AMATYC
43rd Annual Conference
San Diego 2017

Statistical Errors: Finding Them, Avoiding Them

Brian E. Smith

McGill University

Data vs Statistics

"Data don't make any sense,
we will have to resort to statistics."

Outline of Presentation

The presentation will analyze different types of errors that appear in published journals.

The question we ask is what are the sources of these errors and how can we prevent them?

We will provide examples and discuss the causes and potential solutions for different categories of common statistical errors.

PUBLICATION BIAS

Publication bias is a type of **bias** that occurs in published academic research. It occurs when the outcome of an experiment or research study influences the decision whether to **publish** or otherwise distribute it. ... Publishing only results that show a significant finding disturbs the balance of findings.

Publication bias: what are the challenges and can they be overcome?
Joober, R. et al, Journal of psychiatry and neuroscience

In clinical trials in medicine, withholding negative results from publication — publication bias — can have major consequences affecting public health. In clinical and experimental research, publication bias may seriously distort the literature and lead to erroneous research and educational practices.

Subjects who volunteer to participate in research give their consent with the belief that their participation is helping to improve treatment outcomes and to advance scientific knowledge.

In order to honestly and ethically fulfill their contract with research subjects scientific researchers and journal editors should publish both positive and negative outcomes in an equitable manner.

Wise words from the past

Appearances to the mind are of four kinds.
Things either are what they appear to be;
Or they neither are, nor appear to be;
Or they are, and do not appear to be;
Or they are not, and yet appear to be.
Rightly to aim in all these cases
Is the wise man's task.
Epictetus, 2nd century AD

Quoted in the Journal of Psychiatry and Neuroscience

False Research Findings

In a thought-provoking paper, "*Why most published research findings are false*" John Ioannidis has argued, with some justification, that it is possible that most of what is published in the current biomedical literature is false. "Ioannidis claims that most published research is false mainly because the a priori probabilities of most tested relationships in most research fields are very low."

The literature is predominantly biased toward positive results, of which many are likely to be false, whereas negative results that are more likely to be true are less likely to be reported.

"This may explain why, despite thousands of published papers in psychiatry, it is sometimes very difficult to identify solid facts beyond some fundamental observations."

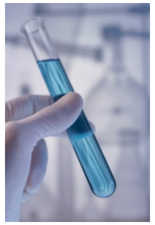
Three categories of negative results

1. Conclusive negative results: clear evidence of an opposite effect (e.g. treatment harms and benefits was expected) or a neutral effect (no effect of a new treatment) in a well-designed study.
2. Exploratory negative results: well-designed and adequately powered study with neutral or opposite results based on exploratory data analysis.
3. Inconclusive negative results: no evidence of an effect in the study that was too small and inadequately powered (e.g., no treatment effect due to small sample size).

Statement of goals from the journal

“The *Journal of Psychiatry and Neuroscience i*s an open access journal committed to publishing high quality manuscripts that will contribute to an unbiased updating of the literature. It welcomes manuscripts reporting negative results, particularly from category 1. We will also consider category 2 manuscripts favourably and encourage authors to be open about the exploratory nature of the results.”

Hannah’s Story

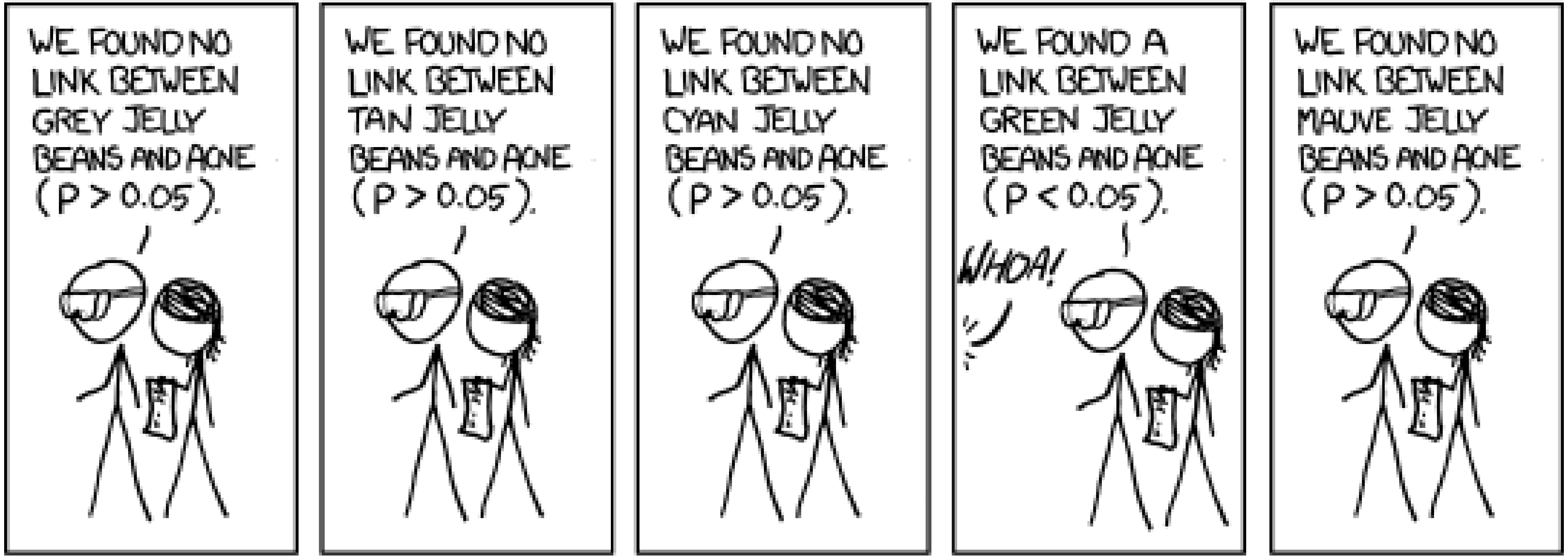


A manufacturer of homeopathic medicines asked Hannah, a microbiologist of international repute, to test a particular compound for its reputed antibacterial effects.

Hannah undertake the research as requested and found that the compound had no therapeutic benefit when compared with a placebo.

The company then told Hannah that she could not publish the findings. When Hannah insisted, the company threatened her with a lawsuit if she published the negative results! Being a scientist who upholds the highest ethical standards in her research she proceeded to publish. Fortunately the company backed off and did not prosecute!

Jelly Beans and statistical significance
Beware of the p-value



Nuijten, M.B. (2016). *Preventing statistics errors in scientific journals*. European Science Editing, 42, 1 8-10

In this paper the authors document inconsistencies in more than 250,000 *p*-values from eight major psychology journals. They discovered that in half of the papers at least one *p*-value was inconsistent with the test statistic and degrees of freedom. Although in most cases the reported *p*-value was only slightly different from those recomputed, nevertheless one in eight papers (12.5%) had “gross inconsistencies” that may have affected the statistical conclusions.

Preventing statistics errors in scientific journals.

The authors found a higher occurrence of major inconsistencies in *p*-values reported as significant, than in those reported as not significant, implying a systematic bias toward statistically significant findings.

It’s not only psychology!

Although these findings were documented in psychology journals, similar consistency rates have been found in the medical sciences in general and psychiatry in particular. Even though some of the reported *p*-values are only slightly different from the correct values, nevertheless this may result in an over reporting of false positive findings.

Causes of errors in reporting *p*-values

1. Sloppy calculations (note that a single psychology research paper contains about ten statistical tests)

1. Publication Bias: papers that report a *p*-value as significant are more likely to get published!

1. Lowenstein et al found that 22% of a sample of over 2000 psychologists admitted to *knowingly* rounding down a *p*-value to report significance thus leading to an excess of false positive findings!

2. Researchers often do not check the results of their co-authors' analyses, and may not even share data with them!

Possible Solutions

1. Promote data sharing.

2. Increase the statistical component of peer reviewing.

3. Award “badges” to authors who share their data (the journal *Psychological Science* awards badges to papers that are accompanied by open data).

4. Use Statcheck – an R package that converts PDF and HTML files to plain text files, extracts data and recomputes p-values, etc.)

Preregistration – a model for avoiding publication bias

The idea is to have researchers submit a detailed research plan *before* collecting data. If the plan is accepted by the journal after peer reviewing, the journal commits to publishing the paper **<u>regardless of whether the results are positive or negative.</u>** This removes the authors' concerns about doing the hard work and not getting a publication if the results are not statistically significant!

Statistical errors in medical research – a review of common pitfalls by Strasak, A.M. et al Swiss Med Wkly, 2007; 137:44-49.

In this paper the authors classify several types of common statistical errors and include suggestions for avoiding them.

They include the following categories:

- Study design
- Data Analysis
- Documentation
- Presentation
- Interpretation

Study Design

Planning and designing a study is a necessary prerequisite for a successful statistical analysis.

Errors occurring in the planning stage can have a negative effect on the validity and reliability of research results.

Major items of concern in this category include:

- Sample size estimation – Type I and Type II error, Power of the test
- Hypothesis Test specification – hypotheses should be pre-specified
- Random Sampling – inferential statistics techniques are valid only for random samples

Data Analysis

Pitfalls in this category

- Use of the wrong statistical test – e.g. Parametric v Nonparametric tests
- Failure to test model assumptions e.g. Student's t requires normality
- Failure to test model assumptions e.g. ANOVA requires normality and homoscedasticity
- In multiple comparison tests failure to include corrections e.g. Bonferroni
- In χ^2 tests failure to use Yates Continuity correction for small samples
- In regression models failure to use multivariate techniques to adjust for confounding factors

Documentation

All statistical methods should be described clearly and with enough detail to permit a reader with access to the study data to recalculate results. Some of the problems here would include:

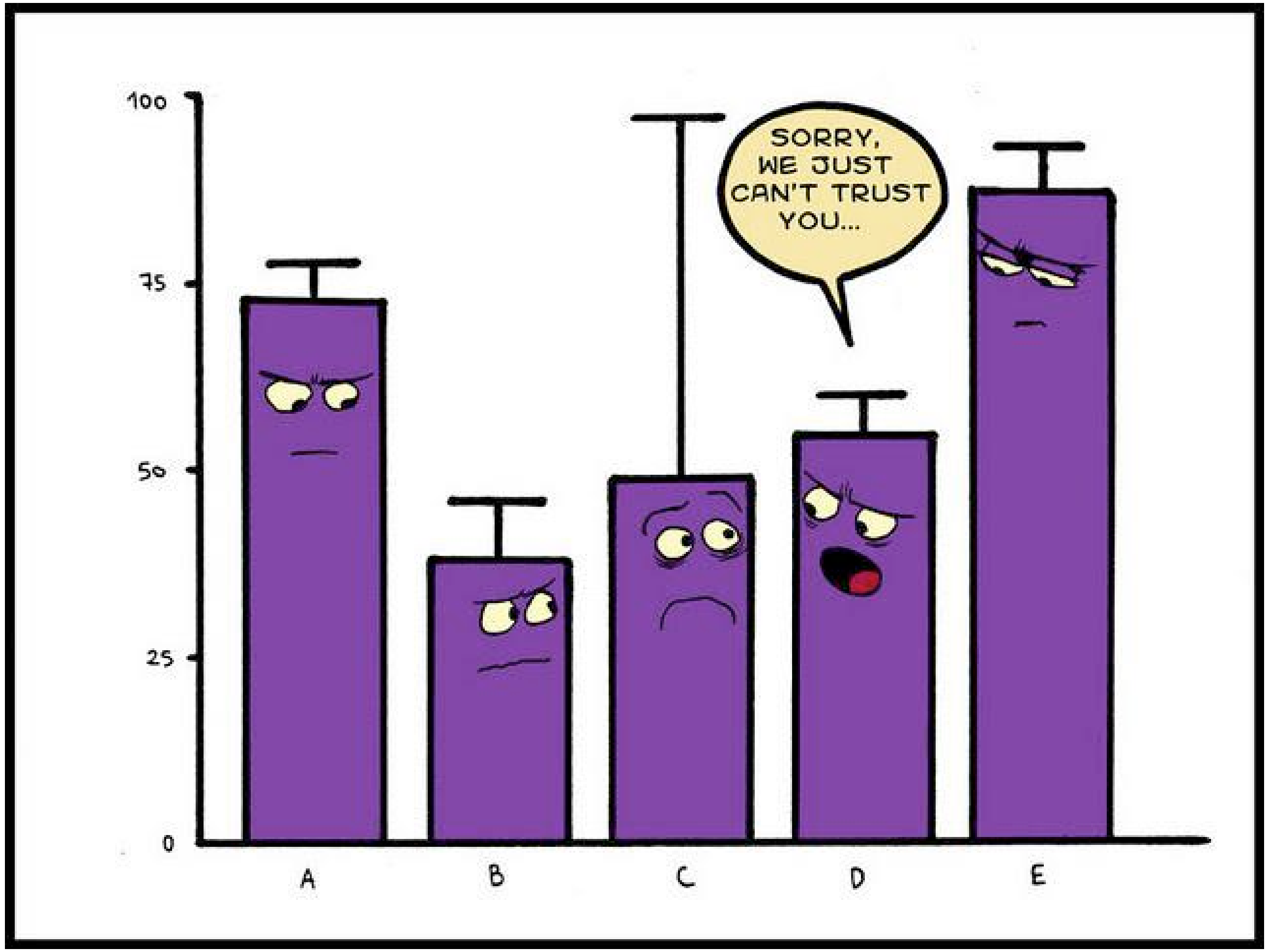
- Failure to specify tests used for analysis
- Failure to specify one-tail vs two-tail tests
- Failure to state if test for equality of population means was paired or unpaired

Presentation

According to Evans good research deserves to be well presented and sound presentation is as much a part of the research as the collection and analysis of data. Here are some points that are of importance in data presentation:

- Inadequate graphical or numerical description of the data set
- Reporting means without adequately describing variability (skewness, outliers)
- Providing SE (standard error) instead of SD (standard deviation)
- Specifying *p*-values without confidence intervals
- Reporting Mean ± SD may be confused with 95% confidence intervals – it is better to write Mean (SD)

Biased Box Plots



Interpretation

Points to ponder:

- Don’t confuse “not significant” with “no difference” or “no effect”
- Don’t draw conclusions not supported by the study data
- Discuss Type II errors when reporting non-significant results
- Discuss sources of potential bias and confounding factors

False Alarms and Pseudo-Epidemics
The limitations of Observational Epidemiology

David A. Grimes MD and Kenneth F, Schulz PhD, MBA

Obstretics & Genecology, Vol. 120, No 4, Oct 2012

Grimes and Schulz make the astonishing claim that "most reported associations in observational clinical research are false, and the minority of associations that are true are often exaggerated".

Their paper then goes on to discuss and justify their assertion.

Bias

They claim that most associations reported in the medical literature are more likely due to bias than to a real causal relationship between variables.

As evidence for this they report on 12 randomized controlled trials that tested 52 claims from observational studies. None of the claims could be substantiated and, amazingly, for 5 of the 52 claims the treatment effect was found to be statistically significant in the opposite direction.

Why so many statistical problems?

1. Most clinical researchers have had little or no formal training in statistical analysis beyond an introductory statistics course. I would note that the ubiquity of computer hardware and sophisticated software makes everyone feels that they can “do” statistics, but without the expertise to recognize inappropriate modeling or incorrect procedures.
2. The availability of huge databases makes “data mining” irresistible – encouraging people to dredge the data in search of something to publish – often without having pre-established hypotheses or guidelines.
3. The imperative to publish. Number (as opposed to quality) of publications is often used as an index of academic achievement. As the authors note “with almost 25,000 biomedical journals in circulation, almost any manuscript gets published somewhere.”

…more

4. The impact of medical journalism and the instant fame that comes from, for example, a breakthrough in cancer research encourages researchers and journalists to report on possible findings long before any real links have been established!
5. Journals will often accept manuscripts of poor quality and which make exaggerated claims. There are huge numbers of journals that need to accept a large number of publications to “stay alive”. Not all of the papers submitted are of a high quality of research and in particular they may not involve well informed statistics. Two negative consequences of publishing weak or false results are (1) sensationalism and (2) litigation – think of some of the huge settlements involved in sensational courts cases which in some case may be based on false statistics (the authors cite the case of silicon breast implants that resulted in a $2.35 billion settlement “despite the scientific evidence refuting an association with autoimmune disease”)

Examples of misleading statistical studies

In 1981, a study was released claiming an association between drinking coffee and pancreatic cancer. Needless to say, this study sparked a media frenzy and caused immense concern among coffee drinkers – and, of course, coffee producers! However, later studies found this association to be false.

In 1974 the drug Reserpine, used for treating hypertension, was drawn from the market after studies claimed a doubling in the risk of breast cancer. Later more carefully constructed statistical studies failed to confirm the link between Reserpine and breast-cancer, but there was no turning back – the drug had been discredited by the public and was never resuscitated.

These types of occurrences cause significant erosion in public confidence in medical research.

Finally, a note on Causality

One cannot discuss statistics without, at least briefly, mentioning causality.

Issues of causation represent a major area of misunderstanding.

It is always essential to distinguish between a statistical relationship and a causal relationship.

It is also necessary to carefully define the direction of causation (a recent studies in New Zealand was designed to establish the causal connection that heavy pot smoking causes early onset psychosis, and not the reverse as some have claimed!)

Regression and Causation

Can a regression analysis establish causality?

Statistical textbooks are full of cute and patently ridiculous examples of false claims of causation. For example, does increased consumption of ice cream cause drowning? (the strong statistical association is explained by the fact that both ice cream consumption and drowning have similar seasonal effects.)

Misleading Headline?

Proximity to freeways increases autism risk, study finds.

Children born to mothers who live close to freeways have twice the risk of autism, researchers reported Thursday.

LA Times, Dec 16 2010

But do they? The report goes on to say “This study isn't saying exposure to air pollution or exposure to traffic causes autism," said Heather Volk, lead author of the paper … ‘But it could be one of the factors that are contributing to its increase.’“

Note how the lead author’s “could be” was converted into a clear causal statement by the LA Times’ headline editor!

Pot and Schizophrenia

Cannabis use in adolescence and risk for adult psychosis: longitudinal prospective study
BMJ 2002; 325 doi: https://doi.org/10.1136/bmj.325.7374.1212 (Published 23 November 2002) Cite this as: BMJ 2002;325:1212

Logistic regression analyses showed that people who used cannabis by age 15 were four times as likely to have a diagnosis of schizophreniform disorder at age 26 than controls.

Food for thought!! Do the statistics hold up?

References

Joober, R, et al. *Publication bias: What are the challenges and can they be overcome?*
J Psychiatry Neurosci. 2012 May; 37(3): 149–152. doi: 10.1503/jpn.120065
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3341407

Ioannidis, J.P. *Why most published research findings are false*. PLoS Med. 2(8) e124. Epub 2005.

Nuijten, M.B. (2016). *Preventing statistics errors in scientific journals*. European Science Editing, 42, 1 8-10

John LK, Lowenstein G, Prelec D. *Measuring the Prevalence of questionable research practices with incentives for truth-telling*. Psychological Science. 2012; 23:524-32.

Strasak, A.M. et al. *Statistical errors in medical research – a review of common pitfalls.* Swiss Med Wkly, 2007; 137:44-49. www.smw.ch

Evans, M. *Presentation of manuscripts for publication in the British Journal of Surgery. Br J Surg. 1989; 76:1311-4*

David A. Grimes and Kenneth F, Schulz. *False Alarms and Pseudo-Epidemics: The limitations of Observational Epidemiology* Obstretics & Genecology, Vol. 120, No 4, Oct 2012