



MIB

**METODOLOGIE
INFORMATICA
BIOSTATISTICA**

Studiul factorilor implicați în patogeneză

TUDOR DRUGAN

2023

Objective

0 Factor de risc sau semn diagnostic.....●

1 RIE, RIN.....●

2 RR, RA, OR.....●

3 Modalitati de efectuare a studiului.....●

4 Statistica.....●

Tabel de contingență

FR+ = factor de risc prezent;

FR- = factor de risc absent;

proporția subiecților dintr-o anumită categorie, respectiv:

- a = bolnavi cu factorul de risc prezent;
- b = indemni de boală cu factorul de risc prezent;
- c = bolnavi fără factorul de risc;
- d = indemni de boală, fără factorul de risc.

Afecțiunea / FR	B+ Bolnavi	B- Sănătoși	Total
FR+	a	b	a+b
FR-	c	d	c+d
Total	a+c	b+d	n

Riscul individual al celor expuși

RIE

probabilitatea ca un subiect să fie bolnav dacă este expus factorului de risc

$$\text{RIE} = p(B^+|FR^+)$$

exprimă proporția de bolnavi din subiecții expuși factorului de risc.

$$\text{RIE} = a/(a+b).$$

Afecțiunea / FR	B+ Bolnavi	B- Sănătoși	Total
FR+	a	b	a+b
FR-	c	d	c+d
Total	a+c	b+d	n

Riscul individual al celor neexpuși

RIN

probabilitatea ca un subiect să fie bolnav fără să fie expus factorului de risc

$$RIN = p(B^+ | FR^-)$$

și exprimă proporția de bolnavi din totalul subiecților neexpuși factorului de risc.

$$RIN = c / (c + d).$$

Afecțiunea / FR	B+ Bolnavi	B- Sănătoși	Total
FR+	a	b	a+b
FR-	c	d	c+d
Total	a+c	b+d	n

Excesul de risc – riscul atribuabil

RA

măsoară specificitatea legăturii între factorul prognostic și îmbolnăvire:

$$RA = RIE - RIN$$

$$RA = p(B^+|FR^+) - p(B^+|FR^-).$$

respectiv partea din riscul absolut care este imputabil expunerii la factorul de risc studiat.

Afecțiunea / FR	B+ Bolnavi	B- Sănătoși	Total
FR+	a	b	a+b
FR-	c	d	c+d
Total	a+c	b+d	n

Riscul relativ

$$RR = \frac{\Pr(B|R)}{\Pr(B|\bar{R})}$$

= raportul dintre probabilitatea condiționată de B a se întâmpla evenimentul A și probabilitatea condiționată de inexistența B a se întâmpla evenimentul A

măsoară forța asocierii:

$$RR = RIE/RIN$$

$$RR = p(B^+|FR^+)/p(B^+|FR^-)$$

$$RR = [a/(a+b)]/[c/(c+d)]$$

Interpretarea riscului relativ

RR=1, înseamnă: nu există factor de risc (factorul presupus nu poate fi incriminat în etiologia bolii luate în studiu).

RR>1, înseamnă: factorul studiat reprezintă un risc pentru boală.

RR<1 înseamnă: factorul studiat este protector.

Recomandare metodologică:

- **Valorile RR cuprinse între 0,9 și 1,1 sunt foarte susceptibile la erori de studiu și nu este recomandabil să fie folosite în publicații fără verificarea riguroasă sau refacerea studiului**

Risc și RR

**RR măsoară asocierea dintre un factor și riscul de apariție a unei reacții adverse, de a face o afecțiune, de a deceda...
într-o perioadă de timp**

RR este practic raportul dintre riscul în grupul expus și cel neexpus, respectiv subiectul expus la factorul studiat are o probabilitate de x ori mai mare de a face boala definită prin criteriul principal de evaluare față de subiectul neexpus.

Pentru clinician RR arată beneficiul bolnavului dacă se înlătură factorul de risc

Intervalul de încredere pentru RR

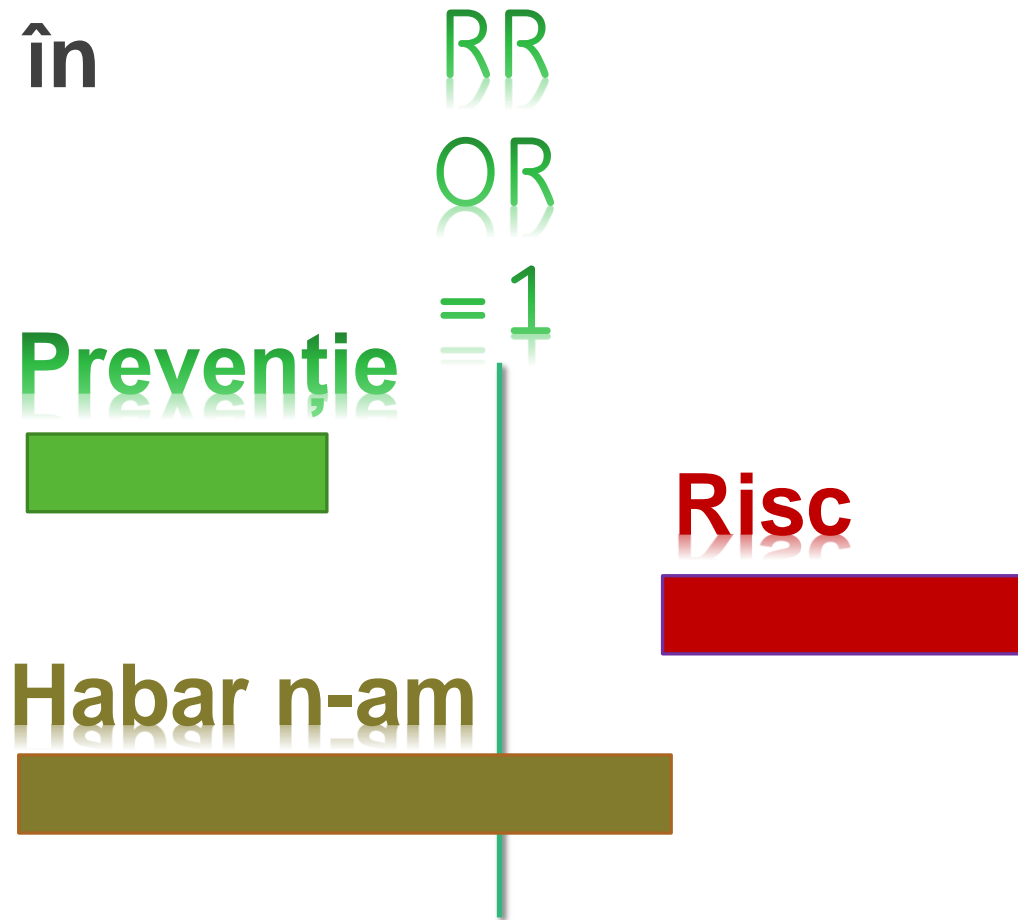
RR, OR iau valori în
intervalul $[0, +\infty]$

Exemple

$[0,2, 0,5]$

$[1,5, 2,4]$

$[0,2, 1,5]$



Fracția etiologică a riscului la expuși

FREE

măsoară specificitatea legăturii:

$$\text{FREE} = (\text{RIE} - \text{RIN}) \times 100 / \text{RIE}$$

Afecțiunea / FR	B+ Bolnavi	B- Sănătoși	Total
FR+	a	b	a+b
FR-	c	d	c+d
Total	a+c	b+d	n

OR = Odds Ratio sau Raportul șanselor

Șansa = Probabilitate / (1-Probabilitate)

Afecțiunea / FR	B+ Bolnavi	B- Sănătoși	Total
FR+	a	b	a+b
FR-	c	d	c+d
Total	a+c	b+d	n

$$\text{Șansa}_{\text{exp}} = \text{Odds}_{\text{exp}} = [a/(a+b)]/[b/(a+b)] = a/b$$

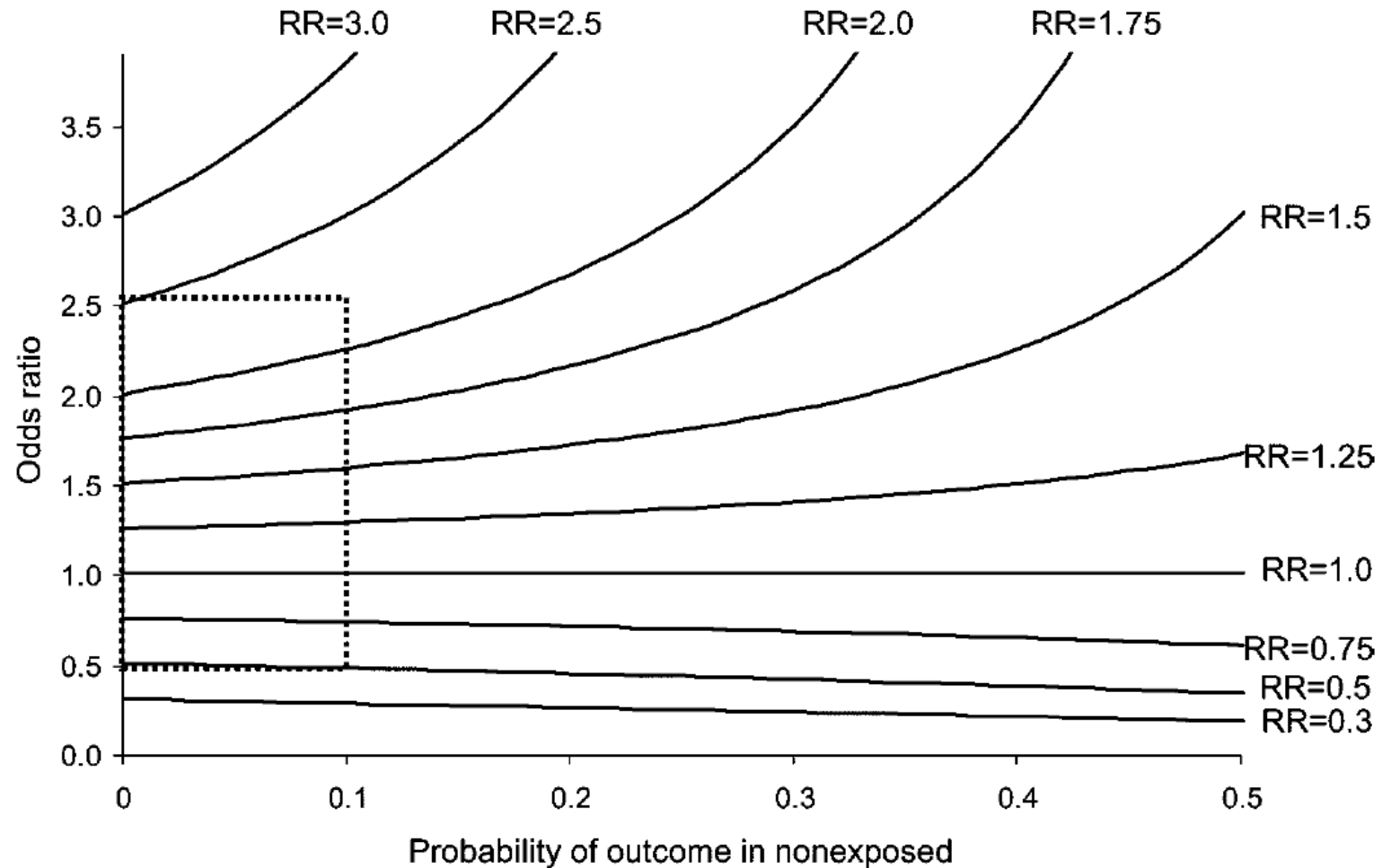
$$\text{Șansa}_{\text{ne-exp}} = \text{Odds}_{\text{ne-exp}} = [c/(c+d)]/[d/(c+d)] = c/d$$

$$\text{OR} = (a/b)/(c/d) = ad/bc$$

OR sau RR?

$RR \approx OR / [(1 - Pr_o) + (Pr_o)(OR)]$ unde Pr_o este probabilitatea de apariție a fenomenului în grupul studiat

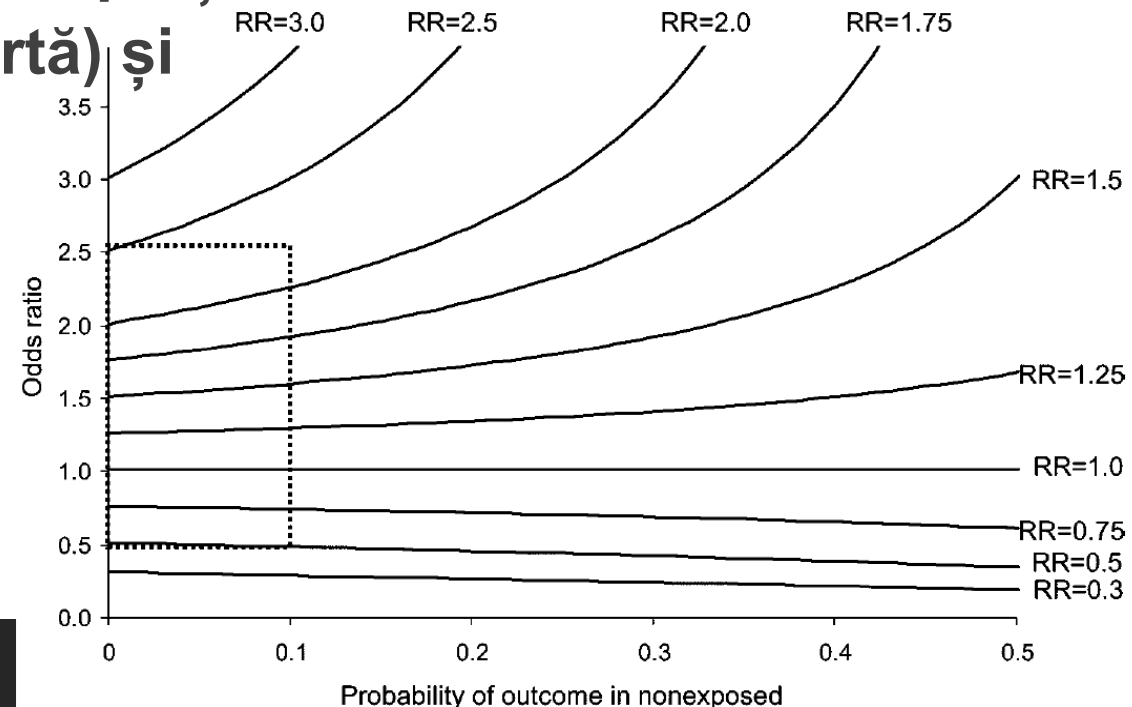
RR este aproximativ egal cu OR numai dacă $Pr_o \ll 10\%$



OR sau RR?

În cazul patologiilor rare OR este aproximarea destul de bună a RR; de aceea acest tip de culegere a datelor se utilizează predilect pentru studiul relației factor de risc – îmbolnăvire în patologia rară,

În rest fiind de preferat culegerea datelor de tip eșantion reprezentativ sau expus – nonexpus (cohortă) și calcularea RR

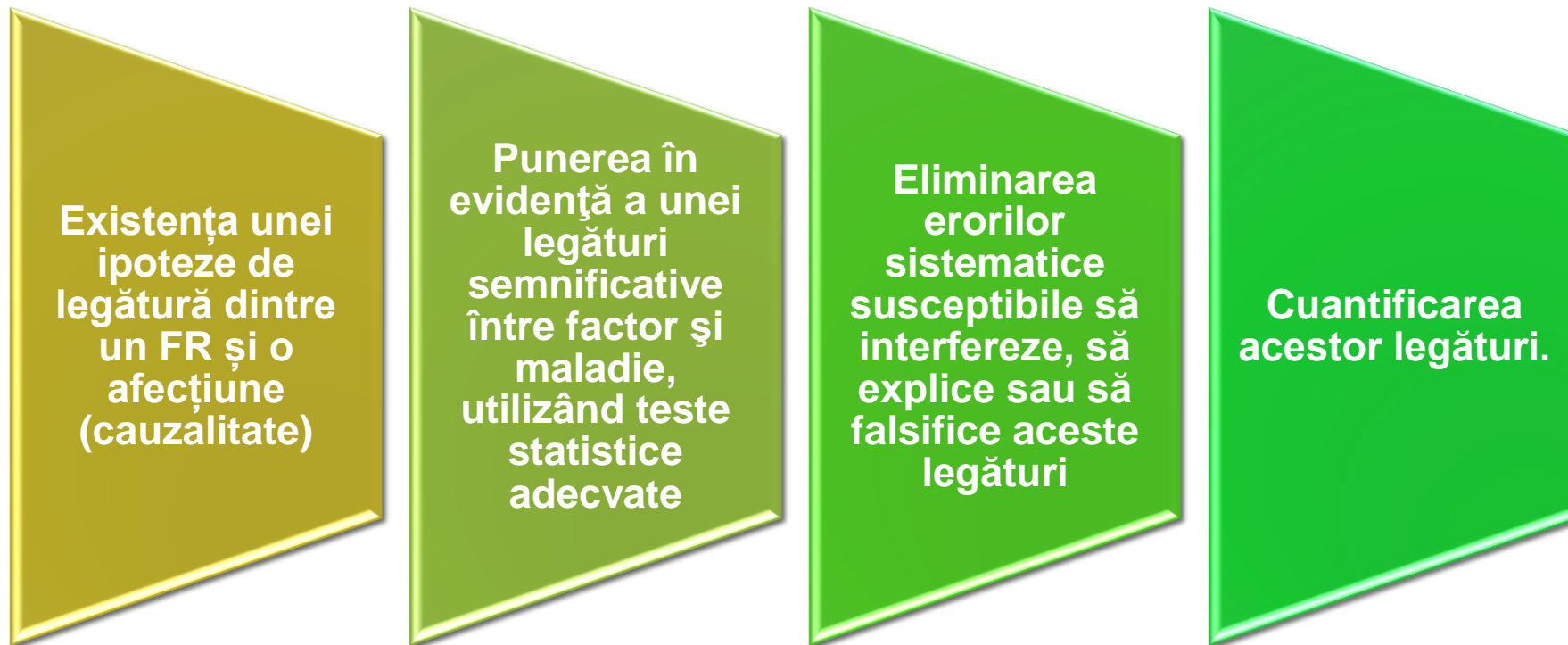


Studiul factorilor implicați în patogeneză

Obiectivul principal:

- demonstrarea importanței unui factor în etiologia unei îmbolnăviri.

4 stadii:



Studiul factorilor implicați în patogeneză

Metodologic studiile compară grupuri de subiecți cu și fără factorul prognostic prezent.

Urmărirea subiecților:

- **Prin observare (fără intervenția cercetătorului în derularea studiului).**
- **Experimentală: în trialul clinic la evidențierea:**
 - **Efectelor adverse**
 - **Efectelor tratamentului**

3 tipuri de culegere a datelor:

- **eșantion reprezentativ,**
- **expus – nonexpus**
- **caz - martor**

Cercetarea
factorilor
prognostici prin
culegere de date tip
eșantion
reprezentativ

	B^+	B^-	
FR^+	a	b	ne^+
FR^-	c	d	ne^-
	nB^+	nB^-	N



Probleme metodologice

Acest tip de culegere a datelor este utilizată cel mai adesea prin analiză transversală,

Permite evaluarea în același timp a factorilor prognostici studiați și a criteriilor de evaluare prestabilite printr-o fotografie instantanee a situației clinice.

Dacă cercetarea factorului prognostic este retrospectivă, apar inconveniente datorate erorilor sistematice de declarare greu de controlat și cuantificat.

Alegerea acestei metodologii de culegere în cazul unei patologii rare, poate pune probleme de eșantionaj.

Tabelul de contingență

prevalența $p(M^+)$

= numărul de îmbolnăviri din populația generală

$$p(M^+) = a+c/N.$$

Extrapolarea rezultatelor la populația generală se face în cadrul unui interval de încredere.

Se pot calcula de asemenea:

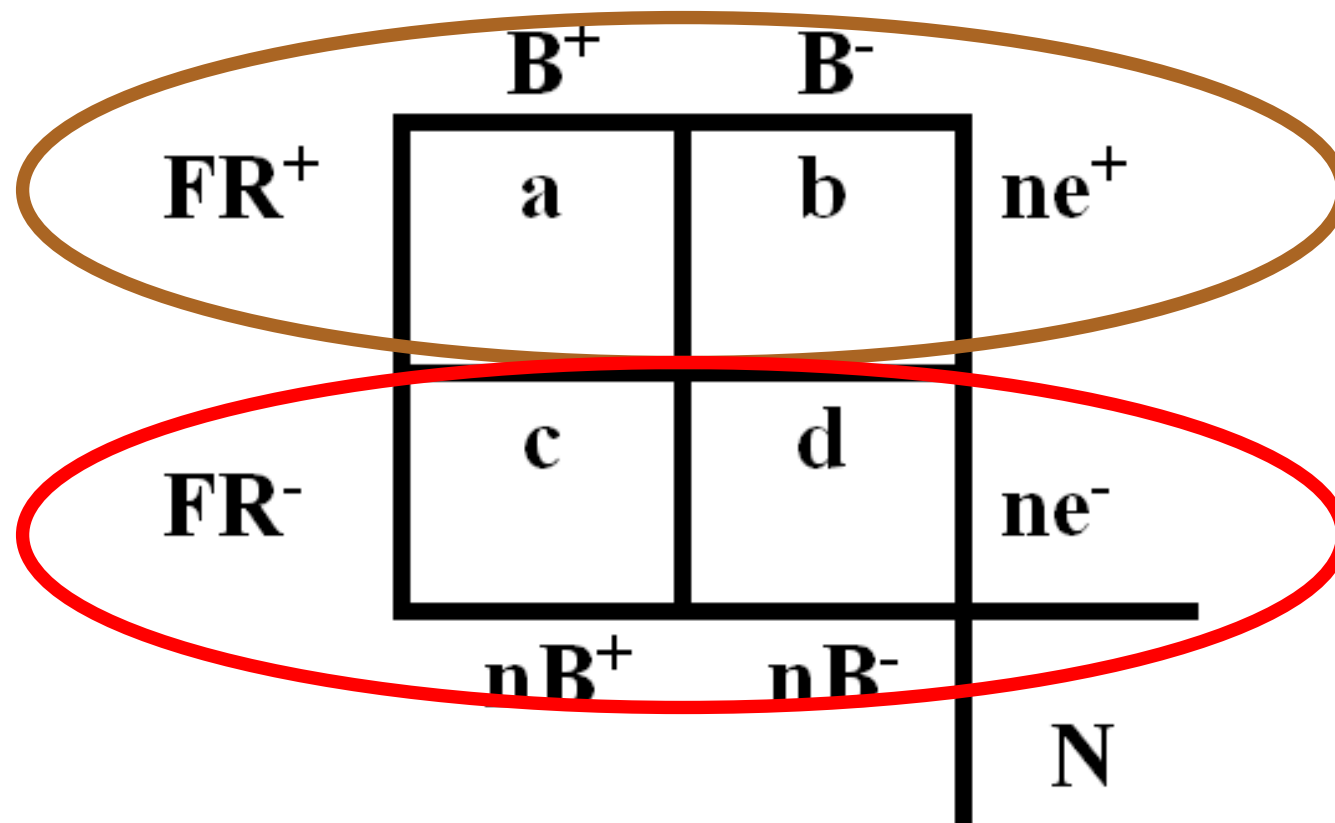
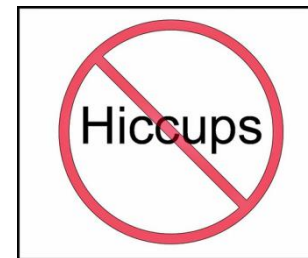
- factorii de risc individuali pentru cei expuși și nonexpuși (RIE și RIN),
- riscul relativ (RR),
- riscul atribuabil (RA)
- fracțiunea etiologică de risc la expuși (FREE).

	B^+	B^-	
FR^+	a	b	ne^+
FR^-	c	d	ne^-
	nB^+	nB^-	N

Cercetarea factorilor prognostici prin culegerea datelor de tip expus – nonexpus



	B^+	B^-	
FR^+	a	b	ne^+
FR^-	c	d	ne^-
	nB^+	nB^-	N



Consideratii metodologice

Este întotdeauna comparativ.

Compararea se face asupra apariției sau nu a unei patologii în funcție de expunere.

Expunerea trebuie înțeleasă ca prezența oricărui factor studiat:

- **expunere la un toxic, la un factor de risc,**
- **apartenență la un grup HLA sau la un tip comportamental**
- **dar și la un grup cu anumită terapie.**

Culegerea prospectiva

Studiu de lungă durată și costisitor

Bogat în date și rezultate.

Această situație necesită o atenție specială privind pierduții din vedere.

Culegerea datelor se face pe măsura derulării evenimentelor.

Nu permite studiul patologiei rare (ar necesita cohorte uriașe pentru a observa un număr redus de subiecți dezvoltând patologia de interes);

Permit însă analiza efectelor unei expuneri rare.

Consideratii metodologice

Postulatul de bază este ca întreaga populație studiată (cohorta) să fie îndemnă de boală la începutul cercetării.

Fiecare subiect este văzut de mai multe ori pe parcursul cercetării (minim 2)

	B^+	B^-	
FR^+	a	b	ne^+
FR^-	c	d	ne^-
	nB^+	nB^-	N

Consideratii metodologice

Prin prestabilirea celor două grupe FR^+ și FR^- pe baza unor criterii clare de incluziune și excluziune (criterii de eligibilitate pentru apartenența unui subiect la un anumit grup de studiu), calculele statistice se pot efectua între componenta de deasupra și dedesubtul liniei îngroșate a tabelului de contingență.

Se pot compara între cele două grupe proporția de subiecți bolnavi dar nu și frecvența expunerii.

	B^+	B^-	
FR^+	a	b	ne^+
FR^-	c	d	ne^-
	nB^+	nB^-	N

Culegerea retrospectiva

Retrospectivă =istorică

cu un risc de erori sistematice mult mai mare

- **de eșantionare,**
- **de culegere,**
- **de declarare**

**În aceste situații experimentatorul “se întoarce în timp”
și se poziționează la începutul expunerii luate în studiu**

Consideratii metodologice

Ca și în cazul alcătuirii grupelor de studiu prin eșantion reprezentativ și în acest tip de culegere pot fi calculate RIE, RIN, RR, RA și FREE.

Acest tip de eșantionare permite cu ceva mai mare ușurință urmărirea anumitor manifestări patologice în raport cu expunerea studiată

un exemplu arhicunoscut este cohorta FRAMINGHAM urmărită timp de peste 30 de ani.

Trebuie făcute însă cele mai riguroase demersuri pentru studiul tuturor argumentelor în favoarea relațiilor cauzale

Studiul de cohorta FRAMINGHAM

Studiul Framingham a inceput in 1948 sub coordonarea U.S. Public Health Service si a fost transferat la National Heart Institute, NIH, in 1949.

Esantionul a fost extras din Framingham, Massachusetts, femei si barbati

A fost primul studiu prospectiv al BCV

A identificat conceptul de factori de risc

Studiul continua printr-un contract al Boston University dat de NHLBI

Studiul de cohorta FRAMINGHAM

Subiecti:

- Au intrat in studiu intre 1948-1952, 5,209 barbati si femei, de 28-62 ani, din Framingham, MA.
- In 28.02.1999, mai erau 993 supravietuitori.

Design:

- Studiul Framingham este o investigatie longitudinala a factorilor constitutionali si de mediu care influenteaza dezvoltarea BCV la barbati si femei
- Examinarea participantilor s-a facut la fiecare 2 ani si cohorta a fost urmarita continuu pentru morbiditate si mortalitate

Studiul de cohorta FRAMINGHAM

Obiective primare:

- **Studiul Framingham**

- A evaluat incidenta si prevalenta bolilor cardiovasculare (BCV) si a factorilor lor de risc,
- Tendintele in evolutia incidentei BCV si a factorilor lor de risc in timp,
- Agregarea familiara a BCV si a factorilor lor de risc.

Obiective secundare:

- **Estimarea ratelor incidentei BCV si descrierea istoriei naturale a BCV,**
- **Evaluarea secventei clinice a semnelor care preced sindromul diagnosticabil clinic precum si consecintele evolutia dupa declansarea bolii manifeste**

Cohorta Framingham



Arată bine pentru 107 ani, nu ?

Cercetarea factorilor prognostici prin culegerea datelor de tip caz martor



	B^+	B^-	
FR^+	a	b	ne^+
FR^-	c	d	ne^-
	nB^+	nB^-	N

Consideratii metodologice

Urmărește în mod analitic (comparativ) două grupe predefinite:

- **Subiecții purtători ai bolii predefinite de criteriul principal de evaluare**
- **Subiecții indemni de boală**

Asemanarea celor două grupuri trebuie să fie cât mai mare exceptând caracterul bolnav/indemn de boală.

Consideratii metodologice

Se caută în cele două grupuri proporția de subiecți expuși și nonexpuși la factorul prognostic înainte de apariția bolii.

Compararea este întotdeauna **retrospectivă** între pacienții bolnavi (cazuri) și un grup de subiecți indemni de boală dar altfel întru totul comparabili cu primul grup (martori).

Avantajele

Cost rezonabil

Durată relativ scurtă a studiului (în comparație cu abordările prospective)

Se pot analiza patologii rare sau cu latență prelungită

Se pot studia concomitent multipli factori prognostici.

Limitele

Imposibilitatea calculului incidenței patologiei în populație

Imposibilitatea calculării frecvenței factorului de expunere

Dificultatea culegerii tuturor datelor privind evenimente din trecut (erori sistematice de culegere și declarare, dependente de memoria subiecților interogați)

Dificultăți în controlul tuturor factorilor care ar putea genera diferențe între bolnavi și martori (erori sistematice de eșantionare)

Consideratii metodologice

Erorile sistematice pot fi multiple, în consecință în acest tip de studiu identificarea și controlul lor trebuie să reprezinte una din prioritățile redactării protocolului și efectuării studiului.

Predefinirea grupelor cu separarea (prin linia verticală îngroșată din tabelul de contingență) cazurilor de martori, nu permite calculul RIE, RIN, deci nu poate fi realizat raportul și diferența acestora, fiind deci imposibilă calcularea RR și RA.

	B ⁺	B ⁻
FR ⁺	a	b
FR ⁻	c	d
	n ₁ =a+c total bolnavi	n ₂ =b+d total indemni de boală

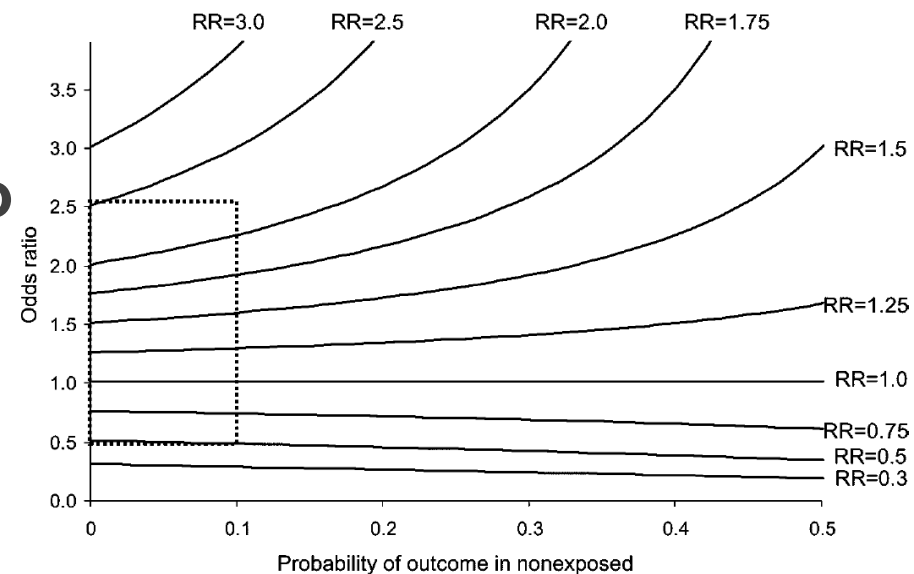
Tabel de contingență

Concluzia acestui tip de studiu se exprimă în riscul de a dezvolta boala în funcție de expunere și se estimează prin rata șansei de a face sau a nu face boala.

Raportul șanselor = Odds Ratio (OR) este

$$OR = (a \times d) / (b \times c)$$

	B ⁺	B ⁻
FR ⁺	a	b
FR ⁻	c	d
	n ₁ =a+c total bolnavi	n ₂ =b+d total indemni de boală





Statistica

VALIDAREA STUDIILOR PENTRU FACTORI PROGNOSTICI

TESTUL CHI PATRAT



Exemplu

Se dorește studierea
asocierii dintre lipsa
unei mese de
dimineață regulate,
consistente și
echilibrate și apariția
bolilor cardio-
vasculare



Exemplu

Pentru aceasta se realizează un studiu retrospectiv pe un eșantion de 400 de subiecți dintre care:

- 160 au făcut afecțiuni cardio-vasculare (B+)
- 130 au declarat la anamneză că nu au avut o masă de dimineață corectă în ultimii 10 ani (FR+). 80 dintre ei au fost diagnosticați cu afecțiuni cardio-vasculare

Afecțiune	B+	B-	Total
Lipsa micului dejun			
FR+	80	50	130
FR-	80	190	270
Total	160	240	400

RR

RR = • $RIE/RIN = (80/130)/(80/270) = 27/13 = 2,08$

Afecțiune	B+	B-	Total
Lipsa micului dejun			
FR+	80	50	130
FR-	80	190	270
Total	160	240	400

Testul Hi-pătrat de independență

Se caută să se stabilească dacă lipsa mesei de dimineață influențează apariția afecțiunilor cardio-vasculare sau dacă apariția acestora este independentă de presupusul factor de risc (validarea RR calculat).

Tabelul de contingență prezentat se numește tabel de contingență observat, iar frecvențele pe care le conține se numesc frecvențe observate.

Afecțiune	B+	B-	Total
Lipsa micului dejun			
FR+	80	50	130
FR-	80	190	270
Total	160	240	400

Definirea ipotezei nule H0

H_0 : afecțiunea și masa de dimineață sunt independente.

H_0 : lipsa mesei de dimineață nu are influență asupra apariției acestor afecțiuni.

ALEGEREA UNUI PARAMETRU

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Testul Hi-pătrat de independență

Se face ipoteza de independență între cele două caractere B și FR (adică ipoteza nulă H_0 , în acest caz) atunci

Se calculează un tabel de contingență teoretic care satisface această ipoteză de independență.

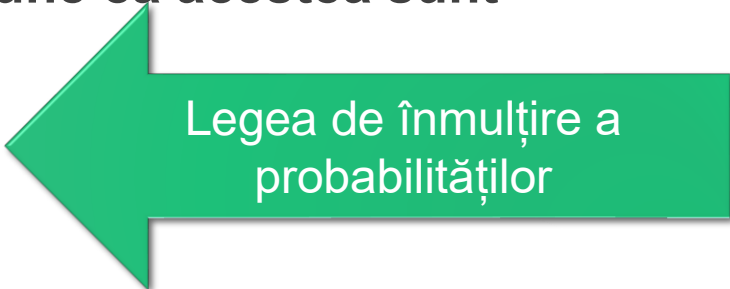
Se determină apoi abaterea (ecartul) dintre cele două tabele de contingență observat și teoretic.

- Dacă această abatere este mică atunci ea este explicată doar prin întâmplare (hazard) și ipoteza de independență este acceptată.
- Dacă această abatere este importantă atunci ipoteza de independență trebuie să fie respinsă.

Testul Hi-pătrat de independență

Problema este următoarea: dispunând de un eșantion de $n = 400$ de subiecți dintre care 160 au maladia B iar 130 nu au consumat micul dejun, să se determine cum sunt repartizați subiecții în funcție de cele două caractere (B și FR) dacă se presupune că acestea sunt independente

Afecțiune	B+	B-	Total
Lipsa micului dejun			
FR+	$\frac{130 \times 160}{400}$	$\frac{130 \times 240}{400}$	130
FR-	$\frac{270 \times 160}{400}$	$\frac{270 \times 240}{400}$	270
Total	160	240	400



Legea de înmulțire a probabilităților

Afecțiune	B+	B-	Total
Lipsa micului dejun			
FR+	52	78	130
FR-	108	162	270
Total	160	240	400

Ecartul dintre cele două tabele de contingență

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Afecțiune	B+	B-	Total
Lipsa micului dejun			
FR+	80	50	130
FR-	80	190	270
Total	160	240	400

Afecțiune	B+	B-	Total
Lipsa micului dejun			
FR+	52	78	130
FR-	108	162	270
Total	160	240	400

$$X^2 = \frac{(80 - 52)^2}{52} + \frac{(50 - 78)^2}{78} + \frac{(80 - 108)^2}{108} + \frac{(190 - 162)^2}{162} = 37,22$$

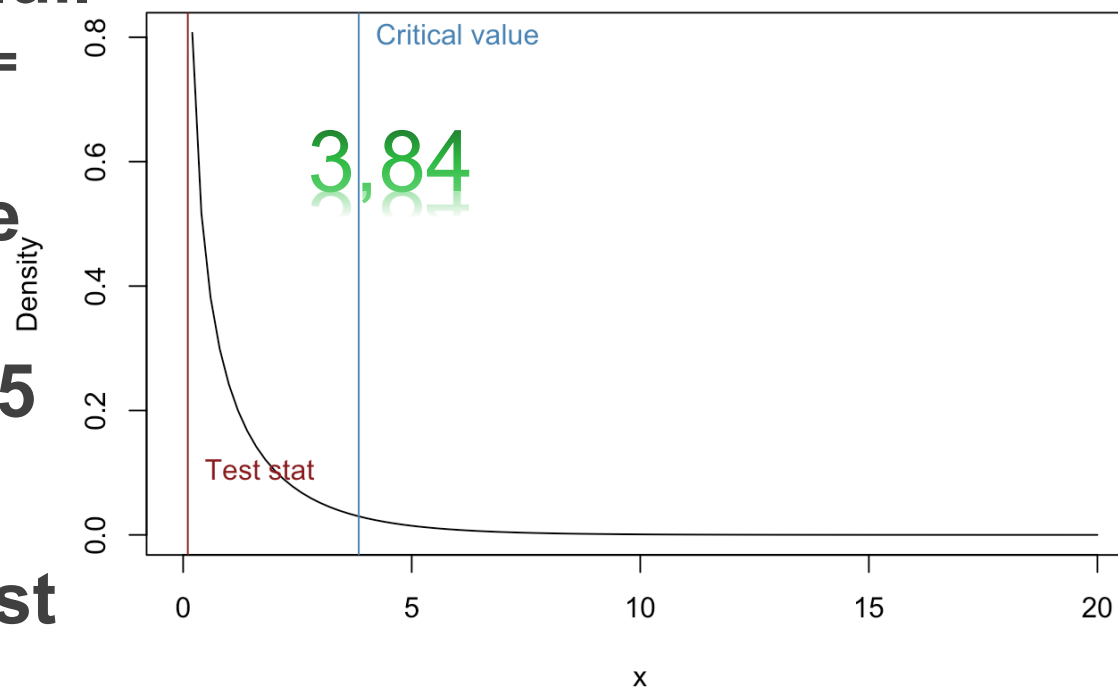
Calcule

α pragul de semnificație al testului.

- pragul de semnificație $\alpha = 0.05$

Regiunea critică este $[X^2_{\alpha}, \infty)$.

Pentru pragul $\alpha = 0.05$ și cu 1 grad de libertate valoarea = 3.84, astfel că în acest caz regiunea critică este intervalul $[3.84, \infty)$.



Decizia

Dacă $X^2 \in [3.84, \infty)$ se respinge H_0 cu un risc de eroare de prima speță $\leq \alpha$.

Dacă $X^2 \notin [3.84, \infty)$ atunci H_0 nu se poate respinge

$X^2 = 37,22 \gg 3.84$ așa că ipoteza nulă H_0 se respinge cu un risc inferior lui 5%.

În concluzie, masa de dimineață neregulată, ne-echilibrată și inconsistentă se asociază cu afecțiunile cardio-vasculare în sensul favorizării acestora ($RR=2,08$).



Cum ne
exprimăm sau
cum redactăm



În studiul nostru, masa de dimineață neregulată, ne-echilibrată și inconsistentă s-a asociat cu afecțiunile cardio-vasculare în sensul favorizării acestora. Riscul relativ de a face astfel de afecțiuni a fost de 2,077 [1,637;2,596], semnificativ statistic $X^2=37,227$, $p<0,001$.

Cum redactăm



Distribuția factorului de risc în lotul de studiu



Testul Chi pătrat

Într-un tabel de contingență distribuția observată se compară cu cea teoretică cu ajutorul testului Chi-pătrat. Condițiile de aplicare sunt:

- Dacă toate valorile din tabelul de contingență **teoretic** sunt peste 5 atunci se folosește Chi-pătrat
- Dacă există cel puțin o valoare în tabelul de contingență **teoretic** între 2 și 5 se va folosi Chi-pătrat cu corecția Yates sau Mantel-Haenszel.
- Dacă există cel puțin o valoare în tabelul de contingență **teoretic** mai mică decât 2 se folosește testul exact al lui Fisher

Testul Chi pătrat

Tabel de contingență
distribuția teoretică

Toate valorile sunt
peste 5

Chi-pătrat

O valoare între 2 și 5

Chi-pătrat cu corecția
Yates sau Mantel-
Haenszel

O valoare mai mică
decât 2

Testul exact al lui Fisher

<http://statpages.info/ctab2x2.html>

Observed Contingency Table

		Condition /disease		
		Present	Absent	Totals
T e s t	Positive	80 = a (TP)	50 = b (FP)	130 = r1
	Negative	80 = c (FN)	190 = d (TN)	270 = r2
	Totals	160 = c1	240 = c2	400 = t

Confidence Level: %

Select from list or enter a more appropriate conf. level

Compute

Download result as CSV

http://statpages.info/ctab2x2.html

Expected Frequencies

52.000	78.000	130.000
108.000	162.000	270.000
160.000	240.000	

Cell Propotions

0.200	0.125	0.325
0.200	0.475	0.675
0.400	0.600	

Row Propotions

0.615	0.385	0.325
0.296	0.704	0.675

Column Propotions

0.500	0.208
0.500	0.792
0.400	0.600

Chi-Square Tests

Type of Test	Chi Square	d.f.	p-value
Pearson Uncorrected	37.227	1	0.000
Yates Corrected	35.909	1	0.000
Mantel-Haenszel	37.134	1	0.000

Fisher Exact Test

Type of comparison (Alternate Hypothesis)	p-value
Two-tailed (to test if the Odds Ratio is <i>significantly different</i> from 1): If you don't know which Fisher Exact p-value to use, use this one . This is the p-value produced by SAS, SPSS, R, and other software.	0.000

<http://statpages.info/ctab2x2.html>

Quantities derived from a 2-by-2 table

Quantities Derived from the 2-by-2 Contingency Table	Value	95% CI	
Odds Ratio (OR) = $(a/b)/(c/d)$;	3.800	2.393	6.045
Relative Risk (RR) = $(a/r1)/(c/r2)$;	2.077	1.637	2.596

Atenție la intervalul de încredere să nu sară peste 1

Factori de risc

**Esantion
reprezentativ**

**RIE, RIN, RR, RA și FREE
Prevalenta**

Caz-martor

OR

**Expus -
nonexpus**

RIE, RIN, RR, RA și FREE

p