Journal of Pediatric Surgery 55 (2020) 2247

Contents lists available at ScienceDirect

Journal of Pediatric Surgery

journal homepage: www.elsevier.com/locate/jpedsurg

Letter to the Editor

Sample size analysis

Dear Editor,

Our journal club enjoyed Nataraja et al.'s randomized trial: lavage vs suction in laparoscopic appendicectomies [1]. Section 1.4.1 considered two distributions (means 4.4 and 5.5 days, sd 1.29), and asked how large, equal sized, samples needed to be such that the chance of getting a p-value > 0.05 was about 20%:— this being the chance of making a Type II error, of wrongly accepting the null hypothesis when the parent populations differ. The Type II error is less well understood than the Type I error, represented by the p-value, being the chance of wrongly rejecting the null hypothesis when there is no difference between the parent populations. The authors determined a need for 50 patients in each sample; we disagree.

For teaching purposes, we generated cumulative normal distributions (means 4.4 and 5.5, sd 1.29) in Microsoft Excel and then generated random samples from each distribution of size n, for $n = 10$ to $n = 200$. We then compared equal sized samples with a Student's t-test to obtain a p-value for each n. We let Excel repeat these trials and t-tests 500 times for each n, and determined the proportion of each 500 which had a $p > 0.05$ (Type II errors). We did the same for $p > 0.01$ and repeated the exercise using lognormal distributions. In this way we estimated how often experimenters would wrongly accept the null hypothesis for each n, for the two levels of statistical significance and two types of distribution. We plotted our results in Fig. 1.

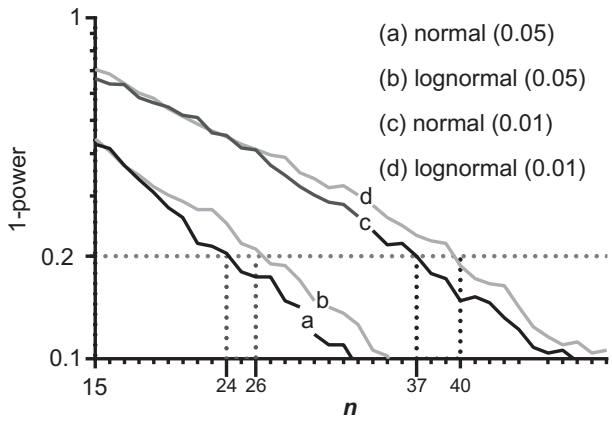

Fig. 1. Using a logarithmic y-axis, (1 − power) is plotted against n.

We found that each group required about 25 patients and not 50. There are mathematical short cuts to estimating sample size but our sledgehammer approach proved to be educational in our journal club. Do the authors, and their ethics committee, disagree with our sample size calculation?

When comparing length of stay the authors accepted the null hypothesis since $p = 0.75$, but what was the risk of them wrongly accepting this null hypothesis, and there really being a difference of means of 1.1 days? That risk is far lower than the 20% they set themselves, but how low? We realized we could estimate a post-hoc risk of making a Type II error if we did more trials. We found the post hoc risk to be about 1/5000, having found only one p-value greater than 0.75 for $n = 42$ in 5000 trials. We now ask if reporting the post hoc risk of a Type II error is helpful.

Anthony Lander*
Birmingham Women's and Children's Hospital, Steelhouse Lane, Birmingham B4 6NH, UK
* Corresponding author.
Email address: t.lander@nhs.net

Waaka Moni-Nwinia
Leeds Children's Hospital, Beckett Street, Leeds LS9 7TF, UK

https://doi.org/10.1016/j.jpedsurg.2020.05.047

Reference

[1] Nataraja RM, Panabokke G, Chang AD, et al. Does peritoneal lavage influence the rate of complications following pediatric laparoscopic appendicectomy in children with complicated appendicitis? A prospective randomized clinical trial. J Pediatr Surg. 2019;54:2524–7. https://doi.org/10.1016/j.jpedsurg.2019.08.039.