Implicit Controversies in Statistical Inference

Vance W. Berger
National Cancer Institute
Gertrude Cox Award Address
W.S.S. Annual Dinner, June 22, 2006
Evidence-Based Medicine

“Medical Guesswork: From heart surgery to prostate care, the medical industry knows little about which treatments really work”.

Front page headline, *Business Week*, 5/29/06.

Replacing physician judgment with evidence (RCTs) is not done nearly often enough.

The lack of evidence to support many medical decisions certainly is a problem.

But when there is no evidence, this state of affairs is easily recognized, and can be fixed.
Evidence-Based Medicine

- It would be a far more sinister problem if the beloved and reliable RCT were tainted, as evidence from randomized trials is generally accepted without being subjected to scrutiny.

- If randomized trials can be subverted, and can mislead, then this would be disastrous, because of their tremendous influence.

- In the grand scheme, which is the greater threat, no evidence at all, or bad evidence?

- I’d rather go hungry than consume toxic food!
Evidence-Based Medicine

- In fact, much can go wrong in RCTs.
- Even random samples may misrepresent the populations from which they were drawn.
- But we consider only the preventable problems that occur routinely in RCTs.
- We focus not on specific problems per se, which represent only the tip of the iceberg, but rather on the climate that embraces them.
- Many problems are readily recognized as such, yet somehow they manage to persist.
Controversies in Statistics

- Explicit Controversies
  - Hotly debated, with articulate arguments on both sides.

- Implicit Controversies (examples to follow)
  - A small group of scholars all argue the same way, and nobody explicitly opposes these views
  - Practice entirely ignores these views, and serves as a *de facto* opposition.

It is incumbent upon all statisticians to do their part to recognize these implicit statistical controversies, and to come down on the right side of them. But the problem of not doing so is larger than statistics.
“If I have seen further than most men, then it is by standing on the shoulders of giants”.
– Isaac Newton, 1643-1727

These giants presumably include Galileo (1564-1642), tried for heresy in 1633.

Socrates (469-399 BC) was put to death for making weaker arguments appear stronger.

Darwin (1809-1882) was ridiculed.

Clearly, many prefer to stomp rather than stand upon the shoulders of these giants.
Choosing Sides

- We now follow Socrates (best to recognize what we don’t know), Galileo (the earth revolves around the sun), and Darwin (evolution should be taught in public schools).
- Eventually, emperor’s new clothes exposed.
- But how long does it take, and at what cost?
- Did we recognize the giants first and then follow their logic, or did we first recognize the valid logic, and so define the giants?
Implicit Controversies

- Implicit controversies pit reason & evidence against political might; when the winds are blowing hard in one direction, few dare to march in the other (logical) direction.
- What is offensive is not that the majority is wrong; it may not be, but at issue is the mode of argument; if the majority is correct, this is tangential to its spurious arguments.
- “If you are not outraged, then you are not paying attention” (car bumper sticker).
The Emperor's New Analyses

• Given that bad medical research inevitably leads to bad treatment decisions that can affect any of us, surely there must be provisions in place to ensure that only the enlightened engage in medical research.

• This may be true in Fantasyland, but now I welcome you to the real world of trials.

• Many common analyses are inappropriate, and there are experts willing to say so, yet these criticisms are never addressed by the many practitioners who ignore them [1].
Run-in Periods

- Many trials pre-treat patients with the active treatment, then select for the study only those patients who respond to it.
- During a carvedilol run-in period, seven deaths and 17 serious events occurred.
- These 24 events were excluded from the analyses by this “do-over” button [2].
- Do we evaluate the advisability of playing Russian roulette by considering the experiences of only those left standing?
Run-in Periods

- There is an intuitive objection in that the best that can be said is that the drug works for those in whom it works (tautology).
- The isolated experiences of the responders are biased for the population to whom the drug will be marketed, because one would need to take the drug (and risk adverse events) to ascertain membership in this group of responders [3], [4], [5], [6].
- The bias has never been addressed, and run-in-based patient selection continues.
Presenting Select Data Only

notify the doctor after taking the drug if you:
1. don't die
2. don't grow extra arms
3. don't have a height change.

i actually feel better from it!

i can't play basketball anymore!

mooooooooooooo
Approximating the Observable

- The second example concerns exact tests.
- Many trials use chi-square (or some other approximate analysis) unless conditions are not met, in which case Fisher’s exact test (or another exact test) is used instead.
- Clearly, then, Fisher is of inherent interest, and the chi-square test is used only as an approximation when conditions allow.
- Otherwise, just use chi-square, regardless.
Approximating the Observable

• With large cell counts, or when the data appear normal, parametric analyses are used (the approximation is close enough).
• That is, we want $p(E)$, the exact $p$-value.
• But we use the approximate $p(A)$, because conditions suggest $D = p(A) - p(E)$ is small.
• So there are two blatant lapses of logic.
• First, we want $p(E)$ but instead use $p(A)$ because it should be close; second, we never even compute $D = p(A) - p(E)$!
Approximating the Observable

• To distinguish the general from specific, why not compute \( D = p(A) - p(E) \), and then use \( p(A) = p(E) + D \) only if \( D \) is small?

• \( D \) may be large even when conditions hold to suggest that \( D \) should be small [7]; it is our “job to be suspicious even when there is no reason to be suspicious” (*The Firm*).

• Moreover, with \( p(E) \) and \( D \) in hand, why add them together, when we want \( p(E) \)?

• These questions [8] are never answered by the masses, who continue using \( p(A) \).
An Approximate World #1

3, 2, 1, and that's it, the clock reached zero and the game is over!

Wait, he shoots, he scores, and it counts! The score clock must have been only approximate!
An Approximate World #2

Yay! 10 years are over!

No! 10 years was just approximate! It's actually 15 years!
An Approximate World #3

But it says only $4.99!

That was just approximate.

$10.00
An Approximate World #4

I only went at 60 MPH! >_<

You have just earned yourself a ticket.

65 MPH was just approximate. The speed limit is really 55 MPH.
Why Do We Randomize?

I'm lifting weights!

What are you doing?
Selection Bias

• If randomization eliminates selection bias and ensures internal validity, then has there ever been a truly randomized trial?
• In the real world, randomized trials are always restricted, usually permuted blocks.
• This is because without restrictions, great imbalance could result, both at the end of the trial and at all points during the trial.
• The cure may be worse than the disease!
The Cure Is Worse than the Disease!

I HATE when that happens, especially so close to home. So close, and yet so far!
Selection Bias

- Randomization without restrictions would allow for chronological bias.
- Blocking entails selecting a block size, and randomizing patients within each block.
- If the block size is four, then the first four patients are a block, with two per group.
- Blocking certainly controls chronological bias, because it forces balance in the numbers in each group at all times.
Selection Bias

- But future treatment assignments can be predicted, and healthier patients can be recruited to one treatment group, sicker patients to the other treatment group.

- This bias can be prevented, detected, and corrected [9], but in practice none of this is done, because it is easier to deny the possibility of bias in randomized trials.

- The masses rely on now obsolete articles to defend their right to do the wrong thing.
C'mon fellow Biases! We can get in through the side door! Follow me!
9-1-1 Operator, can I help you?

You don't need to call the police! Records show that this is a safe neighborhood so there really isn't a burgler.
Scandalous Medical Research

- Altman [10] spoke of the “scandal of poor medical research”, and our three examples represent only the tip of the iceberg.
- Biases can even conceal themselves by creating results that appear so impressive that biases could not seem to explain them.
- Enhancing the results of even effective drugs is a problem because of elasticity.
- The number of treated patients depends on the perceived benefit of the treatment.
Scandalous Medical Research

- Overstating the benefits of a treatment will cause some patients to take the treatment when they otherwise would not have.
- This is a denial of the patient’s right to make an informed treatment decision.
- The scope of medical errors must be broadened to include not only medical malpractice but also research malpractice.
- Poor research is rarely haphazard; being regimented gives an illusion of validity.
Scandalous Medical Research

- Still, poor research leads to distorted results, flawed yet regimented guidelines, and undetected poor treatment decisions.
- The victims are displaced in time and space, and hence cannot unite to trace the problem back to the research malpractice.
- It should be clear not only that misleading information can lead to harmful drugs being found safe and effective, but also that remedies for this problem are ignored.
- How can this situation be allowed to go on?
Scandalous Medical Research

- Interested parties have the power to shift the burden of proof from where it should be.
- Not selecting patients based on run-ins, and testing for selection bias, could reduce profits by clarifying a lack of true benefit.
- The continued use of approximate analyses may reflect precedent and a perceived advantage that in reality does not exist.
- Is the situation much better in your area?
Do for Your Country

• “Ask not what your country can do for you; ask what you can do for your country”.
  – JFK (1917-1963)

• Some among you may go on to follow Socrates, Galileo, and Darwin in creating wonderful new knowledge that will help not only your country, but the world.

• We can all aspire to at least emulate Plato, Newton, and Scopes in recognizing the giants from their ability to reason, and then follow these giants instead of the detractors.
Do for Your Country

• What can you do to ensure that those speaking the truth become the giants upon whose shoulders the rest of us stand?

• Avoiding dishonesty is not enough; one must fight for honesty to prevail; “All that is necessary for the triumph of evil is that good men do nothing” - Edmund Burke.

• It is not easy to break with tradition, even to do what one knows to be the right thing.

• But let us honor Gertrude Cox tonight by heeding her words on this issue.
“The fact that you, as an individual, are classified as a ‘statistician’ does not free you from obligations and responsibilities toward other human beings. Some of you may feel that you … are not to be held responsible for the consequences of your work … [but] you have an obligation to clarify the foundations of your techniques”.

The extent to which foundations are clarified determines the extent to which unfounded techniques are allowed (used) in practice.
Your Role

• It is unsafe to replace common sense with appeal to precedent or powerful authority.
• Credentials should predict, and not replace, reasonable actions and advice.
• I would urge you to demand reason, not only of yourselves, but also of your colleagues, on an ongoing basis.
• This profession, which has been so very good to all of us, and the memory of Gertrude Cox, deserve nothing less.
Further Reading

- More information is available -- just send me a message and I will send you articles.
- Vance Berger
- Vb78c@nih.gov
- (301) 435-5303
Further Reading


Further Reading