Dissertation Module
Research Skills Program

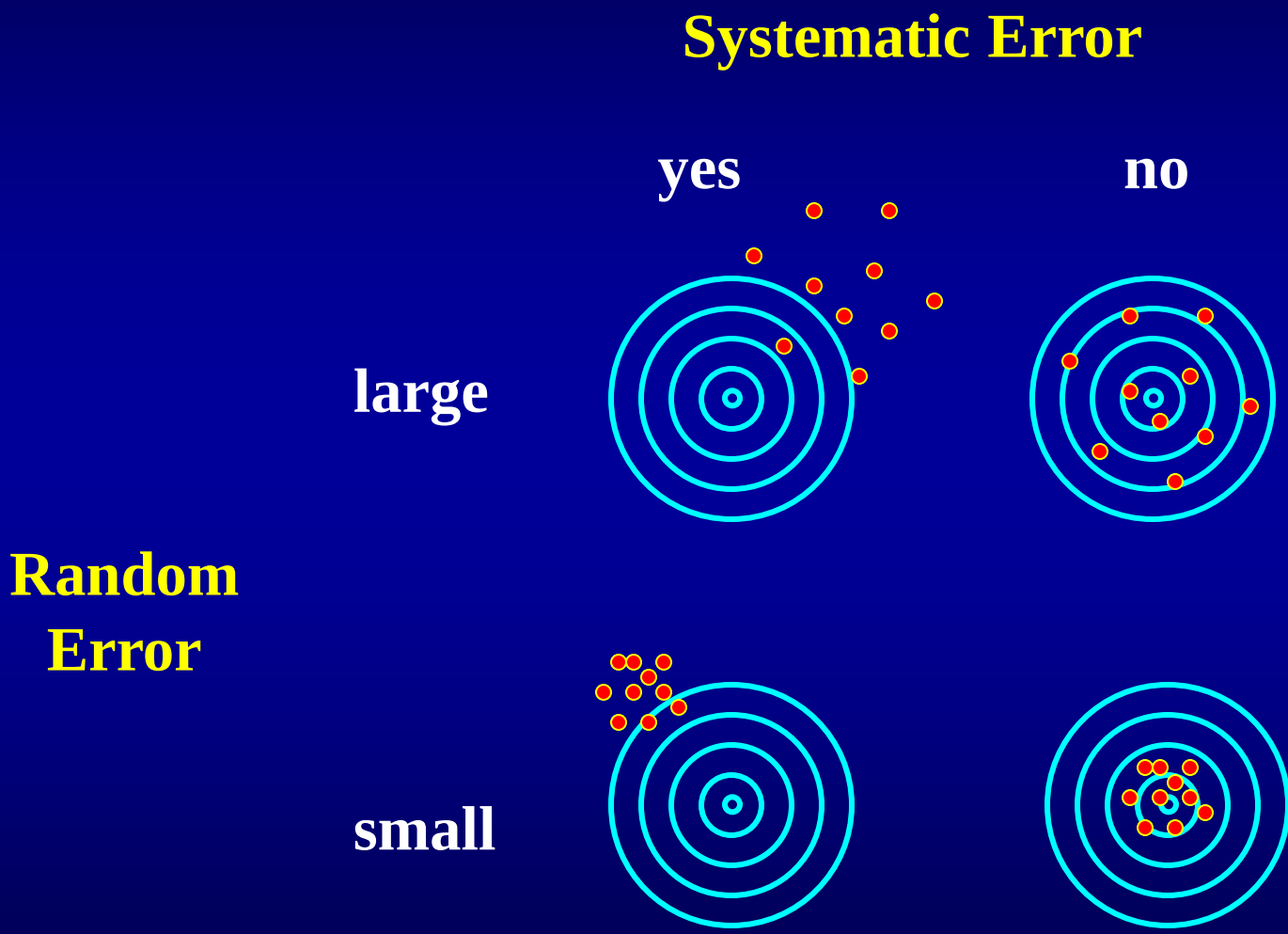
LECTURE 8

SOURCES OF BIAS

LEARNING OBJECTIVES

- Be able to give a definition of bias.
- Be able to identify selection bias, information bias, and potential confounding.
- Be able to advise on how to minimise bias.

Generalizibility - Internal Conformity and Consistency



Generalisibility - Internal Conformity and Consistency

Conformity =	**Absence of Systematic Error (= Bias)**
Consistency =	**Small Random Error**
Generalisibility =	**Conformity & Consistency**

Sound research methods try to avoid systematic error.

Statistics judges random error.

Systematic Error in Analytical Studies: BIAS

If systematic error occurs in the conduct of a study, which leads to misinterpretation of the effect measure (e.g. relative risk) this misinterpretation will be called BIAS.

Only unbiased studies can be regarded as valid for the target population.

Prerequisites for Comparison

Representative Uniformity (avoids selection bias)

The sample(s) taken taken from the actual population has (have) to be representative of the target population

Observational Uniformity (avoids information bias)

The sample(s) under study has (have) to be observed by the same means, with the same intensity, with equal circumstances, equal documentation …..

Structural Uniformity (avoids confounding bias)

The samples under study have to be as alike as possible with respect to structural characteristics and potentially influencing factors

Types of Bias

SELECTION BIAS ☹ **Representative Uniformity**

e.g. **Choice of groups**
Loss to Follow-up

INFORMATION BIAS ☹ **Observational Uniformity**

e.g. **Systematic Measurement Error**
Incorrect Diagnostic Criteria
Misclassification

CONFOUNDING BIAS ☹ **Structural Uniformity**

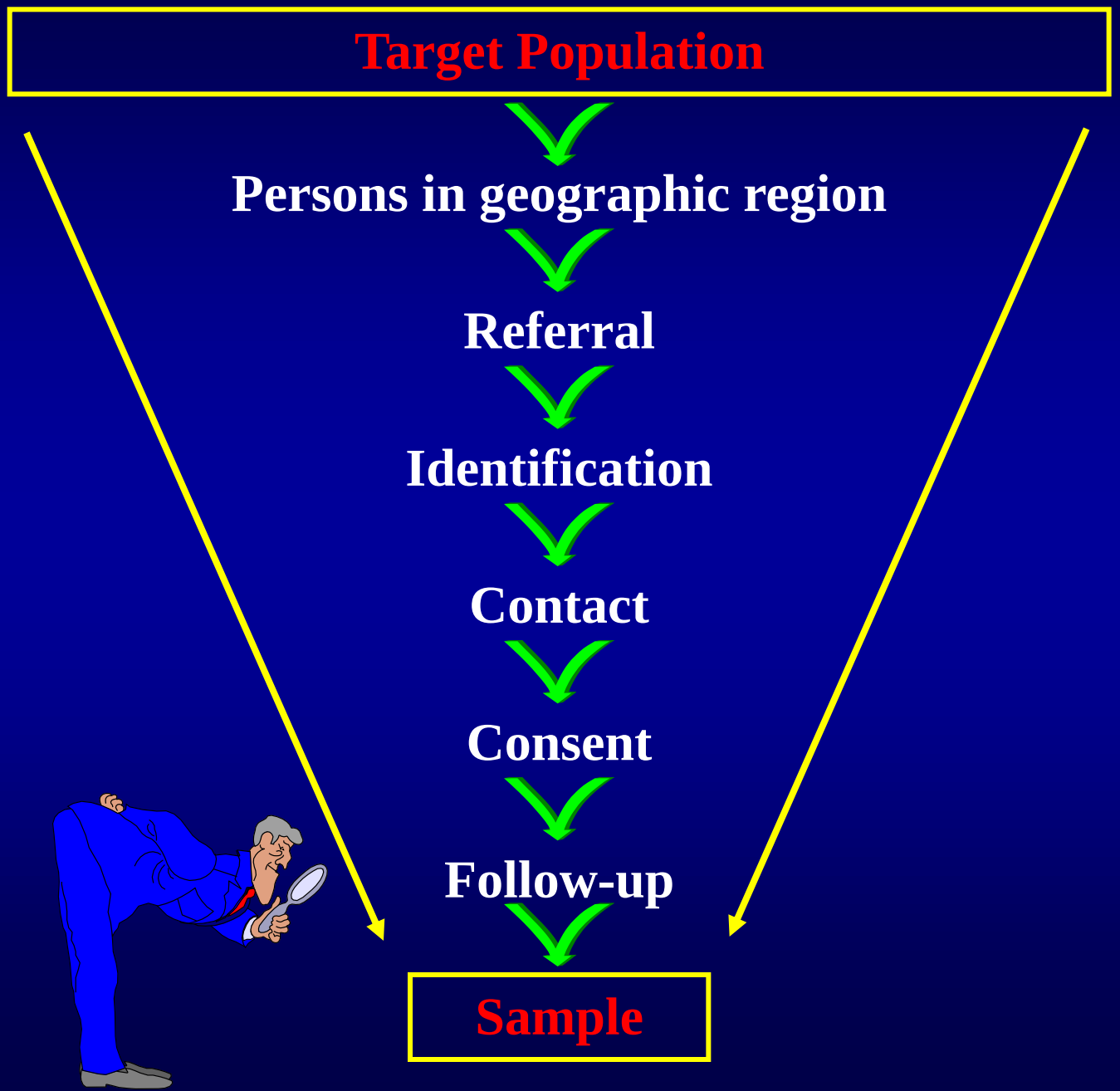
Selection Bias

Selection Bias refers to a distortion in the effect measure, resulting from the manner in which the people are selected for the sample(s).

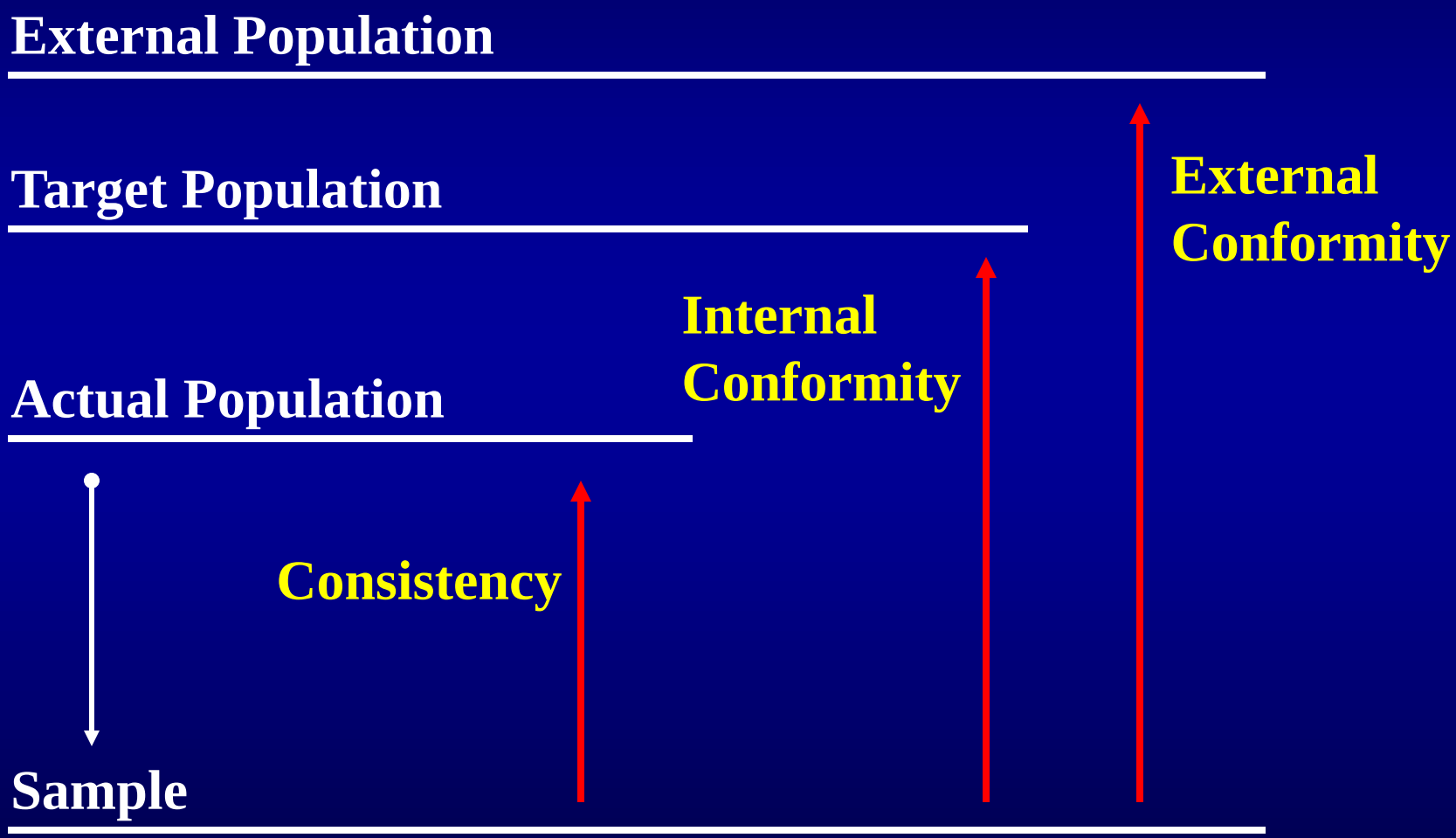
Selection Bias may be introduced by inappropriate sampling techniques.

If selection bias is present, the sample(s) are not representative of the target population.

The Long Way from the Target Population to the Sample



The Hierarchy of Populations



Example: Populations and Sample

A study explored reasons for stuttering in pre-school children (aged between 3 and 6 years) in Australia.

Overall, 330 children aged between 3 and 5 years participated in the study. These children were attending nine-teen pre-schools and day-care centers in Townsville, Brisbane and Cairns. The researchers interviewed the parents and examined the children to collect information on stuttering and potential determinants of stuttering.

Who constitutes the

- **Sample?**
- **Actual Population?**
- **Target Population?**

A study explored reasons for stuttering in pre-school children (aged between 3 and 6 years) in Australia. Overall, 330 children aged between 3 and 5 years participated in the study. These children were attending nine-teen pre-schools and day-care centers in Townsville, Brisbane and Cairns. The researchers interviewed the parents and examined the children to collect information on stuttering and potential determinants of stuttering.

TARGET POPULATION

All Australians children aged 3 to 6 years.

SAMPLE **330 children aged 3 to 5 years, attending day-care or pre-school in Townsville, Brisbane and Cairns.**

ACTUAL POPULATION

All eligible children aged 3 to 6 years who attend one of the 19 pre-schools and day-care centers.

Researchers wanted to investigate sleeping patterns and sleeping problems of elderly Australian residents.

The researchers identified a random sample of 45 nursing homes in Sydney, Melbourne and Brisbane. A total of 33 of these nursing homes, which were home to 1,716 Australian residents aged 65 years or older at the time of the study, agreed to participate in the study. Of these 1,716 elderly Australian residents, 1,191 were eligible for the study and gave their written informed consent.

The description of sleeping patterns and sleeping problems was based on 933 persons who had answered all the respective questions.

- **Sample?**
- **Target population?**
- **Actual population?**

Researchers wanted to investigate sleeping patterns and sleeping problems of elderly Australian residents = TARGET.

The researchers identified a random sample of 45 nursing homes in Sydney, Melbourne and Brisbane. A total of 33 of these nursing homes, which were home to 1,716 Australian residents aged 65 years or older at the time of the study, agreed to participate in the study. Of these 1,716 elderly Australian residents, 1,191 were eligible for the study and gave their written informed consent.

The description of sleeping patterns and sleeping problems was based on 933 persons = SAMPLE who had answered all the respective questions.

Selection procedure for sample

- **Random sample of 45 nursing homes**
- **Sydney, Melbourne, Brisbane**
- **33 nursing homes consented**
- **1191 persons eligible & consented**
- **933 answered questions**

Actual population = all elderly Australians

who live in potentially consenting nursing homes

in large cities

who potentially consent to participate and

who potentially answer all questions

Example: Selection Bias

Research: Reactions to psychological stress tests
Target population: Healthy people in Australia.

Researchers were at James Cook University.
They tried to invite all university students.
They put an ad into the students newspaper, some students responded.
They had individual talks with every student who referred themselves and a group was selected as “qualified”.
Students were informed in detail about the study, and some gave their informed consent.

Selection Bias: Volunteer Bias

Volunteer bias is a potential bias in nearly all epidemiological studies as participants usually have to declare their informed consent and it is quite likely that the *volunteers* differ from people who decline.

Volunteers may differ in being more or less informed, educated, wealthy, desperate ….

In most cases it is extremely complex to quantify or even to determine just the direction of a volunteer bias !

Types of Bias

SELECTION BIAS ☹ **Representative Uniformity**

e.g. **Choice of groups**
Loss to Follow-up

INFORMATION BIAS ☹ **Observational Uniformity**

e.g. **Systematic Measurement Error**
Incorrect Diagnostic Criteria
Misclassification

CONFOUNDING BIAS ☹ **Structural Uniformity**

Types of Bias: Information Bias

Information bias refers to a distortion in the estimation of the effect measure due to measurement error or misclassification of participants on one or more variables.

Some possible sources

- **Invalid measurement**
- **Incorrect diagnostic criteria**
- **Invalid questionnaire / interview etc.**
- **Imprecision / omission in previously recorded data**
- **Unequal diagnostic surveillance**

Example: Misclassification of Disease Status

Hypothetical cohort study using a diagnostic test

True Situation

	D	$\overline{D}$
E	400	600
$\overline{E}$	200	800

Relative Risk = 2

Diagnostic Test:
Sensitivity 0.8 Specificity 0.9

	D	$\overline{D}$
E	380	620
$\overline{E}$	240	760

Relative Risk = 1.6

1st cell: correct positive: 400 x 0.8 = 320
false positive: 60 since 600 x 0.9 = 540
320 + 60 = 380

Information Bias / Misclassification Bias

Non-Differential Misclassification

The same misclassification occurs in diseased and not diseased, and / or in exposed and non-exposed

The resulting bias is always towards the null

Differential Misclassification

Different misclassification occurs in diseased and not diseased, and / or in exposed and non-exposed

The resulting bias is may be in either direction

Example: Differential Misclassification

A cohort study aims at comparing breast cancer incidences (outcome) in users and non-users of oral contraceptives (study factor).

More frequent physical examinations may occur in users of oral contraceptives, who require regular contact with their gynecologists in order to renew their prescriptions.

Because the gynecologists see the users of oral contraceptives more frequently than the non-users, it is likely that users will be earlier and more frequently diagnosed with breast cancer than non-users.

What is the direction of this (potential) bias?

Misclassification: Recall Bias

A case-control study aims to investigate the relationship between life-time sun exposure (study factor) and cutaneous melanoma (outcome).
Participants are asked about their previously experienced sun exposure.
It is likely that people who have experienced a potentially life-threatening disease tend to reflect more carefully about possible causes for their illness and, therefore, will recall for instance severe sunburns more frequently than controls: Recall Bias.

What is the direction of this (potential) bias?

Case-control studies are especially prone to recall bias, which often leads, to an overestimation of the true association between study factor and outcome.

Types of Bias

SELECTION BIAS ☹ **Representative Uniformity**

e.g. **Choice of groups**
Loss to Follow-up

INFORMATION BIAS ☹ **Observational Uniformity**

e.g. **Systematic Measurement Error**
Incorrect Diagnostic Criteria
Misclassification

CONFOUNDING BIAS ☹ **Structural Uniformity**

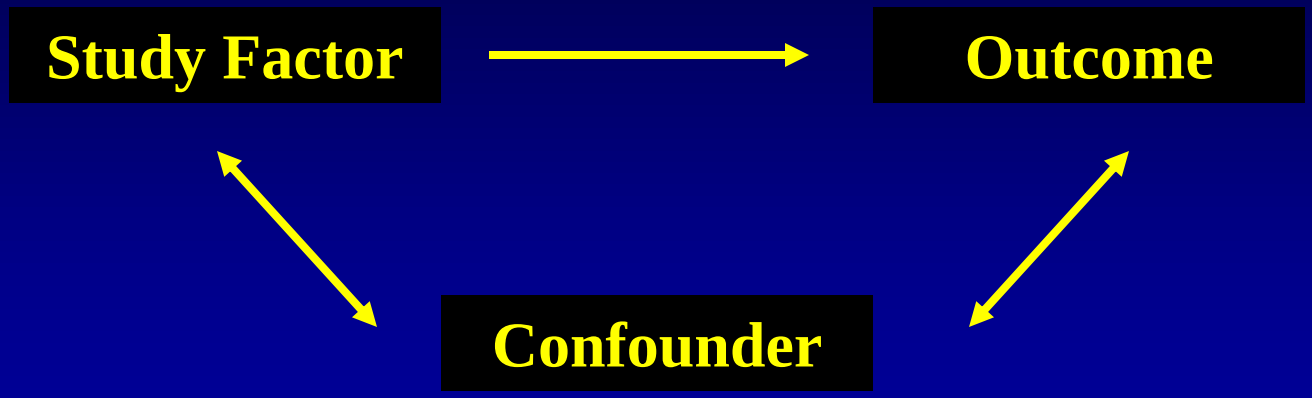
Types of Bias: Confounding Bias

Confounding is a type of bias that *may* occur when the effect of a study factor is mixed in the data with effects of third variables (= confounders) .

Confounding is a very common bias and may occur in all study designs.

The most important issue with respect to confounding is to identify potential confounding variables prior to the onset of the study and record information on the respective variables.

Types of Bias: Confounding Bias



A third variable may be a confounder if:

1. **Confounder and study factor are correlated**
2. **Confounder and outcome are correlated (independent of the study factor)**
3. **Confounder is not an intermediate variable**

Note: All 3 conditions have to be fulfilled!

Intermediate Variable

- Part of the study factor
- Implies same effect as study factor
- Pathologic change directly caused by study factor

Example: **Study factor = Total alcohol consumption**
Intermediate variable = Total wine consumption

Intermediate Variable

Direct cause for the study factor:

Intermediate → **Study Factor** → **Outcome**

Lifetime sun exposure — Number of sunburns — Melanoma

Direct effect of the study factor:

Study Factor → **Intermediate** → **Outcome**

Smoking — Chronic Bronchitis — Lung Cancer

Direct effect of the outcome:

Study Factor → **Outcome** → **Intermediate**

Needle sharing in drug users — HIV — Persistent Lymphadenopathy

What effect does bias have on my results?

Case-Control Study

Seat belt use and head injury

Seat belt use	Head Injury: Yes	Head Injury: No
No = E	10	2
Yes = $\bar{E}$	90	98
	100	100

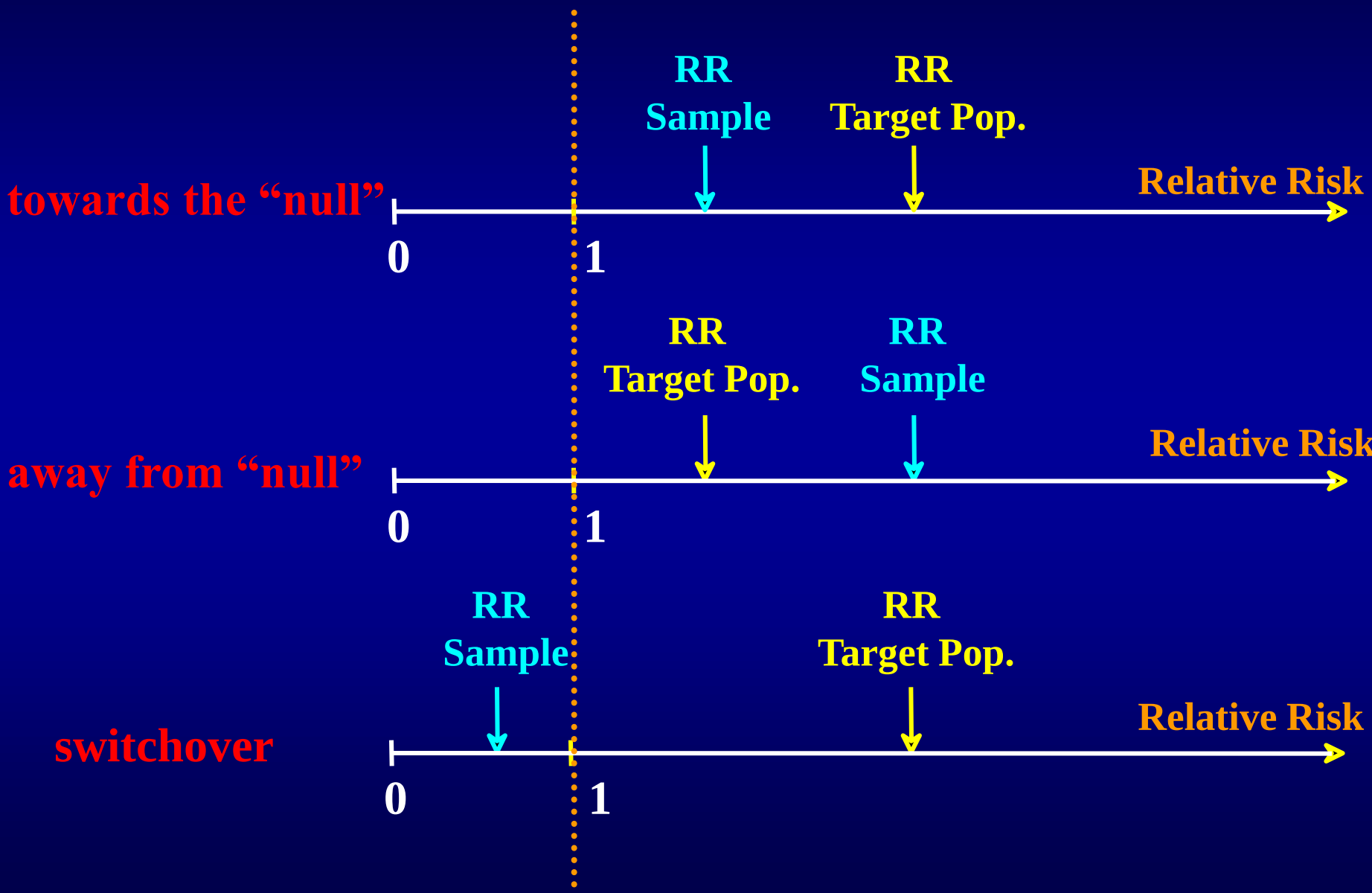
Car accident cases with head injury were 5.4 times more likely to drive without a seat belt than controls.

Exposure Odds Ratio: EOR

$$= \frac{\frac{a}{c}}{\frac{b}{d}} = \frac{a \times d}{c \times b}$$

$= 980 / 180 = 5.44$

Direction of Bias

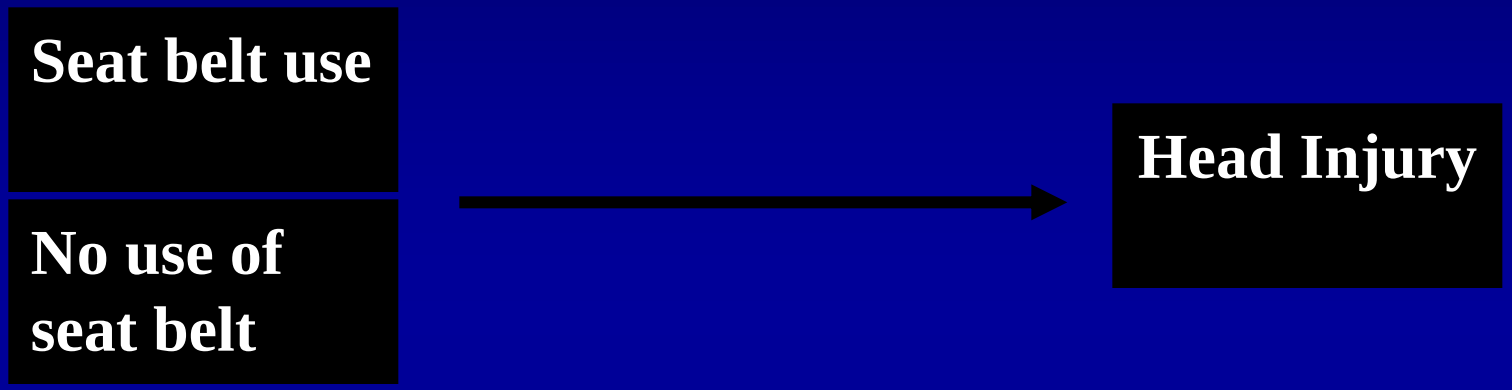


Direction of Bias

Case-control study on seat belt use and head injury in car accidents:

Study factor: Seat belt use (yes versus no)

Outcome: Head injury (No = controls versus Yes = cases)



- **Cases are younger**

CONFOUNDING BIAS

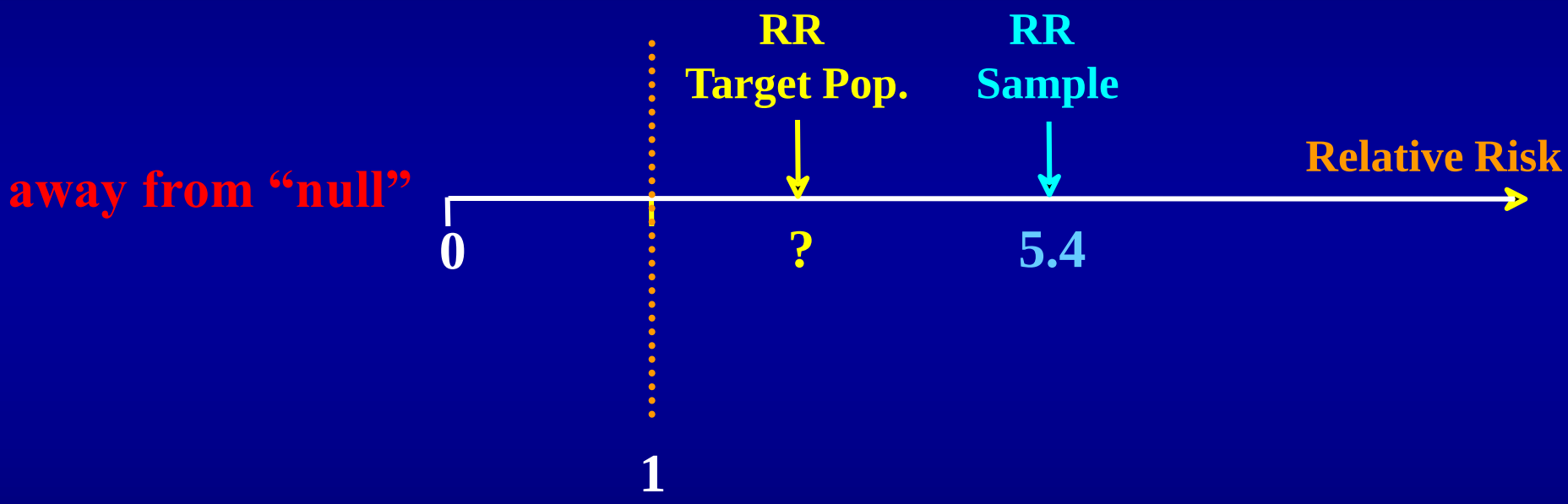
➡ **Younger people more likely NOT to use seat belt.**

➡ **Younger people more likely to have severe injuries.**

Overestimation of true effect of seat belt use.

Bias AWAY FROM the Null

Direction of Bias

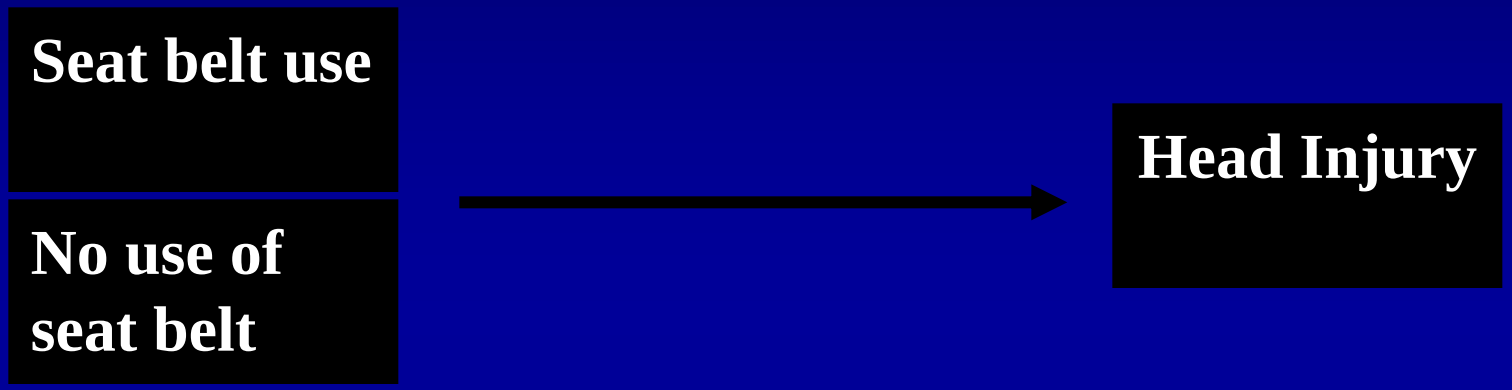


Direction of Bias

Case-control study on seat belt use and head injury in car accidents:

Study factor: Seat belt use (yes versus no)

Outcome: Head injury (No = controls versus Yes = cases)



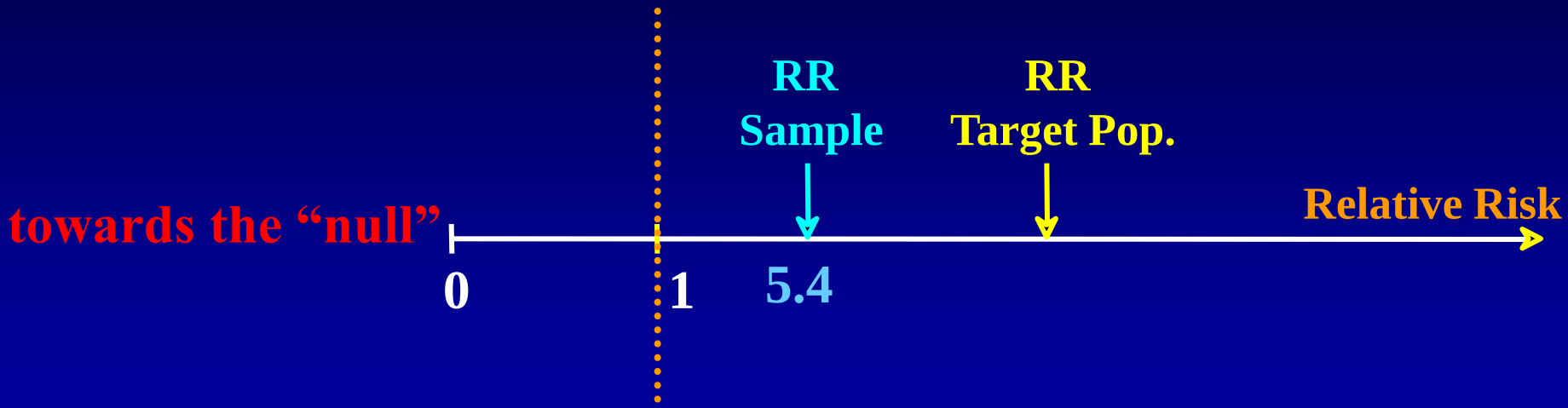
- **Cases do not recall seat belt use correctly INFORMATION BIAS**

➡ **Cases likely to over-report use of seat belt.**

➡ **Underestimation of true effect of seat belt use.**

Bias TOWARS the Null

Direction of Bias



Control of Bias

A priori: **Study Design**

A posteriori: **Statistical Analysis**

SELECTION BIAS	↔	**Representative Uniformity**
INFORMATION BIAS	↔	**Observational Uniformity**
CONFOUNDING BIAS	↔	**Structural Uniformity**

A priori Control of Bias

Experimental Studies

Inclusion / exclusion criteria:	Selection Bias
Randomization:	Confounding Bias
Blinding:	Information Bias
Placebo Controls:	Information Bias

Cohort Studies

Inclusion / exclusion criteria:	Selection Bias
Matching:	Confounding Bias
High follow-up rate:	Selection Bias
Identical intensity of observation:	Information Bias

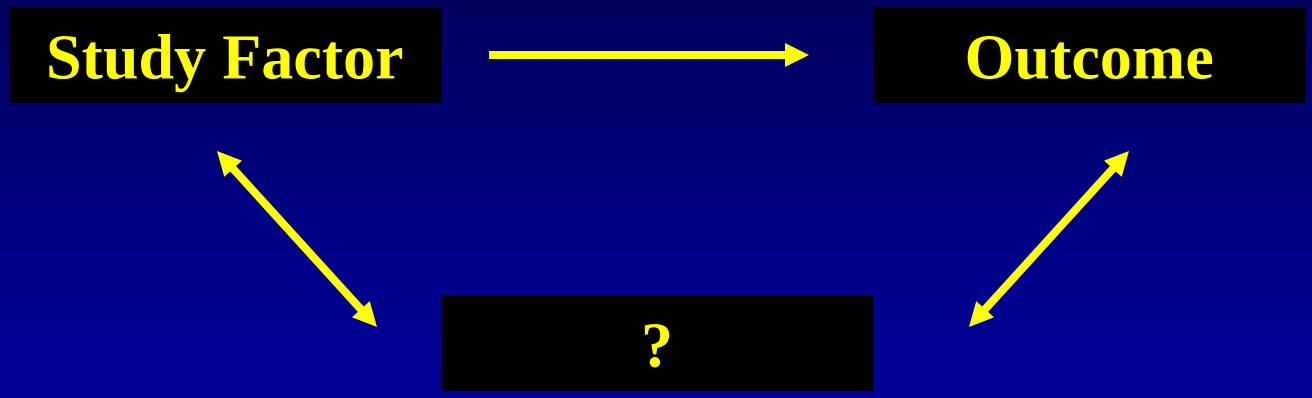
Case-Control Studies

Matching:	Confounding Bias
Selection of control group:	Selection Bias

Cross-Sectional Studies

Random sampling:	Selection Bias
	Confounding Bias

Control of Bias: Confounding Bias



Think about what variables are potential confounders.

Record these variables in your study!!!

A posteriori Control of Bias

Selection Bias

In general ??

The effect measure can be adjusted if information on selection probabilities is available.

Information Bias

In general ??

The effect measure can be adjusted if sensitivity and specificity of misclassification are known.

Confounding Bias

If information on potential confounders has been collected:

Stratification

Standardization

Multivariate statistical techniques

SUMMARY

- Bias is a systematic error which leads to a misinterpretation of the effect measure. Unbiased results are “valid”.
- We differentiate three main types of bias: selection bias, information bias, and confounding.
- Selection bias occurs when the sample does not represent the target population.
- Information bias occurs if exposure or outcome are systematically assessed in an inaccurate way.
- Confounding occurs if the effect of exposure on the outcome is mixed with a third variable (= confounder).
- We can sometimes discuss the likely direction of a bias.