The Statistical Properties of RCTs

Erik van Zwet[1], Simon Schwab[2,3], Stephen Senn[4]

[1] Department of Biomedical Data Sciences, Leiden University Medical Center, the Netherlands
[2] Center for Reproducible Science, University of Zürich, Switzerland
[3] Epidemiology, Biostatistics and Prevention Institute, University of Zürich, Switzerland
[4] Statistical Consultant, Edinburgh, United Kingdom

E-mail for correspondence: E.W.van_Zwet@lumc.nl

Abstract: We abstract a "study" as a triple (β, b, s) where β is the parameter of interest, b is an unbiased, normally distributed estimate of β, and s is the standard error of b. We do not observe β, but we do observe the pair (b, s). We define the z-value $z = b/s$ and the signal-to-noise ratio SNR=β/s. Note that the z-value is the sum of the SNR and independent standard normal "noise". This means that the distribution of the z-value is the convolution of the distribution of the SNR with the standard normal density.

We have collected a very large sample of pairs (b, s) from randomized controlled trials (RCTs) in the Cochrane Database of Systematic Reviews. We used these pairs to estimate the distribution of the z-values. Next, we obtained the distribution of the SNRs by *deconvolution*. Since we already know the conditional distribution of the z-value given the SNR, we now have the joint distribution of the pair (z, SNR).

Many important statistical quantities depend on (β, b, s) only through the pair (z, SNR). In particular, the exaggeration ratio $|b|/|\beta|$ and the indicator variables for the events: $\{|b|/s > 1.96\}$, $\{b - 1.96\, s < \beta < b + 1.96\, s\}$ and $\{\text{sign}(b) \neq \text{sign}(\beta)\}$. These quantities are closely related to the type M (magnitude) error, achieved power, coverage and type S (sign) error, respectively. We have computed their distribution across the Cochrane database both unconditionally and conditionally on the observed z-value. We find that the achieved power is often low and the exaggeration is typically large. However, conditionally on statistical significance, the probability of a type S (sign) error appears to be quite low.

Key words: Power; coverage; type M error; type S error; Cochrane database

Schwab S (2020). Re-estimating 400,000 treatment effects from intervention studies in the Cochrane Database of Systematic Reviews [Data set]. https://doi.org/10.17605/OSF.IO/XJV9G.

van Zwet E, and Gelman A (2020). A Proposal for Informative Default Priors Scaled by the Standard Error of Estimates. http://arxiv.org/abs/2011.15037.

van Zwet E, Schwab S, and Senn S (2020). The Statistical Properties of RCTs and a Proposal for Shrinkage. http://arxiv.org/abs/2011.15004/.