http://ijbsm.zbmu.ac.ir/

doi 10.34172/ijbsm.2021.13

# "*P* Value" or "Number Needed to Treat": Which One Is Better for Evaluating Clinical Treatments' Benefits?

Soudabeh Hamedi-Shahraki[1], Farshad Amirkhizi[2*]

[1]Department of Epidemiology and Biostatistics, Faculty of Public Health, Zabol University of Medical Sciences, Zabol, Iran
[2]Department of Nutrition, Faculty of Public Health, Zabol University of Medical Sciences, Zabol, Iran

*Correspondence to
Farshad Amirkhizi, Department of Nutrition, Faculty of Public Health, Zabol University of Medical Sciences, Zabol, Iran.
Email: amirkhizi.f@gmail.com

**Abstract**
Statistical significance does not necessarily mean clinical significance. A *P* value less than 0.05 does not guarantee the clinical effectiveness of a treatment. To assess the clinical valuable of a treatment, the effect size must be calculated. The number needed to treat (NNT) is an example of an effect size measure that can be very helpful in determining the clinical significance of a treatment. Therefore, it is recommended for all researchers and physicians to look beyond the *P* value and calculate the NNT for assessing the clinical significance of therapeutic measures and agents.
**Keywords:** Number needed to treat, *P* value, Effect size, Clinical trial

## To the Editor,

Most investigators use the existence of a significant difference between the treatment and control groups by considering "*P* < 0.05" to judge the effectiveness of a treatment used in randomized clinical trials (RCTs). However, it should be noted that statistically significant differences between study groups do not certainly mean clinical significance, and unfortunately focusing on "*P* values" may lead to publish irrelevant findings and confuse clinicians about the usefulness of an intervention.[1] For example, suppose we want to compare the effectiveness of drugs A and B in improving hypertension as the outcome of a study. If the percentages of the patients improved by drugs A and B are 42% and 40%, respectively, this 2% difference in the proportion of the patients improved may be statistically noticeable, especially when large sample sizes are recruited. However, despite the statistically significant difference between the efficacy of the two drugs based on "*P* < 0.05", clinicians may not observe any difference between the effectiveness of the two drugs in clinical practice. Indeed, although the

2% difference in the proportion of the patients who improved after receiving drugs A and B is statistically significant, it is not clinically noticeable. Therefore, the question arises as how to determine the clinical significance of interventions in RCTs.

To explore the clinical significance of any result found in RCTs, first, the effect size between the interventions performed in the study should be determined. The effect size of a treatment indicates how large the clinical response observed is in practice.[2,3] There are numerous measures for determining the effect size of binary outcomes in RCTs; one of which is the calculation of number needed to treat (NNT).[4] The NNT is a measure that is not only easy to calculate but also simple to understand and can show the clinical importance of interventions in RCTs.[4,5] The NNT is defined as the number of patients who would have to receive a treatment in order for one individual to benefit from the treatment or to avoid an adverse outcome over a given period of time.[6] The NNT indicates the efficacy by combining the baseline risk without treatment and the risk

reduction with treatment. Also, NNT helps physicians to understand how much attempt is required to forbid one event, thus allowing comparisons with the amount of attempt required to prevent the same events in patients with other disorders.

To calculate the NNT, you need to know the absolute risk reduction (ARR). In trials and other biomedical studies, results are often expressed in terms of the proportions of outcomes in the new treatment ($P_N$) and standard treatment ($P_S$) groups. The difference between $P_N$ and $P_S$ is called the "ARR" (ARR = $P_N$ - $P_S$). The ARR is therefore the amount by which your treatment reduces the risk of a bad outcome. The NNT is the inverse of the ARR: $NNT = \frac{1}{ARR}$. [6,7] It must be kept in mind that a large treatment effect in the ARR, leads to a small NNT. Indeed, a smaller "NNT" represents the more clinical significance of a treatment or an intervention.

For example, if 65% of the patients who received placebo and 90% of the patients who received the new treatment improved, then the absolute risk reduction would be 25% (ARR = 0.90–0.65), resulting in an NNT of 4 $\left(NNT = \frac{1}{0.25}\right)$ It is interpreted as that a clinician will need to treat four patients in order to prevent one adverse outcome.

As another example for the comparison of the effectiveness of two drugs (A and B) in the improvement of hypertension, which was mentioned above, the percentage of improvement with drug A was 42%, and that of drug B was 40%, showing a statistically significant difference. In this example, ARR would be 2%, resulting in an NNT of 50 $\left(NNT = \frac{1}{0.02}\right)$, meaning that for every 50 patients who receive drug A, one more patient would experience improvements in blood pressure compared to drug B, which is not clinically valuable. Nevertheless, although the difference between the remission proportions of the patients receiving the two drugs is statistically significant according to $P < 0.05$; however, the clinical efficacy of the drugs is not significantly different.

The optimal value of NNT is 1, which means all of the patients have a positive response to the treatment, and no patient responds to placebo. When a new treatment has the same effect as a placebo, the NNT would have a value of infinity (ARR = 0 and NNT = $\frac{1}{0}$).[8,9] Thus, in theory, the NNT can vary from 1 to infinity. The NNT may also be a negative value when the response rate for the placebo is higher than the treatment. In practice, the NNT is often used to compare the efficacy of treatments, so negative and infinite values are not usually reported because the treatment is considered ineffective. Therefore, the lower the NNT level (ie, the closer to 1), the more effective the treatment.[9]

Since $P$ value cannot be very helpful in determining the clinical significance, in order to quantify the clinical importance of a statistically significant result, it is better to calculate the NNT. In fact, the NNT is an effect size and is independent of statistical significance. So, it is recommended to all researchers and physicians to look beyond the $P$ value, and to respond to "Whether this result is clinically relevant or not", calculate the NNT. Anyway, clinical decisions in the real world are based on clinical importance, not just statistical significance.

## Ethical Approval
Not applicable.

## Competing Interests
The authors have no conflict of interest to declare.

## References
1. Citrome L. Call for papers for the International Journal of Clinical Practice. http://www.youtube.com/watch?v=KBALRk2hjMs.
2. Citrome L. Compelling or irrelevant? using number needed to treat can help decide. Acta Psychiatr Scand. 2008;117(6):412-419. doi:10.1111/j.1600-0447.2008.01194.x
3. Kraemer HC, Kupfer DJ. Size of treatment effects and their importance to clinical research and practice. Biol Psychiatry. 2006;59(11):990-996. doi:10.1016/j.biopsych.2005.09.014
4. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. N Engl J Med. 1988;318(26):1728-1733. doi:10.1056/nejm198806303182605
5. Kraemer HC, Kupfer DJ. Size of treatment effects and their importance to clinical research and practice. Biol Psychiatry. 2006;59(11):990-996. doi:10.1016/j.biopsych.2005.09.014
6. Citrome L. Show me the evidence: using number needed to treat. South Med J. 2007;100(9):881-884. doi:10.1097/SMJ.0b013e3180f63246
7. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. BMJ. 1995;310(6977):452-454. doi:10.1136/bmj.310.6977.452
8. Andrade C. The numbers needed to treat and harm (NNT, NNH) statistics: what they tell us and what they do not. J Clin Psychiatry. 2015;76(3):e330-333. doi:10.4088/JCP.15f09870
9. Altman DG. Confidence intervals for the number needed to treat. BMJ. 1998;317(7168):1309-1312. doi:10.1136/bmj.317.7168.1309