu au fost hotărâte înainte de începerea studiului (analize *post hoc*) nu au aceeași greutate, ci pot constitui ipoteze pentru studii viitoare.

Variabilele sunt în mare de trei feluri: 1) **numerice cantitative** (*scale*), măsurabile, de exemplu vârsta, greutatea, tensiunea arterială, hemoglobina, NT-proBNP etc; 2) **ordinale**, care sunt variabilele ce sunt clasificate în mai mult de două categorii și la care există o ordine naturală între categorii (de la valoarea cea mai mică la cea mai mare) - de exemplu evoluția bolii (agravat, staționar, ameliorat), stadializări (NYHA insuficiența cardiacă, TNM în cancer), scoruri de activitate sau de calitate a vieții etc; 3) **nominale**, ale căror valori nu pot fi ordonate din punct de vedere medical (culoarea ochilor, rasa, grupa sanguină, diagnosticul etc.) dintre care cele care pot lua doar două valori se numesc dihotomice/bimodale/binare (sexul masculin/feminin, mort/viu, a suferit/nu a suferit efectul, hipertensiv DA/NU. Din ultimul exemplu se vede cum variabilele cantitative sau ordinale pot fi transformate, pe baza unei valori de prag, în variabile nominale dihotomice).

În **Tabelul I** este prezentat modul de rezumare al variabilelor: pentru variabilele cantitative, dacă distribuția este gaussiană, populația se rezumă sub formă de medie și deviație standard; dacă distribuția nu este gaussiană, se rezumă sub formă de mediană și minim-maxim sau cvartila 25%-cvartila 75%. Variabilele ordinale se rezumă ca și cele cantitative cu distribuție nongaussiană (mediană și minim-maxim sau cvartilele 25 și 75%). În rarele cazuri în care variabila ordinală poate lua multe valori și distribuția este gaussiană, poate fi rezumată sub formă de medie și deviație standard. Variabilele nominale sunt rezumate sub formă de număr și procent.

Tabelul I. Rezumarea datelor

Tipul variabilei	Exemplu	Cancer DA	Cancer NU	P
Cantitativă (distribuție Gaussiană)	Vârsta (ani)	Media (SD)		<0,001
		65 (11)	55 (9)	
Cantitativă (distribuție nonGaussiană)	VSH (mm/h)	Mediana (min, max)		0,027
		Mediana (Q25%, Q75%)		
		76 (12, 157)	45 (12, 122)	
		76 (23, 119)	45 (19, 87)	
Ordinală (distribuție nonGaussiană)	Aprecierea mărimii scăderii ponderale (scală Lickert)	Mediana (min, max)		0,09
		Mediana (Q25%, Q75%)		
		3 (0, 4)	4 (0, 5)	
Nominală dihotomică	Sexul masculin	Nr (%)		0,03
		228 (65%)	115 (43%)	

SD=deviație standard; Q25%, Q75%=cvartilele 25 și 75%

În privința testelor statistice, atunci când variabilele au putut fi rezumate sub formă de medie și deviație standard, se aplică teste care compară acești parametri (parametrice, testul t Student); dacă ele nu s-au pretat la rezumarea sub formă de medie și deviație standard, atunci se aplică teste ordinale (*rank tests*), cel mai utilizat fiind Mann-Whitney U. Pentru compararea variabilelor nominale se folosește testul Chi pătrat, iar dacă numărul de indivizi este mic se folosește testul exact al lui Fisher. Pentru compararea a două curbe de supraviețuire se folosește testul log rank (o variantă de Chi pătrat care compară curbele stratificat pe fiecare treaptă, cu putere statistică mai mare decât Chi pătrat).

Atunci când comparăm mai mult de două grupuri, varianta parametrică este testul F (efectuăm analiza varianței, ANOVA), iar cea nonparametrică=ordinală este Kruskal-Wallis.

Pentru comparația valorilor împerecheate (cel mai des, pacientul comparat cu el însuși înainte-după), echivalentul împerecheat parametric este testul t (Student) împerecheat, iar cel nonparametric este testul Wilcoxon (**Figura 1** prezintă algoritmul alegerii testului statistic de aplicat funcție de situație).

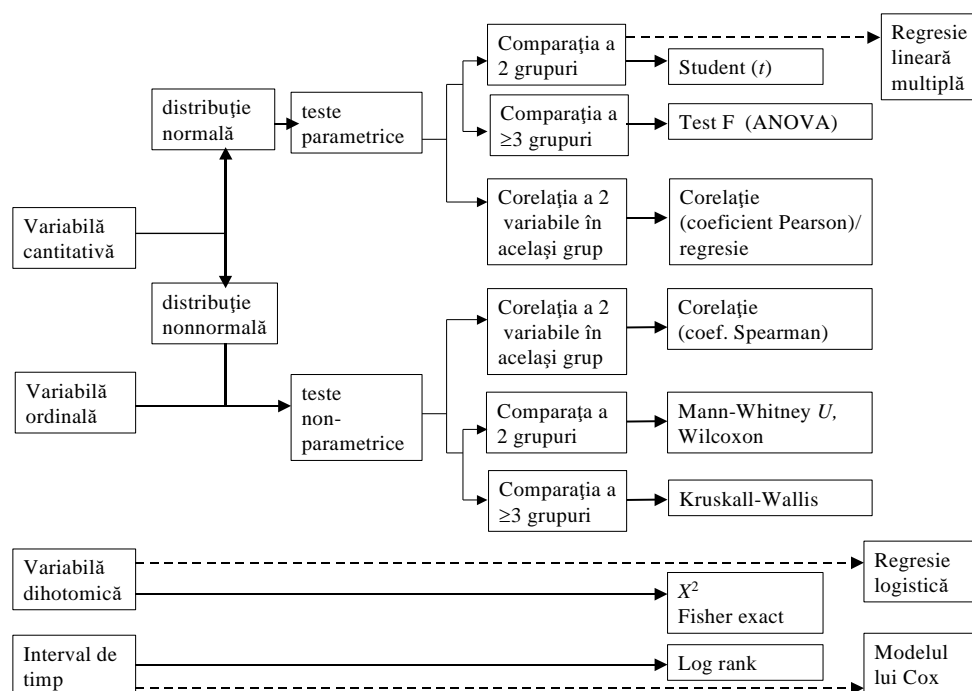


Figura 1. Algoritmul utilizării testelor statistice în funcție de variabile. (- - - - - = analiză multivariabilă)

Corelația se folosește atunci când vrem să vedem cum variază o variabilă cantitativă sau ordinală în funcție de cealaltă. Atunci când ambele variabile din corelație îndeplinesc criteriile de rezumare sub formă de medie și deviație standard se aplică corelația Pearson. Dacă măcar una dintre variabile nu respectă aceste criterii, se aplică corelația nonparametrică Spearman sau Kendall.

Din testele parametrice, pe lângă acel p rezultă și măsuri ale mărimii efectului (diferența dintre medii cu interval de încredere, coeficient de corelație r care ne arată cât de mare este corelația (peste 0,8 importantă, 0,5-0,8 moderată, 0,2-0,5 slabă, sub 0,2 neglijabilă). Din comparațiile nonparametrice nu rezultă decât p care ne arată dacă diferența este semnificativă statistic, fără măsuri ale efectului, cu excepția coeficientului de corelație Kendall τ_b , care ne arată mărimea corelației (4); coeficientul de corelație Spearman ρ nu ne arată nimic.

7. Diagnosticul

În practica clinică lucrăm continuu cu teste diagnostice, chiar dacă uneori o facem empiric și nu conștientizăm. Încă de când vedem prima oară bolnavul, evaluăm sexul și vârsta, și atunci unele posibilități diagnostice cresc în mintea noastră, iar altele scad. După care poate observăm ceva legat de facies, sau dispnee/tahipnee, sau mers etc. și din nou în mintea noastră are loc o re poziționare a probabilităților diagnostice, după care auzim motivele prezentării, facem anamneza, examenul clinic, apoi cerem teste diagnostice paraclinice, și cu fiecare etapă probabilitățile se deplasează în sus sau în jos, la sfârșit rămânând cât mai puține (de preferat unul) diagnostice cu probabilitate mare, după care urmează altele în ordinea scăderii probabilității. În principiu, atunci când un test diagnostic este pozitiv, el crește probabilitatea posttest, iar când este negativ, el scade această probabilitate, mai mult sau mai puțin, în funcție de cât de bun este acel test; în plus, probabilitatea posttest depinde și de probabilitatea pretest (**teorema lui Bayes**).

Cât de bun este un test diagnostic? Aceasta depinde de parametrii testului, care sunt **sensibilitatea**, **specificitatea** și **rapoartele de probabilitate** (*likelihood ratio*), iar pentru testele diagnostice care sunt variabile numerice cantitative, **aria de sub curba ROC**.

Sensibilitatea (S_n) este proporția de indivizi cu boala respectivă au testul pozitiv. De exemplu, dintre pacienții cu lupus eritematos sistemic, 98% au anticorpi antinucleari (ANA), așadar ANA are o sensibilitate de 98%, iar restul de 2% este proporția de fals negativi ($S_n + FN = 100\%$). Dacă S_n este foarte mare (99-100%), atunci proporția fals negativilor este infimă, și putem spune suntem

destul de siguri că un pacient cu testul negativ nu are boala (**SnNout**), indiferent de probabilitatea pretest. Așadar, un test cu Sn maximală ne este util când este negativ, pentru a putea exclude boala. **Specificitatea (Sp)** se calculează la indivizii fără boala respectivă, și este proporția de indivizi fără boală care au testul negativ, ceilalți, cu testul pozitiv, reprezentând rata fals pozitivilor ($Sp+FP=100\%$). Pentru un test cu Sp maximală (99-100%), rata fals pozitivilor este infimă, ceea ce înseamnă că dacă testul este pozitiv, putem fi destul de siguri că pacientul are boala (**SpPin**), indiferent de probabilitatea pretest. Dacă este vorba despre un semn sau simptom foarte specific, îi spunem patognomonic.

Curba ROC (*receiver operator characteristic*) este graficul având ca abscisă rata fals pozitivilor, iar pe ordonată sensibilitatea, pentru un test diagnostic variabilă numerică (de exemplu proNT-BNP pentru insuficiența cardiacă). Acuratețea testului este dată de aria de sub această curbă (AUROC), care poate lua valori între 0,500 și 1 – cu cât este mai mare, cu atât este testul mai bun. Curba ROC ne ajută să comparăm mai multe teste diagnostice pentru aceeași boală (prin intermediul ariei de sub curbă), să alegem o valoare de prag pentru o anumită sensibilitate și specificitate ale testului. Dacă aria de sub curbă a unui test diagnostic este bună ($>0,8$), atunci se poate alege o valoare de prag pentru care să avem în același timp și Sn, și Sp bune. Pentru același test, dacă modificăm valoarea de prag pentru a crește Sn vom pierde din Sp, și invers.

Raportul de probabilitate (*likelihood ratio, LR*) se calculează din sensibilitate și specificitate (5) și are câteva avantaje asupra acestora, dintre care cele mai importante sunt acelea că pot fi calculate pe mai multe nivele ale valorii testului (și nu doar pe două, cum se întâmplă cu Sn și Sp, pentru care testul poate fi doar pozitiv sau negativ), și că știind probabilitatea pretest, este ușor de calculat probabilitatea posttest. Ca și Sn și Sp, nu depind de prevalență (probabilitatea pretest).

Astfel, dacă un test are $LR=1$, el nu modifică probabilitatea; pentru testele având $LR+$ de 2, 5, respectiv 10, pozitivitatea lor crește în mod absolut probabilitatea posttest cu 15, 30, respectiv 45% în valoare absolută. Pentru testele cu $lr-$ (LR atunci când testul este negativ) de 0,5, 0,2 respectiv 0,1, negativitatea lor scade probabilitatea posttest cu 15, 30 respectiv 45%.

De exemplu, dacă prevalența unei boli este de 2%, înseamnă că probabilitatea oricărui individ de a avea boala este de 2%; dacă avem testul X pentru boala respectivă, cu $LR+=2$, și $lr-=0,1$, înseamnă că dacă unui individ din acea populație îi iese testul pozitiv, probabilitatea ca el să aibă boala este de $2\%+15\%=17\%$; dacă testul este negativ, probabilitatea ca totuși să aibă boala este

2% minus 45%, adică sigur nu are boala. Dacă, din anamneză și examenul clinic știm despre individ niște date care duc probabilitatea pretest la 50%, atunci dacă testul este pozitiv probabilitatea posttest este $50\%+15\%=65\%$, adică am mai avea nevoie de un test suplimentar pentru a fi mai siguri de existența bolii, iar dacă testul este negativ $50\%-45\%=5\%$. Acest exemplu nu face decât să sublinieze faptul că probabilitatea posttest depinde de probabilitatea pretest (este direct proporțională) (teorema lui Bayes). Probabilitățile mai exacte se calculează de fapt înmulțind cota (*odds*) pretest cu LR, și apoi transformând cota posttest în probabilitate [cota este probabilitatea de a suferi efectul raportată la probabilitatea de a nu suferi efectul, $p/(1-p)$, iar probabilitatea este $cota/(cota+1)$].

Valorile predictive (pozitivă și negativă, VPP, VPN) sunt probabilitatea ca pacientul respectiv să aibă (să nu aibă) boala, atunci când testul este pozitiv (negativ) (practic, sunt probabilitățile posttest). Pentru aceeași valoare a testului diagnostic, după cum am spus, valorile predictive depind de probabilitatea posttest, și atunci, dacă putem folosi S_n (S_nNout), S_p (S_pPin) sau LR ale unui test, dintr-un studiu, pentru a calcula probabilitatea posttest în funcție de prevalența bolii din locul unde avem noi pacientul, valorile predictive nu pot fi translatate în locuri cu prevalențe diferite (zone geografice cu prevalențe diferite, sau pacienți cu caracteristici diferite, sau din îngrijirea terțiară unde prevalențele sunt cele mai mari în îngrijirea primară unde prevalențele sunt cele mai mici).

Pentru fiecare dintre acești parametri (S_n , S_p , LR, VPP, VPN, AUROC), în studii sunt calculate intervale de încredere care ne estimează valorile lor în realitate, și ne dau informații și cu privire la puterea statistică a studiului (care a fost mică, dacă ele sunt prea largi).

8. Sinteza sistematică și metaanaliza

Sinteza sistematică (*systematic review*) este o sinteză a studiilor care răspund la o anumită problemă, care are o metodologie explicită și reproductibilă privind obiectivele, materialul și metodele. Practic, protocolul este similar oricărui studiu original, numai că acest studiu se desfășoară în baze de date care conțin articole originale (cercetare secundară), și nu direct pe pacienți.

Protocolul trebuie să conțină întrebarea clinică de la care se pornește, de cele mai multe ori în format PICO (pacienți/populație, intervenție, comparator, efect=*outcome*). După care trebuie

menționate tipurile de studii în care vom căuta răspunsul, și bazele de date în care se va efectua căutarea (căutarea trebuie să fie exhaustivă - cu cât mai multe baze de date, cu atât mai bine - însă ele să fie măcar Medline, Embase și registrul Cochrane de studii clinice randomizate. Bineînțeles că trebuie detaliată strategia de căutare, astfel încât oricine vrea să reproducă cercetarea, să ajungă la aceleași rezultate. După care trebuie menționate criteriile de evaluare a validității studiilor găsite, criteriile de alegere a studiilor care vor intra în analiza statistică (**metaanaliza**), cum se va evalua dacă nu există o eroare sistematică de publicare (se știe că studiile pozitive sunt mai degrabă publicate, și atunci oricine va face o sinteză sistematică va găsi doar studii pozitive, iar concluzia va fi că tratamentul este eficient, pentru că studiile “negative” nu au fost publicate), acum se va căuta heterogenitatea (adică dacă sunt diferențe mari între studii în ceea ce privește efectul, ceea ce nu este normal, pentru că efectul real al unui tratament sau factor de risc este unul singur...), ce metode statistice vor fi utilizate pentru metaanaliză etc.

Scopul sintezei sistematice este acela de a decela adevărul cu privire la o terapie, sau o etiologie, sau un prognostic, sau o metodă diagnostică, adevăr pe care nu l-am putut deduce din studii individuale pentru că fie rezultatele lor au fost diferite, fie din cauza puterii statistice mici nu s-a obținut semnificație statistică (punând toate studiile la un loc pentru metaanaliză, puterea statistică crește), fie vrem o mai mare precizie prin îngustarea intervalelor de încredere (tot prin creșterea puterii statistice). Prin evaluarea studiilor individuale găsite, sinteza sistematică poate ajunge și la concluzia că nu ne putem bizui pe rezultatele de până acum, pentru că studiile au fost de proastă calitate, sau pur și simplu că nu există studii care să răspundă întrebări puse.

Concluzionând, atunci când citim un studiu ne gândim de la bun început care este întrebarea de cercetare la care studiul va răspunde, ce tip de studiu s-a făcut știind că studiile sunt diferite ca și validitate, și dacă statistica a fost corectă. Apoi evaluăm mărimea efectului cu intervalele de încredere.

Bibliografie

1. Motulsky H. Intuitive biostatistics. Oxford University Press, New York, 1995.
2. GRADE Working Group. Grading quality of evidence and strength of recommendations. BMJ. 2004, 328: 1490

3. Baicus C, Caraiola S. Effect measure for quantitative endpoints: statistical versus clinical significance, or „How large the scale is?”. Eur J Intern Med.2009. 20:e124-5.
4. Peacock J, Kerry S. Presenting medical statistics from proposal to publication. A step by step guide. Oxford University Press, Oxford, 2008.
5. Băicuș C. Medicina bazată pe dovezi. Cum înțelegem studiile. Editura Medicală, București, 2007.