34. Association, Correlation, and Cause-Effect Relationship

François Noël, December 2025
(Reviewed by Maurício Schüler Nin, UFCSPA)

Two events are said to be associated when their occurrences vary concomitantly, suggesting possible dependence between them. Correlation is a specific measure of association that refines this information by quantifying it (Pereira et al., 2000), measuring the strength and direction of the relationship between two continuous variables.

It is essential to clarify that misinterpretations are quite common in this field, particularly when the demonstration of an association or correlation is interpreted as evidence of a cause-effect relationship. The fact that two variables covary does not imply that changes in both are causally connected; it merely indicates that they influence each other in some way or share an underlying common cause (Marino, 2014). This was ironically illustrated by Helmut Sies (1988) in a letter to *Nature*, where he presented data showing a concomitant decline in the number of newborns in West Germany between 1965 and 1980 and the number of storks hatching their eggs. Provocatively, he remarked that these data "*may suggest a solution that every child knows makes sense*", a surreal conclusion that highlights the danger of interpreting associations or correlations as if they necessarily implied a cause-effect relationship. In this case, a plausible explanation would be that the region's heavy industrialization during that period (a common underlying cause) may have scared off storks, which are very sensitive to air pollution, while simultaneously increasing the population's well-being and leading to greater birth control, an already well-known phenomenon, including in Brazil.

Another example, more serious and current, relates to associations observed in a population-based retrospective study conducted in Seoul, South Korea, between COVID-19 vaccination and cancer incidence (Kim et al., 2025). In this study, the association was based on hazard ratios, a risk measurement discussed in another entry of this glossary. Hazard ratios were calculated using Cox proportional hazards regression analysis, a statistical technique widely used in survival studies to evaluate the effect of variables on the time until the occurrence of an event (such as death, relapse, treatment failure, etc.). The authors' conclusion that "*COVID-19 vaccination could be associated with an increased risk of six specific cancer types*" suggests the possibility of a causal relationship. However, critics of the study highlight the potential presence of a common underlying cause, rather than a cause-effect relationship, as discussed by David Gorski in an article published by the New England Skeptical Society, a nonprofit organization dedicated to promoting science and critical thinking (https://sciencebasedmedicine.org/another-study-misrepresented-as-evidence-covid-vaccines-cause-cancer/). One of his criticisms concerns the "healthy user bias" and the study's failure to adjust for confounding factors that could explain the observed association. According to him, people who engage in healthy behaviors (such as getting

vaccinated against COVID-19) are also more likely to undergo other health interventions, such as cancer screening tests. In the case of this study, this interpretation seems plausible, given that many of the cancer types for which increased risk was observed are exactly those with established screening programs in South Korea.

Before concluding, it is worth recalling that statistical tests such as Pearson's or Spearman's correlation tests are commonly used to determine whether a correlation exists between two independent measurements. The parametric Pearson test has greater statistical power but can only be used when data are normally distributed and the relationship is linear. On the other hand, the nonparametric Spearman test (rank-based) is recommended because it is more robust for small datasets, when the assumption of normality is violated or not evaluated, or when the relationship appears non-linear (Marino, 2014). Even when a statistically significant linear correlation exists between two measurements, it is essential to verify its strength, for example, through the r^2 value in Pearson's test, as well as the p-value, bearing in mind the arbitrariness of the conventional $p = 0.05$ threshold.

In conclusion, we must recognize that, once again, a basic understanding of biostatistics and a critical mindset are necessary to avoid misinterpretations, some of which are disseminated as fake news.

References

Kim H.J., Kim M-H., Choi M.G., Chun E.M. 1-year risks of cancers associated with COVID-19 vaccination: a large population-based cohort study in South Korea. *Biomark. Res.* 13:114, 2025.

Marino M.J. The use and misuse of statistical methodologies in pharmacology research. *Biochem. Pharmacol.* 87:78-92, 2014.

Pereira J.C.R., Paes A.T., Okano V. Espaço aberto: Questões comuns sobre epidemiologia, estatística e informática. *Revista do IDPC*, São Paulo, v.7, p. 12-17, 2000.

Sies H. A new parameter for sex education. *Nature* 332(7):495, 1988.