**Canadian Medical Education Journal**
**Revue canadienne de l'éducation médicale**

# Resist the Randomized Controlled Trials fetish: Different questions require different pyramids of evidence

# Résistez à l'obsession pour les essais contrôlés randomisés : différentes questions nécessitent différentes hiérarchies de preuves

Raywat Deonandan

Volume 16, Number 1, 2025

URI: https://id.erudit.org/iderudit/1117323ar
DOI: https://doi.org/10.36834/cmej.80355

See table of contents

Publisher(s)

Canadian Medical Education Journal

ISSN

1923-1202 (digital)

Explore this journal

Cite this document

Deonandan, R. (2025). Resist the Randomized Controlled Trials fetish: Different questions require different pyramids of evidence. *Canadian Medical Education Journal / Revue canadienne de l'éducation médicale*, *16*(1), 125–126. https://doi.org/10.36834/cmej.80355

Canadian Medical Education Journal

*Commentary and Opinions*

# Resist the Randomized Controlled Trials fetish: different questions require different pyramids of evidence

## Résistez à l'obsession pour les essais contrôlés randomisés : différentes questions nécessitent différentes hiérarchies de preuves

*Raywat Deonandan*[1]

[1]University of Ottawa, Ontario, Canada

Correspondence to: Dr. Raywat Deonandan; email: ray@deonandan.com

Published ahead of issue: Feb 24, 2025; published: Feb 28, 2025. CMEJ 2025, 16(1) Available at https://doi.org/10.36834/cmej.80355



In May of this 2024, a narrative review by Greenhalgh et al.[1] on the effectiveness of masks against respiratory infections made media waves. Their review found that masks work, are not harmful, and that respirators work best. These findings were not surprising to those of us who have been following the data on mask effectiveness for the duration of the COVID-19 pandemic. But they did contradict a prominent narrative among some medical circles, who had been relying upon more traditional meta studies of masking, such as a 2023 Cochrane review that famously found, "uncertainty about the effects of face masks," after relying solely upon randomized controlled trials (RCTs).[2] Thus, for those who treat the standard "evidence pyramid" as sacrosanct, masks were at best unproven and at worst ineffective and harmful—based upon the "best" randomized evidence.

Therein lies the true value of the Greenhalgh review, as it lays bare an intellectual fragility too common in the parsing of medical evidence: an overreliance upon, and indeed fetishization of, the RCT. We teach in medical schools that RCTs are at or near the top of the storied evidence pyramid, with observational studies, lab studies, and personal opinion ranked much lower. This is a dangerously oversimplified view of evidence that few clinicians not regularly engaged in research take the time to consider.

RCTs are most commonly used in medical, economic, and sociological research. They are almost unheard of in physical sciences or engineering. The reason for this is that RCTs are meant to distill a pure effect size in scenarios with poor signal-to-noise ratio, where the noise is the variation in response from individuals in the study. Human behaviour and human biology are amongst the most profound sources of such noise, hence statisticians' reliance upon known distributions, such as the Normal or Poisson curves. The magic of the Central Limit Theorem constrains approximate measurements of almost all human traits to conform to the Normal distribution, allowing us to use traditional statistical tools to compute useful outcomes like the average effectiveness of a drug or social policy.

RCTs are therefore ideal for measuring the safety and efficacy of pharmacological interventions. Different people with different biological histories and genetics will react differently. But with a large enough sample size, an average reproducible signal will emerge that will allow us to prescribe that same intervention to the greater population, with high confidence both that a great majority will see benefits and that a tiny minority will be harmed.

But engineering interventions, such as seatbelts, parachutes, and gasmasks, do not experience individual variation in effectiveness, beyond manufacturing variations. For such interventions, the ultimate test of efficacy is laboratory trials for which sample sizes are miniscule. There are famously no RCTs for parachutes, except tongue-in-cheek ones.[3] Instead, any RCTs for such interventions will not test efficacy per se, but rather

compliance and proper use, because that is how the human variation is expressed herein: through behaviour.

The proper use of RCTs in the masking question, then, is to test whether mask *mandates* or *policies* are effective, which is a very different question from whether masks in and of themselves are effective. After all, no military has ever conducted an RCT to test the effectiveness of battlefield gasmasks, because their worth was already proven in lab tests. Any such RCT would only measure compliance, which would be assumed to be near perfect in a regimented military scenario.

It can be argued that systematic reviews whose included studies are limited to RCTs offer the added benefit of propensity scoring. But much work has been done on the development of such scoring for observational studies, as well,[4] allowing researchers to mimic some of the characteristics of an RCT, including the transparent appreciation and consideration of the innate weaknesses of observational studies, such as potential biases and confounding factors. And it is certainly not unheard of for reviews to include both RCTs and observational studies.[5]

We must stress in medical schools that while Evidence Based Medicine is still the best avenue for clinical decision making, not all medical questions are subject to the same pyramid of evidence. Different fields of science rationally rank different study designs higher. The role of randomization is largely to control for biological and behavioural variation and thus have a trivial role where such variation is absent. The RCT is not magical or sacrosanct and is certainly not always the best design for all instances.

**Conflicts of Interest:** None
**Funding:** None
**Edited by:** Patricia Vohl (section editor); Marcel D'Eon (editor-in-chief)

# References

1. Greenhalgh T, MacIntyre CR, Baker MG, et al. Masks and respirators for prevention of respiratory infections: a state of the science review. *Clin Microbiol Rev*. 2024;37(2):e00124-23. https://doi.org/10.1128/cmr.00124-23
2. Jefferson T, Dooley L, Ferroni E, et al. Physical interventions to interrupt or reduce the spread of respiratory viruses. *Cochrane Database Syst Rev.* 2023;1(1):Cd006207. https://doi.org/10.1002/14651858.CD006207.pub6
3. Yeh RW, Valsdottir LR, Yeh MW, et al. Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial. *BMJ.* 2018;363:k5094. https://doi.org/10.1136/bmj.k5094
4. Austin PC. An Introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res*. 2011;46(3):399-424. https://doi.org/10.1080/00273171.2011.568786
5. Cheurfa C, Tsokani S, Kontouli K-M, Boutron I, Chaimani A. Synthesis methods used to combine observational studies and randomised trials in published meta-analyses. *Syst Rev.* 2024;13(1):70. https://doi.org/10.1186/s13643-024-02464-w