

Como mensurar a precisão dos resultados de estudos clínicos

How to measure the precision of clinical study results

Na fase de elaboração de um estudo clínico, após estabelecer uma pergunta a ser respondida por meio dos dados obtidos pela pesquisa, define-se a população que será estudada. Como, em geral, não é viável avaliar toda a população em estudo, seleciona-se uma amostra. Amostragem é quando a análise é realizada com base numa parte (representativa) da população. Nesta fase, o importante é considerar que a única forma de garantir a representatividade dessa amostra será fazendo sua seleção de forma aleatória, a fim de que cada elemento da população tenha exatamente a mesma probabilidade de ser selecionado, sem que haja influência do pesquisador⁽¹⁻³⁾.

Em amostras selecionadas aleatoriamente (randomizadas) de uma população é possível mensurar a veracidade dos resultados obtidos, por meio de ferramentas como o cálculo do intervalo de confiança (IC). Geralmente, definimos como ideal o IC com 95% de precisão. Isto significa que, com 95% de garantia, o valor da porcentagem de cada variável estudada na amostra corresponde ao real valor da porcentagem da população. Contudo, devemos levar em consideração que mesmo este intervalo não é uma certeza, pois tem uma confiança de 95%, ou seja, há uma probabilidade de 5% de a verdadeira porcentagem na população estudada está fora desses limites. Quanto maior for uma amostra, menor será o intervalo de confiança e, assim, maior será a probabilidade de obtermos extrapolações precisas das verdadeiras porcentagens da população. Amostras pequenas geram dados imprecisos.

Assim, é importante calcular a dimensão da amostra, a fim de dar precisão adequada aos resultados. A dimensão da amostra está relacionada com a precisão do intervalo de confiança desejada, para poder extrapolar os resultados da amostra para a população. Caso a amostra não seja aleatória, não fará sentido falar em intervalos de confiança, porque, nessa situação, a amostra não será representativa da população.

Desta forma, o IC auxilia a verificar se o tamanho da amostra foi suficiente e se os resultados obtidos foram decorrentes da intervenção ou decorrentes do acaso (quanto mais estreito for o IC maior será a certeza de que a amostra representa a população). Além do tamanho da amostra e do IC, o valor “*p*” (nível de significância) ajuda a definir a precisão dos resultados do estudo. O “*p*” representa a probabilidade de as diferenças entre os resultados encontrados terem ocorrido por acaso e não devido à intervenção. Em geral, se tolera o valor de “*p*” < 5%, ou 0,05, o que significa que a chance de os resultados não serem verdadeiros é muito pequena, menos de uma chance em 20.

Pode-se considerar que “*p*” indica a possibilidade de se detectar uma diferença que não existe (falso positivo). Este tipo de erro (tipo I ou α) é prejudicial aos pacientes, pois, no caso de estudos com novas medicações pode sugerir que uma droga que na verdade é ineficaz, faça efeito, prejudicando os usuários, que não estarão sendo adequadamente tratados. Por outro lado, erros falsos negativos (tipo II ou β), que correspondem à possibilidade de não se reconhecer uma diferença que realmente existe, são mais tolerados, em geral, até 20%, pois nesse caso, os prejudicados são as indústrias farmacêuticas, que desenvolvem uma droga eficaz, mas que os estudos consideram ineficaz e, portanto, essa droga não será utilizada.

Assim, a amostragem é uma técnica estatística que significa extrair do todo (população) uma parte (amostra), com o objetivo de avaliar certas características dessa população. A validade do estudo está diretamente ligada à randomização, ao tamanho da amostra, aos indicadores de precisão dos dados (IC e “*p*”) e à magnitude de perda de seguimento. É importante que leitores analisem esses itens, para avaliar a possibilidade de extrapolação dos resultados da amostra para a população em estudo.

Pesquisadores e leitores, também, devem atentar para o fato de que em amostras muito grandes, efeitos pequenos podem ser considerados estatisticamente significantes, com “ p ” < 5%, embora as diferenças sejam clinicamente irrelevantes. Assim, os leitores devem considerar, também, para a análise crítica dos artigos, além dos dados estatísticos, a relevância das características clínicas dos resultados.

Outro fator de confusão é o fato de que, dependendo da quantidade de testes estatísticos utilizados para análise dos dados, algum teste pode eventualmente ter significância positiva, enquanto a grande maioria é negativa, pois como o nível de significância utilizado, comumente, é de < 5%, espera-se que um caso em cada 20, seja falsamente positivo. Se, eventualmente, o pesquisador decidir forçar um resultado positivo, ele pode, por exemplo, realizar diversos testes estatísticos e considerar o único que foi positivo.

De uma maneira geral, pesquisas que incorrem em erros metodológicos de seleção, de aferição ou de interpretação dos dados, tendem a produzir resultados divergentes de estudos semelhantes⁽⁴⁻⁵⁾. Por isso, é importante que o leitor considere possíveis conflitos de interesse expressos pelos autores, além de, também, avaliar diversos estudos semelhantes, antes de formar uma opinião sobre um determinado assunto. Esta é a maneira ideal de se analisar criticamente a literatura.

Newton Kara-Junior

Editor-Chefe da Revista Brasileira de Oftalmologia
Professor Colaborador, Livre-docente e Professor do Programa de Pós-graduação,
Faculdade de Medicina, Universidade de São Paulo, São Paulo, SP, Brasil.

REFERÊNCIAS

1. Siqueira RC. Pesquisa translacional na oftalmologia: o caminho para a medicina personalizada. Rev Bras Oftalmol. 2012; 71(5):338-42.
2. Kara-Junior N, Espindola RF, Gomes BA, Ventura B, Smadja D, Santhiago MR. Effects of blue light-filtering intraocular lenses on the macula, contrast sensitivity, and color vision after a long-term follow-up. J Cataract Refract Surg. 2011;37(12):2115-9
3. Kara-Junior N, Jardim JL, de Oliveira Leme E, Dall'Col M, Susanna R Jr. Effect of the AcrySof Natural intraocular lens on blue-yellow perimetry. J Cataract Refract Surg. 2006;32(8):1328-30.
4. Chamon W. [Plagiarism and misconduct in research: where we are and what we can do]. Arq Bras Oftalmol. 2013;76(6):V-VIII. Portuguese.
5. Chamon W. [Fine prints at the bottom of the page]. Arq Bras Oftalmol. 2013;76(3):v-vi. Portuguese.