

Disegno dello studio epidemiologico I

Definizione

- Studio della malattia e della salute e dei rispettivi determinanti nelle popolazioni umane
- Descrizione
- Individuazione dei determinanti

Informalmente

- L'Epidemiologia consiste nell'impiego di **metodi** quantitativi per comprendere fenomeni sanitari rilevanti

Epidemiologie

- Eziologica

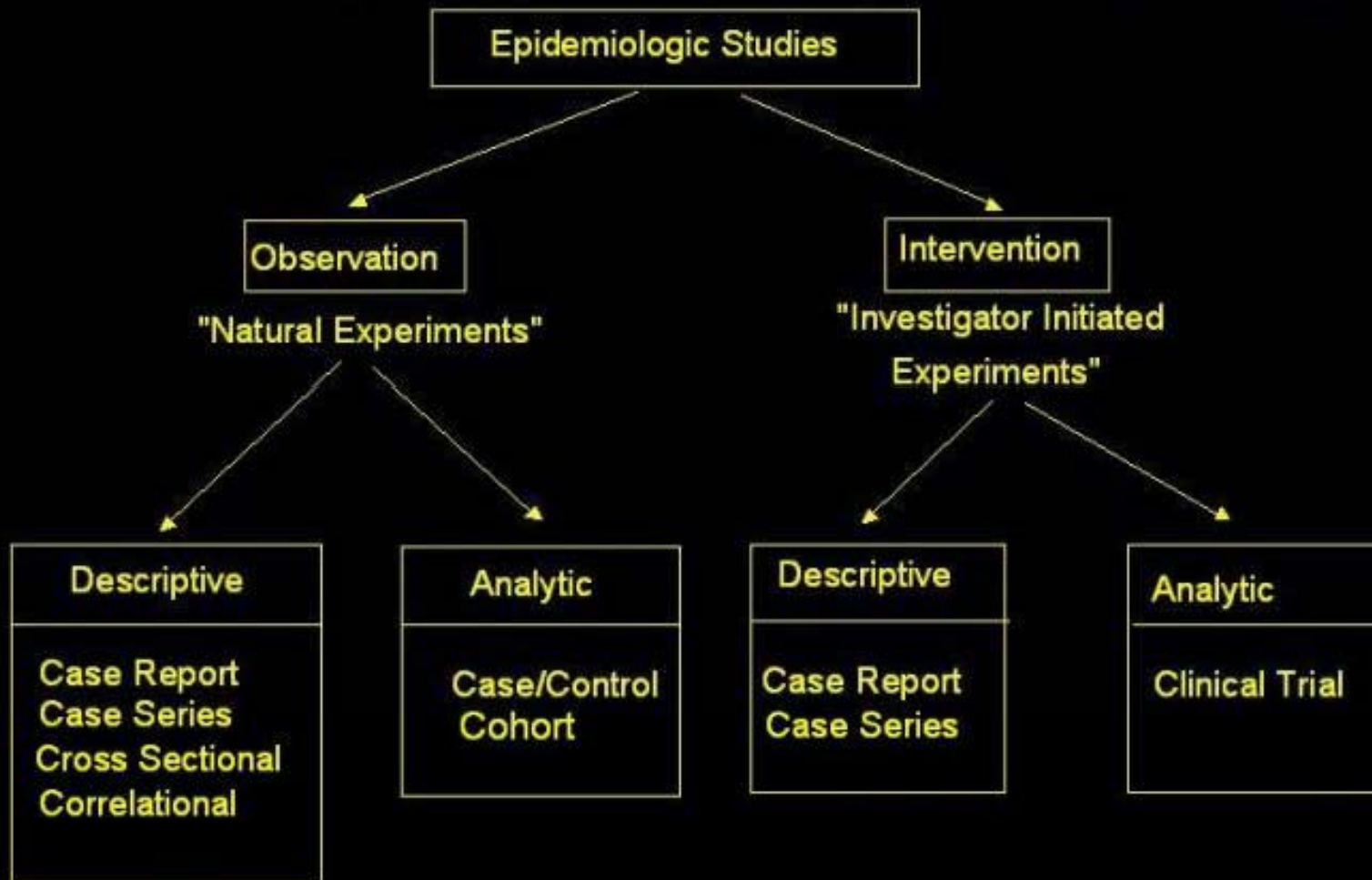
Finalizzata alla identificazione delle cause di malattia e del benessere delle popolazioni

- Clinica e valutativa

Finalizzata alla identificazione di determinanti dei risultati dei trattamenti e, in generale, del decorso della malattia nelle popolazioni dei malati

Tassonomia

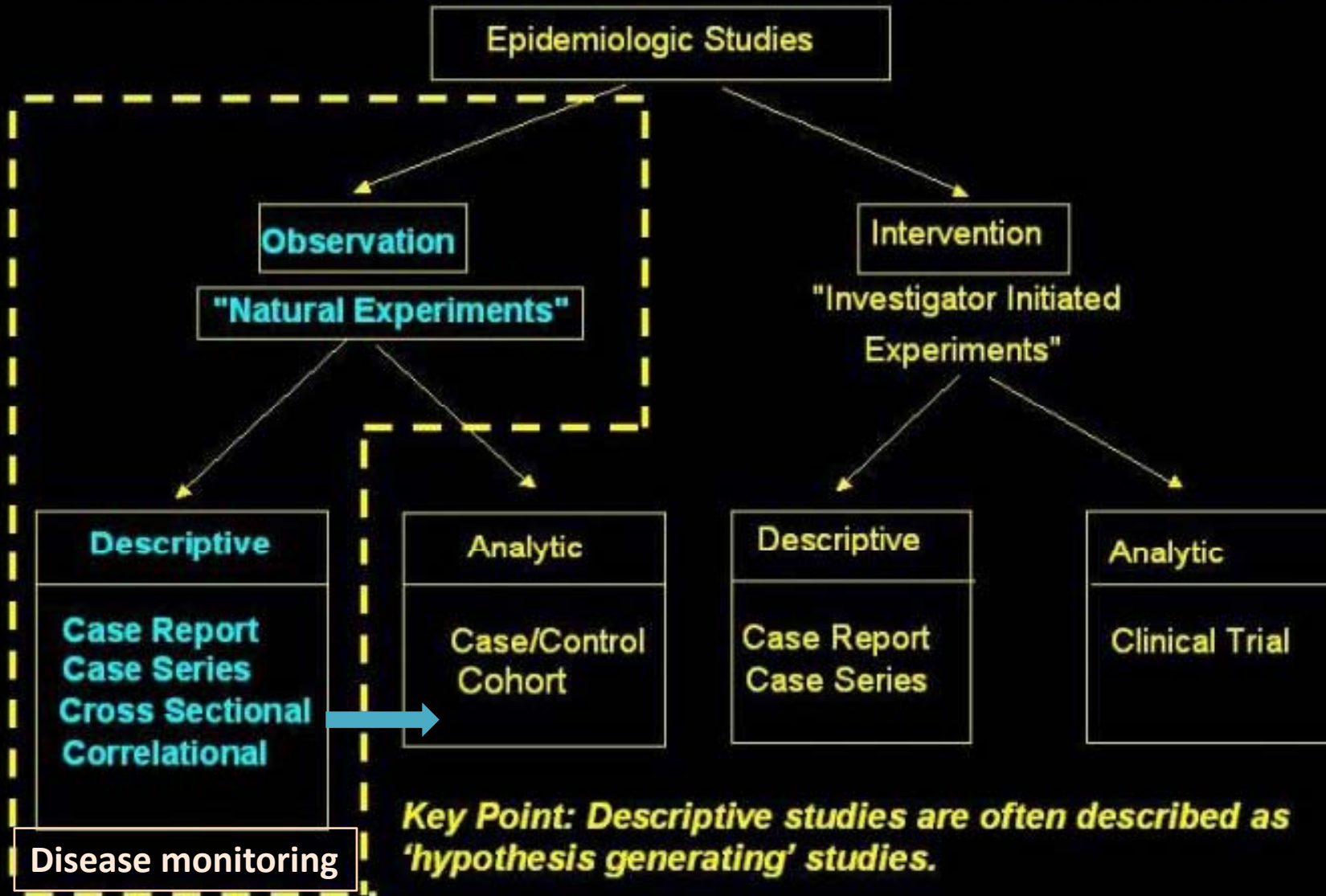
A Taxonomy of Epidemiologic Study Designs



Studi epidemiologici

- Studi Osservazionali
 - Descrittivi
 - longitudinali
 - trasversali
 - Analitici
 - Trasversali
 - Caso-controllo
 - Coorte
 - Storico
 - Prospettico
 - Caso-coorte
- Studi sperimentali

A Taxonomy of Epidemiologic Study Designs



Studi descrittivi

- Di popolazione su campione
- Indagini trasversali (survey)
 - Indagini su numerosi fenomeni sanitari rilevanti (fattori di rischio, partecipazione a interventi di screening)
 - Indagini di prevalenza
- Sistemi di sorveglianza:
 - Mortalità per cause
 - Registri di Patologia (tumori, diabete)

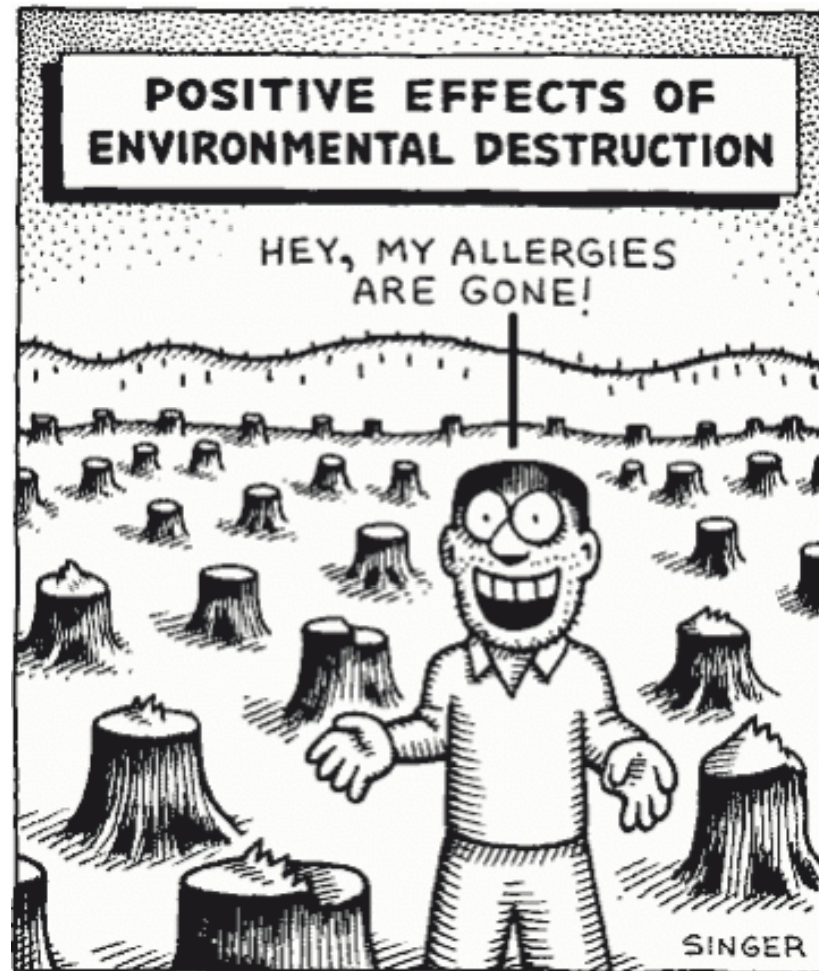
Caratteristiche dei sistemi di sorveglianza

- Utilizzo di regole comuni e sistemi di classificazione standard (confrontabilità)
- Copertura di popolazione (assenza di selezione)
- Durata nel tempo (analisi dei trend temporali)
- Informazioni raccolte limitate
- Assenza di gruppo di confronto definito da fattori di rischio

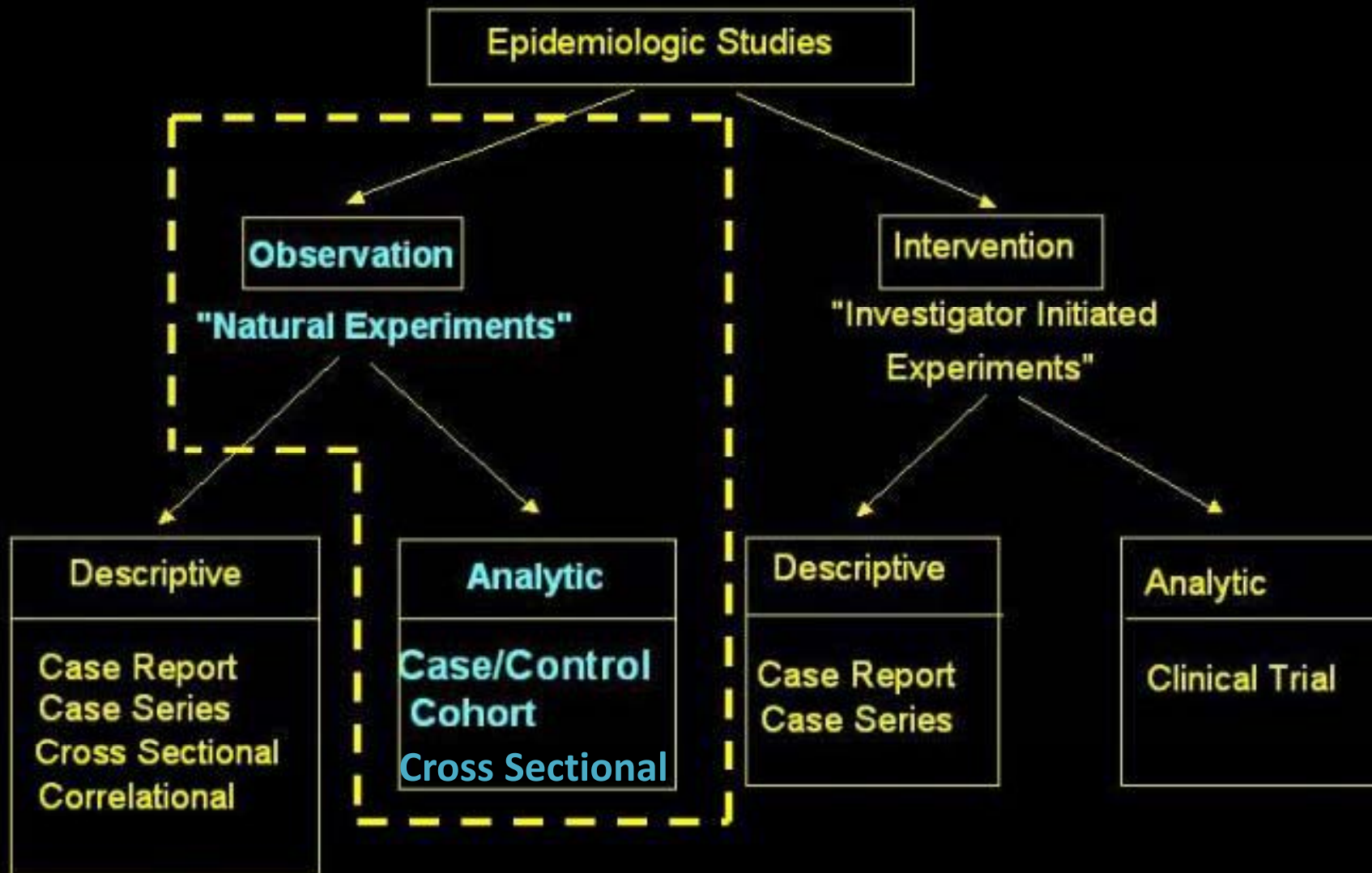
Obiettivi dei sistemi di sorveglianza

- Quantificare il peso delle malattie sulla salute delle popolazioni
- Evidenziare il grado di controllo conseguito dai sistemi sanitari
- Generare ipotesi eziologiche in base a confronti in base a tempo, luogo, e persona
- Semplificare l'esecuzione di studi analitici

Esperimento naturale

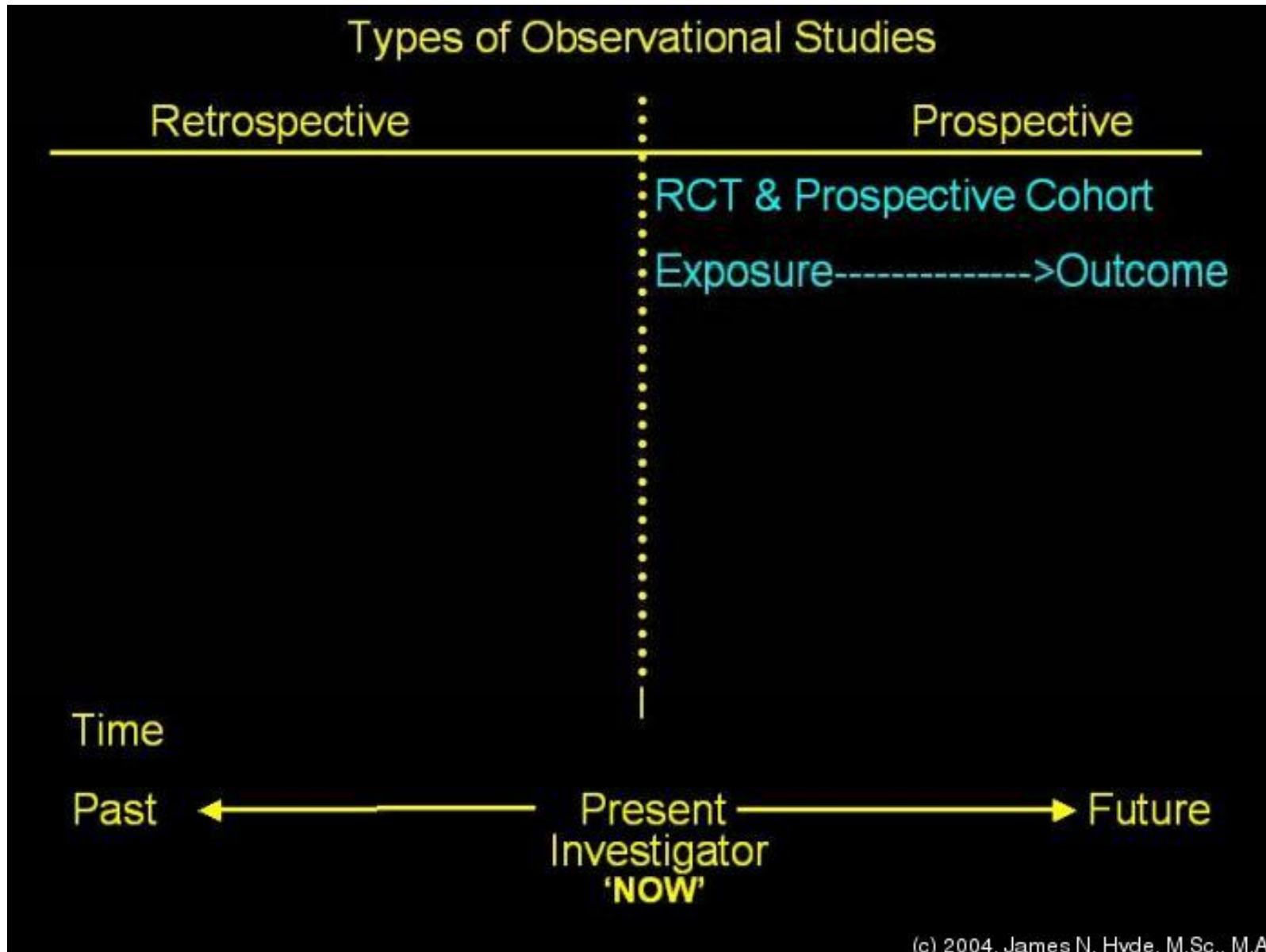


A Taxonomy of Epidemiologic Study Designs

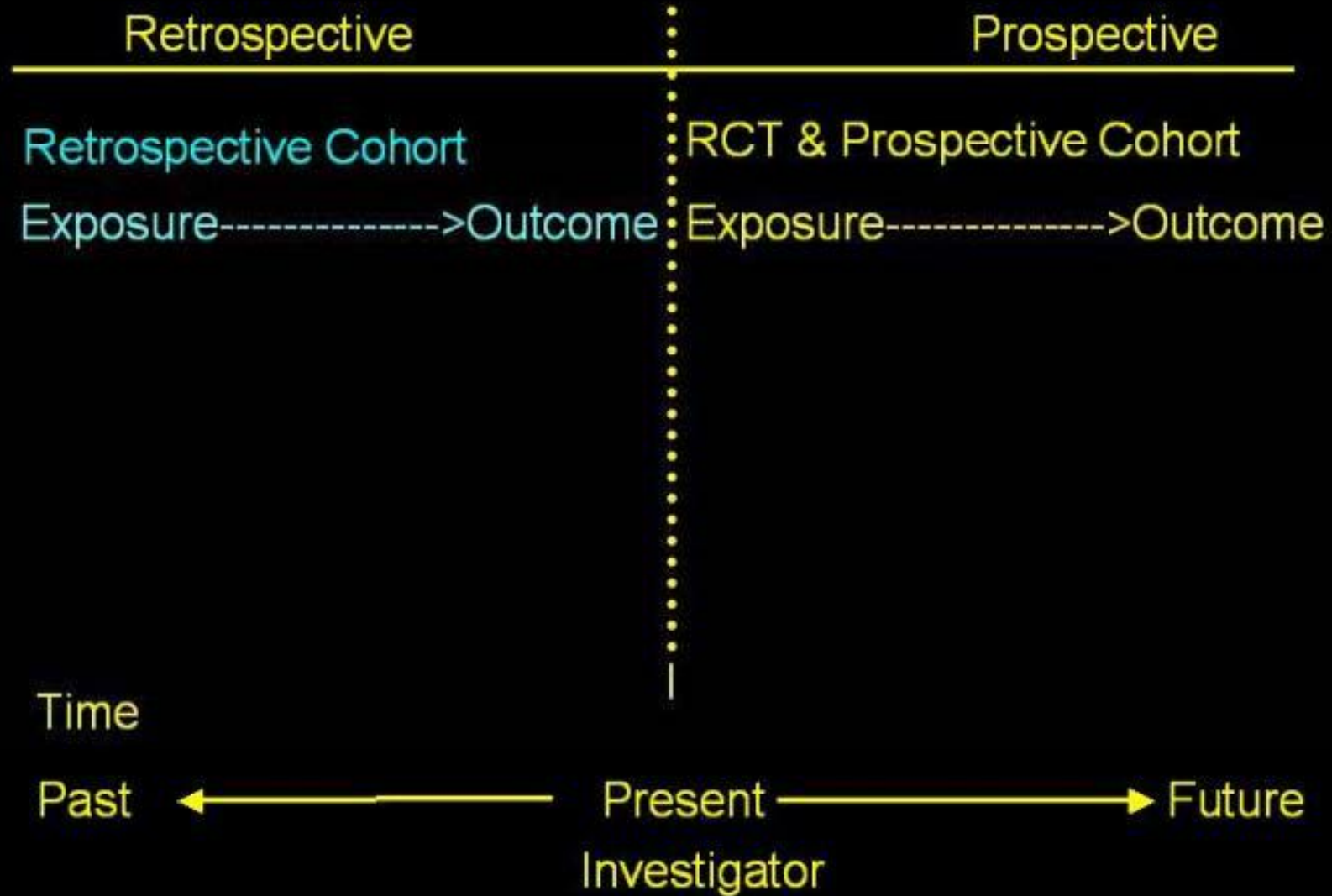


Key Point: Analytic studies-used for exploring causal relationships.

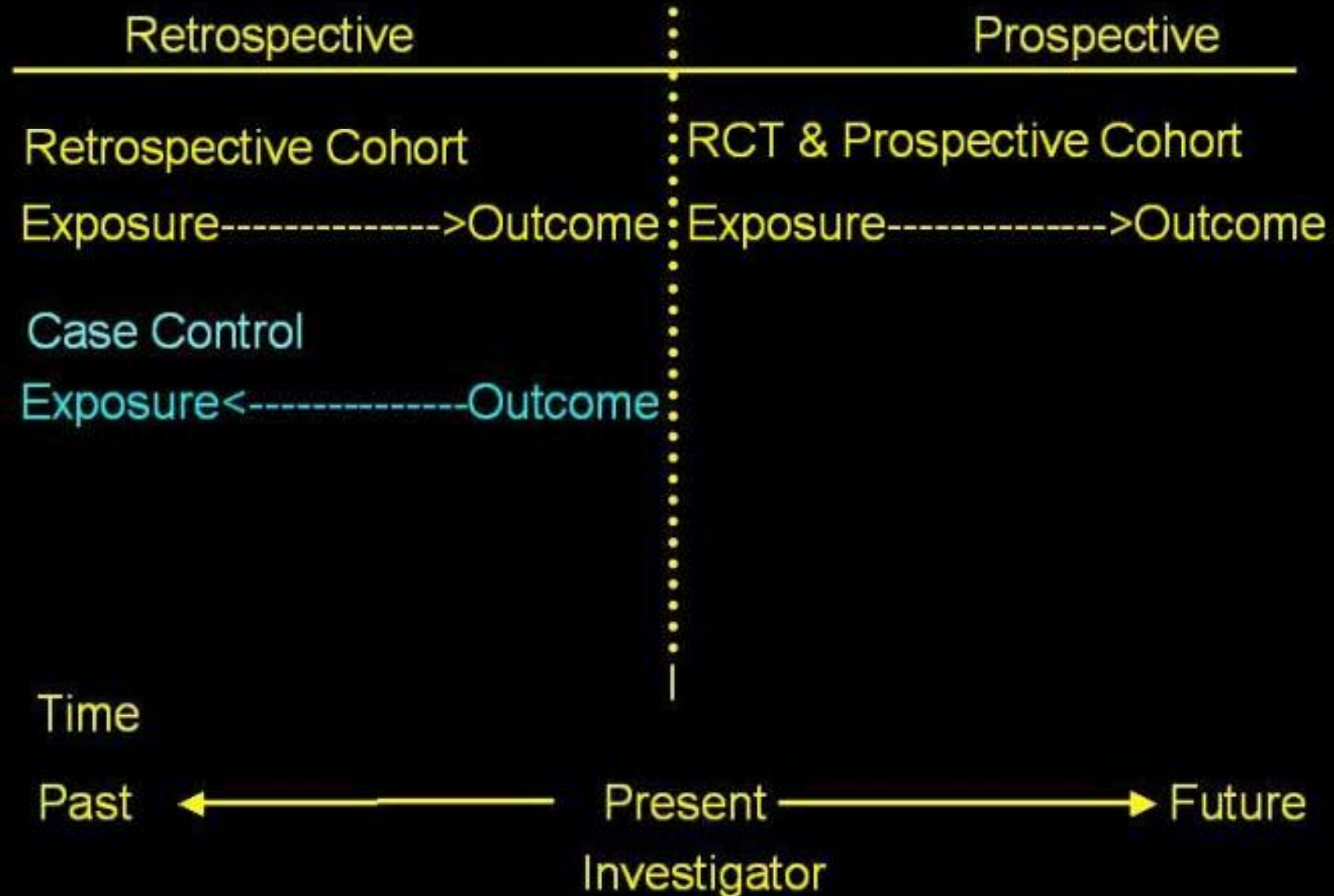
Rispetto all'investigatore



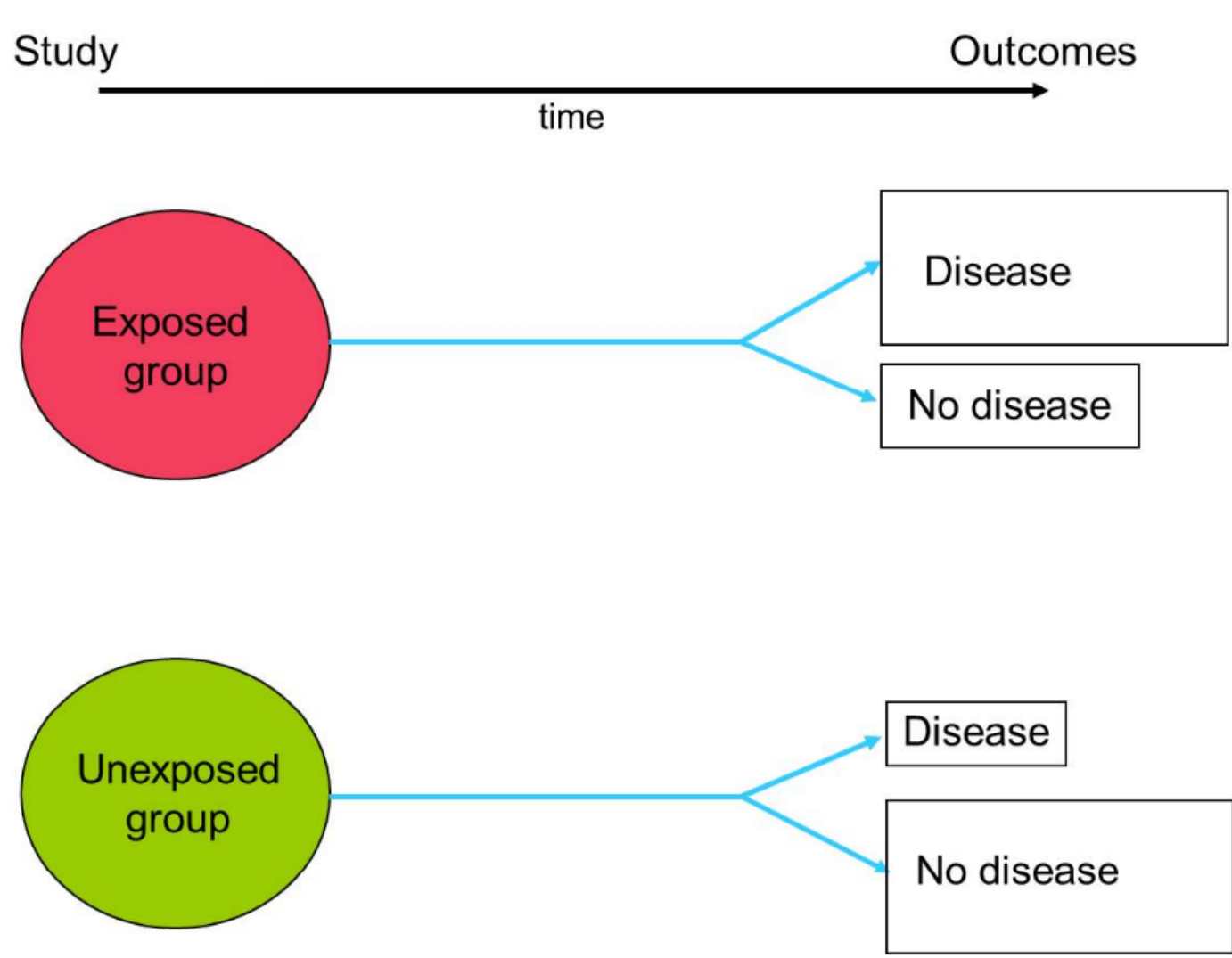
Types of Observational Studies





Types of Observational Studies

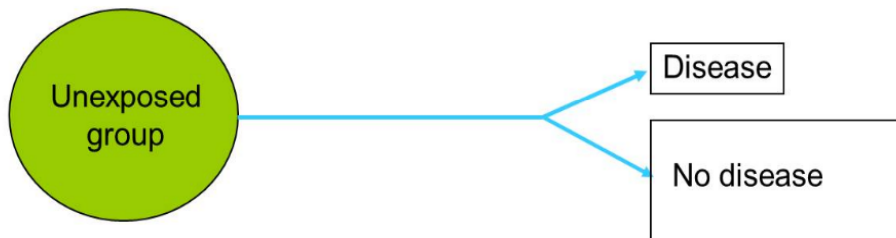
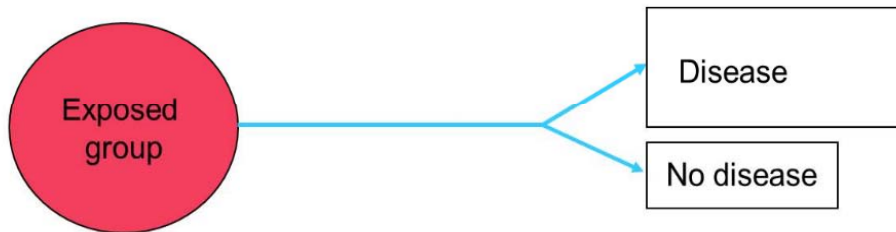



Studi di coorte : prospettici rispetto alla direzione dello studio

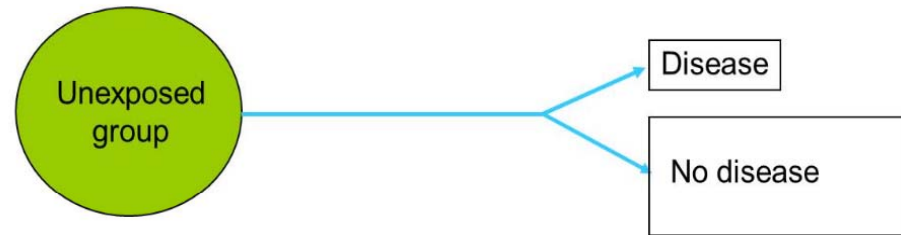
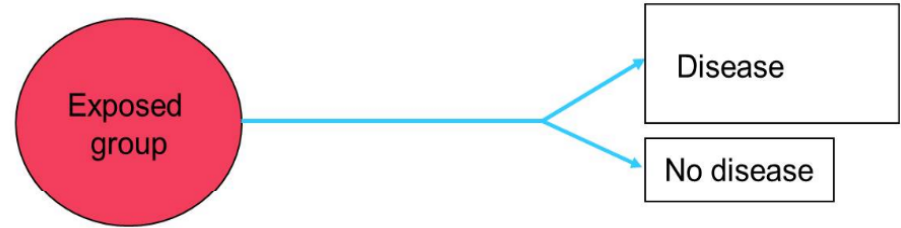


Coorte prospettica
Progettazione dello studio ne precede
la realizzazione

Study   Outcomes
time

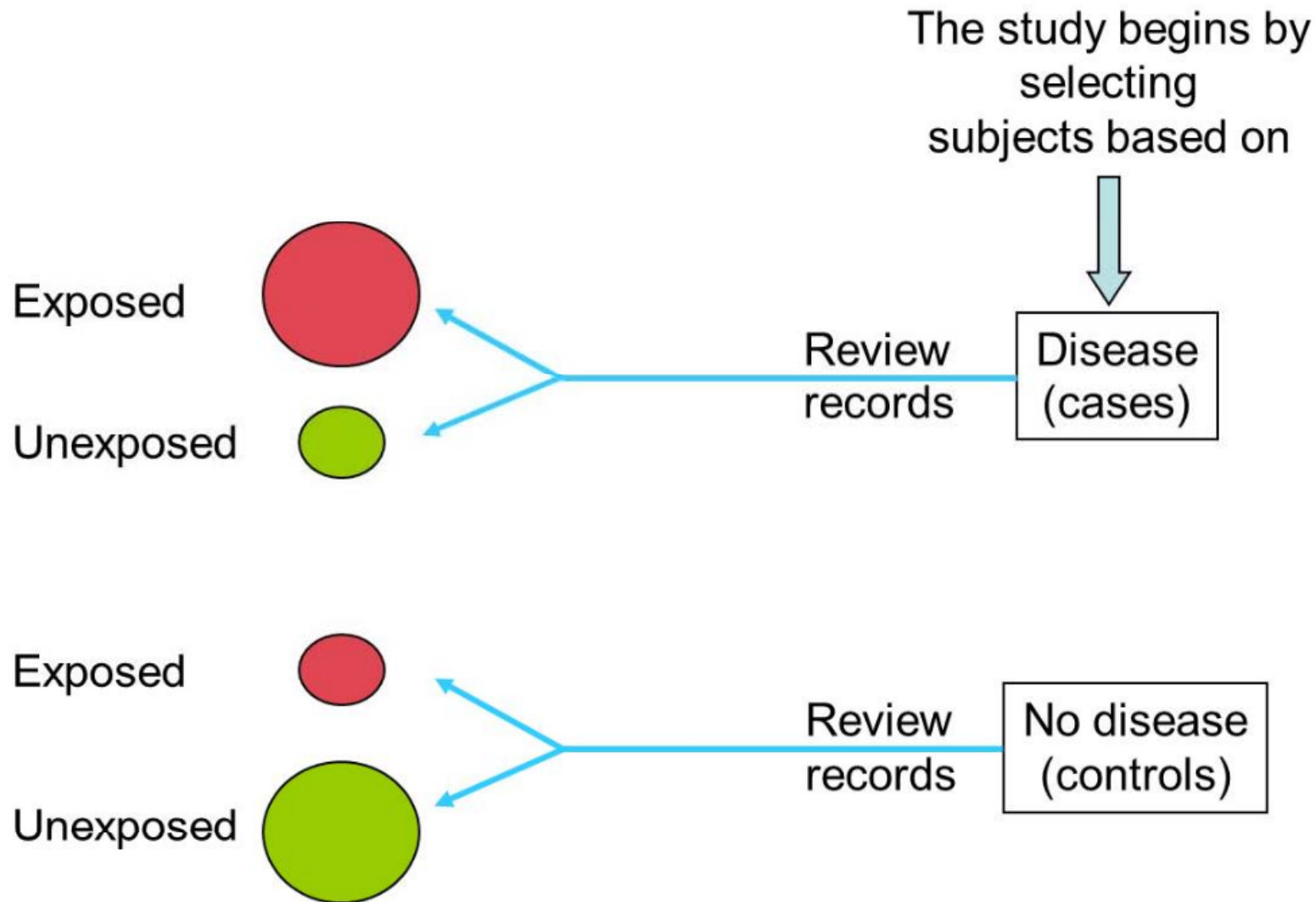


Study begins  Outcomes
time



Coorte storica
Progettazione dello
studio segue
l'esperimento naturale

Retrospectivo o caso controllo



Misure epidemiologiche

Misure epidemiologiche

Frequenza

- *Frequenza assoluta*
- Rischio / probabilità
- Tasso

Associazione

- Rapporto (Rischio relativo, rapporto tra tassi)
- Differenza (tra rischi, tra tassi)
- Rapporto tra odds o odds ratio

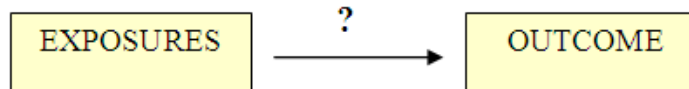
Impatto

- Rischio attribuibile

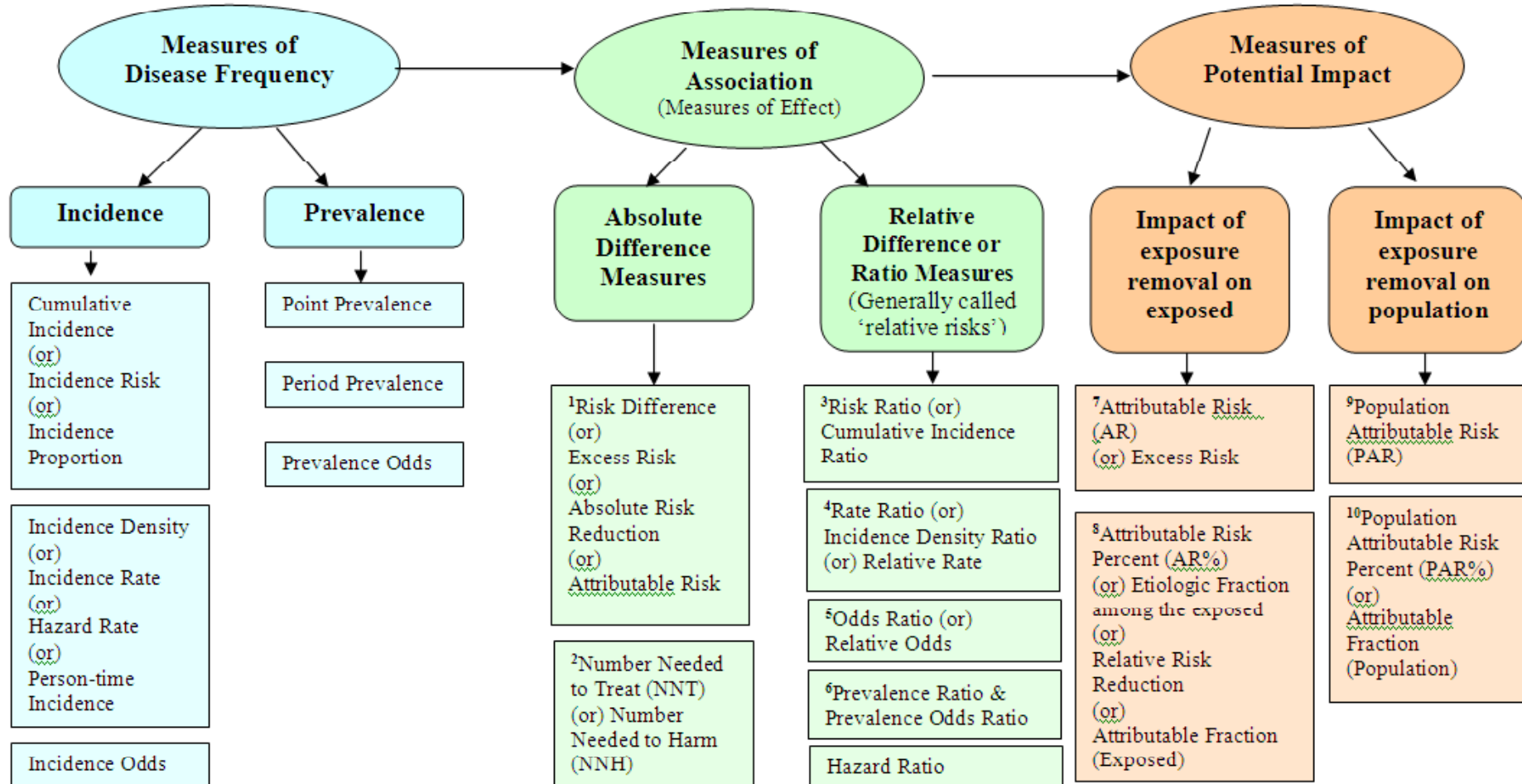
Frequenza assoluta

- Non è compresa tra le misure epidemiologiche dato che non può essere confrontata
- Salvo l'ipotesi piuttosto remota che il numero di eventi derivi da popolazioni di eguale numerosità
- Tuttavia il numero di eventi ha un valore informativo ed intuitivo per chi conosce la dimensione della popolazione
- E' opportuno riportare numeri assoluti a fianco di misure più tecniche come il tasso standardizzato

[AN OVERVIEW OF MEASUREMENTS IN EPIDEMIOLOGY [VER 3, 2007]



Epidemiology is about identifying associations between exposures and outcomes. To identify any association, exposures and outcomes must first be measured in a quantitative manner. Then rates of occurrence of events are computed. These measures are called “*measures of disease frequency*.” Once measured, the association between exposures and outcomes are then evaluated by calculating “*measures of association or effect*.” Finally, the impact of removal of an exposure on the outcome is evaluated by computing “*measures of potential impact*.” In general, measures of disease frequency are needed to generate measures of association, and both these are needed to get measures of impact. There is some overlap between these measures, and terminology is poorly standardized.



Misure di frequenza

Definizione precisa di :

- Numeratore, i “casi”
- Denominatore, la “popolazione a rischio”

Misure di frequenza principali:

- Incidenza
- Prevalenza
- Incidenza: Incidenza cumulativa [incidence risk, incidence proportion]
- Densità di incidenza [incidence rate; force of incidence]
- Prevalenza: prevalenza puntuale
- *Prevalenza di periodo*

Misure di frequenza

Rapporti

Proporzioni:

Il numeratore è incluso nel denominatore o deriva da esso; espresse di solito come %

Tassi: velocità di comparsa di evento in studio; il numeratore non è incluso nel denominatore che è un tempo a rischio

Rischio di malattia

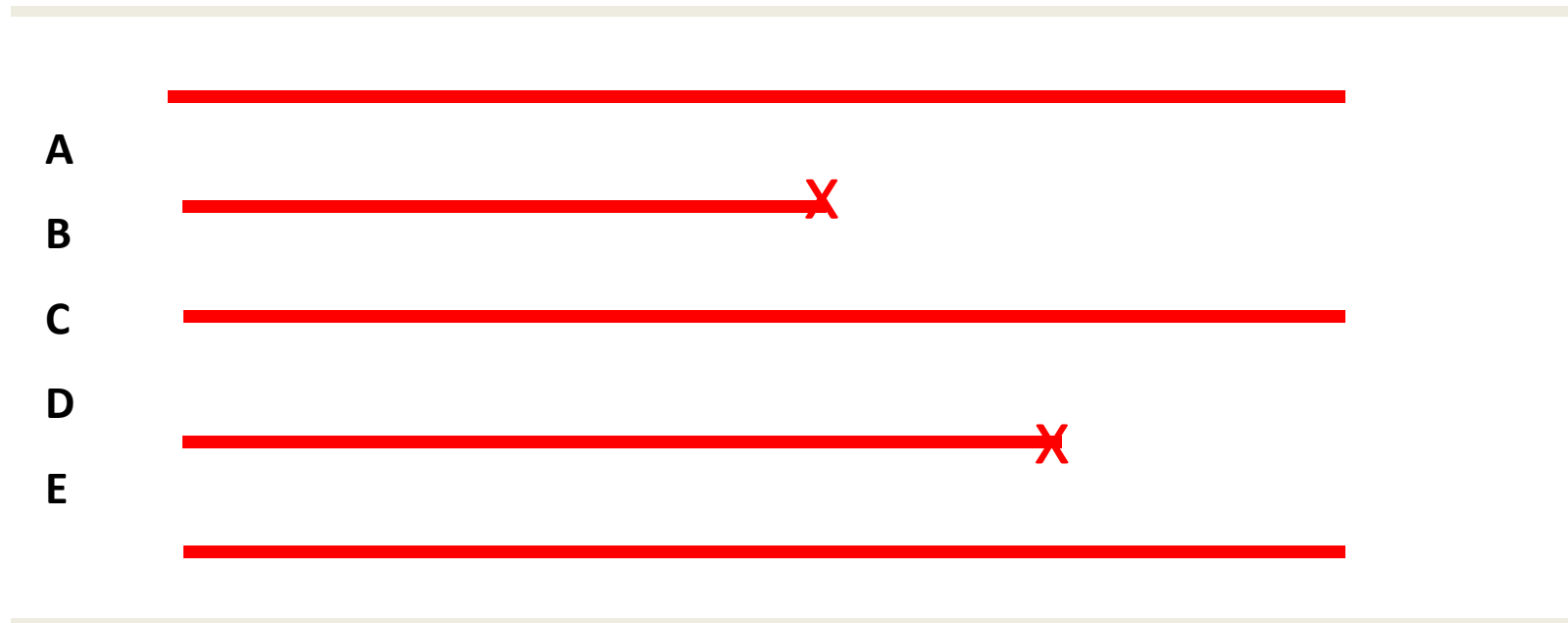
- Nell'ambito della epidemiologia delle patologie infettive si utilizza anche il nome "*tasso di attacco*" dove il termine tasso è inappropriato (*attack rate*)
- Se non consideriamo il tempo di esposizione individuale ma la dimensione iniziale della popolazione abbiamo:

$$R = \frac{n_eventi}{n_popolazione_iniziale}$$

Calcolo del rischio

Admission day

7 days later



Condition: Same follow up for all

$$R = 2 / 5 = 0.4 = 40\%$$

Letalità

- È un rischio
- La calcoliamo tra l'altro per la mortalità post-operatoria convenzionalmente a 30 giorni
- Per malattie infettive caratterizzate da elevato rischio di morte (tetano, botulismo)

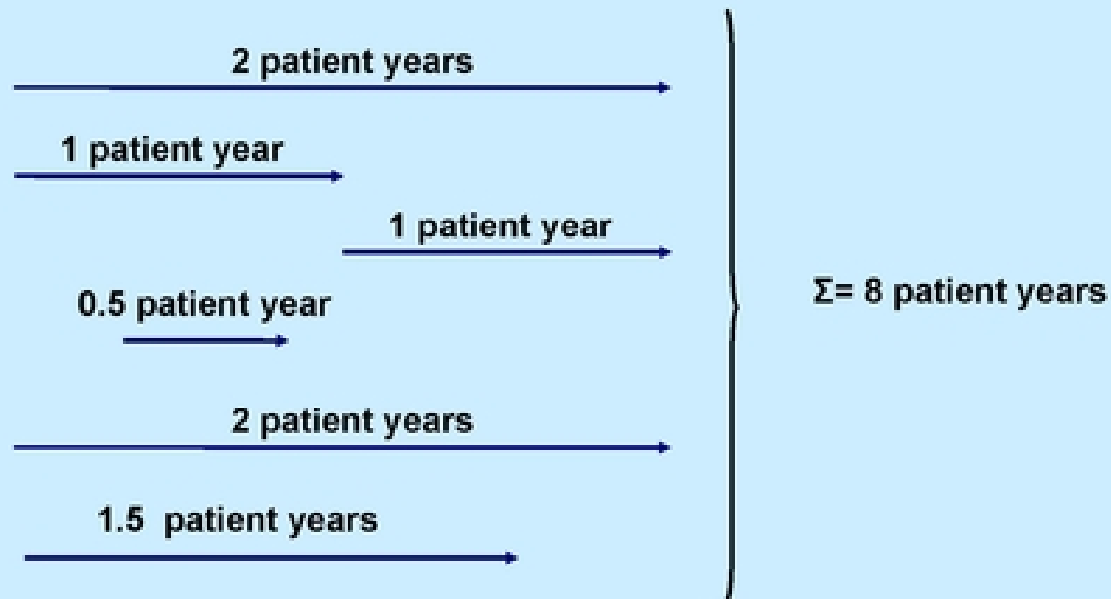
Tasso

- Misura la velocità di comparsa di un evento definito in una popolazione per unità di tempo
- Richiede:
 - la definizione di un evento misurabile e relativa enumerazione
 - La scelta di una popolazione di riferimento rispetto alla quale intendiamo misurare la comparsa dell'evento
 - La misurazione del tempo di osservazione (esposizione a rischio) per ogni componente della popolazione

Tempo (anni) -persona

Person time:

Sum of all individual observation periods of study participants under observation



Per definire un tasso

Specificare:

- il tempo di osservazione o durata dello studio
- il periodo di osservazione cui si riferisce il tasso
- l'unità di misura del tasso : numero di eventi x unità di tempo (ad esempio anno) per numerosità (k) cui si riferisce il tasso (ad esempio 100.000 abitanti)

Specificazione dei tassi

- Grezzo (Crude)
- Specifico (per classe d'età)
- Standardizzato o corretto (di solito per sesso ed età)

Il tasso grezzo

- Esprime la velocità di comparsa dell'evento nell'intera popolazione in relazione agli anni persona di esposizione

$$I = \frac{n_eventi}{\sum anni_persona} * k$$

Densità di incidenza

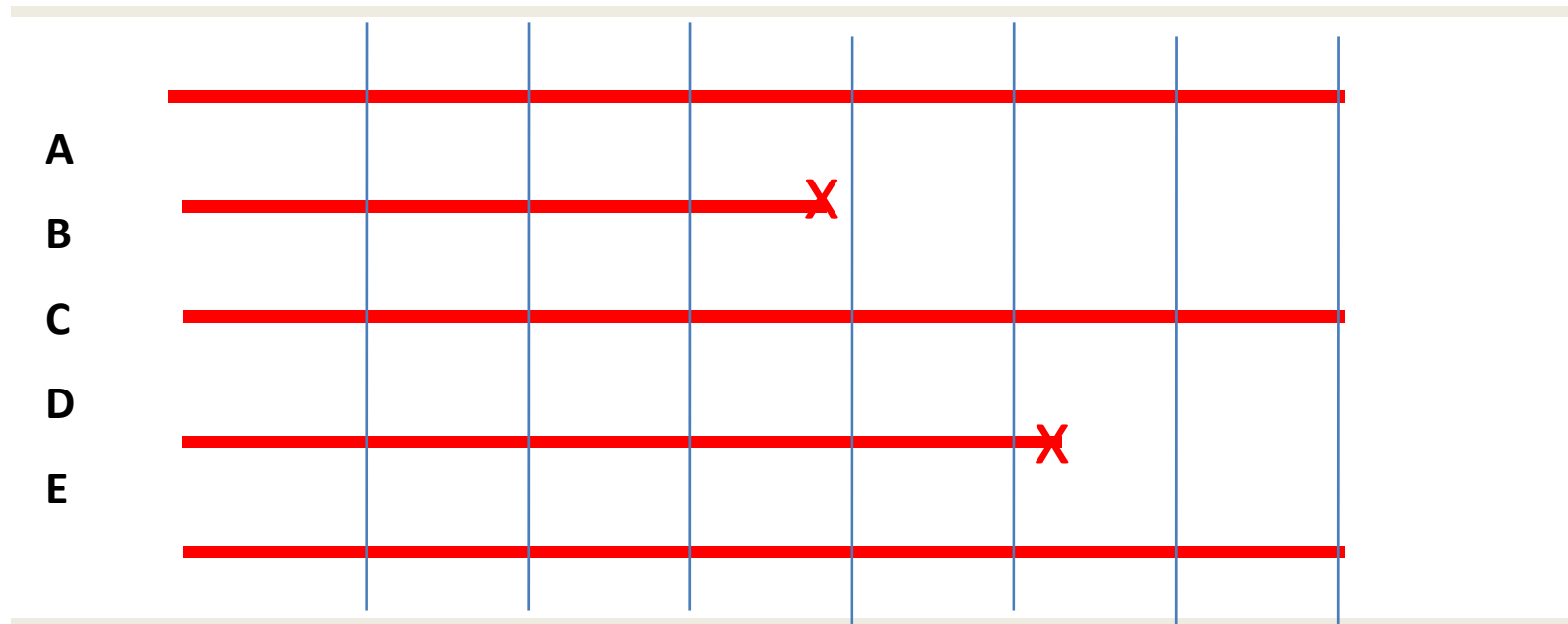
- Il tasso incidenza è anche chiamato:
- Densità di incidenza (*incidence density*)
- Forza di morbosità (*force of morbidity*)

- Il termine *hazard rate* è impiegato più spesso per indicare il valore limite teorico cui il tasso tende al ridursi dell'unità di tempo verso lo 0

Calcolo del tasso per settimane

Admission day

7 days later



Condition: Same follow up for all

$$SP = 3 + 4/7 + 6/7 = 4.43$$

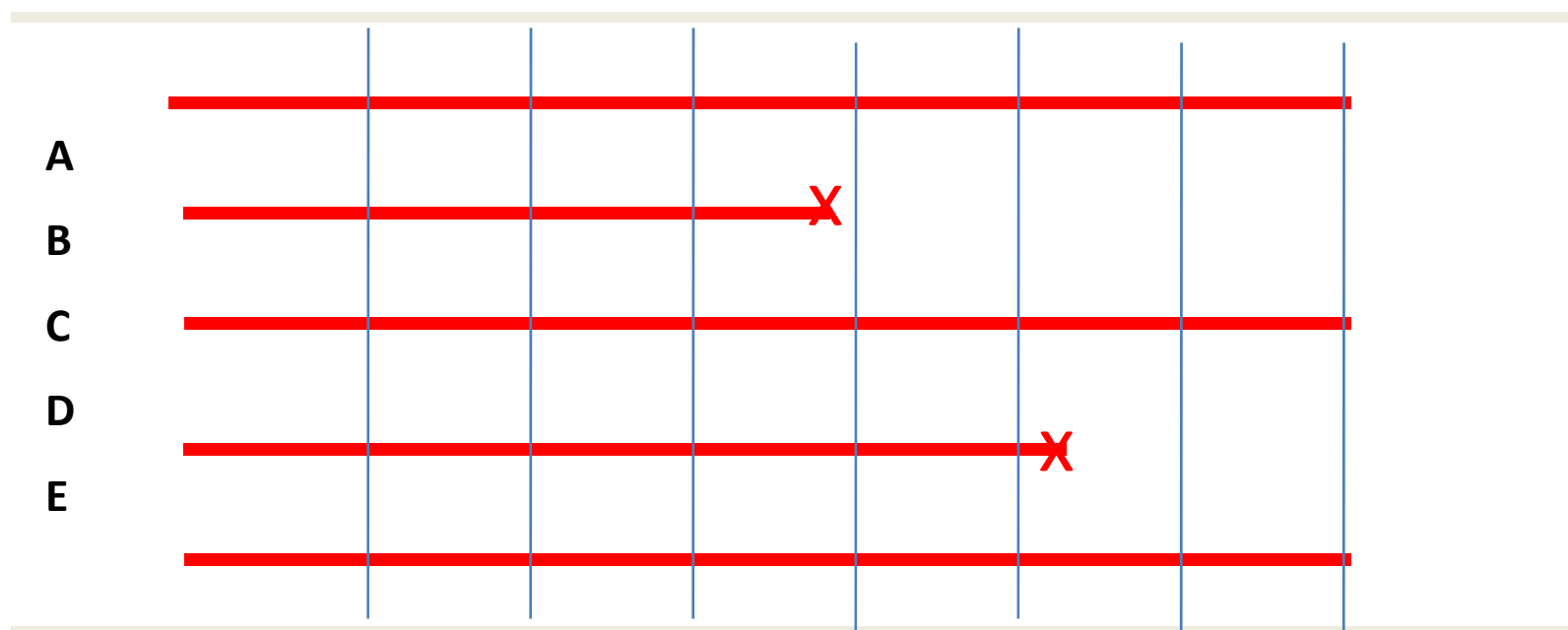
$$T = 2 / 4.43 * 100 = 0.452 * 100 =$$

45.2 casi per 100 settimane persona

Calcolo del tasso per giorno

Admission day

7 days later



Condition: Same follow up for all

Giorni persona $(7*3)+4+6=31$

$T = 2 / 31 * 1000 = 0.065 * 1000 =$

65 casi per 1000 giorni persona

Tasso standardizzato

- Obiettivo: rendere confrontabili tassi osservati in tempi o luoghi diversi eliminando l'effetto delle variabili di standardizzazione
- Per età e sesso
- Metodo diretto: standardizzazione diretta
- Metodo indiretto

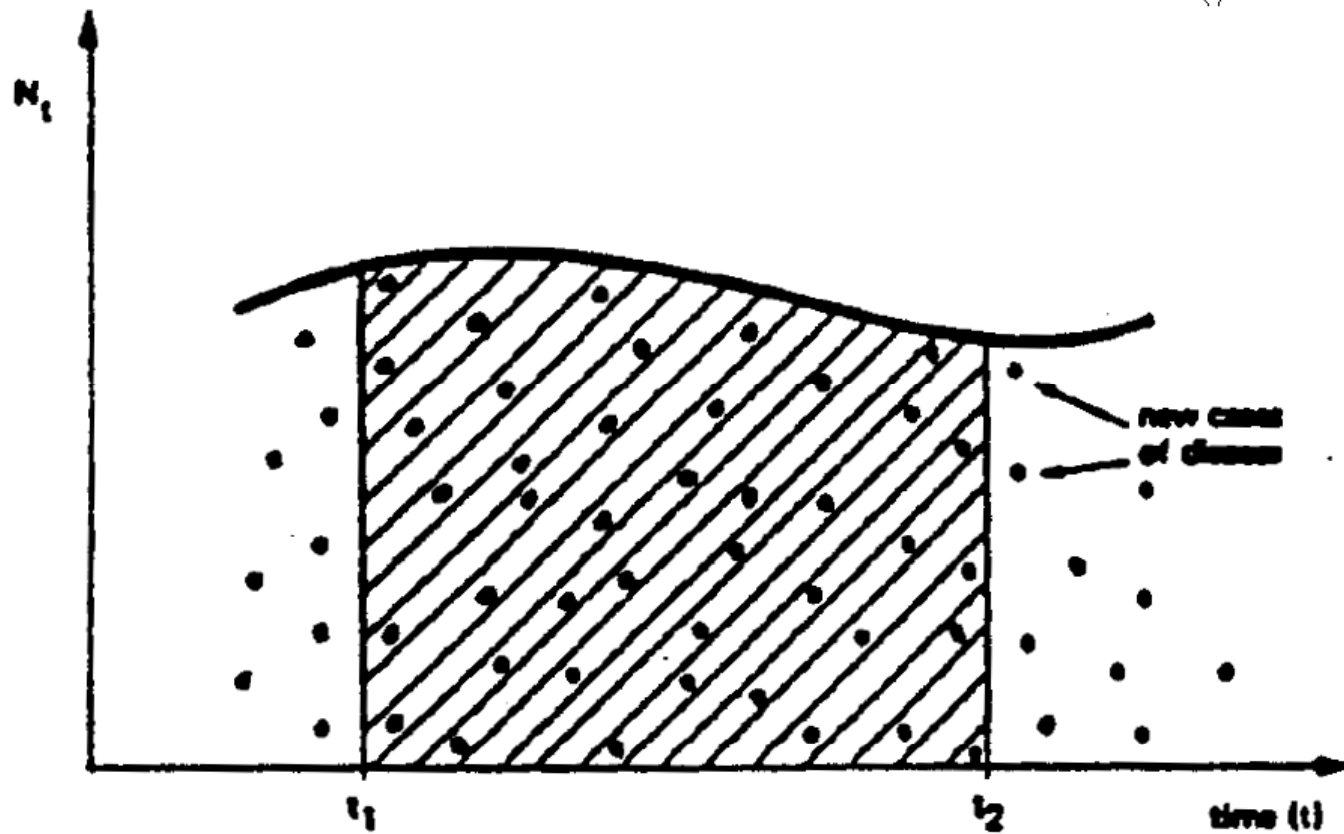


FIGURE 2 Graphical illustration of the occurrence of new (incident) cases over time in a candidate population (of size N_t at time t)

Popolazione umbra 1 gennaio 1996 e 1 gennaio 1997

età	1996		1997	
	m	f	m	f
0-4	16515	15637	16457	15484
5-9	17633	16645	17841	16754
10-14	19059	18035	18585	17721
15-19	22099	20660	21450	19902
20-24	27549	26744	26899	26003
25-29	30260	29480	30266	29561
30-34	32333	31460	32364	31643
35-39	29255	28606	30066	29297
40-44	27301	26920	27608	27287
45-49	28877	28976	28219	28342
50-54	25013	25865	25926	26773
55-59	26970	28258	26661	27654
60-64	24619	27027	24820	27410
65-69	24378	28430	23915	27757
70-74	22075	27840	22319	27795
75-79	13303	18748	14949	21234
80-84	9511	16747	9278	16615
85+	5707	11395	5573	11303
totale	402457	427473	403196	428535

Popolazione umbra 1 gennaio 1996 e stima intermedia (“a metà anno”)

età	1996		1996.5	
	m	f	m	f
0-4	16515	15637	16486	15560.5
5-9	17633	16645	17737	16699.5
10-14	19059	18035	18822	17878
15-19	22099	20660	21774.5	20281
20-24	27549	26744	27224	26373.5
25-29	30260	29480	30263	29520.5
30-34	32333	31460	32348.5	31551.5
35-39	29255	28606	29660.5	28951.5
40-44	27301	26920	27454.5	27103.5
45-49	28877	28976	28548	28659
50-54	25013	25865	25469.5	26319
55-59	26970	28258	26815.5	27956
60-64	24619	27027	24719.5	27218.5
65-69	24378	28430	24146.5	28093.5
70-74	22075	27840	22197	27817.5
75-79	13303	18748	14126	19991
80-84	9511	16747	9394.5	16681
85+	5707	11395	5640	11349
totale	402457	427473	402826.5	428004

Tassi età specifici di incidenza di cancro del polmone in Umbria

età	1996	1996.5	n.	1996	1996.5
m	m	C34-C34.9			
0-4	16515	16486	0	0	0
5-9	17633	17737	0	0	0
10-14	19059	18822	0	0	0
15-19	22099	21774.5	0	0	0
20-24	27549	27224	0	0	0
25-29	30260	30263	0	0	0
30-34	32333	32348.5	1	3.09E-05	3.09E-05
35-39	29255	29660.5	0	0	0
40-44	27301	27454.5	2	7.33E-05	7.28E-05
45-49	28877	28548	11	0.000381	0.000385
50-54	25013	25469.5	11	0.00044	0.000432
55-59	26970	26815.5	43	0.001594	0.001604
60-64	24619	24719.5	67	0.002721	0.00271
65-69	24378	24146.5	90	0.003692	0.003727
70-74	22075	22197	91	0.004122	0.0041
75-79	13303	14126	64	0.004811	0.004531
80-84	9511	9394.5	36	0.003785	0.003832
85+	5707	5640	16	0.002804	0.002837

Tassi età specifici * k

$$r_i = \frac{n_i}{pop_i} * k$$

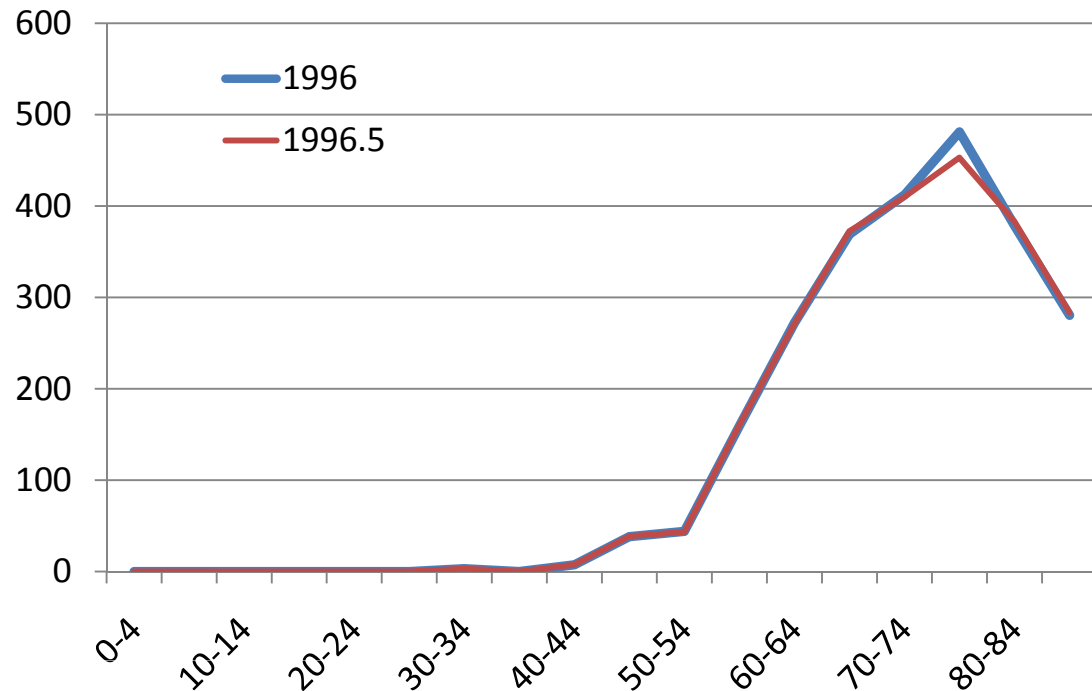
In cui

r_i è il tasso di incidenza
nella i -esima classe
d'età

n_i è il numero di nuovi
casi diagnosticati nella
stessa classe i

pop_i è la numerosità
della popolazione

$K = 100.000$



Tasso grezzo

$$I = \frac{\sum_{i=1}^{18} n_i}{\sum_{i=1}^{18} pop_i} * k = \frac{432}{402457} * 100.000 = 107.3$$

- Il tasso grezzo di incidenza di cancro del polmone è pari a 107 casi per 100.000 abitanti per anno nel 1996 in Umbria
- *Se avessimo utilizzato la popolazione più corretta a metà anno avremmo ottenuto 107.2*

La standardizzazione diretta

età	m+f 2001	r_i	$r_i^* \text{pop st}_i$
0-4	32698	0	0
5-9	33198	0	0
10-14	35176	0	0
15-19	38356	0	0
20-24	45581	0	0
25-29	57069	0	0
30-34	60638	3.09E-05	1.875421
35-39	63021	0	0
40-44	56842	7.33E-05	4.164097
45-49	53684	0.000381	20.44963
50-54	57132	0.00044	25.12501
55-59	50005	0.001594	79.72618
60-64	54154	0.002721	147.3788
65-69	49626	0.003692	183.2119
70-74	48713	0.004122	200.8101
75-79	43155	0.004811	207.6163
80-84	24192	0.003785	91.56892
85+	22586	0.002804	63.32153
totale	825826		1025.248

Il tasso standardizzato

$$I_{st} = \frac{\sum_{i=1}^{18} pop_i * r_i}{\sum_{i=1}^{18} pop_i}$$

- Il tasso di incidenza standardizzato per sesso ed età con la popolazione umbra del 2001 è pari a 124 casi per 100.000 abitanti per anno

$$I_{st} = (1025 / 825826 * 100000)$$

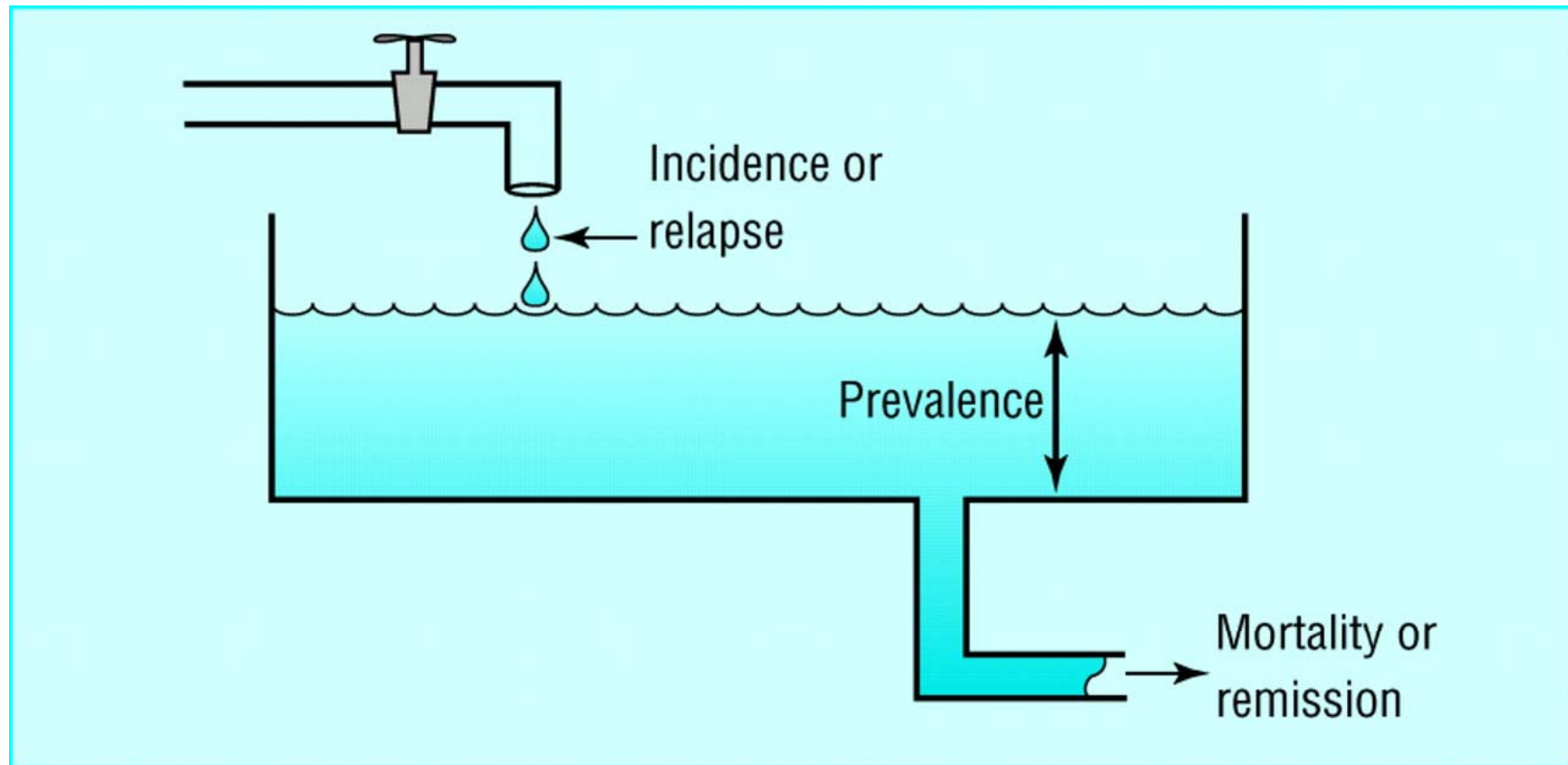
Prevalenza

- Misura centrata sullo **stato di malattia** o sulla presenza di una condizione (ad esempio prevalenza di anticorpi contro un agente infettivo) **in un momento specificato** (a differenza dell'incidenza che misura una transizione)
- Ha qualche valore per patologie soggette ad accumulo nella popolazione
- In particolare se può essere messa in relazione con l'impegno dei servizi sanitari

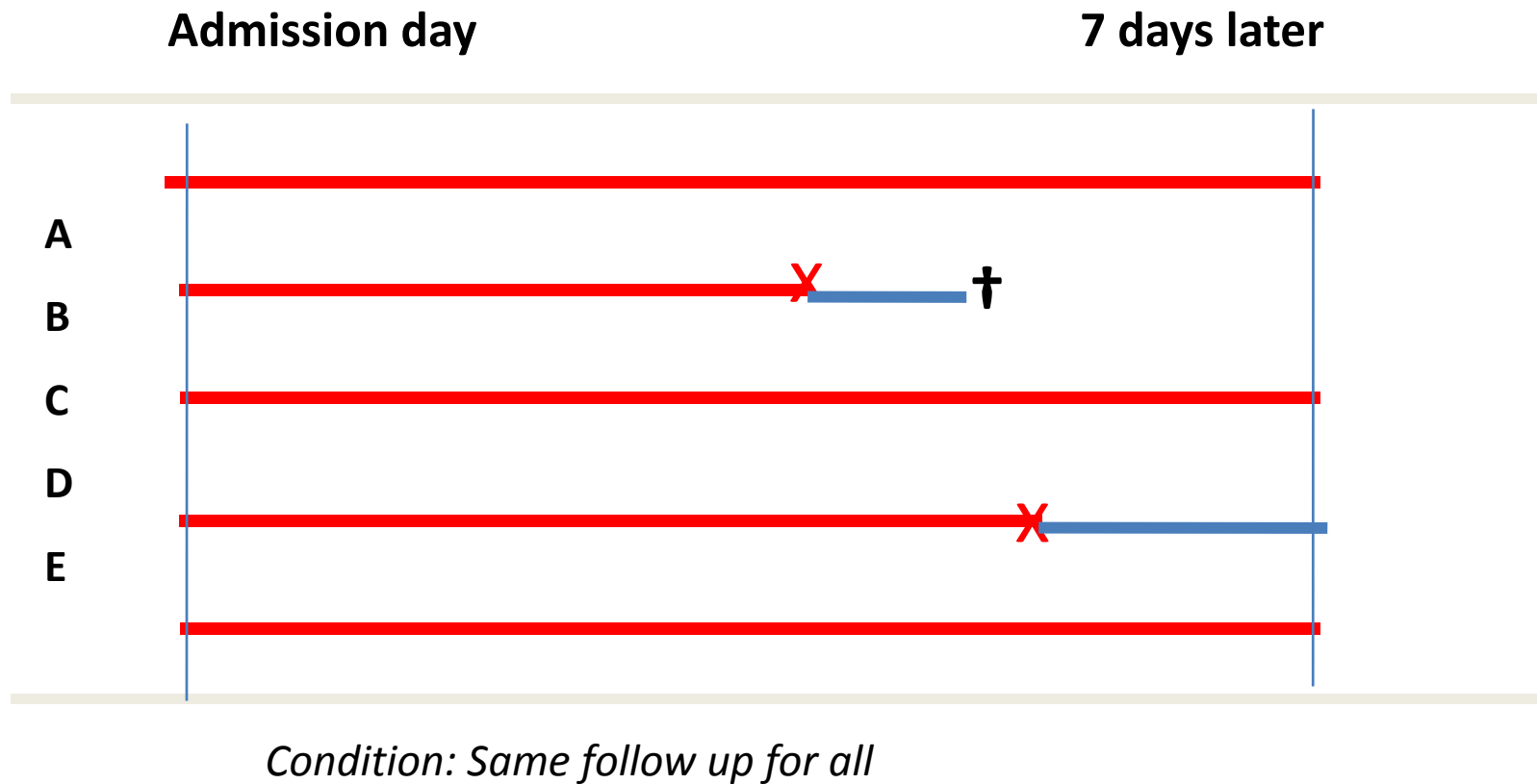
Calcolo della prevalenza puntuale

- Proporzione o rischio
- Prevalenza grezza: numero di malati/popolazione presente
- Il tempo di esecuzione dello studio deve essere trascurabile rispetto alle transizioni in ingresso e in uscita
- Deve essere specificata la data di riferimento della prevalenza
- E' possibile produrre valori età specifici e standardizzati

Relazione tra incidenza e prevalenza

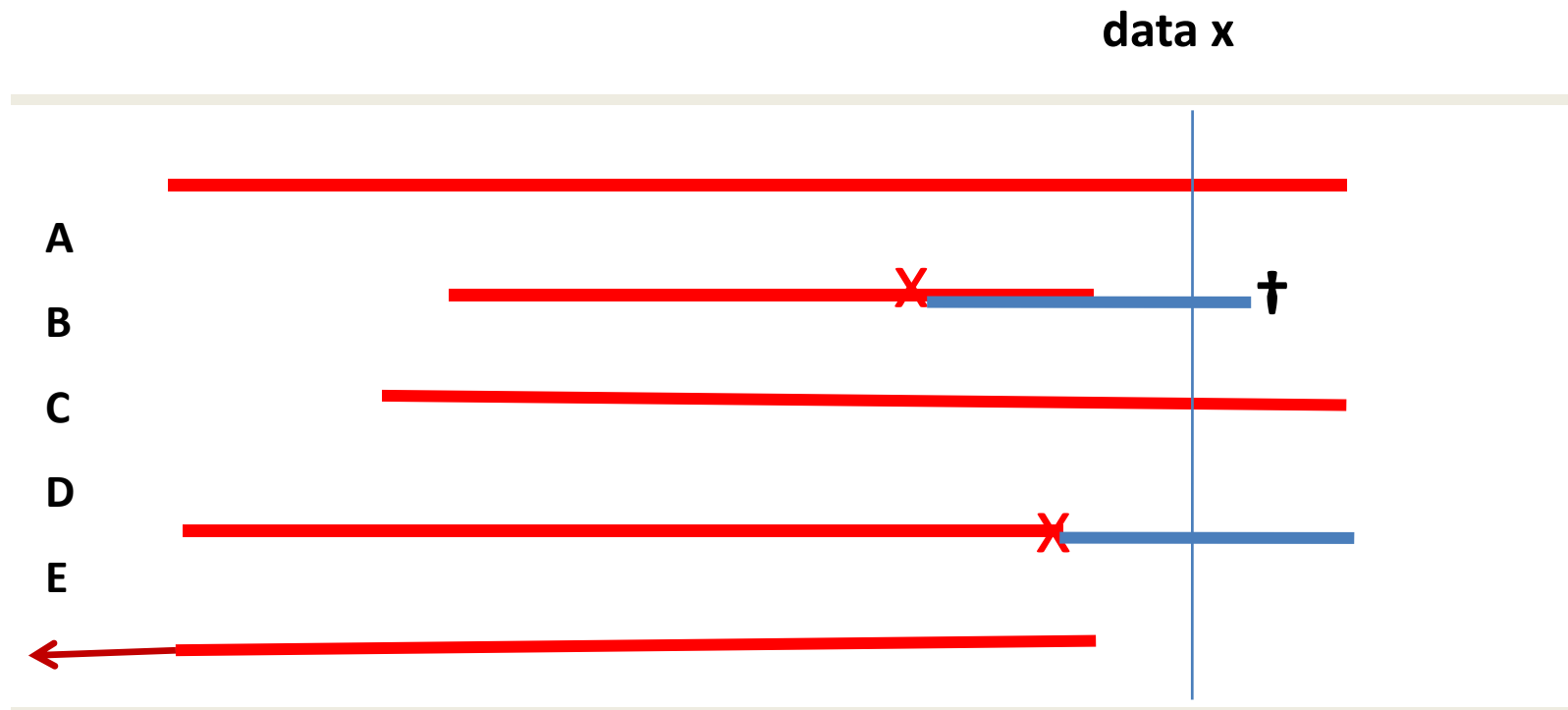


Calcolo della prevalenza puntuale



$$P = 1 / 4 = 0.25 = 25\%$$

Calcolo della prevalenza puntuale

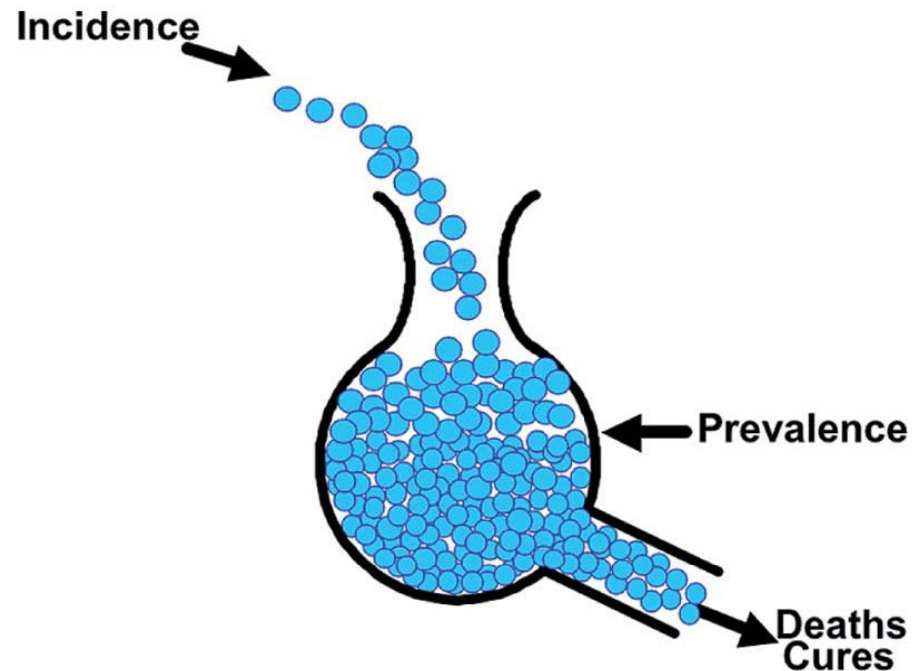


Condition: Same follow up for all

$$P \text{ data } x = 2 / 4 = 0.50 = 50\%$$

Implies a cross-sectional design

Di solito viene calcolata nell'ambito di uno studio trasversale (detto appunto di prevalenza, ingl. Cross-sectional)



Gordis: Epidemiology, 4th Edition.
Copyright © 2008 by Saunders, an imprint of Elsevier, Inc. All rights reserved
Relationship between incidence and prevalence: IV.

Misure di associazione

- Utilizzate per lo più in studi analitici
- Servono a misurare se
 - L'esposizione al fattore x determina un eccesso di rischio di malattia nel gruppo degli esposti rispetto ai non esposti? [fattore di rischio]
 - L'esposizione al fattore x determina un minor rischio di malattia nel gruppo degli esposti rispetto ai non esposti? [fattore protettivo]

Misure relative o rapporti

- Vengono spesso chiamate complessivamente rischi relativi
- Si calcolano come rapporto tra misure di associazione
- Distinguiamo:
 - Rischio relativo (relative risk) o rapporto tra rischi (risk ratio)
 - Rapporto tra tassi (rate ratio)
 - Odds ratio o rapporto tra odds

Differenza tra rischi (*RD risk difference*) o riduzione assoluta del rischio (*ARR absolute risk reduction*)

- Al contrario dei rapporti rischio relativo e odds ratio, dipende dal livello di rischio iniziale
- *[ad esempio se il rischio negli esposti è 2% e nei non esposti 1% RR=2 ARR=1%; se i rischi fossero 20% e 10% avremmo RR=2 ma ARR=10%]*
- Viceversa una variazione del rischio di una data percentuale si traduce in diversi valori delle misure relative a parità di ARR
- Ha come valore corrispondente alla ipotesi di non effetto lo 0

Quanti trattamenti per prevenire un evento? (*NNT number needed to treat*)

- Nel caso di una esposizione nociva l'equivalente è NNH (*number needed to harm*)
- Si ottiene semplicemente calcolando il reciproco della differenza tra rischi ($1/ARR$)
- Esprime appunto quante persone devono essere sottoposte ad un intervento (terapeutico o preventivo) per evitare un evento (decesso o malattia)

Tabella 2x2

		Disease		
		Yes	No	
Exposure	Yes	a	b	a + b
	No	c	d	c + d
		a + c	b + d	N = a + b + c + d

		Disease		
		Yes	No	
Exposure	Yes	a	b	a + b
	No	c	d	c + d
		a + c	b + d	N = a + b + c + d

← Margini della tabella o valori marginali

- Rischio tra gli esposti

$$p(\text{malattia} | \text{exp}^+) = R(\text{exp}^+) = \frac{a}{a + b}$$

- Rischio tra i non esposti

$$p(\text{malattia} | \text{exp}^-) = R(\text{exp}^-) = \frac{c}{c + d}$$

- Rischio relativo

$$RR = \frac{R(\text{exp}^+)}{R(\text{exp}^-)}$$

		Disease		
		Yes	No	
Exposure	Yes	a	b	a + b
	No	c	d	c + d
		a + c	b + d	N = a + b + c + d

← Margini della tabella o valori marginali

- Rischio tra gli esposti

$$p(\text{malattia} | \text{exp}^+) = R(\text{exp}^+) = \frac{a}{a + b}$$

- Rischio tra i non esposti

$$p(\text{malattia} | \text{exp}^-) = R(\text{exp}^-) = \frac{c}{c + d}$$

- Riduzione assoluta del rischio

$$ARR = R(\text{exp}^+) - R(\text{exp}^-)$$

Don't eat potatoes RR

	Diarrheal Disease		Total
	Yes	No	
Ate potato salad	54	16	70
Did not eat potato salad	2	26	28
Total	56	42	98
Probability of disease, ate salad = $54/70 = 0.77$ Probability of disease, no salad = $2/28 = 0.07$ Risk ratio = $0.77/0.07 = 11$			

<http://www.ctspedia.org/do/view/CTSpedia/RateRatio>

Stata RR

	Exposed	Unexposed	Total	
Cases	54	2	56	
Noncases	16	26	42	
Total	70	28	98	
Risk	.7714286	.0714286	.5714286	
	Point estimate		[95% Conf. Interval]	
Risk difference		.7	.5629741	.8370259
Risk ratio		10.8	2.823513	41.31024
Attr. frac. ex.	.9074074		.6458313	.9757929
Attr. frac. pop		.875		

chi2 (1) = 40.02 Pr>chi2 = 0.0000

		Disease		
		Yes	No	
Exposure	Yes	a	b	a + b
	No	c	d	c + d
		a + c	b + d	N = a + b + c + d

- Gli odds

$$Odds(\text{exp}^+) = \frac{a/a+b}{b/a+b} = \frac{p(\text{malattia}^+ | \text{exp}^+)}{p(\text{malattia}^- | \text{exp}^+)} = \frac{p(\text{malattia}^+)}{1 - p(\text{malattia}^+)} = \frac{a}{b} = \frac{\text{eventi}}{\text{non_eventi}}$$

$$Odds(\text{exp}^-) = \frac{c/c+d}{d/c+d} = \frac{p(\text{malattia}^+ | \text{exp}^-)}{p(\text{malattia}^- | \text{exp}^-)} = \frac{p(\text{malattia}^+)}{1 - p(\text{malattia}^+)} = \frac{c}{d} = \frac{\text{eventi}}{\text{non_eventi}}$$

- L'odds ratio

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

Don't eat potatoes OR

	Diarrheal Disease		Total
	Yes	No	
Ate potato salad	54	16	70
Did not eat potato salad	2	26	28
Total	56	42	98
Probability of disease, ate salad = $54/16 = 3.38$			
Probability of disease, no salad = $2/26 = 0.08$			
Odds ratio = $3.38/0.08 = 44$			

Stata OR

	Exposed	Unexposed	Total	Proportion Exposed
Cases	54	2	56	0.9643
Controls	16	26	42	0.3810
Total	70	28	98	0.7143
	Point estimate		[95% Conf. Interval]	
Odds ratio	43.875		8.962168	401.0544 (exact)
Attr. frac. ex.	.977208		.8884199	.9975066 (exact)
Attr. frac. pop	.9423077			

chi2(1) = 40.02 Pr>chi2 = 0.0000

Odds e probabilità

- Sono simili se l'evento è raro
- È possibile calcolare la probabilità dagli odds:

$$P = \text{odds} / (1 + \text{odds})$$

- Gli odds sono meno intuitivi e quindi più difficili da comprendere rispetto alla probabilità ma egualmente leciti come misure epidemiologiche
- Un modello di analisi multivariabile molto utilizzato per variabili di risposta dicotomiche, la regressione logistica, produce come risultato log odds di evento per le variabile esplicative




Rapporto tra tassi di incidenza

- Il rapporto tra tassi (rate ratio)
- Confronta incidenza, mortalità o in generale la velocità di comparsa di un evento tra esposti e non esposti in uno studio di coorte
- Al denominatore dei tassi c'è la somma del tempo trascorso nel corso dello studio a rischio di subire l'evento per ciascuna delle unità sperimentali
- IL RATE RATIO è spesso INDICATO come RISCHIO RELATIVO (Relative Risk RR)

COX-2 selective non-steroidal anti-inflammatory drugs and risk of serious coronary heart disease. *Ray WA 2002*

- Results of premarketing and post-marketing trials have raised doubts about the cardiovascular safety of the non-steroidal anti-inflammatory drug (NSAID) rofecoxib, especially at doses greater than 25 mg. Between Jan 1, 1999, and June 30, 2001, **we did a retrospective cohort study** of individuals on the expanded Tennessee Medicaid programme (TennCare), in which we assessed occurrence of serious coronary heart disease (CHD) in non-users (n=202 916) and in users of rofecoxib and other NSAIDs (rofecoxib n=24 132, other n=151 728). Participants were aged 50–84 years, lived in the community, and had no life threatening non-cardiovascular illness.
- Users of high-dose rofecoxib were 1.70 (95% CI 0.98–2.95, p=0.058) times more likely than non-users to have CHD; **among new users this rate increased to 1.93 (1.09–3.42, p=0.024)**. By contrast, there was no evidence of raised risk of CHD among users of rofecoxib at doses of 25 mg or less or among users of other NSAIDs.

Table 2. Study events by NSAID*

	Person-years	Events	Rate/1000	Adjusted IRR (95% CI)	p
 Non-user	237 975	3085	13.0	1.00	..
Former user	125208	1403	11.2	0.97 (0.90–1.03)	0.294
Current user					
Ibuprofen	16330	190	11.6	0.91 (0.78–1.06)	0.216
Naproxen	21093	245	11.6	0.93 (0.82–1.06)	0.304
Celecoxib	5643	74	13.1	0.96 (0.76–1.21)	0.192
Rofecoxib ≤25 mg	4037	55	13.6	1.03 (0.78–1.35)	0.299
 Rofecoxib >25 mg	618	13	21.0	1.70 (0.98–2.95)	0.058
New user during study					
Ibuprofen	4319	52	12.0	1.01 (0.77–1.33)	0.954
Naproxen	6489	72	11.1	0.92 (0.73–1.16)	0.481
Celecoxib	4609	66	12.2	0.88 (0.67–1.16)	0.361
Rofecoxib ≤25 mg	3430	47	13.7	1.02 (0.76–1.37)	0.888
 Rofecoxib >25 mg	500	12	24.0	1.93 (1.09–3.43)	0.024

* Does not include 16 872 person years of follow-up and 251 events for other or multiple non-selective NSAIDs.

Stata IRR

	Exposed	Unexposed	Total
Cases	12	3085	3097
Person-time	500	237975	238475
Incidence rate	.024	.0129635	.0129867
	Point estimate	[95% Conf. Interval]	
Inc. rate diff.	.0110365	-.0025503	.0246232
Inc. rate ratio	1.851345	.9555646	3.238625 (exact)
Attr. frac. ex.	.4598522	-.0465017	.691227 (exact)
Attr. frac. pop	.0017818		
	(midp) Pr (k>=12) =		0.0247 (exact)
	(midp) 2*Pr (k>=12) =		0.0493 (exact)