

The frequent insignificance of a “significant” P-value

David McGiffin¹, Geoff Cumming², and Paul Myles¹

¹Alfred Hospital

²La Trobe University

September 24, 2021

Abstract

Null hypothesis significance testing (NHST) and p-values are widespread in the cardiac surgical literature but are frequently misunderstood and misused. The purpose of the review is to discuss major disadvantages of p-values and suggest alternatives. We describe diagnostic tests, the prosecutor’s fallacy in the courtroom, and NHST, which involve inter-related conditional probabilities, to help clarify the meaning of p-values, and discuss the enormous sampling variability, or unreliability, of p-values. Finally, we use a cardiac surgical database and simulations to explore further issues involving p-values. In clinical studies, p-values provide a poor summary of the observed treatment effect, whereas the three- number summary provided by effect estimates and confidence intervals is more informative and minimises over-interpretation of a “significant” result. P-values are an unreliable measure of strength of evidence; if used at all they give only, at best, a very rough guide to decision making. Researchers should adopt Open Science practices to improve the trustworthiness of research and, where possible, use estimation (three-number summaries) or other better techniques.

Hosted file

P Value Paper 2021_FINAL.docx available at <https://authorea.com/users/387368/articles/538296-the-frequent-insignificance-of-a-significant-p-value>