Editorial

Common statistical mistakes: Reporting only P-values for results

Researchers in several fields of medicine have found high rates of statistical errors in large numbers of scientific articles, even in the best journals. The problem of poor statistical reporting is, in fact, longstanding, widespread, potentially serious, and not well known, despite the fact that most errors concern basic statistical concepts and can be easily avoided by following a few guidelines.

Reporting only P values for results is one the most common statistical errors. *P* value is often misinterpreted. Even when interpreted correctly, however, they have some limitations. For main results, report the absolute difference between groups (relative or percent differences can be misleading) and the 95% confidence interval for the difference, instead of, or in addition to, P values. The sentences below go from poor to good reporting :

– *"The effect of the drug was statistically significant."*

This sentence does not indicate the size of the effect, whether the effect is clinically important, or how statistically significant the effect is. Some readers would interpret "statistically significant" in this case to mean that the study supports the use of the drug.

– *"The effect of the drug on lowering diastolic blood pressure was statistically significant (P< 0.05)"*

Here, the size of the drop is not given, so its clinical importance is not known. Also, *P* could be 0.049; statistically significant (at the 0.05 level) but so close to 0.05 that it should probably be interpreted similarly to a *P* value of, say, 0.51, which is not statistically significant. The use of an arbitrary cut point, such as 0.05, to distinguish between "significant" and "non significant" results is one of the problems of interpreting *P* values.

– *"The mean diastolic blood pressure of the treatment group dropped from 110 to 92mmHg (P=0.02)."*

This sentence is perhaps the most typical. The pre- and post-test values are given, but not the difference. The mean drop – the 18-mm Hg difference – is statistically significant, but it is also an estimate, and without a 95% confidence interval, the precision (and therefore the usefulness) of the estimate cannot be determined.

– *"The drug lowered diastolic blood pressure by a mean of 18mmHg, from 110 to 92mmHg (95% CI=2 to 34mmHg; P=0.02)."*

The confidence interval indicates that if the drug were to be tested on 100 samples similar to the one reported, the average drop in blood pressure in 95 of those 100 samples would probably range between 2 and 34 mm Hg. A drop of only 2 mm Hg is not clinically important, but a drop of 34 mm Hg is. So, although the mean drop in blood pressures in this study was statistically significant, the expected difference in blood pressures in other studies may not always be clinically important; that is, the study is inconclusive. When a study produces a confidence interval in which *all* the values are clinically important, the intervention is much more likely to be clinically effective. If *none* of the values in the interval are clinically important, the intervention is likely to be ineffective. If only some of the values are clinically important, the study probably did not enroll enough patients .

Asghar Ashrafi Hafez[1, 2]
[1]Proteomics Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[2]Medical Education Research Center, Tabriz University of Medical Sciences, Tabriz, Iran