

Understanding Statistical Significance versus Clinical Significance: From the Perspective of a Study, "Impact of Cell-phone on Human Health"

Md. Nurul Amin¹

ABSTRACT

Background & objective: In clinical research, statistical significance (p -value < 0.05) is often reflexively equated with clinical significance as well. However, statistical significance merely indicates the likelihood that a result occurred by chance, whereas clinical significance reflects the practical value of the finding to patient care. This study clarified the difference between the connotations of the two frequently used terms in medical research – statistical significance and clinical significance.

Methods: The study utilized data from a PhD study (conducted during the period from July 2017– June 2022) involving 118 medical students. We measured hemodynamic variables (pulse, blood pressure, and parotid gland temperature) before and after a mobile phone call (talk-time intervention). Data were analyzed using paired-sample t -tests, Cohen's d effect sizes, and correlation coefficients.

Results: Post-intervention, local temperature increased significantly ($97.4 \pm 0.7^{\circ}\text{F}$ to $98.0 \pm 0.8^{\circ}\text{F}$, $p < 0.001$) with a large effect size ($d = 0.80$). Pulse rate also showed a statistically significant increase (89 ± 17 to 92 ± 17 bpm, $p = 0.009$), but with a negligible effect size ($d = 0.18$). No linear correlation was found between talk time and hemodynamic changes.

Conclusion: The findings demonstrate that while pulse rate changes reached statistical significance due to sample size sensitivity, they lacked clinical significance. Conversely, local temperature changes demonstrated both statistical and clinical significance.

INTRODUCTION:

In the evaluation of drug efficacy, clinicians utilize control groups or placebos to facilitate a comparative analysis between the test drug and the control intervention. When a test drug yields a statistically significant outcome (p -value < 0.05), it often leads to optimism regarding the drug's effectiveness in achieving the desired result. Clinicians frequently equate statistical significance with clinical significance; however, it is crucial to recognize that statistical significance does not

inherently imply clinical significance. This distinction highlights that results may be statistically significant without being clinically meaningful. Again, some researchers, in the absence of any statistically significant results, may still claim clinical significance based on a judgment about the relevance of a finding to his or her practice. In this study, we aim to explore such a scenario.

The case under discussion pertains to my PhD research titled "Impact of Cellphone on Human

Authors' information:

¹**Dr. Md. Nurul Amin**, PhD. Senior Research Fellow, International Diabetes Federation (IDF). Associate Professor (Research & Development) & Executive Editor (Ibrahim Cardiac Medical Journal), Ibrahim Cardiac Hospital & Research Institute, Shahbag, Dhaka, Bangladesh.

Correspondence: Dr. Md. Nurul Amin, Mobile: 01753178452, E-mail: mdamin01@yahoo.com

Health," conducted at the Institute of Environmental Science, University of Rajshahi, Bangladesh, from July 2017 to June 2022. The study included 313 participants, comprising 118 medical students from Rajshahi Medical College and 195 individuals from the University of Rajshahi. A component of this research focused on examining the immediate effects of phone-call intervention on hemodynamic variables (pulse, blood pressure, and local body temperature). Employing a within-sample pre-test/post-test design, the 118 medical college respondents were asked to make a call of their choosing for any duration. Their pulse, blood pressure, and local temperature over the parotid gland were recorded immediately before and after the call. The differences between these two datasets were statistically analyzed using paired-sample t-tests, with a significance level set at 5% (or 0.05) to assess the potential impact of radiofrequency radiation emitted during mobile phone calls on hemodynamic parameters.

Meaning of the word "Significant"

Generally, the word 'significant' means 'important', but in the world of statistics, it is a statement about the likelihood of a result being due to chance, or the amount of uncertainty we are prepared to accept, not its importance. In the world of clinical practice, whether a result is significant is based on its importance to, and implications for, practice; that is, the practical value of any particular result.

Statistical significance is a statement about the likelihood of findings being due to chance. Classical significance testing, with its reliance on p-values, can only provide a dichotomous result statistically significant, or not. Limiting interpretation of research results to p-values means that researchers may either overestimate or underestimate the meaning of their results. Very often the aim of clinical research is to trial an intervention with the intention that results based on a sample will generalize to the wider

population. The p-value on its own provides no information about the overall importance or meaning of the results to clinical practice, nor do they provide information as to what might happen in the future, or in the general population. Clinical significance is a decision based on the practical value or relevance of a particular treatment, and this may or may not involve statistical significance as an initial criterion.

Over the past few decades there has been considerable discussion relating to the interpretation and limitations of significance testing.¹⁻³ In recognition that classical significance testing cannot answer the 'so what' question, there have been calls for researchers to demonstrate the practical, or clinical significance of their findings. Many researchers now provide, in addition to either p-values or alpha levels (and sometimes both), measures of effect size and confidence intervals, which provide additional information as to whether any particular p-value returned by a test may have relevance to practice. A p-value is the probability of finding a result as extreme, or fantastic, or disappointing, as the one returned by a statistical test.

Ways to Solve the Problem?

Confidence intervals are one way for researchers to help decide if a particular statistical result (whether significant or not) may be of relevance in practice. Another way to solve the issue is calculation of effect-size using Cohen's d [Cohen's $d = (\text{Difference between the Two Means}) / (\text{Pooled Standard Deviation})$], where a d-value of ≤ 0.2 indicates a "small" effect, 0.3 – 0.5 represents a "medium" effect, and ≥ 0.8 signifies a "large" effect size. Thus, if the difference between the means of two groups is < 0.2 standard deviations, it is deemed negligible, even if statistically significant.

In the present study, results indicated a significant increase in local temperature, rising from 97.4 ± 0.7 °F at baseline to 98.0 ± 0.8 °F after a median

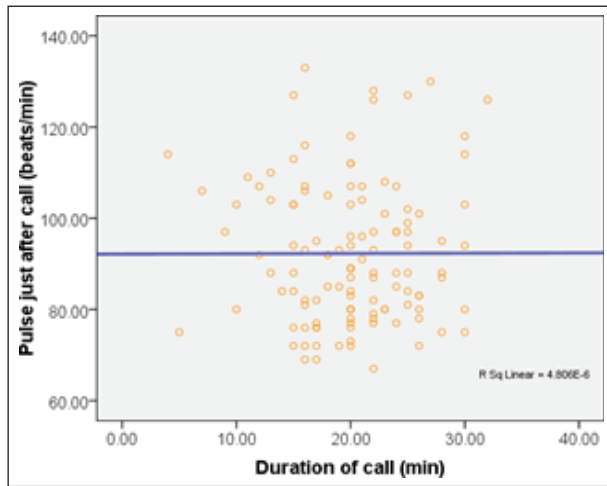


Fig. 1 shows the correlation between talk-time and pulse (n = 118)

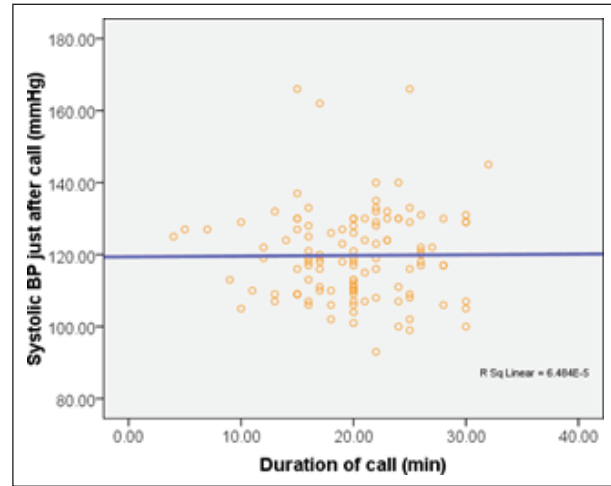


Fig. 2 depicts the correlation between talk-time and systolic blood pressure (n = 118)

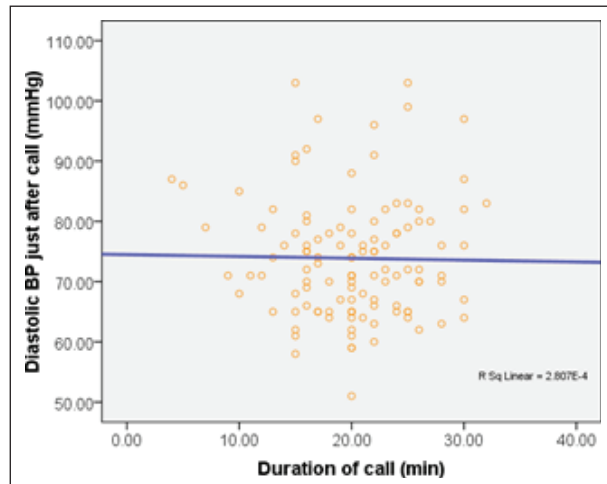


Fig. 3 illustrates the correlation between talk-time and diastolic blood pressure (n= 118)

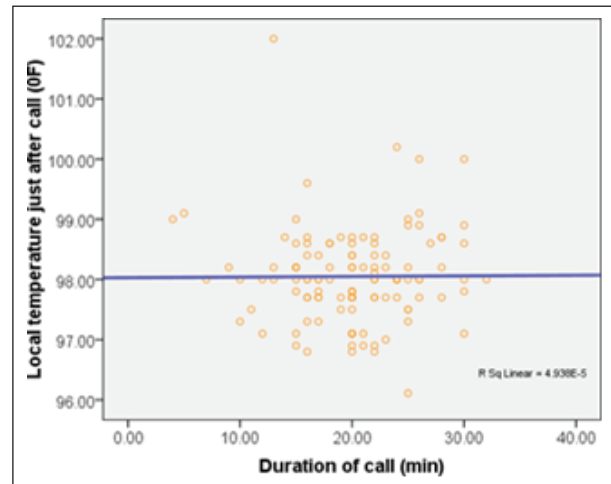


Fig. 4 shows the correlation between talk-time and local temperature (n = 118)

Figures 1-4 demonstrate correlations of talk-time with pulse, blood pressure and temperature (Amin et al¹)

talk time of 20 minutes (range: 4 – 32 minutes, $p < 0.001$). Correspondingly, pulse rates increased from 89 ± 17 beats per minute at baseline to 92 ± 17 beats per minute immediately following the call ($p = 0.009$) (see Table 1). However, none of the haemodynamic variables (local temperature, pulse, systolic and diastolic blood pressures) just after quitting the call were found to bear any significant correlations with talk-time ($r = 0.007$, $p = 0.940$, $r = 0.002$, $p = 0.981$, $r = 0.008$, $p = 0.931$ and $r = -0.017$, $p = 0.857$ respectively) (Fig. 1-4).

The effect sizes (Cohen’s d) for changes in pulse and local temperature post-intervention were calculated as 0.18 and 0.80, respectively as follows¹:

$$\text{Cohen's d pulse} = \frac{92 - 89}{17} = 0.18,$$

$$\text{Cohen's d for Temp.} = \frac{98 - 97.4}{7.5} = 0.8$$

Table 1: Changes in haemodynamic variables before and after a call

| Haemodynamic variables | Measure | | p-value |
|---|-----------------------|----------------------|---------|
| | Before call (n = 118) | After call (n = 118) | |
| Local temperature over parotid gland (0F) | 97.4 ± 0.7 | 98.0 ± 0.8 | < 0.001 |
| Pulse (beats/min) | 89 ± 17 | 92 ± 17 | 0.009 |
| Systolic blood pressure (mmHg) | 120.6 ± 12.5 | 119.3 ± 13.6 | 0.185 |
| Diastolic blood pressure (mmHg) | 74.1 ± 9.4 | 73.2 ± 9.9 | 0.777 |

*Data were analyzed using **Paired-sample t-Tests** and were presented as **mean ± SD**.

DISCUSSION

In this study, we aimed to investigate whether radiofrequency radiation emitted from smartphones during a call leads to an increase in local temperature and alterations in hemodynamic variables among mobile users. The analysis revealed a statistically significant increase in local temperature, averaging 0.60 °F, accompanied by a rise in heart rate of 3 beats per minute. However, no significant correlations were found between local temperature or pulse rate and talk time immediately after ending the call, suggesting variability in individual sensitivity to radiofrequency electromagnetic fields (RF-EMF).

Furthermore, the specific absorption rate (SAR), which quantifies the rate of radiofrequency emissions (measured in W/kg), differs among various mobile phone brands. This variability may contribute to the observed discrepancies in findings. If all individuals were equally sensitive to RF-EMF and all devices emitted uniform radiofrequency levels, one would expect to see a linear correlation between talk time and both local temperature and pulse rate. Additionally, a sample size fallacy was identified during the paired-sample t-test analyses; because the test compares two sets of data from the same 118 individuals, the mathematical power increases significantly, which may have contributed to a Type I error, producing "false positive" statistical significance for trivial variations in pulse and local

temperature following the phone call intervention. Effect size calculations (Cohen's *d*) indicate that while the intervention had a negligible effect on pulse rate ($d < 0.2$), it resulted in a large and significant effect on local temperature ($d = 0.8$).

The intersection of statistics and medical research often sparks debate regarding the statistical versus clinical significance of study findings, particularly in cases that are neither clearly positive nor negative. While studies can achieve statistical significance, they may lack clinical relevance, or conversely, results may be clinically important without being statistically significant. Understanding this complex issue requires a grasp of various aspects of statistical design and result interpretation.²

In contrast, clinical significance is inherently more subjective, encompassing factors such as observed effect size, safety profiles, and quality of life etc. Confidence intervals (CIs) can enhance the interpretation of clinical significance when presented alongside p-values, as they offer a range of possible values consistent with the observed effect size. In a well-structured clinical trial, statistical significance and clinical significance typically align when an appropriately defined target effect and sufficient power are chosen.³ Additionally, methods that expand the sample size based on interim findings⁴ may also result in achieving statistical significance without ensuring clinical significance.⁵

CONCLUSION

Although a clear framework exists for assessing statistical significance, clinical significance remains a subjective judgment made by clinicians and patients. Controversy often arises when clinical trial results are interpreted solely through p-values. Therefore, the use of CIs and consideration of effect sizes are encouraged to clarify the precision of study estimates. It is crucial to recognize that studies with large sample

sizes may more readily identify trivial yet statistically significant differences. If CIs encompass values that reflect meaningful effect sizes for patients, then clinical significance can be established.

REFERENCES:

1. Amin MN, Islam MSU, Haque MJ. Association between mobile phone use and physical wellbeing. *Ibrahim Card Med J* 2023;11(2): 8–15 doi:103329/icmjv11i266572
2. Dahlberg SE, Korn EL, Le-Rademacher J, Mandrekar SJ Clinical versus statistical significance in studies of thoracic malignancies. *J Thorac Oncol* 2020;15(9): 1406–1408 doi:101016/jjtho202006007
3. Cook JA, Fergusson DA, Ford I, et al There is still a place for significance testing in clinical trials. *Clin Trials* 2019;16:223–224.
4. Chen YH, Li C, Lan KK Sample size adjustment based on promising interim results and its application in confirmatory clinical trials. *Clin Trials* 2015;12: 584–595.
5. Freidlin B, Korn EL Sample size adjustment designs with time-to-event outcomes: a caution. *Clin Trials* 2017; 14:597–604.