

ANALIZA STATISTICĂ – Indicatori statistici ai stării de sănătate

Consider că este deosebit de importantă prezentarea schematică a clasificării tipurilor de indicatori epidemiologici, precum și a tipurilor de studii utilizate în epidemiologia clinică, fapt ce permite justificarea protocolului statistic utilizat în prezenta lucrare.

INDICATORI UTILIZAȚI ÎN STUDIILE EPIDEMIOLOGICE

În studiile epidemiologice sunt folosiți, în general, 4 tipuri de indicatori:

- I. Diverse tipuri de **raporturi, proporții și rate**
 1. Raporturi
 2. Proporții
 3. Rate
- II. **Indicatori ai frecvenței bolii** – utilizați pentru a da o măsură a frecvenței bolii sau a numărului de cazuri dintr-o populație
 1. **Prevalența**
 - i. Prevalența la un moment dat
 - ii. Prevalența de-a lungul unei perioade de timp
 - iii. Prevalența de-a lungul vieții
 2. **Incidența**
 - i. Incidența cumulativă
 - ii. Rata incidenței (sau densitatea incidenței)
- III. **Indicatori ai asocierii factori de risc (expunere)-boală**
 1. **Riscul relativ (RR)**
 2. **Rata de șansă (OR –Odds ratio)**
 3. **Rata standardizată a mortalității**
 4. **Rata proporțională a mortalității**
- IV. **Indicatori ai diferenței:**
 1. **Riscul atribuit/populație**

I. RATE ȘI PROPORȚII

1. Raportul - este folosit pentru a compara două cantități (de exemplu **numărul de femei/ numărul de bărbați** dintr-o localitate = $900/802 = 1,122$)

2. Proporția – un tip special de raport, în care numărătorul este inclus și în numitor, rezultând de fapt un procentaj (pentru exemplul anterior procentul femei în localitatea respectivă este **numărul de femei/numărul total de locuitori** (femei + bărbați) din localitatea respectivă = $900/1702 = 52,88 \%$)

3. Rata – un tip special de proporție ce include anumite specificații legate de timp; este unul dintre cei mai utilizați indicatori în epidemiologie deoarece poate exprima clar probabilitatea sau riscul unei boli sau a altor evenimente în cadrul populației studiate pentru a anumită perioadă de timp.

Formula de calcul:

$$\frac{\text{numărul de evenimente în perioada specifică de timp}}{\text{populația expusă la risc în perioada specifică}} * k$$

unde k – folosit pentru unități de populație (cum ar fi 1 000 sau 100 000)

Exemplu: dacă într-un oraș de 120 000 de locuitori într-un an au fost diagnosticate 11 cazuri de leucemie acută (LA) vom avea o rată a îmbolnăvirilor de LA egală cu

$$\frac{11}{120000} * 100000 = 9,166 \text{ la } 100000 \text{ locuitori}$$

II. INDICATORI AI FRECVENȚEI BOLII

1. Prevalența – **proporția de indivizi dintr-o anumită populație ce sunt bolnavi de o anumită boală la un anumit moment în timp** (nu contează faptul că sunt bolnavi diagnosticați recent sau în trecut)

- permite estimarea probabilității sau riscului ca cineva din populația respectivă să devină bolnav la un anumit moment dat.

- dă o imagine a severității problemei respective

- este un indicator folositor pentru planificarea serviciilor de sănătate (personal, aparatură etc)

Formula de calcul:

$$\frac{\text{numărul de cazuri existente într-o anumită boală}}{\text{Totalul populației expuse la risc}} \text{ la un anumit moment în timp}$$

-poate fi și ea exprimată sub forma: la 1000 sau 100000 locuitori etc

Exemplu: dacă într-un oraș de 120 000 de locuitori la un moment dat există 70 de cazuri de leucemie acută (LA) vom avea o prevalență de

$$\frac{70}{120000} = 0,0005833 \text{ sau } 58,33 \text{ de cazuri la } 100000 \text{ locuitori}$$

Tipuri de prevalență:

- *Prevalența la un moment dat* – numărul de cazuri ce există într-o populație la un moment dat
- *Prevalența de-a lungul unei perioade de timp* - numărul de cazuri ce există într-o populație de-a lungul unei perioade specificate de timp
- *Prevalența de-a lungul vieții* – procentul dintr-o populație existentă ce a prezentat un istoric al unei anumite boli într-un moment al vieții

2. Incidența – reprezintă un indicator (o măsură) a cazurilor noi dintr-o anumită boală ce au apărut în populația studiată de-a lungul unei perioade de timp

- este un indicator al probabilității care există ca o persoană neafectată să dezvolte respectiva boală

Practic este folosită **incidența cumulativă – proporția de indivizi ce s-au îmbolnăvit într-o perioadă specificată de timp.**

Perioada de timp poate fi un an calendaristic, 6 luni, 3 ani, 5 ani etc.

Formula de calcul:

$$IC = \frac{\text{numarul de cazuri NOI dintr-o anumită boală de-a lungul unei perioade de timp}}{\text{Totalul populației expuse la risc}}$$

-poate fi și ea exprimată sub forma: la 1000 sau 100000 locuitori etc

Exemplu: dacă într-un oraș de 120 000 de locuitori la un moment dat există 30 de cazuri noi de leucemie acută (LA) în timp de un an, vom avea o incidență de

$$IC = \frac{30 \text{ de cazuri noi de LA}}{120000} = 0,00025 \quad \text{sau } 25 \text{ de cazuri la } 100000$$

locuitori

3. Rata de incidență (densitatea incidenței) – un indicator mai puțin folosit, datorită costurilor relativ mari legate de urmărirea cazurilor de-a lungul unei perioade de timp.

- este o măsură a incidenței capabilă să dea informații chiar dacă respectivele cazuri sunt urmărite pe perioade de timp inegale (datorită decesului sau migrației unor persoane)

- în formula de calcul a indicatorului, similară celei a incidenței, numitorul este o sumă a așa numitului **persoana-timp la risc**

III. INDICATORI AI ASOCIERII FACTORI DE RISC (EXPUNERE)- BOALĂ

- sunt calcule a căror rezultate sunt menite să dea o măsură a frecvenței bolii în funcție de existența altor factori care ar putea influența într-un anumit fel incidența bolii în sânul populației studiate

Extrem de importante în calculul acestor indicatori sunt așa numitele **tabele de contingență 2x2** (sau simplu tabele 2x2).

Acestea sunt folosite pentru:

- a ilustra frecvența bolii, respectiv a expunerii
- calculul asocierii dintre expunerea la factorul (de risc) studiat și boală

Pentru a putea fi folosite în determinarea unor indicatori epidemiologici tabelele de contingență de tip 2x2 trebuie să grupeze corect indivizii din populația studiată, ținând cont de expunere (liniile tabelului), respectiv de starea de boală (coloanele tabelului), cum se poate vedea în exemplul ce urmează.

Tabel VII. Tabel de contingență de tip 2x2

		Boală		Total
		Da	Nu	
Expunere	Da	<i>a</i>	<i>b</i>	$a + b$
	Nu	<i>c</i>	<i>d</i>	$c + d$
Total		$a + c$	$b + d$	$a + b + c + d$

unde:

a = numărul indivizilor din cadrul populației studiate **expuși** (care au suferit o expunere la factorul de risc studiat) și în același timp **bolnavi** (evident, cei suferinzi din cauza bolii studiate)

b = numărul indivizilor din cadrul populației studiate **expuși** dar **sănătoși**

c = numărul indivizilor **neexpuși** (care nu au suferit o expunere la factorul de risc studiat) și în același timp **bolnavi**

d = numărul indivizilor **neexpuși** și **sănătoși**

a + b = **numărul total al indivizilor** (bolnavi + sănătoși) din cadrul populației studiate **expuși** la factorul de risc

c + d = **numărul total indivizilor** (bolnavi + sănătoși) din cadrul populației studiate **neexpuși** la factorul de risc

a + c = **numărul total al indivizilor bolnavi** din cadrul populației studiate

b + d = **numărul total al indivizilor sănătoși** din cadrul populației studiate

a + b + c + d = Mărimea totală a populației studiate (eșantionului), practic suma tuturor celor patru celule ale tabelului

Există 2 indicatori epidemiologici foarte importanți pentru măsurarea unei asocieri expunere (factor de risc)-boală:

- **Riscul relativ** (Relative Risk - RR)
- **Rata de șansă** (Odds Ratio –OR)

Ambii indicatori sunt calculați *folosind gruparea datelor în tabele de contingență de tip 2x2*.

Riscul relativ (Relative Risk - RR)

- servește la măsurarea magnitudinii asocierii între incidența bolii studiate și factorul de risc cercetat
- reprezintă în fapt un raport dintre rata incidenței persoanelor expuse și rata incidenței în rândul celor neexpuși la factorul de risc
- folosit în cercetările legate de etiologia unor boli
- sinonime: Rata de risc (Risk ratio)

Formula de calcul:

$$RR = \frac{\text{Incidența bolii în rândul subiecților expuși la factorul de risc}}{\text{Incidența bolii în rândul subiecților neexpuși la factorul de risc}}$$

sau ținând cont de modul de grupare a datelor într-un tabel de contingență de tip 2x2:

$$RR = \frac{a/(a + b)}{c/(c + d)}$$

Interpretarea valorilor riscului relativ:

- Riscul relativ poate avea, teoretic vorbind, valori între 0 și infinit
- $RR = 1$: Nu există nici un fel de asociere între expunere (factorul de risc) și boală; practic incidența (rata incidenței) este identică între cele două grupuri (expuși, respectiv neexpuși la factorul de risc)
- $RR > 1$ = Există o asociere pozitivă între expunere (factorul de risc) și boală
- $RR < 1$ = Există o asociere negativă între expunere (factorul de risc) și boală sau, în unele cazuri putem spune că expunerea la factorul respectiv are chiar un “efect protector” față de boală

Rata de șansă (Odds Ratio –OR)

- **Singura modalitate de măsurarea a asocierii expunere (factor de risc)-boală disponibilă în cazul studiilor de tip Case-Control**

Dă o măsură a riscului relativ (are valori apropiate de acesta) când:

- valoarea incidenței este mică ($< 5\%$ din populația totală) și
- grupul de control (martor) este reprezentativ pentru întreaga populație (inclusiv din punctul de vedere al expunerii la factorul de risc)

Trebuie definită în doi pași:

- **Șansă (ODDS)** – Raportul dintre probabilitatea ca un eveniment să se întâmple și probabilitatea ca respectivul eveniment să nu aibă loc

- **Rata de șansă** (Odds Ratio –OR) – Raportul dintre rata (șansa) expunerii în rândul celor bolnavi și rata (șansa) expunerii în rândul celor sănătoși

Formula de calcul:

- ținând cont de modul de grupare a datelor într-un tabel de contingență de tip 2x2, formula de calcul este:

$$OR = \frac{a/c}{b/d} = \frac{a * d}{b * c}$$

Interpretarea valorilor ratei de șansă (OR):

- Rata de șansă (OR) poate avea, teoretic vorbind, valori între 0 și infinit

OR = 1 : Nu există nici un fel de asociere între expunere (factorul de risc) și boală;

OR > 1 : Expunerea reprezintă un factor de risc în cazul bolii respective (crește riscul apariției bolii)

OR < 1 : Expunerea reprezintă un factor „protector” în cazul bolii respective (scade riscul apariției bolii)

Interpretarea rezultatelor analizei statistice. Interpretarea valorii lui RR (OR) și P în cazul studiilor ce utilizează tabele de contingență de tip 2x2

Interpretarea existenței unei asocieri între factorul de risc și apariția bolii, în cazul unui studiu retrospectiv de tip cohortă, necesită o analiză atentă, ținând cont de următorii indicatori **(13, 145)** :

1. **valoarea lui P (cel mai important parametru)**
2. **valoarea minimă și maximă** a intervalului de încredere, calculat pentru o probabilitate de 95 % (CI 95 %)
3. **valoarea ratei de șansă (Odds ratio-OR)**

Într-un astfel de studiu ne interesează atât valoarea lui P (cel mai important parametru cât și intervalul de încredere pentru un anumită rată de șansă (OR) sau risc relativ (RR).

Valoarea lui P ne răspunde la următoarea întrebare: Dacă, de exemplu, nu există nici o asociere între variabilele care definesc coloanele și variabilele care definesc liniile unui tabel de contingență (practic în cazul studiului nostru între expunere și boală), care este șansa (probabilitatea) ca luând la întâmplare indivizi (așa numita *eșantionare aleatorie - random sampling*) dintr-o populație să descoperim o asociere (mai) semnificativă decât cea din studiul în cauză **(145)**? Cu alte cuvinte,

dacă într-adevăr nu există deloc nici o asociație între linii și coloane care este probabilitatea ca o eșantionare aleatorie să ducă la obținerea unei valori a OR (foarte) diferită de 1,0 precum și diferită de cea obținută în studiul în cauză ?

Pentru a explica mai bine cele afirmate mai sus, să luăm un exemplu:

Să presupunem că pentru studiul nostru asupra influenței radiațiilor ionizante asupra incidenței leucemiilor acute (LA) am obținut o valoare a **OR = 0,923** și o valoare a lui **P = 0,991**. Conform celor afirmate mai sus, aceasta valoare a lui P ne spune că printr-o eșantionare aleatorie, la modul teoretic, avem **99,1 %** șanse de a obține un OR diferit atât de valoarea obținută (0,923) de noi cât și de 1,0. De aceea putem afirma că ceea ce am obținut noi în studiu este *statistic nesemnificativ* dacă **P = 0,991**. Invers, dacă din analiza statistică ar rezulta o valoare a **OR = 4,204** și o valoare a lui **P < 0,0001**, aceasta ar însemna că avem mai puțin de 0,0001 % șanse de a obține, prin eșantionare aleatorie (alegând deci la întâmplare subiecți) un OR mult diferit atât de valoarea obținută (4,204) de noi cât și de 1,0. De aceea putem afirma că ceea ce am obținut noi în studiu este *statistic semnificativ* dacă **P < 0,0001**.

Din nefericire, termenul "*semnificativ statistic*" nu este de multe ori același lucru cu "*important din punct de vedere științific*".

Înainte de a interpreta din punct de vedere științific o anumită valoare a lui P sau un anumit interval de încredere trebuie să ne gândim atent la ordinul de mărime a ratei de risc (RR) sau ratei de șansă (OR). Cât de mare trebuie să fie această valoare a lui OR pentru a o considera importantă din punct de vedere științific (chiar dacă am obținut un rezultat considerat statistic semnificativ)? Cât de mică trebuie să fie acea valoare pentru a o putea considera "normală" și a putea trage concluzia că expunerea studiată nu poate fi considerată factor de risc în boala respectivă ?

Modul în care interpretăm rezultatele unui studiu depinde foarte mult de valoarea lui P: mică (**P < 0,05**) sau mare (**P > 0,05**).

A. Dacă valoarea lui P este mică (P < 0,05) = rezultat semnificativ statistic:

În acest caz este improbabil că asocierea expunere-boală observată să fie întâmplătoare. Putem deci respinge ideea că asocierea este o coincidență datorată eșantionării și putem concluziona că populația studiată are un RR sau OR diferit de 1,0. Asocierea este deci statistic semnificativă (dacă **P < 0,05**). Dar este acest rezultat important și din punct de vedere științific?

Intervalul de încredere (CI 95 %) este cel care ne poate da un răspuns.

Orice analiză statistică include întotdeauna, în proporție mai mică sau mai mare, o serie de erori datorate eșantionării (așa numitele effects of random sampling), astfel că valoarea *RR* sau *OR* calculată din datele studiului probabil nu este identică cu valoarea “adevărată” a *RR* sau *OR*. Dar, din nefericire, nu există posibilitatea determinării exacte a acestei valori “adevărate”. În schimb analiza statistică ne dă o măsură a acestei valori prin intermediul intervalului de încredere la 95 %. *Practic putem fi 95 % siguri că acest interval conține valoarea “adevărată” a RR sau OR.*

Pentru a putea interpreta rezultatele analizei statistice într-un context științific pertinent, trebuie să privim la ambele capete ale intervalului de încredere și să ne întrebăm dacă acestea reprezintă valori ce au importanță din punct de vedere științific sau sunt valori obișnuite, comune, fără importanță științifică.

Tabelul VIII sistematizează variantele posibile:

Tabel VIII. Interpretarea rezultatelor analizei statistice

Valoarea minimă a CI 95 %	Valoarea maximă CI 95 %	Interpretare
Comună (relativ apropiată de 1)	Comună (relativ apropiată de 1)	Chiar dacă <i>RR</i> sau <i>OR</i> nu este 1,0 , asocierea expunere-boală este mică și relativ lipsită de interes științific , chiar dacă este statistic semnificativă
Comună (relativ apropiată de 1)	Importantă (mult mai mare decât 1)	Deoarece CI 95 % variază între o valoare a <i>OR</i> comună din punct de vedere biologic (științific) și o valoare semnificativă, nu putem trage o concluzie certă cu privire la importanța științifică a rezultatului. <i>La modul ideal, o nouă serie de date ar trebui analizate pentru a putea avea o concluzie certă, atât din punct de vedere statistic cât și științific.</i>
Importantă (mult mai mică decât 1)	Comună (relativ apropiată de 1)	Deoarece CI 95 % variază între o valoare semnificativă și o valoare a <i>OR</i> comună din punct de vedere biologic (științific), nu putem trage o concluzie certă cu privire la importanța științifică a rezultatului. Apare o problemă legată însă de designul studiului nostru , expunerea pare să influențeze în sens negativ incidența bolii, putând fi considerată chiar ca un “factor de protecție” (13). <i>O nouă serie de date ar trebui analizate pentru a putea avea o concluzie certă, atât din punct</i>

		<i>de vedere statistic cât și științific, iar dacă rezultatele sunt similare trebuie verificat modul în care am conceput designul studiului și gruparea datelor în tabelele de contingență.</i>
Importantă	Importantă	De vreme ce până și valoarea minimă a CI 95% reprezintă o asociere expunere-boală suficient de mare pentru a fi considerată biologic importantă, se poate trage concluzia indubitabilă că există o asociere expunere-boală semnificativă atât din punct de vedere statistic, cât și științific.

B. Dacă valoare lui P este mare ($P > 0,05$) = rezultat nesemnificativ statistic:

Dacă valoarea lui P este mare, nu avem nici un motiv plauzibil să concluzionăm că RR sau OR diferă de 1,0. Aceasta **nu înseamnă însă că putem spune că RR sau OR sunt egale cu 1,0. Pur și simplu nu avem dovezi consistente că valorile RR sau OR diferă de 1,0.**

Care este oare valoarea “adevărată” a riscului relativ sau a ratei de șansă ?

Cum datele noastre includ o serie de erori datorate eșantionării (effects of random sampling) valoarea RR sau OR calculată din datele studiului probabil nu este identică cu valoarea “adevărată” a RR sau OR. Neexistând posibilitatea determinării exacte a acestei valori “adevărate”, analiza statistică ne dă o măsură a acestei valori prin intermediul intervalului de încredere la 95 %. Practic putem fi 95 % siguri că acest interval conține valoarea “adevărată” a RR sau OR.

Când însă $P > 0,05$, **intervalul de încredere este mult lărgit**: el include așa numita ipoteză de nul (null hypothesis – OR sau RR = 1,0) și se întinde pe o gamă de valori care ar putea sugera **de la existența unei asocieri negative** expunere-boală (valoarea minimă a CI 95 % este RR <1,0 sau OR <1,0) până la **existența unei asocieri pozitive** expunere-boală (valoarea maximă a CI 95 % este RR >1,0 sau OR >1,0)

Pentru a putea interpreta rezultatele analizei statistice într-un context științific pertinent, trebuie să privim la ambele capete ale intervalului de încredere, situațiile ce pot fi întâlnite fiind prezentate schematic în tabelul IX:

Tabel IX. Interpretarea rezultatelor analizei statistice

Valoarea minimă a CI 95 %	Valoarea maximă CI 95 %	Interpretare
Comună (relativ apropiată de 1)	Comună (relativ apropiată de 1)	Nu există o asociere semnificativă statistic între expunere-boală. Cu alte cuvinte nu există o nici asociere între liniile și coloanele tabelului de contingență sau dacă există ea este cu certitudine foarte mică și lipsită de interes. Un rezultat ce poate fi considerat semnificativ și din punct de vedere științific.
Comună (relativ apropiată de 1)	Diferită de 1 (mai mare decât 1)	Nu se poate trage o concluzie foarte clară. Deși analiza statistică demonstrează inexistența unei asocieri statistice semnificative , expunerea la factorul de risc duce la apariția unui interval de încredere ce se întinde de la o valoare minimă specifică unei slabe asocieri negative expunere-boală, până la o valoare maximă specifică unei asocieri pozitive puternice expunere-boală; Cât de apropiată de 0,05 este valoarea lui P ? <i>Pentru a putea trage concluzii clare este necesară repetarea experimentului folosind mai mulți subiecți.</i>
Diferită de 1 (mai mică decât 1)	Comună (relativ apropiată de 1)	Nu se poate trage o concluzie foarte clară. Deși analiza statistică demonstrează inexistența unei asocieri statistice semnificative , expunerea la factorul de risc duce la apariția unui interval de încredere ce se întinde de la o valoare minimă specifică unei puternice asocieri negative expunere-boală, până la o valoare maximă specifică unei slabe asocieri pozitive expunere-boală; Cât de apropiată de 0,05 este valoarea lui P ? Dacă P are o valoare relativ apropiată de 0,05 – atenție la designul studiului ! <i>Pentru a putea trage concluzii clare este necesară repetarea experimentului folosind mai mulți subiecți.</i>

Modul în care trebuie interpretate rezultatele analizei statistice în cazul studiilor ce utilizează tabele de contingență de tip 2x2 este cel de mai sus, clar prezentat într-un număr mare de surse bibliografice (13, 145, 241, 242) din care aș

dori să menționez, pentru relevanța și claritatea expunerii, cartea - **The InStat Guide to Choosing and Interpreting Statistical Tests**, © 1990-1998, GraphPad Software (145):

- **valoarea lui P** ca prim parametru, din care rezultă **semnificația statistică** a analizei
- **valorile CI 95 %** pentru **explicitarea sensului și magnitudinii asocierii**, din care rezultă **semnificația științifică a analizei**

Cu toate acestea, o serie de articole apărute în ultimii ani în prestigioase reviste internaționale (**7, 93, 153**) pe teme similare prezentului studiu, prezintă valorile OR sau RR fără a preciza valoarea lui P, ci numai intervalul de încredere la 95 % (CI 95 %).

Aceasta situație m-a pus într-o dilemă importantă, având de ales între două opțiuni:

- interpretarea rezultatelor analizei statistice după cele mai stricte reguli, conform lucrării **The InStat Guide to Choosing and Interpreting Statistical Tests** și celorlalte surse bibliografice, *ceea ce ar fi pus o umbră de îndoială asupra comparației dintre rezultatele studiile anterior menționate și rezultatele acestei cercetări.*
- interpretarea rezultatelor analizei statistice conform articolelor ce prezintă studii similare celui efectuat de mine (OR sau RR, CI fără prezentare valorii lui P), *ceea ce ar fi pus o umbră de îndoială asupra semnificației statistice reale a rezultatelor obținute de mine.*

Din fericire, sper că am reușit să găsesc o cale de mijloc, dată fiind și o particularitatea a studiului meu, care a necesitat împărțirea datelor primare pe intervale de studiu (6 intervale, unul ca **martor**, caracterizat prin **lipsa expunerii - 1983-1985**, respectiv 5 **intervale post-expunere** 1986-1988, 1989-1991, 1992-1994, 1995-1997, 1997-2000).

Evident, este puțin probabil (dar nu imposibil) ca o asocierie expunere-boală observată să fie statistic semnificativă **simultan** pe toate aceste intervale și în același timp cu valori minime și maxime ale CI mai mari decât 1,0 (dacă ar exista o astfel de situație am putea spune că există o asocierie expunere-boală **semnificativă** atât din punct de vedere **statistic** cât și **biologic** sau **științific**).

De aceea am luat în considerare din punct de vedere al concluziilor certe doar valorile cu $P < 0,05$, dar am realizat și o cuantificare a situațiilor în care valoarea lui P este apropiată de 0,05.

Aceasta deoarece pot fi întâlnite patru situații distincte:

I. În cazul existenței unei asocieri semnificative statistic ($P < 0,05$):

Cazul 1:

- valoarea OR (RR) > 1
- valoarea minimă a CI 95 % > 1 ; valoarea maximă a CI 95 % > 1 (de obicei mult mai mare decât 1)
- evident, $P < 0,05$
- **Interpretare:** În acest caz putem spune că există o asociere **POZITIVĂ, SEMNIFICATIVĂ STATISTIC**, între factorul de risc și boala studiată.

Cazul 2:

- valoarea OR (RR) < 1
- valoarea minimă a CI 95 % < 1 (de obicei mult mai mic decât 1); valoarea maximă a CI 95 % < 1
- $P < 0,05$

Interpretare: În acest caz putem spune că există o asociere **NEGATIVĂ, SEMNIFICATIVĂ STATISTIC**, între factorul de risc și boala studiată. Cu alte cuvinte, în funcție de expunerea studiată, unele surse bibliografice susțin că, în acest caz, putem afirma că expunerea nu numai că nu influențează apariția bolii, ci are chiar o acțiune “protectoare”. (13)

II. În cazul inexistenței unei asocieri semnificative statistic ($P > 0,05$):

Cazul 3:

- $P > 0,05$, dar valoarea lui P este mică (în aceasta lucrare am considerat acest caz pentru $P < 0,1$)
- valoarea OR (RR) > 1

- valoarea minimă a CI 95 % < 1; valoarea maximă a CI 95 % > 1

Interpretare: În acest caz putem spune că **poate fi observată o asociere POZITIVĂ, NESEMNICATIVĂ STATISTIC**, între factorul de risc și boala studiată.

Cazul 4:

- $P > 0,05$, dar valoarea lui P este mică (în aceasta lucrare am considerat acest caz pentru $P < 0,1$)
- valoarea OR (RR) < 1
- valoarea minimă a CI 95 % < 1; valoarea maximă a CI 95 % > 1 (de obicei apropiată de 1)

Interpretare: În acest caz putem spune că **poate fi observată o asociere NEGATIVĂ, NESEMNICATIVĂ STATISTIC**, între factorul de risc și boala studiată.

Există și o a cincea situație, cea în care $OR (RR) \cong 1$, ceea ce înseamnă între factorul de risc (expunere) și boală **nu poate fi făcută nici un fel de asociere**, factorul de risc (expunerea) nu are nici un fel de influență în cazul studiat (în această lucrare am considerat ca validă o astfel de situație pentru $0,95 < OR < 1,05$).

Practic, după cum se poate vedea, valoarea OR, respectiv CI 95 %, ne furnizează informații despre **sensul influenței factorului de risc (expunerii) asupra bolii** (în mod evident o valoare a OR egală cu 1, însemnând faptul că factorul respectiv nu influențează în nici un fel boala), în timp ce **valoarea lui P** ne dă informații despre **semnificația statistică** a acestei influențe ($P < 0,05$ – **asocierea descoperită este semnificativă statistic**).

Cele prezentate mai sus sunt sistematizate în tabelul X:

Tabel X. Interpretarea rezultatelor statistice privind asocierea factor de risc-boală

ASOCIERE factor de risc - boală STATISTIC SEMNIFICATIVĂ				
OR (RR)	Val. minimă CI 95 %	Val. maximă CI 95 %	P	Interpretare
> 1	> 1	>> 1	< 0,05	Există o asociere POZITIVĂ, SEMNIFICATIVĂ STATISTIC între factorul de risc și boala studiată
Asociere pozitivă			Statistic semnificativ	
OR (RR)	Val. minimă CI 95 %	Val. maximă CI 95 %	P	Interpretare
< 1	<<1	< 1	< 0,05	Există o asociere NEGATIVĂ, SEMNIFICATIVĂ STATISTIC între factorul de risc și boala studiată
Asociere negativă			Statistic semnificativ	
ASOCIERE factor de risc - boală STATISTIC NESEMNIFICATIVĂ				
OR (RR)	Val. minimă CI 95 %	Val. maximă CI 95 %	P	Interpretare
> 1	< 1	> 1	> 0,05	Poate fi observată o asociere POZITIVĂ, NESEMNIFICATIVĂ STATISTIC între factorul de risc și boala studiată
Asociere pozitivă			Statistic ne semnificativ	
OR (RR)	Val. minimă CI 95 %	Val. maximă CI 95 %	P	Interpretare
< 1	< 1	> 1	> 0,05	Poate fi observată o asociere NEGATIVĂ, NESEMNIFICATIVĂ STATISTIC, între factorul de risc și boala studiată
Asociere negativă			Statistic ne semnificativ	
OR (RR) $\cong 1$: între factorul de risc (expunere) și boală nu poate fi făcută nici un fel de asociere (în cazul studiului nostru am considerat această situație pentru $0,950 < OR (RR) < 1,050$)				

Riscul atribuit (RA)

- numărul de cazuri din populația expusă ce ar putea fi „eliminate” dacă expunerea (la factorul de risc) ar putea fi îndepărtată. (239)

Formula de calcul

$$RA = \frac{\text{Incidența în rândul populației expuse} - \text{Incidența în rândul populației neexpuse (la factorul de risc)}}{\text{Incidența în rândul populației expuse}}$$

TIPURI DE STUDII EPIDEMIOLOGICE

Este cunoscut faptul că deosebim 2 tipuri mari de studii epidemiologice (239)

I. **Descriptive** (persoane, locuri și timp)

- » Servesc la **generarea** de ipoteze de lucru

II. **Analitice** (cauzale)

- » Servesc la **testarea** ipotezelor de lucru

I. STUDII DESCRIPTIVE

CARACTERISTICI GENERALE:

- **Sunt, în general, studii de tip corelational**
 - » Examinează caracteristici ale întregii populații
 - Exemple: Studierea unor relații între vânzările de țigări pe teritoriul unei țări și mortalitatea datorată BPOC în acea țară
 - » Reprezintă *primul pas* în studierea unei *relații de tip factor de risc-boală*
 - » **Avantaje**
 - Metodă rapidă și ieftină, poate fi utilizată ca un prim pas în efectuarea unui studiu mai complex
 - » **Limitări**
 - **Nu poate furniza informații certe cu privire la existența sau nonexistența unor relații de tip cauză-efect**
 - **Nu poate oferi cercetătorului o modalitate de control detaliat al potențialelor surse de eroare**

SUBTIPURI ALE STUDIILOR DESCRIPTIVE:

A. Descriere de caz sau serii de cazuri (Case Reports and Case Series)

- » Descriu patologia unui singur bolnav sau a unui grup de bolnavi
- » Cea mai utilizată formă de studiu publicată în revistele medicale
- » **Avantaje**
 - Poate duce la formularea de noi ipoteze de lucru
 - Constituie un element major de legătură între medicina clinică și epidemiologie

» **Limitări**

- Nu poate fi folosit pentru verificarea unor ipoteze de lucru

B. Studiu descriptiv de tip urmărire a prevalenței (Cross-Sectional or prevalence Survey)

- » Expunerea la factorii de risc și starea de boală sunt simultan cuantificate în cadrul unei populații

- » Furnizează informații legate de frecvența și caracteristicile unei anumite boli

» **Avantaje**

- Foarte folositor în domeniul sănătății publice
- Poate furniza informații utile legate de prevalența (incidența) unei boli în anumite grupuri de populație (expuse la anumiți factori de risc specifici profesiei de exemplu)

» **Limitări**

- De cele mai multe ori nu poate să dea informații legate de faptul dacă expunerea la factorii de risc a precedat boala sau a apărut în timpul ei
- Datorită folosirii prevalenței ca indicator, nu se face o deosebire între cazurile noi și cele deja existente la momentul începerii studiului
- Nu poate fi utilizat pentru studierea factorilor etiologici (13)

C. Studiu descriptiv de tip ecologic

- » Studii care în care unitățile de analiza sunt mai degrabă populații întregi sau grupe de populații decât subiecți individuali (cazuri)

» **Avantaje**

- Este extrem de util în generarea de ipoteze de lucru

» **Limitări**

- Fiecare individ aparținând unei populații este caracterizat de media calculată pentru întreaga populație studiată
- Pot apărea erori datorate faptului că o asociere observată la nivelul întregii populații studiate nu reprezintă în mod obligatoriu asocierea existentă la nivel individual

- Aceste două limitări pot determina apariția unor erori ce constituie așa numitul “Ecologic Fallacy”

ASPECTE CE TREBUIE LUATE ÎN CONSIDERARE ÎN CAZUL EMITERII DE IPOTEZE ȘI A STABILIRII UNUI ANUMIT DESIGN PENTRU STUDIU:

- » **Ce boală va fi studiată și cum?**
 - **Bine fundamentată teoretic:** boală, respectiv studiu bine fundamentat teoretic, pe bază unor surse bibliografice numeroase
 - **Studiată empiric,** pe baza criteriilor de diagnostic
- » **Ce tip de expunere va fi studiată ?**
 - **Bine fundamentată teoretic:** există numeroase surse bibliografice și studii similare
 - **Studiată empiric,** pe baza criteriilor și metodelor de examinare alese de către cercetător
- » **Care este timpul de inducție ?** (Cu alte cuvinte care este intervalul mediu de timp dintre expunerea la factorul de risc și efect – apariția bolii ?)
- » **Ce alți factori pot influența această asociere ?**
 - **Factori bine studiați (fundamentați teoretic)**
 - **Empiric,** pe baza criteriilor și metodelor de examinare alese de către cercetător

II. STUDII ANALITICE

CARACTERISTICI GENERALE:

- Permit efectuarea unei comparații explicite între expunere (la factorii de risc) și boală
- Utilizează gruparea populației studiate în tabele de contingență realizate în așa fel încât să se poată observa clar (și determina statistic mai apoi) dacă riscul este diferit între populația expusă și cea neexpusă
- Sunt singurele studii ce pot fi utilizate pentru **TESTAREA DE IPOTEZE**

SUBTIPURI ALE STUDIILOR ANALITICE:

Studiilor analitice sunt grupate în două subtipuri mari:

- Studii analitice de tip **observațional** – urmează cursul natural al evenimentelor
- Studii analitice de tip **intervențional** – cercetătorul determină expunerea (de exemplu un anumit tratament) și urmărește subiecții (de exemplu așa numitele studii clinice - clinical trials)

În ceea ce urmează mă voi referi numai la prima categorie, studii analitice de tip **observațional**, *acesta fiind tipul de studii folosit de mine în cadrul acestei lucrări.*

STUDII ANALITICE OBSERVAȚIONALE – SUBTIPURI:

A. Studii de tip Case-Control (Case-Control studies)

- » Gruparea în tabele de contingență se face în funcție de prezența sau absența bolii:
 - Persoane care prezintă boala
 - Grup de comparație
- » Practic un astfel de studiu compară expunerea în rândul persoanelor bolnave (grup - cazuri) cu expunerea în rândul persoanelor sănătoase (grup - control)

Este studiul epidemiologic cel mai des utilizat, în ciuda potențialelor erori, ce pot apărea relativ frecvent fără o grupare atentă în tabele de contingență potrivite scopului propus (13)

Avantaje:

- » Tipul de studiu cel mai potrivit în cercetări legate de boli cu perioada lungă de incubare
- » Eficient din punct de vedere al timpului necesar și al costurilor aferente
- » Potrivit pentru studiul unor boli rare

Poate fi utilizat pentru studierea simultană a mai multor factori de risc, în cadrul aceleași boli

Dezavantaje:

- » Lipsit de eficiență în cazul studierii unor expuneri la factori de risc rar întâlniți (cum ar fi studierea efectelor unor accidente industriale)
- » Nu permite calculul direct al ratei de incidență a bolii

- » Uneori poate fi extrem de grea stabilirea unor relații temporale între expunere și apariția bolii

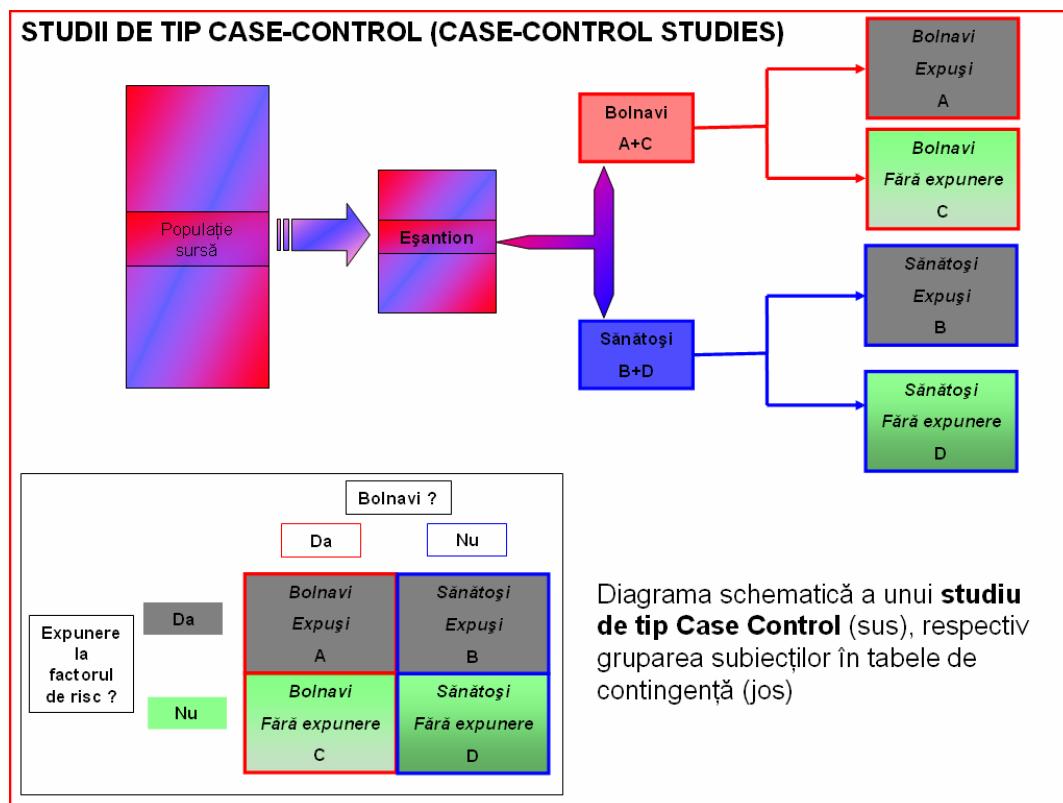


Fig.4 Diagrama schematică a unui studiu de tip Case-Control

Exemplu: - Studiul asocierii dintre apariția limfoamelor nonhodgkin și infecția cu virusul Epstein-Bar sau cu *Helicobacter pylori* (factori biologici de risc studiați și în cadrul acestei lucrări)

ASPECTE CE TREBUIE LUATE ÎN CONSIDERARE ÎN CAZUL UNUI STUDIU DE TIP CASE-CONTROL:

- **definirea și selecția cazurilor**
- **selecția cazurilor pe baza incidenței sau a prevalenței** (cu alte cuvinte cazurile studiate erau diagnosticate înainte de începerea studiului sau sunt cazuri noi diagnosticate recent sau chiar la momentul studiului)
- **validitatea studiu versus generalizarea studiului** (Poate fi studiul considerat valid? – *probleme legate de designul studiului*; Pot fi concluziile lui extinse la totalul populației? – *există alte studii similare ce au ajuns la concluzii asemănătoare?*)

- **sursa loturilor control:** clinici, eșantioane din populație în general, serii speciale de loturi martor etc.
- **raportul cazuri/control** (martor)
- **felul cum se face constatarea stării de boală, respectiv a expunerii la factorul de risc**

B. Studii de tip cohortă (Cohort studies)

- » Subiecții studiului sunt clasificați pe baza expunerii la un anumit factor de risc
- » Populația este urmărită pentru a determina prezența (apariția) bolii în cadrul acesteia ;
- » În funcție de modul de urmărire a prezenței bolii în cadrul populației deosebim:
 - Studii de tip cohortă **prospective**, respectiv **retrospective**
 - **Studiul de tip prospectiv** are ca punct de pornire timpul prezent, cercetătorul investigând populația (eșantionul) prospectiv (în viitor); **avantajul** major al unui astfel de studiu constă în faptul că cercetătorul poate colecta orice informație crede el de cuviință că are fi utilă.
 - În cazul studiului retrospectiv, cercetătorul **identifică** eșantionul (populația studiată – cohorta) și **studiază** acea populație de-a lungul unui anumit interval de timp **precedent** momentului în care a decis să înceapă studiul; **dezavantajul** unei astfel de abordări constă în faptul că cercetătorul are la dispoziție doar informații colectate în trecut, fără a mai putea avea acces la eventuale informații utile, dacă acestea nu au fost deja colectate și arhivate în vreun fel.

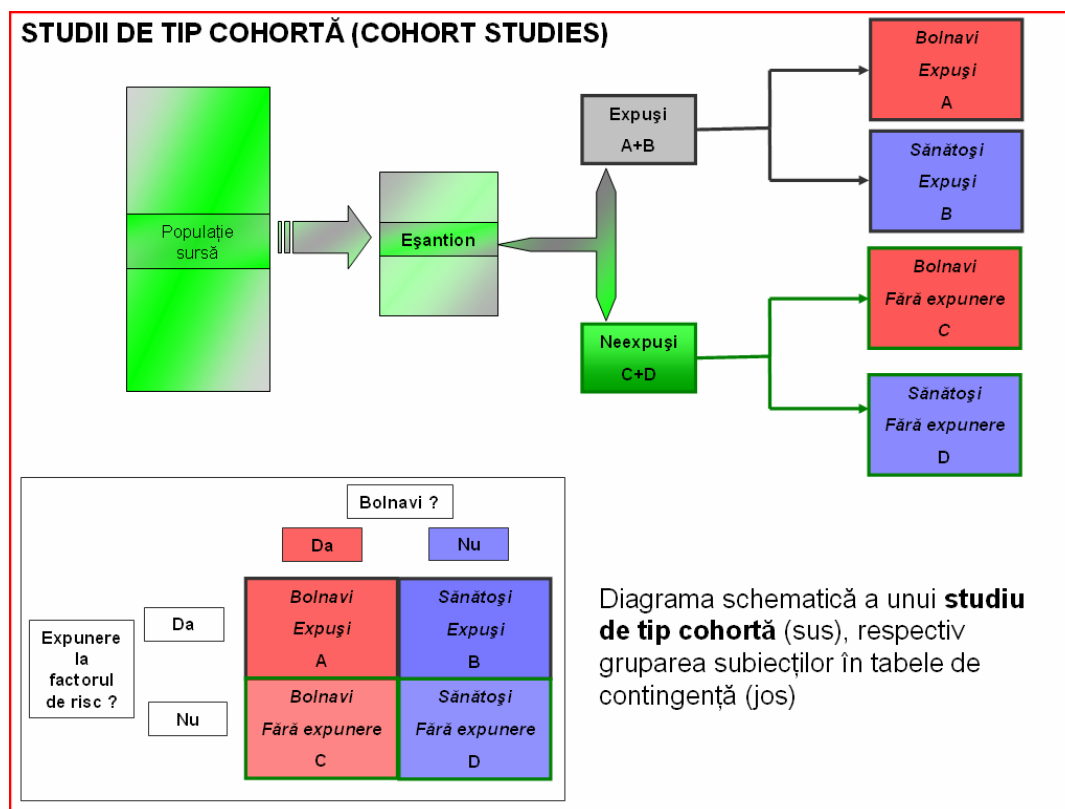


Fig.5 Diagrama schematică a unui studiu de tip Cohortă

Avantaje:

- » Astfel de studii pot măsura incidența bolii și deci și cuantifica riscul
- » Expunerea precede apariția bolii
- » Poate servi la studiul unei mari categorii de boli
- » Este foarte eficient în cazul unor **expuneri la factori de risc rar întâlniți**, cum este cazul expunerilor datorate unor **accidente industriale**, aceasta deoarece *investigatorul poate să aleagă eșantionul avînd ca principal criteriu expunerea la factorul de risc*

Dezavantaje:

- » Necesită un număr mare de subiecți
- » Relativ lipsit de eficiență în cazul unor boli rare
- » Perioadă lungă de urmărire a subiecților studiului
- » Subiecții pot suferi schimbări în starea de sănătate de-a lungul perioadei de studiu, fără ca investigatorul să aibă cunoștința de acest lucru
- » Necesită, în general, costuri mari pentru realizare

Exemplu:- Expuneri neobișnuite (accidente industriale), cum este și cazul accidentului nuclear de la Cernobîl (1986), **care ne-a permis efectuarea, în cadrul acestei lucrări, a unui studiu retrospectiv de tip cohortă** în scopul *studierii rolului radiațiilor ionizante ca factor de risc în apariția unor leucemii și limfoame*

ASPECTE CE TREBUIE LUATE ÎN CONSIDERARE ÎN CAZUL UNUI STUDIU DE TIP COHORTĂ:

- **modul de selecție (grupare) a populației expuse**
- **modul de selecție (grupare) a populației neexpuse** (grup de comparație sau martor)
- **datele legate de sursa de expunere la factorul de risc**
- **datele legate de variabilele de ieșire** (apariția bolii, indicatorii folosiți, eventualele surse de eroare)

SELECȚIA UNUI ANUMIT TIP DE STUDIU (STUDY DESIGN)

Trebuie să țină cont de următorii factori și criterii:

❖ Criteriul gradului de cunoaștere științifică a fenomenului studiat (bolii)

- » Studiu ce **aduce informații noi** sau studiază o **boală nouă** (vezi Severe Acute Respiratory Syndrome – SARS)
- » Studiu ce **adaugă** informații relevante despre o boală
- » Studiu ce **confirmă** informații deja existente despre o boală

❖ Criteriul ipotezei de lucru

- » Studiu ce servește la **generarea** de ipoteze
- » Studiu ce servește la **testarea** unor ipoteze de lucru

❖ Criteriul epidemiologic / statistic

- » Studiu ce servește la verificarea unor ipoteze privitoare la **boală** (boala – subiect principal de studiu)
- » Studiu ce servește la verificarea unor ipoteze privitoare la **expunere** (expunerea la factorul de risc- subiect principal de studiu)

»Mărimea eșantionului și reprezentativitatea lui

❖ **Criteriul resurselor disponibile (de obicei factorul limitant)**

»Costuri

»Timp

ASOCIERE VERSUS CAUZALITATE

"The presence of an association ... in no way implies that the observed relationship is one of cause and effect." (Hennekens and Buring)

Principala idee care trebuie să răzbată în urma unei analize statistice efectuate în cadrul unui studiu epidemiologic, este faptul că *asocierea expunere (factor de risc)-boală* se referă strict la **existența unei dependențe statistice între două variabile**.

Pentru a putea considera, eventual, această **asociere** ca o **relație cauză-efect**, există o serie de criterii de interpretare a unui studiu epidemiologic a căror parcurgere este o condiție sine qua non.

Aceste criterii minimale sunt prezentate în cele ce urmează:

- *Există o asociere expunere (factor de risc)-boală, validă din punct de vedere statistic?*

DACĂ DA ...

- *Se poate datora această asociere șansei sau nu...?*

- *Se poate datora această asociere unor erori ale studiului, legate de bias...?*

- *Se poate datora această asociere unor erori ale studiului, legate de anumite confuzii?*

CRITERII PENTRU STABILIREA UNEI RELAȚII DE TIP CAUZĂ-EFECT

Pentru a răspunde la întrebarea: *Poate fi judecată o asociere statistică ca o relație de tip cauză-efect ?* – trebuie luate în considerare următoarele criterii:

1. Magnitudinea asocierii:

- i. Dacă a fost relevată existența unei asocieri statistice semnificative, cât de mare este magnitudinea acesteia ?

2. Credibilitatea biologică a ipotezei de lucru:

- i. Există un mecanism deja cunoscut care să explice științific asocierea ?

3. Consistența rezultatelor studiului vis-à-vis de alte studii similare:

- i. Există alte studii care să confirme concluziile noastre?

4. Secvența de timp:

- i. Pentru cazurile studiate, precede expunerea apariția bolii cu o perioadă de timp consistentă (a se vedea timpul de inducție al bolii) pentru un mecanism biologic credibil?

5. Relația “doză”- răspuns:

- i. Putem vorbi de un gradient al riscului ce poate fi relaționat cu nivelul expunerii?

6. Poate fi considerată asocierea ca fiind specifică?

- i. Câți factori cauzează boala?
- ii. Câte boli pot fi cauzate de factorul respectiv?