Dear Editor,

Reporting and interpretation of results of classical statistical tests is widespread among applied researchers, most of whom erroneously believe that such tests are prescribed by a single coherent theory of statistical inference. This is not the case; classical statistical testing is a hybrid of the two different approaches formulated by Fisher on the one hand and Jerzy Neyman and Egon Pearson on the other.[1,2]

In Fisher’s approach, the researcher sets up a null hypothesis that a sample comes from a hypothetical infinite population with a known sampling distribution. The null hypothesis is said to be “disproved,” as Fisher called it, or rejected if the sample estimate deviates from the mean of the sampling distribution by more than a specified criterion, the level of significance. According to the theory, p value is the conditional probability than P (data/null hypothesis).[1‑3]

The Neyman–Pearson approach formulates two competing hypotheses, the null hypothesis and the alternative hypothesis. In a not-so oblique reference to Fisher, Neyman commented on the rationale for an alternative hypothesis. Specification of an alternative hypothesis critically distinguishes between the Fisherian and Neyman–Pearson methodologies.[1‑3]

p-value actually indicates the probability that whatever the difference came in assumed value (null hypothesis) and empirical value (data) is due to chance or sampling variation.[4,5] It can be found in a structure of clinical trial, suppose you analyzed a clinical trial data and found the p value is 0.2. It means that the probability of observed difference in efficacy of two drugs, which came out from clinical trial is resulting due to chance is 20%. If you repeat this experiment 100 times, then the difference in efficacy of two drugs reported in 20% of studies are due to the chance factor. In a clinical trial or any study, if the sample size is increased, then it will have less sampling error or error due to chance in result, which clearly indicates the lowering of p value. In general, increasing the sample size of study causes a decrease in the p value.[6,7]

In any experiment, it is mandatory to fix a level of sampling error or chance factor or p value up to the permissibility of sampling variability (whatever the difference come in study is due to chance factor), because majority of studies are based on the sample. If the difference in study is there it is said that, if the original p value is less than the prefixed level, then it means the difference occurred from the study has the less chance that the difference is due to sampling variability. It means the difference is the real and statistically valid. In general, the pre-fixed levels of p values in studies are 0.10, 0.05, 0.01 and 0.001. These prefixations depend on the researcher choice, literature of concern area and applicability of experiment. Suppose researcher describing their statistically significant (p < 0.05) result found a rate ratio 1.5 that is in new drug, the cure rate of disease is 1.5 times higher as compared to the standard drug group.

The above-mentioned paragraphs are very easy to understand the p value concept and how, by using this concept, we can easily determine the conclusion in dichotomous way that result is statistically significant or statistically in significant. Suppose researcher prefix the statistical significance on 0.05 and performed the experiment/trial, researcher intervene an Ayurvedic formulation in patient of study to increase the level of hemoglobin as researcher wants to test the efficacy of new Ayurvedic drug formulation in increasing the hemoglobin level. Researcher took sample of 120 patients and performed pre-post analysis after intervene of the new formulation. A difference of 0.50 was found in hemoglobin level with p = 0.06. Researcher thinks that the sample size is low that is why he did not get significant results and hence, he could not publish article or claim for efficacy of new formulation. Researcher again performed the same experiment with 900 patients and found the difference in hemoglobin is 0.40 with p = 0.001. Now, he claimed that a significant difference is found in hemoglobin level that is why this new formulation is effective to increase the hemoglobin level.

The lacuna behind these conclusions is P value misconception. Here, researcher is not bothered about the biological significance, it means a change of 0.4 or 0.5 in hemoglobin level is biologically significant or not. The criteria for minimum sample size and many more things are related to the study design and planning. In standard clinical trials, researchers have to think not only on the p value but also discuss a range of other potential explanations for result. In any scientific experiment, the explanation of the result should be scientific and that goes far beyond the to statistical significance or p value. Other potential contributors in result such as background evidence, study design, sample size, data quality and mode of action of drug are often more important than statistical significance. In Ayurveda, where a product contains different chemical compound and the patient’s Prakriti (Kapha, Vata, Pitta and variants)[8] also determines the result, hence, it is quite tedious to conclude only on statistical significance.

The objection behind p value theory is that it is being used for yes and no decision in certain experiment for testing the statistical significance.[9,10] It can change with sample size, study design and even wrong selection of statistical test

Illusion of the “p-value” theory in Ayurveda research: A need for perceivable alternative

Dear Editor,

Reporting and interpretation of results of classical statistical tests is widespread among applied researchers, most of whom erroneously believe that such tests are prescribed by a single coherent theory of statistical inference. This is not the case; classical statistical testing is a hybrid of the two different approaches formulated by Fisher on the one hand and Jerzy Neyman and Egon Pearson on the other.[1,2]

In Fisher’s approach, the researcher sets up a null hypothesis that a sample comes from a hypothetical infinite population with a known sampling distribution. The null hypothesis is said to be “disproved,” as Fisher called it, or rejected if the sample estimate deviates from the mean of the sampling distribution by more than a specified criterion, the level of significance. According to the theory, p value is the conditional probability than P (data/null hypothesis).[1‑3]

The Neyman–Pearson approach formulates two competing hypotheses, the null hypothesis and the alternative hypothesis. In a not-so oblique reference to Fisher, Neyman commented on the rationale for an alternative hypothesis. Specification of an alternative hypothesis critically distinguishes between the Fisherian and Neyman–Pearson methodologies.[1‑3]

p-value actually indicates the probability that whatever the difference came in assumed value (null hypothesis) and empirical value (data) is due to chance or sampling variation.[4,5] It can be found in a structure of clinical trial, suppose you analyzed a clinical trial data and found the p value is 0.2. It means that the probability of observed difference in efficacy of two drugs, which came out from clinical trial is resulting due to chance is 20%. If you repeat this experiment 100 times, then the difference in efficacy of two drugs reported in 20% of studies are due to the chance factor. In a clinical trial or any study, if the sample size is increased, then it will have less sampling error or error due to chance in result, which clearly indicates the lowering of p value. In general, increasing the sample size of study causes a decrease in the p value.[6,7]

In any experiment, it is mandatory to fix a level of sampling error or chance factor or p value up to the permissibility of sampling variability (whatever the difference come in study is due to chance factor), because majority of studies are based on the sample. If the difference in study is there it is said that, if the original p value is less than the prefixed level, then it means the difference occurred from the study has the less chance that the difference is due to sampling variability. It means the difference is the real and statistically valid. In general, the pre-fixed levels of p values in studies are 0.10, 0.05, 0.01 and 0.001. These prefixations depend on the researcher choice, literature of concern area and applicability of experiment. Suppose researcher describing their statistically significant (p < 0.05) result found a rate ratio 1.5 that is in new drug, the cure rate of disease is 1.5 times higher as compared to the standard drug group.

The above-mentioned paragraphs are very easy to understand the p value concept and how, by using this concept, we can easily determine the conclusion in dichotomous way that result is statistically significant or statistically in significant. Suppose researcher prefix the statistical significance on 0.05 and performed the experiment/trial, researcher intervene an Ayurvedic formulation in patient of study to increase the level of hemoglobin as researcher wants to test the efficacy of new Ayurvedic drug formulation in increasing the hemoglobin level. Researcher took sample of 120 patients and performed pre-post analysis after intervene of the new formulation. A difference of 0.50 was found in hemoglobin level with p = 0.06. Researcher thinks that the sample size is low that is why he did not get significant results and hence, he could not publish article or claim for efficacy of new formulation. Researcher again performed the same experiment with 900 patients and found the difference in hemoglobin is 0.40 with p = 0.001. Now, he claimed that a significant difference is found in hemoglobin level that is why this new formulation is effective to increase the hemoglobin level.

The lacuna behind these conclusions is P value misconception. Here, researcher is not bothered about the biological significance, it means a change of 0.4 or 0.5 in hemoglobin level is biologically significant or not. The criteria for minimum sample size and many more things are related to the study design and planning. In standard clinical trials, researchers have to think not only on the p value but also discuss a range of other potential explanations for result. In any scientific experiment, the explanation of the result should be scientific and that goes far beyond the to statistical significance or p value. Other potential contributors in result such as background evidence, study design, sample size, data quality and mode of action of drug are often more important than statistical significance. In Ayurveda, where a product contains different chemical compound and the patient’s Prakriti (Kapha, Vata, Pitta and variants)[8] also determines the result, hence, it is quite tedious to conclude only on statistical significance.

The objection behind p value theory is that it is being used for yes and no decision in certain experiment for testing the statistical significance.[9,10] It can change with sample size, study design and even wrong selection of statistical test
although it gives a single figure for judgment in easy and impressive way. One possible option for cope up with this problem is to use the Bayesian measures\(^{(11)}\) of evidence. Bayesian theory is based on the Bayes’ theorem which analyzes the data by incorporating the prior information about parameter and resultant posterior distribution of parameter will give the required answer. Bayesian analyses can be carried out with and without incorporating any external or prior information.

Even in case of Ayurveda, if we want to add input or prior information (Prakriti) in the analysis, then the result will be better and according with the principles of Ayurveda. According to classical theory of statistical analysis (p-value theory), it is not possible to consider the prior information in the analysis, according to Bayesian theory, it is possible. Bayesian theory is considered the prior status of parameter (in the form of Prakriti) and adds up this information with current data with Bayes law. Hence, it is better to analyze the Ayurvedic data by Bayesian theory not by classical theory (p-value theory) of statistics.

Financial support and sponsorship
Nil.

Conflicts of interest
There are no conflicts of interest.

Arunabh Tripathi, Saket Ram Trigulla
NIIMH, CCRAS, Ministry of AYUSH GOI, Hyderabad, Telangana, India

Address for correspondence: Dr. Arunabh Tripathi, NIIMH, CCRAS, Ministry of AYUSH, GOI, Hyderabad - 500 036, Telangana, India. E-mail: ashoka07bhu@yahoo.in

References
1. Hubbard R, Bayarri MJ. Confusion over measures of evidence (P’s) versus errors (α’s) in classical statistical testing. Am Stat 2003;57:3, 171-8.
2. Lenhard J. Models and statistical inference: The controversy between Fisher and Neyman–Pearson. Br J Phil Sci 2006;57:69-91.
3. Good IJ. Some logic and history of hypothesis testing. In: Pitt JC, editor. Philosophy in Economics. 1st ed. Boston: Riedel; 1981. p. 149-74.
4. Armitage P, Berry G, Matthews JN. Statistical Methods in Medical Research. 4th ed. Massachusetts, USA: Blackwell Science Publishing; 2002. p. 47-82.
5. Daniel WW. Biostatistics: A Foundation for Analysis in Health Sciences. 7th ed. Delhi, India: John Wiley and Sons; 2006. p. 210.
6. Casella G, Berger RL. Statistical Inference. 2nd ed. Belmont, CA, USA: Duxbury; 2002. p. 374.
7. Lehmann EL., Fisher, Neyman. The Creation of Classical Statistics. 1st ed. New York, USA: Springer-Verlag; 2011. p. 9.
8. Tripathy R, Shukla V. editors. Vaidyamanorama, Hindi Commentary Charak Samhita: Part-I. Varanasi: Chaukhambha Surbharati Publication, India; 2000. p. 645.
9. Amrhein V, Greenland S, McShane B. Retire statistical significance. Nature 2019;567:305-7.
10. Wasserstein RL, Lazar NA. The ASA's statement on P values: Context, process and purpose. Am Stat 2016;70:2, 129-33.
11. Bernardo JM, Smith AF. Bayesian Theory. 2nd ed. Chichester: England Wiley; 1994. p. 2-7.

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

How to cite this article: Tripathi A, Trigulla SR. Illusion of the “p-value” theory in Ayurveda research: A need for perceptible alternative. AYU 2019;40:134-5.