Corresponding author: Dr. Leila Janani, leila_janani@yahoo.com

1. Department of Biostatistics, School of Public Health, Iran University of Medical Sciences, Tehran, Iran.
2. Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran.
P-value: What is and what is not

Table 1. Artificial case-control study

<table>
<thead>
<tr>
<th></th>
<th>Study A</th>
<th></th>
<th>Study B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Exposed</td>
<td>Unexposed</td>
<td>Exposed</td>
</tr>
<tr>
<td>Cases</td>
<td>8</td>
<td>2</td>
<td>800</td>
</tr>
<tr>
<td>Controls</td>
<td>7</td>
<td>3</td>
<td>700</td>
</tr>
</tbody>
</table>

Note: Hypothetical data

computing the p-value is that any deviation of the observed data from the null hypothesis was produced by chance, so it is clear that when only chance affects the deviation of the null hypothesis in the calculation of the p-value, it cannot be the probability of operating of the chance (7, 8).

The p-value is the probability of the observed data given that the null hypothesis is true, which is a probability that measures the consistency between the data and the hypothesis being tested if, and only if, the statistical model used to compute the p-value is correct (9). The smaller the p-value the greater the discrepancy: “If p is between 0.1 and 0.9, there is certainly no reason to suspect the hypothesis tested, but if it is below 0.02, it strongly indicates that the hypothesis fails to account for the entire facts. We should not be off-track if we draw a conventional line at p = 0.05” (2).

P-values have often been claimed or taken to imply the presence or absence of an effect or the importance of a result, which is certainly NOT true. In a clinical trial study in which the differences of an index are evaluated before and after an intervention and the mean difference is small (for example, 0.5 unit and p-value = 0.03), the p-value just implies the statistical significance (of an effect or correlation presence) and does not explain the importance of an effect or result. Thus, statistical significance is not equivalent to clinical significance and vice versa (10).

P-value does not measure the size of a effect. Suppose in 2 artificial case-control studies, the relationship between disease and exposure was examined by odds ratio. The data in Table 1 demonstrates that ratio of being exposed or unexposed; and as a result, the odds ratio (OR = 1.71) is the same across the studies and the only difference between them is the sample size. However, the p-value for testing the hypothesis that the true odds ratio is equal to 1 is 0.605 for study A and less than 0.0001 for study B. Apparently, the same effect can produce very different p-values. As it is clearly mentioned in ASA statement, any effect (large or small) can produce a small or large p-value depending on the sample size or measurement precision, so conclusion about a hypothesis should not be based on p-value only, but also other aspects such as measurement precision, sample size, study design, and assumptions should be taken into account (6). For example, in cross-over design and parallel design, we may have the same effect but a different p-value, or the same effect will have different p-values if the precision of the estimates differs.

Other incorrect interpretations of the p-value are as follows:

- If p-value = 0.2, there is a 20% chance that the null hypothesis is correct.
- Same p-values should have the same observed effects.
- P-value = 0.02 means that the probability of a type I error is 2%.
- P-value is a statistical index and has its own strengths and weaknesses, which should be considered to avoid its misuse and misinterpretation (12). Reporting the descriptive statistics, using confidence intervals of the measurement indexes alongside the p-value, and its true interpretation are of paramount importance (1).

Conclusion

A good report contains describing the data by suitable numerical and graphical summaries of data, dominance on the setting of the study, and logical and clinical interpretation of quantitative indexes. As a whole, summarizing statistical comparisons to statistical significance or non-significance is one of the highly popular statistical misinterpretations of p-values and hypothesis testing. The p-value like other indexes should be used and interpreted appropriately and should not be a scientific reason just by itself.

Conflict of Interests

The authors declare that they have no competing interests.

References