

AB034. The danger of relying on the interpretation of P values in single studies: irreproducibility of results from clinical studies

Ronald L. Thomas, Paul R. Barach,
James D. Wilkinson, Ahmad A. Farooqi,
Steven E. Lipshultz

Department of Pediatrics, Wayne State University School of Medicine, Children's Hospital of Michigan, Children's Research Center of Michigan, Detroit, MI, USA

Correspondence to: Ronald L. Thomas, PhD. Department of Pediatrics, Wayne State University School of Medicine, Children's Hospital of Michigan, Children's Research Center of Michigan, 3901 Beaubien, Detroit, MI 48202, USA. Email: rthomas@med.wayne.edu.

Abstract: P values are a common component and outcome measure in most every published observational or randomized clinical trial. However, many physicians, researchers, journalists, and policy makers have little or no training in statistics and are forced to rely on the interpretation of results based solely on the authors or secondary sources. Statistical analysis of data often involves the calculation and reporting of the P value as statistically significant or not, without much further thought. But P values are highly un-replicable and their definition is not directly associated with reproducibility.

Findings from clinical studies are not valid if they cannot be reproduced. Although other methodological issues relate to reproducibility the P value is arguably at the root of the problem. Many common misinterpretations and misuses of the P value are practiced. The American Statistical Association (ASA) recently published its first ever policy statement concerning their proper use and interpretation of P values for scientists and researchers. This policy statement addresses the misguided practice of interpreting study results based solely on the P value, given that it is often irreproducible in subsequent, similar studies. We investigated the irreproducibility of the P value by using simulation software and results reported from a published randomized control trial. We show that the probability of attaining another statistically significant P value varied quite widely on replication. We also show that power alone determines the distribution of p, and will vary with sample size and effect size. In conclusion, P values interpreted solely by themselves, can be misleading potentially leading to biased inferences from clinical studies.

Keywords: Interpreting P values; misunderstanding and misconceptions of P values; irreproducibility of P values; null hypothesis significance testing (NHST); Exploratory software for confidence intervals (ESCI); dance of the P values

doi: 10.21037/pm.2020.AB034

Cite this abstract as: Thomas RL, Barach PR, Wilkinson JD, Farooqi AA, Lipshultz SE. The danger of relying on the interpretation of P values in single studies: irreproducibility of results from clinical studies. *Pediatr Med* 2020;3:AB034.