

## Epidemiologia analitică

**Studiile analitice** investighează relația dintre două categorii de **variabile**, **factorul(ii) de risc** și efectul (boală, complicație, deces), căutând să evidențieze eventualele asocieri dintre acestea. Practic, pentru un studiu analitic trebuie să existe întotdeauna un grup martor – fie că este cohorta martor, neexpusă, din **studiile de cohortă și experimentale**, fie că este grupul care nu a suferit efectul din **studiile caz-martor**). De fapt, analiza din aceste studii o putem descrie în două moduri, fie ca evidențiere a unei asociații epidemiologice (între factorul de risc și efect), fie ca evidențiere a unei diferențe, între grupul de cercetat și grupul martor.

Exemple. Studiu de cohortă în care vrem să demonstrăm că diabetul este factor de risc pentru apariția infarctului de miocard – o cohortă de indivizi cu diabet = expuși și una de indivizi fără diabet = cohorta martor. Dovedim că diabetul zaharat și infarctul de miocard sunt asociate prin demonstrarea faptului că există o diferență între frecvența apariției (incidența, riscul) infarctului de miocard între cohorta de diabetici și cea martor.

Studiu caz-martor prin care vrem să arătăm același lucru – cazurile, care sunt indivizi internați cu infarct miocardic acut și martorii, care sunt indivizi asemănători ca vârstă, sex, alți factori de risc, doar că nu au infarct miocardic. Pe toți îi evaluăm în privința prezenței diabetului. Dovedim asocierea dintre diabet și infarctul de miocard prin demonstrarea faptului că există o diferență în ceea ce privește frecvența prezenței (prevalența) diabetului zaharat la cazuri față de martori.

În urma acestor studii vor rezulta niște măsuri ale **forței asocierii** dintre factorul de risc și efect, reprezentate prin **riscul relativ** (după studiile de cohortă) sau **odds ratio** (după studiile caz-martor), atunci când **variabilele** studiate (factorul de risc și efectul) sunt nominale, dihotomice. Dacă variabilele a căror asociere vrem să o evidențiem sunt numerice, măsura asocierii va fi redată prin mărimea **coeficientului de corelație** (pentru detalii privind variabilele, vezi Capitolul 11. Pregătirea, prezentarea, analiza și interpretarea rezultatelor).

Înafara **forței asocierii**, studiile vor măsura și impactul expunerii, prin intermediul **riscului atribuibil** (la studiile observaționale de cohortă), sau al **reducerii absolute a riscului** (la studiile experimentale).

Forța asocierii (redată prin rapoarte – **risc relativ**, **odds ratio**, **reducerea relativă a riscului**) reprezintă un argument cu privire la etiologie, pe când impactul expunerii (redat prin diferențe – **risc atribuibil**, **reducerea absolută a riscului**) este un argument privind latura de sănătate publică a asocierii respective.

Pentru fiecare dintre parametrii care cuantifică asociația epidemiologică (risc relativ, **odds ratio**, coeficient de corelație, reduceri relativă sau absolută a riscului etc.) avem o semnificație statistică, exprimată ca  $p$  și/sau interval de încredere (vezi Capitolul 11. Pregătirea, prezentarea, analiza și interpretarea rezultatelor).

Trebuie precizate și definite încă dinaintea începerii studiului (în protocol): obiectul investigației, populația și cele două tipuri de variabile (expunerea = variabila independentă și efectul = variabila dependentă), precum și modul de măsurare al variabilelor.

## Studiile de cohortă

În epidemiologie, „cohorta” este un grup de indivizi desemnat pe baza unor caracteristici comune, care este urmărit de-a lungul timpului. Studiul de cohortă, care este arhetipul tuturor studiilor epidemiologice, implică măsurarea apariției efectului (bolii) în două sau mai multe cohorte (dacă este urmărită o singură cohortă, studiul este o **serie de cazuri**, și este de tip descriptiv). Așadar, evenimentele se produc exact ca în realitate, pornind de la expunere și ajungând la efect și din acest motiv studiile de cohortă sunt singurele care permit calcularea incidenței bolii, și deci previziunea riscului.

Sunt două scopurile studiilor de cohortă, unul **descriptiv** (descrierea incidenței unor efecte = boli, complicații) de-a lungul timpului, iar celălalt **analitic** (analiza asocierilor dintre **factori de risc** sau **factori prognostici** și efectele respective).

Studiile de cohortă se mai numesc și studii de urmărire (*follow-up*), de incidență, longitudinale, etiologice, prospective. În privința ultimelor două denumiri, pe de o parte mai există și alte studii al căror scop este să caute etiologia (studiile caz-martor) - deși cele de cohortă sunt cele mai valide, iar pe de altă parte există și studii de cohortă retrospective.

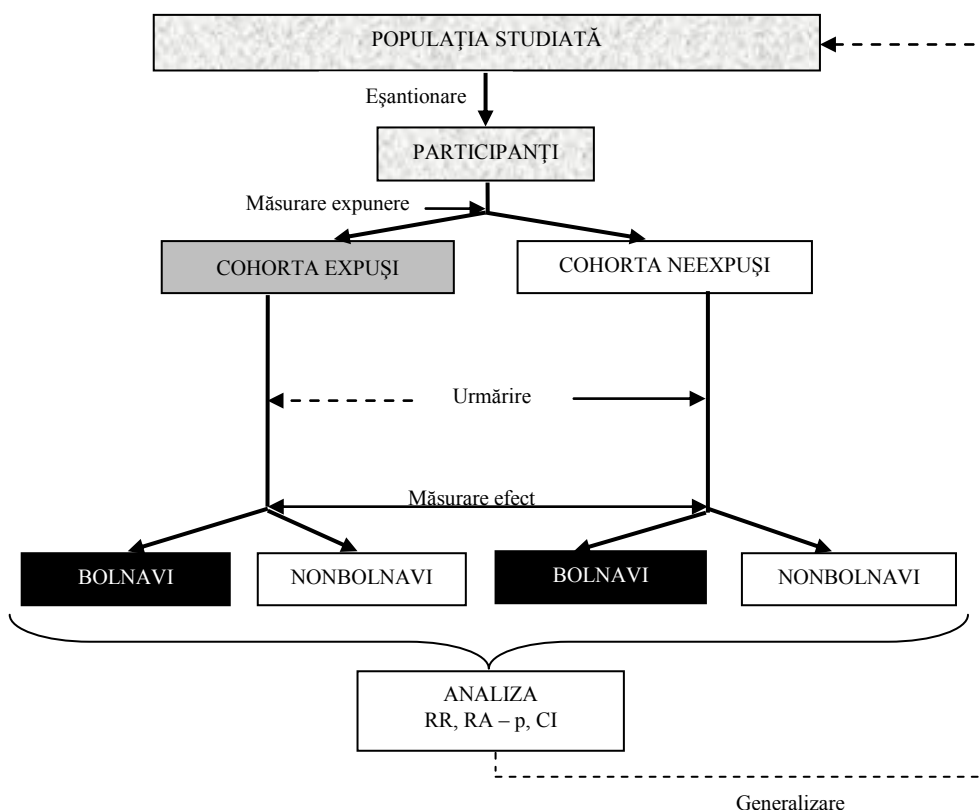
În studiile de cohortă, investigatorul selecționează un grup (cohortă) de indivizi expuși și unul (cohortă) de neexpuși, urmărind ambele grupuri pentru a compara incidența bolii (sau mortalitatea din cauza bolii).

Dacă există o asociere pozitivă între expunere și boală, ne așteptăm ca incidența bolii să fie mai mare în grupul expus decât în cel neexpus.

În Figura .4 este prezentată schema unui **studiu de cohortă de tip I**. În acest tip de studiu, factorul de risc este frecvent răspândit în populație, astfel încât pentru studiul nostru includem indivizi din populația generală la care evaluăm prezența/absența factorului de risc, pe baza căreia împărțim populația în două cohorte, de expuși și de martori.

De exemplu, vrem să evaluăm fumatul ca factor de risc pentru apariția bolii arteriale periferice. Selecționăm prin tragere la sorți un eșantion de indivizi, pe care-i întrebăm apoi dacă fumează. Pe baza răspunsului la întrebare, din eșantionul nostru se vor autoselecționa două cohorte, una de fumători (expuși) și una de nefumători (martori).

Figura .4. Schema unui studiu de cohortă de tip I.



La început trebuie să ne asigurăm, pe de o parte, că indivizii incluși în studiu nu au suferit deja efectul urmărit (nu vom include indivizi care deja au boală arterială periferică), iar pe de altă parte că sunt la risc, adică pot suferi efectul (nu vom include indivizi care au membrele inferioare amputate).

După aceasta urmează perioada de urmărire, care trebuie să fie suficient de lungă astfel încât fumatul, dacă este implicat în etiologia arteriopatiei, să aibă timp să-și producă efectul - studiile de cohortă se întind, de obicei, pe durata multor ani, de obicei peste 10 (în cazul Framingham, studiul continuă și acum, la 60 de ani, asupra descendenților celor incluși inițial).

La sfârșitul perioadei studiului, îi numărăm pe cei care au suferit efectul (care au arteriopatie) din ambele cohorte, iar datele le prelucrăm conform Tabelului .5, calculând **riscul relativ** și verificând semnificația statistică a asociației epidemiologice (în acest exemplu fiind vorba despre variabile calitative, dihotomice – fumează sau nu, au sau nu arteriopatie - vom aplica un test statistic tip  $X^2$ ).

**Studiul de cohortă de tip II** se efectuează atunci când prevalența factorului de risc nu este îndeajuns de mare încât din eșantionul inițial să rezulte o cohortă destul de numeroasă de expuși, sau când expunerea afectează numai anumite categorii profesionale. Să presupunem că vrem să evidențiem asociația anticoncepțională orale - tromboembolism.

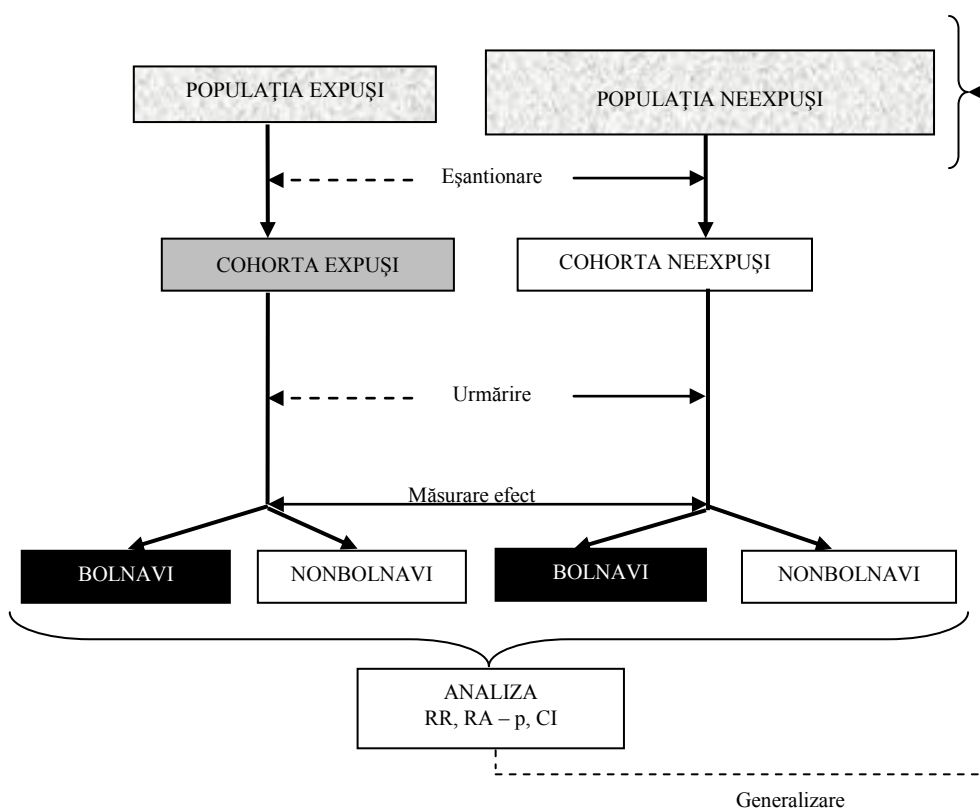


Figura .5. Schema unui studiu de cohortă de tip II

Dacă studiul s-ar desfășura într-o țară dezvoltată, am putea face un studiu de tip I, pentru că ar exista destule utilizatoare de anticoncepționale, în România însă trebuie să selecționăm, separat, o cohortă de utilizatoare de anticoncepționale (vare vor fi din mediul urban, cu studii măcar medii) și o cohortă de femei care nu folosesc anticoncepționale. La fel cu expunerile care afectează anumite populații sau categorii profesionale – pentru radiațiile ionizante, de exemplu, se poate lua cohorta de locuitori ai orașului Hiroshima ca expuși, și o altă populație ca martori; sau expunerea la un toxic profesional, o cohortă reprezentată de lucrătorii din secția care folosește aceste toxice ca expuși și pe cei din altă secție sau altă fabrică îi vom lua ca și cohortă martor (Figura .5).


În rest, studiul de cohortă de tip II se desfășoară, iar rezultatele sunt analizate la fel ca la studiul de tip I. Studiile de tip II sunt, însă, mai expuse factorilor de confuzie, pentru că

expușii și martorii nu provin din aceeași populație, și atunci pot exista și alte diferențe între grupuri înafara expunerii.

Să presupunem că vrem să studiem consumul de anticoncepționale orale ca și factor de risc pentru tromboza venoasă profundă. Pentru aceasta, selectăm o cohortă de utilizatoare din mediul urban și o cohortă martor, de femei care nu le utilizează, din mediul rural. După ce le urmărim timp de câțiva ani, constatăm că incidența trombozei la utilizatoare este de câteva ori mai mare față de neutilizatoare. Problema este că sunt mult mai multe diferențe între cele două cohorte decât utilizarea sau nu de anticoncepționale – diferențe de alimentație, de mod de viață etc. Să ne gândim numai la faptul că utilizatoarele, femei de la oraș cu studii măcar medii, lucrează la birou, se deplasează cu autoturismul propriu și nu fac mișcare, în timp ce cele neexpușe lucrează în agricultură. Este, așadar, posibil ca anticoncepționalele să nu fie adevăratul factor de risc al trombozei, chiar dacă a rezultat în urma studiului o asociație epidemiologică semnificativă, iar adevăratul factor de risc să fie modul de viață (în acest caz, modul de viață este un **factor de confuzie**).

**Măsurarea expunerii** se face cu ajutorul chestionarelor, măsurătorilor clinice și paraclinice, foilor de observație și a registrelor medicale.

**Măsurarea efectului (bolii)** se face identic la expuși și la neexpuși (procedurile de identificare a bolii trebuie să fie aceleași), ideal practicându-se „orbirea”. Oricum, trebuie definite detaliat efectele de interes (criteriile diagnostice), înainte de începerea studiului, iar dacă se studiază efecte multiple, trebuie definit fiecare.

		EFECT (BOALĂ)		
		DA	NU	
Factor de risc	DA	a	b	a+b
	NU	c	d	c+d
Direcția studiului 				

Tabelul .5. Tabel de contingență 2 x 2 reprezentând un studiu de cohortă: (a+b) expuși comparați cu (c+d) neexpuși.

Luând datele din Tabelul .5, a+b este cohorta expusă la factorul de risc, c+d este cohorta neexpusă,  $a/(a+b)$  este incidența evenimentului la expuși (= riscul la expuși),  $c/(c+d)$  este incidența (riscul) la neexpuși, iar **riscul relativ** (RR) este raportul dintre cele două riscuri: riscul la expuși (R<sub>exp</sub>) / riscul la neexpuși (R<sub>nexp</sub>),

$$RR = \frac{a/(a+b)}{c/(c+d)} = \frac{a(c+d)}{c(a+b)} \quad (1).$$

**Riscul atribuibil** este diferența dintre riscul la expuși și riscul la neexpuși, adică:

$$RA = \frac{a}{a+b} - \frac{c}{c+d} \quad (2),$$

și ne arată cu cât este mai mare riscul la expuși față de neexpuși, sau cât din risc este datorat (atribuibil) expunerii.

Exemplu: Sunt urmărite două cohorte de câte 200 de bărbați de 40-50 de ani, fără boală coronariană, una de expuși (diabet zaharat tip II) și una de neexpuși (nondiabetici). După 10 ani se constată că în cohorta de diabetici, 20 au boală coronariană, iar în cohorta fără diabet, 12 au boală coronariană.

După cum se vede în Figura .6, la sfârșitul studiului vom avea în cohorta cu diabet 20 de bărbați care au făcut boală coronariană și 180 de bărbați (200-20) care nu au făcut, iar în cohorta martor vom avea 12 bărbați care au făcut boala coronariană și 188 (200-12) care nu au făcut. **Incidența** bolii coronariene la diabetici (**riscul** la expuși) este de 0,10 (sau 10%), în timp ce **incidența** la nondiabetici (**riscul** la neexpuși) este de 0,06 (6%), așadar **riscul relativ** = 10/6 = 1,66, care se interpretează ca riscul de a face boală coronariană este de 1,6 ori mai mare la diabetici decât la nondiabetici.

**Riscul atribuibil** este riscul la expuși minus riscul la neexpuși, adică 10-6=4%, ceea ce se interpretează ca 4% dintre bărbații de 40-50 de ani fac coronaropatie din cauza diabetului.

După cum am mai spus, **riscul relativ** (RR) ne arată de câte ori este mai mare riscul la expuși față de neexpuși, și este o măsură a forței asocierii dintre factorul de risc (expunere) și efect (boală). Logic, cu cât este mai mare RR, cu atât argumentul că

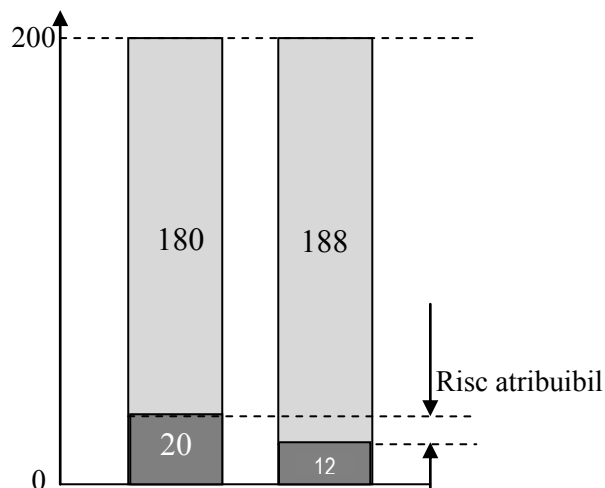


Figura .6. Studiu pe două cohorte a câte 200 de persoane (exemplul din text)

expunerea și efectul sunt asociate este mai puternic (de altfel, mărimea RR poate fi un argument și în favoarea relației de cauzalitate dintre expunere și efect).

**Riscul atribuibil**, în schimb, ne arată câți indivizi vom salva de la efectul urmărit dacă eradicăm expunerea (în exemplul de mai sus, câți bărbați am salva de boală coronariană dacă am eradica diabetul...). Cu cât riscul atribuibil este mai mare, cu atât efectul evitării expunerii este mai mare.

[Dacă **riscul** este probabilitatea de a suferi efectul, **cota (odds)** este raportul dintre probabilitatea de a suferi efectul și probabilitatea de a nu-l suferi. Așadar,  $\text{riscul} = \text{probabilitatea}$ , iar  $\text{cota (odds)} = \frac{\text{probabilitatea}}{1 - \text{probabilitatea}}$ . Luând exemplul de mai sus (Figura .6), cota la expuși este  $\frac{20}{200-20} = 0,111$ , iar cota la neexpuși este  $\frac{12}{200-12} = 0,064$ . Comparând cotele cu riscurile, se vede cum cotele sunt mai mari (la expuși,  $\text{riscul} = 0,10$ ,  $\text{cota} = 0,111$ , la neexpuși  $\text{riscul} = 0,06$ ,  $\text{cota} = 0,064$ ), iar explicația este cât se poate de simplă, aritmetică: pentru același numărător, întotdeauna numitorul este mai mic la cotă decât la risc, iar diferența dintre cotă și risc crește cu cât incidența (prevalența) bolii este mai mare, pentru că aceasta este de fapt diferența dintre numitori la cele două fracții. Iar **raportul cotelor (odds ratio)** este  $\frac{0,111}{0,064} = 1,73$ , mai mare decât **riscul relativ** (1,66). Așadar, după cum se vede, cota este întotdeauna mai mare decât riscul, iar **raportul cotelor (odds ratio)** este întotdeauna mai mare decât **riscul relativ**.

Dacă facem calculul cotelor și al raportului acestora din tabelul de contingență (Tabelul .5), atunci cota la expuși este  $a/b$ , cota la neexpuși este  $c/d$ , iar raportul cotelor este

$$OR = \frac{a/b}{c/d} = \frac{ad}{bc} \quad (3).$$

Atenție! În studiile de cohortă, unde se poate calcula direct incidența bolii = riscul, se folosește întotdeauna riscul relativ și nu avem de ce să calculăm *odds ratio*, care ne folosește pentru estimarea riscului relativ atunci când nu îl putem calcula. Am vorbit însă despre cote și raportul cotelor la studiile de cohortă pentru a înțelege cum se calculează, și care sunt diferențele conceptuală și aritmetică între riscuri și cote.]

Exemple de studii de cohortă:

Exemplul 1: Unul dintre cele mai cunoscute studii de cohortă este Studiul Framingham privind bolile cardiovasculare, care a început în 1948. Framingham este un oraș din Massachussetts, la distanță de aproximativ 40 de kilometri de Boston (unde se află Universitatea Harvard). Cum cea mai mare problemă a studiilor de cohortă este urmărirea fără prea mulți pierduți din vedere, s-a considerat că această localitate cu de 30.000 de locuitori este cea mai potrivită, deoarece acești locuitori nu erau foarte mobili: pe de o parte avea destule locuri de muncă în industria locală pentru a nu trebui să plece aiurea ca să-și câștige pâinea, iar

pe de lată parte aveau un spital local unde se internau în momentul în care sufereau unul dintre efectele urmărite (infarct miocardic, accident vascular cerebral), astfel încât datele puteau fi ușor strânse.

Locuitorii au fost considerați eligibili dacă aveau între 30 și 60 de ani (cei mai tineri aveau puține șanse să sufere efectul în cei 20 de ani, cât fusese proiectat studiul să dureze, iar cei peste 62 de ani aveau în majoritate deja boală coronariană. Investigatorii au căutat un eșantion de 5.000 de indivizi, și au avut în cele din urmă 5.127 de bărbați și femei între 30 și 62 de ani, fără boală cardiovasculară la intrarea în studiu (s-a extras un eșantion de 6.507 indivizi, dintre care au acceptat să participe 4.469; la aceștia s-au adăugat 740 de voluntari, în total 5.209; la numărul final de 5.127 s-a ajuns după evaluarea inițială și excluderea celor care aveau deja boală coronariană). În acest studiu au fost evaluate multe expuneri, printre care fumatul, obezitatea, hipertensiunea, nivelul ridicat al colesterolului, nivelul scăzut al activității fizice etc. Evenimentele coronariene noi au fost identificate prin examinarea populației studiului la fiecare 2 ani și prin supravegherea zilnică a internărilor la singurul spital din Framingham. Studiul a fost proiectat pentru a testa mai multe ipoteze (câte expuneri): incidența bolii coronariene crește cu vârsta și apare mai devreme și mai frecvent la bărbați; persoanele cu hipertensiune au risc mai mare de a dezvolta boală coronariană decât normotensivii; nivelul crescut al colesterolului seric este asociat cu un risc crescut de boală coronariană; fumatul și consumul habitual de alcool sunt, de asemenea, asociate cu un risc crescut de boală coronariană; activitatea fizică este, dimpotrivă, asociată cu o scădere a apariției bolii coronariene; obezitatea este asociată cu un risc crescut de boală coronariană; și, în sfârșit, boala coronariană apare mai frecvent la indivizii cu diabet zaharat.

Când ne uităm astăzi pe această listă, ne putem întreba de ce a fost nevoie de așa un studiu laborios pentru a examina aceste relații evidente și binecunoscute. Tocmai, de aceea aceste relații sunt astăzi „evidente” și „binecunoscute”, pentru că a existat studiul Framingham, un studiu clasic de cohortă care a adus contribuții fundamentale la înțelegerea epidemiologiei bolii cardiovasculare și din același motiv este, acest studiu, celebru.

Studiul Framingham este un studiu de cohortă de tip I, selectând o populație definită pe baza locului de reședință sau a altor factori fără legătură cu expunerea. Populația a fost apoi observată de-a lungul timpului pentru a vedea care dintre indivizi sunt sau devin expuși, iar mai apoi care dintre ei fac boala cardiovasculară. Această abordare a permis studiul simultan al mai multor expuneri, precum hipertensiunea, fumatul, obezitatea, colesterolul și altele.

Exemplul 2: Studiul privind sănătatea asistentelor (*Nurses Health Study*) a evaluat incidența și factorii de risc pentru bolile frecvente ale femeilor. Etapele de bază ale acestui studiu au fost:

1. Asamblarea coortei: În 1976, investigatorii au obținut, de la Colegiul Asistentelor, lista cu asistentele înregistrate în primele 11 state ale SUA ca mărime a populației, având vârstele între 25 și 42 de ani, cărora li s-a trimis prin poștă o invitație de a participa la studiu.

2. Măsurarea **variabilelor independente** (expuneri) și a posibiloilor **factori de confuzie**. A fost trimis un chestionar privind dieta și alți potențiali factori de risc și au primit chestionarele completate de la 121.700 de asistente. Apoi, au retrimis chestionarele la fiecare 2 ani timp de 20 de ani, actualizând astfel expunerile care se modificaseră între timp și au evaluat și alți factori de risc, neevaluați inițial.

3. Au urmărit coorta și au măsurat efectele. Chestionarele periodice includeau și întrebări privind apariția mai multor efecte, care erau apoi confirmate prin cercetarea documentelor medicale.

Mărimea coortei și durata lungă de urmărire a furnizat o ocazie unică de studiu al factorilor de risc pentru diferite forme de boli cardiovasculare, cancer, și alte boli frecvente. De exemplu, investigatorii au evaluat ipoteza că alimentația bogată în fibre este asociată cu scăderea riscului de cancer colorectal. Aportul de fibre a fost evaluat în 1980, și 787 de cazuri de cancer colorectal au fost confirmate între 1980 și 1994. Incidența cancerului colorectal printre femeile cu consumul cel mai mic de fibre a fost similară incidenței cancerului printre femeile cu consumul cel mai mare (risc relativ = 1,0; intervalul de încredere 95% între 0,7 și 1,4), deci aportul alimentar crescut de fibre nu previne cancerul de colon.

**Studiul de cohortă retrospectiv** are, conceptual, același design ca și cel prospectiv: un grup de subiecți este urmărit de-a lungul timpului, măsurarea factorilor de risc efectuându-se mai întâi, iar apoi evaluându-se efectele. Diferența constă în faptul că asamblarea coortei, măsurătorile inițiale, urmărirea și evaluarea efectelor a avut loc în trecut (înaintea începerii studiului).

Unui investigator i-a venit ideea de a evalua asocierea dintre anticorpii anticardiolipinici (aCL) și incidența accidentului vascular cerebral. Pe această temă sunt nenumărate studii caz martor, pentru un studiu de cohortă prospectiv ar fi trebuit mulți ani, așa că investigatorii au efectuat un studiu de cohortă retrospectiv pe coorta Framingham. Factorul de risc, care trebuia măsurat de la început, este reprezentat de aCL, și cum din fericire, pentru fiecare participant la studiu căruia i se măsuraseră colesterolul, glicemia etc. se păstraseră la congelator și eprubete cu ser pentru determinări viitoare, s-au putut determina, iar coorta Framingham a putut fi împărțită, pe baza lor, în expuși și neexpuși. Înregistrarea accidentelor vasculare cerebrale oricum se

făcuse, pentru că fusese unul dintre efectele urmărite de studiul Framingham, așa că nu a mai rămas decât de făcut statisticile (măsurarea riscului relativ cu interval de încredere și p). Același lucru s-ar putea face și cu proteina C reactivă, la care nu se gândea nimeni la începuturile studiului Framingham că ar fi factor de risc cardiovascular.

### **Studiile de cohortă – avantaje și dezavantaje**

Avantaje:

- Permit măsurarea directă a **incidenței (riscului absolut)** al bolii la expuși și neexpuși
- Pot elucida temporal relația dintre expunere și boală.
- Deosebit de utile când expunerea este rară (studiu de cohortă de tip II).
- Pot evalua efecte multiple ale unei singure expuneri.
- Atunci când sunt prospective, minimizează erorile sistematice de măsurare a expunerii

Dezavantaje:

- Ineficiente, deoarece trebuie înrolați mult mai mulți decât cei care suferă evenimentul de interes (De exemplu, studiul Framingham asupra bolilor cardiovasculare, cele mai frecvente din SUA, a fost cel mai mare de acest tip atunci când a început. Cu toate acestea, a fost nevoie de urmărirea a peste 5000 de oameni, timp de mai mulți ani, până când au putu fi publicate primele concluzii. Doar 5% dintre indivizi avuseseră un eveniment coronarian în timpul primilor 8 ani). Așadar, nu pot fi folosite pentru studiul bolilor rare.
- Atunci când sunt prospective, durează mult timp (cel puțin zece ani) (cu excepția cazurilor în care riscul este foarte mare, iar factorul de risc acționează repede)
- Atunci când sunt retrospective: necesită documente medicale de bună calitate
- Validitatea rezultatelor poate fi afectată serios de pierderile din vedere (un studiu este valid când sunt pierduți din vedere mai puțin de 5% dintre pacienți, și aproape sigur invalid când sunt pierduți peste 20%).