Government of Western Australia
Department of Health

Perth Children's Hospital

CAHS Research Education Program

2023 Research Skills Seminar Series

Statistical Tips for Interpreting Scientific Claims

Presented by

27th October 2023

Michael Dymock

Biostatistician
Telethon Kids Institute

TELETHON KIDS INSTITUTE
Discover. Prevent. Cure.

Perth Children's Hospital Foundation

Acknowledgement of Country

The Child and Adolescent Health Service acknowledge Aboriginal people of the many traditional lands and language groups of Western Australia.
We acknowledge the wisdom of Aboriginal Elders both past and present and pay respect to Aboriginal communities of today.

- Over 20 topics across the research process
  - 1h overview
  - Handouts are provided
- Recorded and uploaded
- Feedback
  - Back of handout
  - Emailed link
- Please hold questions to the end
  - Use provided microphone

Twenty tips for interpreting scientific claims

In 2013 Sutherland, Spiegelhalter & Burgman published a list to

"help non-scientists interrogate advisers and grasp the limitations of evidence."

The 20 Tips

1. Differences & Chance Cause Variation
2. No Measurement is Exact
3. Bias is Rife
4. Bigger is Usually Better for Sample Size
5. Correlation does not Imply Causation
6. Regression to the Mean can Mislead
7. Extrapolating Beyond the Data is Risky
8. Beware the Base-Rate Fallacy
9. Controls are Important
10. Randomisation Minimises Bias
11. Seek Replication
12. Scientists are Human
13. Significance is Significant
14. Separate No Effect from Non-Significance
15. Effect Size Matters
16. Study Relevance Limits Generalisation
17. Feelings Influence Risk Perception
18. Dependencies Change the Risks
19. Data can be Dredged, or Cherry Picked
20. Extreme Measurements may Mislead

The 20 Tips

1. **Differences & Chance Cause Variation**
2. **No Measurement is Exact**
3. **Bias is Rife**
4. **Bigger is Usually Better for Sample Size**
5. **Correlation does not Imply Causation**
6. **Regression to the Mean can Mislead**
7. **Extrapolating Beyond the Data is Risky**
8. **Beware the Base-Rate Fallacy**
9. **Controls are Important**
10. **Randomisation Minimises Bias**
11. **Seek Replication**
12. **Scientists are Human**
13. **Significance is Significant**
14. **Separate No Effect from Non-Significance**
15. Effect Size Matters
16. Study Relevance Limits Generalisation
17. **Feelings Influence Risk Perception**
18. Dependencies Change the Risks
19. Data can be Dredged, or Cherry Picked
20. Extreme Measurements may Mislead

Differences & chance cause variation (1)

Perth Children's Hospital Foundation

What caused the variation?

- Suppose that you are comparing two treatments and you observe a *difference*
  - Is there *truly* a difference or did it occur by chance?
- Don't forget that rare events **DO** occur!

"The main challenge of research is teasing apart the importance of the process of interest from the innumerable other sources of variation."

What can we do about it?

Observational vs Randomised
Inclusion/exclusion criteria
Control group

Frequentist analysis
State hypotheses
Compute test statistic / p-value

Statistical Toolkit

Study Design

Hypothesis testing

Posterior distributions

Independent replication

Pillar of science
Controls for hidden variation
Builds evidence base

Bayesian analysis
Probability distribution of parameter
Conveys uncertainty

No measurement is exact (2)

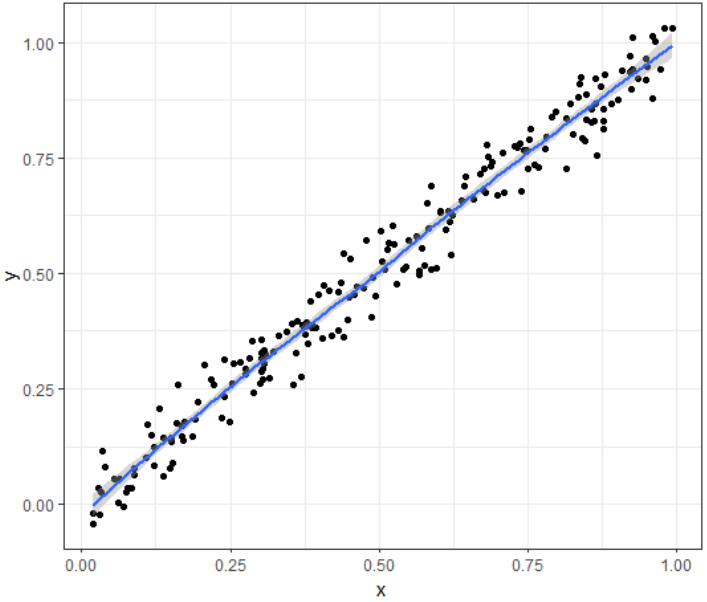
Perth Children's Hospital Foundation

No measurement is exact

All (health) measurements have some error

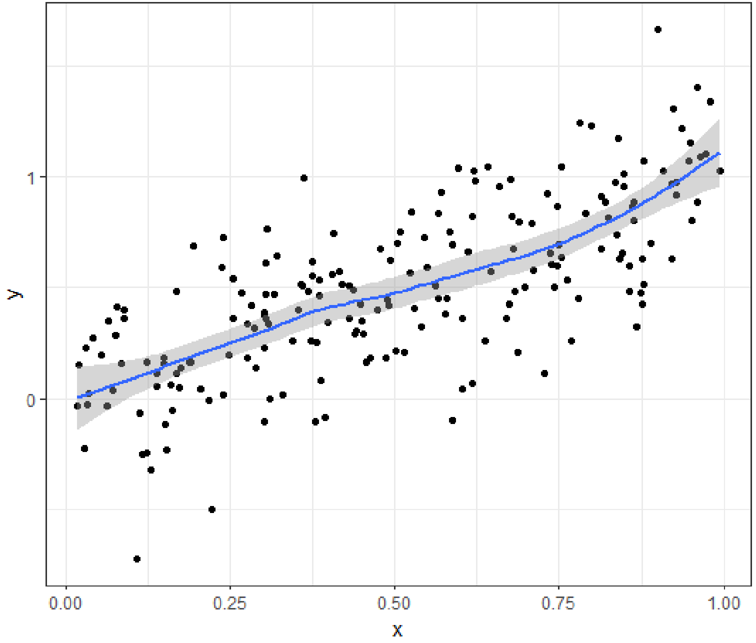
- equipment and/or operator error
- time fluctuations (daily, seasonal)

If the measurement error is large relative to the effect size then the precision will be low.



No measurement is exact

- Standardise how measurements are obtained
- Always present an estimate of the effect size (magnitude & direction) with associated precision (often a 95% confidence interval)



Bias is rife (3)

Bias is rife **within** studies

- Bias can exist due to a systematic error in the
  - Design
  - Recruitment
  - Data Collection
  - Analysis
- Results in an incorrect estimation of the true effect of the exposure/intervention on the outcome

Read beyond the abstract - strengths & weaknesses of the study are in the methods and discussion sections

Bias is rife **within** studies

- Generally, we assign more credibility to results from a study that selects participants based on an appropriate **sampling scheme** rather than a study based on observational data.
- Consider these sources of potential bias:
  - Selection
  - Recall
  - Survival
  - Study deviations

Bias is rife **across** studies

- **Study Publication Bias:**

Studies are published or not depending on their results and leads to inflated or exaggerated effect sizes in early meta-analyses.

- **Time Lag Bias** *(or "pipeline bias"):*

Non-significant research results can take longer to achieve publication

- **Outcome Reporting Bias:**

Study outcomes that are statistically significant have a higher chance of being fully reported and leads to over-estimation of the effect size

Bigger is usually better for sample size (4)

Bigger is Usually Better for Sample Size

Average efficacy can be more reliably and accurately estimated from a study with hundreds of participants than from a study with only a few participants.

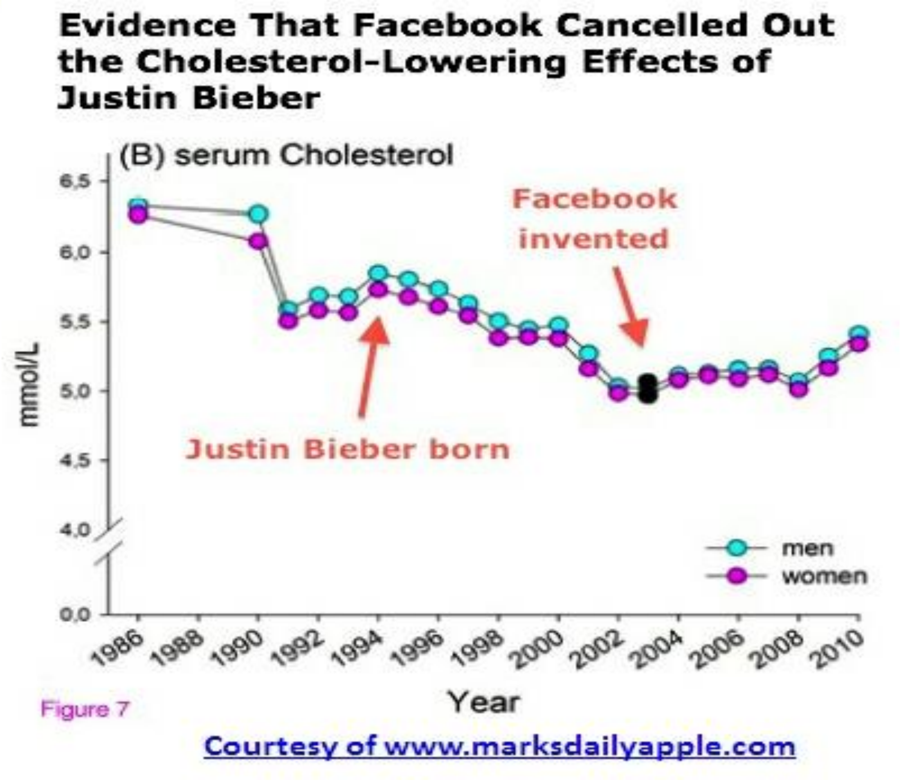
Reduces chance of **Type I Error**

Ensure that **subgroup** analyses are adequately powered (i.e., able to detect any group differences)

Correlation does not imply causation (5)

Perth Children's Hospital Foundation

Correlation does not imply causation



Consider whether the detected association may be due to a third unmeasured/ unknown confounding (*lurking*) factor or whether it may simply be incidental

Correlation does not imply causation

- Review existing literature for common confounders and account for these in the design stage if possible
- Check for imbalance in subgroups or adjustment for these in the analysis
- Check whether the association is biologically plausible
- **Draw a causal diagram!**

Bradford Hill Criteria

Generalisability (external validity)

"At its best a trial shows what can be accomplished with a medicine under careful observation and certain restricted conditions. The same results will not invariably or necessarily be observed when the medicine passes into general use."

Austin Bradford Hill, 1984

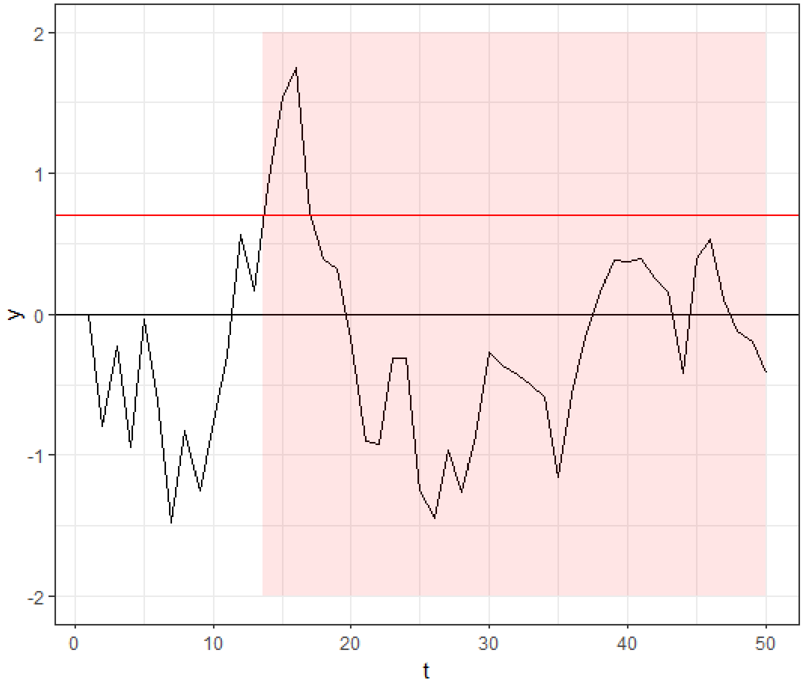
Regression to the mean can mislead (6)

Perth Children's Hospital Foundation

Regression to the mean can mislead

Commonly in clinical trials individuals are recruited based on their baseline assessment
(e.g., SBP > 160mmHg,
CD4 count < 350 cells/mm^3)

Patients often present when their symptoms have worsened, some only temporarily, so over time their average score may fall back to the true value.



Regression to the mean can mislead

Studies over time should always include:

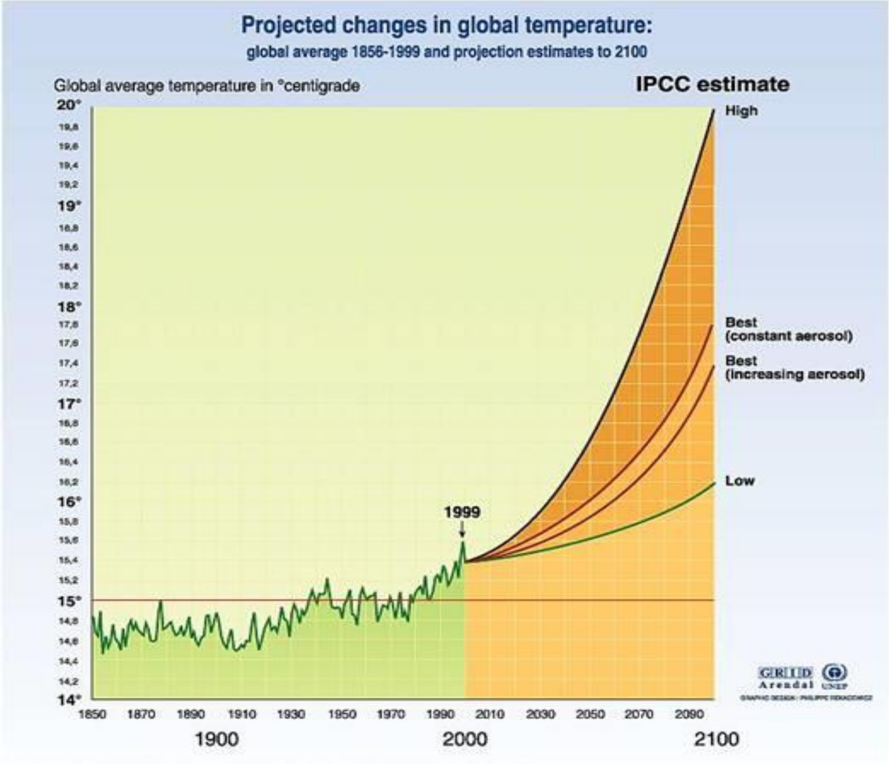
- **comparator/control group**
- **and record baseline measurements**

There is a tendency to jump to conclusions when there has been a cluster of rare events, however, random processes tend to return to their base rate over time (if left untouched)

Extrapolating beyond the data is risky (7)

Perth Children's Hospital Foundation

Extrapolating beyond the data is risky



Source : Temperatures 1856 - 1999: Climatic Research Unit, University at East Anglia, Norwich UK. Projections: IPCC report 95.

"Patterns found within a given range do not necessarily apply outside that range.

Thus, it is very difficult to predict the response of ecological systems to climate change, when the rate of change is faster than has been [previously] experienced."

Extrapolating beyond the data is risky

- Generally, we assign more credibility to predictions within the range of the data.

- When forecasts are necessary, they should be
  - robust to underlying methods
  - model a range of assumptions
  - and be presented with uncertainty intervals

Beware of the base-rate fallacy (8)

Beware of the base-rate fallacy

“The ability of an imperfect test to identify a condition depends upon the likelihood of that condition occurring (the base rate).”

Don’t be overly influenced by high sensitivity or specificity rates (true test positives and negatives)

Suppose you test **positive** for a disease with 1/1000 prevalence (test has 99% *sensitivity* and 98% *specificity*)

Are you truly positive?

Beware of the base-rate fallacy

D = true infection
T = positive test result

$P(D) = 0.001$ (prevalence)
$P(T|D) = 0.99$ (sensitivity)
$P(\bar{T}|\bar{D}) = 0.98$ (specificity)

$$P(D|T) = \frac{P(D \cap T)}{P(T)} = \frac{P(D)P(T|D)}{P(D)P(T|D) + P(\bar{D})P(T|\bar{D})} = \frac{0.001 \times 0.99}{0.001 \times 0.99 + (1-0.001) \times (1-0.99)} = 0.047$$

4.7%

Controls are important (9)

Controls are important

For all the reasons previously covered

Randomisation minimises bias (10)

Randomisation minimises bias

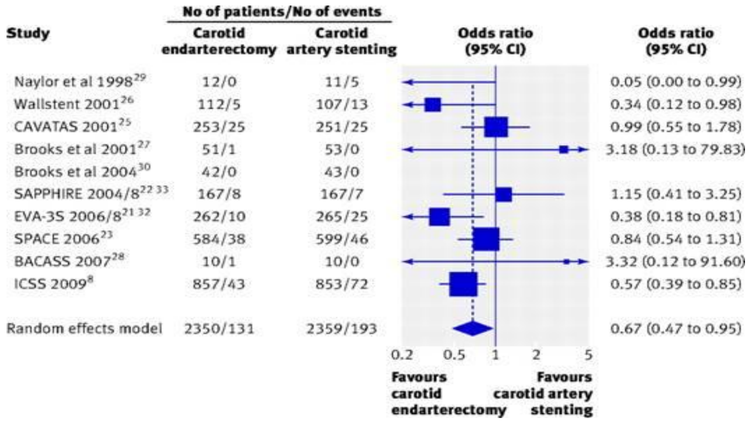
- Ideally individuals (or units) should be randomised to intervention to minimise systematic differences between the groups due to factors other than the intervention

- The randomisation process should be checked for balance at baseline across treatment groups for confounding variables.

Seek replication (11)

Seek replication

"Results consistent across many studies, replicated on independent populations, are more likely to be solid."



"The results of several such experiments may be combined in a meta-analysis to provide an overarching view of the topic with potentially much greater statistical power than any of the individual studies."

Seek replication

Look carefully at the:

- study design
- outcomes
- inclusion/exclusion criteria
- and statistical methods to determine if the studies should be compared

Scientists are human (12)

Scientists are human

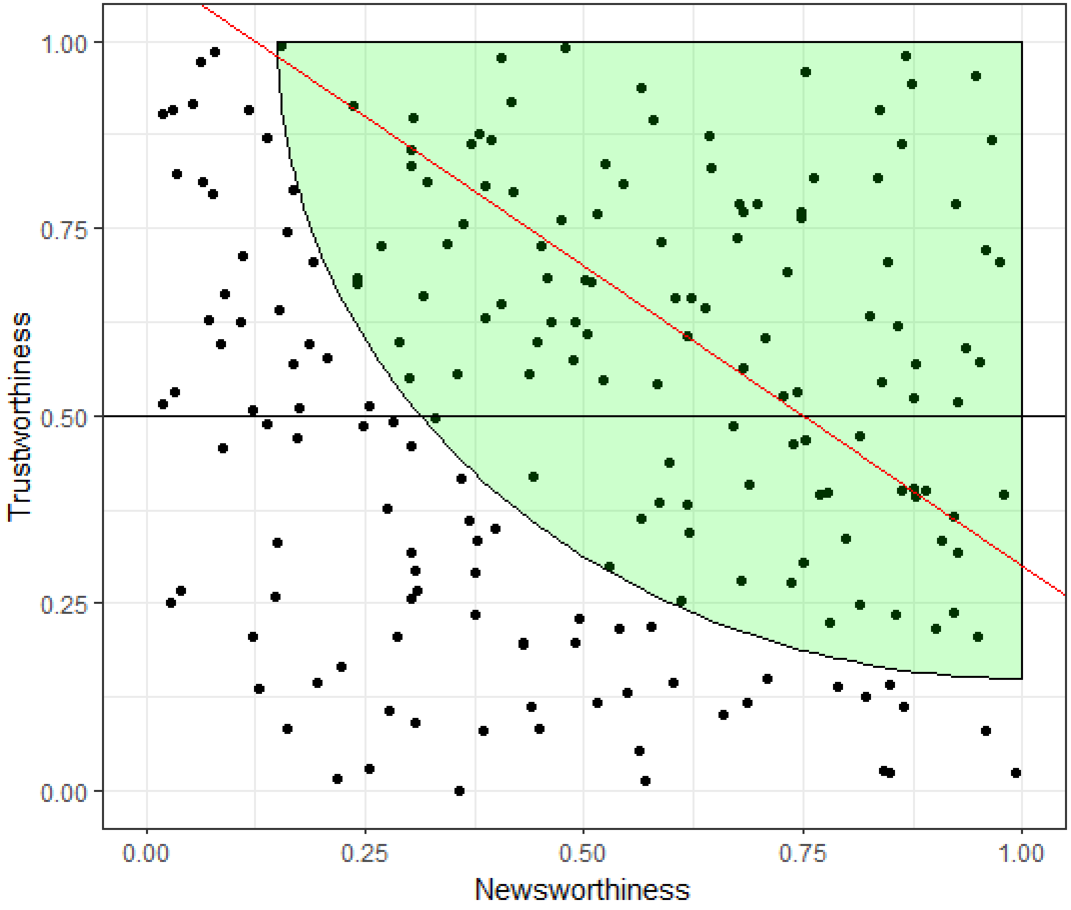
“Peer review is not infallible: journal editors might favour positive findings and newsworthiness.”

Researchers may have a vested interest in promoting their research or be prone to exaggeration.

Statistical Tools

Reporting Guidelines:

CONSORT, TREND, STROBE, REMARK, STREGA, PRISMA



Trustworthiness
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
Newsworthiness

Significance is significant (13)

Significance is significant

1. P-values can indicate how **incompatible** the data are with a specified statistical model
2. P-values **do not** measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone
3. Scientific conclusions, business or policy decisions **should not be** based only on whether a p-value passes a specific threshold (e.g., p-value < 0.05)
4. Proper inference requires **full reporting and transparency**
5. A p-value or statistical significance, **does not** measure the size of an effect or the importance of a result
6. By itself, a p-value **does not** provide a good measure of evidence regarding a model or hypothesis

Separate no effect from non-significance (14)

Perth Children's Hospital Foundation

Separate no effect from non-significance

- Proof by **contradiction**
  - Suggest Theory X
  - Find a contradiction (or counter example) to Theory X

*“**The lack of a statistically significant result** (say a P-value > 0.05) does not mean that there was no underlying effect: it means that **no effect was detected**.”*

- With a hypothesis test, we aim to assess evidence that **counters** the claim of the null hypothesis, thus **supporting** the alternative hypothesis
- **BUT** the failure to find counter evidence **does not** prove the null hypothesis

Feelings influence risk perception (17)

Perth Children's Hospital Foundation

Feelings influence risk perception

*"Broadly, risk can be thought of as the **likelihood of an event occurring** in some time frame, **multiplied by the consequences should the event occur**. People's risk perception is influenced disproportionately by many things, including the **rarity** of the event, how much **control they believe they have**, the **adverseness of the outcomes**, and whether the risk is **voluntarily or not**."*

Upcoming Sessions

17 Nov Ethics Processes for Clinical Research in WA

Dr Natalie Giles, Manager Ethics & Compliance CAHS

21 Nov **WORKSHOP**

Navigating Research Ethics and Governance in WA

Dr Natalie Giles and the CAHS Ethics and Governance Team

Register → **researcheducationprogram.eventbrite.com.au**

We love feedback

A survey is included in the back of your handout, or complete online

https://tinyurl.com/surveyStatTips

ResearchEducationProgram@health.wa.gov.au cahs.health.wa.gov.au/ResearchEducationProgram

© 2023 CAHS Research Education Program

Child and Adolescent Health Service Department of Research
Department of Health, Government of Western Australia

✉ ResearchEducationProgram@health.wa.gov.au
cahs.health.wa.gov.au/ResearchEducationProgram