

# ‘Counting people’ endpoints

- Example of a simple survey in dental undergraduates (cross-sectional survey)
- First look at study design, display of data and interpreting results
- Mainly interested in describing the characteristics of a single group of people

- What is the aim of the study?
- Pick any result, and interpret it.
- Say one thing you liked about the study.
- Say one thing you did not like about the study.
- What does the study contribute to clinical practice?

## Study objectives:

- To examine the smoking, alcohol and drug habits of **all** UK dental undergraduates

## Outcome measures:

To estimate the **prevalence** of:

- Smoking
- Alcohol
- Recreational drug use

# Selecting the sample

- The study was performed in one dental school
- Another way is to take a **random sample** of all dental schools
- For this we need a **sampling frame** (eg a list of all dental students)
- **Simple random sample**: every student has the same chance of being included in the study

- **Prevalence** of a disease (or attribute): the proportion of people with the disease (or attribute) measured at one point in time
- **Incidence rate**: the proportion of people who are **new** cases of the disease (or attribute) within a specified time period

Aim: To obtain the prevalence of smoking, alcohol use and drug use in dental undergraduates in the UK

Advantages of the study:

- All students in the dental school were surveyed; across all 5 years of study
- Anonymous – students are less likely to lie about sensitive questions
- Pre-labelled and pre-paid envelopes – increases response rate

Aim: To obtain the prevalence of smoking, alcohol use and drug use in dental undergraduates in the UK

### Disadvantages of the study:

- Only one dental school out of the 13 in the UK was included
- Assumes that students in this particular school are representative of all UK dental students
- Study done in 1998 and published in 2000; will the results and the conclusions still apply to students enrolling now?

- Look at the simple ways of displaying the data
  - Bar chart (Figure 1)
  - Frequency table (Tables 1 and 2)
- Results section subdivided into well-defined specific sections
- Think of the outcome measures and their definitions, eg
  - Current smoker or not
  - Binge drinker
  - Alcohol: number of units consumed in previous week, or average number of units per week
  - Regular user of drugs



# Question

- One endpoint of interest is ‘binge drinking’
- How was this defined?
- “Consumes at least half the weekly limit in one session ( $\geq 7$  units for females and  $\geq 10$  units for males)”
- Do you agree with the definition

## Binge drinking in students who drink alcohol

---

Gender & year	N	% who binge drink
---------------	---	-------------------

---

---

All students
--------------

---

172
-----

57
----

---

## Binge drinking in students who drink alcohol

Gender & year	N	% who binge drink*
<i>All males</i>	72	55
<i>All females</i>	100	59
All students	172	57

\*Consumes at least half the weekly limit in one session ( $\geq 7$  units for females and  $\geq 10$  units for males)

## Binge drinking in students who drink alcohol

Gender & year	N	% who binge drink*
Males 1-3	42	45
Males 4-5	30	70
<i>All males</i>	72	55
Females 1-3	64	69
Females 4-5	36	40
<i>All females</i>	100	59
All students	172	57

\*Consumes at least half the weekly limit in one session ( $\geq 7$  units for females and  $\geq 10$  units for males)

## Cannabis use since becoming an undergraduate

Gender & year	N	Ever used %	Regularly use %
Males 1-3	53	55	3.8
Males 4-5	34	74	14.7
<i>All males</i>	<i>87</i>	<i>62</i>	<i>8.0</i>
Females 1-3	73	45	4.1
Females 4-5	38	58	10.5
<i>All females</i>	<i>111</i>	<i>50</i>	<i>6.3</i>
All students	198	55	7.1

# What are the implications of conducting the study on a *sample* of people?

55% (109/198) had ever used cannabis.

If the study had included all students in all 13 UK dental schools, would the estimate still be 55%?

(we need to assume that subjects in this one dental school had similar characteristics to all UK students)

# What are the implications of conducting the study on a *sample* of people?

55% (109/198) had ever used cannabis.

If the study had included all students in all 13 UK dental schools, would the estimate still be 55%?

(we need to assume that subjects in this one dental school had similar characteristics to all UK students)

95% confidence interval 48% to 62%

# Interpreting these results

All dental students in the UK

True prevalence= ??

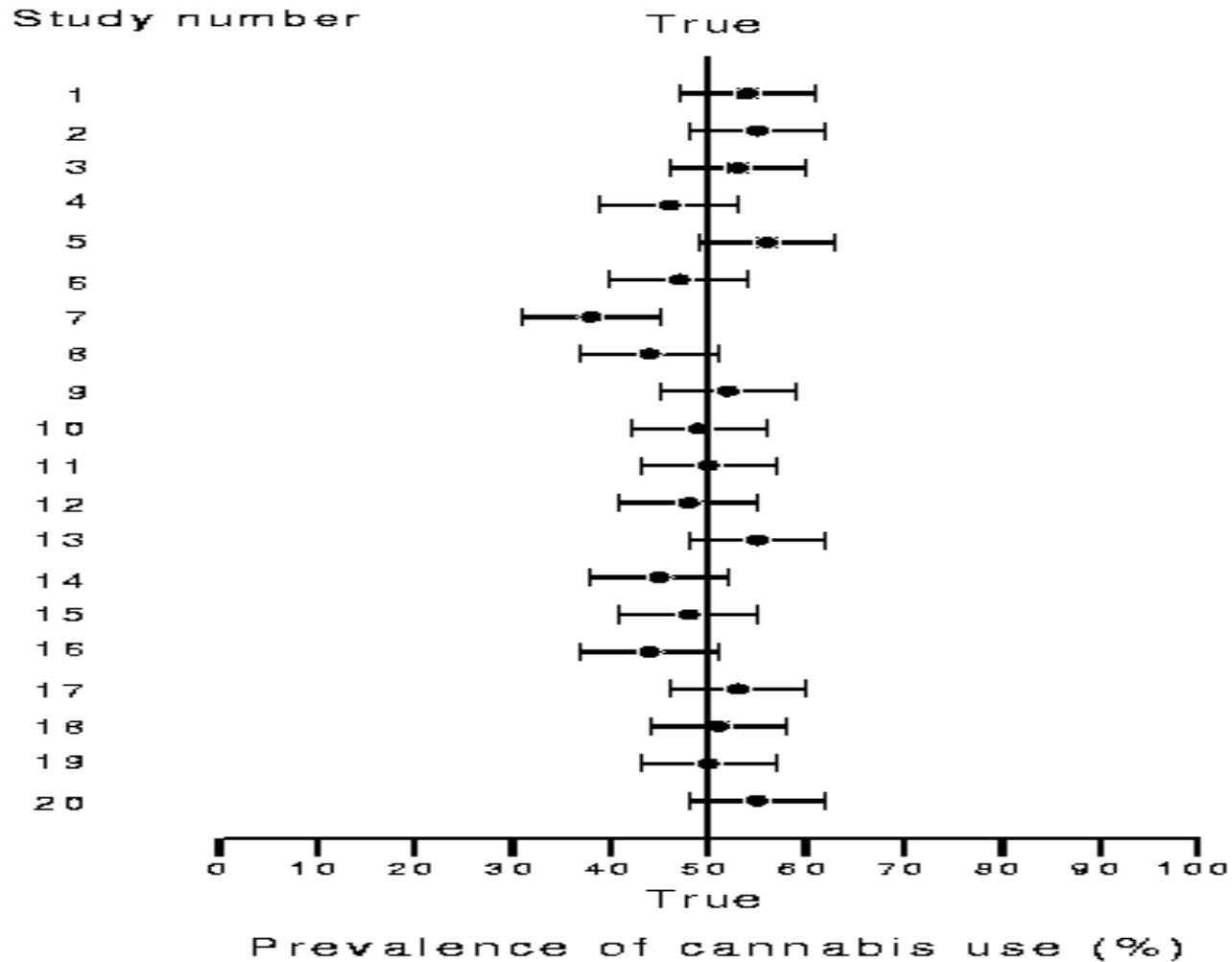
Sample of students in study (198)

Observed prevalence=55%

- Based on the data from this one dental school we think the true prevalence is 55%
- However, we cannot be entirely sure of this
- But there is a 95% chance that whatever the true value is, the range 48 and 62% contains it
- [This is not technically accurate, but we use this as an approximate guide to what the true value is (eg a conservative and optimistic estimate)]
- The 'true' value is always unknown & a fixed number (so either it is within the given 95% CI or it is not – ie there is no *probability* of this).
- 95% CIs technically mean that if we had several studies from the same population, then 95% of such confidence intervals would contain the true value]



20 studies – each based on 198 students



- Because we are using 95% CIs, it means we expect to get the wrong answer 5% of the time (ie 1 in 20)
- It does not necessarily indicate that something was wrong with a particular study
- It could simply mean we were unlucky enough to pick a sample that had different characteristics to most of the others in the population of interest (natural variation)
- Bear this in mind if your results are unexpected

- With  $n=198$  students, the 95% CI for the true prevalence is 48 to 62%
- What do you think happens to this interval if:
  - The study was bigger?
  - The study was smaller?

Number of students  
in the study

All 5000 in the UK

True prevalence

4000

3000

2000

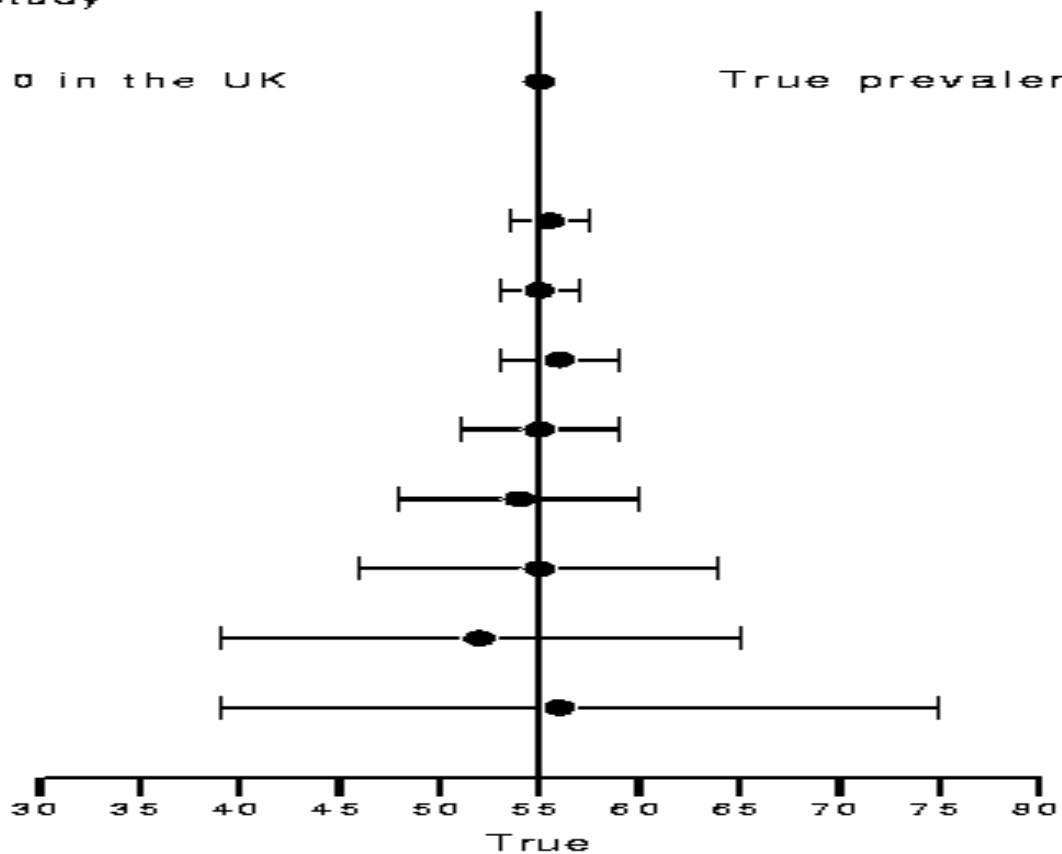
1000

500

198

100

50



Prevalence of cannabis use (%)

- When we conduct a study on a sample of people there will always be some uncertainty of how far our observed estimate is from the true value
- This uncertainty is quantified by the **standard error**
- If the study were based on the whole population (eg all dental students in the UK in 1998) we would have the true prevalence. There would be no uncertainty; the standard error would be 0
- By taking several different samples of the same size, the standard error tells us how much the prevalences, on average, spread about the true prevalence

Standard error of a prevalence is a measure of the uncertainty associated with trying to estimate the **true** prevalence when we only have a sample.

If observed prevalence =  $p$ , and sample size =  $n$ :

$$\text{Standard error} = \sqrt{\frac{p(1-p)}{n}}$$

Example: a prevalence of 55% based on 198 students:  $p=0.55$ ,  $n=198$

$$\text{Standard error} = \sqrt{\frac{0.55(1-0.55)}{198}} = 0.0354$$

Calculating a confidence interval for the **true** prevalence (or other proportion)

Lower limit = observed prevalence - 1.96 x standard error of prevalence

Upper limit = observed prevalence + 1.96 x standard error of prevalence

Example: Prevalence of 55% based on 198 students

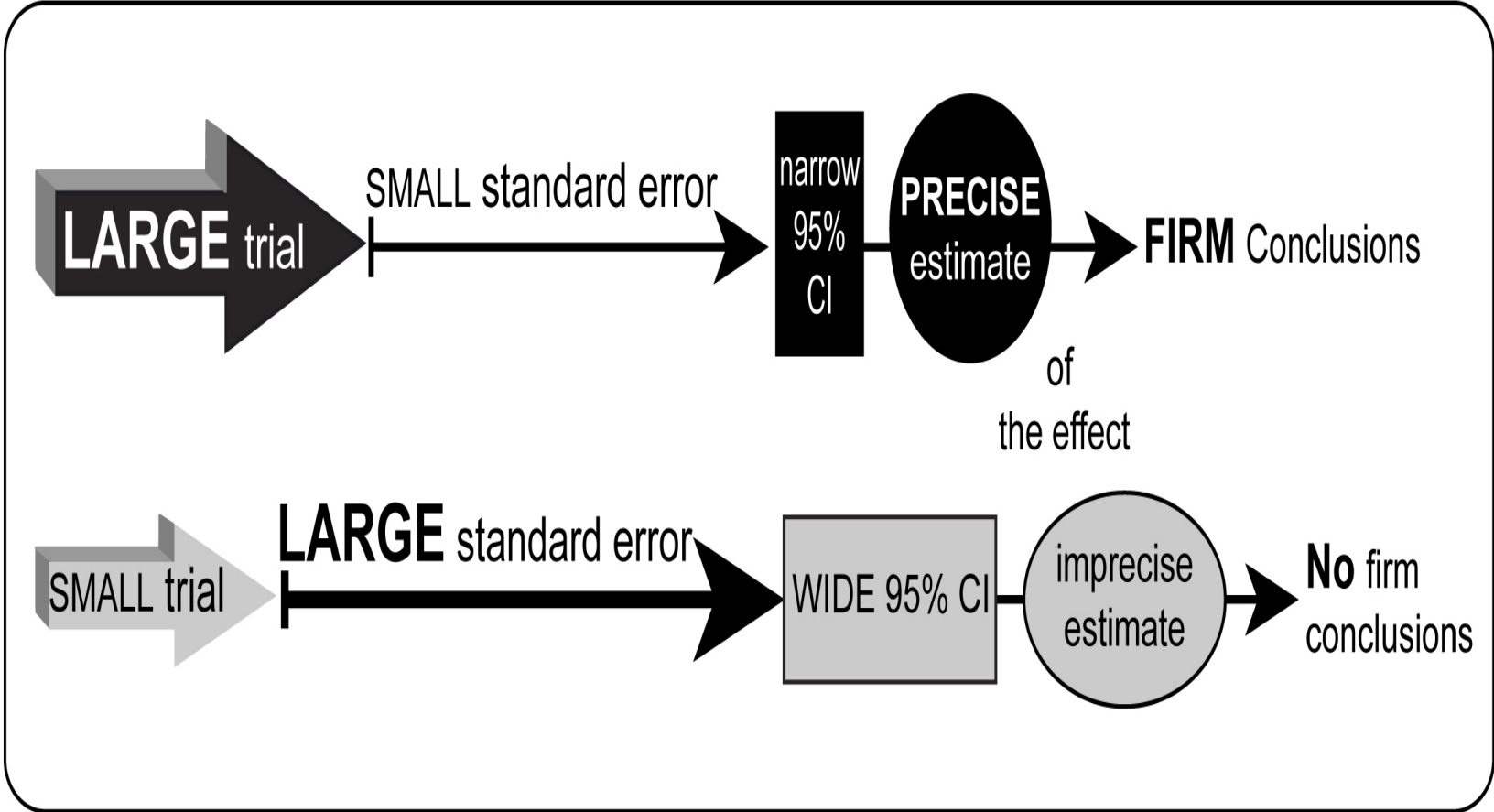
$p=0.55$ ,  $n=198$ , standard error = 0.0354

95% Confidence Interval =  $0.55 \pm \underline{1.96} \times 0.0354 = 48$  to 62%

If we want a 95% CI then we use the multiplier “1.96”

If we want to be more relaxed (eg 90% CI) the multiplier becomes 1.645

If we want to be more strict (eg 99% CI) the multiplier becomes 2.575





# Other considerations

- The response rate was 76% (is this high enough?)
- Could the 24% non-responders have very different characteristics?
- Does drinking and drug use affect exam performance?
- Does drinking and drug habits persist after graduating? What could the follow-on study be?
- Are the main results of this study what you might expect for:
  - (i) dental/medical undergraduates
  - (ii) similarly aged people from the general population?

# Possible issues in the dental survey

- Characteristics (eg ethnic group) of the 24% non-responders differ from the 76% responders
- The smoking, drinking and drug habits of non-responders differs to responders
- Some people misreport their consumption (eg heavy drug or alcohol users may report lower levels)
- **How could any of the above lead to the observed prevalence being over- or underestimated? (called biases)**