35. Associação, correlação e relação causa-efeito

François Noël, dezembro 2025
(Revisado por Maurício Schüler Nin, UFCSPA)

Dois eventos se dizem associados quando suas ocorrências variam concomitantemente, sugerindo possível dependência entre eles. A correlação é uma medida específica de associação que permite refinar a informação, quantificando-a (Pereira e cols., 2000), medindo a força e a direção da relação entre duas variáveis contínuas.

É fundamental esclarecer que interpretações equivocadas são bastante comuns nessa área, quando a demonstração de uma associação ou correlação é interpretada como evidência de relação causa-efeito. O fato de duas variáveis covariarem não sugere que as mudanças em ambas estejam causalmente conectadas, mas apenas indica alguma influência entre elas ou a existência de uma causa subjacente comum (Marino, 2014), como ilustrado de forma irônica por Helmut Sies (1988) em carta à revista *Nature*. Nela, o autor apresenta dados de diminuição concomitante do número de recém-nascidos na Alemanha Ocidental entre 1965 e 1980 e do número de cegonhas chocando seus ovos. De forma provocadora, menciona que esses dados "podem sugerir uma solução que toda criança sabe que faz sentido", conclusão surreal que evidencia o perigo de interpretar associações ou correlações como se indicassem, necessariamente, uma relação de causa e efeito. Neste caso, uma explicação plausível seria que a forte industrialização da região nesse período (causa subjacente comum) tenha afugentado as cegonhas, aves muito sensíveis à poluição do ar, e, por outro lado, elevado o nível de bem-estar da população, levando a um maior controle da natalidade, fenômeno bem conhecido, inclusive no Brasil.

Outro exemplo, mais sério e atual, remete às associações observadas em um estudo retrospectivo de base populacional realizado em Seul, na Coreia do Sul, entre a vacinação contra a COVID-19 e a incidência de câncer (Kim *e cols.,* 2025). Neste estudo a associação foi baseada nas razões de riscos (*hazard ratio*), uma medida de risco discutida em outra matéria deste glossário (https://sbfte.org.br/wp-content/uploads/2022/08/30.-Risco-relativo-relative-risk-razao-de-chances-odds-ratio-e-razao-de-riscos-final.pdf). Neste caso, as razões de risco foram calculadas usando a análise de regressão de riscos proporcionais de Cox, uma técnica estatística amplamente usada em estudos de sobrevivência, para avaliar o efeito de variáveis sobre o tempo até a ocorrência de um evento (como morte, recaída, falha de tratamento, etc.). A conclusão dos autores de que "a vacinação contra a COVID-19 pode estar associada a um risco aumentado de seis tipos específicos de câncer" sugere a possibilidade de uma relação causal. No entanto, críticos do estudo apontam a possibilidade de existir uma causa subjacente comum, e não uma relação de causa e efeito, como destacou David Gorski em uma matéria publicada pela *New England Skeptical Society*, uma organização sem fins lucrativos dedicada à promoção da ciência e do pensamento crítico

(https://sciencebasedmedicine.org/another-study-misrepresented-as-evidence-covid-vaccines-cause-cancer/). De fato, uma das suas críticas aponta o "viés do usuário saudável" e a falha em corrigir fatores de confusão para explicar a associação observada. Segundo ele, é provável que pessoas que adotam um comportamento saudável (como vacinar-se contra a COVID-19) também tenham maior probabilidade de adotar outras intervenções de saúde, como exames de rastreamento de câncer. No caso deste estudo, essa interpretação parece plausível, já que muitos dos tipos de câncer para os quais se observou um risco aumentado são justamente aqueles que contam com programas de rastreamento na Coreia do Sul.

Antes de concluir, queria lembrar que é comum usar testes estatísticos como os testes de Pearson ou de Spearman para verificar se existe correlação entre duas medidas independentes. O teste paramétrico de Pearson tem maior poder estatístico, mas só pode ser utilizado quando a distribuição dos dados é normal e a relação é linear. Por outro lado, o teste não paramétrico de Spearman (baseado em ordens) é indicado por ser mais robusto para conjuntos de dados pequenos, quando a suposição de normalidade é violada ou não é avaliada, ou quando a relação parece ser não linear (Marino, 2014). Mesmo havendo correlação linear estatisticamente significativa entre duas medidas, é fundamental verificar a sua força, por exemplo pelo valor de r^2 no caso do teste de Pearson, assim como o valor de P, lembrando da arbitrariedade em termos fixado o valor de P=0,05 como divisor de água.

Em conclusão, devemos reconhecer que, mais uma vez, um entendimento básico de bioestatística e uma postura crítica são necessários para evitar interpretações equivocadas, às vezes divulgadas na forma de *fake news*.

Referências

Kim H.J., Kim M-H., Choi M.G. and 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