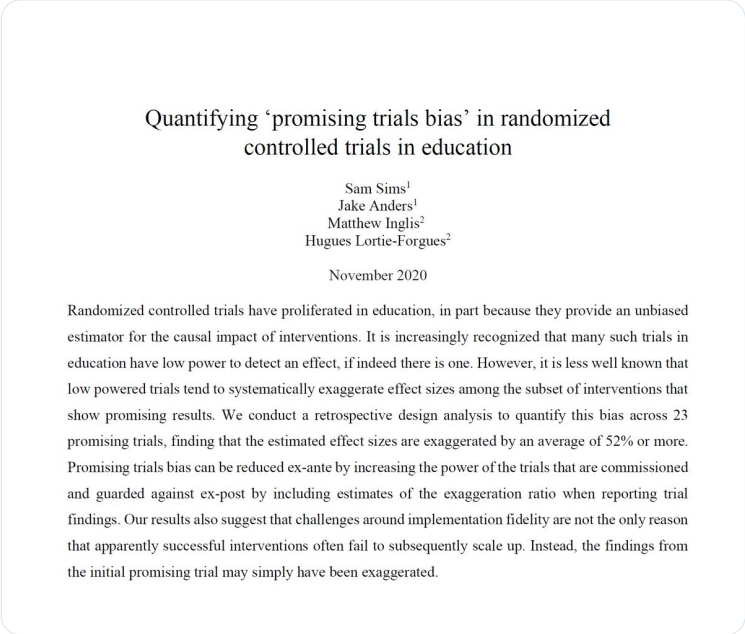
Sam Sims @DrSamSims

17 Nov 20 • 7 tweets • DrSamSims/status/1328762506232406018

RCTs tell us 'what works', right? They give us unbiased causal estimates, right?

New paper with @jakeanders @mjinglis & Hugo Lortie-Forgues: bit.ly/32VdEvL

👇



Quantifying 'promising trials bias' in randomized controlled trials in education

Sam Sims[1]
Jake Anders[1]
Matthew Inglis[2]
Hugues Lortie-Forgues[2]

November 2020

Randomized controlled trials have proliferated in education, in part because they provide an unbiased estimator for the causal impact of interventions. It is increasingly recognized that many such trials in education have low power to detect an effect, if indeed there is one. However, it is less well known that low powered trials tend to systematically exaggerate effect sizes among the subset of interventions that show promising results. We conduct a retrospective design analysis to quantify this bias across 23 promising trials, finding that the estimated effect sizes are exaggerated by an average of 52% or more. Promising trials bias can be reduced ex-ante by increasing the power of the trials that are commissioned and guarded against ex-post by including estimates of the exaggeration ratio when reporting trial findings. Our results also suggest that challenges around implementation fidelity are not the only reason that apparently successful interventions often fail to subsequently scale up. Instead, the findings from the initial promising trial may simply have been exaggerated.

It's increasingly recognised that RCTs are often underpowered (too small) to detect effects of the size we find in education: bit.ly/3pDeiYA

It's less well known that underpowered RCTs also lead to exaggerated effect sizes (ES) among RCTs with promising results...

We ask: If an RCT yields 'promising' results (eg, ES=0.15, p<0.05), what does that really tell us?

Our answer: ES from such trials are exaggerated by 50% on average, and the probability that the true ES=0 is surprisingly large - certainly much larger than 5%.

Besides these empirical findings, our paper tries to provide a clear explanation of how this happens in low powered trials - including nice diagrams 📈

You can read the @CEPEO_UCL working paper here: https://repec-cepeo.ucl.ac.uk/cepeow/cepeowp20-16.pdf

This is a problem for edu research:

- Clearinghouses/toolkits of effective practices often rely on results from single promising RCTs. Are they reliable?

- Exaggerated ES are a recipe for RCT findings that don't replicate. Are we misallocating funding to scale-up trials?

So what should we do?

1) Increase RCT power (hard but essential)

2) Use other designs (e.g. matching) alongside RCTs

3) Use 'design analysis' when planning and when reporting results from RCTs: bit.ly/3fapnv2

@threadreaderapp unroll

• • •