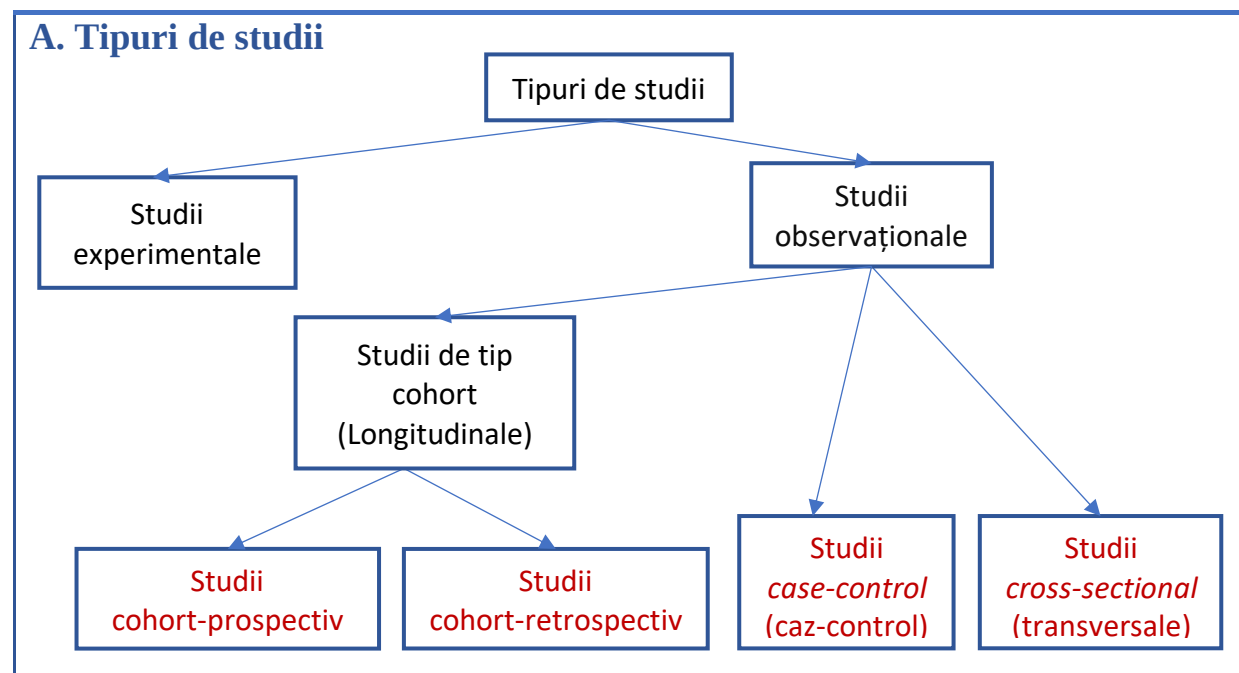


Studii observaționale. Analiza riscului

I. Repere teoretice

Principalele tipuri de studii întâlnite în domeniul medical sunt studiile experimentale, în care investigatorul are controlul complet asupra studiului, și studiile observaționale, în care investigatorul nu intervine în studiu.



Studiile observaționale includ studii ce se întind pe un interval mare de timp (studii longitudinale) și studii ce se desfășoară la un moment dat (studii transversale).

Aceste studii fac parte din categoria studiilor populaționale (studiile epidemiologice). O analiză des întâlnită în cadrul acestor studii este **analiza riscului**.

Factorul de risc reprezintă o cauză ipotetică ce determină creșterea probabilității de a prezenta o anumită afecțiune. Exemplele de factori de risc includ: factori de mediu (e.g. factori poluanți, toxine, microorganisme), factori comportamentali (e.g. alcool, "life style"), factori sociali (e.g. evenimente personale, șomaj, divorț), factori genetici.

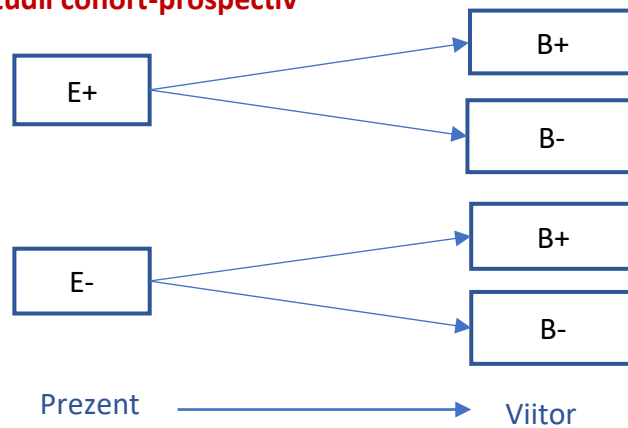
Tipurile de expunere: expunere punctuală, sau expunere cronică.

Analiza riscului studiază relația **factor de expunere (E) – boala (B)**.

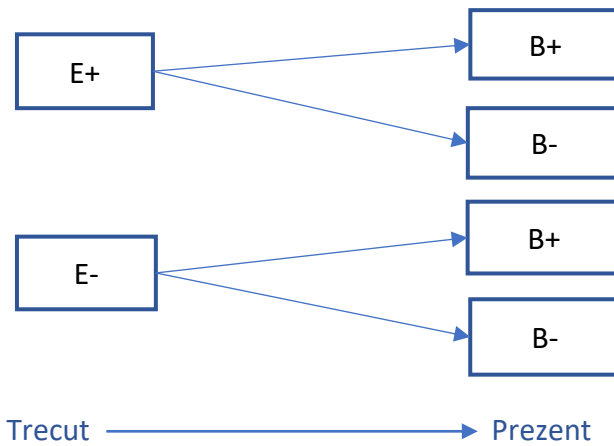
B. Studii observaționale

- interval mare de timp
- moment inițial (prezent) - niciun grup nu prezintă afecțiunea urmărită
- moment final (viitor) - se identifica cine prezintă boala

Studii cohort-prospectiv

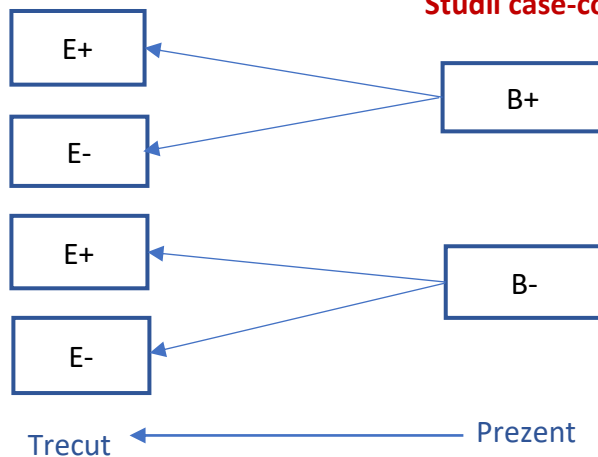


Studii cohort-retrospectiv



- interval mare de timp
- moment inițial (trecut) - niciun grup nu prezintă afecțiunea urmărită
- moment final (prezent) - se identifica cine prezintă boala

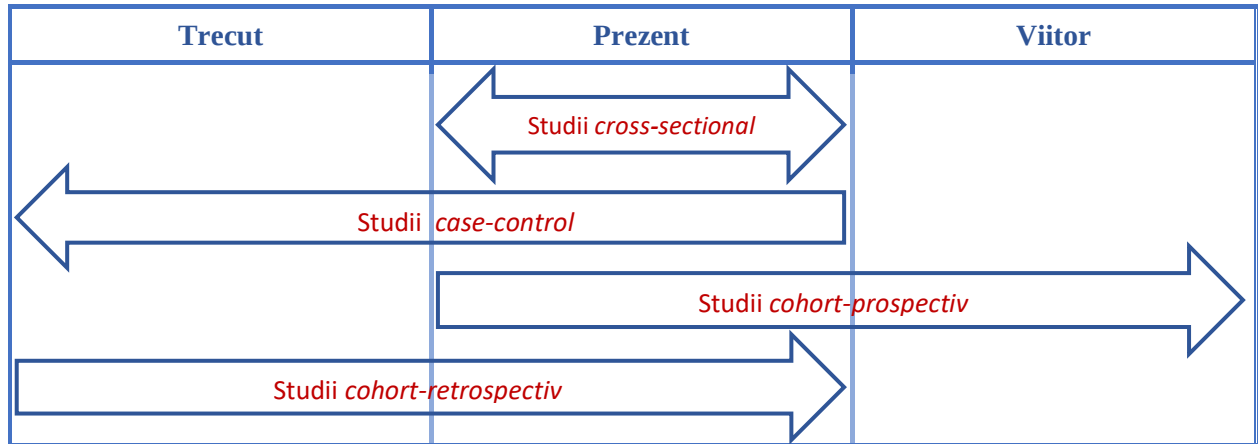
Studii case-control



- studiu retrospectiv clasic
- moment inițial (prezent) - exista un grup care prezintă afecțiunea urmărită
- moment final (trecut) - se identifica cine a fost expus la factorul studiat

Tăria studiilor:

experimental > cohort-prospectiv > cohort-retrospectiv > case-control > cross-sectional



Tabel de contingență

	Boala +	Boala -	Total
Expunere +	a	b	a+b
Expunere -	c	d	c+d
Total	a+c	b+d	a+b+c+d

Relative Risk (RR)

raportul între probabilitățile de a prezenta boala între două grupuri – unul expus și unul neexpus.

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

Odds Ratio (OR)

raportul între probabilitatea de a prezenta boala pentru un grup expus și probabilitatea de a prezenta boala pentru un grup neexpus.

$$OR = \frac{ODD(E+)}{ODD(E-)} = \frac{a/b}{c/d} = \frac{ad}{bc}$$

C. Interpretare - indicatori și intervale de confidență (IC) (n.î. 95%)

RR, OR > 1 factor de risc	limita inferioară IC _{RR} , IC _{OR} > 1 : factor de risc semnificativ
RR, OR = 1 factor indiferent	IC _{RR} , IC _{OR} conțin 1 : factor nesemnificativ
RR, OR < 1 factor de protecție	limita superioară IC _{RR} , IC _{OR} < 1 : factor nesemnificativ

II. Probleme propuse:

Problema 1

S-a observat că accidentele vasculare cerebrale (AVC) au loc cu preponderență la tinerii ce consumă droguri recreaționale¹. Pentru a investiga această problemă, ei au identificat toți pacienții cu vârste între 15 și 44 de ani admiși la un anumit spital și pacienții din grupul de control pentru care consumul de droguri recreaționale nu reprezenta un factor de risc. Datele sunt prezentate în tabelul 1.

Tabel 1. Datele pentru grupul cu AVC și grupul de control.

	AVC	Control
Abuz de droguri	73	18
Fără abuz de droguri	141	196
Total	214	214

Cerințe:

1. Este abuzul de droguri recreaționale un factor de risc pentru apariția AVC? Calculați OR.
2. Este factorul respectiv semnificativ statistic? Formulați ipotezele statistice. Aplicați testul statistic corespunzător.
3. Calculați intervalul de încredere pentru OR. Considerați un nivel de încredere de 95%.
4. Interpretați statistic intervalul obținut.

Problema 2

Un studiu *cross-sectional* își propune să analizeze dacă există o relație între fumat și anxietate. Pentru aceasta, un eșantion de 1000 de persoane au fost simultan clasificate în fumători și nefumători, precum și în funcție de nivelul de anxietate. Astfel s-a observat că 300 de persoane au fost identificate ca având un nivel ridicat de anxietate, 500 identificați drept fumători și 200 de persoane au fost identificate drept fumători și cu nivel ridicat de anxietate.

¹ Kaku DA, Lowenstein DH: Emergency of recreational drug abuse as a major risk factor for stroke in young adults. Ann Intern Med 1990;113:821–827.

Tabel 2. Datele cu privire la studiul legat de relația dintre fumat și anxietate

	Nivel de anxietate	
	Ridicat	Scăzut
Fumător	200	300
Nefumător	100	400
Total	300	700

Cerințe:

1. Care este prevalența în grupul de studiu a nivelurilor ridicate de anxietate.
2. Care este OR pentru un nivel ridicat de anxietate pentru fumători în comparație cu cei nefumători.
3. Interpretați rezultatul.

Problema 3

Un studiu își propune să investigheze posibila asociere între stilul de viață sedentar și apariția în timp a diabetului zaharat de tip 2 (DZ 2). Pentru aceasta, în studiu sunt înrolați 1200 de adulți (sedentari și nesedentari), care inițial nu prezintă DZ 2. Aceștia sunt urmăriți timp de 10 ani. La sfârșitul studiului, se constată că din cei 300 de sedentari, 50 s-au îmbolnăvit de DZ 2, iar în grupul nesedentariilor, s-au îmbolnăvit 70 de adulți.

Cerințe:

1. Precizați ce tip de studiu este cel prezentat.
2. Este stilul de viață sedentar un factor de risc pentru apariția DZ 2? Construiți tabelul de contingență. Calculați RR și OR.
3. Este factorul respectiv semnificativ statistic? Formulați ipotezele statistice. Aplicați testul statistic corespunzător.
4. Calculați intervalul de încredere pentru RR și OR. Considerați un nivel de încredere de 95%.
5. Interpretați statistic intervalul obținut.

Presupuneți că în studiu sunt înrolați 120 de adulți, dintre care 30 sunt sedentari. Din rândul sedentariilor se îmbolnăvesc 5 adulți, iar dintre nesedentari 7.

6. Calculați în acest caz RR și OR.
7. Este riscul calculat semnificativ statistic?
8. Găsiți o justificare pentru noile rezultate obținute.

In continuare, rezolvările vor fi prezentate sub forma de explicații scrise cu capturi de ecran, dar aveți disponibile și rezolvările în format de înregistrare video.

III. Aplicarea testelor statistice cu ajutorul aplicațiilor software:

D. EpiInfo

Calcul RR si OR

Aplicarea testului X^2

Se folosește comanda *STATCALC* -> *2x2 Tables*; *Exposure* - [variabila factorului de expunere]; *Outcome* - [variabila ce exprima afecțiunea/boala studiată].

E. OpenEpi <https://www.openepi.com>

Calcul RR si OR

Aplicarea testului X^2

Se alege opțiunea *Counts* -> *Two by Two Table*

Pentru completarea tabelului de contingenta, se alege **Enter New Data**. Se completează tabelul si, la final, se alege comanda **Calculate**.

IV. Soluții:

Problema 1

1. Identificam variabilele din problema: *factorul de expunere* este consumul de droguri recreaționale, care este o variabila categorie cu doua categorii (variabila dihotomica), si *boala* este reprezentata de apariția AVC, care este tot o variabila categorie cu doua categorii (variabila dihotomica).

Calcul OR

Varianta EpiInfo

Se folosește comanda *STATCALC* -> *2x2 Tables*; *Exposure* - Abuz de droguri; *Outcome* - apariția AVC.

		Outcome	
		Yes	No
Exposure	Yes	73	18
	No	141	196

Odds-based and Risk-based Parameters

	Estimate	Lower	Upper
Odds Ratio	5.6375	3.2222	9.8632
MLE OR	5.6151	3.2508	10.0501
Fisher Exact		3.1506	10.4602
Relative Risk	1.9173	1.6305	2.2546
Risk Difference	38.38	28.6473	48.1127

Observăm ca $OR = 5.6375 > 1$, prin urmare abuzul de droguri reprezintă un factor de risc pentru apariția AVC.

Varianta OpenEpi

Se accesează aplicația web OpenEpi la adresa: <https://www.openepi.com>

Din mediul din stânga, se alege comanda *Counts -> Two by Two Table*

[Expand All](#) | [Collapse](#)

- Home
- Info and Help
- Language/Options/Settings
- Calculator
- Counts
 - Std.Mort.Ratio
 - Proportion
 - Two by Two Table
 - Dose-Response
 - R by C Table
 - Matched Case Control
 - Screening
- Person Time
 - 1 Rate
 - Compare 2 Rates
- Continuous Variables
 - Mean CI
 - Median/%ile CI
 - t test
 - ANOVA

Start	Enter	Results	Examples	Help
Open Source Statistics for Public Health				
Enter New Data				
Add Stratum Stratum 1 Delete Stratum				
Open Epi 2 x 2 Table				
		'Exposure'		Totals
		(+)	(-)	
'Disease'	(+)	66	36	102
	(-)	28	32	60
Totals		94	68	162

Pentru completarea tabelului de contingenta, se alege **Enter New Data**. Se completează tabelul si se alege **Calculate**.

Start	Enter	Results	Examples	Help
--------------	--------------	----------------	-----------------	-------------

[Settings](#)

Open Epi 2 x 2 Table				
	Disease			Totals
		(+)	(-)	
Exposure	(+)	73	18	91
	(-)	141	196	337
Totals		214	214	428

Odds-Based Estimates and Confidence Limits

Point Estimates		Confidence Limits
Type	Value	Lower, Upper
CMLE Odds Ratio*	5.615	3.251, 10.05 ¹
		3.151, 10.46 ¹
Odds Ratio	5.638	3.222, 9.863 ¹
Etiologic fraction in pop.(EF _p OR)	28.06%	20.54, 35.59
Etiologic fraction in exposed(EF _e OR)	82.26%	68.97, 89.86

2. Pentru a studia daca factorul de risc este semnificativ (daca poate fi generalizat la nivel de populație), vom formula mai întâi ipotezele statistice:

H0: abuzul de droguri nu reprezintă un factor de risc semnificativ pentru apariția AVC.

Ha: abuzul de droguri reprezintă un factor de risc semnificativ pentru apariția AVC.

a) Aplicam testul Chi-pătrat (X^2)

Varianta EpiInfo

Statistical Tests		
	X2	2 Tailed P
Uncorrected	42.218	0
Mantel-Haenszel	42.1194	0
Corrected	40.6968	0

	1 Tailed P	2 Tailed P
Mid P Exact	0	
Fisher Exact	0	0

Observație: valoarea lui p nu este 0, dar având mai mult de 5 zecimale de 0 după virgula, s-a afișat 0.

Varianta OpenEpi

Chi Square and Exact Measures of Association			
Test	Value	p-value(1-tail)	p-value(2-tail)
Uncorrected chi square	42.22	<0.000001	<0.000001
Yates corrected chi square	40.7	<0.000001	<0.000001
Mantel-Haenszel chi square	42.12	<0.000001	<0.000001
Fisher exact		<0.000001	<0.000001
Mid-P exact		<0.000001	<0.000001

All expected values (row total*column total/grand total) are ≥ 5
OK to use chi square.

Observăm că valoarea lui p (2-tailed, testul X^2 necorectat) este < 0.001 , prin urmare vom respinge ipoteza inițială și acceptăm ipoteza alternativă; deci abuzul de droguri reprezintă un factor de risc extrem semnificativ pentru apariția AVC.

b) Interpretăm intervalul de încredere pentru OR (nivel de încredere de 95%).

Intervalul de confidență este $IC_{OR} = (3.222; 9.863)$. Observăm că valoarea 1 nu aparține intervalului de confidență ($1 \notin IC_{OR}$), și $IC_{OR} > 1$, prin urmare factorul de expunere este un factor de risc semnificativ.

3. Concluzie: abuzul de droguri reprezintă un factor de risc pentru apariția AVC.

Problema 2

Identificăm variabilele din problema: *factorul de expunere* este fumatul, care este o variabilă categorică cu două categorii (fumător și nefumător - variabilă dihotomică), și *boala* este reprezentată de nivelul crescut de anxietate, care este tot o variabilă categorică cu două categorii (variabilă dihotomică).

Calcul OR

Varianta EpiInfo

Vom folosi comanda *STATCALC* -> *2x2 Tables*; *Exposure* - Fumat; *Outcome* - Nivel de anxietate.

		Outcome	
		Yes	No
Exposure	Yes	200	300
	No	100	400

Odds-based and Risk-based Parameters

	Estimate	Lower	Upper
Odds Ratio	2.6667	2.0096	3.5386
MLE OR	2.664	2.0103	3.5433
Fisher Exact		1.9914	3.5787
Relative Risk	2	1.6284	2.4564
Risk Difference	20	14.4563	25.5437

Varianta OpenEpi

Se accesează aplicația web OpenEpi la adresa: <https://www.openepi.com>

Din mediul din stânga, se alege comanda *Counts -> Two by Two Table*.

Pentru completarea tabelului de contingenta, se alege **Enter New Data**. Se completează tabelul și se alege **Calculate**.

Start	Enter	Results	Examples	Help
--------------	--------------	----------------	-----------------	-------------

[Settings](#)

Open Epi 2 x 2 Table				
	Disease			Totals
		(+)	(-)	
Exposure	(+)	200	300	500
	(-)	100	400	500
Totals		300	700	1000

Odds-Based Estimates and Confidence Limits

Point Estimates		Confidence Limits
Type	Value	Lower, Upper
CMLE Odds Ratio*	2.664	2.01, 3.543 ¹
		1.991, 3.579 ¹
Odds Ratio	2.667	2.01, 3.539 ¹
Etiologic fraction in pop.(EF _p OR)	41.67%	31.61, 51.72
Etiologic fraction in exposed(EF _e OR)	62.5%	50.24, 71.74

Observăm ca $OR = 2.667 > 1$, prin urmare fumatul reprezintă un factor de risc pentru un nivel ridicat de anxietate.

2. Pentru a studia dacă factorul de risc este semnificativ (dacă poate fi generalizat la nivel de populație), vom formula mai întâi ipotezele statistice:

H0: fumatul nu reprezintă un factor de risc semnificativ pentru un nivel ridicat de anxietate.

Ha: fumatul reprezintă un factor de risc semnificativ pentru un nivel ridicat de anxietate.

a) Aplicăm testul Chi-pătrat (X^2)

Varianta EpiInfo

Statistical Tests		
	X2	2 Tailed P
Uncorrected	47.619	0
Mantel-Haenszel	47.5714	0
Corrected	46.6714	0

	1 Tailed P	2 Tailed P
Mid P Exact	0	
Fisher Exact	0	0

Varianta OpenEpi

Chi Square and Exact Measures of Association

Test	Value	p-value(1-tail)	p-value(2-tail)
Uncorrected chi square	47.62	<0.0000001	<0.0000001
Yates corrected chi square	46.67	<0.0000001	<0.0000001
Mantel-Haenszel chi square	47.57	<0.0000001	<0.0000001
Fisher exact		<0.0000001	<0.0000001
Mid-P exact		<0.0000001	<0.0000001

All expected values (row total*column total/grand total) are ≥ 5
OK to use chi square.

Observăm că valoarea lui p (2-tailed, testul X^2 necorectat) este < 0.001 , prin urmare vom respinge ipoteza inițială și acceptăm ipoteza alternativă; deci fumatul reprezintă un factor de risc extrem semnificativ pentru un nivel crescut de anxietate.

b) Interpretăm intervalul de încredere pentru OR (nivel de încredere de 95%).

Intervalul de confidență este $IC_{OR} = (2.010; 3.539)$. Observăm că valoarea 1 nu aparține intervalului ($1 \notin IC_{OR}$), și $IC_{OR} > 1$, prin urmare factorul de expunere este un factor de risc semnificativ.

3. Concluzie: fumatul reprezintă un factor de risc pentru un nivel ridicat de anxietate.

Problema 3

1. Studiul prezentat este de tip cohort-prospectiv deoarece studiul se desfășoară pe o perioadă lungă de timp (10 ani), iar la începutul studiului sunt două grupuri, unul expus la factorul studiat și celălalt neexpus, iar niciun grup nu prezintă DZ2.

2. Pentru a construi tabelul de contingență, mai întâi, identificăm variabilele din problemă: *factorul de expunere* este stilul de viață sedentar, care este o variabilă categorie cu două categorii (variabilă dihotomică), și *boala* este reprezentată de prezenta DZ2, care este tot o variabilă categorie cu două categorii (variabilă dihotomică).

	DZ2 +	DZ2 -	Total
Stil de viață sedentar +	a	b	a+b
Stil de viață sedentar -	c	d	c+d
Total	a+c	b+d	a+b+c+d

Calcul RR și OR

	DZ2 +	DZ2 -	Total
Stil de viață sedentar +	50	250	300
Stil de viață sedentar -	70	830	900
Total	120	1080	1200

$a + b = 300$
 $c + d = 900$

Grupul sedentarelor : $a + b = 300$, din aceștia 50 s-au îmbolnăvit de DZ2 $\Rightarrow a = 50$ și $b = 250$

Grupul nesedentarelor: $c + d = 900$, din aceștia 70 s-au îmbolnăvit de DZ2 $\Rightarrow c = 70$ și $d = 830$

Varianta EpiInfo

Se folosește comanda *STATCALC* -> *2x2 Tables*; *Exposure* - stil de viață sedentar; *Outcome* - apariția DZ2.

		Outcome	
Exposure		50	250
		70	830

Odds-based Parameters			
	Estimate	Lower	Upper
Odds Ratio	2.3714	1.6064	3.5008
MLE OR	2.3694	1.5988	3.4959
Fisher Exact		1.5697	3.5577

Risk-based Parameters			
	Estimate	Lower	Upper
Relative Risk	2.1429	1.5274	3.0063
Risk Difference	8.8889	4.3231	13.4547

Varianta OpenEpi

Risk-Based* Estimates and 95% Confidence Intervals
(Not valid for Case-Control studies)

Point Estimates		Confidence Limits	
Type	Value	Lower, Upper	Type
Risk in Exposed	16.67%	12.86, 21.32	Taylor series
Risk in Unexposed	7.778%	6.193, 9.721	Taylor series
Overall Risk	10%	8.423, 11.83	Taylor series
Risk Ratio	2.143	1.527, 3.006 ¹	Taylor series
Risk Difference	8.889%	4.323, 13.45°	Taylor series
Etiologic fraction in pop.(EFp)	22.22%	11.12, 33.32	
Etiologic fraction in exposed(EFe)	53.33%	34.53, 66.74	

Odds-Based Estimates and Confidence Limits

Point Estimates		Confidence Limits	
Type	Value	Lower, Upper	
CMLE Odds Ratio*	2.369	1.599, 3.496 ¹	
		1.57, 3.558 ¹	
Odds Ratio	2.371	1.606, 3.501 ¹	
Etiologic fraction in pop.(EF _p OR)	24.1%	12.35, 35.84	
Etiologic fraction in exposed(EF _e OR)	57.83%	37.75, 71.43	

Observam ca $RR = 2.143$ si $OR = 2.371$; $RR, OR > 1$, prin urmare stilul de viață sedentar reprezintă un factor de risc pentru apariția DZ2.

3. Pentru a studia daca factorul de risc este semnificativ (daca poate fi generalizat la nivel de populație), vom formula mai întâi ipotezele statistice:

H_0 : stilul de viață sedentar nu reprezintă un factor de risc semnificativ pentru apariția DZ2.

H_a : stilul de viață sedentar reprezintă un factor de risc semnificativ pentru apariția DZ2.

a) Aplicam testul Chi-pătrat (X^2)

Varianta EpiInfo

Statistical Tests		
	X2	2 Tailed P
Uncorrected	19.7531	0
Mantel-Haenszel	19.7366	0
Corrected	18.7778	0
	1 Tailed P	2 Tailed P
Mid P Exact	0	
Fisher Exact	0	0

Varianta OpenEpi

Chi Square and Exact Measures of Association			
Test	Value	p-value(1-tail)	p-value(2-tail)
Uncorrected chi square	19.75	0.000004406	0.000008812
Yates corrected chi square	18.78	0.000007343	0.00001469
Mantel-Haenszel chi square	19.74	0.000004444	0.000008888
Fisher exact		0.00001726	0.00003453
Mid-P exact		0.00001211	0.00002422

Observam ca valoarea lui p (2-tailed, testul X^2 necorectat) este < 0.001 , prin urmare vom respinge ipoteza inițială și acceptăm ipoteza alternativă; deci stilul de viață sedentar reprezintă un factor de risc extrem semnificativ pentru apariția DZ2.

b) Interpretăm intervalul de încredere pentru OR (nivel de încredere de 95%).

Intervalul de confidență pentru RR este $IC_{RR} = (1.527; 3.006)$, iar pentru OR este $IC_{OR} = (1.527; 3.006)$. Observăm că valoarea 1 nu aparține intervalelor ($1 \notin IC_{RR}$, $1 \notin IC_{OR}$), $IC_{RR} > 1$ și $IC_{OR} > 1$, prin urmare factorul de expunere (stilul de viață sedentar) este un factor de risc semnificativ.

Concluzie: stilul viață sedentar reprezintă un factor de risc pentru apariția DZ2.

Partea a doua a problemei:

Noul tabel de contingență este:

	DZ2 +	DZ2 -	Total
Stil de viață sedentar +	5	25	30
Stil de viață sedentar -	7	83	90
Total	12	108	120

Observăm că sunt aceleași valori, de 10 ori mai mici. În acest caz, RR și OR sunt:

Outcome	
Exposure	5
	25
7	83

Odds-based Parameters			
	Estimate	Lower	Upper
Odds Ratio	2.3714	0.692	8.127
MLE OR	2.3517	0.6323	8.2776
Fisher Exact		0.5393	9.4988

Risk-based Parameters			
	Estimate	Lower	Upper
Relative Risk	2.1429	0.7345	6.2516
Risk Difference	8.8889	-5.5496	23.3273

Observăm ca $RR = 2.143$ și $OR = 2.371$; $RR, OR > 1$, prin urmare stilul de viață sedentar reprezintă un factor de risc pentru apariția DZ2.

Studiem dacă factorul de risc este semnificativ (dacă poate fi generalizat la nivel de populație).

Statistical Tests		
	X2	2 Tailed P
Uncorrected	1.9753	0.1599
Mantel-Haenszel	1.9588	0.1616
Corrected	1.1111	0.2918
	1 Tailed P	2 Tailed P
Mid P Exact	0.0956	
Fisher Exact	0.1461	0.1721

Observăm că valoarea lui p (2-tailed, testul X^2 necorectat) este 0.16 (> 0.05), prin urmare vom accepta ipoteza inițială și respingem ipoteza alternativă; deci stilul de viață sedentar nu reprezintă un factor de risc semnificativ pentru apariția DZ2.

Justificare: noile rezultate nu mai sunt semnificative statistic (deci nu mai pot fi generalizate la nivel de populație) din cauza numărului redus de persoane cuprinse în studiu.