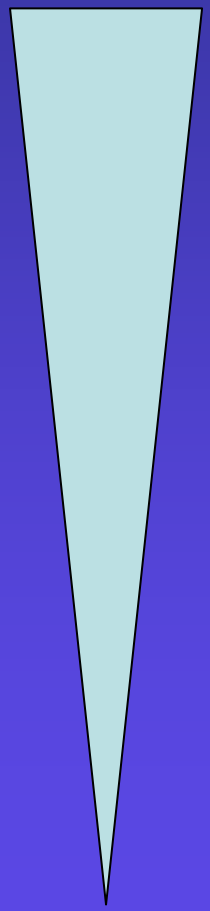
Study design

András Keszei

1st Budapest Clinical Epidemiology Course – organized jointly with the
15th Budapest Nephrology School

Elements of design

- Theoretical design
  - Research question
  - Occurrence relation
- Design of data collection
  - Time
  - Study population
  - Experimental or non-experimental
  - Choice of measurement
- Design of analysis



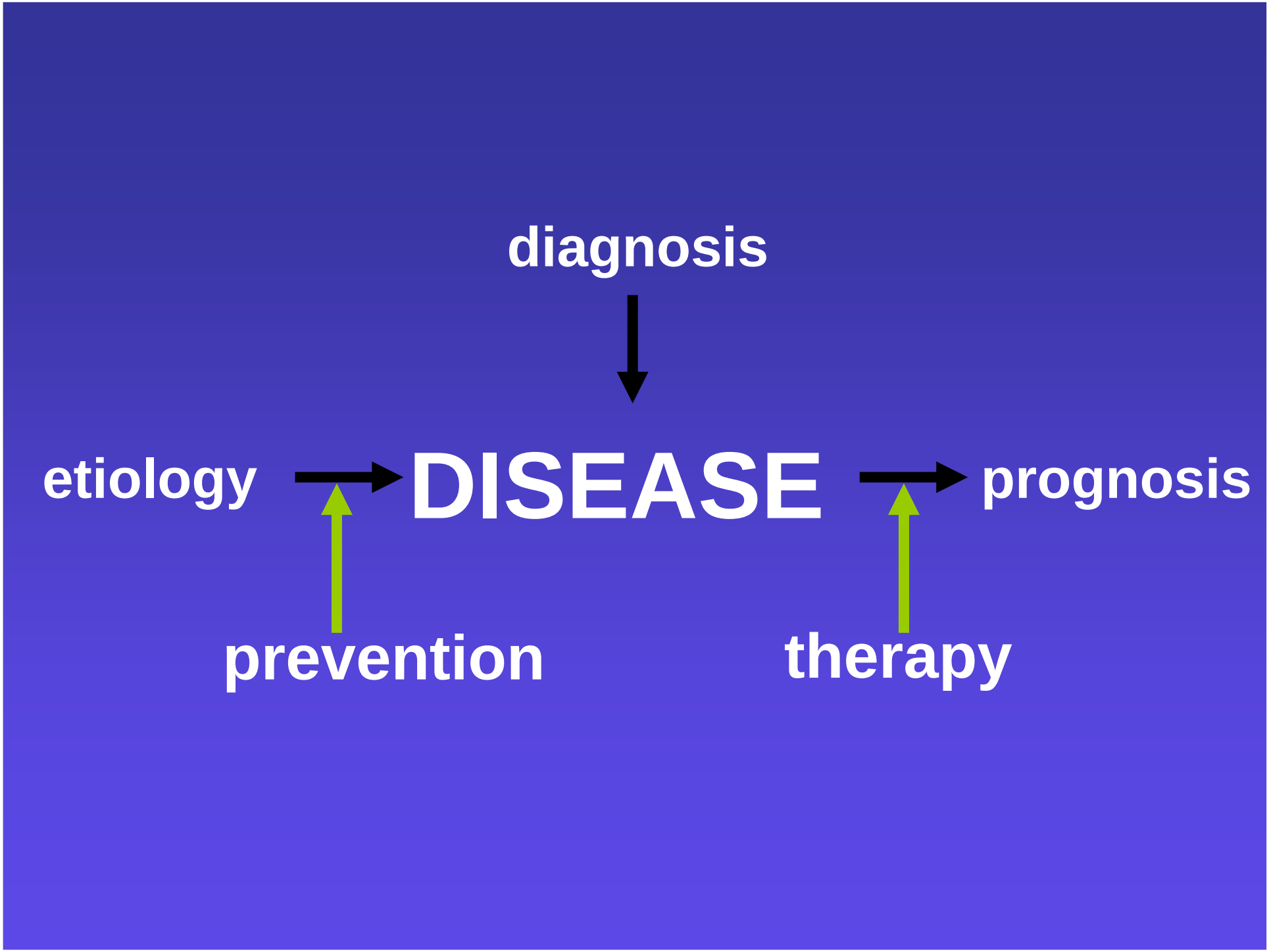
Theoretical design

Does five-day treatment with penicillin in children with acute otitis media reduce the duration of symptoms?

- Outcome
- Determinants
- Domain

Occurrence relation

Outcome ← Determinant(s)


diagnosis
etiology
DISEASE
prognosis
prevention
therapy

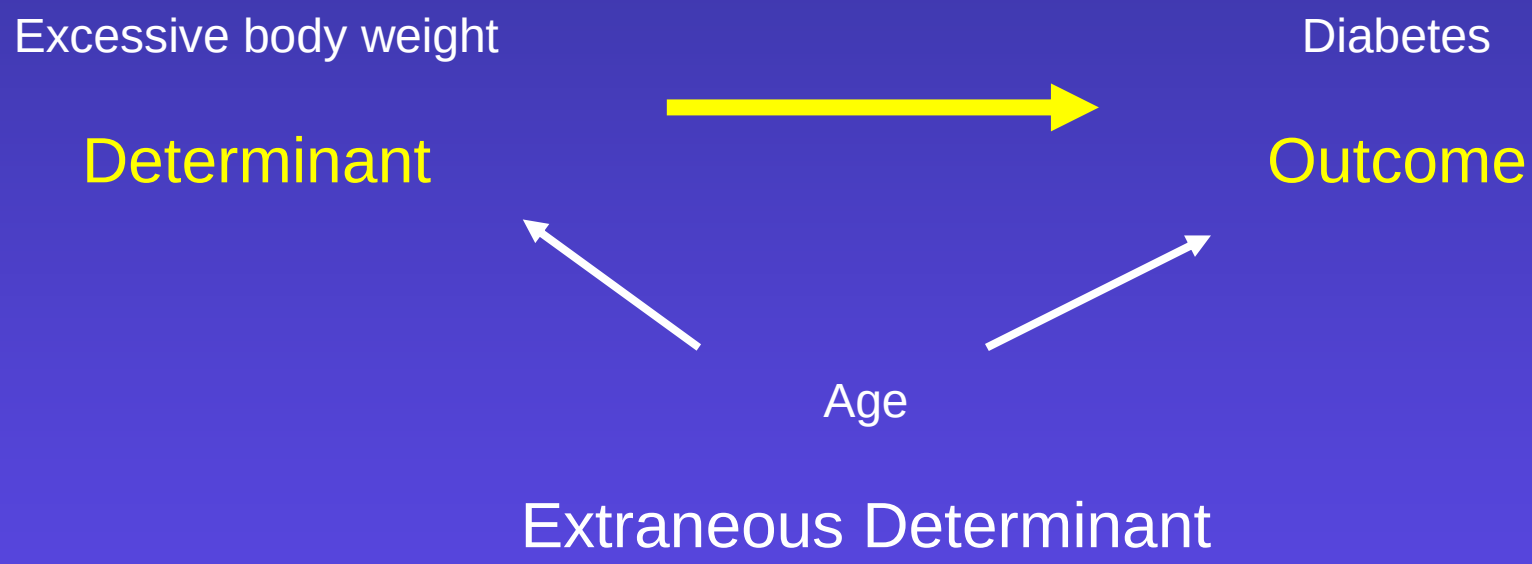
Types of epidemiologic research

- Etiologic
  - Does *exposure* cause *disease?*
  - One or many extraneous determinants
- Diagnostic
  - Predict probability of disease from (non)clinical profile
- Prognostic
  - Predict course of disease
- Intervention prognostic
  - Predict course of disease given treatment and profile

Etiologic research

- Theoretical design
  - Does *exposure* cause *disease?*
  - One or many extraneous determinants
  - outcome as a function of determinant conditional on confounders

Etiologic research



Etiologic research

- Theoretical design
  - Does *exposure* cause *disease?*
  - One or many extraneous determinants
  - outcome as a function of determinant conditional on confounders
  - longitudinal in nature

Etiologic research

- Design of data collection
  - Temporal relationship must be incorporated
  - Cohort study
  - Case-control study
  - Experimental or non-experimental

Cross-sectional studies: Exposure and disease are measured at the same moment among study participants

Examples of national surveys:

- NHANES (National Health and Nutrition Examination Study) USA
- MORGEN (Monitoren Risicofactoren Gezondheid en Ziekte in Nederland) Monitoring Risk factors for Health and Disease The Netherlands
- Hungarostudy 2002 Hungary

Ecological studies: correlation between (frequency of) exposure and disease at population-level

Examples:

- Meat consumption and risk of colon cancer (Armstrong and Doll, 1975)
- Wine consumption and coronary heart disease mortality (St Leger et al., 1979)
- Helicobacter Pylori infection and gastric cancer mortality (Forman et al., 1990)

Ecological studies: correlation between (frequency of) exposure and disease at <u>population-level</u>

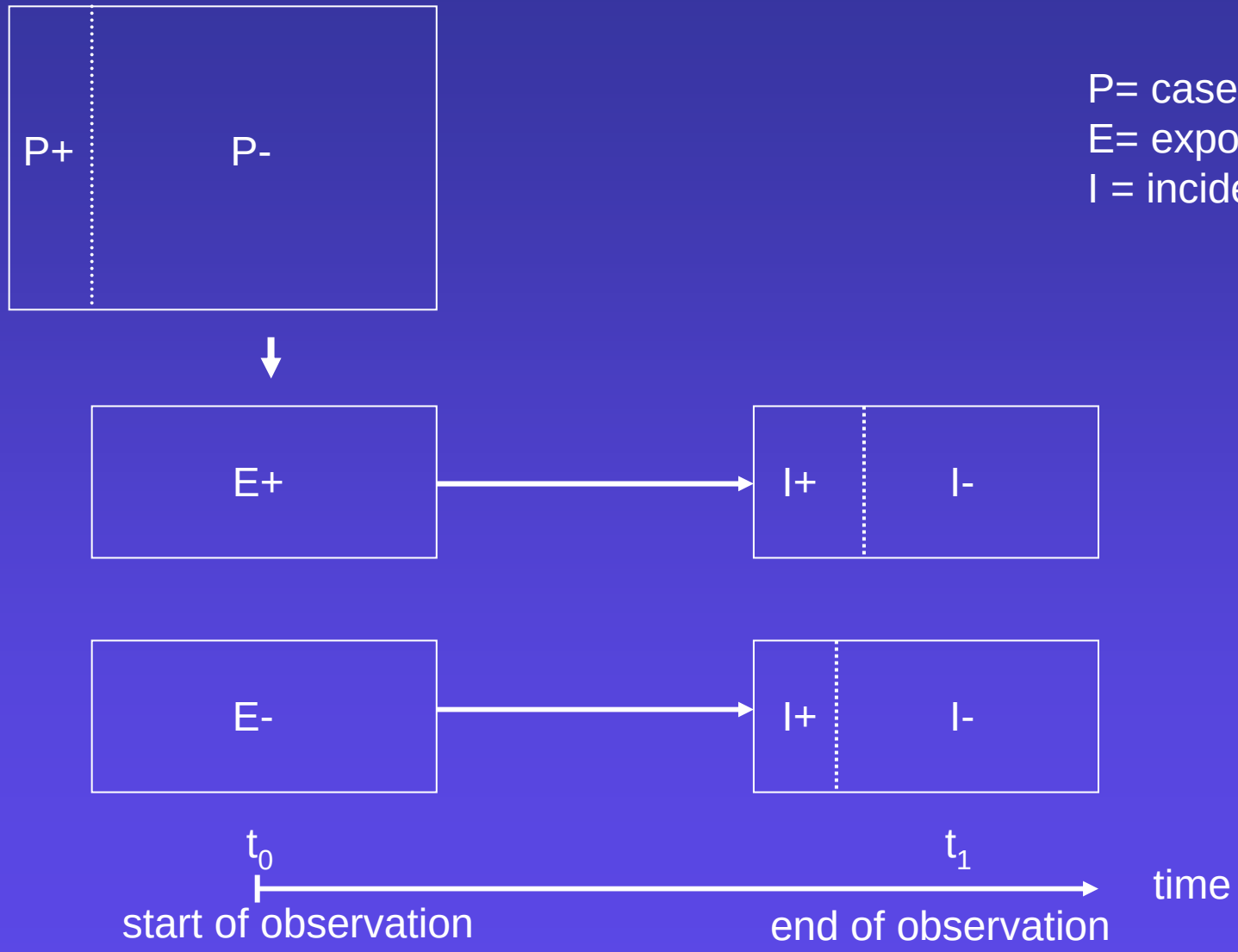
Advantages:

- Costs very low
- Quick results, short study duration
  data from population statistics and other registries

Disadvantages:

- Data at population level: *are the exposed the ones that are diseased?*
  ecological fallacy
- Regional differences in diagnostic procedures, exposure measurement
- Often lack of information on important variables (other exposures)

Cohort design



P= case; Outcome +/-
E= exposure +/-
I = incident cases

Cohort design

- Start with population not experiencing the outcome
- Risk of developing outcome should be present
- Exposure measurement at t_0
- Follow-up
- Determine incidence of outcome

Risk of coronary heart disease in men aged 47-55 years

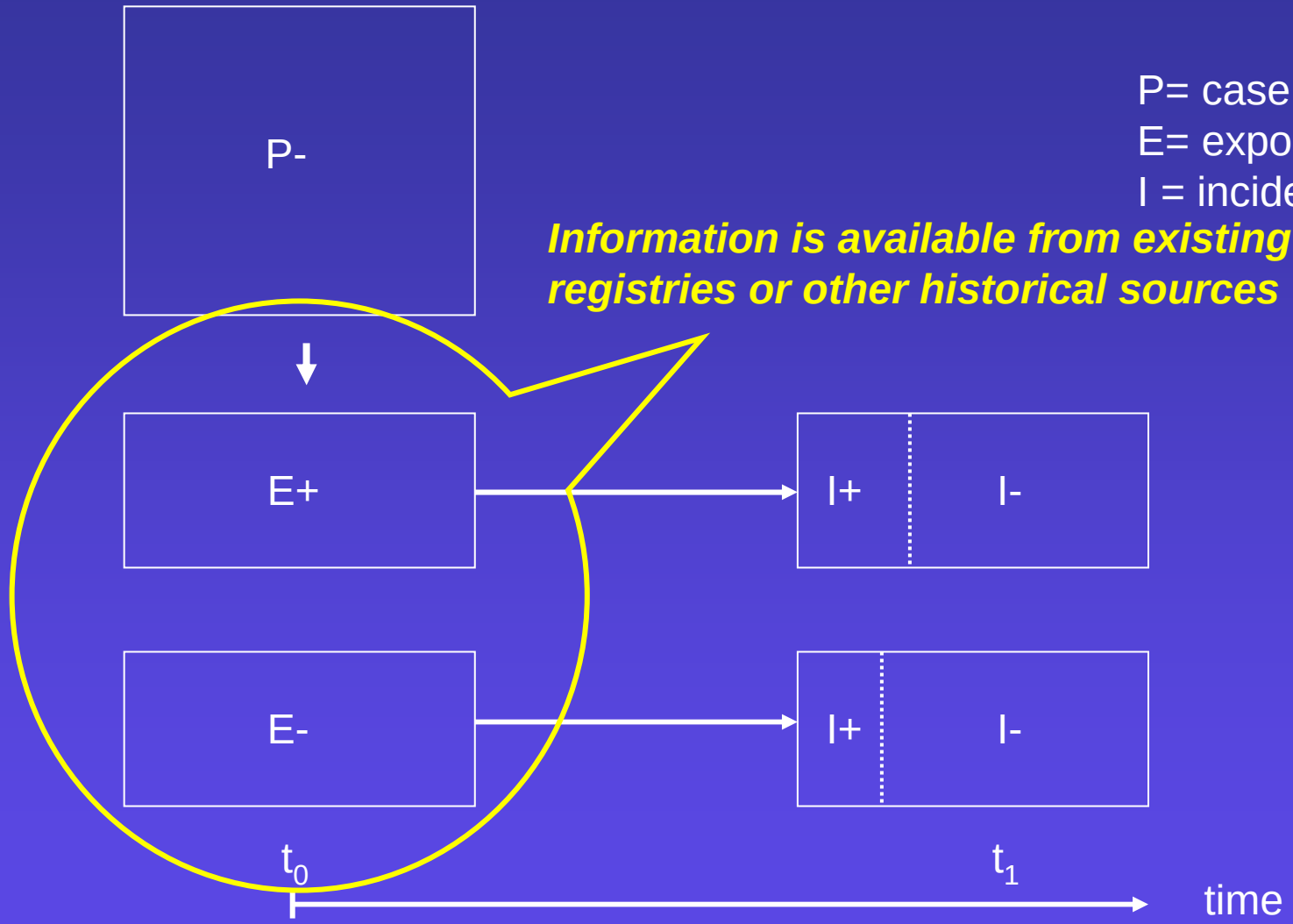
	N	Case	Cumulative incidence (%)
Never smoked	2212	88	4
Ex-smoker	1530	92	6
Current smoker	2806	281	10

11.8 year follow-up

Incidence of myocardial infarction among vasectomized and nonvasectomized men

Age (years)	Case	Person years	Incidence per 1000 person years
35-44	14	16 806	0,8
45-54	24	8 133	3,0
55-64	7	1 700	4,1
Total	45	26 639	1,7
35-44	56	83 057	0,7
45-54	110	40 971	2,7
55-64	49	8 570	5,7
Total	215	132 598	1,6

Retrospective Cohort design



Chest tube placement, video-assisted thoracoscopic surgery (VATS), or thoracotomy: effect on length of hospital stay. Shah SS et al. 2008.

Case-control design

> "The sophisticated use and understanding of case-control studies is the most outstanding methodological development of modern epidemiology."
> Rothman, 1986

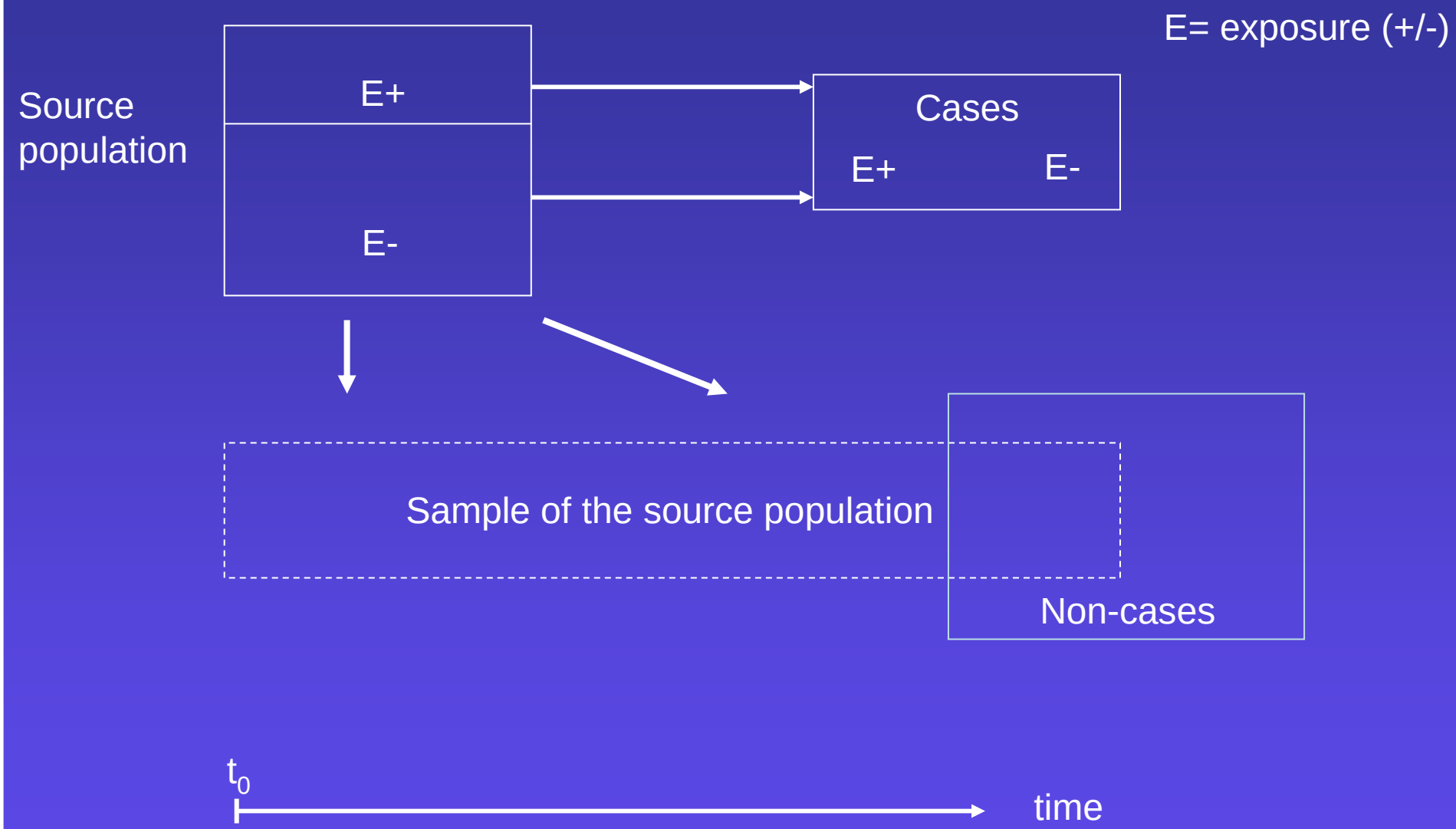
- Different terms used:
  - Case-referent study, TROHOC, Retrospective study
- Case-control studies have proven their potential value
  - Aspirin use and Reye syndrome (Hurwitz ES. et al. 1987)
  - Diethylstilboestrol use by pregnant women and occurrence of clear-cell vaginal carcinoma in their daughters (Herbst AL. et al. 1971)

Case-control design

- Efficiency
  - outcome of interest is rare
  - time between exposure and outcome is long
  - measurement of determinant is expensive, burdening or time-consuming
- Essence

Study *cases* and a *sample* of the source population

Case-control design



Case-control design

- Proper sampling of controls !
  - to ensure that measures of association will be similar to what would be obtained from a cohort study
  - controls should be representative sample of the source population from which cases are drawn

Case-control design

E+
E-

d_1 cases in T_1 person time

d_2 cases in T_2 person time

$$RR = \frac{d_1/T_1}{d_2/T_2}$$

↓ f sampling fraction

e+
e-

t_1 person time ~ fT_1

t_2 person time ~ fT_2

$$RR_s = \frac{d_1/t_1}{d_2/t_2} \sim \frac{d_1/fT_1}{d_2/fT_2} = RR$$

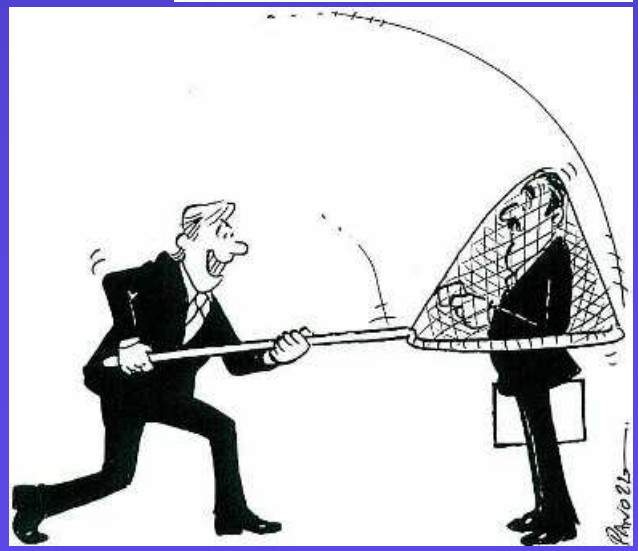
Sample of the source population

Swimming pool analogy

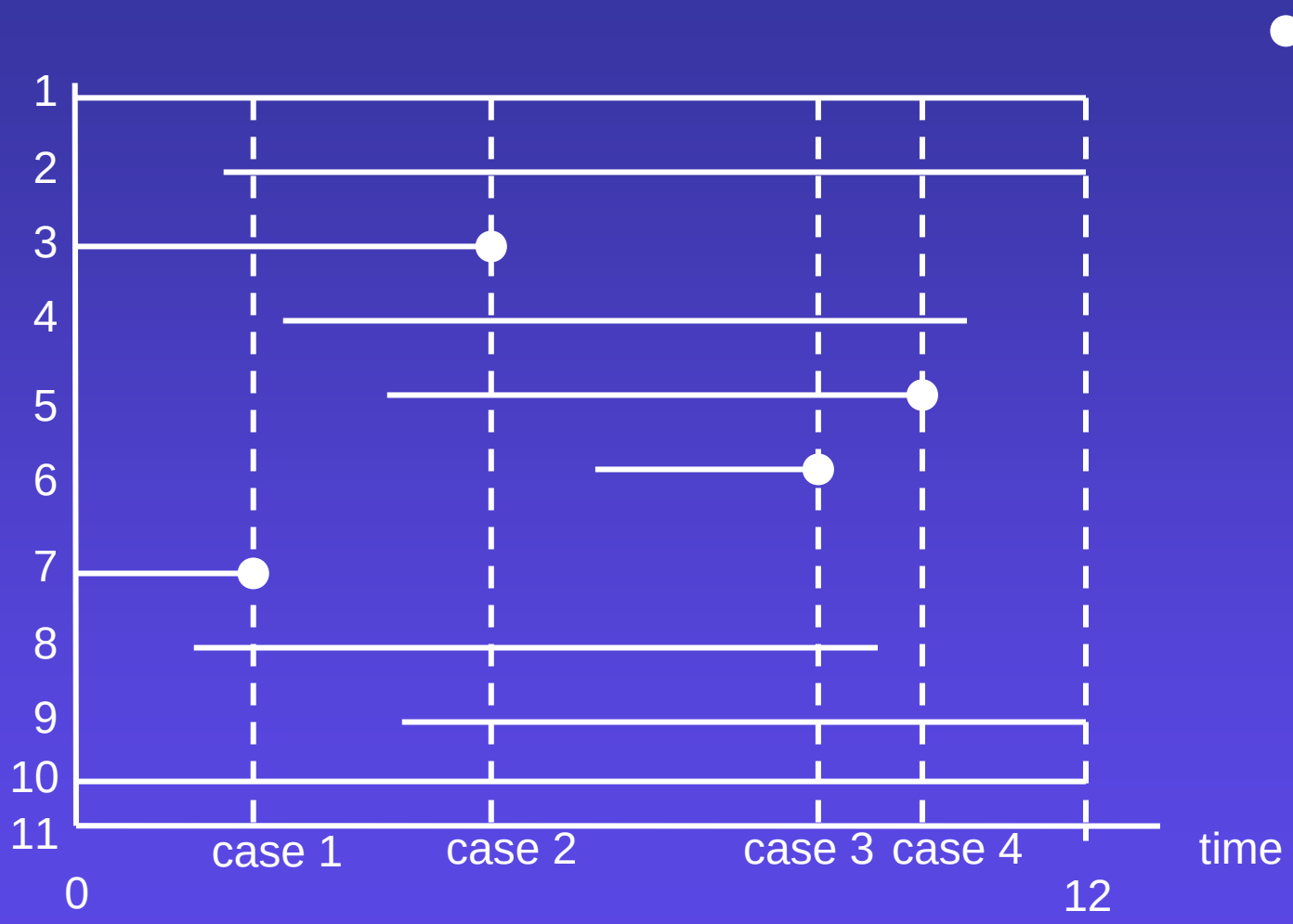
Pool - Source population

Lifeguard - Researcher

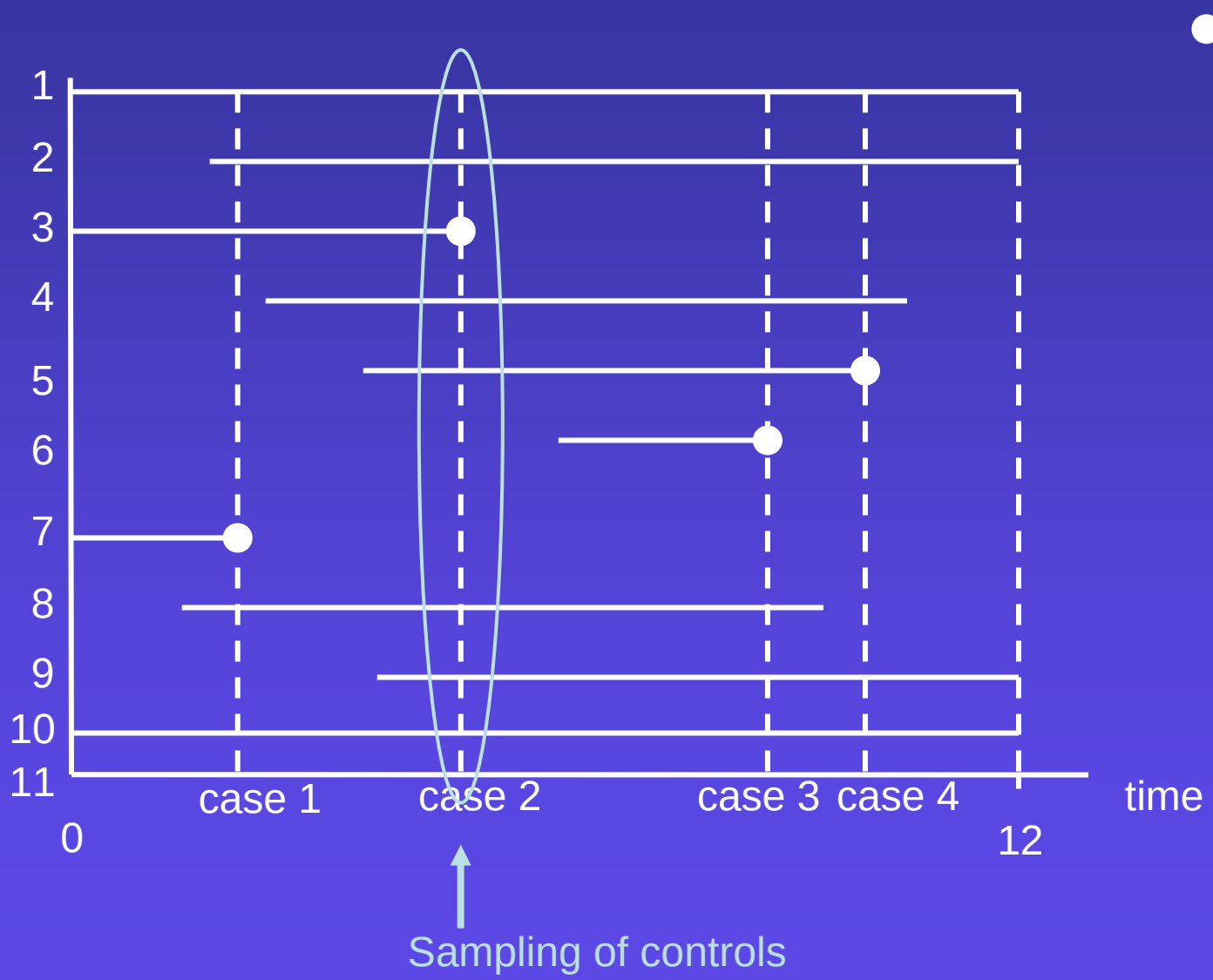
Net is used to catch those who become cases and to randomly sample other swimmers.



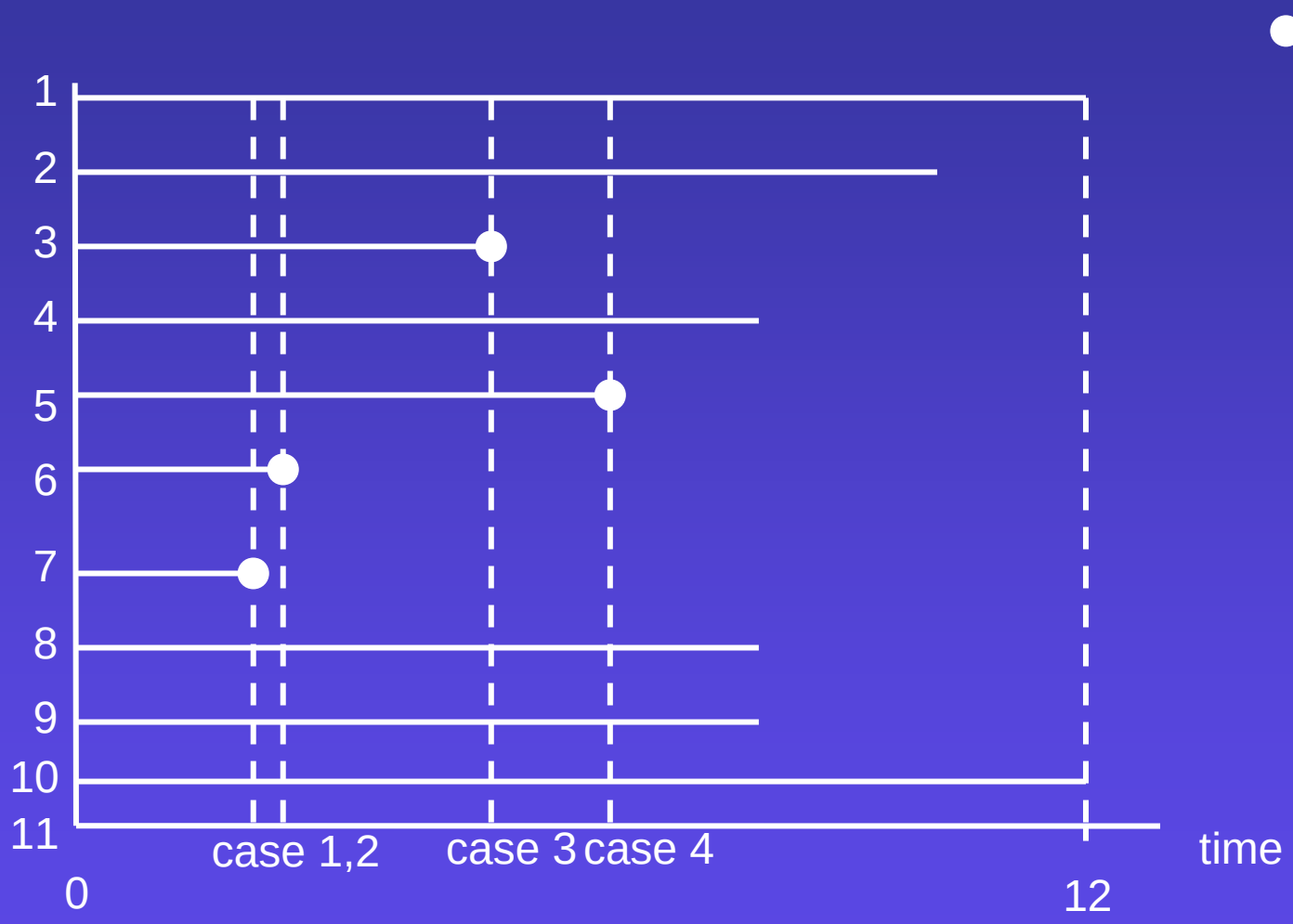
Control sampling - dynamic population



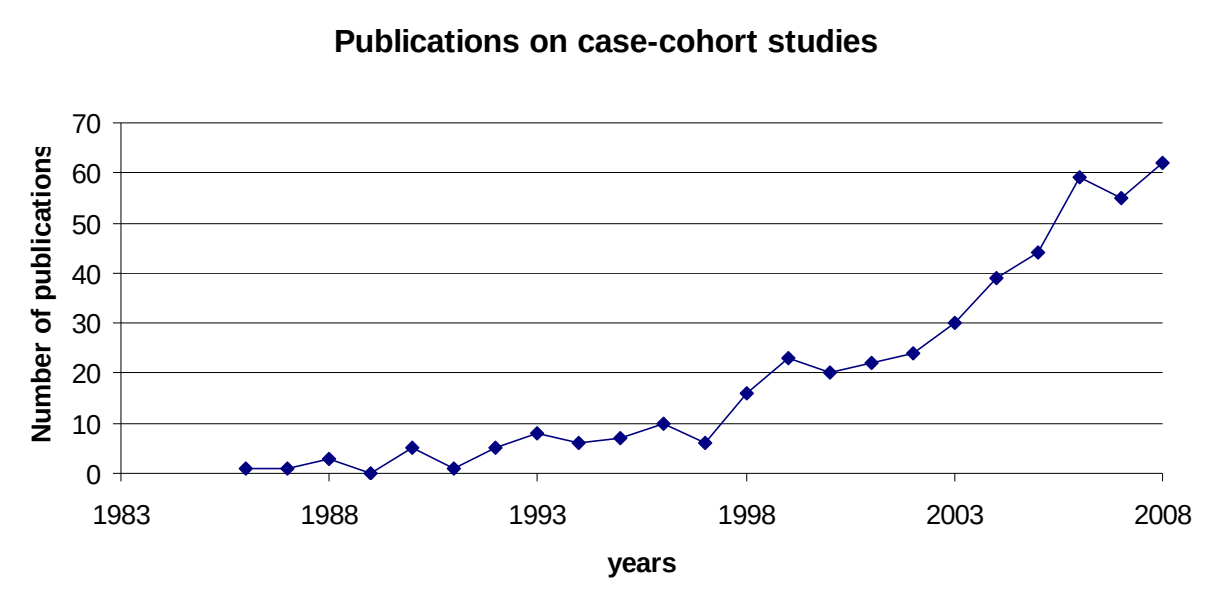
Control sampling - dynamic population



Control sampling - closed population



Case-cohort design



Case-control design

- Retrospective
  - All data on outcome, determinants and other factors are available when study is initiated
- Prospective
  - Cases are identified and recruited until enough cases have been included
  - A control group is sampled during the same time period

Etiologic research

- Data collection
  - outcome
  - determinant
  - confounders

Does dietary nitrate cause bladder cancer? Zeegers MP,2006

Data collection

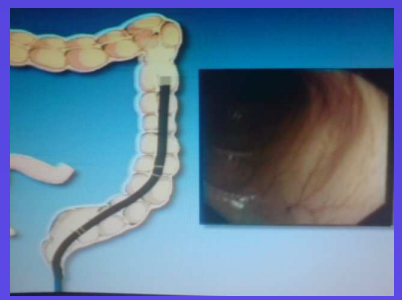
- Outcome
  - Registry
  - Hospital or physicians' charts
  - Morbidity Sentinel Programs
- Determinants
  - Questionnaire
  - Diary
  - Interview
  - Physical measurement

Instruments

1. non-invasive not harmful

2. non-invasive, but potentially harmful

3. invasive



Diagnostic research

- Theoretical design
  - Probability of disease as a function of multiple diagnostic tests
  - Domain
  - Goal
    - Optimal diagnostic strategy
    - Whether new diagnostic test provides added diagnostic value

N-terminal pro B-type natriuretic peptide in the detection of heart failure

	NT-proBNP positive (T+)	NT-proBNP negative (T-)	Total
Heart failure (D+)	9	0	9
Heart failure absent (D-)	69	55	124
	78	55	133

Sensitivity P(T+|D+)=100%
Specificity P(T- |D-)=44%

Positive predictive value P(D+|T+)=12%
Negative predictive value P(D-|T-)=100%

Likelihood ratio of positive test P(T+|D+)/P(T+|D-)=1.8
Likelihood ratio of negative test P(T-|D+)/P(T-|D-) =0

N-terminal pro B-type natriuretic peptide in the detection of heart failure

$$P(D+|T+) = \frac{P(D+)\cdot \text{sensitivity}}{P(D+)\cdot \text{sensitivity} + [1-P(D+)]\cdot(1-\text{specificity})}$$

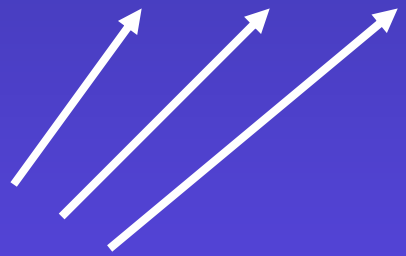
$$P(D-|T+) = \frac{P(D-)\cdot (1-\text{specificity})}{P(D+)\cdot \text{sensitivity} + [1-P(D+)]\cdot(1-\text{specificity})}$$

$$\frac{P(D+|T+)}{P(D-|T+)} = \frac{P(D+)}{P(D-)} * LR+$$

posterior odds

Posterior odds in sequential testing

$$\frac{P(D+|T_1+, T_2+, T_3+)}{P(D-|T_1+, T_2+, T_3+)} = \frac{P(D+)}{P(D-)} * LR(T_1+|D+) * LR(T_2+|D+) * LR(T_3+|D+)$$



3 positive tests

Assumes that the results of tests are independent of each other

Diagnostic research

- Design of data collection

Cross-sectional design

Case-control design

Diagnostic research

- Design of data collection
  - Case-control design

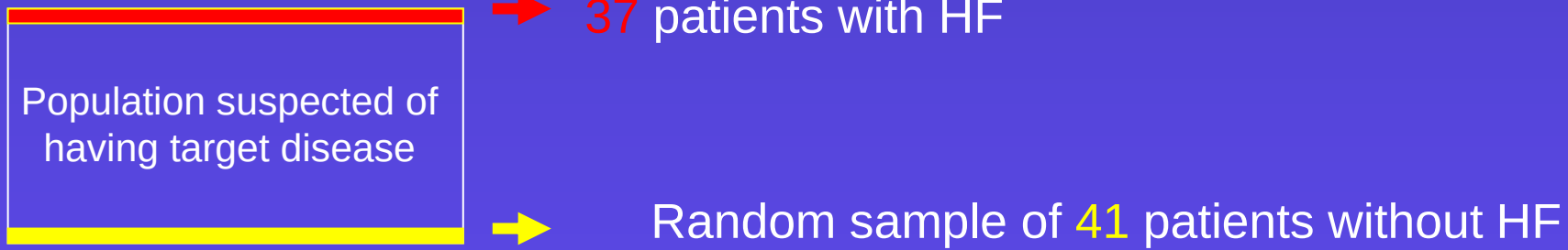
Attractive if measurement of determinant is time consuming or expensive

Diagnostic research

- Example

  Does results from cardiovascular magnetic resonance imaging (CMR) provide additional diagnostic information for identifying heart failure (HF) in patients with stable COPD. (Rutten FH. 2005)

n=405



Bias in diagnostic study

- Knowledge of results from diagnostic test may influence patient selection
  - Selective disease verification leads to bias
- Determinant influences outcome assessment
  - Incorporation bias, diagnostic review bias

Design of analysis diagnostic study

- Objective will guide analysis
  - Which determinant contributes to estimation of probability of disease
  - To what extent does it change the probability
  - To develop and validate a diagnostic model

Prognostic research

- Theoretical design
  - Outcome in the future as function of determinants measured at one (or more) time points before the outcome occurs.
- Domain
  - Individuals at risk of outcome
- Goal
  - Which determinant contributes to prediction
  - Does a marker provide *additional* predictive value

Prognostic research

- Design of data collection
  - longitudinal
  - cohort design Johansson, 1992 – prostate cancer
  - case-control design does not allow estimation of absolute risk
  - observational
  - prognostic study within a trial

Prognostic research

- Prognostic study within trial
  - Predictors of prolonged course in children with acute otitis media (Rovers MM. et al. 2007)
    - Control arm of randomized control trials
    - Outcome: fever and/or pain at 3-7 days

Prognostic research

- Bias in prognostic study
  - Determinant influences outcome assessment
  - Selective loss to follow-up
    - value of determinant is underestimated if poor prognosis is related to loss to follow-up

Intervention prognostic research

Effect of intervention

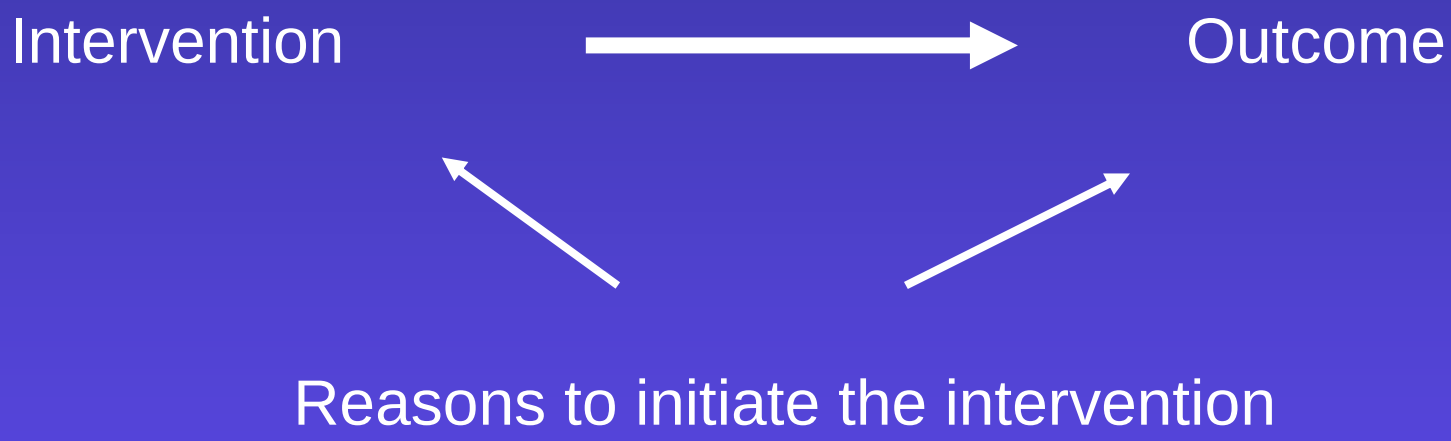
+

Natural history + Observer effect + Extraneous effects

Intervention prognostic research

- Comparability of natural history, extraneous effects and observer effects
  - Quasi-experiment
    - Comparison of two drugs in the treatment of leukemia
      - Select groups to be similar on age, proportion of males, severity, etc.

Intervention prognostic research



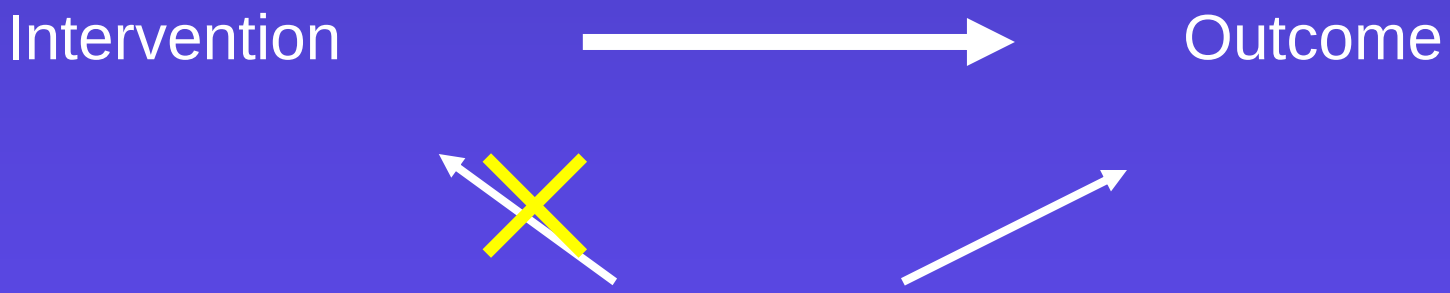
Confounding by indication

Comparison of death from cardiovascular causes in untreated and drug treated hypertensive women

	Rate ratio (95% Confidence interval
Crude value	1.0 (0.6 to 1.5)
Adjusted for: Age	0.7 (0.4 to 1.1)
+ Body mass index, pulse rate	0.6 (0.4 to 1.0)
+ Smoking, lipid levels	0.6 (0.4 to 0.9)
+ Diabetes	0.5 (0.3 to 0.9)

Intervention prognostic research

- Randomization 
  - Treatment is allocated at random
  - Differences in prognosis without treatment is result of random imbalances



Intervention prognostic research

- Randomized controlled trial
- Cohort study in which allocation of exposure is made by the investigator

Randomized Controlled Trial

Base population

Study population

In- / excusion criteria
Informed consent
Sampling

Measurement outcome other factors

R

Check prognostic factors

Intervention

Measurement outcome

Compa-rable?

Contrast?

Effect?

Check prognostic factors

Control intervention

Measurement outcome

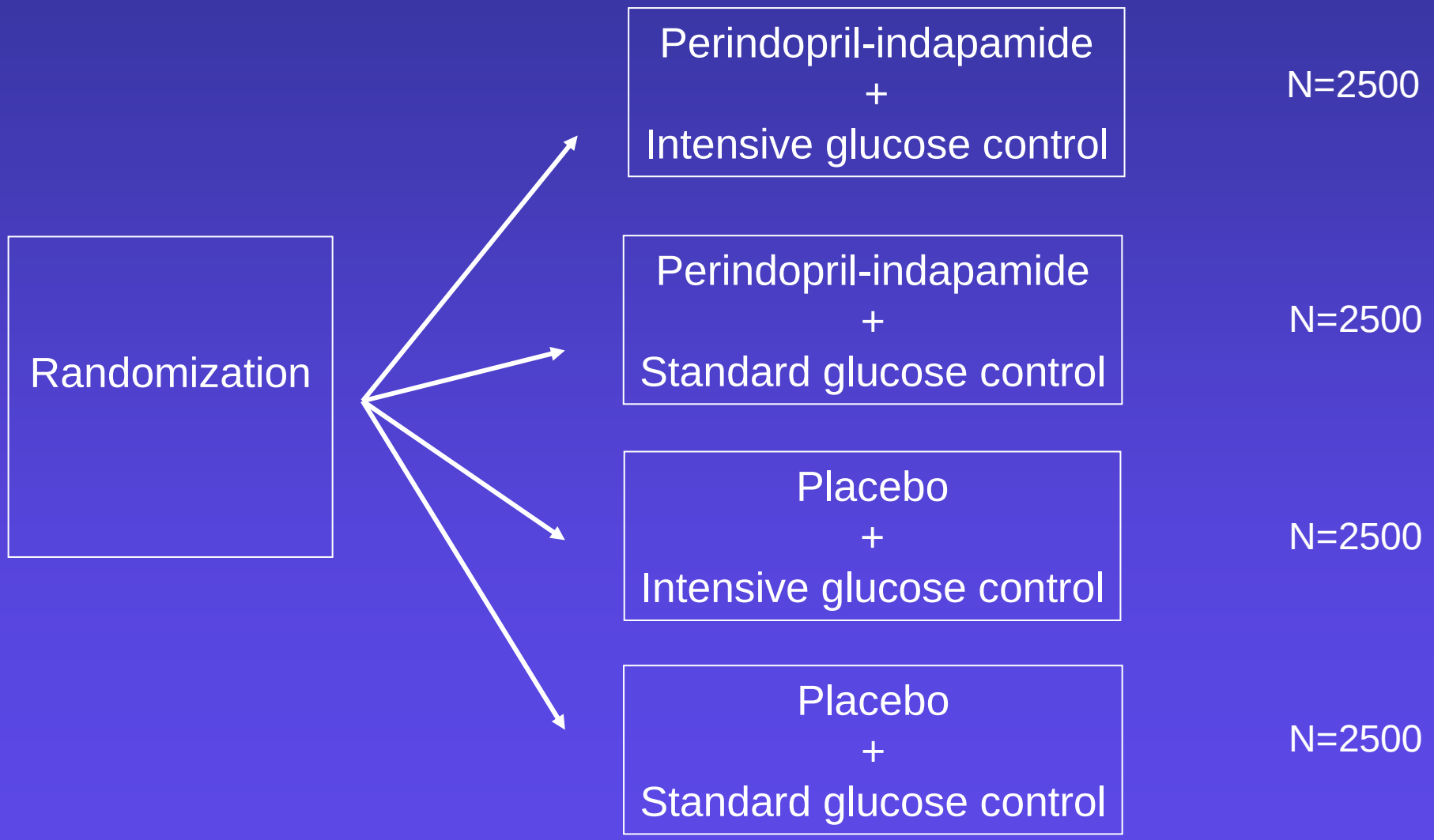
time

Intervention prognostic research

- Factorial design
  - Two treatment contrasts are studied
  - Efficient
  - Study interactions

A randomized factorial trial of vitamins C and E and beta carotene in the secondary prevention of cardiovascular events in women: results from the Women's Antioxidant Cardiovascular Study. Cook NR et al., 2008

ADVANCE study



Intervention prognostic research

- Crossover design
  - Comparison of effects within individual
  - Efficient
    - Between patient variability is not an issue
  - Limitations
    - Period effect
    - Carry-over effect
  - Suitable for quick and reversible treatment effects

Intervention prognostic research

- Cluster randomized design
  - Randomization at group level, ex. GP
    - Minimal intervention strategy in the treatment of low back pain in general practice (Jellema P. et al. 2005)

Intervention prognostic research

- Participants
  - Generalizability of findings
    - Inclusion/exclusion criteria
- Comparability of extraneous and observer effects
  - placebo, blinding

Intervention prognostic research

- Analysis
  - categorical data:
    - RR reduction
    - absolute risk reduction, number needed to treat
  - continuous data
    - e.g t-test
  - survival data
    - relative hazard
  - “intention to treat” or “ per protocol”

Key questions regarding the sample size

- What is the main purpose of the trial?
- What is the **principal** measure of patients **outcome**?
- How will the data be analyzed to detect a treatment difference? (The test statistic: **t-test , X^2**)
- What type of results does one anticipate in the non-exposed group?
- H_o and H_A, How small a **difference** is it important to detect and with what degree of **certainty**? (**δ, α and β.)**
- How to deal with treatment withdraws and protocol violations.

Sample size

Incidence in non-exposed group	Prevalence of exposure in population %	RR	Sample size per group	
			Cohort	Case-control
1/1000	50	1.2	576732	2535
		2.0	31443	177
		4.0	5815	48
1/100	50	1.2	57100	2535
		2.0	3100	177
		4.0	567	48
1/10	50	1.2	5137	2535
		2.0	266	177
		4.0	42	48

$\alpha=0.05$; $\beta=0.1$

J. Tonascia