

Como interpretar um artigo científico?

How to read a scientific article?

Marcos Roberto de Sousa¹, Gláucia Fernandes Cota²

INTRODUÇÃO

Há pouco mais de 20 anos, uma guinada em direção à evidência redirecionou a lógica do pensar e agir e vem desde então orientando a formação profissional e a conduta médicas. Os residentes e especializando em medicina nestes anos, ávidos por evidências e diretrizes, procuram a objetividade mais do que nunca. Porém, encontram dificuldades em separar boas informações em uma literatura muito extensa. Uma análise da base SCOPUS mostrou que, entre 1997 e 2007, o número de artigos/ano de revistas da área cardiovascular indexados aumentou 57% e o número de revistas cresceu 75%, chegando a 222 jornais e 164.377 artigos publicados no período. Destes, 15,6% não receberam nenhuma citação após serem publicados e 46% receberam menos de 5 citações em 5 anos. Portanto, 60% da produção publicada nesta base no período não foi relevante, de acordo com este critério que avalia a utilização da informação produzida.¹

Uma análise de 3271 recomendações de 19 diretrizes americanas publicadas em 2013 mostrou que 50% delas, embora classificadas como fortemente indicadas, eram baseadas em evidências fracas ou advinham de consenso de *experts* e menos de 10% eram sustentadas por evidência de alta qualidade.² Na mesma linha, avaliação de 461 metanálises revelou que de um quarto a um terço delas tinha resultados falso-positivos ou exagerados.³ Diante disso, como interpretar um artigo científico?

Para tentar responder à pergunta, propomos uma perspectiva filosófica, uma análise crítica sistematizada geral e específica e, por fim, a aplicação do pensamento Bayesiano. É muito importante que o leitor tente formar uma opinião própria que pode ser diferente das conclusões do autor do artigo, à luz de sua formação e conhecimentos prévios. Para uma análise crítica das informações apresentadas, um roteiro organizado para avaliação da qualidade e quantificação dos resultados de uma pesquisa pode ser instrumento útil. A formação básica de um aluno ou especializando passa primeiro pelo seu curso, suas experiências e pela leitura de livros-texto. Considere que os artigos são uma atualização e não uma fonte primária de formação.

VISÃO FILOSÓFICA

Uma pesquisa original, aquela que produz um conhecimento que não existia, está na fronteira entre o conhecimento estabelecido e o conhecimento ainda incerto. Se o conhecimento é incerto ainda, então é sujeito a questionamentos, pode haver falhas metodológicas e vieses. Por isto, é fundamental utilizar o próprio julgamento crítico. Vários foram os filósofos que se debruçaram sobre o tema “o que podemos saber?” (René Descartes, 1596-1650), ou se “há relação de causalidade entre dois eventos?” (David Hume, 1711-1776). Um alerta para pesquisadores e leitores de artigos é “não vemos as coisas como elas são, mas como nós somos” (Immanuel Kant, 1724-1804). Muitas vezes, antes de 1840, o termo filosofia foi usado como sinônimo de ciência. Segundo Patrícia Fara, o termo “cientista” não existia, foi criado em cerca de 1840 e as palavras usadas antes para expressar “cientista” eram: “naturalista”, “homem da ciência” e “filósofo experimental”.⁴

Também na história recente, a filosofia exerce seu criticismo reflexivo sobre o pensar científico, mantendo acessa a inquietude necessária para a busca, mas também a lucidez irrefutável sobre a limitação humana: “O sucesso das histórias da ciência nasce do caráter abstrato dos enclausuramentos de percepção e linguagem envolvidos e da incorporação de relações matemáticas idealizadas”.⁵ Apesar destas considerações, decisões médicas precisam ser tomadas todos os dias. O método científico e a produção de evidência seguem

¹ Universidade Federal de Minas Gerais-UFMG, Hospital das Clínicas. Belo Horizonte, MG - Brasil.

² Fundação Oswaldo Cruz-FIOCRUZ, Centro de Pesquisas René Rachou; UFMG, Hospital das Clínicas. Belo Horizonte, MG - Brasil.

Instituição:

Universidade Federal de Minas Gerais-UFMG, Hospital das Clínicas. Belo Horizonte, MG - Brasil.

* Autor Correspondente:

Marcos Roberto de Sousa
E-mail: mrsousa@cardiol.br

Recebido em: 20/12/2015.

Approved em: 15/01/2016.

como instrumentos muito úteis, desde que devidamente avaliados sob o crivo atento do leitor.

Do ponto de vista prático, o relato de um estudo científico exige um rigoroso método de investigação. De maneira simplificada, podemos assumir que o cientista é o observador e que o objeto de investigação é a relação entre duas variáveis: exposição e desfecho. Em estudos de prevalência e incidência, podemos ter apenas desfecho, sem exposição. Por sua vez, a exposição pode ser um fator de risco ou uma variável associada (estudos de coorte e caso-controle) ou uma intervenção (ensaio clínico). Para que o cientista investigue, primeiro é preciso definir de forma clara a exposição e o desfecho de interesse. É exatamente esta definição (ou indefinição) que exploramos sob a perspectiva filosófica. Frequentemente, a definição de exposição difere entre autores, mesmo tratando-se da mesma condição estudada.

Da mesma forma, dada a dificuldade de se investigar um desfecho clínico de interesse, frequentemente se aceita um desfecho substitutivo, por exemplo, um marcador laboratorial correlato ao desfecho clínico, baseado em deduções e extrapolações lógicas, mas ainda assim limitadas. Isto tem utilidade, mas também riscos, pois nem sempre a redução no número de eventos de um desfecho correlato substitutivo significa que houve impacto sobre o desfecho clínico. É importante lembrar que definimos os problemas que se apresentam de acordo com nossos recursos de linguagem, nossas experiências passadas e influências culturais. Tais recursos determinam a maneira de ver o mundo e interpretar os fatos, também explicam a busca por padrões onde eles não existem e a crença nos resultados que coincidem com nossas convicções prévias, o que é chamado de “viés de confirmação”.⁶⁻⁸

Vamos aqui simplificar o artigo científico em introdução, métodos, resultados e discussão. Para a avaliação crítica do leitor, interessam mais os métodos e resultados. Uma sugestão inicial, extraída da reflexão crítica da filosofia: desconfie da introdução e da discussão do artigo. Não estamos falando de má-fé do autor, não se trata disso. Mas lembre-se de que, na introdução, o autor exercita sua retórica expondo uma justificativa com o viés da tese ali defendida e do próprio resultado alcançado. Como já foi dito, o ser humano é propenso a um “viés de confirmação”, e por mais que tente ser imparcial e judicioso (como obviamente deve ser), é frequentemente influenciado por aquilo que pensa *a priori*, sendo a retórica um recurso legítimo para a defesa de uma tese. Na discussão, o autor deve usar um raciocínio dialético, apresentando evidências da literatura que corroboram, mas também as que contradizem seus resultados.

Bons artigos se antecipam às críticas indicando e explicando as divergências. Mas isto nem sempre acontece e a dialética da discussão é frequentemente influenciada pela tese defendida pelo autor, cuja perspectiva deve ser observada. Não se trata de tarefa fácil, mas são inúmeras as informações imbuídas de significado: observe a afiliação dos autores, seu tipo de vínculo institucional, suas outras publicações, sua área de atuação, a seção de agradecimentos e o financiamento informado para o estudo. A terceira sugestão seria: avalie a revista em que o artigo foi publicado. Há regras rigorosas para que uma produção científica seja divulgada. Quanto maior é o impacto e a relevância da revista, teoricamente, maior o rigor para aceitação do trabalho para publicação.

Esta última assertiva não quer dizer que estudos de revistas boas não tenham falhas metodológicas.

A última questão filosófica que gostaríamos de abordar é que, dado que o ser humano apresenta naturalmente viés de confirmação, e dadas as condições culturais de hipérbole em que vivemos, resultados positivos tendem a ser mais frequentemente publicados, mais frequentemente em língua inglesa e em revistas de maior impacto. É o chamado viés de publicação,⁹ que faz surgir, em metanálises, resultados exageradamente positivos como já mencionado na introdução. Além disso, ao lembrar-se deste viés ao ler artigos com resultados positivos, facilita para o leitor considerar a possibilidade de um estudo neutro ou negativo avaliando a mesma questão no futuro.

VISÃO GERAL SOBRE AS REGRAS PARA INTERPRETAÇÃO DE ESTUDOS

As seções que interessam primordialmente em uma pesquisa original são a seção de métodos e a seção de resultados. Os leitores devem priorizar estas seções sobre as demais.

SEÇÃO DE MÉTODOS E PACIENTES DO ARTIGO

O delineamento metodológico e estatístico para os diversos tipos de estudos são claramente estabelecidos e devem ser seguidos, havendo diretrizes publicadas para quase todos os tipos de delineamentos. Se já definimos que o objeto de estudo é a relação entre uma exposição e um desfecho e o cientista como observador, podemos então avaliar a relação temporal entre o objeto e o observador (Figura 1).

O tempo é uma variável adicional na relação exposição e desfecho. Estudos conduzidos no passado podem não encontrar os mesmos resultados de hoje porque as outras variáveis (especialmente as outras terapias associadas) sofreram modificações ao longo do tempo. O tempo de seguimento diferente entre estudos determina taxas de desfechos diferentes, por exemplo, é bom lembrar que “em longo prazo, todos estaremos mortos”. Obviamente, estudos que seguiram pacientes com doenças graves por um ano ou por cinco anos encontrarão taxas muito diferentes de mortalidade. Por outro lado, ao comparar estudos, taxas de mortalidade diferentes com tempos de seguimento semelhantes sugere que há diferenças de gravidade clínica dos pacientes ou diferenças de abordagem clínica da doença.

Estudos prospectivos, como demonstrado na Figura 1, partem da exposição e acompanham os pacientes medindo o desfecho como “incidência” ou como “tempo até o evento”. Depois, faz-se uma comparação das taxas entre expostos e não expostos através do Risco Relativo (RR) ou da Hazard Ratio (HR), respectivamente. Estudos caso-controle, por sua vez, partem, por definição metodológica, dos casos e selecionam um grupo controle, e por isto não podem falar em incidência, uma vez que não tem o denominador da equação. Se o estudo é caso-controle, não é possível medir incidência ou tempo até o desfecho. Por isto, a medida estatística do objeto de estudo (ou seja, da relação entre exposição e desfecho), é a “odds ratio” (OR). Já os estudos transversais, aqueles que avaliam os pacientes em um dado momento e não fazem seguimento, só podem determinar a prevalência pontual e a associação transversal (sejam medidas diagnósticas, sejam razões de prevalência, diferença de prevalência ou *odds ratio*),

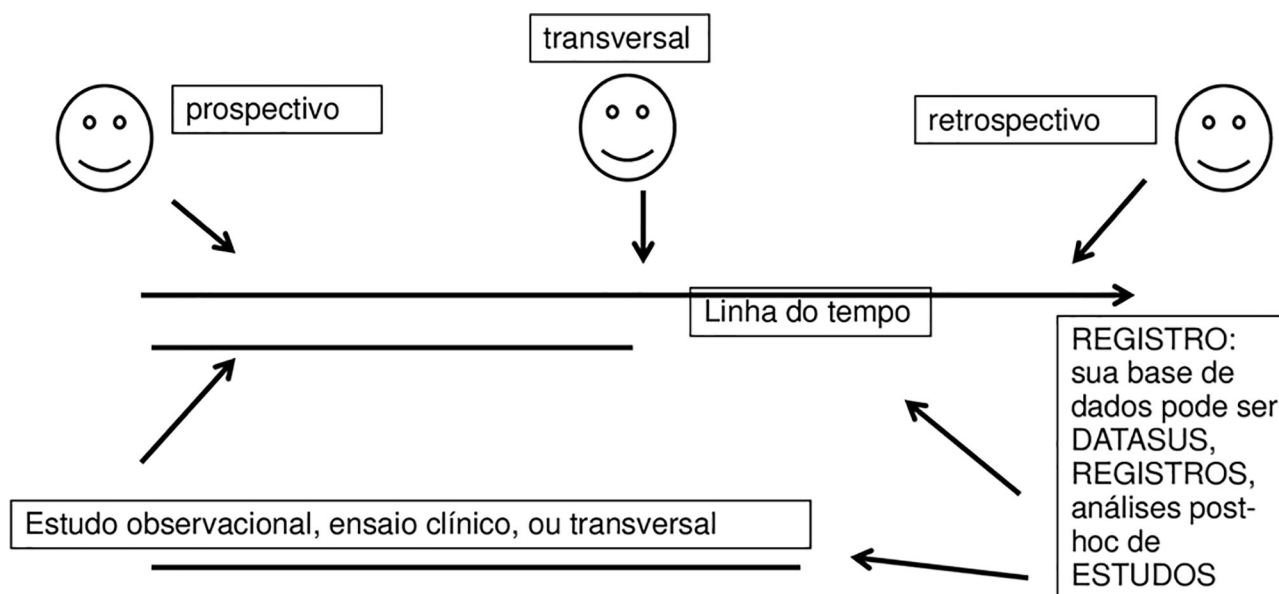


Figura 1: A relação temporal entre o observador e as associações observadas no artigo original determina o delineamento metodológico e estatístico.

mas não estabelecem qualquer relação de causa e efeito desta associação. Estudos transversais são também usados para comparação de métodos diagnósticos, como demonstrado na Figura 2.

Assim, o método de delineamento do estudo está atrelado ao método estatístico que mede a associação entre as variáveis “exposição” e “desfecho”. É necessário que o investigador siga os passos de um modelo teórico de produção de uma observação científica. Tais regras devem ser seguidas porque os pressupostos metodológicos determinam os modelos estatísticos a serem utilizados.

De forma simplificada, podemos ter uma visão geral sobre os diversos tipos de estudos científicos: o objeto de estudo deve ser uma questão simples e bem definida, seja uma medida de incidência ou prevalência, seja uma medida de associação entre uma exposição e um desfecho (que estamos chamando aqui de objeto de estudo). A randomização ou aleatorização permite que os pacientes sejam distribuídos aleatoriamente nos dois grupos de modo que as variáveis de confusão se distribuam igualmente ao acaso.

Atualmente, as técnicas de aleatorização devem ser realizadas de forma que o pesquisador não possa prever a sequência (“*concealment allocation*”), garantindo a ausência de viés de seleção. O objeto de estudo pode ser analisado, do ponto de vista temporal, prospectivamente, transversalmente ou retrospectivamente, com suas respectivas medidas de objeto de estudo (que são as medidas de “tamanho de efeito” ou de associação: RR, HR, OR, ou medidas de acurácia). Partindo desta visão geral sobre os artigos, podemos apreender uma série de regras práticas, definindo as circunstâncias da investigação (Tabela 1). Identificar e caracterizar clinicamente a população do estudo ajuda o leitor a pensar em quais pacientes de sua prática o artigo se aplica.

O cálculo amostral, em geral apresentado em “análise estatística” dentro da seção de métodos, é um requisito importante de qualidade do estudo. Nem todo estudo requer cálculo amostral, por exemplo, quando o autor apresenta os casos consecutivos de um serviço de referência para

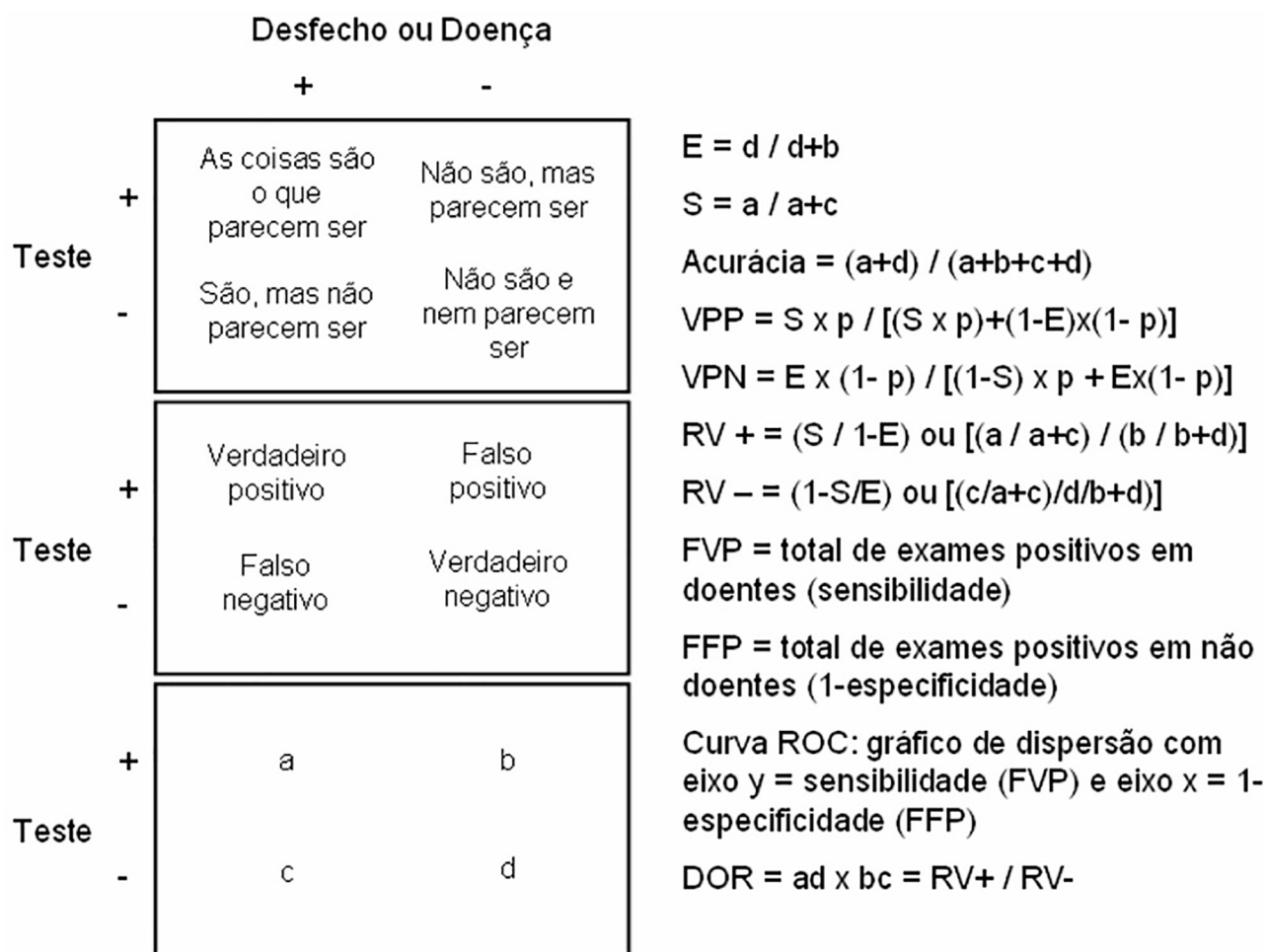
uma doença rara. Neste caso, trata-se de uma amostra de conveniência justificada pela dificuldade de se obter amostra de pacientes suficiente para a investigação. Ainda assim, é possível estimar *a posteriori* o poder estatístico do estudo (de acordo com o número de pacientes e de eventos obtidos) para as inferências realizadas.

Nos estudos com cálculo amostral, é importante observar se a amostra necessária foi atingida e seguida pelo tempo planejado. Além disso, é importante lembrar que frequentemente a amostra é calculada para a análise do desfecho primário do estudo, de modo que inferências sobre resultados de desfechos secundários e análises de subgrupo são apenas observações geradoras de hipóteses, já que a chance de falso positivo (erro alfa, mais frequentemente usado 5%) e o poder do estudo (1 - beta, sendo beta a chance de falso negativo do estudo, mais frequentemente usado poder de 80%) foram considerados para o desfecho primário.

Os testes estatísticos para comparar cada variável nos dois grupos em geral podem ser simplificados assim: teste T para comparação de médias; estatística F para análise de variância; teste do Qui-Quadrado para comparação de variáveis categóricas; testes condicionais quando os grupos são pareados; e testes não paramétricos (Wilcoxon ou Kruskal-Wallis) para variáveis com distribuição assimétrica. Os testes diagnósticos foram apresentados na Figura 2. Em ensaios clínicos e coortes prospectivas, são utilizados o modelo de regressão de Cox com estimativa da HR e a curva de sobrevivência de Kaplan Meyer com teste do log-rank; e pode ser usado também o RR (menos usado atualmente por não considerar o tempo e sim apenas a proporção ou incidência de eventos durante o período total de seguimento). A Tabela 2 apresenta a evolução da metodologia de pesquisa nas últimas décadas.

RESULTADOS

Em geral, a seção de resultados começa com uma descrição de como a amostra foi selecionada e quantos



Conceitos e medidas de desempenho de um teste diagnóstico ou prognóstico; Toda decisão clínica é baseada, conscientemente ou não, em probabilidade; Testes diagnósticos podem ser utilizados para avaliar presença ou ausência de doença, para avaliar a gravidade do quadro clínico, para monitorar a resposta a uma intervenção e para estimar o prognóstico; a - número de resultados verdadeiro-positivos (VP); b - número de resultados falso-positivos (FP); c - número de resultados falso-negativos (FN); d - número de resultados verdadeiro-negativos (VN); Especificidade (E) - probabilidade de exame negativo nos não doentes; Sensibilidade (S) - probabilidade de exame positivo nos doentes; Acurácia do exame - proporção de resultados corretos; Valor preditivo negativo (VPN) - probabilidade de não haver a doença em pessoas com teste negativo; Valor preditivo positivo (VPP) - probabilidade de doença em pessoas com teste positivo; P - na fórmula dos valores preditivos, significa prevalência da doença na população; Razão de verossimilhança de um teste positivo (RV+) - mede o quão mais provável ser o teste positivo nos doentes que nos não-doentes; Razão de verossimilhança de um teste negativo (RV-) - mede o quão mais provável ser o teste negativo nos doentes que nos não doentes; Fração de verdadeiro-positivos (FVP) - total de exames positivos em doentes; Fração de falso-positivos (FFP) - total de exames positivos em não doentes; Curva ROC - curva Receiver Operator Characteristic. É usada para comparar um exame com resultado contínuo em relação a um “padrão-ouro” ou a um desfecho. Trata-se de um gráfico de dispersão com eixo y = sensibilidade (FVP) e eixo x = 1-especificidade (FFP). O ponto do gráfico no canto mais alto superior esquerdo é o ponto ideal de desempenho do exame, com sensibilidade = 100% e especificidade = 100%; DOR: razão de chances de diagnóstico, difícil de ser interpretada clinicamente, mas muito útil do ponto de vista estatístico para avaliar o desempenho global do teste e também muito útil na meta-análise, porque ajuda na construção da curva sROC (summary ROC, resultados agrupados de vários estudos na forma de curva ROC). Fonte: Sousa MR, Ribeiro AL. Systematic review and meta-analysis of diagnostic and prognostic studies: a tutorial. Arq Bras Cardiol. 2009;92(3):229-38, 235-45.¹⁰

Figura 2: Métodos estatísticos de avaliação diagnóstica.

Tabela 1. Acrônimo para iniciar a avaliação de um artigo: PICOTS.

P	População estudada, características clínicas, gravidade ou estágio da doença
I	Intervenção terapêutica nos ensaios clínicos, ou exposição nas coortes, ou teste índice em estudos diagnósticos
C	Comparação com placebo nos ensaios clínicos, ou com grupo comparador não exposto nas coortes, ou com teste de referência em estudos diagnósticos
O	<i>Outcome</i> em inglês, desfecho. Observar se é desfecho clínico ou substitutivo
T	Tempo de seguimento
S	<i>Study design</i> em inglês, delineamento do estudo

Tabela 2. Evolução da metodologia de pesquisa nas últimas décadas.

Marcos históricos em intervenções e tratamentos e a evolução da metodologia de pesquisa		
Década	Intervenções e tratamentos	Métodos relacionados
1940-50	Antibióticos (penicilina e estreptomicina), diálise renal, anestesia geral, radioterapia, circulação extracorpórea, vacina para influenza, exame de Papanicolau, corticosteroides, lente intraocular para catarata.	Primeiros ensaios clínicos
1950-60	Ressuscitação cardiopulmonar, transplante renal, vacina para poliomielite, clotpromazina para esquizofrenia, microscopia fluorescente, tratamento para tuberculose, marca-passo, valva cardíaca artificial, cirurgia cardíaca aberta.	Metodologia de caso-controle e curva de sobrevivência de Kaplan-Meier
1960-70	Substituição de quadril, enxerto coronariano, transplante cardíaco, contraceptivo oral, diagnóstico pré-natal de síndrome de Down.	Ensaios clínicos pragmáticos versus explanatórios, comitês externos independentes de monitorização de estudos, crescimento de comitês de métodos de pesquisa observacional.
1970-80	Cura para alguns cânceres infantis, cuidado intensivo neonatal, tomografia computadorizada, medidas de qualidade no cuidado à saúde, cirurgia ambulatorial, vacina contra caxumba, varicela, rubéola, pneumonia.	Modelo de riscos proporcionais de Cox, metanálises, ascendência de ensaios clínicos aleatorizados, regras estatísticas de interrupção de estudos.
1980-90	Terapia com insulina para diabetes, trombolíticos, anti-hipertensivos, ressonância magnética, cirurgia robótica, implante de coração artificial permanente, estimulação cerebral elétrica profunda, cirurgia a laser na córnea, vacina para hepatite B.	Escore de propensão, ensaios clínicos simples e de grandes amostras, modelos prognósticos (por exemplo, Framingham), análises de decisão e custo-efetividade.
1990-00	Stents coronarianos, terapia tripla para HIV, transplante facial, vacina para hepatite A, vacina para rotavírus.	Medicina baseada em evidências, metanálise cumulativa, CONSORT (diretriz para ensaios clínicos), registros eletrônicos, cadeias de Markov e amostragem de Monte Carlo para inferência Bayesiana.
2000-10	Projeto Genoma Humano completo, stents farmacológicos, guia do FDA para desfechos relatados pelos pacientes, técnicas cirúrgicas minimamente invasivas, vacina para HPV.	Registro de ensaios clínicos no clinicaltrials.gov , pesquisa comparativa de efetividade, pesquisa genômica de larga escala, pesquisa em reprodução.
2010-	Genômica, epigenômica, medicina individualizada, telessaúde.	Pesquisa com resultados centrados no paciente.

foram os pacientes recrutados. A primeira tabela usualmente descreve as características clínicas dos pacientes. Os testes que comparam variáveis entre os dois grupos geram um valor de *p*, que é a probabilidade de ocorrência daquele resultado sob a hipótese de igualdade.¹² Quando o valor de *p* é muito baixo, em geral abaixo de 0,05, consideramos que se trata de uma diferença estatisticamente significativa.

Nem tudo que é estatisticamente diferente tem relevância clínica.¹³ A relevância clínica é dada pelo tamanho da diferença (ver Tabela 3), o intervalo de confiança da medida de tamanho de efeito e pelo julgamento clínico. Portanto, os leitores devem estar atentos ao tamanho da diferença absoluta entre os dois grupos e aos extremos dos intervalos de confiança. O intervalo de confiança diz que o efeito real pode estar dentro daquela faixa de valores e nos mostra que o efeito é menos relevante quando um dos extremos está muito próximo da neutralidade.

Se há dois grupos sendo comparados no estudo, é muito importante observar se se tratam de grupos comparáveis, ou seja, se são indivíduos semelhantes em relação às variáveis que podem influenciar o desfecho. Estudos aleatorizados em geral comparam dois grupos semelhantes, mas pode haver alguma variável que não ficou igualmente distribuída mesmo com uma boa aleatorização e o ideal é fazer ajustes estatísticos para as variáveis que são diferentes entre os dois grupos.

“Variáveis de confusão” são variáveis que influenciam o desfecho e que não são o foco de interesse ou a exposição que está sendo avaliado. Por exemplo, se o desfecho é mortalidade entre pacientes tratados ou não com droga nova, a idade diferente entre os dois grupos pode ser uma variável determinante de morte além do efeito da droga em estudo. Tais variáveis são chamadas covariáveis em modelos de regressão (Cox, Linear, Logística, outros). Com foco

no objeto de estudo - ou seja, a relação entre exposição e desfecho - precisamos lembrar e tentar controlar outras variáveis que podem influenciar aquele desfecho de interesse.

Como leitores de artigo avaliam se o pesquisador se preocupou corretamente em avaliar isto? Primeiro, pela presença de aleatorização em estudos comparativos. Segundo, pelas tabelas de comparação de grupos, que geralmente são a primeira tabela de resultados, onde se pode ver se os grupos são semelhantes em relação às outras variáveis que podem influenciar o desfecho. E terceiro, se o investigador controlou, através de modelos estatísticos, as variáveis de confusão presentes desproporcionalmente nos grupos e outras variáveis relevantes, mesmo que igualmente distribuídas nos dois grupos. O controle estatístico dos fatores e covariáveis de confusão é feito por modelos de regressão, com ajuste (nos quais podem ser utilizados escores de propensão).

Outra regra geral que precisa ser lembrada é o “tamanho ou força de associação”. Em estudos com baixo risco de viés, quanto maior é a força de associação, mais peso damos à interpretação. Quanto maior for a HR, ou o RR, ou a OR, ou quanto maior for a medida de acurácia de um teste diagnóstico, maior será o impacto daquela exposição sobre aquele desfecho. No caso de ensaios clínicos, podemos estimar melhor o impacto individual de uma intervenção através do “número necessário de pacientes tratar para evitar um desfecho” (do inglês *number needed to treat* ou NNT). O NNT pode ser calculado a partir de um único ensaio clínico ou através de uma metanálise, dividindo-se 100 pela redução absoluta de risco (Tabela 3).

Além do efeito esperado/benéfico entre a exposição e o desfecho, é necessário considerar os eventos adversos do teste ou do tratamento em questão, no caso de estudos diagnósticos ou de ensaios clínicos. No caso de estudos de

Tabela 3. Exemplos de estudos com mesma redução relativa do risco e diferentes redução absoluta do risco.

Exemplo	Taxa de eventos no grupo controle	Taxa de eventos no grupo intervenção	Redução relativa do risco	Redução absoluta do risco	NNT
1	0,6%	0,3%	50%	0,3%	333
2	16%	8%	50%	8%	13

Observe que a redução relativa do risco é menos informativa individualmente do que a redução absoluta do risco. No exemplo 1, seria necessário tratar 333 pacientes para reduzir um desfecho primário, enquanto que, no exemplo 2, seria necessário tratar 13 pacientes para reduzir um desfecho primário.

teste diagnósticos, há discussões importantes na literatura sobre o risco de falso positivo ou falso negativo de exames em contextos específicos¹⁴ e sobre o problema do excesso de diagnóstico (*overdiagnosis*) em exames de triagem ou “*check-up*”.¹⁵ No caso de ensaios clínicos, o rigor do investigador em observar e relatar eventos adversos é frequentemente negligenciado pelos leitores na avaliação do impacto negativo de uma nova terapia. Além de calcular o NNT, precisamos calcular o “número necessário de pacientes tratar para adicionar um evento adverso” (do inglês *number needed to harm* ou NNH) da mesma maneira que se calcula o NNT.

No caso de estudos diagnósticos (vide Figura 2), medidas globais que dão uma ideia de acurácia global de um teste são a acurácia (soma de verdadeiros positivos mais verdadeiros negativos dividida pela soma do total de resultados), a *odds ratio* diagnóstica (multiplicação de verdadeiros positivos vezes verdadeiros negativos dividida pelo produto de falsos positivos vezes falso negativos) e a área sob a curva ROC. Quanto maior for qualquer destas três medidas de efeito, maior é a acurácia do teste. Em revisões sistemáticas de métodos diagnósticos, podemos comparar a acurácia de exames através da curva *summary ROC*¹⁰ ou *Hazard Ratio*.

Dois conceitos gerais em interpretação de metanálises são heterogeneidade e viés de publicação, este já mencionado. A heterogeneidade pode ser clínica ou estatística. A heterogeneidade clínica pode ser observada facilmente pelas diferenças de características clínicas entre os pacientes dos estudos incluídos na metanálise e pelos diferentes resultados dos estudos. A heterogeneidade estatística refere-se à dispersão dos resultados individuais dos estudos em relação à medida sumarizada do efeito medido. A heterogeneidade estatística é usualmente avaliada pela medida de inconsistência (I^2). Esta medida é simples de interpretar e dá uma ideia sobre o quanto os resultados dos estudos incluídos em uma metanálise são diferentes entre si. Como regra geral, I^2 em torno de 25% significa heterogeneidade leve, 50% heterogeneidade moderada e em torno de 75% heterogeneidade acentuada. O leitor deve ser cauteloso com metanálises muito heterogêneas.

VISÃO ESPECÍFICA SOBRE ARTIGOS CIENTÍFICOS

O que chamamos aqui de visão específica tem duas dimensões: a) conhecimento do conteúdo específico ou tema a que se refere determinado artigo e b) regras específicas para cada tipo de estudo. O conteúdo específico do tema remete o leitor a pensar por si mesmo. Em geral, antes de ler um artigo, a pessoa deve ter algum conhecimento sobre o assunto abordado, seja pela sua formação acadêmica, livros e tratados sobre o tema. Repetindo: primeiro é preciso ler o livro, para depois ler artigos científicos de atualização. Este conhecimento auxilia a pensar nas covariáveis que podem influenciar o desfecho, na plausibilidade e na consistência dos resultados. Ao comparar artigos sobre um mesmo tema,

seja de forma simplificada ou através de uma metanálise, podemos observar que os vários estudos sobre uma mesma questão científica podem ter resultados diferentes.

Um bom conhecedor do tema pode averiguar que os resultados são diferentes por causa da diferença clínica entre os pacientes, o que chamamos em metanálise de heterogeneidade clínica. Por exemplo, no caso de apenas dois estudos avaliando a mesma questão, os resultados podem ser diferentes por causa da idade média de um estudo mais avançada do que a idade média de outro estudo. Ou por causa de formas adicionais de tratamento diferentes. Há vários vieses diferentes entre os estudos que podem provocar resultados diferentes, mas, em geral, um leitor que conhece o tema pode ver diferenças clínicas e demográficas entre os grupos que estão sendo comparados. Se há diferenças clínicas entre os grupos que estão sendo comparados, isto pode explicar resultados diferentes não relacionados ao efeito da exposição (fator de risco ou tratamento) de interesse sobre o desfecho.

As regras específicas para os vários delineamentos de estudo não podem ser resumidas em um único artigo. Existem diretrizes para executar e descrever cada tipo de estudo e aqui oferecemos referências e endereços eletrônicos ao leitor no Quadro 1. É rápido e fácil consultar tais referências em situações específicas de leitura de artigos científicos, estando disponíveis rápidos *check-lists* produzidos para auxiliar na averiguação da qualidade do estudo. Brevemente, nos ensaios clínicos é muito importante o cegamento, o cálculo amostral, o rigor de que o grupo controle receba o melhor tratamento disponível. Nos estudos diagnósticos, quem realiza o teste índice deve estar “cego” para o resultado do teste de referência e vice-versa. A seleção dos pacientes para estudos diagnósticos é um problema comum quando um dos testes é invasivo, por exemplo, fazer cateterismo em quem já tem teste funcional não invasivo positivo e não fazer em quem o resultado do teste funcional foi negativo. Isto cria um viés que aumenta a probabilidade de doença em quem faz o teste invasivo. Nos estudos transversais, os métodos comparados ou desfechos e exposições devem ter sido avaliados em momentos próximos para evitar mudanças no quadro clínico com o tempo.

PENSAMENTO BAYESIANO

As metanálises e revisões sistemáticas têm mostrado, cada vez mais, que apenas um grande ensaio clínico não é suficiente para responder uma questão. Se um grande ensaio clínico, em geral caro e de difícil realização, não é suficiente para responder uma questão, imaginem estudos de qualidade menor.¹⁶ Sendo assim, como vamos avaliar se aplicamos ou não, na prática, os resultados de um estudo que lemos? Para demonstrar tal dificuldade, vou resumir uma publicação recente, em que 29 equipes de pesquisadores foram convocadas para responder a mesma pergunta de

Quadro 1. Endereços eletrônicos úteis para diretrizes ou orientações de como escrever ou como realizar os diferentes tipos de estudo.

Como relatar estudos observacionais: http://www.strobe-statement.org/
Análise de qualidade de estudos não aleatorizados: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp
Diretrizes de pesquisa qualitativa: http://onlinelibrary.wiley.com/doi/10.1348/014466599162782/full
Diretriz de avaliação estatística de escalas e questionários: http://www.ncbi.nlm.nih.gov/pubmed/11480471
Itens essenciais para avaliar estudos diagnósticos: http://www.stard-statement.org/
Padrões consolidados para realização e descrição de ensaios clínicos: http://www.consort-statement.org/
Análise de qualidade de revisões sistemáticas de métodos diagnósticos: http://www.bris.ac.uk/quadas/
Diretriz de como realizar revisão sistemática e metanálise: http://www.prisma-statement.org/
Avaliação e hierarquização da qualidade de evidências para fazer recomendações e diretrizes baseadas em valores e preferências dos pacientes: http://www.gradeworkinggroup.org/

Todos os endereços foram acessados em 05/03/2016.

pesquisa com o mesmo conjunto de dados de mais de 2000 participantes.

Mais de um terço dos grupos de pesquisa concluiu que o resultado da associação entre duas variáveis era neutro, ao mesmo tempo em que a metade concluiu que o resultado era positivo. Quando os pesquisadores foram reunidos e puderam discutir seus métodos, chegaram a um consenso de uma associação levemente positiva. Isto mostra que pesquisas quantitativas, aparentemente objetivas, sobre uma mesma questão, podem divergir. A conclusão é que precisamos discutir, encontrar e corrigir falhas metodológicas e explicar divergências com o uso de diferentes modelos estatísticos para nos aproximar da “verdade” em uma questão científica.¹⁷

Diante das incertezas e dentro de nossos enclausuramentos de percepção e linguagem, o pensamento Bayesiano fornece uma abordagem probabilística para a aprendizagem. O pensamento Bayesiano é aquele que considera o cenário prévio em conjunto com os dados novos acrescentados por um estudo para estimar a probabilidade de que o cenário possa ser modificado a partir daquela informação. É a visão dos estudos científicos como testes diagnósticos da realidade.¹⁸

A probabilidade de um teste positivo ser verdadeiramente positivo é dada pelo teorema de Bayes. Por este teorema, o valor preditivo de um teste depende da probabilidade pré-teste e da acurácia do teste. Analogamente, o valor de um estudo depende do conhecimento prévio (plausibilidade e consistência) ao estudo e da qualidade (acurácia, risco de viés) do estudo. Há um julgamento clínico subjetivo na incorporação de novas evidências e é importante que o profissional desenvolva a capacidade de julgar e avaliar para não ceder às pressões e interesses. O pensamento Bayesiano incorpora alguma subjetividade da interpretação do contexto de conhecimento prévio ao artigo simultaneamente ao resultado do próprio artigo, para uma tomada de decisão.

Uma vez que o leitor tem uma opinião formada, pode discutir com colegas e reformular sua opinião. Diante de situações de incerteza e diante de mais de uma opção de tratamento ou diagnóstico com diferentes resultados e riscos, é importante apresentar o cenário para o paciente. O paciente autônomo, livre e esclarecido, participa do processo de decisão. Os valores e preferências individuais dos pacientes podem determinar escolhas diferentes conforme suas prioridades. As pessoas podem ter preferências diferentes por graus de resultado e risco diferentes. O planejamento da assistência a um paciente individual deve

ser compartilhado com o próprio interessado, definindo objetivos, acompanhando resultados e redefinindo metas.¹⁹

Assim, como recomendações práticas, propomos o seguinte roteiro de 10 itens para interpretar um artigo científico:

1. Quem é o autor, a instituição, a revista?;
2. Qual é o objeto de estudo? (incidência, prevalência, medida de associação entre qual exposição e qual desfecho clínico ou substitutivo?);
3. Qual era a amostra calculada? (Foi atingida em “resultados?”);
4. PICOTS;
5. Qual é o risco de viés do estudo conforme a avaliação de qualidade para cada tipo específico de estudo?;
6. Qual é o impacto estatístico ou tamanho de efeito ou força de associação, NNT, NNH, ou acurácia?;
7. É biologicamente e fisiopatologicamente plausível?;
8. É consistente com outros estudos da literatura?;
9. Quais são as implicações de se aplicar os resultados do estudo na prática e quais são as implicações de não se aplicar os resultados do estudo na prática?;
10. Discussão com outros especialistas na tentativa de estabelecer um protocolo de abordagem do problema até que novos estudos surjam e mais importante: o paciente deve participar da decisão de acordo com seus valores e preferências.

Para encerrar, uma reflexão de David Sackett: “o especialista, aquele no ápice da expertise sobre um tema, precisa saber o momento certo de se calar sobre suas crenças arraigadas em sua experiência e deixar o conhecimento evoluir através dos questionamentos e novas evidências apresentadas pelos mais jovens na especialidade”.²⁰ Tal reflexão, vinda de um profissional que decidiu se calar em dois momentos de expertise ao longo de sua vida profissional reconhecida e valorizada, alerta os jovens leitores para o grande valor de suas próprias reflexões, de suas próprias observações.

Ao ler um texto, a retórica deste texto pode estar fundamentada no *ethos*, no *logos* e/ou no *pathos*. O *ethos* é a circunstância de quem fala, ou seja: os jovens são muito influenciados pela autoridade do falante e não pela força e qualidade de seus argumentos. O *pathos* é a motivação pelo sentimento, pela gravidade da situação do paciente, que pode nos levar a tentativas fúteis simplesmente por desespero. O *logos* é o uso da boa argumentação na defesa de uma conduta, articulada através de argumentos válidos, baseada

em plausibilidade, consistência, qualidade da evidência e força de associação, aliadas aos valores e preferências do paciente.

REFERÊNCIAS

1. Ranasinghe I, Shojaee A, Bikdeli B, Gupta A, Chen R, Ross JS, *et al.* Poorly cited articles in peer-reviewed cardiovascular journals from 1997 to 2007: analysis of 5-year citation rates. *Circulation*. 2015;131(20):1755-62.
2. Jacobs AK, Anderson JL, Halperin JL; ACC/AHA TASK FORCE MEMBERS, Anderson JL, Halperin JL, Albert NM, *et al.* The evolution and future of ACC/AHA clinical practice guidelines: a 30-year journey: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines. *Circulation*. 2014;130(14):1208-17.
3. Pereira TV, Ioannidis JP. Statistically significant meta-analyses of clinical trials have modest credibility and inflated effects. *J Clin Epidemiol*. 2011;64(10):1060-9.
4. Fara P. Uma Breve História da Ciência. São Paulo: Fundamento; 2014.
5. Lawson H. Closure: a History of Everything. Oxford: Routledge; 2001.
6. Kanehman D. Rápido e Devagar: Duas Formas de Pensar. São Paulo: Objetiva; 2013.
7. Silver N. O Ruído e o Sinal. Rio de Janeiro: Intrínseca; 2013.
8. Taleb NN. A lógica do Cisne Negro. Rio de Janeiro: Best Seller; 2007.
9. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997;315(7109):629-34.
10. Sousa MR, Ribeiro AL. Systematic review and meta-analysis of diagnostic and prognostic studies: a tutorial. *Arq Bras Cardiol*. 2009;92(3):229-38, 235-45.
11. Gabriel SE, Normand SL. Getting the methods right--the foundation of patient-centered outcomes research. *N Engl J Med*. 2012;367(9):787-90.
12. Goodman S. A dirty dozen: twelve p-value misconceptions. *Semin Hematol*. 2008;45(3):135-40.
13. Willenheimer R. Statistical significance versus clinical relevance in cardiovascular medicine. *Prog Cardiovasc Dis*. 2001;44(3):155-67.
14. Cota GF, de Abreu Ferrari TC, de Sousa MR. HIV testing strategies: test performance is important, but not sufficient. *Expert Rev Anti Infect Ther*. 2011;9(2):207-14.
15. Qaseem A, Alguire P, Dallas P, Feinberg LE, Fitzgerald FT, Horwitch C, *et al.* Appropriate use of screening and diagnostic tests to foster high-value, cost-conscious care. *Ann Intern Med*. 2012;156(2):147-9.
16. Ioannidis JP. Why most published research findings are false. *PLoS Med*. 2005;2(8):e124.
17. Silberzahn R, Uhlmann EL. Crowdsourced research: Many hands make tight work. *Nature*. 2015;526(7572):189-91.
18. Ioannidis JP, Tarone R, McLaughlin JK. The false-positive to false-negative ratio in epidemiologic studies. *Epidemiology*. 2011;22(4):450-6.
19. Peterson ED, Ho PM, Barton M, Beam C, Burgess LH, Casey DE Jr, *et al.* ACC/AHA/AACVPR/AAFP/ANA concepts for clinician-patient shared accountability in performance measures: a report of the American College of Cardiology/American Heart Association Task Force on Performance Measures. *Circulation*. 2014;130(22):1984-94.
20. Sackett DL. The sins of expertness and a proposal for redemption. *BMJ*. 2000;320(7244):1283.