Making comparisons

- Previous sessions looked at how to describe a single group of subjects
- However, we are often interested in comparing two groups

Data can be interpreted using the following fundamental questions:

- Is there a difference? Examine the effect size
- How big is it?
- What are the implications of conducting the study on a sample of people (confidence interval)
- Is the effect real? Could the observed effect size be a chance finding in this particular study? (p-values or statistical significance)
- Are the results clinically important?

Effect size

- A single quantitative summary measure used to interpret research data, and communicate the results more easily
- It is obtained by comparing an outcome measure between two or more groups of people (or other object)
- Types of effect sizes, and how they are analysed, depend on the type of outcome measure used:
  - Counting people (i.e. categorical data)
  - Taking measurements on people (i.e. continuous data)
  - Time-to-event data

Example

Aim:

- Is Ventolin effective in treating asthma?

Design:

- Randomised clinical trial
- 100 micrograms vs placebo, both delivered by an inhaler

Outcome measures:

- Whether patients had a severe exacerbation or not
- Number of episode-free days per patient (defined as days with no symptoms and no use of rescue medication during one year)

Main results

Treatment group	No. of patients	proportion of patients with severe exacerbation	Mean No. of episode-free days during the year
GROUP A Ventolin	210	0.30 (63/210)	187
GROUP B placebo	213	0.40 (85/213)	152

Definition	**Type of outcome measure**
Proportion with severe exacerbation	Counting people (binary data) Exacerbation = Yes or No
Mean episode-free days	Taking measurements on people (continuous data) You measure the number of episode-free days for each patient

Outcome measures based on 'counting people'

- The proportion (or percentage) in each group can be called a risk
- Because we work with proportions or percentages it doesn't matter if the groups have the same number of subjects or not
- The risk of having a severe exacerbation is 30% in the Ventolin group and 40% in the placebo group
- An effect size involves quantitatively comparing these 2 risks

Effect sizes

- Take the ratio: 30% ÷ 40% = 0.75
- Called relative risk or risk ratio

•The risk with Ventolin (i.e. 30%) is 75% of the risk with placebo (i.e. 40%)

•Or, the risk with Ventolin is *reduced* by 25%, compared to placebo.

•Is this clinically important? Small, moderate or large effect?

- Note that it is not enough just to say the “risk associated with Ventolin is reduced by 25%”
- There is always a comparison group
- The correct statement is that the risk is reduced by 25% compared to placebo

Effect sizes

- Take the difference: 30% - 40% = -10 percentage points
- Called absolute risk difference
- Among patients given Ventolin there are 10% fewer with a severe exacerbation than patients given placebo (the minus sign above indicates fewer)
- Or, in every 100 patients given Ventolin there are 10 fewer with an exacerbation, compared to 100 patients given placebo.
- Is this clinically important?

‘No effect’ value

If there were no difference in the outcome measure between the two groups, what would be relative risk and risk difference be?

Effect size	No effect value
Relative risk	
Absolute risk difference	

‘No effect’ value

If there were no difference in the outcome measure between the two groups, what would be relative risk and risk difference be?

Effect size	No effect value
Relative risk	1
Absolute risk difference	0

The comparison or reference group

- Effect sizes almost always involve comparing two groups
- Therefore, which is made the reference group must always be clear
- In the example, we examine Ventolin versus placebo
- Relative risk = 30% ÷ 40% = 0.75
- Risk difference = 30% - 40% = -10 percentage points

The comparison or reference group

- However, we could examine placebo versus Ventolin
- Relative risk = 40% ÷ 30% = 1.33
- Risk difference = 40% - 30% = +10 percentage points

Interpreting relative risks

- A relative risk of 0.5 or 2.0 is easy to explain:
  The risk is *halved* or *doubled*
- Relative risks of say 0.75 or 1.33 are not so intuitive
- We therefore convert them to a **percentage change in risk**

Converting relative risks

No effect value:

RR=0.75

How far is it from 1?

Answer: 0.25
Multiply by 100, to turn into a percentage
Because it is below 1, we say the risk is reduced

Percentage change in risk = ‘25% reduction’

Converting relative risks

No effect value: 1

RR=0.75

How far is it from 1?

Answer: 0.25
Multiply by 100, to turn into a percentage
Because it is below 1, we say the risk is reduced

Percentage change in risk = ‘25% reduction’

Converting relative risks

No effect value: 1

RR=1.33

How far is it from 1?

Answer: 0.33
Multiply by 100, to turn into a percentage
Because it is above 1, we say the risk is increased

Percentage change in risk = ‘33% increase’ (sometimes called the ‘excess risk’)

- Generally, if the relative risk is <2.0, then convert to a percentage change in risk to interpret
- If above 2.0, leave as it is
- E.g., RR=10, is the same as an excess risk of (10-1)x 100 = 900%
- This looks cumbersome, so it is appropriate just to say that the risk has increased 10-fold

Defining risk

- The word ‘risk’ implies something bad, e.g. the risk of dying or the risk of developing a disease
- However, it can be used for any ‘counting people’ outcome measure, e.g. the risk of surviving to 1 year, the risk of not experiencing pain after surgery

Relative Risk or risk difference indicates the magnitude of the effect, but not whether the effect is beneficial or harmful; that depends on the definition of the outcome measure

GOOD OUTCOME ➡ benefit if RR>1

Outcome measure:

percentage of patients who recover from gingivitis after 1 month

Antibiotic A (90%)	Antibiotic B (70%)	RR 1.3 (90/70)

Antibiotic A better than B

BAD OUTCOME ➡ harm if RR>1

Outcome measure:

percentage of patients who experience pain after surgery

Treatment C (40%)	Treatment D (29%)	RR 1.4 (40/29)

Treatment C worse than D

Relative risk vs risk difference

Risk in Group A	Risk in Group B	Relative risk	Absolute risk difference
40%	80%	0.5	40 percentage points
10%	20%	0.5	10
1%	2%	0.5	1

Relative risk tends to be similar across different populations, and so does not depend on the background risk

Risk difference does depend on the background risk, and so is expected to vary between different populations

Risk versus odds

- Odds is another way of expressing chance
- If risk is 1 in 10 (ie 1/10)
- Then odds is 1:9 (ie 1/9)

- The denominator for risk is everyone
- The denominator for odds is everyone without the event of interest

Relative risk and odds ratio

To look at the difference between the risk and odds ratio consider the same example (Ventolin vs placebo) but in a group of patients where severe exacerbation was less common.

	Severe exacerbation		
	Yes	**No**	**Total**
Group A Ventolin	11 (a)	199 (b)	210 (n1)
Group B Placebo	22 (c)	191 (d)	213 (n2)

Risk of severe exacerbation in Group A = 11/210 = 5.2% (i.e. a/n1)
Risk of severe exacerbation in Group B = 22/213 = 10.3% (i.e. c/n2)
Relative risk = 5.2 ÷ 10.3 = 0.50

Odds of severe exacerbation in Group A = 11/199 (i.e. a/b)
Odds of severe exacerbation in Group B = 22/191 (i.e. c/d)
Odds ratio = 11/199 ÷ 22/191 = 0.48 [i.e. (axd) ÷ (bxc)]

Relative risk and odds ratio

Now consider the same example (Ventolin vs placebo) but in a group of patients where severe exacerbation much MORE common.

	Severe exacerbation		
	Yes	**No**	**Total**
Group A Ventolin	84 (a)	126 (b)	210 (n1)
Group B Placebo	170 (c)	43 (d)	213 (n2)

Risk of severe exacerbation in Group A = 84/210 = 40.0% (i.e. a/n1)
Risk of severe exacerbation in Group B = 170/213 = 79.8% (i.e. c/n2)
Relative risk = 40.0 ÷ 79.8 = 0.50

Odds of severe exacerbation in Group A = 84/126 (i.e. a/b)
Odds of severe exacerbation in Group B = 170/43 (i.e. c/d)
Odds ratio = 84/126 ÷ 170/43 = 0.17

Relative risk and odds ratio

- Odds ratios are used in several statistical analyses because they have useful mathematical properties that make some analyses easier to do
- When the disease is uncommon (say <20%), relative risk and odds ratio are similar, so we can interpret them in the same way
- Otherwise, odds ratio means something else. In the second example above, RR=0.50 and OR=0.17.
- The OR of 0.17 does not mean that the risk has been reduced by 83%.
- It means the odds in Group A is 0.17 times the odds in Group B
- The RR of 0.50 means that the risk has been reduced by 50% (halved).

Main results

Treatment group	No. of patients	proportion of patients with severe exacerbation	Mean No. of episode-free days during the year
GROUP A Ventolin	210	0.30 (63/210)	187
GROUP B placebo	213	0.40 (85/213)	152

Outcome measures based on ‘taking measurements on people’

- The mean episode-free days were
- 187 - Ventolin
- 152 – placebo

•Effect size is the difference between the means
•Mean difference = 187 – 152 = +35 days.

•Is this clinically important?

Interpretation:

- Patients given Ventolin had more episode-free days than those given placebo
- On average the difference is +35 days per year

Remember that:

•The mean difference of 35 days indicates the average for the group as a whole

•For some individual patients the difference will be less than 35 days, some more than 35 days

- The difference between 2 mean values often has nice mathematical properties (i.e., it follows a Normal distribution), and therefore easy to analyse
- The ratio between 2 means often does not have a Normal distribution, so is not usually specified as an effect size.
- Also, when looking at paired values for a patient (e.g. value at Time 0 and value at Time 1) the ratio between these two is impossible to get if one of the patient's value is zero
- The no effect value for the mean difference = 0

What is the true effect given we only have a sample of asthma patients in the study?

- Severe exacerbation
  - Relative risk = 0.75 (risk reduced by 25%)
  - Absolute risk difference = -10 percentage points
- Episode-free days
  - Mean difference = +35 days

If the study were conducted on a different group of patients, would we see identical results?

Effect size	**Estimate**	**95% confidence interval (CI)**
Risk difference	-0.10	-0.19 to –0.01
Relative risk	0.75	0.58 to 0.98
Percentage change in risk (minus sign indicates risk is reduced)	-25%	-2 to -42%
Mean difference	+35 days	22 to 48 days

NB: the first 3 above all relate to ‘risk’, the ‘mean difference’ has nothing to do with risk (it is simply a measurement)

Every asthma patient ever	True risk difference = ??
Trial of 423 patients	Observed difference = -0.10 95 % CI : -0.19 to -0.01

Interpretation:

•We think the true difference is that there are 10% fewer patients with an exacerbation using Ventolin

•But whatever the true effect is, we are 95% certain that it is somewhere between 1 and 19 percentage points (these give a conservative and optimistic estimate of the true effect)

•The range does not contain the no effect value (of 0), so we can be confident that the true risk difference is unlikely to be 0, i.e. there is likely to be a real effect.

Confidence interval (CI)

Relative Risk (RR) : 0.75

95 % CI : 0.58 to 0.98

Interpretation:

- We think the true relative risk is 0.75. But we are 95% certain that it is likely to lie somewhere between 0.58 and 0.98
- The range does not contain the no effect value (i.e. 1), so we can be confident that the true risk is unlikely to be 1.

Confidence interval (CI)

Percentage change in risk: -25%

95 % CI : -2% to –42%

Interpretation:

- We expect that the risk of having an exacerbation in patients given Ventolin is reduced by 25%.
- The true risk reduction is likely to lie somewhere between 2% and 42%.
- The range does not contain the no effect value (i.e. 0), so we can be confident that the true percentage change in risk is unlikely to be 0.

Confidence interval (CI)

Mean difference: +35 days

95% CI : +22 to +48 days

Interpretation:

•On average patients have 35 *more* episode-free days in the Ventolin group when compared to the placebo group.

•We are 95% sure that the true mean difference is between +22 and +48 days

•The range does not contain the no effect value (i.e. 0), so we can be confident that the true mean difference is unlikely to be 0.

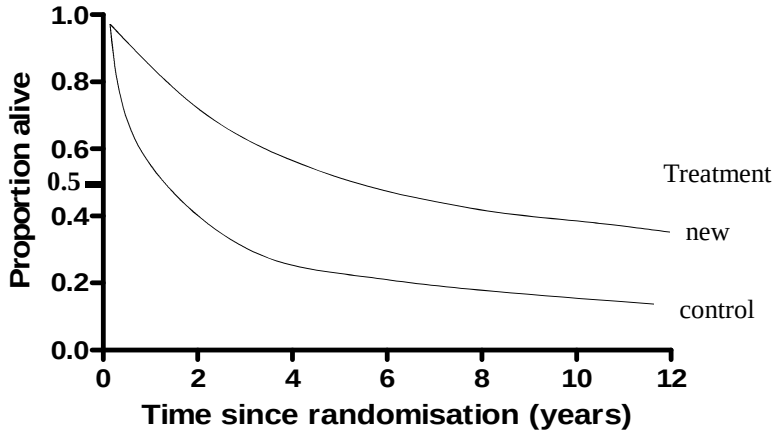
Outcome measures based on time-to-event data

- For a single group, a Kaplan-Meier curve can be drawn
- For 2 or more groups, we simply overlay these curves on the same diagram

•Effect size:

–Hazard ratio (the risk of having an event in Group 1 divided by the risk in Group 2, at the same point in time)

–Difference in survival or event rates at a specific time point



- The hazard ratio is interpreted like a relative risk
- e.g: a hazard ratio of 0.80 indicates that the chance of having an event is
- e.g: a hazard ratio of 1.40 indicates that the chance of having an event is

What do you think the no effect value for the hazard ratio is?

- The no effect value for a hazard ratio is 1

- The hazard ratio is interpreted like a relative risk
- e.g: a hazard ratio of 0.80 indicates that the chance of having an event is reduced by 20%
- e.g: a hazard ratio of 1.40 indicates that the chance of having an event is increased by 40%.

What do you think the no effect value for the hazard ratio is?

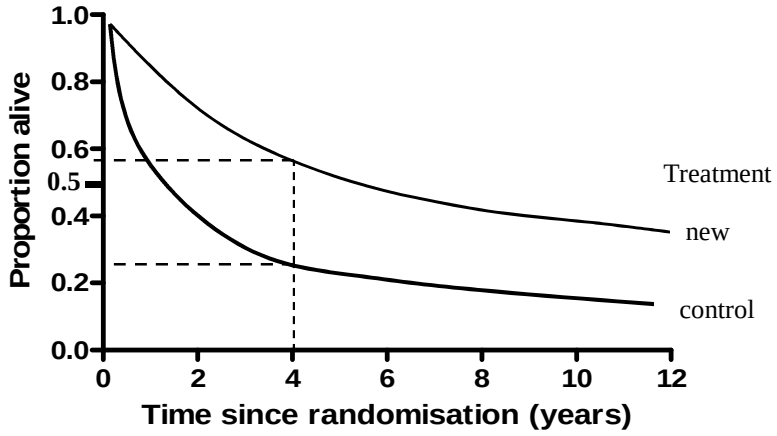
- The no effect value for a hazard ratio is 1

Relative risk or hazard ratio?

- They can be interpreted in the same way
- But in a *specific* study, they may be different:

	RR (had the event or not)	**HR (how long it took to get the event, or censored otherwise)**	**Comments**
Study 1	0.75 Risk reduced by 25%	0.55 Risk reduced by 45%	RR and HR very different. But HR is a more sensitive effect size because it has allowed for time (here time has mattered). So use HR here
Study 2	0.60 Risk reduced by 40%	0.58 Risk reduced by 42%	RR and HR quite similar, hence can use one or the other (time didn't really matter)

We can also get risk difference from Kaplan-Meier curves and they are interpreted in exactly the same way as before (i.e. with 'counting people' endpoints)



Risk difference at 4 years = 58% - 26% = +32 percentage points
The no effect value = 0

- But you should specify the time point before you look at the data
- It should be one that is clinically relevant
- Don’t choose the time point just because that is where the largest difference is!

- A hazard ratio is a good effect size for this type of data because it compares the whole curve in one group with the curve in another group
- The difference between two survival rates only applies to one time point, and can therefore be more influenced by variability

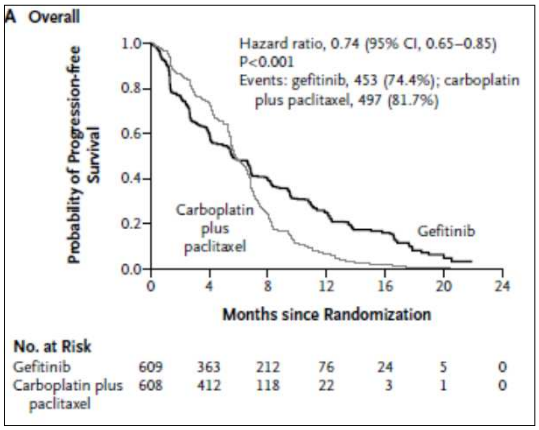
•Hazard ratios assume that the percentage difference in risk is the same over time, i.e. if there is a 20% reduction at 1 year, there is also a 20% reduction at 5 years

•This is called an ‘assumption of proportional hazards’

•When this is clearly not the case, the hazard ratio may not be appropriate (use risk difference at a time point instead)

An example of non proportional hazards.

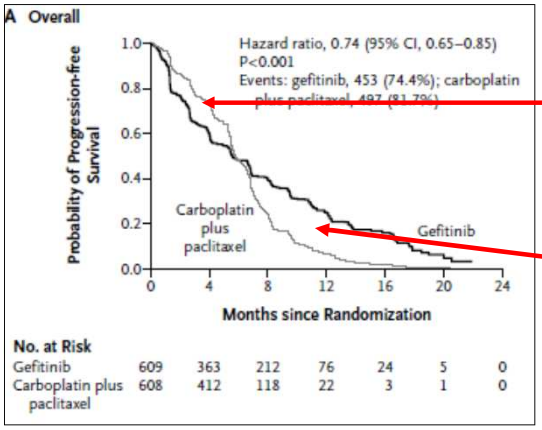
Gefitinib or Carboplatin–Paclitaxel in Pulmonary Adenocarcinoma



Curves that cross do not indicate proportional hazards.

Perhaps use risk difference instead

Gefitinib or Carboplatin–Paclitaxel in Pulmonary Adenocarcinoma



HR>1; gefitinib worse

HR<1; gefitinib better

Summarising data – comparing two groups of people (or things)

Type of outcome measure	Summary measure for each group	Effect size	What are the implications of conducting the study on a sample of people*?
Counting people (binary/ categorical data)	Percentage (proportion)	Relative risk Risk difference	95% confidence interval 95% confidence interval
Taking measurements on people (continuous data)	Mean & standard deviation	Difference between 2 means	95% confidence interval
	Median	Difference between 2 medians	95% confidence interval
Time-to-event measures	Kaplan-Meier curve	Hazard ratio	95% confidence interval
	Event rate at a specific time point	Risk difference	95% confidence interval
	Median time	Difference in median	

* And all confidence intervals are calculated using the standard error of the effect size

Type of outcome measure	Effect size	No effect value
Counting people (binary or categorical data)	Relative risk (risk ratio); odds ratio	1
	Percentage change in risk	0
	Absolute risk difference	0
Taking measurements on people (continuous data)	Difference between 2 means	0
	Difference between 2 medians	0
Time-to-event data	Hazard ratio	1
	Difference between 2 event rates at a specific time point	0

If the 95% CI for the effect size does not contain the appropriate ‘no effect’ value, then we can conclude there is likely to be a real effect