



The good, the bad and the ugly: What makes research 'valid'?

Dr Christina Marel^{1,2}

Dr Lexine Stapinski^{1,2}

¹Centre for Research Excellence in Mental Health and Substance Use

²National Drug and Alcohol Research Centre, University of New South
Wales



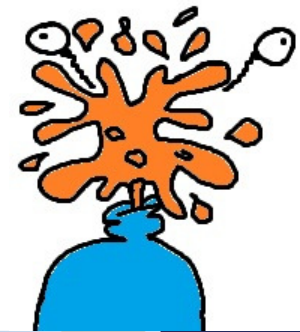
NHMRC CENTRE OF RESEARCH EXCELLENCE
in MENTAL HEALTH and SUBSTANCE USE



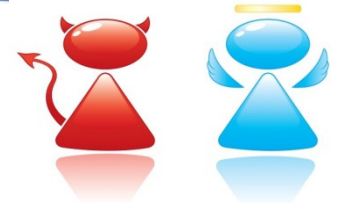
Critiquing Research: Overview



- ❖ The 'good' versus the 'bad'
- ❖ Essential research ingredients
 - Where to look?
- ❖ Evaluating treatments – example
- ❖ Stairway to research strength
- ❖ Hidden traps of bad research: what to look out for
- ❖ The 'ugly' – a brief comment of the use of statistics



The 'good' versus the 'bad'



- ❖ 'Good' research is cautious about drawing conclusions, acknowledges uncertainties and avoids exaggerated claims.
- ❖ It demands multiple types of evidence to reach a conclusion.
- ❖ It does not assume correlation = causation (one thing causes another).



The 'good' versus the 'bad'



- ❖ 'Bad' research may use accurate data, but manipulate and misrepresent the information to support a particular conclusion.

-E.g. research questions may be defined and analysis selected to reach a desired outcome



- ❖ Alternative perspectives and data may be ignored, distorted or ridiculed
- ❖ Conclusions might be based on faulty logic, unreliable data or insufficient evidence.
- ❖ The limitations of the analysis may be ignored and the implications of results exaggerated.



What to look for: Essential ingredients



❦ Who?

- Who conducted the research? Who was the research funded by?

❦ What?

- What is being studied? Is there a clear research question? Is the research question/s carefully defined?

❦ Why?

- Why is the research being conducted? Is the research question justified by the evidence provided / background information / literature?

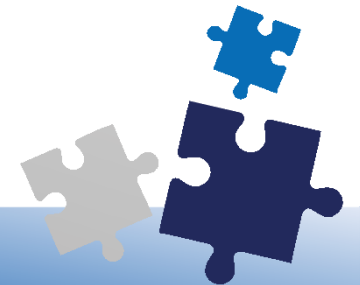


What to look for: Essential ingredients



How?

- Are the methods justified? Is there a discussion about why particular methods of analyses or data are used or not used? Is the sample size adequate?
- Does the research fairly present all perspectives or does it focus on promoting one side only? Are important findings in the field highlighted or pointed out?
- Are the study methods and analysis clearly communicated so they are understandable and replicable by other researchers?
- Are there clear logical links between research results, conclusions and implications?
- Are the limitations of the study discussed?



What to look for: Essential ingredients



❦ Where is it published?

- Research publications may include: journal articles, book chapters, technical/government reports
- Has the publication been peer-reviewed?
 - This refers to the critical assessment by qualified experts (preferably not known by the authors) which enhances research quality.
- For journals: Is the journal well-respected in the field? What is the impact factor?

Impact Factor? A calculation of the number of number of times papers in the journal have been cited in the two preceding years



What to look for: Essential ingredients



❦ Referencing

- Good research provides a comprehensive overview of an issue and discusses the context.
- Not all sources are equal. Better sources are published by reputable presses (e.g. University Presses), are refereed (peer reviewed), or are by reputable organizations.
- Wikipedia, ask.com, yahoo, nine.msn and other similar web sites are typically not appropriate final research sources.

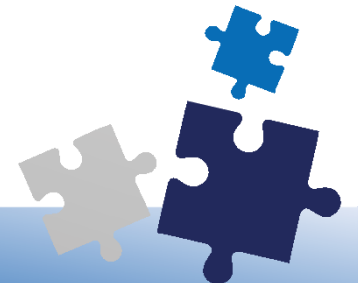


Where to look?



✦ Activity 1 (10 mins):

- Mills KL, Lynskey M, Teesson M, Ross J, Darke S. (2005) Post-traumatic stress disorder among people with heroin dependence in the Australian treatment outcome study (ATOS): prevalence and correlates. *Drug and Alcohol Dependence*, 77(3): 243-249
- ✦ Read through the article (handout) and assess:
 - Who?
 - What?
 - Why?
 - How?
 - Where?
 - Evaluation of this paper? Any concerns?



Where to look?



❖ Who?

- Researchers from NDARC, UNSW, Washington University

❖ Where is it published?

- Drug and Alcohol Dependence = good journal. Impact factor 3.141; ranked 4/16 (Substance Abuse); 45/135 (Psychiatry)

- Peer reviewed

❖ Journal rankings can be found on the journal home page, or via journal citation reports

❖ Note: Due to copyright, the annotated paper from the talk cannot be published in these slides, however some of you might be able to access it depending on your library access.

- ❖ Mills KL, Lynskey M, Teesson M, Ross J, Darke S. 2005 Post-traumatic stress disorder among people with heroin dependence in the Australian treatment outcome study (ATOS): prevalence and correlates, Drug and Alcohol Dependence, 77(3): 243-249



Where to look?



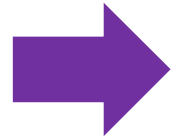
- Who funded the research?
 - Look at acknowledgements
 - National Health and Medical Research Council (NHMRC)
- 
- Commonwealth Department of Health and Ageing

✦ Think about:

- Are there any conflicts of interest?
- Eg. Was the research funded by a pharmaceuticals company wanting to evaluate their own medications?
- Are the researchers only interested in pursuing one theory / research agenda? Look to see whether methods justified.



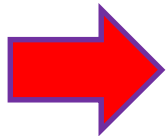
What is being studied and why?



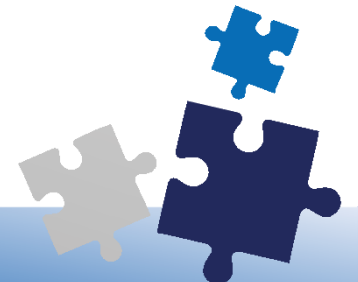
Look at introduction / background of paper



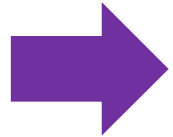
- Is there a clear research question?
- Is there enough evidence to support the reasons for the study?
- Is the research question justified by the evidence provided / background information / literature?



Red flag: If you have made it to the methods section of a paper and can't figure out what is being studied or why – the aims / evidence and justification may not have been clearly articulated – a red flag.



How is it being studied?



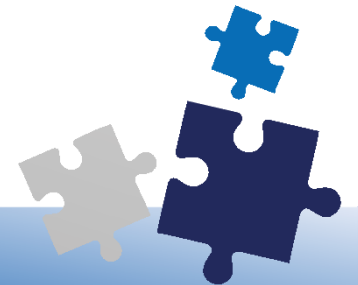
- Look at the methods sections
- Are the methods justified?
 - Are the study methods and analysis clearly communicated so they are understandable and replicable by other researchers?
 - Have any stakeholders been consulted?
 - ⌘ If so, details of this consultation will be found in the methods section



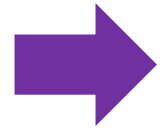
How is it being studied?



- ➡ Look at the analysis section within the methods.
- Is there a discussion about why particular methods of analyses (statistics / qualitative analyses) or data are used or not used?
 - Are the study methods and analysis clearly communicated so they are understandable and replicable by other researchers?



How is it being studied?



- Look at the discussion section
- Does the research fairly present all perspectives or does it focus on promoting one side only?
 - Are there clear logical links between research results, conclusions and implications?
 - Are the limitations of the study discussed?



Specific issues when evaluating treatments



- ❖ Who was excluded from the study? Are the results generalisable to the real world?
 - E.g., if participants with multiple substance use problems are excluded from a study, the sample may be less severe than typically seen in practice
- ❖ What is the treatment being compared to?
 - An appropriate comparison group increases confidence that it is the treatment itself that is effective, rather than other factors (e.g. placebo effect, effect of assessment, natural course of the disorder)
- ❖ When comparing two treatments –important to establish that the two groups are not significantly different at baseline (i.e. *before* treatment begins)
 - E.g., are there more severe symptoms in one group, more males vs females



Specific issues when evaluating treatments



❖ Stronger conclusions can be drawn from random allocation compared to observational design

- In a randomised control trial, participants are *randomly allocated* to one treatment or another
- Observational design makes use of existing differences in treatment received – this may mean there are differences between the groups.

E.g., more severe patients more likely to choose a particular treatment



❖ Are the assessments appropriate?

- Are validated assessment tools used? (Look for references)
- Ideally, those responsible for doing the assessments should be “blind” to the treatment received.



Specific issues when evaluating treatments



- Are participants followed up after treatment?
 - Results may look good immediately after treatment, but important to check whether this is sustained
 - How is missing data dealt with? Participants who did better are more likely to complete follow-up assessments, this may paint an inaccurate picture of effectiveness of the treatment



Evaluating Treatments: Example



Example paper: Bucknam, W. 2006 Supression of symptoms of alcohol dependence and craving using high dose baclofen, Alcohol & Alcoholism, 42(2): 158-160

Note. Due to copy-right issues the annotated paper cannot be published in these slides, however the paper itself can be freely downloaded here:

<http://alcalc.oxfordjournals.org/content/42/2/158.full>

Abstract

Treatment with baclofen was associated with a reduction in cravings and preoccupation for alcohol

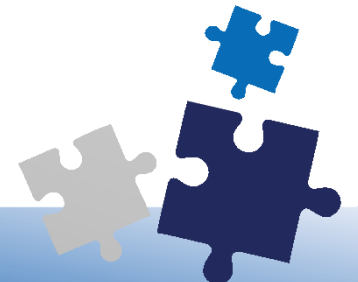


Evaluating Treatments



⚙️ Activity 2 (10 mins) :

- Bucknam, W. (2006) Suppression of symptoms of alcohol dependence and craving using high dose baclofen. Alcohol & Alcoholism, 42(2): 158-160
- ⚙️ Read through the article (handout) and assess:
 - Who?
 - What?
 - Why?
 - How?
 - Where?
 - Evaluation of this paper? Any “red flags”?



Evaluating Treatments: Example

Red Flag in
INTRODUCTION:
Justification for the
study

Page 158: “The data presented thus far was previously reported in a letter to the editor of Journal of the American Medical Association (Ameisen, 2005a) and in a case study authored by Olivier Ameisen, MD, who used himself as the subject of study (Ameisen, 2005b). Dr. Ameisen had previously tried recommended dosages of disulfiram, oral naltrexone, acamprosate, and topiramate and had had extended periods of abstinence utilizing CBT and extensive involvement in alcoholics anonymous (AA). He nevertheless persisted to have alcohol cravings and anxiety symptoms, which had predated his alcohol dependence, despite trials of buspirone, selective serotonin reuptake inhibitors, valproate and carbamazepine. Hypothesizing that the dose-dependent suppression of alcohol consumption (3 mg/kg body wt) in animals could be transposed in humans, he subjected himself to a trial. He self-prescribed baclofen up to 270 mg/day (3.6 mg/kg body wt) during the first 37 days and experienced, for the first time in his alcoholic life, the absence of craving for alcohol. Indeed, he reported a state of complete and persistent indifference to alcohol, along with substantial reduction of anxiety, for a duration of 9 months at the time of his report.”

Taken from: Bucknam, W. 2006 Supression of symptoms of alcohol dependence and craving using high dose baclofen, Alcohol & Alcoholism, 42(2): 158-160

Evaluating Treatments: Example



The research questions seek to examine the benefits of baclofen in terms of reducing craving, depression and anxiety

Page 159: "Scales to evaluate craving and laboratory parameters were not used...."

Dr. Ameisen's experience, use of a selective serotonin reuptake inhibitor (SSRI; paroxetine) did appear to be necessary as baclofen by itself did not satisfactorily reduce Mr. A's anxiety or depression."

Red Flag in
METHOD:

But, no standardised assessment tools are used to assess craving, anxiety and depression..
Or alcohol intake



Taken from: Bucknam, W. 2006 Supression of symptoms of alcohol dependence and craving using high dose baclofen, Alcohol & Alcoholism, 42(2): 158-160

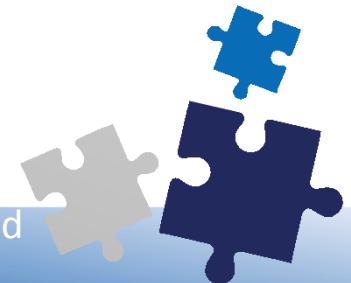


Evaluating Treatments: Example



Page 159: “Over the first month he gradually increased his dosage to 100 mg/day, taken on a t.i.d schedule, and reported a completely satisfactory response....On occasion, when stressed, he increased his dosage to 140 mg/day.”

Red Flag in
METHOD:
Can future studies
replicate based on this
description of methods?



Taken from: Bucknam, W. 2006 Supression of symptoms of alcohol dependence and craving using high dose baclofen, Alcohol & Alcoholism, 42(2): 158-160



Evaluating Treatments: Example



Page 159: "If he chose to drink his consumption was never more than 12 per week, or 3 per occasion, and his sense of euphoria from that was dulled."

Red Flag in

METHOD /RESULTS

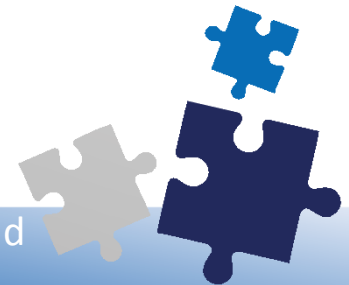
What are the actual data?

How much did the participant actually drink before vs after treatment?

How was the participant's drinking assessed? Was there a follow-up period?



Taken from: Bucknam, W. 2006 Supression of symptoms of alcohol dependence and craving using high dose baclofen, Alcohol & Alcoholism, 42(2): 158-160



Evaluating Treatments: Example



Red Flags in DISCUSSION

Can we really assume
Mr A is representative
of this population?

Are these conclusions
justified by the data?

Factual statements
without references

Page 9: "Mr. A is an individual whom I believe represents a very large number of patients who do not experience a satisfactory anti-craving response to either the current FDA-approved medications for alcohol dependence or to topirimate.

...

Placebo response is a possibility. If that is the case, however, there is no apparent explanation for why it did not appear in trials of either naltrexone or acamprosate, alone or in combination, or with topirimate. Given the nearly four decades of use of high-dose baclofen for the long-term comfort care of patients with muscular spasticity from various neurological conditions (spinal injuries, multiple sclerosis), and the absence of report of serious or irreversible adverse effect, baclofen may be a safe, effective and well-tolerated adjunct to our treatment efforts with this population. Hypotension, changes in glucose control in diabetics, sedation and changes in seizure control are potential side-effects. Randomized trials of high-dose baclofen should be conducted to test elimination of alcohol craving and its potential consequences."



Taken from: Bucknam, W. 2006 Supression of symptoms of alcohol dependence and craving using high dose baclofen, Alcohol & Alcoholism, 42(2): 158-160

Evaluating Treatments: Example



DISCUSSION
Acknowledges some
limitations...

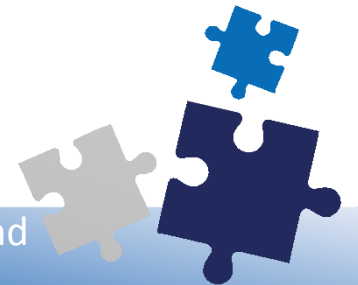
Case studies can be an
appropriate starting
point for a more
systematic
investigation.

Page 159: "Being a case study, this report is obviously limited...."

Randomized trials of high-dose baclofen should be conducted to test elimination of alcohol craving and its potential consequences."



Taken from: Bucknam, W. 2006 Supression of symptoms of alcohol dependence and craving using high dose baclofen, Alcohol & Alcoholism, 42(2): 158-160



Activity 3: Q & A



❦ Which of the following alternate research designs would provide the strongest evidence?

- a) A summary of 8 case reports on patients receiving baclofen treatment
- b) A trial where 100 participants were randomly allocated to receive either baclofen or naltrexone
(naltrexone = another medication for alcohol dependence intended to reduce craving)
- c) An observational study comparing 100 patients prescribed either baclofen or naltrexone



Stairway to Research Strength



- ❧ Larger number of participants in a study = more representative of the population, reduced error
- ❧ Even so, we need to be cautious about drawing strong conclusions from one single study
- ❧ Systematic review articles allow us to draw conclusions with greater confidence
 - They synthesise results from multiple studies
 - When conducted appropriately (e.g. independent authors, appropriate methods and reporting) they allow for more confident conclusions than one single study.
 - The Cochrane library is an excellent collection of well-conducted reviews: <http://www.thecochranelibrary.com>

Stairway to Research Strength



**Well conducted
systematic review of
multiple studies**

**Several well-conducted
studies (e.g. RCTs)**

**Well conducted study
(with adequate sample
size)**

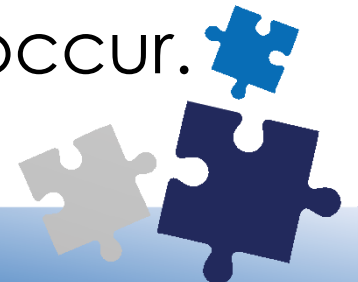
Single case study

Hidden traps of 'bad' research



1. Association doesn't prove causation: Example

- ❖ People sometimes die in hospitals.
- ❖ Research just focusing on this outcomes could confuse causation and 'prove' that hospitals 'cause' people to die.
- ❖ Causation is confused here because other factors and the wider context may not be considered.
- ❖ The wider context makes it more likely that hospitals significantly reduce death rates compared with what would otherwise occur.



Association doesn't prove causation



Scientific evidence proves that bunnies are a major causation for peace.

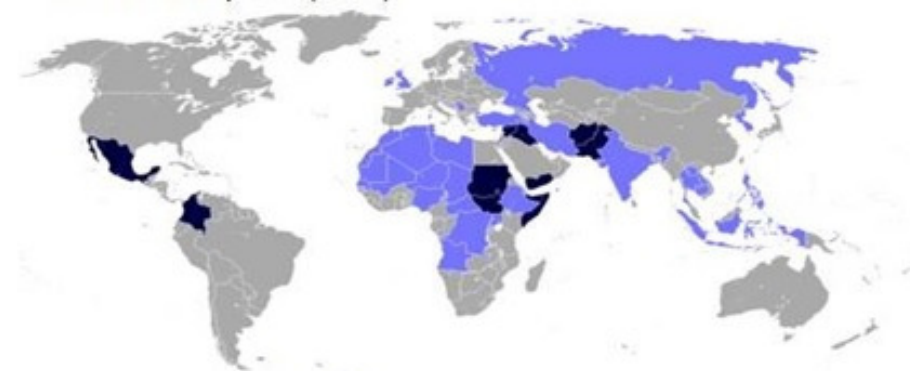
Wild bunny range

Source: Wikipedia (2012)



List of ongoing military conflicts

Source: Wikipedia (2012)



Major wars, 1,000+ deaths per year
Minor wars and conflicts, 10-1,000 deaths per year



Source: I * love science; retrieved 26 June 2014



Hidden traps of 'bad' research



2. Application of legitimate sounding statistics

'Smell of baked bread may be a health hazard' (headline from Cincinnati Enquirer; 1998)



Source: www.geoffmetcalf.com/bread.html, retrieved 5 March 2014



2. Application of legitimate-sounding statistics



✦ Sometimes research can use legitimate sounding numbers and data to push an agenda

✦ For example:

- More than 98% convicted prisoners are bread eaters
- In the 18th century, when the majority of bread was baked in the home, the average life expectancy was less than 50 years; infant mortality rates were significantly higher, and more women died in childbirth
- More than 90% of murders were committed within 24 hours of bread consumption
- Bread has been proven to be addictive. Subjects deprived of bread and given only water to eat begged for bread after only two days.
- Bread is often a “gateway” food item, leading the user to “harder” items such as butter, jam , peanut butter and even meat.

✦ Although these data may be accurate, they are being misused to present the view that bread is dangerous



Source: www.geoffmetcalf.com/bread.html, retrieved 5 March 2014



DANGER!



THE DEADLY FACTS ABOUT WATER!

FACT!

WATER CAN BE CHEMICALLY
SYNTHESIZED BY BURNING
ROCKET FUEL!!!

FACT!

OVER CONSUMPTION CAN CAUSE
EXCESSIVE SWEATING, URINATION,
AND EVEN DEATH!!!

FACT!

100%
OF ALL SERIAL KILLERS,
RAPIST AND DRUG DEALERS HAVE
ADMITTED TO DRINKING WATER!!!



FACT!

WATER ONE OF THE PRIMARY INGREDIENTS
IN HERBICIDES AND PESTICIDES!!!

FACT!

WATER IS THE LEADING
CAUSE OF DROWNING!!!

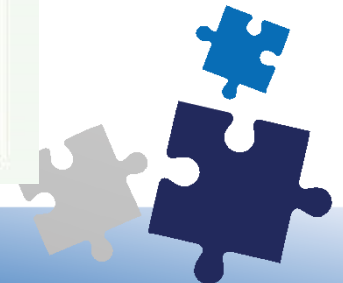
FACT!

100 PERCENT OF ALL PEOPLE
EXPOSED TO WATER WILL DIE!

© 2014 OUTREACH FILM



Source: I * love science; retrieved 26 June 2014

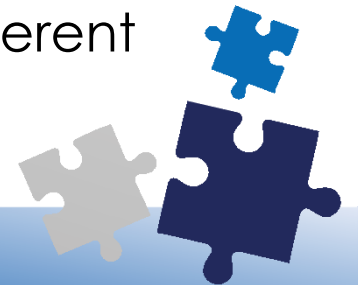


Hidden traps of 'bad' research



3. Sampling

- ✦ In the mid-1980s, Midway Airlines promoted its New York-to-Chicago service with ads in the *New York Times* and the *Wall Street Journal* claiming that "84 percent of frequent business travelers to Chicago prefer Midway Metrolink to American, United, and TWA."
- ✦ This figure seemed very surprising, considering Midway was carrying only 8 percent of the traffic between New York and Chicago at the time.
- ✦ The fine print at the bottom of the ad revealed that the survey was "conducted among Midway Metrolink passengers between New York and Chicago." This is, the only passengers eligible for inclusion in the survey were the 8 percent who were already flying Midway.
- ✦ The surprise in the results was that one in six Midway passengers apparently preferred to be flying on a different airline.



3. Sampling



- ❖ Many studies have found that alcohol use 'protects' against coronary heart disease
- ❖ However, many studies combined the classification of occasional drinkers and abstainers, and former drinkers and abstainers, potentially including more people with illness in those categories
- ❖ As they age, many people abstain or cut down on their alcohol intake for health reasons (e.g., disability, frailty, medication use etc.) Including these people as 'abstainers' in these studies might give impression that they are less healthy than light drinkers and at risk of early death
 - I.e., light drinking might be a sign of good health in older people, not a cause of it
- ❖ There may have been over-estimation of the estimates of the extent of impact that light drinking has on mortality from coronary heart disease



Source: Fillmore, et al., 2006

Mixed bag...



Association
doesn't prove
causation

Drought trigger clear as day

WHEN I was a kid we never had drought after drought.

Then we started with daylight saving. We started with a little bit, but now we have six months of the year daylight saving.

It has just become too much for the environment to cope with.

It is so logical, for six months of the year we have an extra hour each day of that hot afternoon sun.

I read somewhere that scientific studies had shown there is a lot less moisture in the atmosphere which means we get less rain.

I believe this one hour extra sun is slowly evaporating all the moisture out of everything.

Why can't the Government get the CSIRO to do studies on this, or better still, get rid of daylight savings.

They have to do something before it is too late.

— CHRIS HILL,
Albury

Pushing own
agenda?

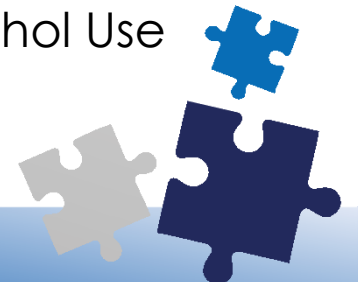
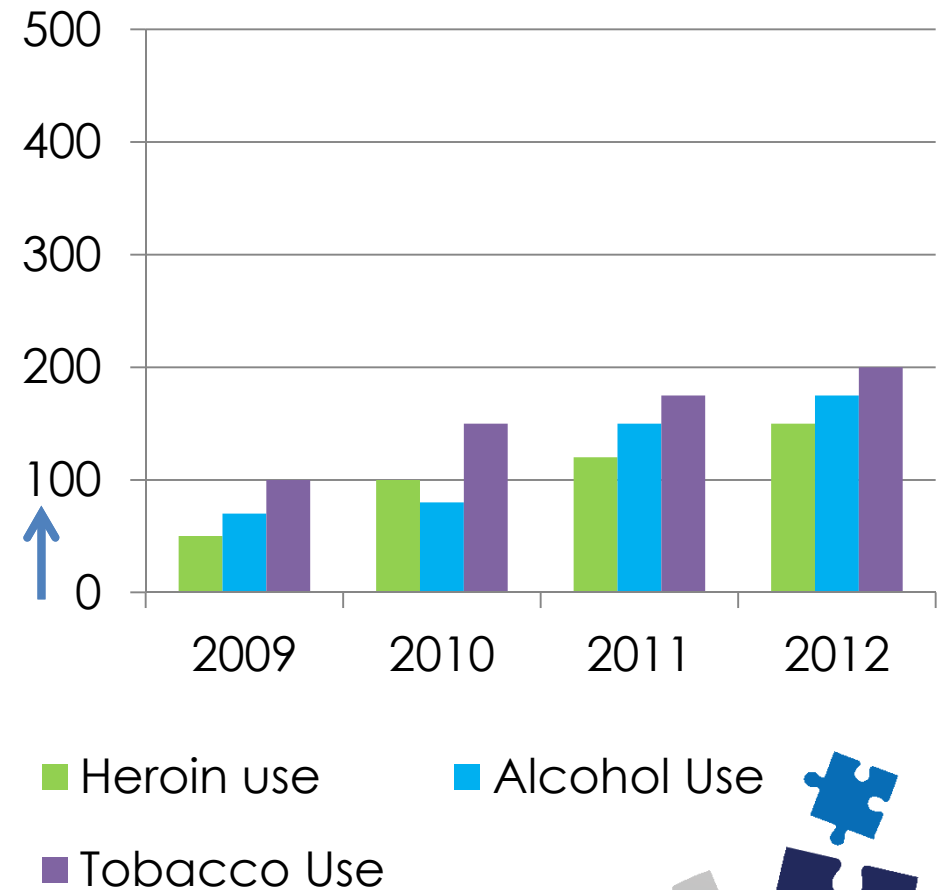
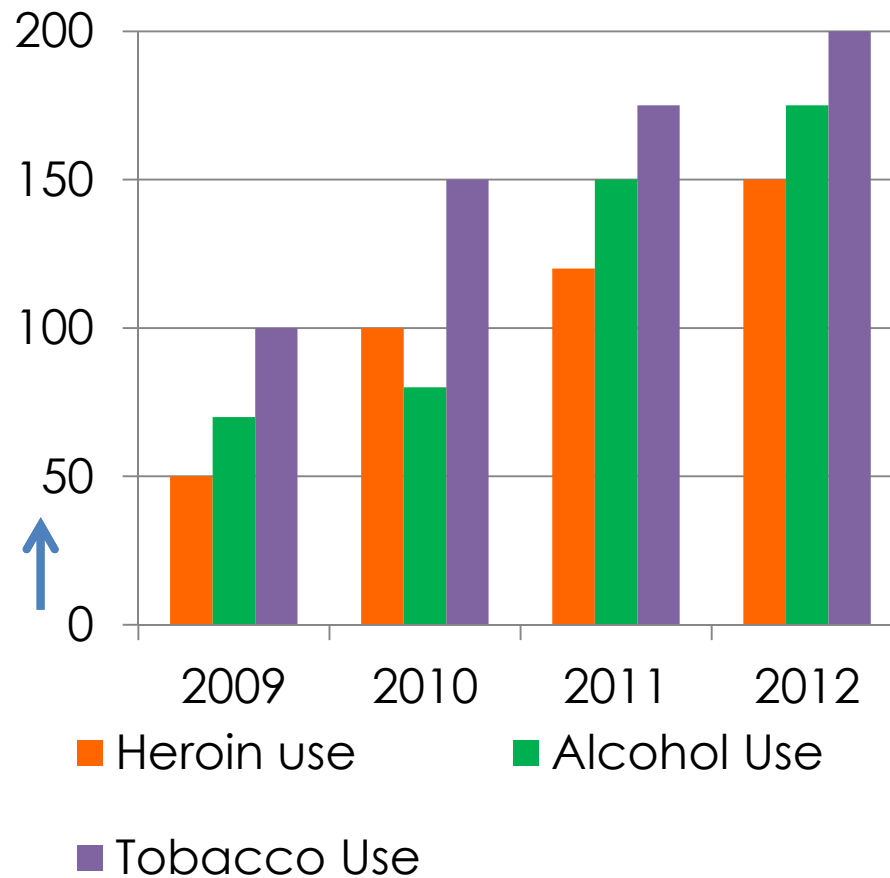


Source: I * love science; retrieved 26 June 2014

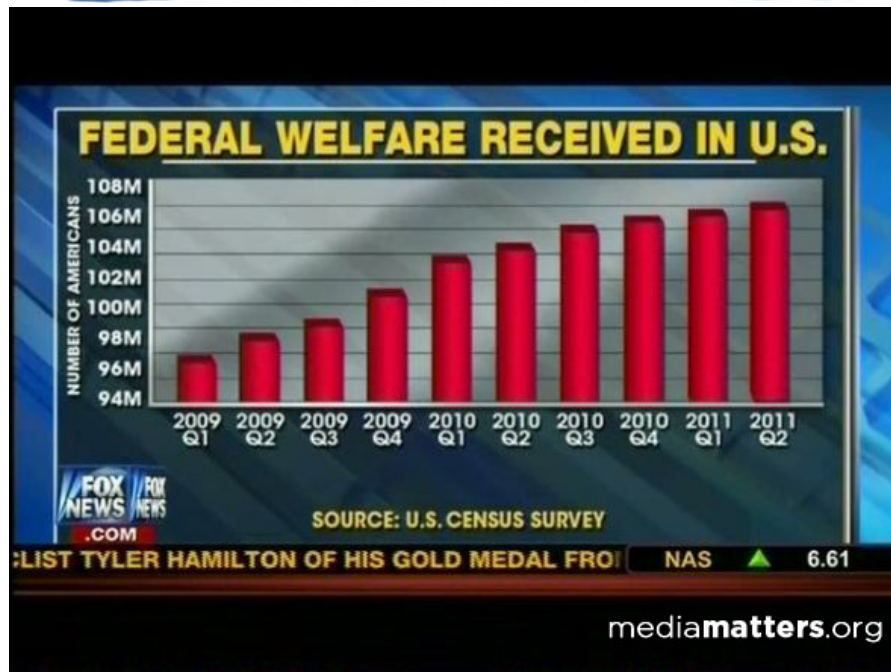
4. Presentation of data



- Altering the axis levels

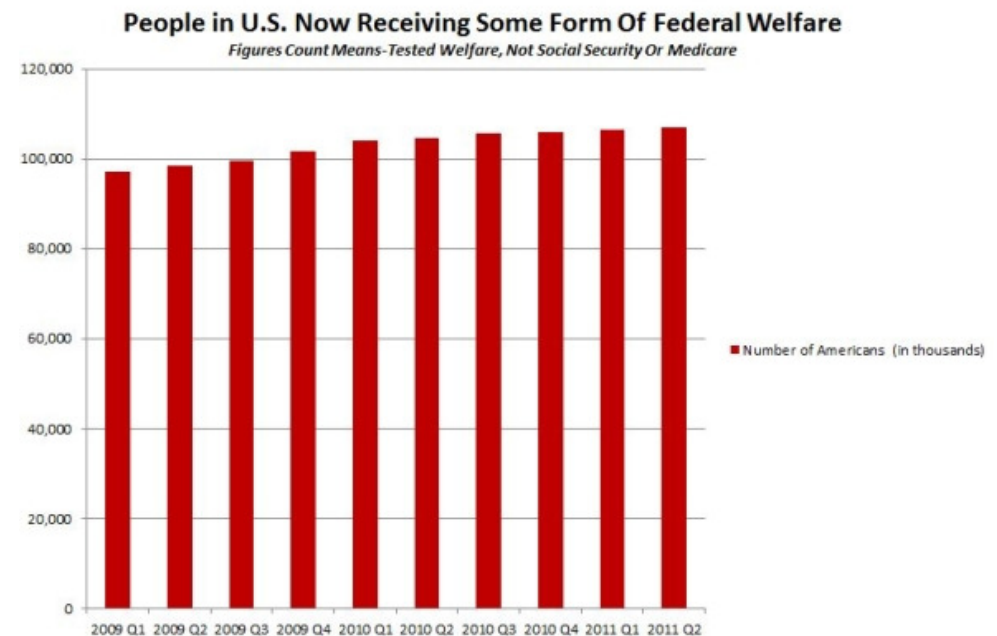


Presentation of data: Truncating axes



Graph A (left): Fox News graph illustrating welfare reform, amplifying Obama's demolition of welfare reform. The Y-axis is skewed and starts at 94,000,000 (not at 0).

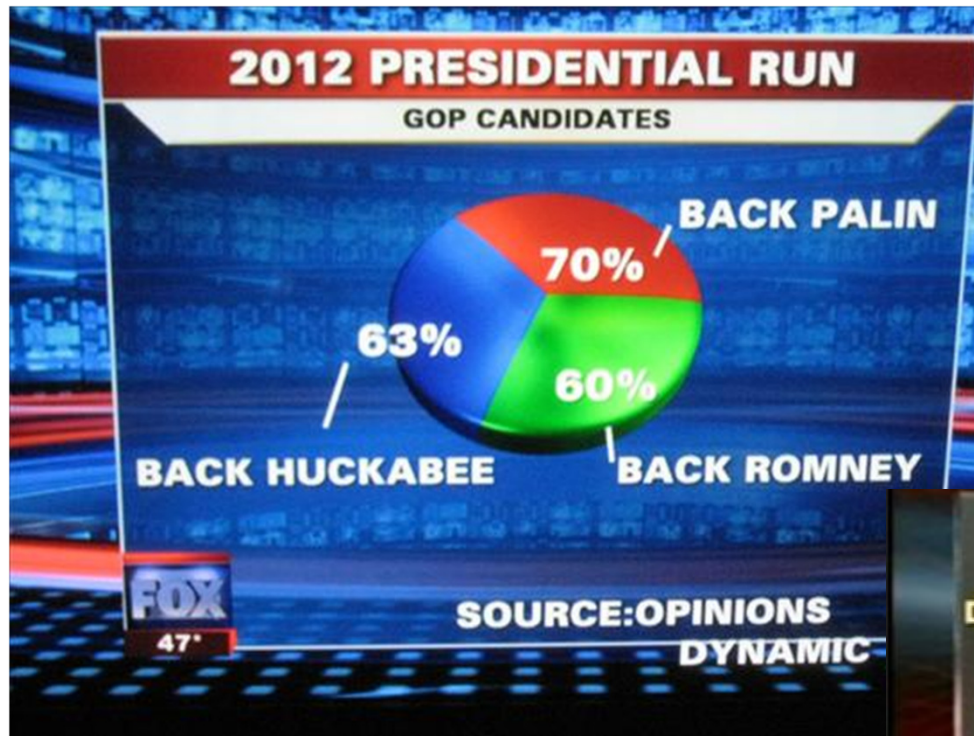
Graph B (right): Realistic depiction of these same data, with axes starting at 0



Source: mediamatters.org; retrieved 5 March 2014



Presentation of data: Numbers don't add up



$$70 + 63 + 60 = ? 193\%$$



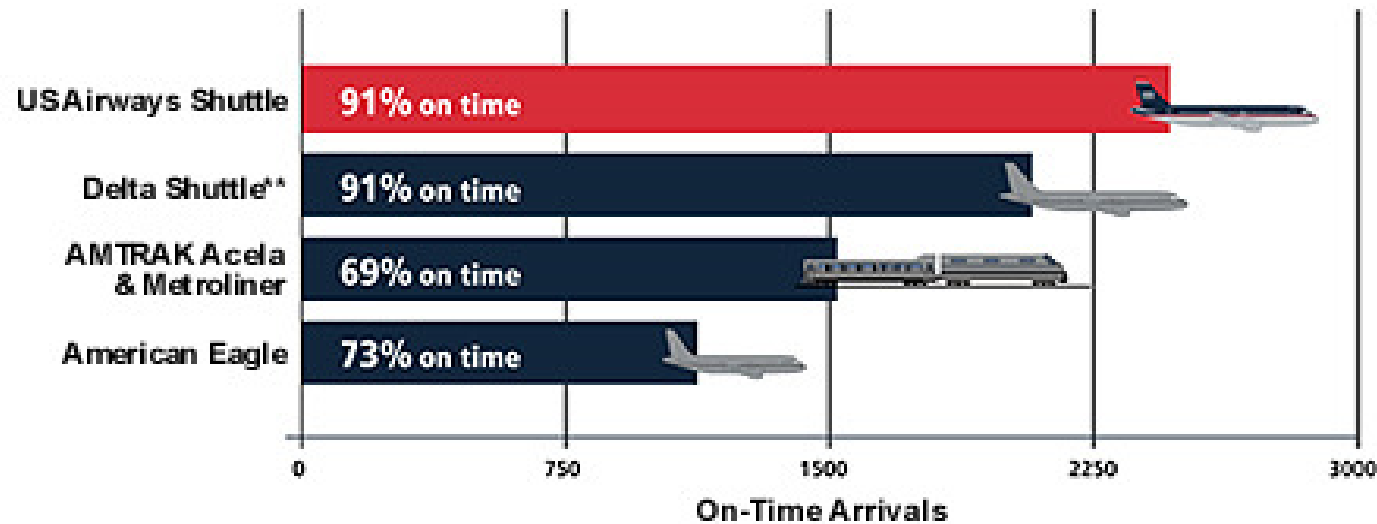
$$59 + 35 + 26 = ? 120\%$$



Source: simplystatistics.org; retrieved 5 March 2014



Presentation of data: Numbers don't match scale



ACTIVITY 4: What is an accurate conclusion from this data?

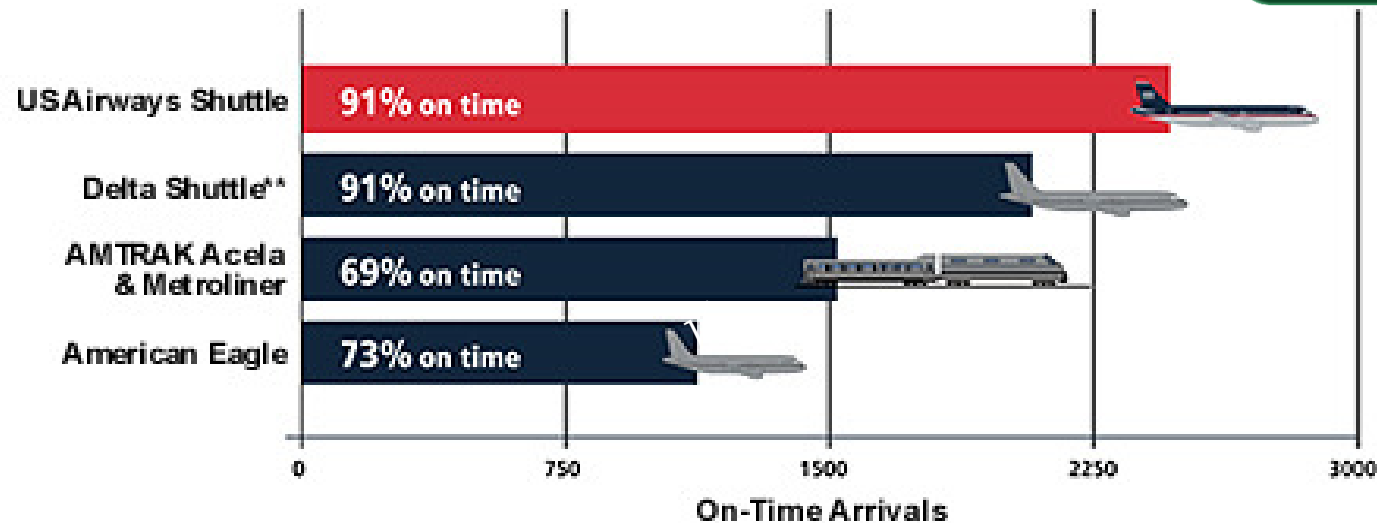
- a) US Airways Shuttle is clearly the most punctual airline
- b) Delta and US Airways are equally punctual
- c) There is not enough information to make a conclusion
- d) None of these

Source: saylee.worldpress.com; retrieved 5 March 2014

Presentation of data: Numbers don't match scale

91% more than
91%?

73% is less
than 69%?



ACTIVITY 4: What is an accurate conclusion from this data?

- a) US Airways Shuttle is clearly the most punctual airline
- b) Delta and US Airways are equally punctual
- c) There is not enough information to make a conclusion
- d) None of these

Source: saylee.worldpress.com; retrieved 5 March 2014

Presentation of data: Misleading numbers



Employment in Sydney



■ Employed ■ Unemployed

❗ What do categories include?

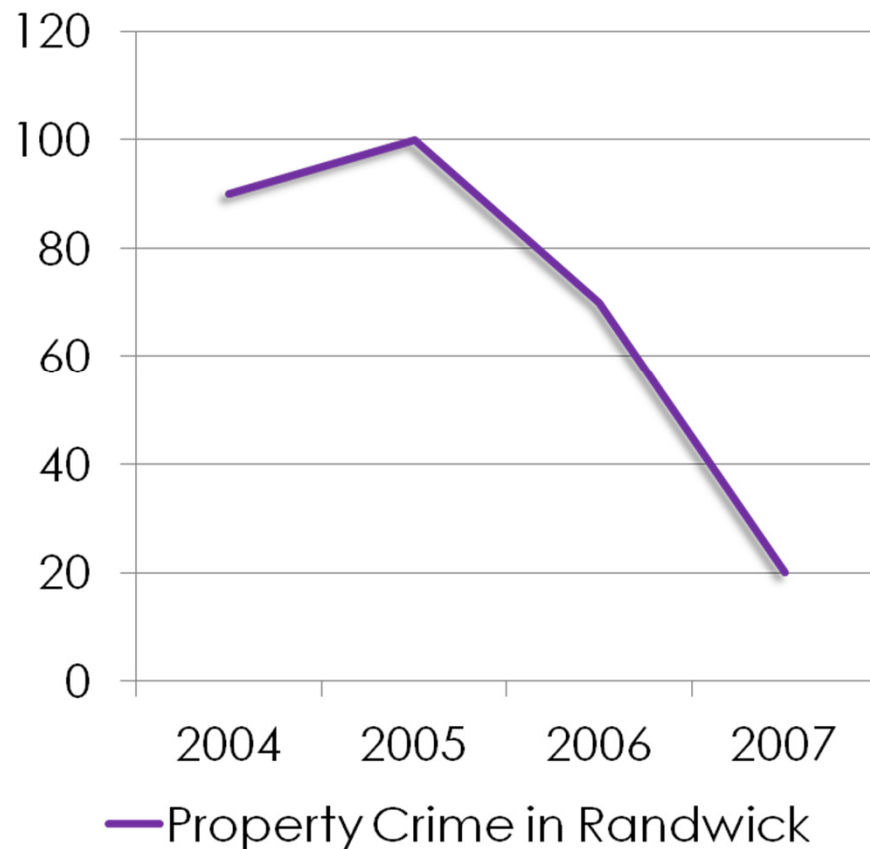
- Are casual / part-time employees considered?
- How many hours per week included in 'employment' category?
- Are categories consistent with previous measures (if comparisons are made)?
- Over what time period?
- Do numbers add up to 100?



Presentation of data: What does it mean?



**Property crime in
Randwick**



- This looks like property crime has decreased from 2004-2007, right?
- Definition of property crime?
- Recording property crime?
- Reporting property crime?
- All these factors can vary the data without the actual crime rate changing
 - It's crucial to know what exactly is being measured and presented.



The UGLY... interpreting significance

- ✿ Some research overestimates or exaggerates its effect or significance.
- ✿ As such, examine the results yourself and draw your own conclusions.
- ✿ It's good to know a little about p values:
 - p values refer to significance, and tells you how confident you can be that the difference being measured is a true difference in the broader population
 - Most tests are conducted with 95% confidence interval: $p < 0.05$
 - If the p value is less than 0.05, this means you can be 95% confident that difference observed reflects a real difference. I.e., If the study was repeated, the difference would be expected in 95/100 replications
 - The **lower** the p value, the more confident you can be that the difference observed in the study reflects a true difference

E.g. we are more confident when we find $p = 0.001$ than $p = 0.049$

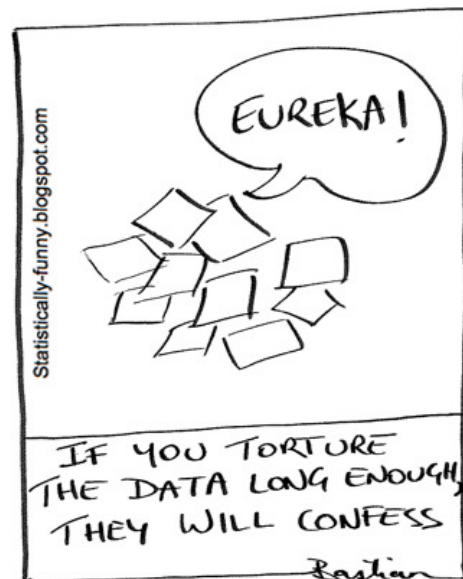


The UGLY... interpreting significance



✿ It's good to know a little about p values:

- We also need to think about the size of the difference and whether it is meaningful. E.g., 1% change in developing mood disorder, is that meaningful?
- Beware of studies that include large numbers of statistical tests without adjusting the threshold for significance (e.g. to $<.001$, "Bonferroni" or other adjustment).
- The more statistical tests that are conducted, the greater the likelihood that something will emerge as "significant". When conducting multiple tests, this may be capitalising on chance rather than reflecting a true difference



Thank you!



AMM



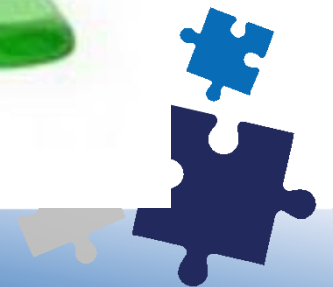
Normal cone



Sili-cone



Source: I * love science; retrieved 26 June 2014




Visit us... comorbidity.edu.au



NHMRC CENTRE OF RESEARCH EXCELLENCE
in MENTAL HEALTH and SUBSTANCE USE



Home About Us Research Our People Publications Resources Training Contact Us



Our research streams: Prevention

Latest News

Jun 09	New substance use and PTSD therapist guide and...
May 30	Read about Dr Tonelle Handley's suicide...
May 30	Australian Rotary Health awards scholarship to...
May 26	Registration is now open for the 2nd National...

[View More](#)

Latest Events

Aug 25	15th International Mental Health Conference, 25-...
Jul 23	The 2014 Australian Winter School Conference, 23-...
Jul 03	CMHDARN Rural Research Forum, 3 July 2014
Jun 14	Annual Meeting of the College on Problems of Drug...

[View More](#)



[Privacy Policy](#) [Copyright and Disclaimer](#) [Accessibility](#) [About This Site](#) [Log in](#)



Australian Government



Further reading



- ❖ Howell, D. C. (2012). Statistical Methods for Psychology. Belmont, CA : Thomson/Wadsworth.
- ❖ Smithson, M. (2000). Statistics with Confidence. London: Sage.
- ❖ Keppel, G., & Wickens, T. D. (2004). Design and Analysis: A Researcher's Handbook. (4th Ed.). Upper Saddle River, NJ: Pearson.
- ❖ The Cochrane library <http://www.thecochranelibrary.com>

Papers reviewed in workshop:

- ❖ Mills KL, Lynskey M, Teesson M, Ross J, Darke S. (2005) Post-traumatic stress disorder among people with heroin dependence in the Australian treatment outcome study (ATOS): prevalence and correlates. Drug and Alcohol Dependence, 77(3): 243-249
- ❖ Bucknam, W. (2006) Suppression of symptoms of alcohol dependence and craving using high dose baclofen. Alcohol & Alcoholism, 42(2): 158-160

