Introduction to Statistics

A/Prof Allan McRae
The University of Queensland

Overview

A lot of research on MND is reported by the media with varying degrees of sensationalism.
This talk will help you understand the research results and assess their impact.

A large proportion of MND research is comparing two groups

- **Observational study:** e.g. taking people with MND and people without-MND (often called "controls") and comparing them
- **Clinical Trials:** e.g. giving one group of people a treatment, and another a placebo and comparing the outcomes

- Statistics allow us to interpret the differences between the two groups
- Concepts we will cover:
  - P-value
  - Effect size
  - Bias
  - Replication

Study finds ex-rugby internationals' risk of neurodegenerative diseases is between two and 15 times higher than the public's

Posted Wed 5 Oct 2022 at 8:30am

Anecdotal vs Statistical Evidence

Anecdotal evidence for causes/treatments of disease is widespread

> ***"Sportspeople are at a higher risk of MND"***
>
> ***"This region of the country has a higher MND risk"***
>
> ***"This supplement is helping my MND symptoms"***

Anecdotal evidence may be correct, and it may not…

Policy makers & clinicians will not act on anecdotal evidence because it may be harmful on closer inspection

Statistical studies provide a way to disentangle fact from fiction

Comparing Two Groups of People

Many studies focus on comparing two groups – disease/healthy, treatment/placebo

Some differences are expected when comparing two groups of people

- If you go talk to 10 people in the street, the number of males and females will vary
- Even if males and females were 50:50, chances are you will not meet exactly 5 males and 5 females

- This is called sampling variation or “chance variation”

- If we repeated our talking to 10 people in two different locations, is any male:female ratio difference due to chance or real?
- We uses statistics to ask if the differences between groups is more than you would expect by chance

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

Statistical P-values

Is the difference between our two groups more than we would expect by chance?

- The main metric for determining the statistical evidence for a difference between groups is called a P-value
- Can interpret a P-value as:

> ***What is the probability of obtaining my observed results if there is really no difference between the two groups?***

- A P-value of 0.05 means there is a 5% chance of observing a difference as extreme between the two groups by chance
- 1 in 20 studies will give a P-value of 0.05 by chance!
- Want to see “small” P-values to conclude the two groups are different
  - What is small? $P<0.01$? $P<0.001$?

- P-values get reported in scientific literature, but never in the media!

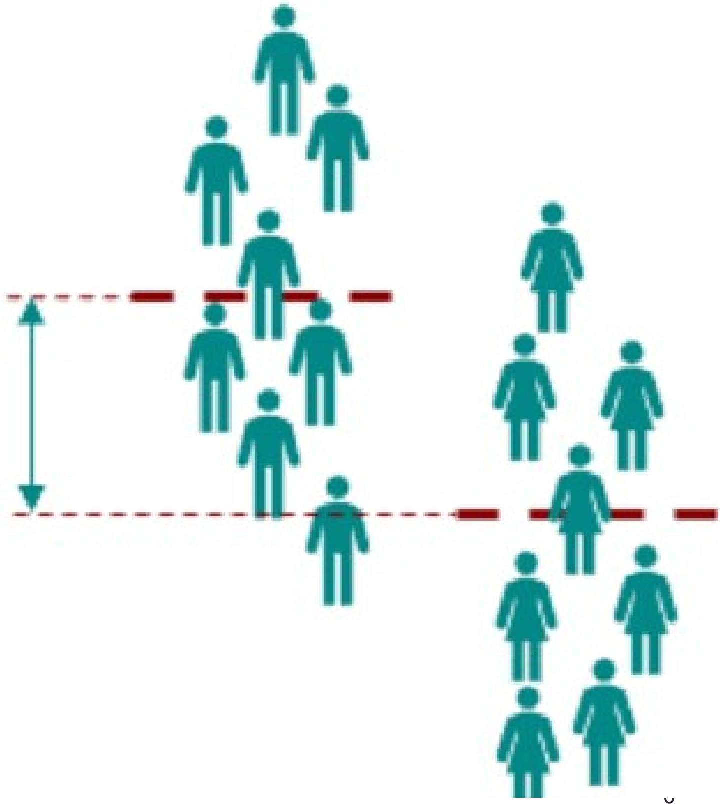
Effect Size

Putting significant effects into context

There is a difference between being statistically significant and biologically relevant

- A small P-value only means the difference between groups is non-random.
- It does not tell you how big that difference is

- Effect size for risk of disease:
    - 1 → the risk is the same between the two groups
    - < 1 → there is a protective effect (i.e. lowers risk of disease)
    - > 1 → the exposure is increasing the risk of disease
- Effect size for comparing amounts/time/…
    - 0 → the amount is the same between the two groups
    - > 0 → the amount is greater in the disease/treatment group
    - < 0 → the amount is lesser in the disease/treatment group



Effect Size

Not all significant effects are meaningful

With a large enough study, even really small effects on disease can become significant

- Is an exposure that gives a 1.01x risk of disease worth legislating against?

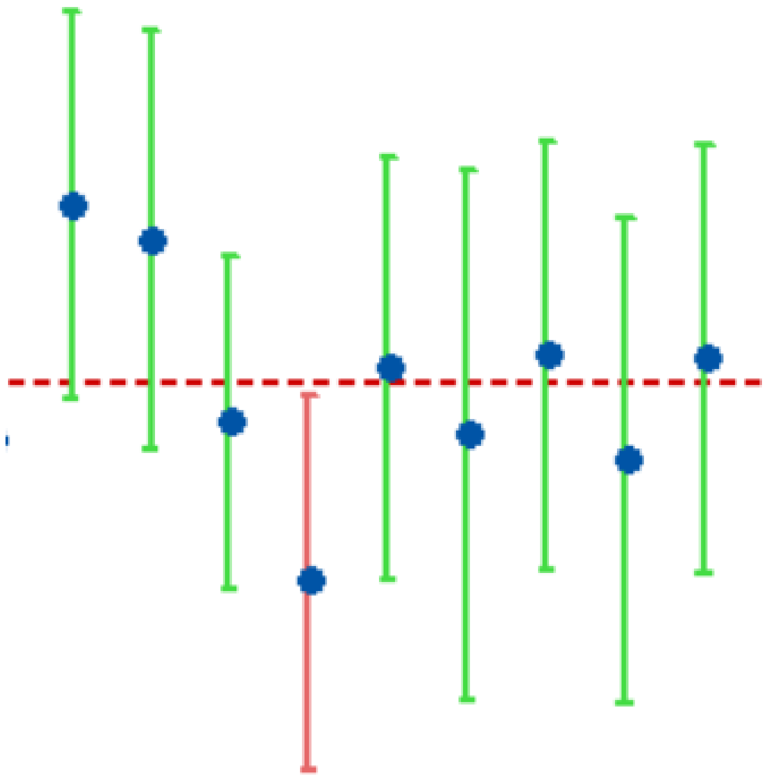
- Interpretation of effects also depends on disease frequency
- E.g. compare an exposure that increases the risk of disease 2x for a disease
    - that happens at a rate of 1:2,000,000
    - that happens at a rate of 1:100

Confidence Intervals

Establishing the precision of the effect estimate

Effect sizes are estimated with error

- Confidence intervals tell us the range of values consistent with our observed sample
- Generally presented as 95% confidence intervals
    - “We are 95% sure the effect of this exposure on the risk of disease is between 2 and 15”
- Confidence intervals reflect significance of the result
    - A non-significant risk estimate will have a confidence interval that includes 1
    - For the same risk estimate, smaller confidence intervals reflect stronger signficance



Bias

Is everything else the same?

Do the two populations match in every other way apart from what is being tested?

- Often referred to as “matched” groups
  - Age
  - Male/Female ratio
  - Urban/rural
  - …

- Often very difficult to assess as you do not know what has not been considered

Replication

So good we did it twice!

Replication is the gold standard for any scientific study

- Can we get the same result in an independent sample of people?

- Prevents conclusions being made due to an unknown bias in a single population

Rugby and MND – Lets look at the evidence

Combining everything we have covered in this talk!

The full results are not presented in the news media articles

- Need to find the original study
- Authors of the study will be happy to provide you a copy

Study finds ex-rugby internationals' risk of neurodegenerative diseases is between two and 15 times higher than the public's

Posted Wed 5 Oct 2022 at 8:30am

Rugby and MND – Lets look at the evidence

Risk of neurodegenerative disease being primary cause of death

Table 2 Primary cause of death among former international rugby players and a matched population comparison group

Primary cause of death	Former international rugby players (n=412)	Matched population comparison group (n=1236)	HR (95% CI)	P value*
Any cause†	121 (29.4%)	381 (30.8%)	0.86 (0.68 to 1.08)	0.19
Cardiovascular disease	54 (13.1%)	155 (12.5%)	0.89 (0.63 to 1.28)	0.534
Cancer	30 (7.3%)	108 (8.7%)	0.77 (0.50 to 1.19)	0.242
Neurodegenerative disease	11 (2.7%)	18 (1.5%)	2.43 (0.92 to 6.42)	0.073
Respiratory disease	11 (2.7%)	43 (3.5%)	0.62 (0.28 to 1.39)	0.247

*Cox proportional hazards regression.
†All-cause mortality analysis did not fulfil the proportional hazards assumption and showed time-dependent variability.

- 11/412 (2.7%) of former international rugby players
- 18/1236 (1.5%) of matched population comparison group
- Risk: 2.43 (95% CI 0.92 to 6.42)
- P = 0.073

Rugby and MND – Lets look at the evidence

A secondary look...

Considering both primary and secondary cause of death

- Risk 2.60,
- 95% CI 1.44 to 4.70
- P = 0.002

Table 4 Neurodegenerative disease subtypes among former international rugby players compared with a matched population comparison group

Neurodegenerative disease	HR/OR (95% CI)	P value
Dementia (NOS)	2.17 (1.26 to 3.72)	0.005*
Parkinson's disease	3.04 (1.51 to 6.10)	0.002*
MND/ALS‡	15.17 (2.10 to 178.96)	0.009†

*Cox proportional hazards regression.
†Baptista-Pike mid-p odds ratio.
‡MND/ALS adjusted to accommodate zero events by transferring one general population observation from no event to event.
ALS, amyotrophic lateral sclerosis; MND, motor neuron disease; NOS, not otherwise specified.

Rugby and MND – Lets look at the evidence

Is there a bias?

It is difficult to identify potential biases, but there are some clues that are tough to spot

- No-one in the matched population comparison group (1236 people) was diagnosed with MND
- Lifetime risk of MND in the UK is about 1 in 300
- That is a bit weird…

- International rugby players show better survival rates into their 80s
- Is this a reflection of the players living longer?

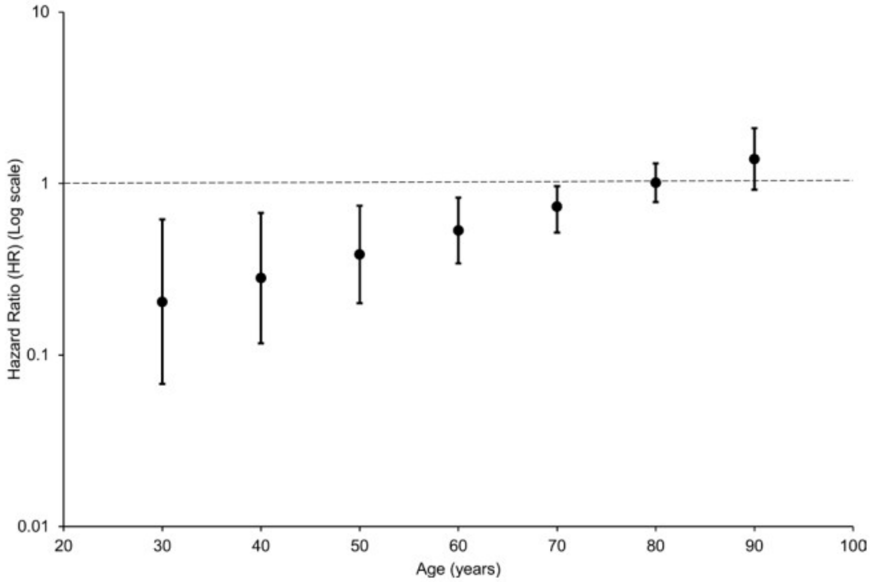


Figure 2 Time-varying hazard ratios for all-cause mortality among former international rugby players compared with a matched population

Look beyond the news media

Headlines are designed to make you click!

Look at other sources to help you interpret new research

- e.g. MND Australia Facebook group

MND Australia
October 5 ·

Research published this week from the University of Glasgow suggests there may be a 15-fold increase in MND among international rugby players, compared to the general population.

The research team compared health outcomes among 412 male, Scottish, former international rugby players and over 1,200 similar individuals from the general population, and also found an increased risk of other neurodegenerative diseases, such as dementia and Parkinson's.

While the study does suggest there is an increased risk of MND among elite level rugby players, the sample size for the study was quite small for an uncommon disease like MND, and more extensive research will need to be undertaken to further establish this link and to investigate the exact causes of the increased risk.

Read more: https://bit.ly/3V0iIlx

https://www.abc.net.au/.../study-finds-brain.../101501052

Summary

Key concepts

- **P-value** – what is the probability of getting the observed result by chance if there was no effect?
- **Effect size** – how big is the effect?   And its associated **confidence interval**.
- **Bias** – is there anything else that could be driving this result?
- **Replication** – has this result been reproduced?