

**Can research funding bodies adequately  
take methodological issues into account?  
— an open question**

**Jason Grossman**

University of Queensland

<http://tumble.tk>

The ability of research funding bodies to make good assessments of the methodology of proposed empirical projects is extremely limited. This raises two major problems for the health of research.

1. This limits funders' ability to fund the best research.
2. This could kill innovation in research methodology, especially in social research.

# Contents

- Introduction
- Six contentious things we can't seriously expect funding bodies to make informed judgements on
- When we subcontract methodological decisions
- What else can we do?

# Introduction

# Introduction

- Research power brokers:
  - funding bodies

# Introduction

- Research power brokers:
  - funding bodies
  - ethics committees

# Introduction

- Research power brokers:
  - funding bodies
  - ethics committees
  - legal & policy bodies

# Introduction

- Research power brokers:
  - funding bodies
  - ethics committees
  - legal & policy bodies
  - journal editors

# Introduction

- Research power brokers:

- funding bodies

- ethics committees

- legal & policy bodies

- journal editors

Coming up: six contentious things we can't seriously expect funding bodies to make informed judgements on.

# Introduction

- Research power brokers:

- funding bodies

- ethics committees

- legal & policy bodies

- journal editors

Coming up: six contentious things we can't seriously expect funding bodies to make informed judgements on.

My main claims (for today, anyway):

1. **These areas are contentious, with only ad hoc “solutions”.**

# Introduction

- Research power brokers:
  - funding bodies
  - ethics committees
  - legal & policy bodies
  - journal editors

Coming up: six contentious things we can't seriously expect funding bodies to make informed judgements on.

My main claims (for today, anyway):

1. **These areas are contentious, with only ad hoc “solutions”.**
2. **These areas are important.**

# Introduction

- Research power brokers:
  - funding bodies
  - ethics committees
  - legal & policy bodies
  - journal editors

Coming up: six contentious things we can't seriously expect funding bodies to make informed judgements on.

My main claims (for today, anyway):

1. **These areas are contentious, with only ad hoc “solutions”.**
2. **These areas are important.**
3. **Funding bodies are not equipped to evaluate the arguments.**

# Introduction

- Research power brokers:
  - funding bodies
  - ethics committees
  - legal & policy bodies
  - journal editors

Coming up: six contentious things we can't seriously expect funding bodies to make informed judgements on.

My main claims (for today, anyway):

1. **These areas are contentious, with only ad hoc “solutions”.**
2. **These areas are important.**
3. **Funding bodies are not equipped to evaluate the arguments.**

# Example 1: significance tests

# Example 1: significance tests

- What is a P value?

# Example 1: significance tests

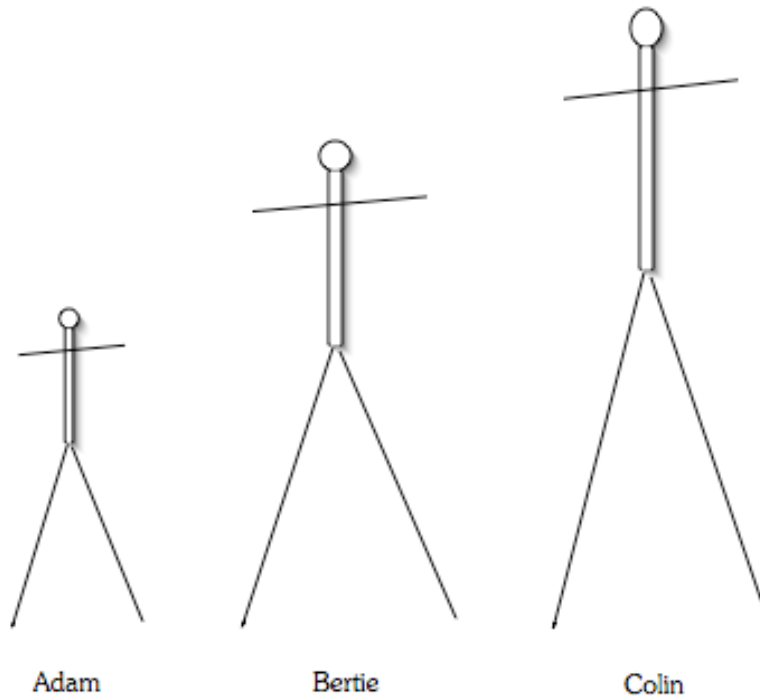
- What is a P value?
- $P = p(\text{observed result} \mid H_0)$ ?

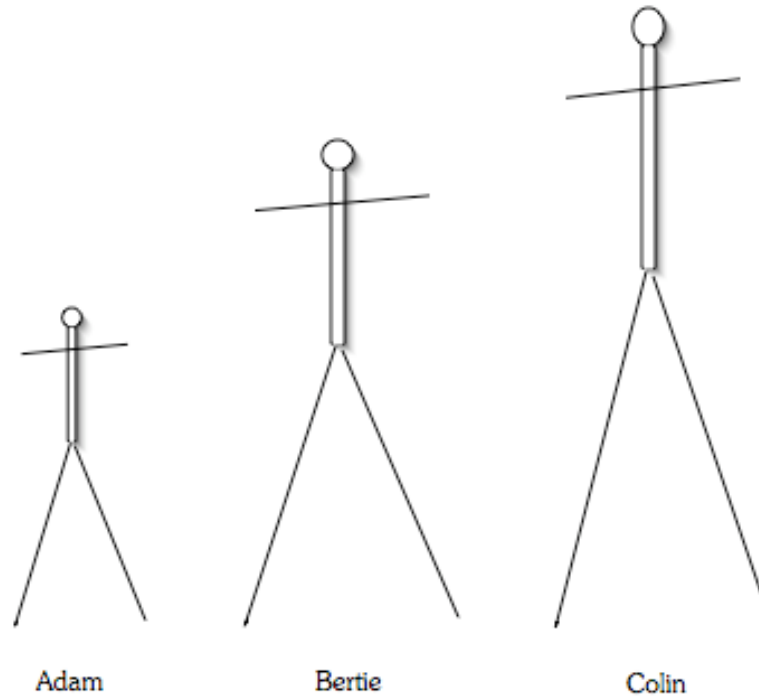
# Example 1: significance tests

- What is a P value?
- $P = p(\text{observed result} \mid H_0)$ ?
- $P(h_0, T(x_a))$  is defined as the proportion of an infinite sequence of hypothetical experiments, each duplicating the experiment we have actually conducted, on the assumption that the hypothesis  $h_0$  is true, that would result in a value of  $T(x_i)$  greater than or equal to  $T(x_a)$
- . . . where  $x_i$  is the observation made in each hypothetical experiment,  $x_a$  is the observation made in the actual experiment, and  $T$  is an arbitrary function from the space of possible observations to the real numbers.

# Example 2: confidence intervals

I am studying bonobo chimpanzees in the wild. Researchers further up the river have told me that three new bonobos have moved into my study area. Two of them, Adam and Colin, are indistinguishable apart from size; the third, Bertie, is unusually pale and exactly intermediate in size between Adam and Colin. I know that Colin is two metres taller than Adam, and I expect to get two measurements of Adam and/or Colin. My task is to estimate  $x_B$ , the height of the middle one.





The best 75% confidence interval for  $x_B$  is:

$$C = (x_1 - 1, x_1 - 1) \quad \text{if } x_1 = x_2$$
$$= \left( \frac{x_1 + x_2}{2}, \frac{x_1 + x_2}{2} \right) \quad \text{otherwise.}$$

“ $p(C) = 75\%$ ”

# Example 3: multiplicity I

We test 300

A group sequential clinical trial tests its main hypothesis typically 6 times.

We want a 5% “rate” of false positive errors.

There is no standard method for this.

# Example 3: multiplicity II

# Example 3: multiplicity II

<age of subject 1: **801 months**,

initial tumour histology for subject 1: **t35**,

initial treatment for subject 1: **radiotherapy**,

size of subject 1's tumour at 6 months: **unknown**,

side-effects at 6 months: **unknown**,

size of subject 1's tumour at 13 months: **11 mm**,

site of subject 1's secondary tumours: **leukemia**,

. . .

age of subject 2: **684 months**,

initial tumour histology for subject 2: **q+**,

initial treatment for subject 2: **none**,

. . .

age of subject 3: **787 months**. . . >

# Example 4: validated measures

# Example 4: validated measures

- over-stating correlation coefficients

# Example 4: validated measures

- over-stating correlation coefficients
- significance tests of correlation coefficients

# Example 4: validated measures

- over-stating correlation coefficients
- significance tests of correlation coefficients
- non-transitivity of validation

# Example 4: validated measures

- over-stating correlation coefficients
- significance tests of correlation coefficients
- non-transitivity of validation
- use of “validated” as a success word

## Example 5: intention to treat analysis

An experimental weight-loss program recruits 100 subjects, 50 of whom are randomized into a control group. Of the other 50, only 10 complete the program. An intention-to-treat analysis would (by definition) work out average results for all 50, *including* the 40 who did not complete the intervention.

# Example 6: randomised controlled trials (RCTs)

**JASON GROSSMAN AND FIONA J. MACKENZIE**

---

**TABLE 1** SAMPLE EVIDENCE HIERARCHIES

<i>MERGE (New South Wales Department of Health)</i> <sup>1</sup>	<i>U.S. Preventive Task Force</i> <sup>2</sup>	<i>Medicare Services Advisory Committee</i> <sup>3</sup>
<b>Level I:</b> systematic review of all relevant randomized controlled trials; large multi-center randomized controlled trials	<b>Level I:</b> evidence obtained from at least one properly randomized, controlled trial	<b>Level I:</b> evidence obtained from a systematic review of all relevant randomized controlled trials
<b>Level II:</b> one or more randomized controlled trials and studies	<b>Level II-1:</b> evidence obtained from well-designed controlled trials without randomization	<b>Level II:</b> evidence obtained from at least one properly designed randomized controlled trial
<b>Level III:</b> controlled trials without randomization; cohorts; case-control analytic studies; multiple time series; before and after studies (preferably from more than one center or research group)	<b>Level II-2:</b> evidence obtained from well-designed cohort or case-control analytic studies, preferably from more than one center or research group	<b>Level III-1:</b> evidence obtained from well-designed pseudo-randomized controlled trials (alternate allocation or some other method)
<b>Level IV:</b> other observational studies	<b>Level II-3:</b> evidence obtained from multiple time series with or without the intervention. Dramatic results in uncontrolled experiments (such as the results of the introduction of penicillin treatment in the 1940s) could also be regarded as this type of evidence	<b>Level III-2:</b> evidence obtained from comparative studies with concurrent controls and allocation not randomized (cohort studies), case-control studies, or interrupted time series with a control group
		<b>Level III-3:</b> evidence obtained from comparative studies with historical control, two or more single-arm

# Example 6: problems for RCTs

# Example 6: problems for RCTs

- comparison with case-control studies

# Example 6: problems for RCTs

- comparison with case-control studies
- small RCTs useless (at least in isolation)

# Example 6: problems for RCTs

- comparison with case-control studies
- small RCTs useless (at least in isolation)
- incorrect level of analysis

# Example 6: problems for RCTs

- comparison with case-control studies
- small RCTs useless (at least in isolation)
- incorrect level of analysis
- intention-to-treat analysis

# Example 6: problems for RCTs

- comparison with case-control studies
- small RCTs useless (at least in isolation)
- incorrect level of analysis
- intention-to-treat analysis
- blinding and placebos

# Example 6: problems for RCTs

- comparison with case-control studies
- small RCTs useless (at least in isolation)
- incorrect level of analysis
- intention-to-treat analysis
- blinding and placebos
- insensitive outcome measures

# Example 6: problems for RCTs

- comparison with case-control studies
- small RCTs useless (at least in isolation)
- incorrect level of analysis
- intention-to-treat analysis
- blinding and placebos
- insensitive outcome measures
- . . . and all the problems from examples 1 to 5 as well!

When we subcontract methodological decisions

# When we subcontract methodological decisions

- “Standard” answers are given.

# When we subcontract methodological decisions

- “Standard” answers are given.
- e.g. the 5% significance level cut-off

# When we subcontract methodological decisions

- “Standard” answers are given.
- e.g. the 5% significance level cut-off
- science as a space shuttle

# When we subcontract methodological decisions

- “Standard” answers are given.
- e.g. the 5% significance level cut-off
- science as a space shuttle
- suppression of innovation on the methodological level

What else can we do?

