*Corresponding author

Abdulrahman Alturki, Montreal Neurological Institute and Hospital, McGill university, Canada, National Neurosciences Institute, King Fahad Medical City, Saudi Arabia, Email: abdulrahman.alturki@mail.mcgill.ca

Editorial

The Value of P Value in the Medical Field

Abdulrahman Alturki[1,2]*

[1]*Montreal Neurological Institute and Hospital, McGill University, Canada*
[2]*National Neurosciences Institute, King Fahad Medical City, Saudi Arabia*

Editorial

Almost every statistical test generates a P value (or several). Yet, many physicians don't really understand what P values are. It is mentioned that P value is probably the most ubiquitous and at the same time, misunderstood, misinterpreted, and occasionally miscalculated index [1].

It is disappointing as some of our surgical and medical communities still use P values on many levels of result reporting (from departmental meetings and journal clubs to grand rounds, even in publications). To me, this disappointment reached its peak when we were asked in one of our clinical board exams to define P value and to interpret a clinical trial results based on it!

When Fisher first introduced it as a formal research tool, it was meant to be used as a rough numerical guide of the strength of evidence against the null hypothesis and to be used flexibly within the context of a given problem. He proposed the use of the term "significant" to be attached to small P values. In his nicely detailed review [2], Goodman discussed the historical debate and implications of P values and hypothesis tests.

Usually P value is used to determine the presence or absence of statistical significance despite it telling us nothing more than whether the P value is less than an arbitrary number "0.05". Statistically significant P value (P<0.05) is not informative about the data that we are analyzing. A P value may be viewed as the probability of obtaining an estimate at least as far from a specified value (most often the null value, i.e., the value of no effect) as the estimate we have obtained, if the specified (null or test) value were (note the subjunctive) the true value [3]. An operational interpretation of a P value less than 0.05 is that one should repeat the experiment, if subsequently repeated experiments also showed "significant" P values then one could reach a conclusion that the observed effects were unlikely to be the result of chance alone.

The critical point is not knowing what a P value actually is but understanding what a P value is not. Elaboration on the misconceptions of the P value is beyond the scope of this paper as it is discussed in depth in many publications. [1,4,5]. Table 1 lists some of the common misunderstandings about P values in the clinical setting and table 2 shows some uses of P values where they are of limited use.

Reporting and interpreting data in its clinical relevance is of particular importance in the medical field as the current methods that statisticians recommend express the results of studies in terms that are directly relevant to the clinical use to which they may be put. Many journals have endorsed the new orthodoxy of expecting authors to calculate confidence intervals whenever the data warrant this approach.

Confidence intervals will appear again and again in statistics. They are of the general form: estimate +/- margin of error. The estimate (and typically the margin of error as well) is computed from the sample data.
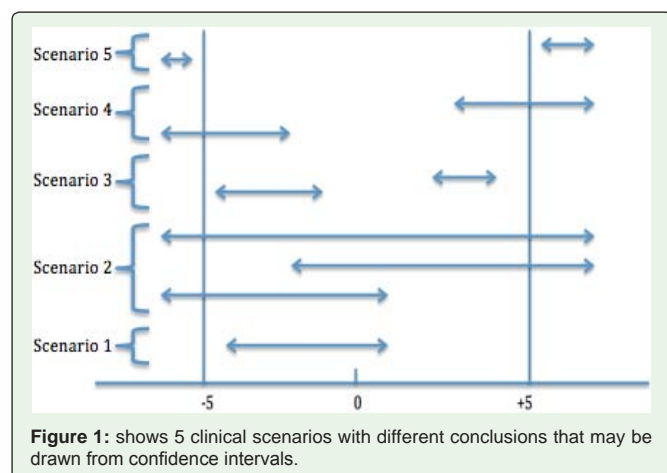
The confidence level chosen affects the width of the confidence interval through the size of the margin of error, and corresponds to the probability that the interval covers (includes) the true value of the parameter. Table 3 some properties of confidence intervals.

The obvious advantage of a confidence interval is that it expresses results in the units in which the measurements were made, and so allows the reader to consider critically the clinical relevance of the results.

There are 5 different conclusions that may be drawn from confidence intervals (figure 1):

1.   The CI includes point estimate (μ0), and neither end of the CI suggests anything clinically that might be interesting to me. I can conclude that this variable does not have any effect big enough to interest me. Scenario 1, no to minimal reduction (not clinically interesting) in systolic blood pressure in stroke patients receiving a newly released medication.

2.   The CI includes μ0, and one or both of the CI limits interests me clinically. The study might be important, but further study is needed. Scenario2, change (clinically interesting (harm or

Figure 1: shows 5 clinical scenarios with different conclusions that may be drawn from confidence intervals.

Table 2: Some uses of P values where they are of limited use.

Baseline characteristics table for studies
Describing the strength of an effect
Choosing what to adjust for based on the observed associations in the data
Post-hoc analyses
Multiple comparisons
Stepwise regression

Table 3: Some properties of confidence intervals.

Decreasing the confidence level, $1 - \alpha$, decreases the width of the CI
Increasing the significance level, α, decreases the width of the CI
Increasing sample size, decreases the width of the CI
To halve the width of the CI, sample size must be increased by a factor of 4
Decreasing the population variance, decreases the width of the CI

benefit) at some point, ±5 mm hg) in systolic blood pressure in stroke patients receiving a newly released medication.

3. The CI does not include μ0, but both limits of the CI are too near zero to interest me clinically. Conclude that this variable does not have any effect big enough to interest me. Scenario3, minimal change (did not reach clinical interest, ±5 mm hg) in systolic blood pressure in stroke patients receiving a newly released medication.

4. The CI does not include μ0, and some of the values inside the limits interest me clinically. The study suggests the variable may be important to me, but further study is needed. Scenario4, definitive change (clinically interesting at some point, ±5 mm hg) in systolic blood pressure in stroke patients receiving a newly released medication.

5. The CI does not include μ0, and all of the values inside the limits interest me clinically. Conclude that this variable is important. Scenario 5, definitive change (clinically interesting, ±5 mm hg) in systolic blood pressure in stroke patients receiving a newly released medication.

The advantage of confidence intervals in comparison to giving P values after hypothesis testing is that the result is given directly at the level of data measurement.

The confidence intervals are used and interpreted commonly in the medical field as surrogates of statistical significance, fixating on "statistical significance" could lead to ignoring the quantitative information provided by them i.e. the direction and strength of the effect.

Table 1: Common misunderstandings about P value in the clinical setting.

Statistically significant should be clinically significant " and vice versa"
Significance level, such as 0.05, is determined by the P value
It indicate the size or importance of the observed effect
It is the probability that the null hypothesis is true
It is the probability that the alternative hypothesis is false
If P value is less than "X" then null hypothesis has less than "X" % chance of being true

P-values are clearer than confidence intervals. It can be judged whether a value is greater or less than a previously specified limit. This allows a rapid decision as to whether a value is "statistically significant" or not. However, this type of "on spot diagnosis" can be misleading, as it can lead to clinical decisions solely based on "presumptive" statistics.

Statistical significance must be distinguished from medical relevance or biological importance. If the sample size is large enough, even very small differences may be statistically significant; on the other hand, even large differences may lead to non-significant results if the sample is too small. At the end, the clinician should be more interested in the size and direction of the difference e.g. in therapeutic effect between two treatment groups in clinical studies, as this is what is important for successful treatment, rather than whether the result is statistically significant or not.

In many cases published medical literature requires no firm decision: it contributes incrementally to an existing body of knowledge [4]. Working in a field where the life and well being of people is at stake, one should aim for carful data description and good interpretation of estimated effect measures rather than concrete significance testing with dichotomized answers which lead inherently to misleading interpretation of data.

References

1. Goodman S. A dirty dozen: twelve p-value misconceptions. Seminars in hematology. Elsevier. 2008; 135-140.

2. Goodman SN. P values, hypothesis tests, and likelihood: implications for epidemiology of a neglected historical debate. American Journal of Epidemiology. 1993; 137: 485-496.

3. Rothman KJ, Greenland S, Lash TL. Modern epidemiology. Lippincott Williams & Wilkins. 3rd edition. 2008.

4. Sterne JA, Smith GD. Sifting the evidence-what's wrong with significance tests? Physical Therapy. 2001; 81: 1464-1469.

5. Schervish MJ. P values: what they are and what they are not. The American Statistician. 1996; 50: 203-206.

Citation: Alturki A. The Value of P Value in the Medical Field. SM J Public Health Epidemiol. 2015;1(4):1020.

Page 2/2