Testing for Blindness at the end of an RCT

don’t?

must?

can’t?

needn’t?
Definition of **BLINDING** used here

- The testing of human participants under conditions of *intentional ignorance* about the exact nature or timing of an intervention.
Double-Blind (as used here)

Both the patient and their clinician are intentionally ignorant about the exact nature of the experimental and control interventions.
Purpose of Blinding

To isolate “hard” knowledge and material causality from contamination by mental delusion, enthusiastic bias, or even calculated deceit.

Eg: Homeopathy Trial

• Active Rx: “potentized” salt solution (1 crystal dissolved in 1 drop of distilled snow water, which was then diluted 29 times at a ratio of 1:100, repeatedly shaken)

• prepared by 2 “uncontaminated” pharmacists (who had been kept away from allopathic drugs for 2 days and then carefully bathed).

• Placebo: distilled snow water
Homeopathy Trial, cont.

• 25 bottles of each were randomized (code locked away) and given by a blinded investigator to blinded volunteers (thus a double-blind RCT)

• Treatments taken for 3 weeks, after which participants were asked whether they had experienced “anything unusual”
Homeopathy Trial, cont.

• 42 of the 50 participants reported “nothing unusual”

• Only 5 on “active” Rx and 3 on placebo experienced “anything unusual” during the trial.

• Conclusion: No effect
In what year was this trial done?  

1835!

Authorship: “a society of truth-loving men”

It is the earliest double-blind trial unearthed by the James Lind Library

www.jameslindlibrary.org/
Historic uses of blinding

18th-19th centuries: To challenge the “bogus” claims of unconventional medicine (Mesmerism, telepathy, hypnosis)

Only recently (20th century) employed to evaluate conventional medicine

Logically followed the introduction of concurrent controls and placebo/active comparison Rx
Given the importance traditionally assigned to blinding

It is “vital” to protect it during a trial

It is “vital” to validate whether it has prevented bias at the end of a trial
So we are taught to test for blindness before, during, and after an RCT . . .

. . . by asking study participants and their clinicians to speculate about the treatment they’ve received during the trial.
That’s what Textbook #1 says…

“After the study is over, it is a good idea to assess whether the participants and investigators were unblinded by asking them to guess which treatment the participant was assigned to . . . .

“... if a higher than expected proportion guesses correctly, the published discussion of the findings should include an assessment of the potential biases that partial unblinding may have caused.”
That’s what Textbook #2 says...

“... Ask the participant and the clinic staff to guess to which group the participant was assigned.

To the degree that 50% is exceeded, the amount of unblinding can be estimated.

If substantially fewer than half of the guesses are correct, one might suspect that people did know but were trying not to admit it.”

-Friedman, Furberg, and DeMets, 3rd Edn, 1998
That’s what Textbook #3 says...

“Q: Should the effectiveness of the treatment masking be assessed when the trial is over?

A: Yes . . . Guesses made by clinic staff and patients regarding treatment assignments can be used to make the assessments.”

Meinert, 1986
That’s what Textbook #4 says...

“The degree to which the double blind was truly effective should be evaluated by also asking the investigator to guess which treatment each patient received...

... The data... will indicate how well the blind was maintained and whether investigators, patients, or both were able to break it... “

Spilker, 1991
..which led me to pose 4 questions

1) Don’t trialists test for blindness and report its success in trial publications?

2) Must trialists test for blindness and report its success in trial publications?

3) Can’t trialists accurately test for blindness during and after trials?

4) Needn’t trialists test for blindness during and after trials?
For purposes of this discussion

I’m talking about superiority trials of drugs that either:

⇒ Don’t cause “tell-tale” side-effects, or

⇒ Their “tell-tale” side-effects can be replicated in their corresponding placebo.
I’m NOT talking about trials in which

- Blinding is impossible (most surgical and behavioural trials)

- Blinding is deemed so cumbersome as to obstruct other, more important elements of trial execution (debates over blinding our oral vs. I-V anticoagulation trials)

- . . . but we’ll re-visit unblinded trials later in this talk
Q1. Don’t trialists test for blindness and report its success in trial publications?

3 recent systematic reviews:
- Dean Fergusson et al: BMJ 2004;328:432
Sources for these SRs

- “Blinded trials from top journals”
- Medline, Cochrane registries and “high-impact factor” journals
- Cochrane Central Register of Controlled Trials
- Surprising non-overlap, with just 3 of 1880 trials appearing in more than one review.
How often did blinded trials report testing for blindness? 

only 2% to 8% of them!!
When carried out, how often did these tests conclude that blinding was **successful**?

💀💀 about half the time !!!

A1. Don’t trialists test for blindness and report its success in trial publications?

No, they usually don’t.

And when they do, blinding is judged unsuccessful half the time !!!
DON’T PANIC!

- We still have 3 questions to go

- Their answers might provide both an explanation for - and a solution to - these “failures”
Q2. **Must** trialists test for blindness and report its success in trial publications?

All 3 reviews **favored** “improved testing and reporting” . . . .

. . . . but **none** demanded it.
Q2. Must trialists test for blindness and report its success in trial publications?

CONSORT checklist formerly included:
“If [blinding] done, how the success of blinding was evaluated”

That item now removed from their checklist
A2. **Must** trialists test for blindness and report its success in trial publications?

Reviewers and authorities encourage it, but don’t demand it for all trials.
Q3. Can’t trialists accurately test for blindness during and after trials?

Why might testing for blindness be inaccurate (or even biased)?

Indeed, might there be an alternative explanation for an apparent loss of blindness when, in fact, it is preserved?
My alternative explanation for the apparent loss of blindness:

As soon as events/successes/failures start to occur, what if blindness becomes confounded with hunches about efficacy?
How hunches about efficacy might work:

 Might patients who do **well** get a hunch that they’re on the experimental treatment?

 Might patients who fare **badly** get a hunch that they’re on the control treatment?

 Might their clinicians do the same?
If this alternative explanation is correct

- When a trial is *positive* (Experimental Rx *works*),

  - but the end-of-study test for blindness really tests for *hunches* about efficacy,

  - and truly blind patients who did *well* guess *correctly* that they were on experimental Rx,

  - then Blinding will *always appear* to have been lost in a “positive” trial, even when it *hasn’t*
Popper challenges us to falsify this “hunch” hypothesis.

- We unwittingly tested it at the end of the first ever trial of aspirin for preventing stroke or death.
RRPCE ("Canadian Aspirin Trial")

- PICOT: Among patients with transient cerebral ischemic attacks (TIAs), can daily aspirin and/or sulfinpyrazone (Anturan) reduce the risk of stroke and death over the next 2 years?

- 585 patients (lost 1) in 26 centers suffered 85 strokes and 42 deaths

- Result: Aspirin was superior, both to the placebo and to Anturan (RRR = 31%)
At the end of the aspirin trial, but before we gave them the results

We gave each study clinician a list of their patients (\(n\) per clinician ~ 20)

And asked them to tell us which drug(s) they thought each patient had received
In this 2x2 factorial trial

Patients received:

1) **Active** Aspirin + Sulfinpyrazone’s **Placebo**
2) Aspirin’s **Placebo** plus **Active** Sulfinpyrazone
3) **Both** active drugs
4) **Both** placebos

So the probability of correctly guessing their patients’ treatment on the basis of chance = 25%
Conventional wisdom in interpreting our clinicians’ guesses

If correct guesses were >25%, conventional wisdom would conclude a loss of blinding.

Our clinicians’ correct guesses weren’t 25%

Our clinicians’ correct guesses were 18%

They were statistically significantly wrong!
Why our study clinicians were wrong

We'd also asked our clinicians to predict the relative efficacy of aspirin and sulfinpyrazone.

Most of them had a hunch that sulfinpyrazone was superior to aspirin.

In fact, the actual result was the exact opposite (aspirin was superior!).
As a result of their incorrect hunches about efficacy:

When a patient had done well, their clinician tended to guess that they’d been randomized to sulfinpyrazone.

When a patient had stroke or died, their clinician tended to guess that they’d been randomized to aspirin or the double placebo.
End-of-study tests for blindness can’t

We had not been testing our clinicians’ for blindness

We’d been testing, at least in part, for their (incorrect) hunches about efficacy

If their hunches had been correct, they’d have correctly predicted a greater than chance (P<0.05) number of regimens, and we’d have concluded that our trial had become unblinded
Trial participants have hunches too

Tom Chalmers tested **Vitamin C vs. placebo** for URI prophylaxis and symptoms  
(JAMA 1975;231:1038-42)

“No time to design, test, and manufacture a **placebo** indistinguishable from **Vitamin C**”

Participants: NIH employees  
(in trouble from the start)
Trouble:

“Early in the study, we discovered that some of the volunteers had tasted the contents of their capsules . . . .

. . . . and professed to know whether they were taking the Vitamin C or the placebo.”
Trouble:

By the end of the trial, 45% of the Placebo group claimed they knew their Rx.
- and 44% of the Placebo group had dropped out.

By the end of the trial, 52 % of the Vitamin C group claimed they knew their Rx.
- and only 34% of the Vitamin C group had dropped out.
Of those who thought they knew:

<table>
<thead>
<tr>
<th></th>
<th>51 thought they were on Vitamin C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Really were on</td>
<td>40</td>
</tr>
<tr>
<td>Vitamin C</td>
<td></td>
</tr>
<tr>
<td>Really were on</td>
<td>11</td>
</tr>
<tr>
<td>Placebo</td>
<td></td>
</tr>
</tbody>
</table>
Of those who thought they knew:

<table>
<thead>
<tr>
<th>Thought they were on</th>
<th>Really were on</th>
<th>Really were on</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vitamin C</td>
<td>Vitamin C</td>
<td>Placebo</td>
</tr>
<tr>
<td>51</td>
<td>40</td>
<td>12</td>
</tr>
<tr>
<td>Placebo</td>
<td>11</td>
<td>39</td>
</tr>
</tbody>
</table>
Reported Severity of Symptoms:
Reported Severity of Symptoms:

<table>
<thead>
<tr>
<th>And thought they were on Vitamin C</th>
<th>And thought they were on Placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Really were on Vitamin C</td>
<td>Less severe vs. More severe</td>
</tr>
</tbody>
</table>
Reported Severity of Symptoms:

<table>
<thead>
<tr>
<th>And thought they were on Vitamin C</th>
<th>And thought they were on Placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Really were on Vitamin C</td>
<td>Less severe vs. More severe</td>
</tr>
<tr>
<td>Really were on Placebo</td>
<td>Less severe vs. More severe</td>
</tr>
</tbody>
</table>
Reported mean duration of URIs:
Reported mean duration of URIs:

<table>
<thead>
<tr>
<th></th>
<th>Blinded Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Really were on</strong></td>
<td></td>
</tr>
<tr>
<td>Vitamin C</td>
<td>6.6 days</td>
</tr>
<tr>
<td><strong>Really were on</strong></td>
<td></td>
</tr>
<tr>
<td>Placebo</td>
<td>6.3 days</td>
</tr>
</tbody>
</table>
Reported mean duration of URIs:

<table>
<thead>
<tr>
<th></th>
<th>Blinded Participants</th>
<th>Unblinded Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Really were on Vitamin C</td>
<td>6.6 days</td>
<td>4.7 days</td>
</tr>
<tr>
<td>Really were on Placebo</td>
<td>6.3 days</td>
<td>8.1 days</td>
</tr>
</tbody>
</table>
A3. Can’t trialists accurately test for blindness during and after trials?

No, they can’t.

Tests for blindness fail as soon as:

i. study patients start having successes or failures, or know/think they know their Rx

ii. study patients and their clinicians start developing corresponding hunches about efficacy
A3. Trialists can’t accurately test for blindness during and after trials

Because end-of-study tests for blindness confound blindness with hunches about efficacy . . .

. . . an unblinded trial is indistinguishable from a blinded trial whose participants’ hunches about efficacy are correct.
Q4. Needn’t trialists test for blindness during and after trials?

Let’s return to a consideration of first principles

Why do we blind study patients and their clinicians in RCTs?
Why do we blind patients and clinicians in RCTs?

1. To prevent control patients from receiving the experimental Rx = **CONTAMINATION**.

Eg: A patient (or their clinician) learns that they are receiving a placebo, and seek the experimental Rx outside the trial.
Why do we blind patients and clinicians in RCTs?

2. To prevent one group (esp. controls) from receiving more extraneous efficacious Rx than the other group = Co-INTERVENTION

Eg: A patient (or their clinician) learns that they are receiving a placebo and they seek a non-study efficacious Rx.
Why do we blind patients and clinicians in RCTs?

3. To prevent hunches about efficacy from affecting the reporting (by study patients) or documentation (by study clinicians) of events =

BIASED OUTCOME REPORTING

Eg: When hunch = Active Rx, the under-reporting of “softer” or transient events, or reporting them to be due to some extraneous cause.
BIASED OUTCOME REPORTING

RCT of cyclophosphamide, prednisone, and plasma exchange in multiple sclerosis:

Treatment benefits (lots of “soft” signs) determined by both unblinded and blinded neurologists.

Unblinded neurologists found a benefit from triple Rx at 6, 12, and 24 months.

Blinded neurologists did not.

My proposition: Why not regard blinding as a mere process?

My contention: Once an RCT is underway, the measurement of blinding is not important in itself.

The important measurements test for the consequences of its loss: contamination, co-intervention, and biased outcome reporting.
If there is no risk of:

- **Contamination** (can only receive the experimental Rx within the trial)
- **Co-intervention** (doesn’t exist)
- **Biased outcome reporting** (total mortality)

⇒ You don’t need to blind anybody!
But when blinding *is* important

We must test for the consequences of its *loss*: contamination, co-intervention, and biased outcome reporting.
Testing for the consequences of the loss of blindness

1. **Contamination** in the Aspirin Trial?
   - Tested for a pathognomonic effect of aspirin on platelet aggregation.
   - Tested for a pathognomonic effect of sulfinpyrazone on serum uric acid.

Conclusion: No contamination of any group.
Testing for the consequences of the loss of blindness

2. **Co-intervention** in the NASCET Carotid Endarterectomy Trial?

   Concern: More vigorous antihypertensive Rx for patients who didn’t get the operation.

   I monitored blood pressures of all patients at all follow-up visits. Pestering letters for ↑BP.

   Conclusion: No BP co-intervention
Testing for the consequences of the loss of blindness

3. **Biased Outcome Reporting** of strokes in the NASCET Carotid Endarterectomy Trial?

Check-list including mild signs and symptoms (as well as events) at every follow-up visit.

Records of events and some non-events purged of Rx and sent to blind adjudicators.

Included a “rigor”ous outcome: rigor mortis.

Conclusion: No biased outcome reporting.
A4. *Needn’t* trialists test for blindness during and after trials?

*Don’t need to* (and usually *can’t*) test for blindness during/after a trial.

*Do need to* test for the consequences of its loss (contamination, co-intervention, biased outcome reporting).
If you need to blind in your RCT:

1. Achieve blindness in pre-trial pilots.

2. Don’t test for blindness during the trial (test for contamination and co-intervention)

3. Don’t kid yourself that you’re testing for blindness at the end of the trial (you’re also testing for hunches about efficacy)
RAPPORT
DES COMMISSAIRES
CHARGÉS PAR LE ROI,
DE L'EXAMEN
DU
MAGNÉTISME ANIMAL.
Imprimé par ordre du Roi.

À PARIS,
DE L'IMPRIMERIE ROYALE.
M. DCC LXXXIV.