# Buying external validity at the expense of internal validity: A deal with the devil?

The lack of 'external validity' of randomised clinical trials – their inability to support conclusions that are generalisable beyond the study – has become a popular argument against the absolute epistemic superiority of RCTs. However, it is generally acknowledged that external validity only comes at the expense of 'internal validity' – the trustworthiness of the causal conclusion within the study. Moreover, the internal validity of RCTs is the very reason why this method ranks at the top of the hierarchies. The criticism of poor external validity is thus only convincing to the extent that it is substantiated by an argument that external validity is preferable to internal validity. In my contribution, I will provide a methodological account of the internal-external validity distinction and show that for many clinical purposes it is rational to buy external validity at the expense of internal validity.

Although the internal-external validity distinction has been introduced about six decades ago (Campbell 1957), the notion of 'external validity' has undergone little systematic development. Researchers primarily focus on aspects that limit external validity (Rawlins 2008; Rothwell 2006, 2005; Zuidgeest et al. 2017), while philosophers mostly regard external validity as synonymous with the idea of extrapolation. Thus, philosophical literature on external validity primarily discusses inferential strategies to generalise clinical trial results to a broader population (Cartwright 2007; Guala 2003; Reiss 2019; Fuller 2019; Stegenga 2018). These discussions have certainly revealed important shortcomings of wide-spread reasoning practices, yet some also fuelled the idea that internal validity should be prioritised: Francesco Guala, for example, claimed that it does not even make sense to bring up the question of external validity without being confident about internal validity (Guala 2003).

In my contribution, I first clarify the internal-external validity distinction by arguing that they are constituted by clusters of design properties of an experimental design. I show that the properties constituting internal validity justify highly accurate causal conclusions, while those constituting external validity justify the similarity of the experimental context to a target context. Moreover, being a cluster of properties, researchers can consequently decide for each of the properties whether to increase internal or external validity (Godwin et al. 2003). In the second part, I discuss a variety of design choices that increase external validity at the expense of internal validity. I will argue that some of these experimental designs can be understood as implementing variations of an ideal intervention that are discussed in the literature as 'fat-handed' or 'soft' interventions (Korb et al. 2004; Eberhardt and Scheines 2007; Scheines 2005). I further show that such experimental designs license inferences to causal hypotheses that are less accurate, in the sense that they identify an interactive cluster of causes instead of a single cause, while inferences to such cluster causes are nevertheless highly justified. I finally defend that clinical researchers should embrace the loss of accuracy because it amounts to an important benefit: such cluster causes constitute causal entities that are meaningful for clinical practice like, for example, 'the taking of a treatment, or 'the prescription of a treatment'. I conclude by pointing out that buying external validity at the cost of internal validity is particularly desirable for clinical research on medical treatments for broad populations.

## References


Campbell, D. T. (1957). *Factors relevant to the validity of experiments in social settings* (Vol. 54).

Cartwright, N. (2007). Are RCTs the Gold Standard? *BioSocieties, 2*, 11–20. doi:10.1017/S1745855207005029.

Eberhardt, F., & Scheines, R. (2007). Interventions and Causal Inference. *Philosophy of Science, 74*, 981–995. doi:10.1086/525638.

Fuller, J. (2019). The myth and fallacy of simple extrapolation in medicine. *Synthese, 15*, 15. doi:10.1007/s11229-019-02255-0.

Godwin, M., Ruhland, L., Casson, I., MacDonald, S., Delva, D., Birtwhistle, R., et al. (2003). Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity. *BMC Medical Research Methodology, 3*, 28. doi:10.1186/1471-2288-3-28.

Guala, F. (2003). Experimental Localism and External Validity. *Philosophy of Science, 70*, 1195–1205. doi:10.1086/377400.

Korb, K. B., Hope, L. R., Nicholson, A. E., & Axnick, K. (2004). Varieties of Causal Intervention. In C. Zhang, H. W. Guesgen, & W.-K. Yeap (Eds.), *PRICAI 2004: Trends in Artificial Intelligence*: *8th Pacific Rim International Conference on Artificial Intelligence, Auckland, New Zealand, August 9-13, 2004. Proceedings* (Vol. 3157, pp. 322–331, Lecture Notes in Computer Science, Vol. 3157). Berlin, Heidelberg: Springer.

Rawlins, M. (2008). De Testimonio: on the evidence for decisions about the use of therapeutic interventions. *Clinical medicine (London, England), 8*(6), 579–588.

Reiss, J. (2019). Against external validity. *Synthese, 196*, 3103–3121. doi:10.1007/s11229-018-1796-6.

Rothwell, P. M. (2005). External validity of randomised controlled trials: "To whom do the results of this trial apply?". *The Lancet, 365*, 82–93. doi:10.1016/S0140-6736(04)17670-8.

Rothwell, P. M. (2006). Factors that can affect the external validity of randomised controlled trials. *PLoS clinical trials, 1*, e9. doi:10.1371/journal.pctr.0010009.

Scheines, R. (2005). The Similarity of Causal Inference in Experimental and Non-experimental Studies. *Philosophy of Science, 72*, 927–940. doi:10.1086/508950.

Stegenga, J. (2018). *Medical Nihilism*. New York.

Zuidgeest, M. G. P., Goetz, I., Groenwold, R. H. H., Irving, E., van Thiel, G. J. M. W., & Grobbee, D. E. (2017). Series: Pragmatic trials and real world evidence: Paper 1. Introduction. *Journal of clinical epidemiology*, 7–13. doi:10.1016/j.jclinepi.2016.12.023.