New study designs

OMEGA-NET, 30 March 2022

Manolis Kogevinas

manolis.kogevinas@isglobal.org

ISGlobal Barcelona Institute for Global Health

Institució CERCA Centres de Recerca de Catalunya

A partnership of:

CLÍNIC BARCELONA Hospital Universitari · UNIVERSITAT DE BARCELONA · Parc de Salut MAR · upf. Universitat Pompeu Fabra Barcelona · Generalitat de Catalunya · Ajuntament de Barcelona · FUNDACIÓN RAMÓN ARECES

Outline

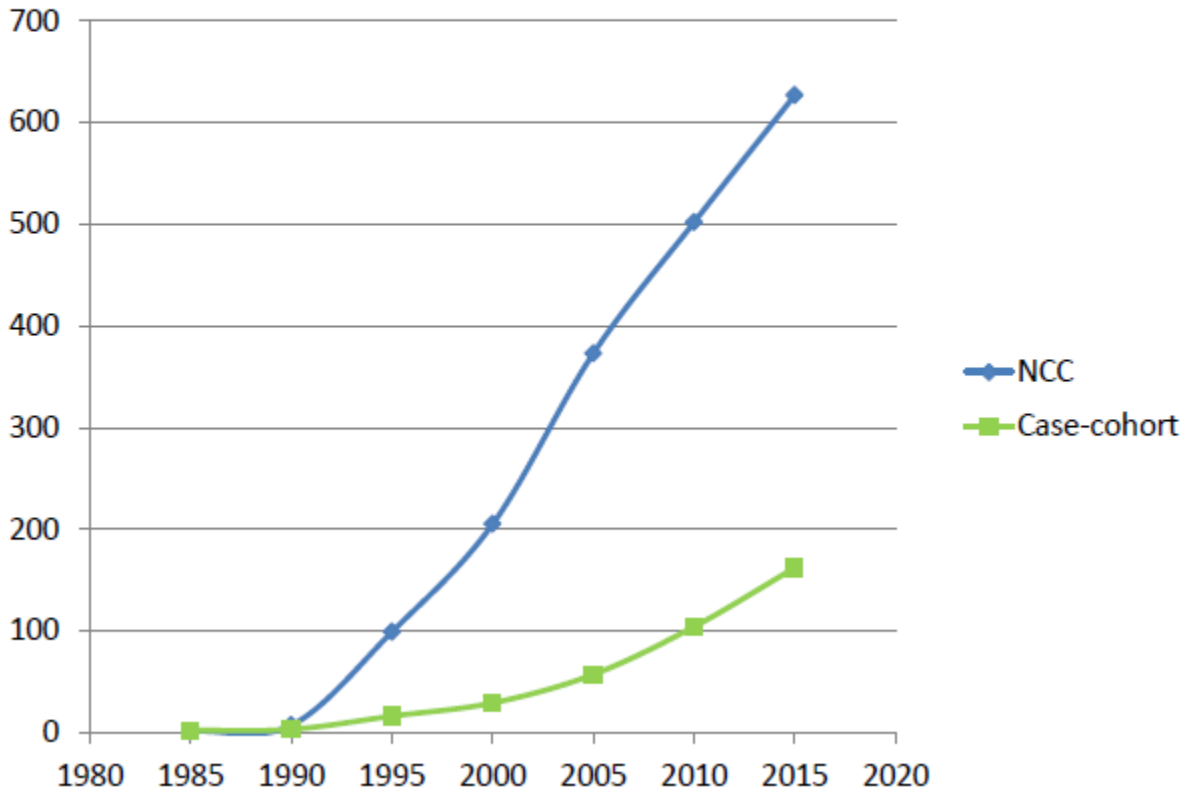
Present variations in “classical” case-control study design, focusing on rationale and design issues and less on analysis

- Nested case-control studies (*D Thomas 1977; Prentice and Breslow 1978*)
- Case-cohort studies (*Prentice 1986; Kupper et al 1975; Miettinen 1982*)
- Case-crossover studies (*Maclure M 1991*)

(some slides from A Johansson, H Checkoway, J Olsen and P Vineis)

ISGlobal Barcelona Institute for Global Health

References to nested case-control and case-cohort in Web of Science



(Slide from Anna LV Johansson, KI, Stockholm)

OMEGA-NET

DNA adducts in non-smokers and lung cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC)

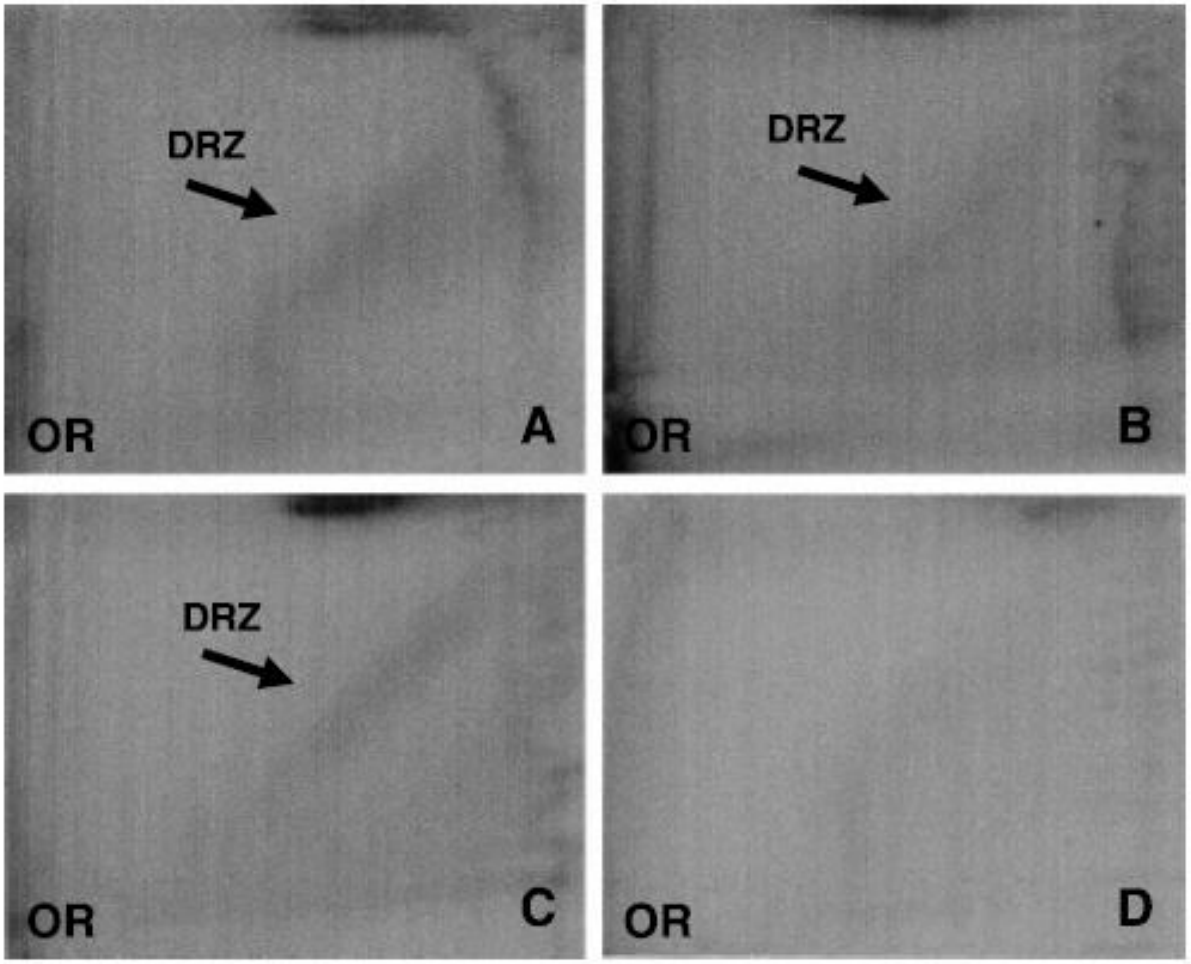
- Use a sensitive biomarker evaluating internal dose (DNA adducts) related to exposure to polycyclic aromatic hydrocarbons (PAH) in non-smokers
- PAHs are one of the major classes of carcinogens present in atmosphere capable to form DNA adducts leading to DNA damage after metabolic activation. When unrepaired, DNA adducts can cause mutations, which may ultimately induce cancer formation

(Peluso M, Cancer Res 2005)

OMEGA-NET

ISGlobal Barcelona Institute for Global Health

DNA adduct patterns (DRZ – diagonal radioactive zone) in the chromatograms of two never smokers who developed lung cancer during follow-up (top) and two never smokers who did not develop (bottom)



(Peluso M, Cancer Res 2005)

ISGlobal Barcelona Institute for Global Health

DNA adducts in non-smokers and lung cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC)

Cases included newly diagnosed lung cancer (n = 115), upper respiratory cancers (pharynx and larynx; n = 82), bladder cancer (n = 124), leukemia (n = 166), and chronic obstructive pulmonary disease or emphysema deaths (n = 77) accrued after a median follow-up of 7 years among the EPIC former smokers and never-smokers.

(Peluso M, Cancer Res 2005)

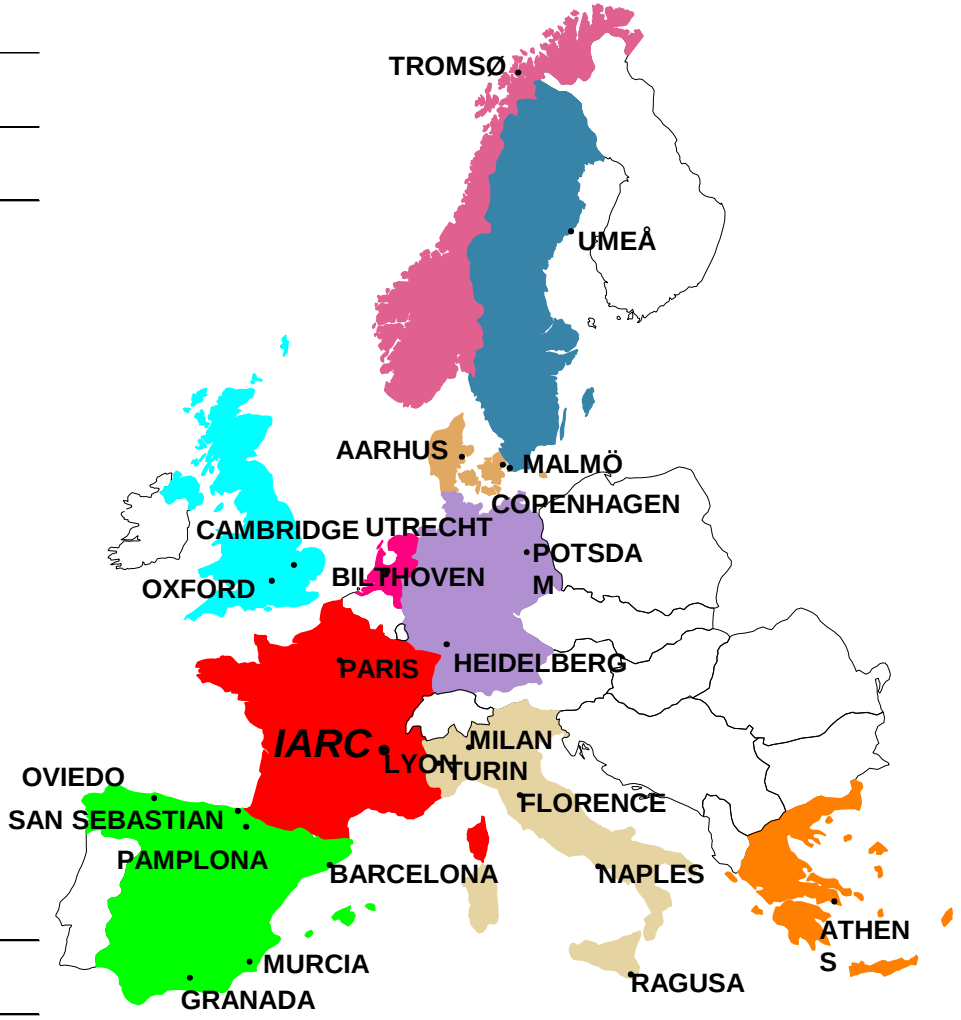
OMEGA-NET

ISGlobal Barcelona Institute for Global Health

EPIC

Collaborating centres and cohort subjects

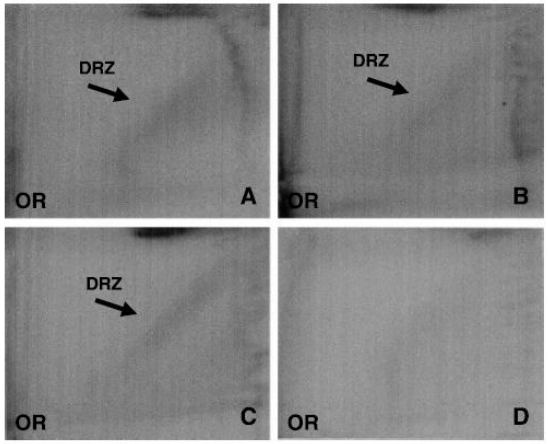
	Subjects included	
	Questionnaire	Q + Blood
France	74 524	21 053
Italy	47 749	47 725
Spain	41 440	39 579
UK	87 942	43 141
Netherlands	40 072	36 318
Greece	28 555	28 483
Germany	53 091	50 678
Sweden	53 826	53 781
Denmark	57 054	56 131
Norway	37 215	11 000
All	521 468	387 889



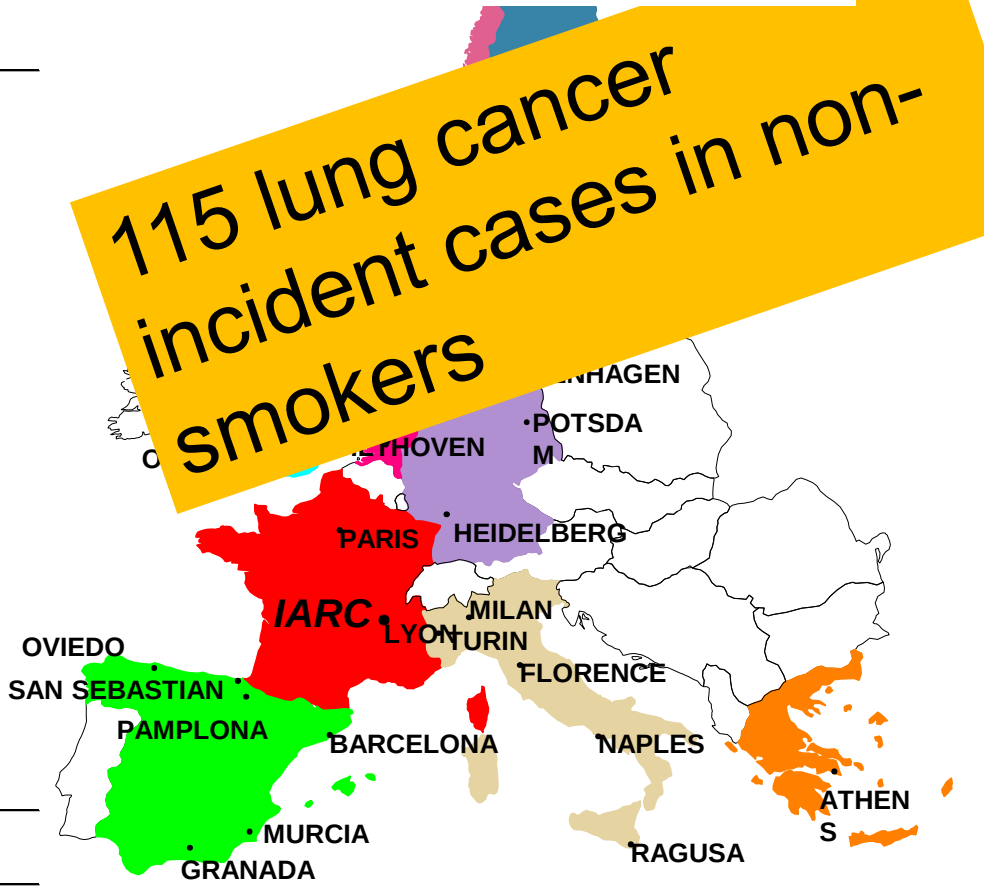
EPIC

Collaborating centres and coh

	Subjects included	
	Questionnaire	Q + Blood
France	74 524	21 053
Italy	47 749	47 725
Spain	41 440	39 579
UK	87 942	43 141
Netherlands	40 072	36 318
Greece	28 555	28 483
Germany	53 091	50 678
Sweden	53 826	53 781
Denmark	57 054	56 131
Norway	37 215	11 000
All	521 468	387 889



115 lung cancer incident cases in non-smokers



Cohort study on DNA adducts in non-smokers and lung cancer

	No lung cancer	Lung cancer	Total
Exposed (high DNA adducts) *	a	b	40.000
Non-exposed (low DNA adducts	c	d	160.000
	199.885	115	200.000

* Assume 20% of the population with high levels of DNA adducts

ISGlobal Barcelona Institute for Global Health

Cohort study on DNA adducts in non-smokers and lung cancer

	No lung cancer	Lung cancer	Total
Exposed (high DNA adducts)	a		
Non-exp (low D adduct		d	160.000
	199.885	115	200.000

Not efficient!

ISGlobal Barcelona Institute for Global Health

Nested case-control study on DNA adducts in non-smokers and lung cancer

	No lung cancer	Lung cancer	Total
Exposed (high DNA adducts)	a’	b	69
Non-exposed (low DNA adducts	c’	d	276
	230	115	345

* Assume 20% of the population with high levels of DNA adducts

ISGlobal Barcelona Institute for Global Health

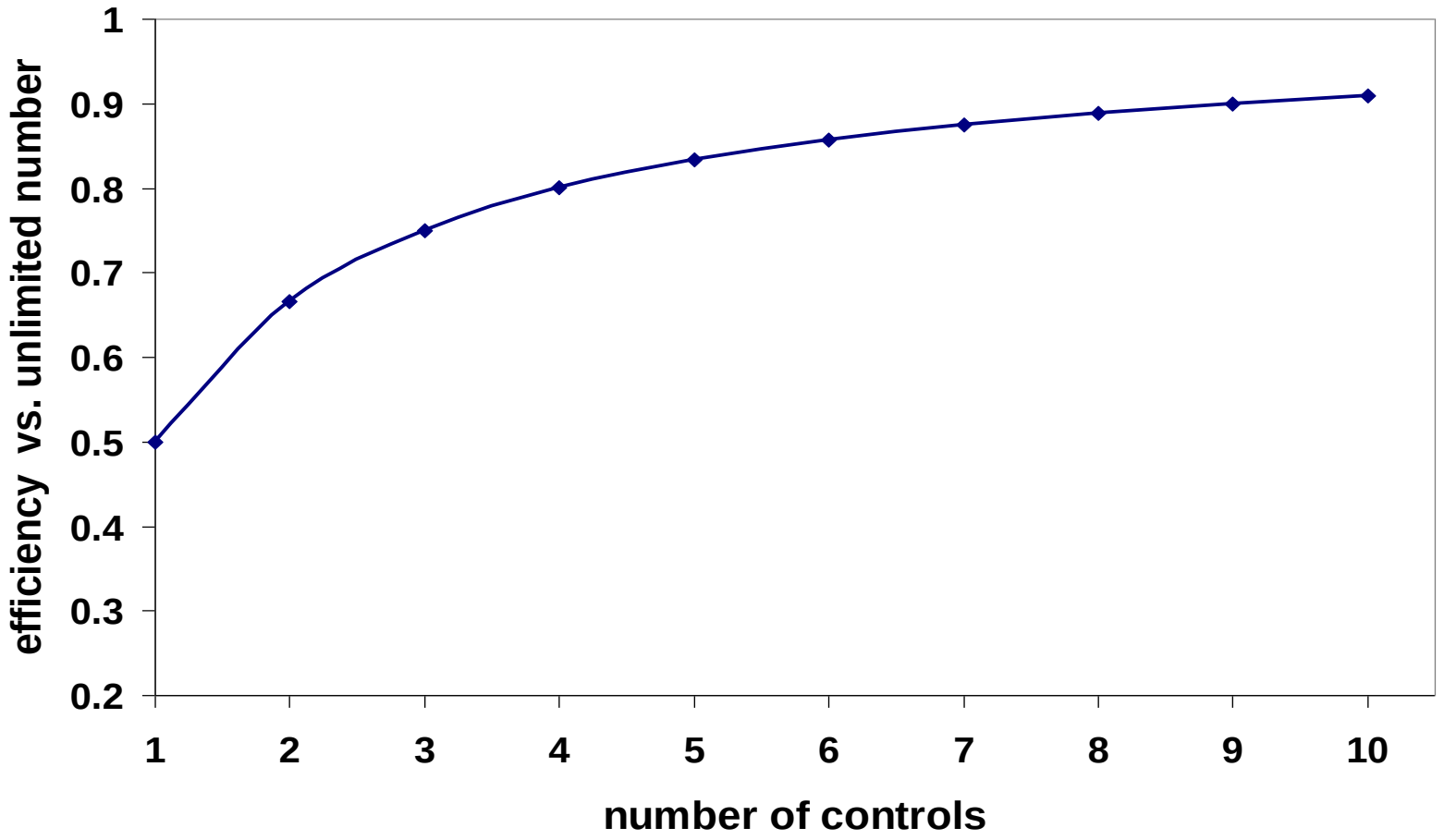
How many controls per case?

Very small gain for more than 5 controls per case

If there are *m* times controls per case, the precision of the nested case-control study study compared to the cohort study is given by the formula:

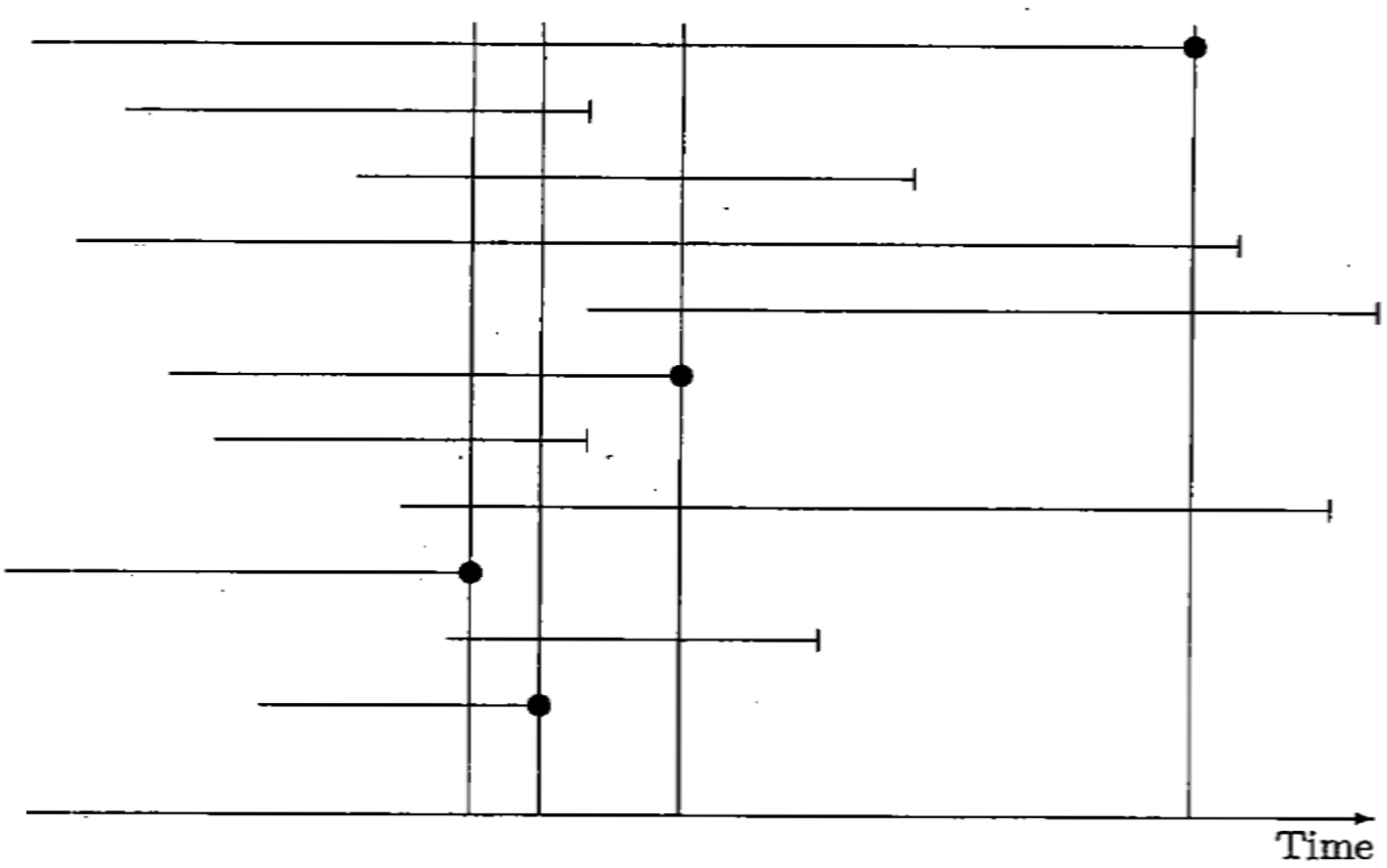
$$\sqrt{1+1/m}$$

Number of controls per case



Efficiency for a fixed number of cases: M/(M+1), where M is the number of controls, compared to a study with unlimited number of controls

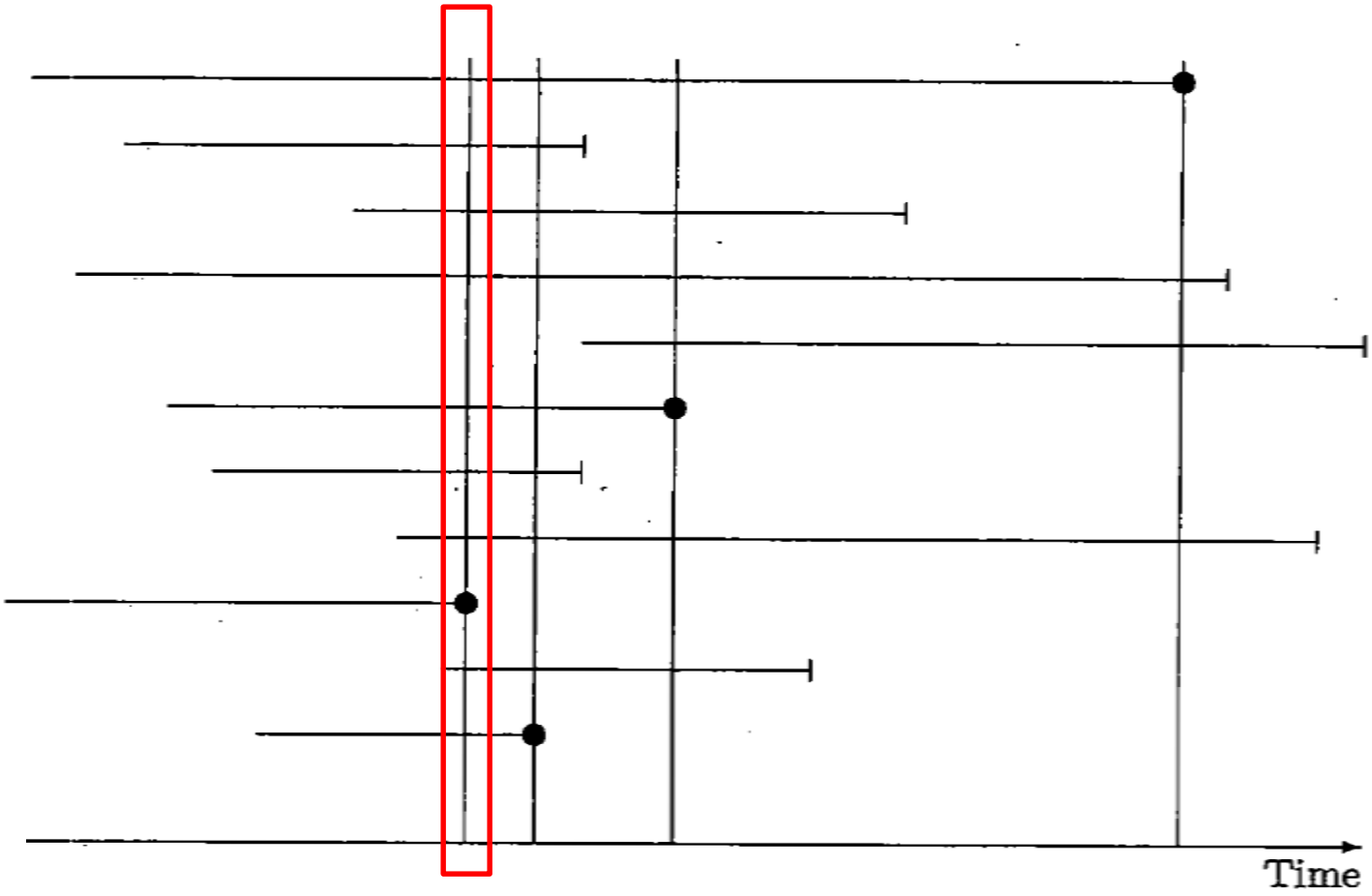
Risk sets for a follow-up study (n=11)
Black dots indicate the 4 subjects who “fail”



OMEGA-NET

ISGlobal Barcelona Institute for Global Health

Risk sets for a follow-up study (n=11)
Black dots indicate the 4 subjects who "fail"

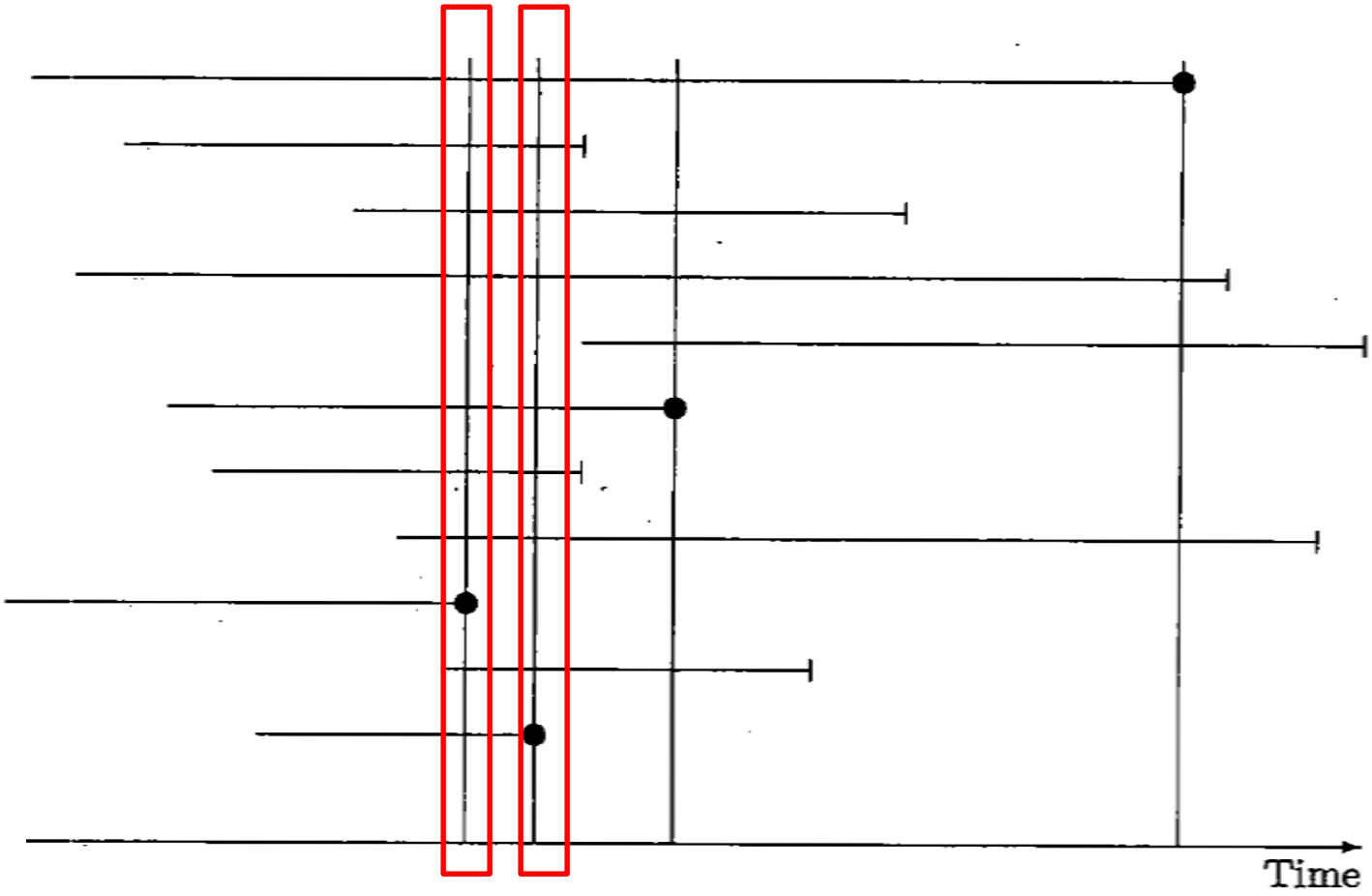


OMEGA-NET

ISGlobal Barcelona Institute for Global Health

Risk sets for a follow-up study (n=11)

Black dots indicate the 4 subjects who "fail"

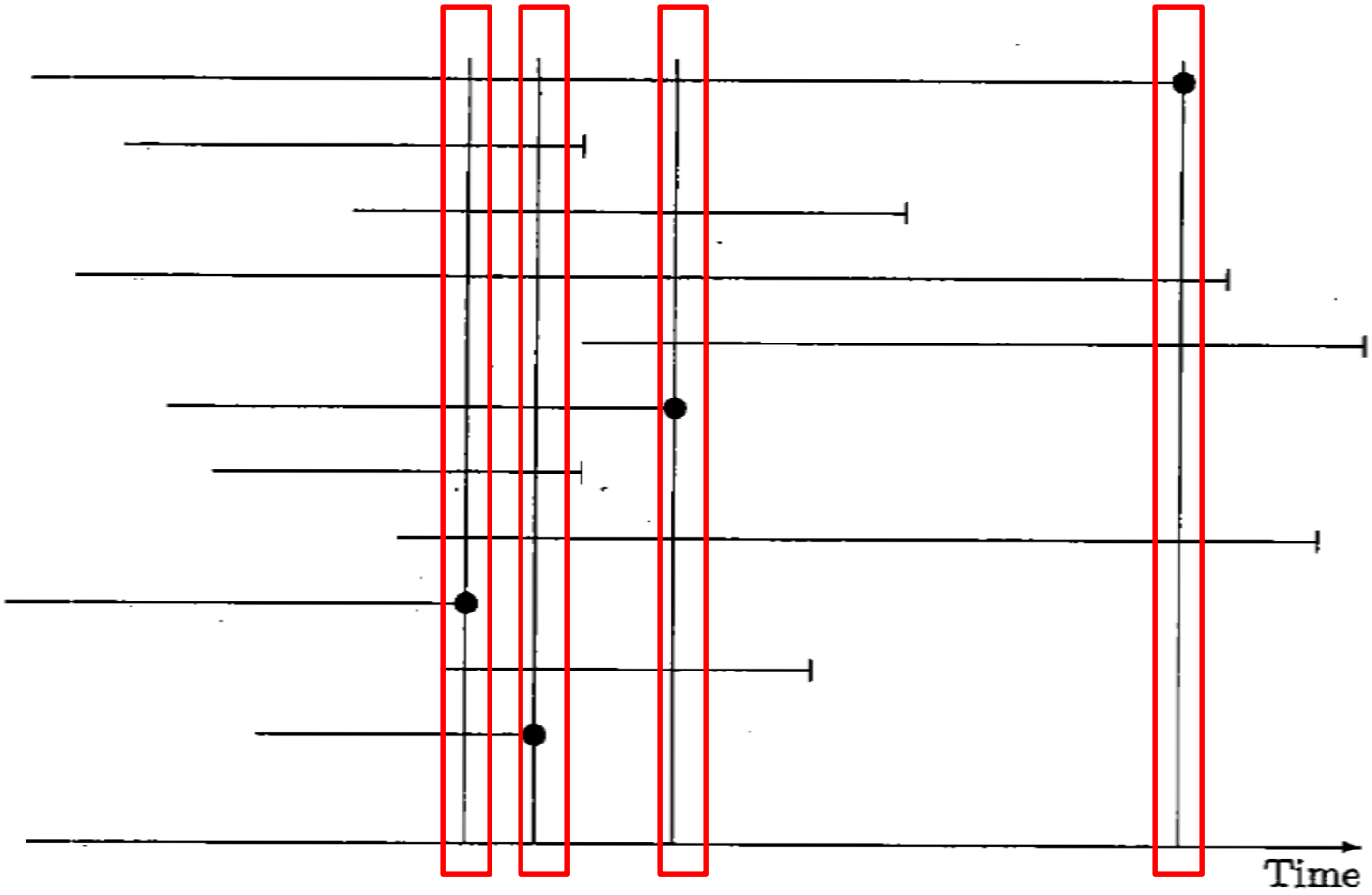


OMEGA-NET

ISGlobal Barcelona Institute for Global Health

Risk sets for a follow-up study (n=11)

Black dots indicate the 4 subjects who "fail"



OMEGA-NET

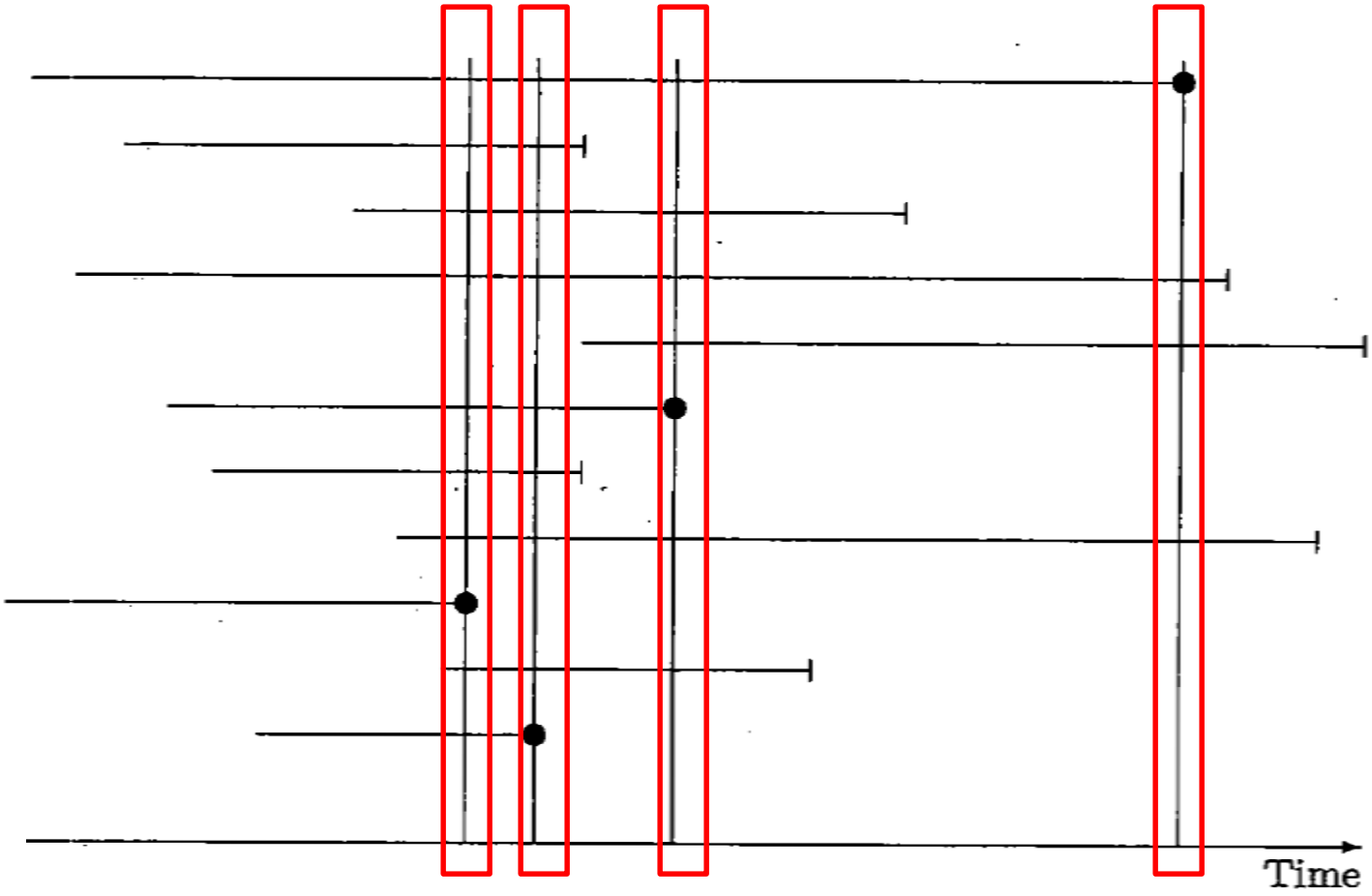
ISGlobal Barcelona Institute for Global Health

Risk sets for a follow-up study (n=11)

Black dots indicate the 4 subjects who "fail"

How many subjects do you have in each risk set?

How would you select a single control per case?



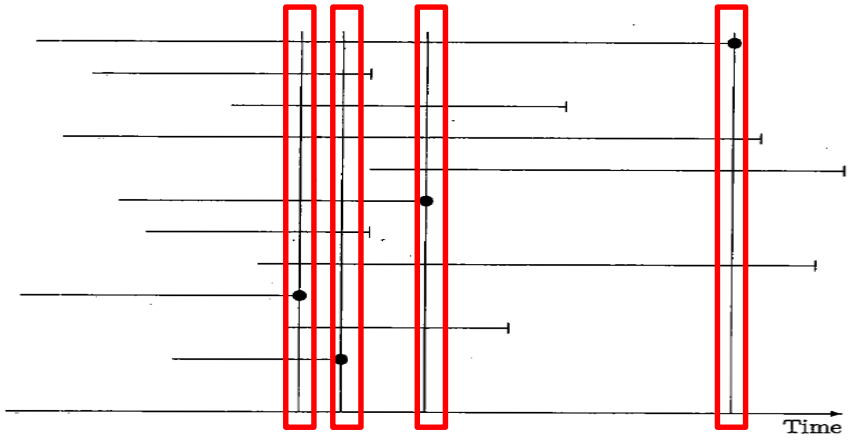
OMEGA-NET

ISGlobal Barcelona Institute for Global Health

Nested case-control studies

- Sampling within a cohort; select all cases
- Controls are **time-matched** to cases
  - a control can be selected more than once
  - a control can become a case
- Controls can be used only for one outcome

(with some exceptions if sampling fraction known)

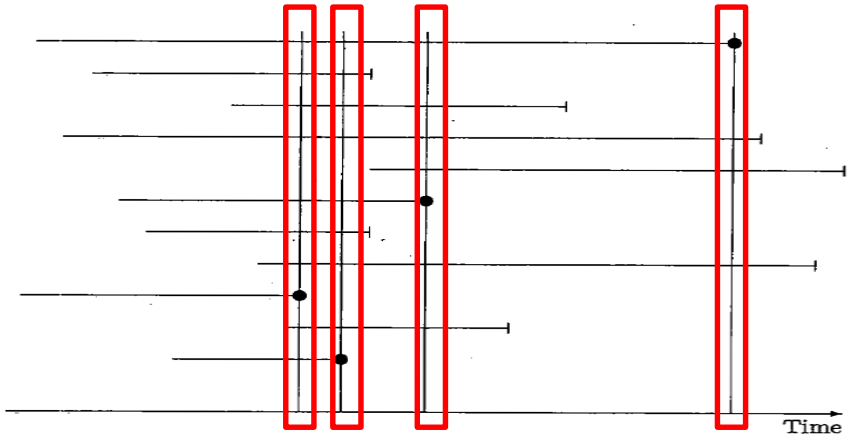


OMEGA-NET

ISGlobal Barcelona Institute for Global Health

Nested case-control studies

- selection of controls frequently involves matching on confounders (e.g. length of follow-up, age, gender etc)
- analysis using conditional logistic regression, conditioning on riskset
- the odds ratio (OR) estimates the underlying Hazard Ratio in the cohort



OMEGA-NET

ISGlobal Barcelona Institute for Global Health

Summary: why do a nested case-control study?

- The main reason is to reduce labour and cost of data collection; complete data only for subjects chosen for the nested study
- Other reasons:
  - new hypotheses not in original design that involve collection of new data on exposures/confounders
  - avoid computational burden (e.g. time-dependent variables)

Concerns with the Nested Case-Control Design

Due to the risk set matching,

- the control series is not intuitively understood
- controls are not representative of the cohort population
- the control series has few other uses, so the investment in biomarker analyses cannot be leveraged for other research

- Growing interest in case-cohort analyses

Case-Cohort Design

OMEGA-NET

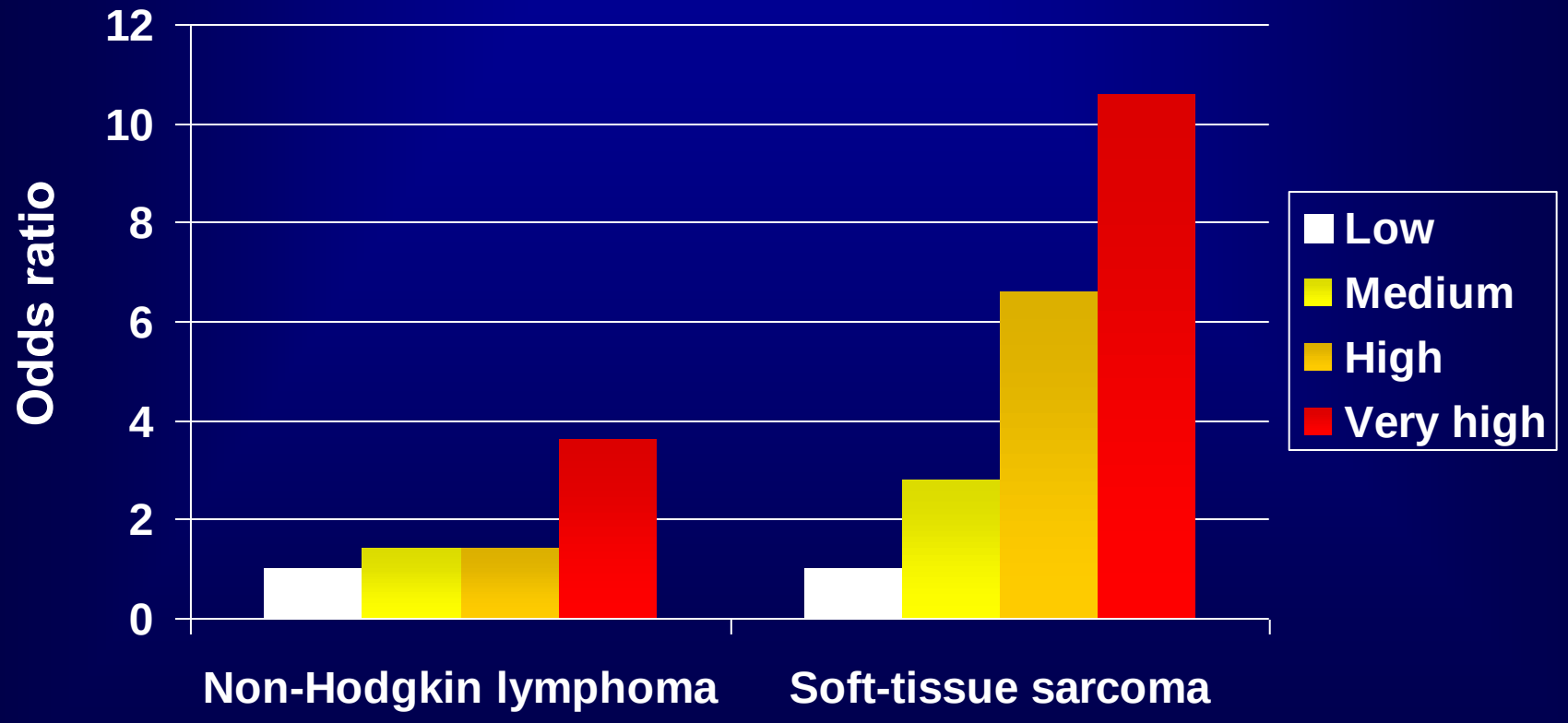
ISGlobal Barcelona Institute for Global Health

SMRs (Standardized Mortality Ratio) for different cancers. IARC study on 21,863 workers exposed to phenoxy herbicides, chlorophenols and dioxins (TCDD), 1939-1992 (Kogevinas AJE 1997)

Cause of death (ICD-9 codes) *	Workers exposed to TCDD or higher chlorinated dioxins		
	No. of deaths	SMR *	95% CI *
All causes	2728	1.00	0.97-1.04
All malignant neoplasms	710	1.12	1.04-1.21
Connective and other soft-tissue 171	6	2.03	0.75-4.43
Non-Hodgkin lymphoma 200, 202	24	1.39	0.89-2.06
Hodgkin's lymphoma 201	8	1.29	0.56-2.53

ISGlobal Barcelona Institute for Global Health

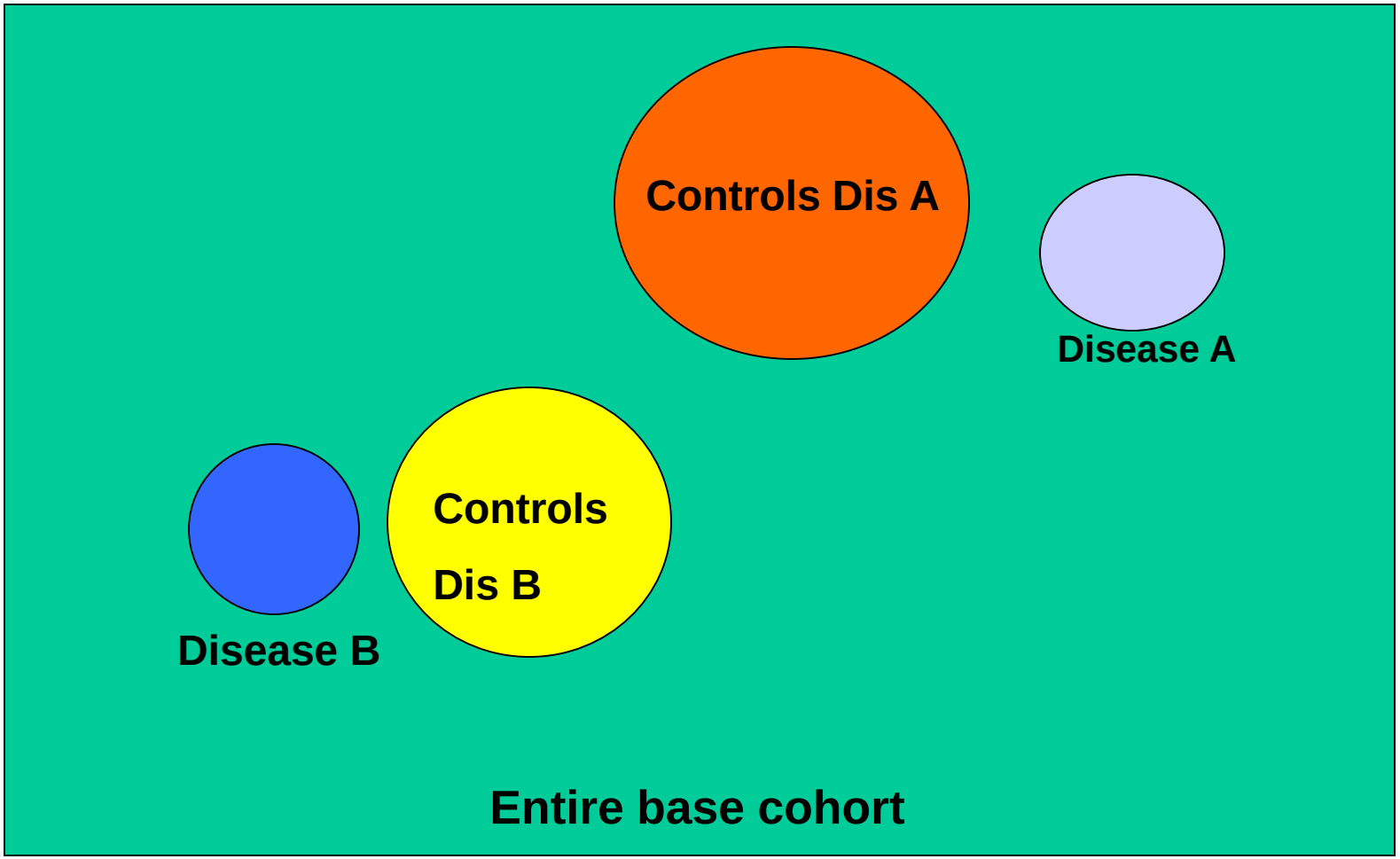
Risk for non-Hodgkin lymphoma and soft-tissue sarcoma. IARC cohort (nested case-control study). TCDD exposure matrix.



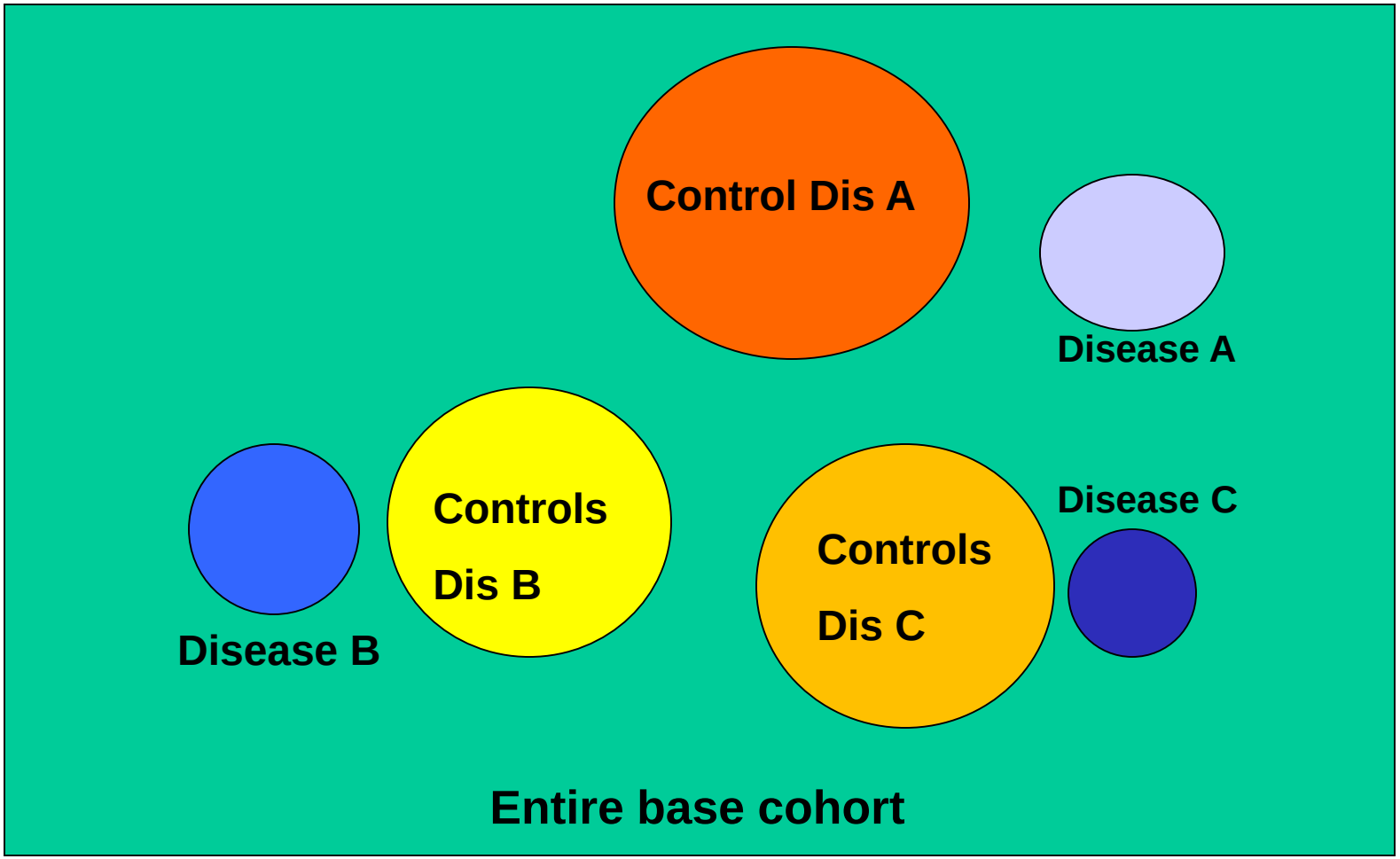
(Kogevinas et al 1995, Epidemiology)

ISGlobal Barcelona Institute for Global Health

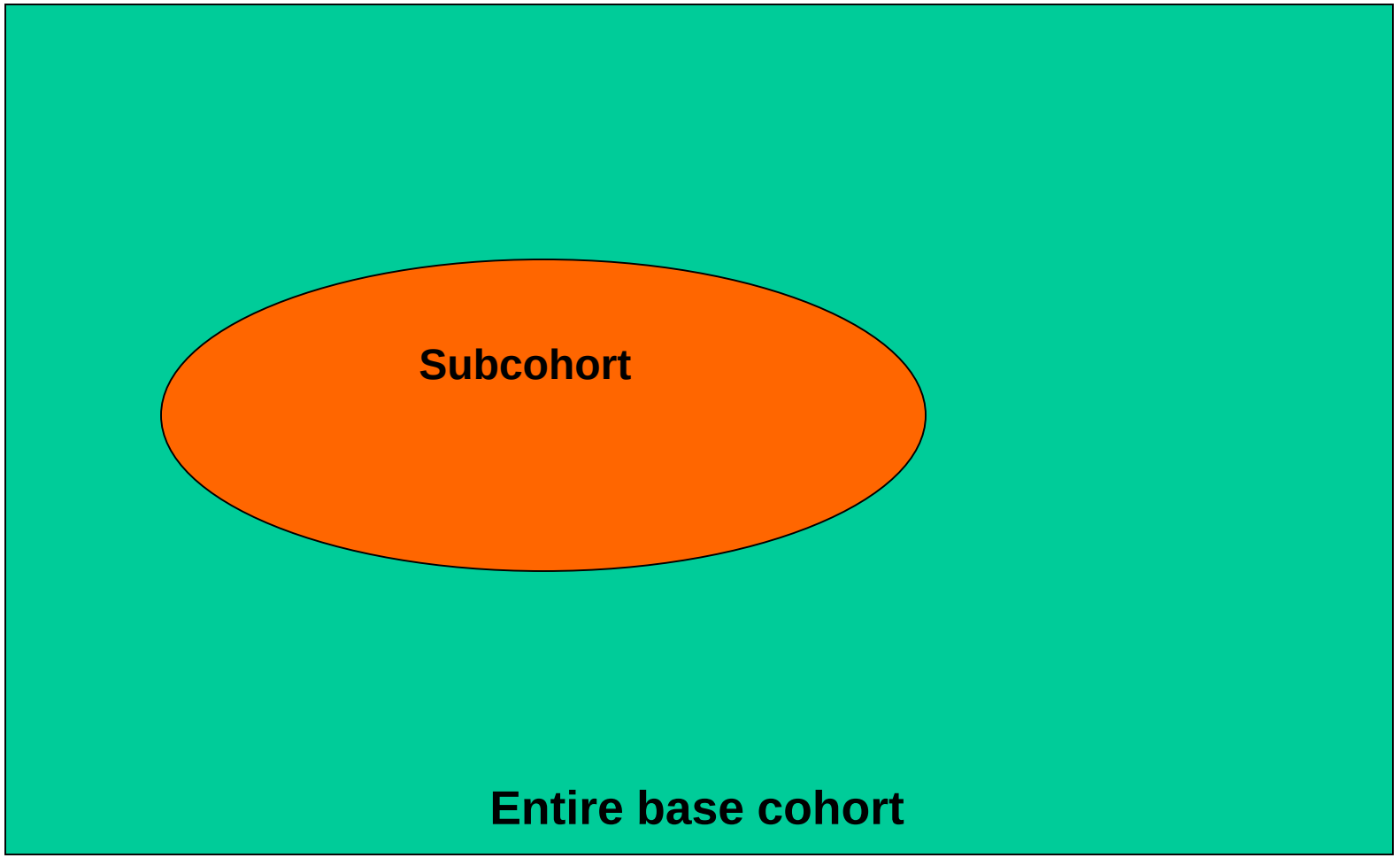
Selection of controls in nested case-control study



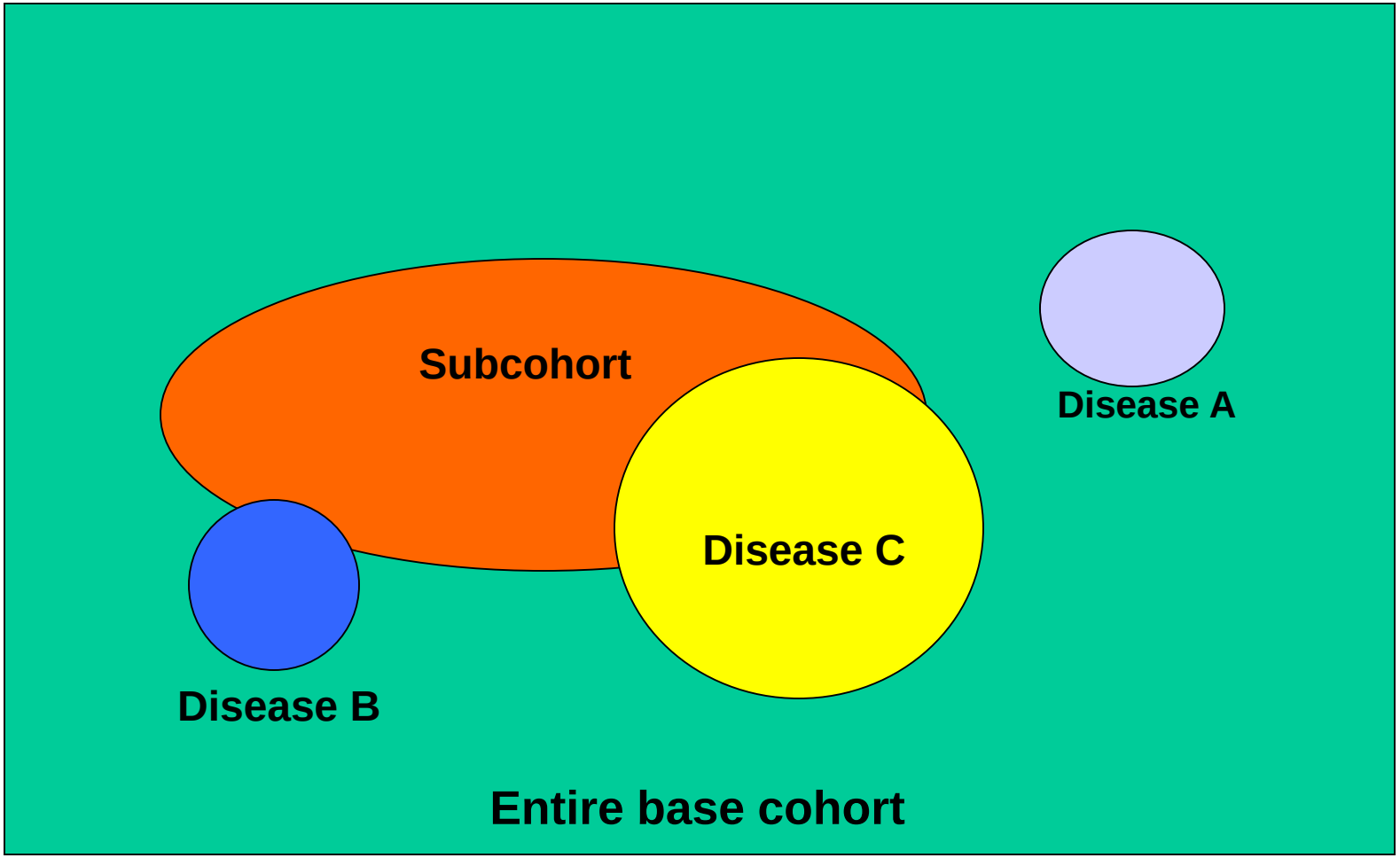
Selection of controls in nested case-control study



Selection of reference subcohort in case-cohort studies



Selection of reference subcohort in case-cohort studies

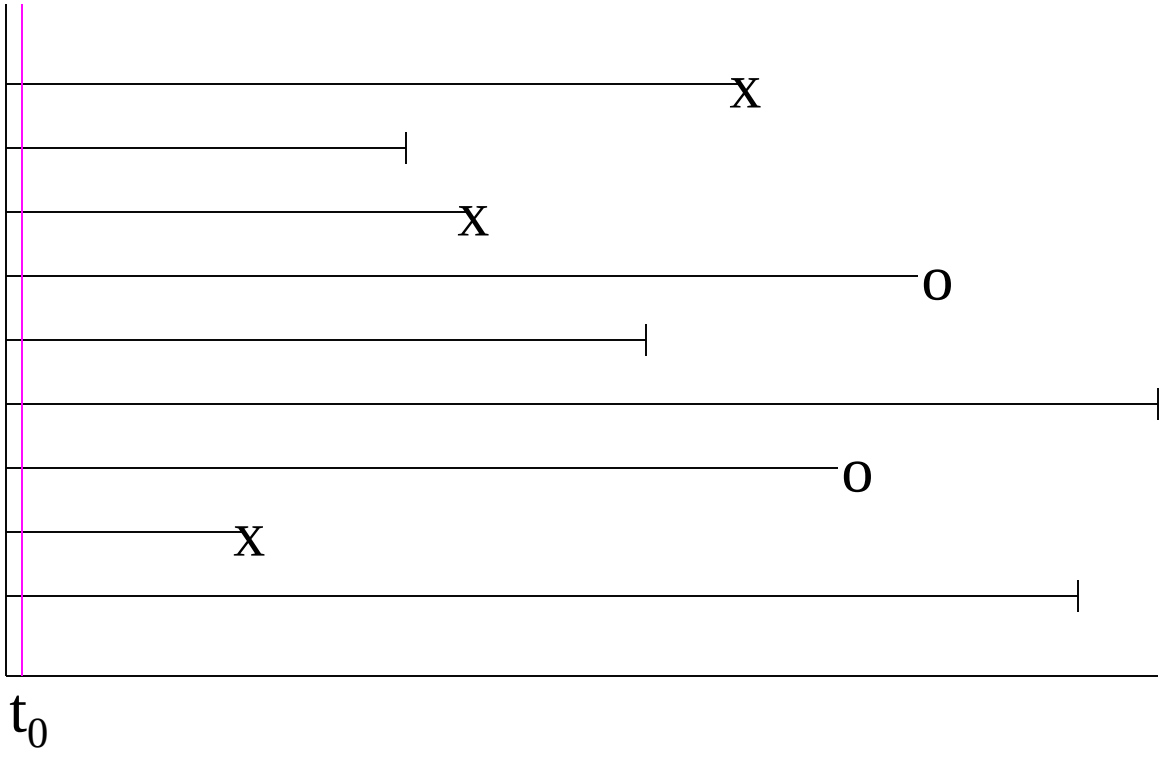


Slide from H Checkoway

Case-cohort design

- Select a single subcohort from the initial cohort at its inception with a pre-specified sampling fraction, either randomly or using stratified random sampling
- Add all cases (one or more outcomes) that occur in the cohort outside the subcohort

Selection of subcohort in a closed cohort

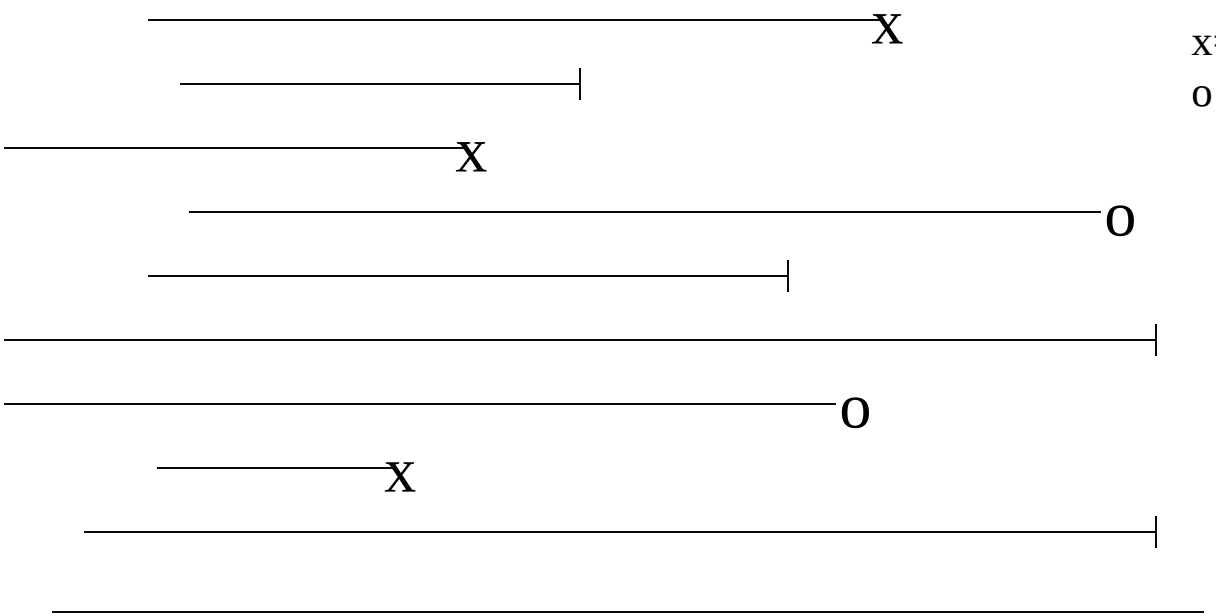


x= diseased
o = lost to follow-up

OMEGA-NET

ISGlobal Barcelona Institute for Global Health

Selection of subcohort in an open cohort



x= diseased
o = lost to follow-up

Selection of subcohort in closed or open cohorts

- In *closed* cohort (everybody enters cohort at t_0), a sample of all subjects ("sub-cohort") is randomly selected from cohort members at start of follow-up t_0
- In *open* cohort (time of entry into cohort is variable), a sample of all subjects ("sub-cohort") is randomly selected from members of cohort as it is followed over time (i.e., regardless of when subjects entered the cohort)

Case-cohort (inclusive) sampling

- Controls can be selected from those at risk at the *beginning* of the follow-up period, i.e. from the entire *source population*
- i.e. controls are selected from the denominators for (cohort) *risk ratio* analyses

A Hypothetical Incidence Study

	Exposed	Non-exposed	Total
Cases	1,813	952	2,765
Total	10,000	10,000	20,000

			Risk ratio
Rsik	1813/10000	952/10000	1.90

OMEGA-NET

ISGlobal Barcelona Institute for Global Health

A Hypothetical Case-Control Study (case-cohort design)

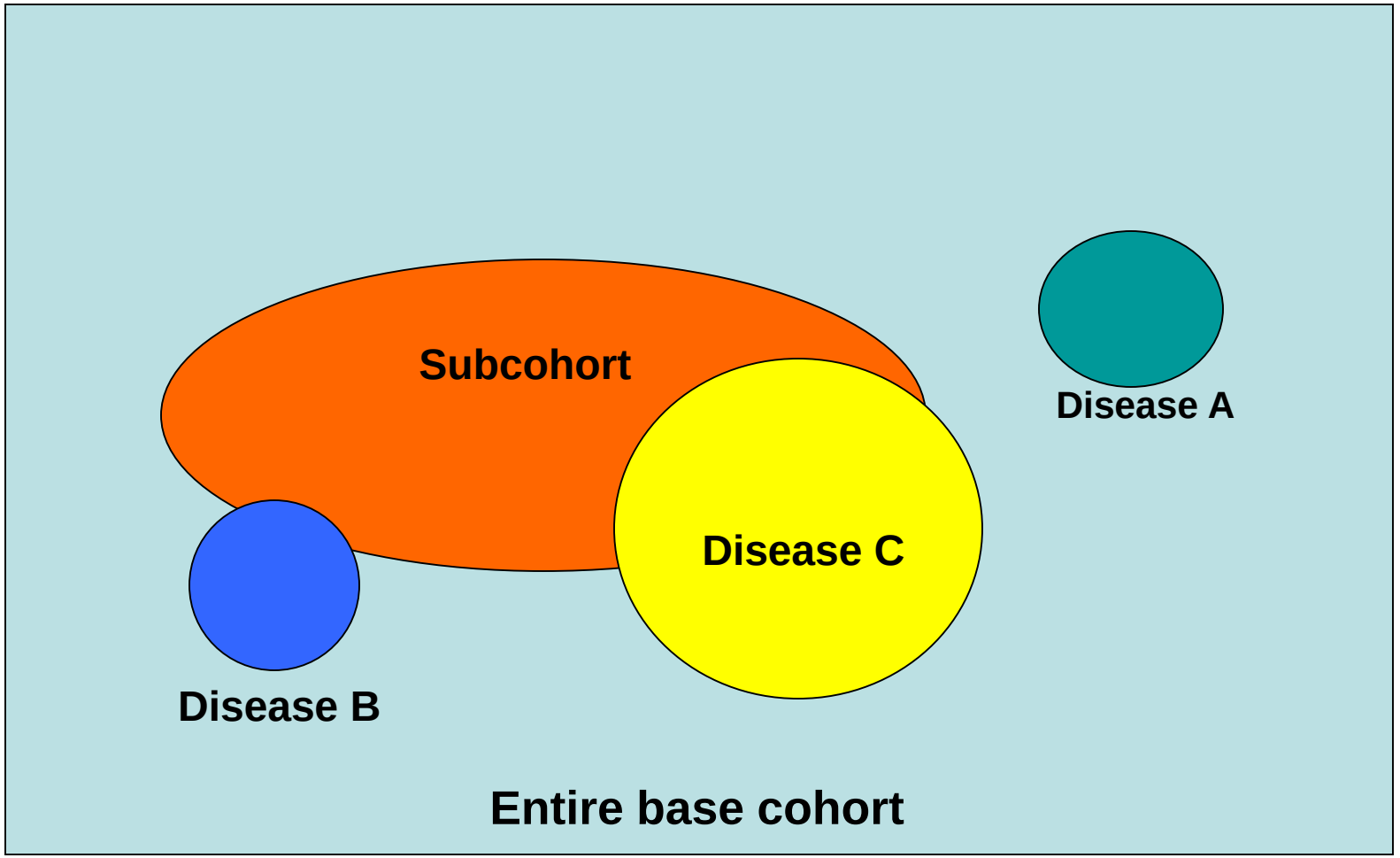
	Exposed	Non-exposed	Total
Cases	1,813	952	2,765
Controls	1,383	1,383	2,766

			Odds ratio
Odds	1813/1383	952/1383	1.90

OMEGA-NET

ISGlobal Barcelona Institute for Global Health

Selection of reference subcohort in case-cohort studies



Slide from H Checkoway

Comparisons in case-cohort analysis

- **Disease A vs. Subcohort (no overlap, no problem)**
- **Disease B vs. Subcohort (leave overlap cases in?—minor effect)**
- **Disease C vs. Subcohort (leave overlap cases in?—non-trivial effect possibly – can be adjusted in analysis)**

Analysis Case-Cohort Design

- The subcohort is a random or stratified random sample from the cohort and can be analyzed by itself as a cohort study using the ordinary Partial Likelihood approach (Cox Regression)

- This is inefficient because of the loss of cases that occur outside the subcohort

- Key feature of the case–cohort design is the inclusion of all cases that occur in the cohort regardless of whether they are in the selected subcohort

Notation for a Stratum in a Case-Cohort Study

	Exposed	Unexposed	Total
Case, not control	A_{11i}	A_{01i}	M_{11i}
Case and control	A_{10i}	A_{00i}	M_{10i}
Noncase control	B_{1i}	B_{0i}	M_{0i}
Total	N_{1i}	N_{0i}	N_{+i}

OMEGA-NET

from Rothman et al, 2008

Analysis Case-Cohort Design

Standard cohort analysis with some adjustments

Must adjust for bias introduced because the case–cohort sample is not population-based because of over-sampling of cases

The bias produced by including cases outside the subcohort is corrected by not allowing those cases to contribute to risk sets other than their own (risk sets comprise cohort subjects still under observation and at risk at the times cases occur).

Analysis includes:
Weighting: Due to oversampling of cases
Adjustment of variance: Because the same control population is upweighted and used repeatedly over time

Biomarkers Cause Logistical Problems in Case-Cohort Studies

The case-cohort study relies on the assumption that exposure can be equally well measured in the sub-cohort as in the cases, and subsequent case series.

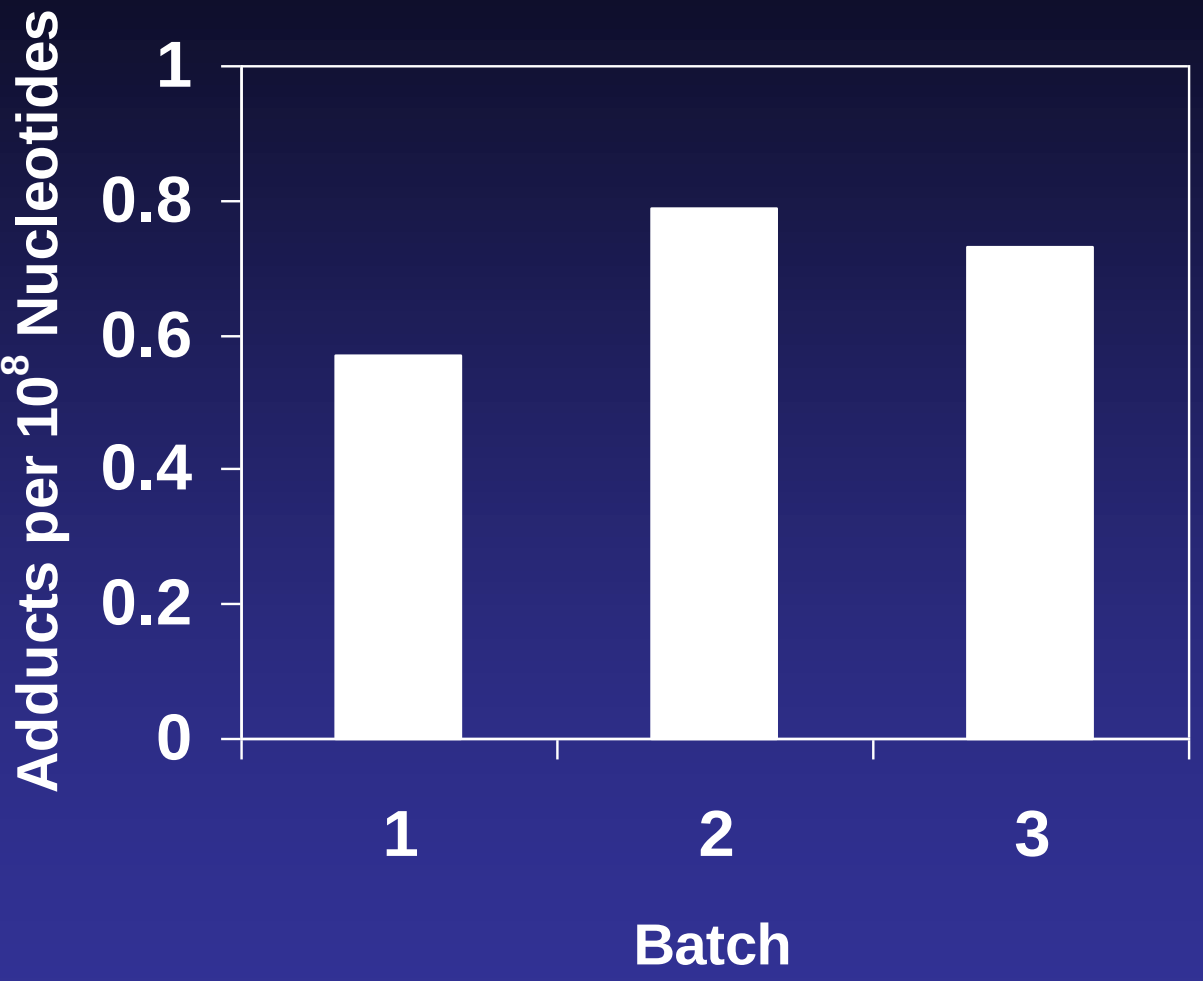
Three issues with biomarkers make this assumption questionable.

- Batch effects
- Storage effects
- Freeze-thaw cycles

OMEGA-NET

(Slides on biomarkers from P Vineis)

Batch Effects In EPIC/GEN-AIR



P < 0.01 for difference between batch 1 and 3 after control for gender, smoking (never vs. ex-smoker), country and EPIC center.

Storage Effects

Biological samples are typically stored at -70^{o}C or lower. However, not all biomarker targets are stable at this temperature.

- **Evidence that antioxidant micronutrients in serum, cotinine and B[a]P-DNA adducts are stable**
- **Evidence that serum cholesterol, free PSA, serum sex hormones, salivary Ab, and IH targets in tissue sections are not stable.**

Freeze-Thaw Cycles

As biological samples freeze and thaw the pH and ionic balance of the liquid phase of the sample can be very different from the natural condition of the sample. Changes in pH and ionic balance can degrade biomarker targets.

- **There is evidence that lipoprotein (a), antibodies, endogenous antioxidants, saliva cortisol, EGFR and DNA quality degrade during freeze-thaw cycles.**

Summary biomarkers case-cohort design

For biomarkers not affected by batch, storage, and freeze-thaw cycles (e.g. genotypes) use the case-cohort design.

- **The sub-cohort can be used as a referent for multiple case-series**
- **Simple random sample allows for valid cross-sectional analyses and external comparisons**

The case-cohort design best leverages the investment in biomarker analyses

Summary, case-cohort design

- In a case-cohort study, cases are defined as those participants of the cohort who developed the disease of interest
- Controls are identified before the cases develop and are randomly chosen from all cohort participants regardless of whether they have the disease of interest or not
- Baseline data are collected early in the study

OMEGA-NET

ISGlobal Barcelona Institute for Global Health

Case-cohort design: advantages

- Only need to select one reference group
- Evaluate multiple outcomes
- Like nested case-control study, minimizes data collection requirements
- Can estimate absolute risks

Summary, differences between case-control and case-cohort design

- Case-cohort studies very similar to nested case-control studies
- Main difference is the way in which controls are chosen
  Main advantage of case-cohort design over nested case-control design is that the same control group can be used for comparison with different case groups in a case-cohort study
- Main disadvantages of the case-cohort design is that it requires a more complicated statistical analysis and it can be less efficient than a nested case-control study under some circumstances

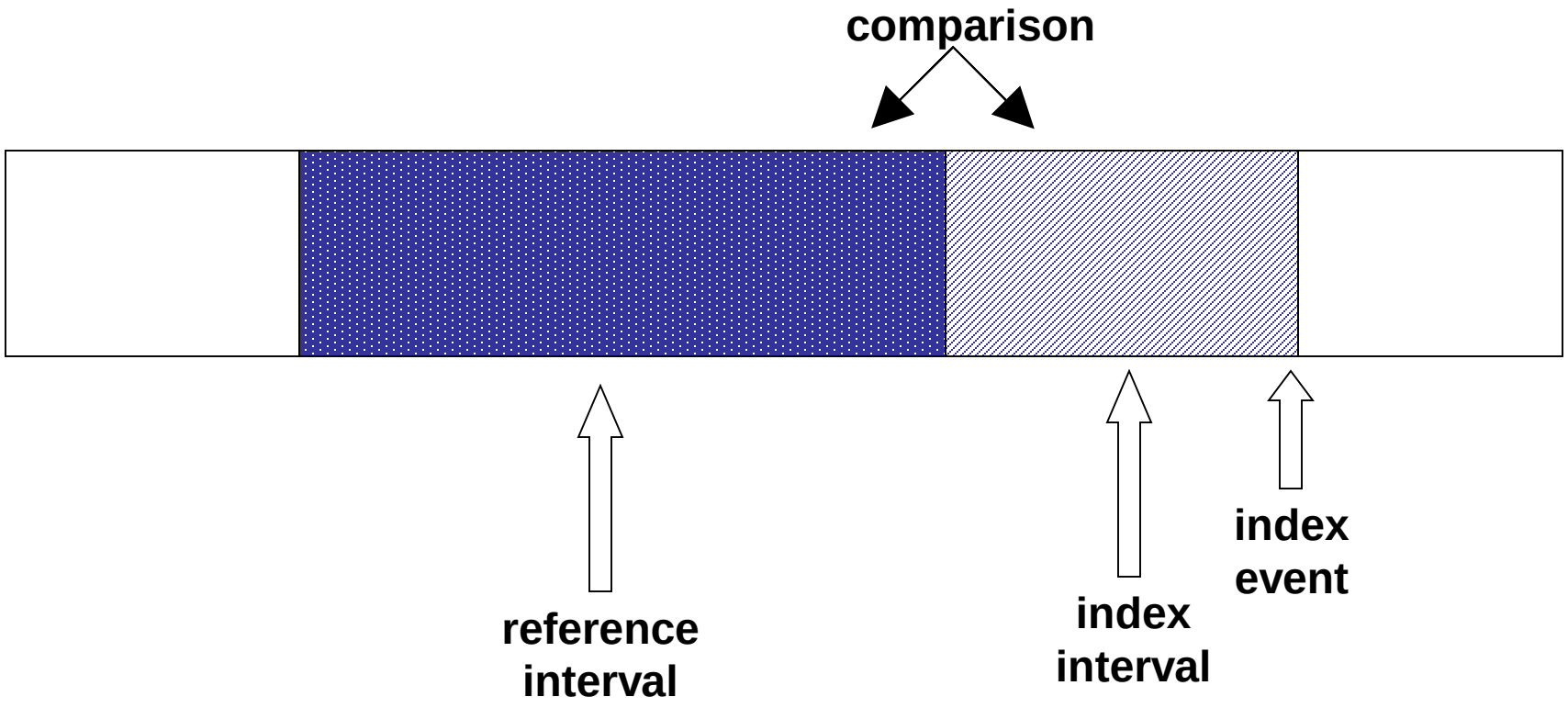
Case-Crossover studies

Case-crossover design

- "Case only" study, i.e., cases serve as their own controls (special type of matched case-control study)
- Comparisons of exposures during cases' "index" intervals with exposures during "reference" intervals

- The principal advantage of the case-crossover design, relative to a conventional case-control study, is that matching each case with himself or herself greatly facilitates control of potential confounders that are time invariant and possibly difficult to measure, such as genetic factors.

Case-crossover design: unidirectional reference interval



OMEGA-NET

ISGlobal Barcelona Institute for Global Health

Index and reference intervals in case-crossover design

- Index interval: time period preceding disease (or injury) onset when exposures may have etiologic relevance

- Reference interval: time interval of typical exposure, usually preceding disease onset

  - Unidirectional (before index interval)

  - *Bidirectional (before and after index interval)*

Reference interval matching in case-crossover studies

- A single day, hour or other time interval (1:1 matching)
- Multiple days, hours, or other time intervals within the reference interval (K:1 matching)

OMEGA-NET

ISGlobal Barcelona Institute for Global Health

Examples of case-crossover studies

- Physical exercise and AMI
- Drug intake and skin rash
- Intake of red wine and migraine
- Use of mobile phones and car accidents
- Use of protection measures and occupational injuries

OMEGA-NET

ISGlobal Barcelona Institute for Global Health

Example: study of risk and protective factors among for acute hand injuries

- Cases provided details on the extent and timing of transient work factors during the 90 min preceding their injuries
- Transient factors studied:
- using a machine, tool, or work material that performed differently than usual e.g. a jammed machine, malfunctioning hand tool, a recently sharpened knife-
  - wearing gloves;
  - performing an unusual task;
  - doing a task using an unusual work method;
  - being distracted
  - being rushed;
  - feeling ill.

(Sorock et al, OEM 2004)

Example: study of risk and protective factors among for acute hand injuries

- Cases provided details on the extent and timing of transient work factors during the 90 min preceding their injuries
- Classified as exposed if they experienced these factors at the time of the injury.
- Reference period exposures were estimated as averages for the month preceding the injury

OMEGA-NET

(Sorock et al, OEM 2004)

Case-cross over study. Relative risks for hand injury associated with transient workplace conditions

Worker group	Unusual work task or equipment	Glove use
All subjects	11.0 (9.4 to 12.8)*	0.4 (0.3 to 0.5)
Occupational group		
Machine/assembly	10.6 (18.5 to 13.3)	0.5 (0.4 to 0.6)
Construction trades	17.4 (9.3 to 32.3)	0.3 (0.2 to 0.4)
Packaging	5.1 (3.8 to 6.9)	0.4 (0.3 to 0.7)
Service, prof, mgmt	15.7 (11.1 to 22.3)	0.4 (0.3 to 0.6)
Job experience (years)		
⩽1	13.2 (9.1 to 19.2)	0.4 (0.3 to 0.6)
1 to 3	16.3 (11.4 to 23.3)	0.2 (0.2 to 0.4)
>3	9.0 (7.4 to 11.0)	0.5 (0.4 to 0.6)

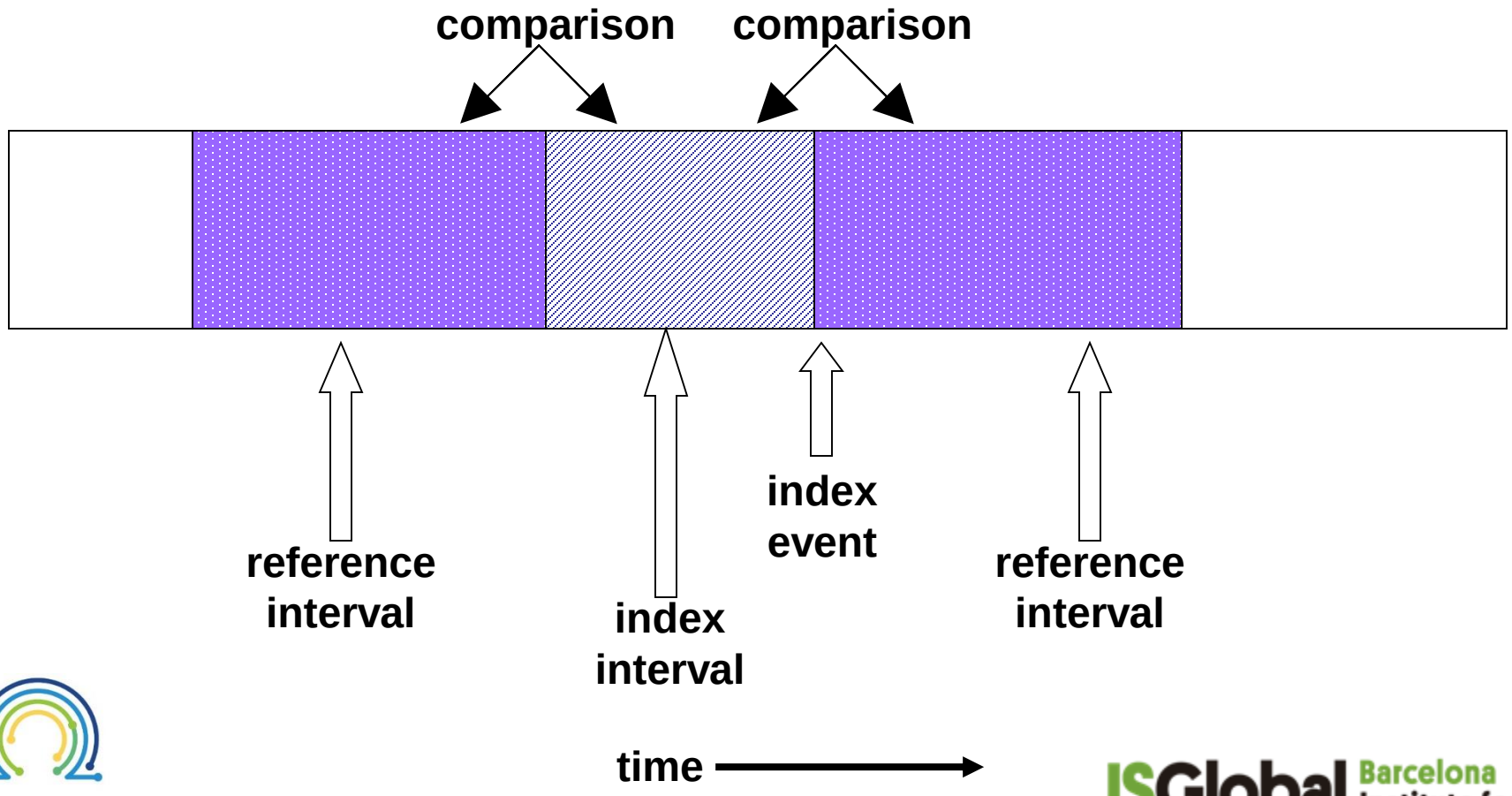
*Relative risk (95% CI) exposed versus non-exposed.

(Sorock et al, OEM 2004; table form Checkoway et al OEM 2007)

Case-crossover: selection of index and reference intervals

- The selection of index and reference intervals is not necessarily clear cut
- Width of the index interval will depend on the characteristics of the exposure and the health outcome and the nature of their presumed relation
- Simplest case of a very acute severe injury, the index interval can be as short as several minutes or hours
- More complex: effects of sensitising chemicals may appear hours or days after relevant exposures.
- Placement and width of reference intervals can be sources of uncertainty.

Case-crossover design: bidirectional reference interval



OMEGA-NET

ISGlobal Barcelona Institute for Global Health

Directionality of referent interval in case-crossover design

- Bi-directional preferred if known temporal trend of exposure (e.g., reductions over time)

- BUT, bi-directionality requires that health outcome unrelated to subsequent exposure
  - Can be violated if occurrence of illness or injury alters workplace safety policy (e.g., exposures reduced)

OMEGA-NET

ISGlobal Barcelona Institute for Global Health

Case types and information

Type	Ref. w	Causal w	Information
1	-	-	None
2	-	+	For
3	+	-	Against
4	+	+	None

$$OR = \frac{\sum \text{type 2}}{\sum \text{type 3}}$$

OMEGA-NET

ISGlobal Barcelona Institute for Global Health

Advantages of case-crossover design (relative to conventional case-control design)

- Suitable for studies of acute onset outcomes
- Logical/convenient choice of controls

  (e.g., Who would controls be in a study of acute injuries? How would they be identified)
- Better control of confounding by fixed variables (e.g., medical history, genetics)

OMEGA-NET

ISGlobal Barcelona Institute for Global Health

References

- Prentice RL. A case-cohort design for epidemiologic cohort studies and disease prevention trials. Biometrika, 1986; 73:1-11. 1986.
- Cologne et al. Conventional case–cohort design and analysis for studies of interaction. International Journal of Epidemiology 2012;1–13
- Maclure M. The case-crossover design: a method for studying transient effects on the risk of acute events.Am J Epidemiol. 1991 Jan 15;133(2):144-53.
- Checkoway H, Pearce N, Kriebel D. Selecting appropriate study designs to address specific research questions in occupational epidemiology. Occup Environ Med. 2007 Sep;64(9):633-8.

OMEGA-NET

ISGlobal Barcelona Institute for Global Health