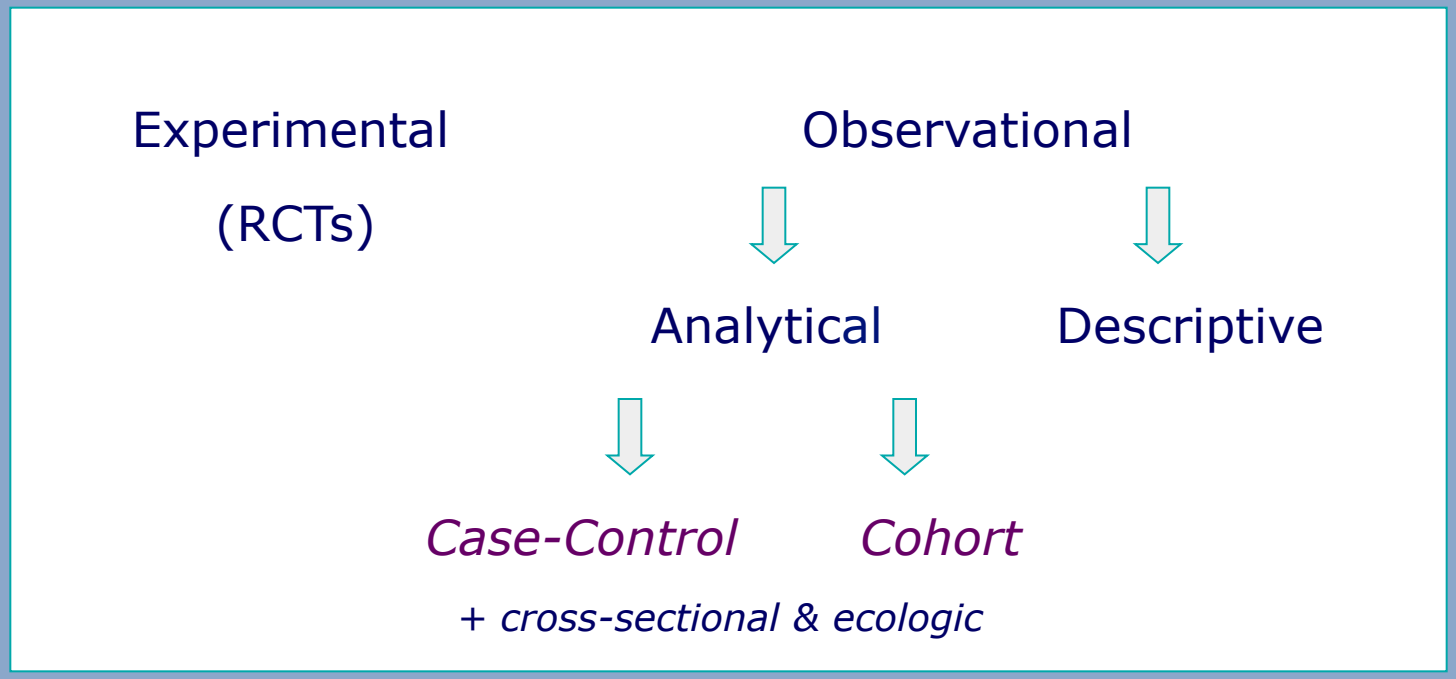
RESEARCH DESIGN

What Is Research Design?

The *structure* of research

Epidemiologic Study Designs



Epidemiologic Study Designs

Descriptive studies

Examine patterns of disease

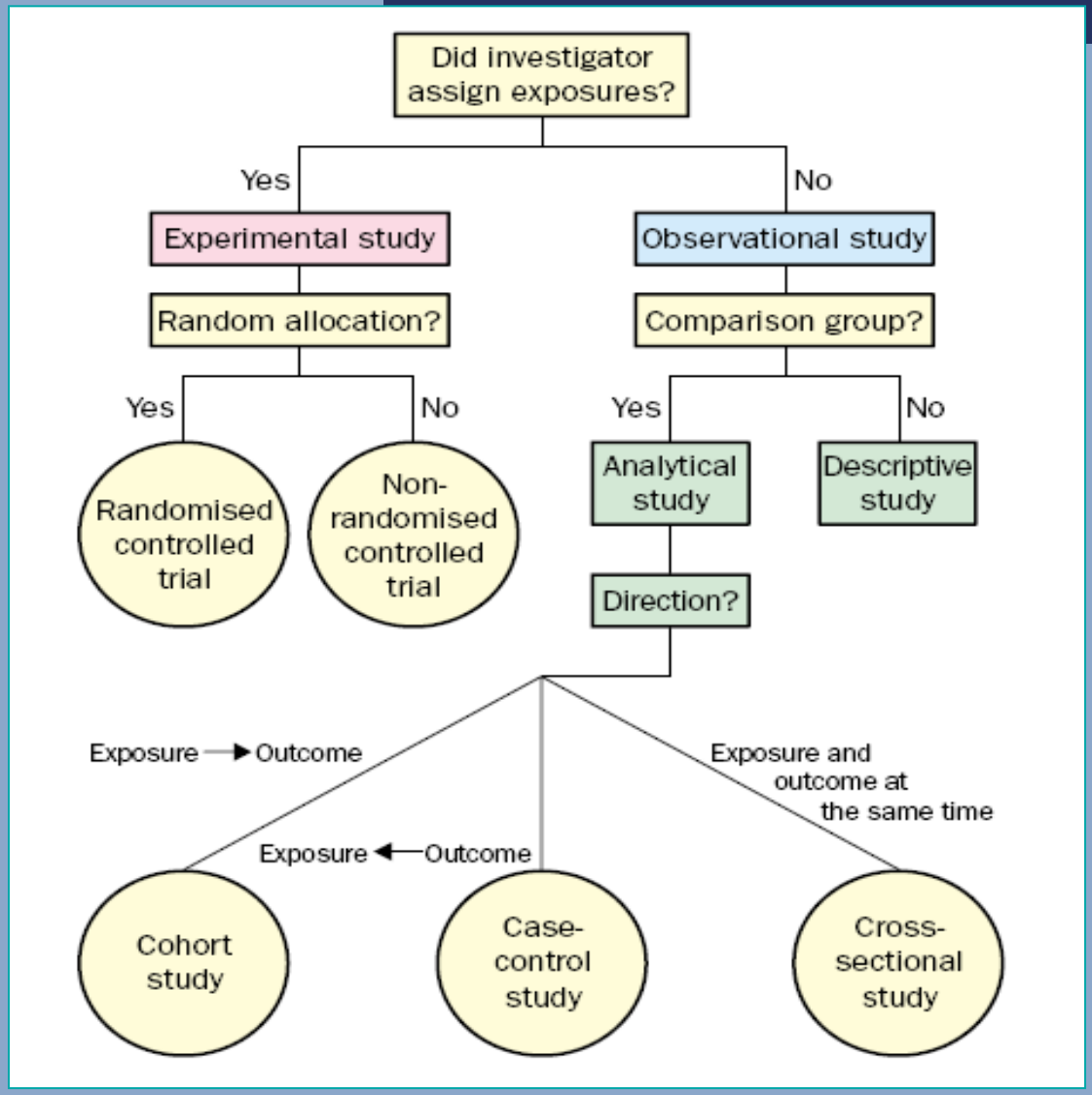
Analytical studies

Studies of suspected causes of diseases

Experimental studies

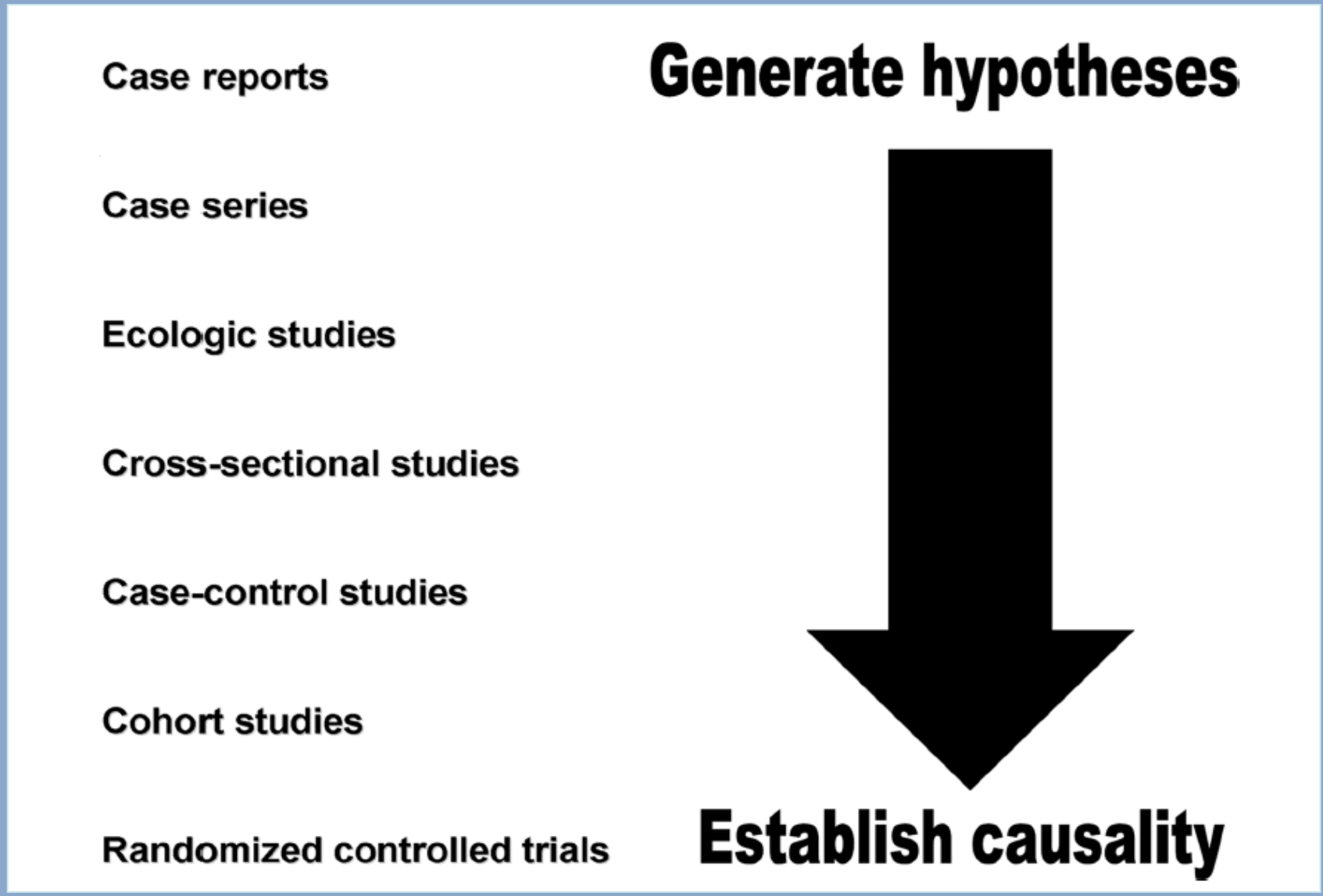
Compare treatment modalities

Epidemiologic Study Designs



Grimes & Schulz, 2002 (www)

Hierarchy of Epidemiologic Study Design



Tower & Spector, 2007 (www)

Observational Studies

(no control over the circumstances)

- Descriptive: Most basic demographic studies
- Analytical: Comparative studies testing an hypothesis
    * cross-sectional
      (a snapshot; no idea on cause-and-effect relationship)
    * cohort
      (prospective; cause-and-effect relationship can be inferred)
    * case-control
      (retrospective; cause-and-effect relationship can be inferred)

Epidemiologic Study Designs

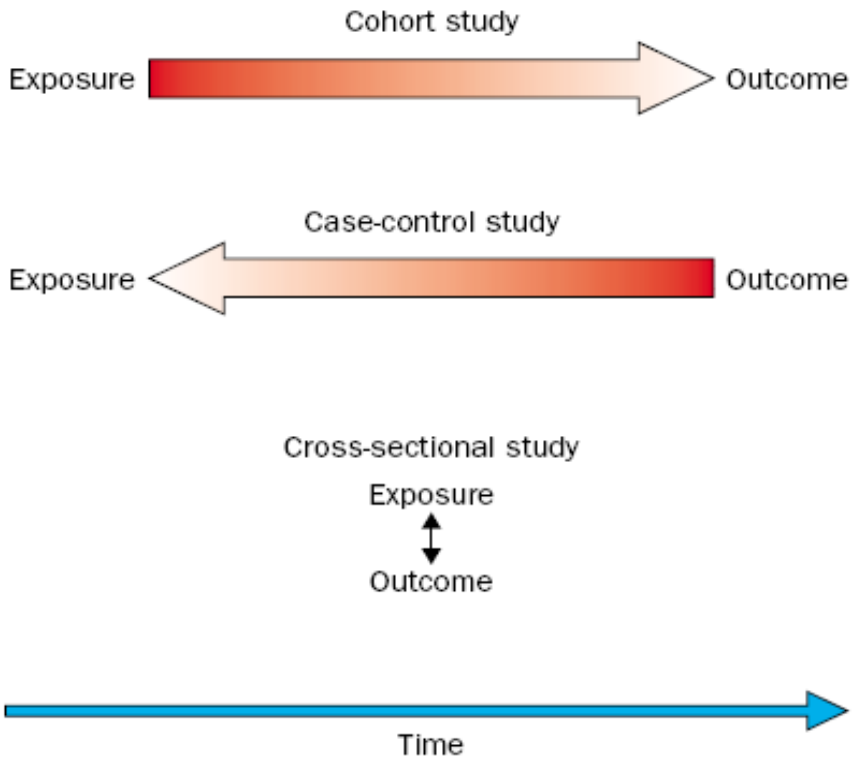


Figure 2: **Schematic diagram showing temporal direction of three study designs**

Grimes & Schulz, 2002 (www)

Analytical Studies

(comparative studies testing an hypothesis)

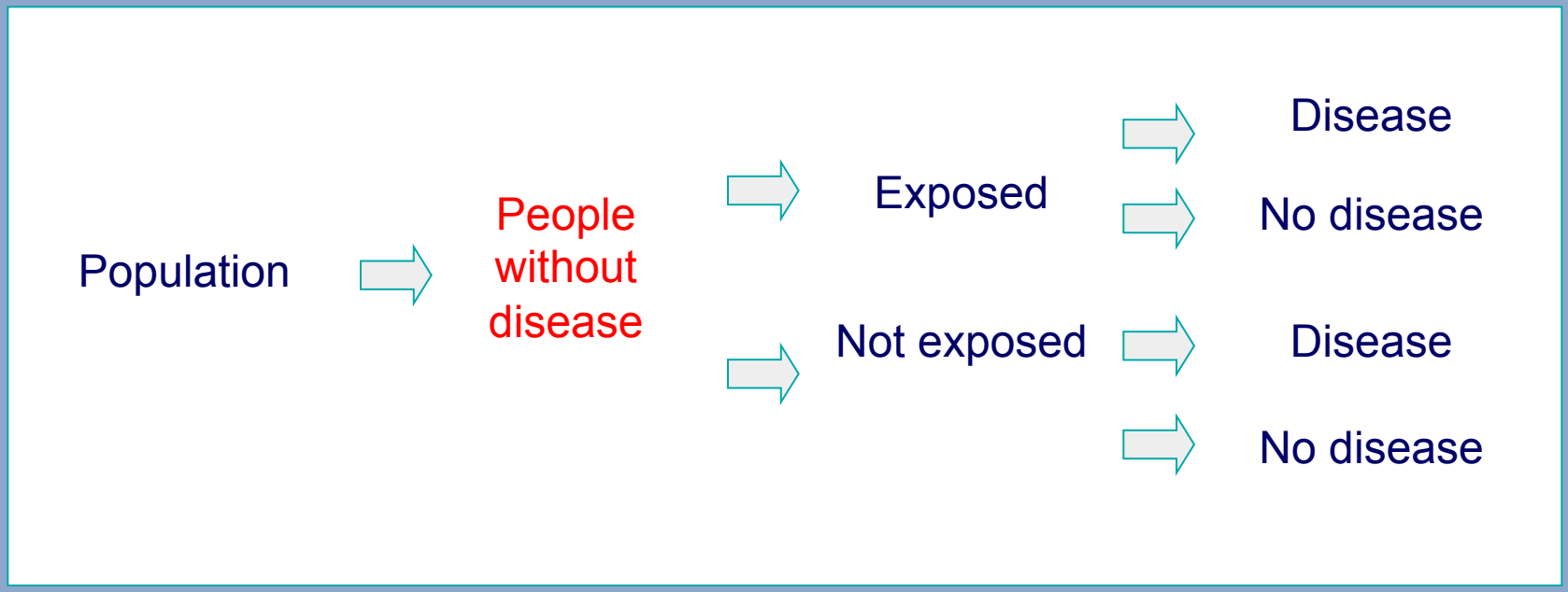
* cohort (prospective)

Begins with an exposure (smokers and non-smokers)

* case-control (retrospective - trohoc)

Begins with outcome (cancer cases and healthy controls)

Cohort Studies



Examples of Cohort Studies

* *Framingham Heart Study* (www)
* *NHANES Studies* (www)
* *MACS* (www)
* *Physicians' Health Study* (www)
* *Nurses' Health Study* (www)
* *ALSPAC* (www)

Advantages of Cohort Studies

- Can establish population-based incidence
- Accurate relative risk (risk ratio) estimation
- Can examine rare exposures (asbestos > lung cancer)
- Temporal relationship can be inferred (prospective design)
- Time-to-event analysis is possible
- Can be used where randomization is not possible
- Magnitude of a risk factor's effect can be quantified
- Selection and information biases are decreased
- Multiple outcomes can be studied
  (smoking > lung cancer, COPD, larynx cancer)

Disadvantages of Cohort Studies

- Lengthy and expensive
- May require very large samples
- Not suitable for rare diseases
- Not suitable for diseases with long-latency
- Unexpected environmental changes may influence the association
- Nonresponse, migration and loss-to-follow-up biases
- Sampling, ascertainment and observer biases are still possible

Presentation of cohort data: Population at risk

Does HIV infection increase risk of developing TB among a population of drug users?

	Population (follow up 2 years)	Cases
HIV +	215	8
HIV -	289	1

Source: Selwyn et al., New York, 1989

Does HIV infection increase risk of developing TB among drug users?

Exposure	Population (f/u 2 years)	Cases	Incidence (%)	Relative Risk
HIV +	215	8	3.7	11
HIV -	298	1	0.3	

Presentation of cohort data: Person-years at risk

Tobacco smoking and lung cancer, England & Wales, 1951

	Person-years	Cases
Smoke	102,600	133
Do not smoke	42,800	3

Source: Doll & Hill

Presentation of data: Various exposure levels

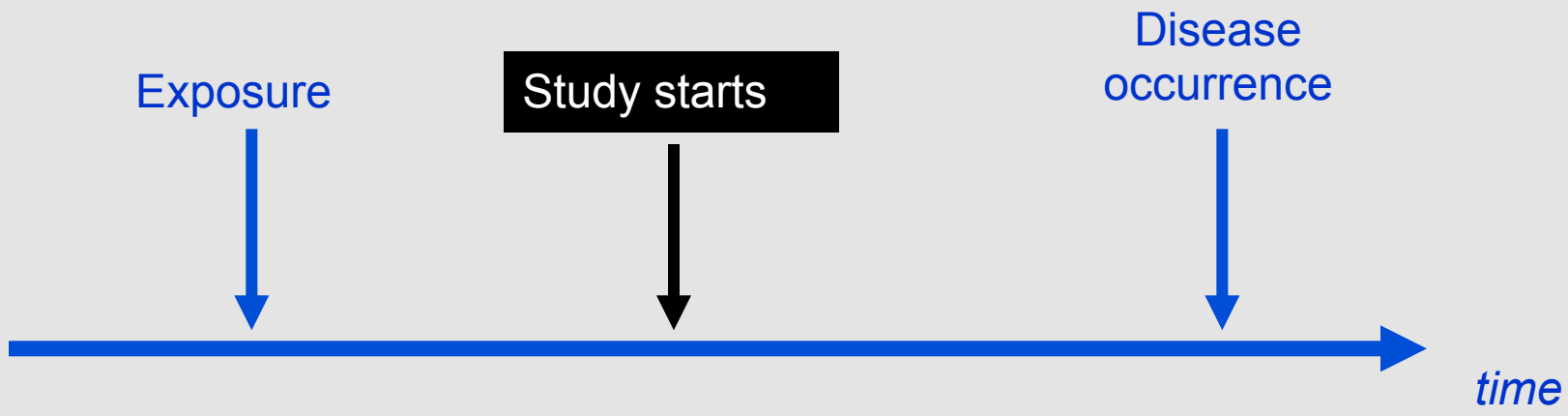
Daily number of cigarettes smoked	Person-years at risk	Lung cancer cases
> 25	25,100	57
15 - 24	38,900	54
1 - 14	38,600	22
none	42,800	3

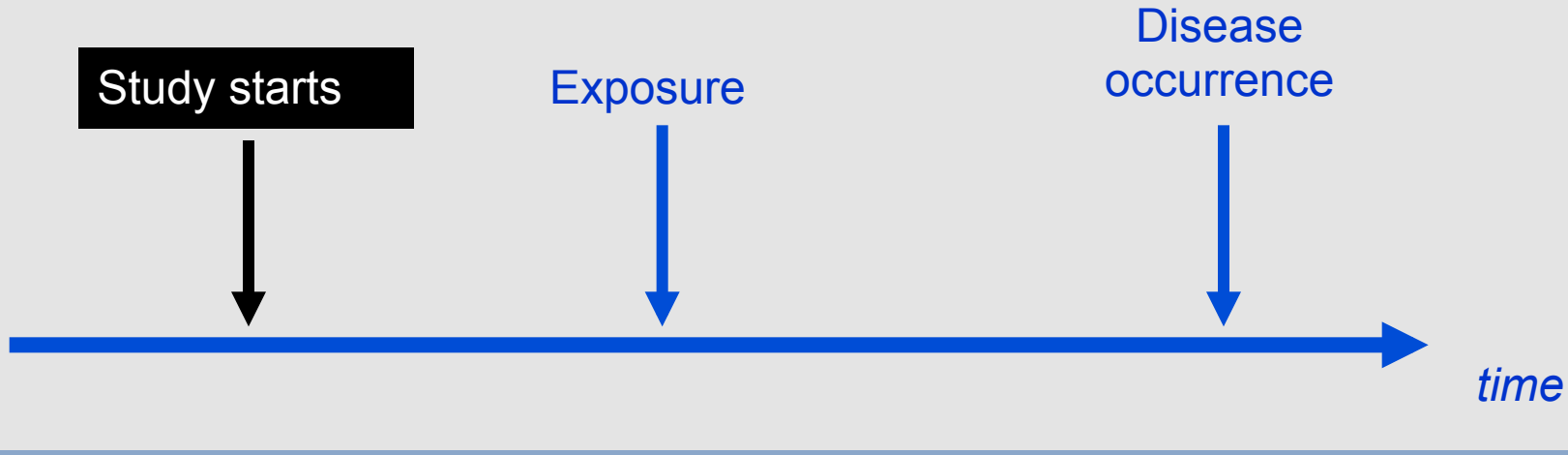
Cohort study: Tobacco smoking and lung cancer, England & Wales, 1951

Cigarettes smoked/d	Person-years at risk	Cases	Rate per 1000 p-y	Rate ratio
> 25	25,100	57	2.27	32.4
15 - 24	38,900	54	1.39	19.8
1 - 14	38,600	22	0.57	8.1
none	42,800	3	0.07	Ref.

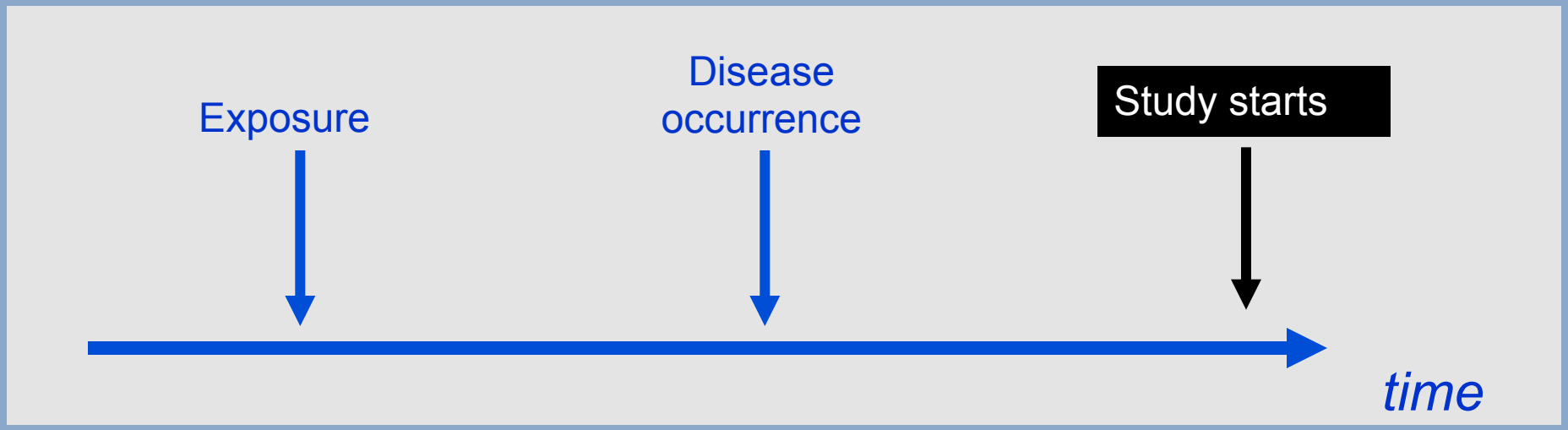
Source: Doll & Hill

Prospective cohort study





Retrospective cohort studies



Cohort Studies

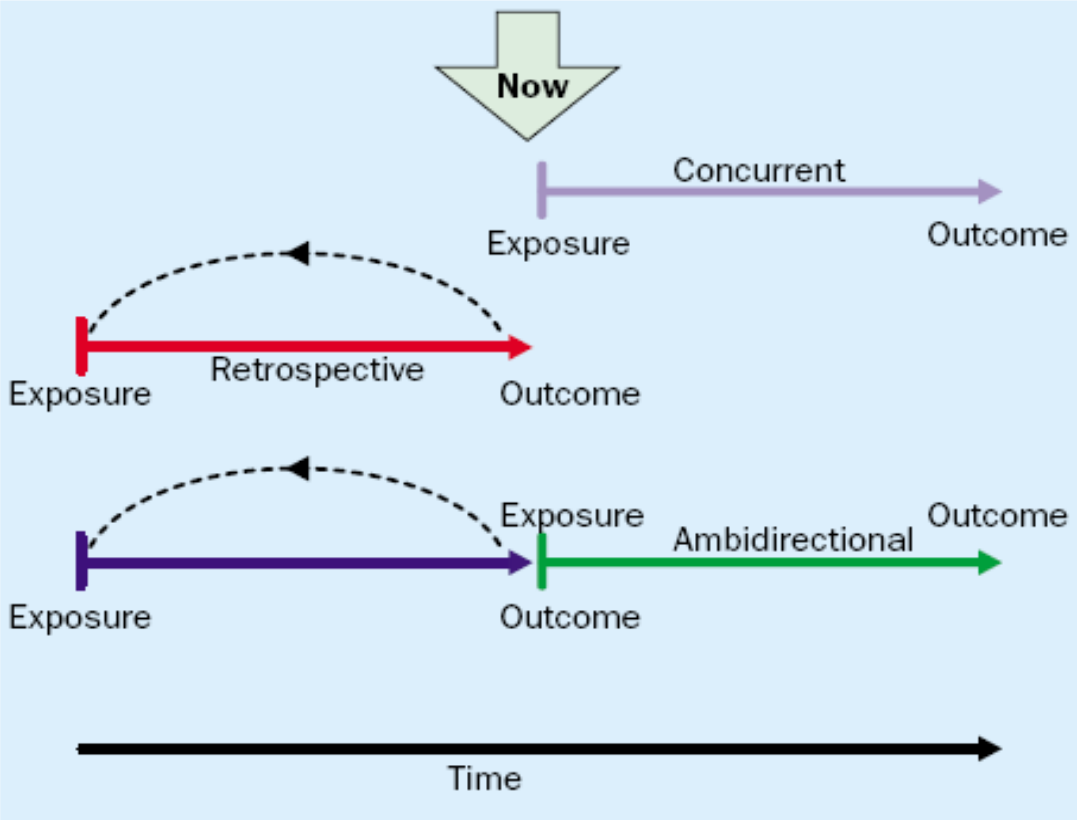


Figure 2: **Schematic diagram of concurrent, retrospective, and ambidirectional cohort studies**

Cohort Studies

Panel 2: **Features to look for in a cohort study**

How much selection bias was present?

1 Were only people at risk of the outcome included?
1 Was the exposure clear, specific, and measurable?
1 Were the exposed and unexposed groups similar in all important respects except for the exposure?

What steps were taken to minimise information bias?

1 Was the outcome clear, specific, and measurable?
1 Was the outcome identified in the same way for both groups?
1 Was determination of outcome made by an observer blinded as to treatment?

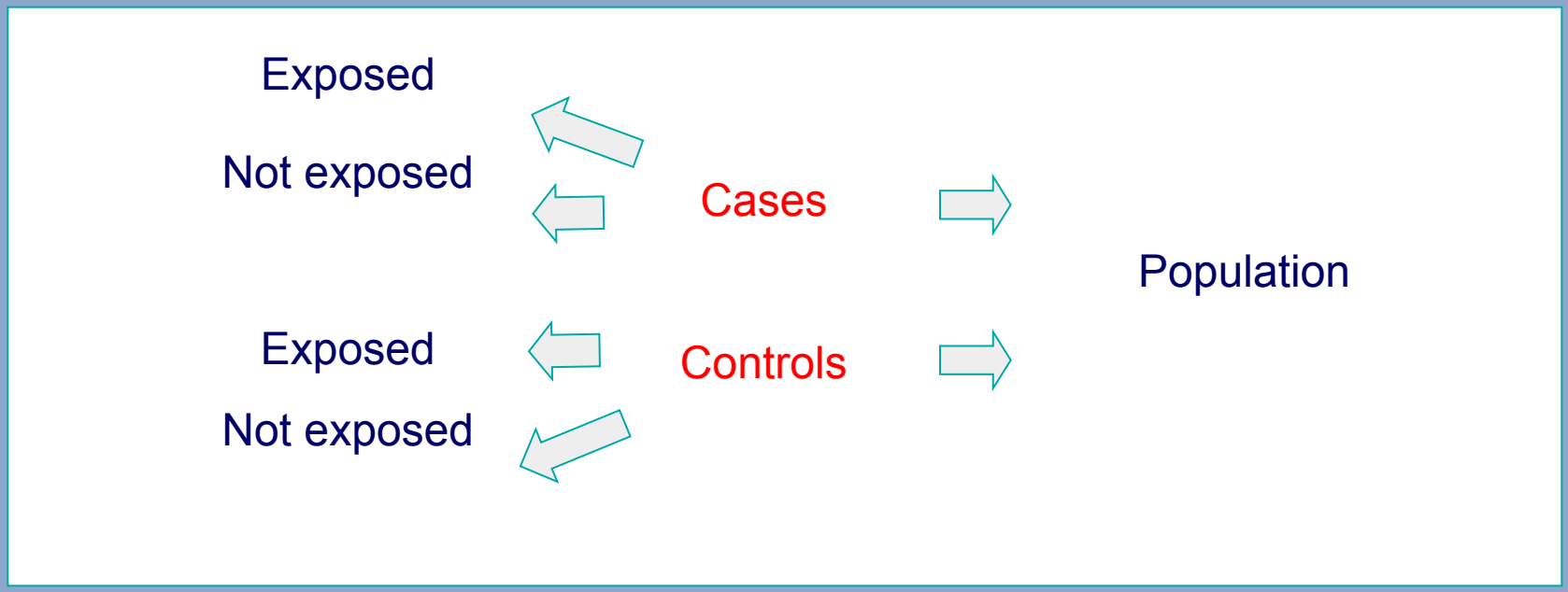
How complete was the follow-up of both groups?

1 What efforts were made to limit loss to follow-up?
1 Was loss to follow-up similar in both groups?

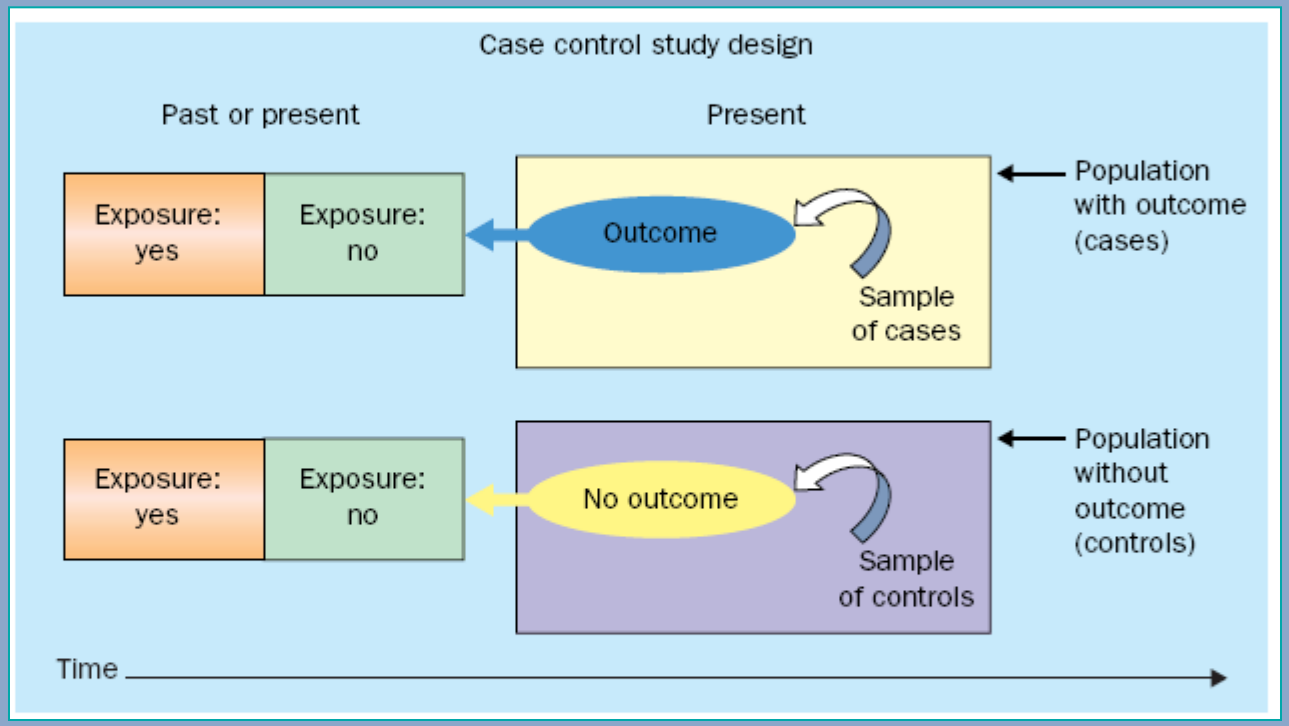
Were potential confounding factors sought and controlled for in the analysis?

1 Did the investigators anticipate and gather information on potential confounding factors?
1 What method(s) were used to assess and control for confounding?

Case-Control Studies



Case-Control Studies



Schulz & Grimes, 2002 (www) (PDF)

Advantages of Case-Control Studies

- Cheap, easy and quick studies
- Multiple exposures can be examined
- Rare diseases and diseases with long latency can be studied
- Suitable when randomization is unethical
  (alcohol and pregnancy outcome)

Disadvantages of Case-Control Studies

- Case and control selection troublesome
- Subject to bias (selection, recall, misclassification)
- Direct incidence estimation is not possible
- Temporal relationship is not clear
- Multiple outcomes cannot be studied
- If the incidence of exposure is high, it is difficult to show the difference between cases and controls
- Not easy to estimate attributable fraction
- Reverse causation is a problem in interpretation - especially in molecular epidemiology studies

Case-Control Studies: Potential Bias

Panel 2: **Introduction of bias through poor choice of controls**

Cases	Control selection	Non-representativeness	Selection bias
Colorectal cancer patients admitted to hospital	Patients admitted to hospital with arthritis	Controls probably have high degrees of exposure to NSAIDs	Would spuriously **reduce** the estimate of effect (odds ratio)
Colorectal cancer patients admitted to hospital	Patients admitted to hospital with peptic ulcers	Controls probably have low degrees of exposure to NSAIDs	Would spuriously **increase** the estimate of effect (odds ratio)

NSAIDs=non-steroidal anti-inflammatory drugs.

Schulz & Grimes, 2002 (www) (PDF)

Cause-and-Effect Relationship

Temporal sequence
Did exposure precede outcome?

Strength of association
How strong is the effect, measured as relative risk or odds ratio?

Consistency of association
Has effect been seen by others?

Biological gradient (dose-response relation)
Does increased exposure result in more of the outcome?

Specificity of association
Does exposure lead only to outcome?

Biological plausibility
Does the association make sense?

Coherence with existing knowledge
Is the association consistent with available evidence?

Experimental evidence
Has a randomised controlled trial been done?

Analogy
Is the association similar to others?

Grimes & Schulz, 2002 (www) (PDF)

Cause-and-Effect Relationship

Panel 1: **What to look for in observational studies**

Is selection bias present?
In a cohort study, are participants in the exposed and unexposed groups similar in all important respects except for the exposure?

In a case-control study, are cases and controls similar in all important respects except for the disease in question?

Is information bias present?
In a cohort study, is information about outcome obtained in the same way for those exposed and unexposed?

In a case-control study, is information about exposure gathered in the same way for cases and controls?

Is confounding present?
Could the results be accounted for by the presence of a factor—eg, age, smoking, sexual behaviour, diet—associated with both the exposure and the outcome but not directly involved in the causal pathway?

If the results cannot be explained by these three biases, could they be the result of chance?
What are the relative risk or odds ratio and 95% CI?[11,12]

Is the difference statistically significant, and, if not, did the study have adequate power to find a clinically important difference?[13,14]

If the results still cannot be explained away, then (and only then) might the findings be real and worthy of note.

Grimes & Schulz, 2002 (www) (PDF)

Elements of a Design

- Observations or measures
- Treatments or programs
- Groups
- Assignment to group
- Time

Learning Objectives

- The major descriptors of research design
- The major types of research designs
- The relationships that exist between variables in causal designs and the steps for evaluating those relationships

Research Design

- Blueprint
- Plan
- Guide
- Framework

The Degree of Structure

Exploratory Study

- Loose structure
- Expand understanding
- Provide insight
- Develop hypotheses

Formal Study

- Precise procedures
- Begins with hypotheses
- Answers research questions

The Topical Scope

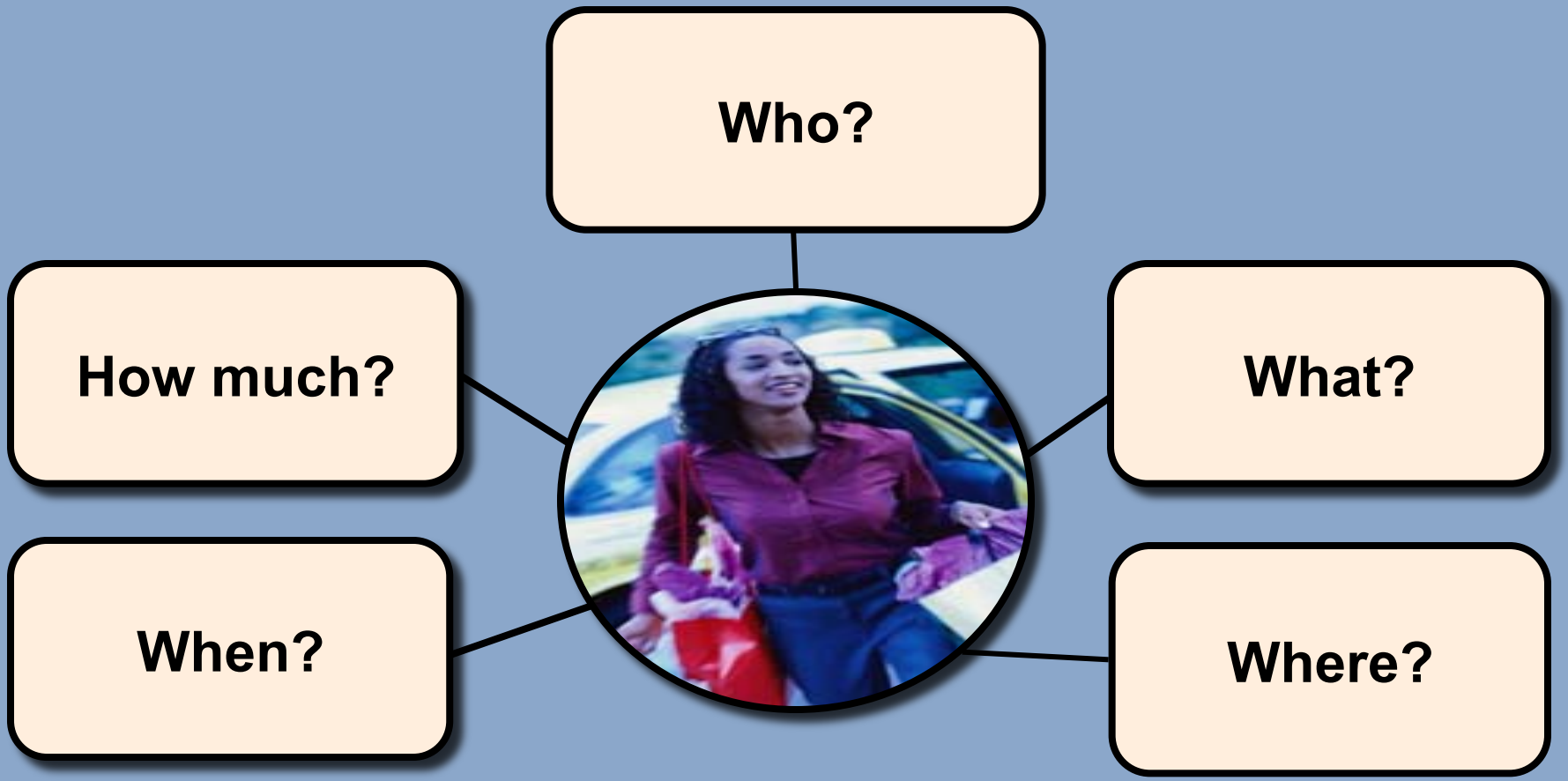
Statistical Study

- Breadth
- Population inferences
- Quantitative
- Generalizable findings

Case Study

- Depth
- Detail
- Qualitative
- Multiple sources of information

Descriptive Studies



Causal Studies

Experiment

- Study involving the manipulation or control of one or more variables to determine the effect on another variable

Ex Post Facto study

- After-the-fact report on what happened to the measured variable

Methods of Data Collection

Monitoring

Communication

The Time Dimension

Cross-sectional

Longitudinal

The Research Environment

Field conditions

Lab conditions

Simulations

Participants' Perceptions

No deviation perceived

Deviations perceived as unrelated

Deviations perceived as researcher-induced

Approaches for Exploratory Investigations

- Interviewing
- Participant observation
- Film, photographs
- Projective techniques
- Psychological testing

- Case studies
- Street ethnography
- Elite or expert interviewing
- Document analysis
- Proxemics and Kinesics

Common Exploratory Techniques for Research

Secondary Data Analysis

Experience Surveys

Focus Groups

Experience Surveys

- What is being done?
- What has been tried in the past with or without success?
- How have things changed?
- Who is involved in the decisions?
- What problem areas can be seen?
- Whom can we count on to assist or participate in the research?

Focus Groups

- Group discussion
- 6-10 participants
- Moderator-led
- 90 minutes-2 hours

Descriptive Studies

Descriptions of population characteristics

Estimates of frequency of characteristics

Discovery of associations among variables

Evidence of Causality

Covariation between A and B

Time order of events

No other possible causes of B

Selected Issues in Study Design

Most problems in studies are due to poor design (not poor analysis)

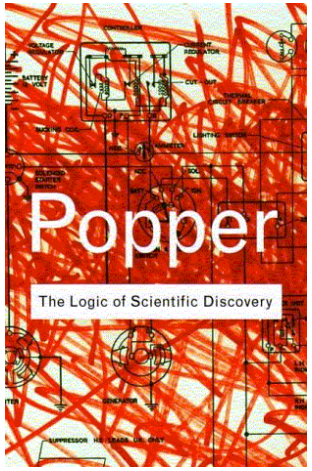
The Research Question

When I came to practice I was looking for answers like everybody else. For years I asked "what's the right answer?" Now I am learning "What is the right question?"

- Science is the holding of multiple working hypotheses (Thomas Huxley)
- A study is only as good as its hypothesis
- But where do hypothesis come from?

  observation + biological understanding + social understanding + intuition → **causal hypothesis**

  Admittedly, creative action can never be fully explained. (Popper)



Hypothesis Refinement

- Research is an ongoing process of hypothesis generation, refutation, refinement, and corroboration
- Results from a single study are seldom definitive (or even clear)
- So how do you know whether a hypothesis is correct?
- Good scientific practice . . . places the emphasis on reasonable scientific judgment and the accumulation of evidence and not dogmatic insistence of the unique validity of a certain procedure (Jerome Cornfield cited in Vandenbroucke & de Craen, 2001)
- There is no such as "proof" (in the mathematical sense in science), but there is "proof" that it "works":

> *When you ask people what made the modern West different from other cultures around the world, most of the answers are terribly negative: the disenchantment of the world, the destabilization of the earth, the death of God, the death of the Goddess, nightmare after nightmare. These naysayers tend to overlook the 40 years of life extension that the West has given us, the wonders of modern physics, modern medicine, the abolition of slavery, the rise of democracies, the rise of feminism, and so on. Until we honor both the good and bad news of modernity, we're not going to see our situation clearly*. -- Ken Wilber

Beautiful Theory, Ugly Fact

Science is organized common sense where many a beautiful theory is killed by an ugly fact (Thomas Huxley)

- Our job is to draw conclusions based on "ugly fact"
- Illustrative example: "Whole language learning education theory"
  - Educational theorists long pushed the "whole language" approach to teaching reading and talked down the need for breaking words into basic sounds called "phonics."
  - In 2000, a national panel reviewed ugly facts from 52 randomized studies.
  - Conclusion: no matter what the theory says, phonics is essential in teaching reading.

How do we create a study to gather ugly facts?

- There is no recipe for study design
- However, it helps to know
  - Elements of design
  - Where studies tend to go astray

Selected Elements of Study Design

- Measurement accuracy (variables)
- Effects can only be gauged relative to baseline (provided by a control group)
- Experimental studies differ from non-experimental studies (of course)
- The unit of recorded measure - individual or aggregate (ecological)
- Upstream and downstream causes should be considered
- Measurements may be longitudinal in individuals over time
- Cohort or case-control samples
- Hypothesis testing ("analytic") or hypothesis generating ("descriptive") studies
- Is the exposure randomized?
- Are groups comparable at baseline (confounding)
- Will you use prospective or retrospective measurements?
- Incident or prevalent cases?
- Matched or independent samples?
- Will you blinded subjects and/or observers?
- Is the study based in an open- or closed-population?
- There are too many design elements to discuss in a single week. We can't cover them all!

Comparative studies may be classified as:

I. Experimental - investigator assigns an intervention to see if he or she can influence a response

Randomized experiments

Non-randomized experiments

II. Observational – no investigator intervention *per se*

Cohort

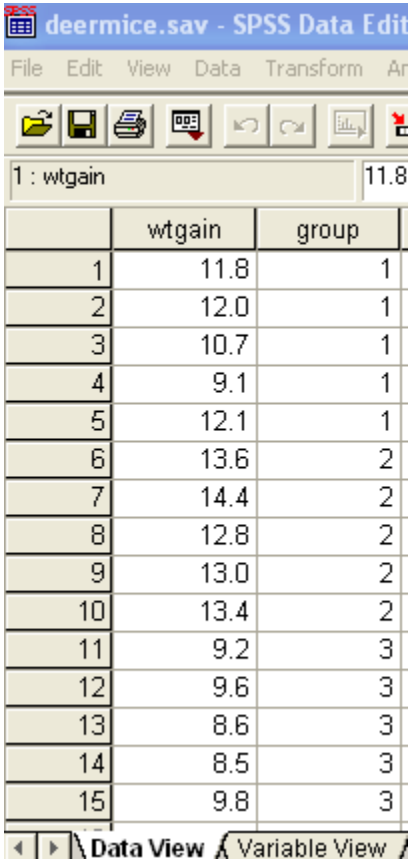
Case-Control

Cross-sectional

Ecological

Weight Gain on Different Diets

Explanatory variable = diet group (1=standard, 2=junk, 3=health)
Response variable = weight gain (grams)

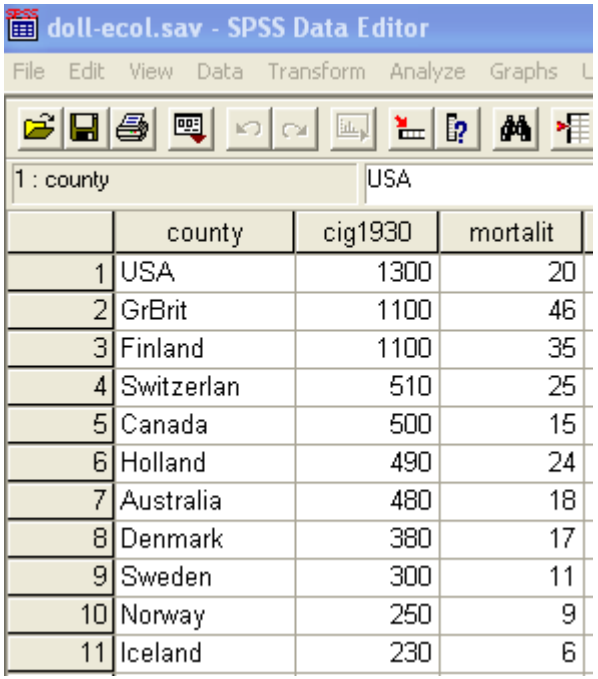


	wtgain	group
1	11.8	1
2	12.0	1
3	10.7	1
4	9.1	1
5	12.1	1
6	13.6	2
7	14.4	2
8	12.8	2
9	13.0	2
10	13.4	2
11	9.2	3
12	9.6	3
13	8.6	3
14	8.5	3
15	9.8	3

*Data are **experimental** because the investigator assigned the explanatory variable*

Cigarettes and Lung Cancer Mortality

Explanatory var = per capita cigarette consumption (cig1930)
Response var = lung cancer mortality per 100,000 (mortalit)

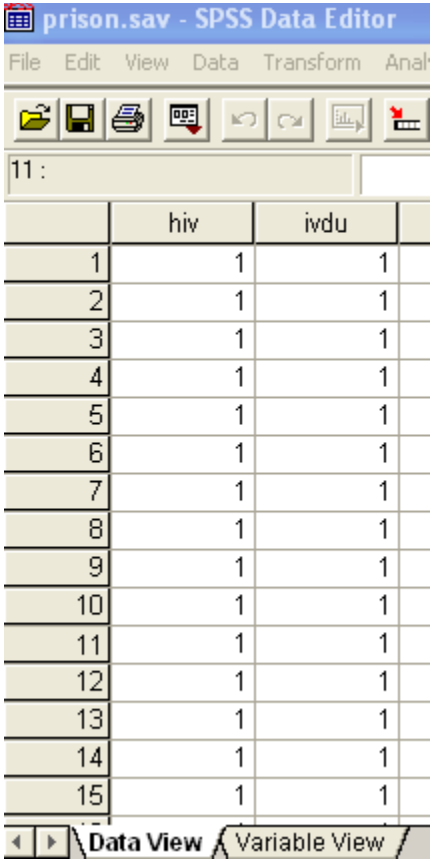


	county	cig1930	mortalit
1	USA	1300	20
2	GrBrit	1100	46
3	Finland	1100	35
4	Switzerlan	510	25
5	Canada	500	15
6	Holland	490	24
7	Australia	480	18
8	Denmark	380	17
9	Sweden	300	11
10	Norway	250	9
11	Iceland	230	6

Data are observational with data on aggregate-level. This is an ***ecological study***

HIV in a Women's Prison

Explanatory var = IV drug use (1 = users, 2 = non-user)
Response var = HIV serology (1 = positive, 2 = negative)

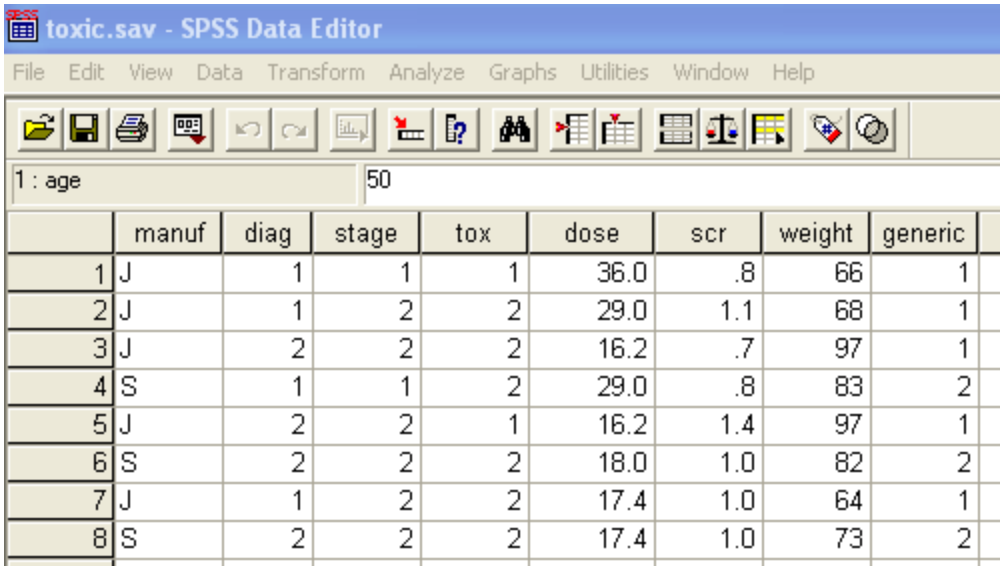


	hiv	ivdu
1	1	1
2	1	1
3	1	1
4	1	1
5	1	1
6	1	1
7	1	1
8	1	1
9	1	1
10	1	1
11	1	1
12	1	1
13	1	1
14	1	1
15	1	1

*Data are observational on the individual-level. But onset data cannot be unraveled. Thus, data are **cross-sectional***

Toxicity in Cancer Patients

Explanatory variable = generic drug use (generic: 1 = yes, 2 = no)
Response variable = cerebellar toxicity (tox: 1 = yes, 2 = no)



	manuf	diag	stage	tox	dose	scr	weight	generic
1	J	1	1	1	36.0	.8	66	1
2	J	1	2	2	29.0	1.1	68	1
3	J	2	2	2	16.2	.7	97	1
4	S	1	1	2	29.0	.8	83	2
5	J	2	2	1	16.2	1.4	97	1
6	S	2	2	2	18.0	1.0	82	2
7	J	1	2	2	17.4	1.0	64	1
8	S	2	2	2	17.4	1.0	73	2

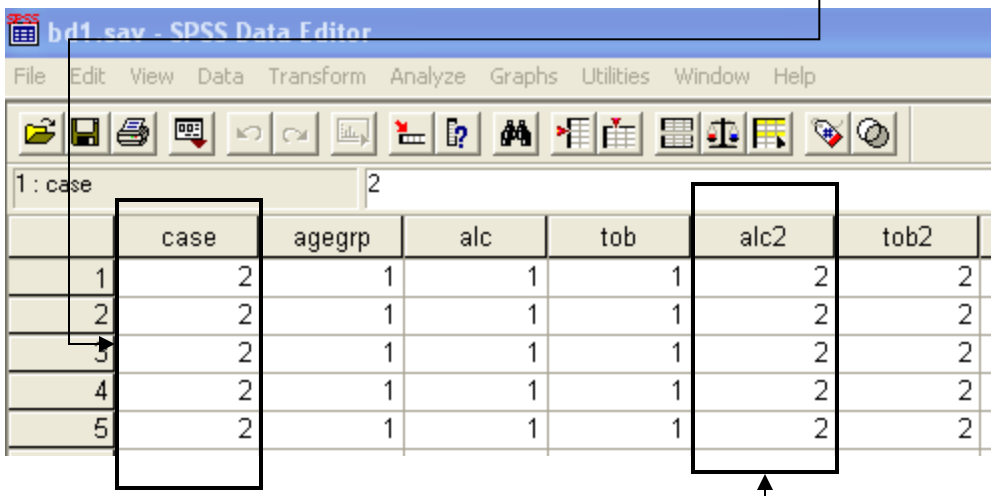
*Data are observational, individual-level, longitudinal, with all individuals followed over time. Thus, data are **cohort**.*

Comment: This is a *retrospective* cohort based on data abstracted data from medical records.

Esophageal Cancer and Alcohol Consumption

Explanatory var = alcohol consumption (alc2: 1 = high, 2 = low)

Response var = esophageal cancer (case: 1 = case, 2 = control)



	case	agegrp	alc	tob	alc2	tob2
1	2	1	1	1	2	2
2	2	1	1	1	2	2
3	2	1	1	1	2	2
4	2	1	1	1	2	2
5	2	1	1	1	2	2

Data are observational, individual-level, with study of all population cases but only a sample of non-cases. Thus, data are ***case-control****.*

Error in Research

- All research has errors
- Two types of errors
  - Random error
  - Systematic error