

<https://doi.org/10.22491/2357-9730.96394>

Bioestatística

BIOESTATÍSTICA E EPIDEMIOLOGIA: PERGUNTAS QUE VOCÊ SEMPRE QUIS FAZER, MAS NUNCA TEVE CORAGEM

BIOSTATISTICS AND EPIDEMIOLOGY: QUESTIONS YOU ALWAYS WANTED TO ASK BUT NEVER HAD THE COURAGE TO

Stela Maris de Jezus Castro^{1,2}, Aline Castello Branco Mancuso¹,
Vanessa Bielefeldt Leotti^{1,2}, Vânia Naomi Hirakata¹,
Suzi Alves Camey^{1,2}

RESUMO

Dando continuidade aos artigos da série “Perguntas que você sempre quis fazer, mas nunca teve coragem”, que tem como objetivo responder e sugerir referências para o melhor entendimento das principais dúvidas estatísticas levantadas por pesquisadores da área da saúde, este terceiro artigo aborda o contexto epidemiológico. Neste contexto, foram diferenciadas as principais medidas como prevalência, incidência, Odds Ratio (OR), Risco Relativo (RR), Razão de Prevalência (RP) e Hazard Ratio (HR), foi esclarecido o uso de análises por intenção de tratar e análise por protocolo, e também discutidos alguns dos termos comumente utilizados e pouco compreendidos como tipo de amostra, nível de evidência, relevância clínica e estatística, entre outros.

Palavras-chave: *Delineamentos; nível de evidência; medidas de associação; ensaios clínicos; análise por intenção de tratar; análise por protocolo; amostra aleatória; amostra por conveniência; randomização em blocos*

ABSTRACT

Continuing the series of articles “Questions you always wanted to ask but never had the courage to,” which aims to answer key statistical questions raised by health researchers and suggest references for a better understanding, this third article addresses the epidemiological context. In this context, important measures such as prevalence, incidence, odds ratio (OR), relative risk (RR), prevalence ratio (PR) and hazard ratio (HR) were differentiated; the use of intention-to-treat analysis and per-protocol analysis was clarified; and some terms commonly used and poorly understood were discussed, such as type of sample, level of evidence, clinical and statistical relevance, among others.

Keywords: *Designs; level of evidence; measures of association; clinical trials; intention-to-treat analysis; per-protocol analysis; random sample; convenience sample; block randomization*

o contexto estatístico, na área epidemiológica, é fonte de diversas confusões por parte dos pesquisadores. Fato preocupante, visto que muitas destas afetam diretamente o desenvolvimento de estudos e pesquisas, assim como suas conclusões. Como pôde ser evidenciado, via enquête, a maior parte das dúvidas dos pesquisadores não se deve a cálculos matemáticos, mas sim à compreensão teórica dos termos e conceitos utilizados.

Visando dar continuidade aos artigos da série “Perguntas que você sempre quis fazer, mas nunca teve coragem”, que tem como objetivo responder e sugerir referências para o melhor entendimento das principais dúvidas estatísticas levantadas em uma enquête com pesquisadores da área da saúde¹, este artigo pretende esclarecer algumas das dúvidas levantadas no contexto epidemiológico. São respondidas 11 questões envolvendo conceitos gerais de Bioestatística e Epidemiologia.

Clin Biomed Res. 2019;39(3):258-265

1 Unidade de Bioestatística, Grupo de Pesquisa e Pós-graduação (GPPG), Hospital de Clínicas de Porto Alegre (HCPA). Porto Alegre, RS, Brasil.

2 Departamento de Estatística, Instituto de Matemática e Estatística, Universidade Federal do Rio Grande do Sul (UFRGS). Porto Alegre, RS, Brasil.

Autor correspondente:

Stela Maris de Jezus Castro
l-bioestatistica@hcpa.edu.br
Grupo de Pesquisa e Pós-Graduação (GPPG), Hospital de Clínicas de Porto Alegre (HCPA)
Rua Ramiro Barcelos, 2350.
90035-903, Porto Alegre, RS, Brasil.

O QUE É E QUAIS SÃO OS TIPOS DE DESFECHOS?

Desfecho, do inglês *outcome*, pode ser traduzido como resultado de um problema, de um momento difícil. Em pesquisas epidemiológicas, o desfecho tem sido tradicionalmente definido como a variável que representa o resultado de uma doença, de um período de tratamento ou uma consequência na saúde, mas também pode ser o resultado de uma simples observação. Em toda pesquisa epidemiológica é possível definir um ou mais desfechos como, por exemplo: indicadores de saúde em uma população específica; medidas de status funcional relacionados à idade; ou ocorrência de recidiva e óbito em estudos oncológicos².

Em termos estatísticos, pode-se definir **desfecho** como a variável mais relevante do estudo, que serve para testar a hipótese em questão. Ele pode ser medido de forma categórica (severidade da doença, ocorrência de óbito etc.) ou de forma contínua (nível de pressão sanguínea, níveis de glicemia, tempo até determinado evento etc.).

A determinação do desfecho (ou dos desfechos) de um estudo depende primordialmente do objetivo a ser alcançado, pois é ele que responderá o mesmo. Logo, na fase de elaboração do projeto de pesquisa, delinear com clareza e precisão o objetivo (ou objetivos) é imprescindível para definir qual (is) o(s) desfecho(s) a ser(em) medido(s) na amostra em estudo. O Quadro 1 apresenta alguns exemplos de objetivos de estudos e seus respectivos desfechos.

Quadro 1: Exemplos de objetivos e desfechos de estudos.

Objetivo do Estudo	Desfecho	Tipo de Desfecho
Estimar a prevalência de sífilis congênita em recém-nascidos ³ .	Resultado do exame de sífilis	Categórico binário $\begin{cases} 0 \text{ não} \\ 1 \text{ sim} \end{cases}$
Avaliar a eficácia do opicinumab em pacientes com esclerose múltipla ⁴ .	Melhoria da deficiência	Categórico binário $\begin{cases} 0 \text{ não} \\ 1 \text{ sim} \end{cases}$
Identificar os efeitos da reposição hormonal com estrógenos e progesterona (comparados ao placebo) nos níveis de Lipoproteína em mulheres pré-menopáusicas ⁵ .	Diferença no nível de Lipoproteína entre o início e fim do tratamento	Quantitativo contínuo
Comparar dois tratamentos imunoterápicos quanto a sua capacidade de aumentar a sobrevida de pacientes que sofrem de melanoma ⁶ .	Tempo até a ocorrência de óbito	Quantitativo contínuo

QUAIS OS TIPOS DE DELINEAMENTOS?

Um dos tópicos identificados em várias das respostas recebidas na enquete foi a confusão e desconhecimento dos delineamentos de pesquisa. Enquadrado como uma das principais questões a serem definidas em qualquer projeto de pesquisa, o delineamento deve contemplar o planejamento da pesquisa em um sentido amplo, compreendendo não apenas os objetivos e o contexto no qual os dados são analisados, mas também a metodologia de desenvolvimento, coleta e análise. Em termos gerais, o delineamento é responsável por caracterizar o estudo.

Existem diversos tipos de delineamentos, mas podem ser classificados em dois grandes grupos: estudos observacionais ou estudos experimentais. Dentre os principais estudos observacionais,

destacam-se o estudo de coorte, o estudo de caso-controle e os estudos transversais (estes ainda podem ser classificados como estudos de prevalência, estudos de teste diagnóstico ou estudos ecológicos, entre outros). Já dentre os estudos experimentais, os mais conhecidos na área da saúde são os chamados ensaios clínicos. Os estudos observacionais também podem ser classificados em analíticos ou observacionais, enquanto os estudos experimentais também são classificados em randomizados ou não. A Figura 1 ilustra esse conceito.

Em edições anteriores desta revista já foram detalhados os diferentes tipos de delineamentos observacionais⁷ e os ensaios clínicos^{8,9}. Num contexto mais teórico, Hulley et al.⁵ e Rothman et al.¹⁰ descrevem e contextualizam alguns dos delineamentos.

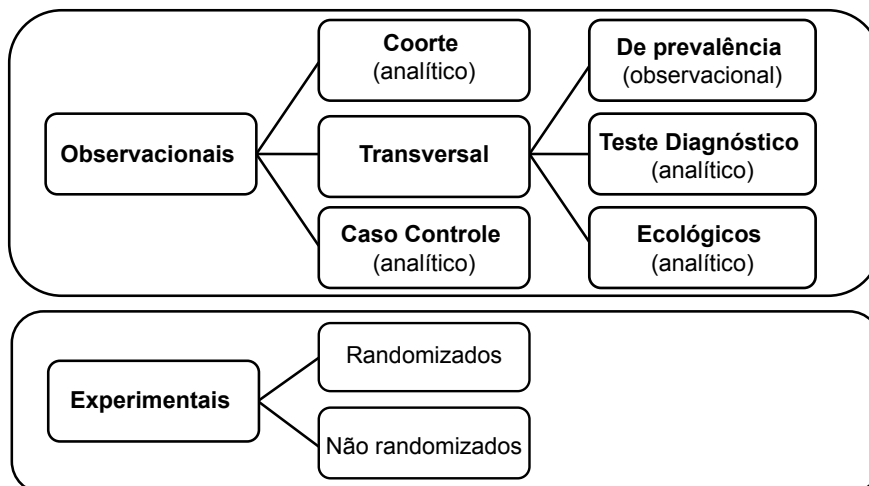


Figura 1: Classificação dos Delineamentos.

O QUE É NÍVEL DE EVIDÊNCIA?

De acordo com Guyatt et al.¹¹, o termo evidência pode receber diferentes interpretações conforme o contexto utilizado. Em termos científicos, o Sistema GRADE (do inglês *Grading of Recommendations Assessment, Development and Evaluation*), adotado pelo Ministério da Saúde, define como *evidência* os resultados (conjunto de informações) de pesquisas clínicas claras, concisas e relevantes utilizados para confirmar ou refutar uma hipótese científica^{12,13}.

O sistema GRADE¹² fornece diretrizes sobre a qualidade da evidência e sua força de recomendação para tomada de decisões em saúde. Neste manual, a avaliação da qualidade da evidência é realizada para cada desfecho analisado para uma dada tecnologia, a partir do conjunto de evidências disponíveis para a questão de pesquisa, podendo ser classificada em quatro níveis. Estes *níveis de evidência* representam a confiança que se tem na estimativa dos resultados apresentados¹³.

A classificação se inicia a partir do delineamento dos estudos. Por exemplo, ensaios clínicos randomizados são considerados delineamentos mais adequados para estudos de intervenções e, por isso, a qualidade da evidência já inicia alta. No caso de

estudos observacionais, a qualidade da evidência apresentada inicia como baixa. Já a classificação como nível de evidência muito baixo é utilizada para recomendações oriundas de opiniões de especialistas. No entanto, a classificação inicial do nível de evidência pode ser alterada após a avaliação de outros fatores. Ela pode ser reduzida devido a limitação metodológica, inconsistência, evidência indireta, imprecisão ou viés de publicação, entre outros. Ou elevada quando apresenta fatores que aumentem a confiança nos resultados apresentados como, por exemplo, grande magnitude de efeito¹³. A Figura 2 ilustra este conceito. Para melhor compreender esta forma de qualificação dos trabalhos, pode ser visto um exemplo em Camanho¹⁴ que traduz a metodologia utilizada para o cotidiano.

Contudo, existem diversos trabalhos sobre a qualidade e a quantidade dos níveis de evidência, variando entre as diferentes áreas da saúde. Galvão (2006)¹⁵ destaca alguns dos sistemas de classificação utilizados na área de enfermagem. Anvisa (2015)¹⁶ discrimina os níveis de evidências de estudos clínicos com medicamentos oncológicos em função dos desfechos clínicos. Já o site da CEBM (do inglês *Centre for Evidence-Based Medicine*)¹⁷ apresenta 10 níveis para medicina baseada em evidências.



Figura 2: Qualidade de Evidência dos Delineamentos.

QUAL A DIFERENÇA ENTRE ENSAIO CLÍNICO DE COMPARAÇÃO, ENSAIO CLÍNICO DE SUPERIORIDADE, DE EQUIVALÊNCIA E DE NÃO-INFERIORIDADE?

Na questão referente aos tipos de delineamentos, foi visto que os ensaios clínicos são estudos experimentais, cujo objetivo principal é a avaliação de um novo tratamento⁸. No entanto, o estudo desse novo tratamento pode querer comprovar que ele é melhor (estudo de superioridade) do que o padrão, fixado um limite superior; que ele não é inferior (estudo de não-inferioridade) ao padrão, fixado um limite inferior; ou tão eficaz (estudo de equivalência) quanto o padrão, fixados um limite inferior e um superior.

Há ainda os ensaios clínicos de comparação randomizados, que são, tradicionalmente, descritos como experimentos médicos planejados para avaliar qual de duas ou mais intervenções é a mais eficaz. Em geral, a hipótese do pesquisador é que exista uma diferença entre os grupos comparados, sendo que, na maioria dos ensaios, um dos grupos tratados é considerado como controle, o que pode significar a ausência de tratamento (placebo) ou um tratamento considerado padrão (de eficácia já reconhecida). No entanto, o fato de não se encontrar uma diferença estatisticamente significativa entre os grupos não implica na equivalência das intervenções. Este é um erro clássico cometido na interpretação dos resultados de um ensaio clínico de comparação randomizado¹⁸. Conforme Altman e Bland¹⁹ e Alderson²⁰: “a ausência de evidência não é evidência de ausência”.

Nos ensaios de equivalência, a hipótese do pesquisador é a de que NÃO existe diferença entre

dois ou mais tratamentos. Assim, o objetivo do estudo é provar que a eficácia dos tratamentos é igual, em outras palavras, equivalente. Um exemplo onde os estudos de equivalência são muito utilizados é na área da indústria farmacêutica, para a aprovação regulatória dos medicamentos genéricos. A Anvisa (Agência Nacional de Vigilância Sanitária)²¹ regula e fiscaliza os estudos desta área no Brasil.

Já em outros contextos, a hipótese do pesquisador pode ser a de que um tratamento NÃO é menos eficaz que outro. Para testar hipóteses deste tipo, existem os estudos de não-inferioridade, cujo objetivo é mostrar que um tratamento não é, dentro de certos critérios, menos eficaz que outro já existente. Estes estudos são bastante utilizados em situações nas quais as comparações com placebo são inviáveis ou antiéticas¹⁸.

Como já foi visto anteriormente, o ensaio clínico randomizado é considerado o melhor delineamento para avaliar os efeitos de intervenções clínicas²², partindo de altos níveis de evidência^{12,13}.

O QUE SIGNIFICA E QUANDO UTILIZAR ANÁLISE ITT E ANÁLISE PP?

A análise por intenção de tratar, ITT (do inglês *intention-to-treat*), se aplica a ensaios clínicos randomizados, com dois ou mais grupos de comparação, com o intuito de preservar a distribuição aleatória e minimizar os possíveis erros sistemáticos. Caracterizando-se por incluir todos os indivíduos nos grupos que foram inicialmente alocados, independente da adesão, do desvio de protocolo, da retirada ou troca de tratamento ou do que ocorrer com cada um deles ao longo do estudo. Essa estratégia preserva o benefício da randomização, permitindo a distribuição balanceada de fatores prognósticos nos grupos comparados e evitando vieses devido a exclusão de pacientes. Assim, o efeito observado será realmente devido ao tratamento designado. No entanto, essa será uma estimativa conservadora, visto a inclusão de pacientes que não concluíram ou até mesmo não realizaram o tratamento²³⁻²⁵.

Já a análise por protocolo, PP (do inglês *per-protocol*), caracteriza-se por considerar apenas um subconjunto da amostra inicial, que contém apenas os pacientes que completaram o tratamento originalmente alocado, sem nenhum desvio grave de protocolo^{22,23}. Consequentemente, o efeito observado será sob condições ótimas, dado que os pacientes foram totalmente compatíveis com o tratamento.

Conforme Shah²³, a análise ITT é preferível nos ensaios de superioridade, já para os ensaios de não-inferioridade ainda não há um consenso. Neste tipo de ensaio, o autor recomenda tanto a análise ITT como a análise PP como abordagens, pois ambas devem apoiar as evidências encontradas. Uma comparação entre a análise ITT e análise PP pode ser encontrada na página de Obermeier²⁴.

QUAL A DIFERENÇA ENTRE PREVALÊNCIA E INCIDÊNCIA?

Tanta incidência (incidência cumulativa) quanto prevalência são medidas de frequência de ocorrência de um determinado evento (por exemplo, ocorrência de óbito, ocorrência de uma doença, etc.) em uma determinada população em risco. Em termos matemáticos, é a relação entre as partes de um todo, conforme a Expressão 1.

$$\frac{\text{n}^\circ \text{ de eventos entre os indivíduos observados}}{\text{n}^\circ \text{ total de indivíduos observador}}$$

(Expressão 1)

A diferença crucial entre as medidas está no delineamento do estudo e, portanto, no numerador da Expressão 1. No caso da incidência, conta-se o número de eventos novos (oriundo, por exemplo, de um estudo de coorte), logo deve ser calculada conforme a Expressão 2. Já para prevalência, conta-se o número de eventos existentes (em um estudo transversal, por exemplo), conforme a Expressão 3.

$$\text{Incidência} = \frac{\text{n}^\circ \text{ de eventos novos entre os indivíduos observados}}{\text{n}^\circ \text{ total de indivíduos observador}}$$

(Expressão 2)

$$\text{Prevalência} = \frac{\text{n}^\circ \text{ de eventos exixtentes entre os indivíduos observados}}{\text{n}^\circ \text{ total de indivíduos observador}}$$

(Expressão 3)

Exemplificando, numa população hipotética de 100 indivíduos, 10 estão sendo tratados para hepatite C desde 2015 e 3 foram diagnosticados em 2016. Em 2016, a prevalência de hepatite C será de 13/100 = 0,13 (13%) enquanto a incidência será de 3/100 (3%).

QUANDO UTILIZAR E COMO CALCULAR A INCIDÊNCIA PESSOA-ANO?

Taxa de incidência pessoa/tempo, densidade de incidência, força de morbidade (ou mortalidade), taxa de azar ou intensidade de doença são medidas de incidência que incorporam o tempo da exposição no denominador¹⁰.

A taxa de incidência pessoa-ano (ou pessoa/tempo) é calculada, geralmente, em estudos de coorte,

nos quais indivíduos sem a doença pesquisada são acompanhados ao longo do tempo e a ocorrência de novos casos da doença é observada. Normalmente, cada indivíduo é observado a partir de um determinado momento até que ocorra uma das seguintes situações: início da doença, morte, saída do estudo (perda de seguimento) ou final do estudo.

Semelhante ao cálculo da incidência, discutido anteriormente, o *numerador* da taxa de incidência pessoa-tempo é o número de novos casos identificados durante o período de observação. No entanto, o denominador é diferente. Enquanto no cálculo da incidência o denominador é o número de indivíduos expostos, na taxa de incidência pessoa-tempo o *denominador* é a soma do tempo que cada pessoa foi observada, para todos os indivíduos da amostra. Este denominador representa o tempo total que a população estava em risco e sendo monitorada para o surgimento da doença. Assim, a taxa de incidência pessoa-tempo é dada pela Expressão 4^{10,25}:

$$\text{Taxa de incidência pessoa - tempo} = \frac{\text{n}^\circ \text{ de casos novos identificados durante determinado período}}{\text{tempo total em risco de doença}}$$

(Expressão 4)

Uma das vantagens desta taxa é que os indivíduos podem entrar e sair do estudo em momentos diferentes, como no caso do estudo de acompanhamento da NHANES, onde alguns participantes foram arrolados em 1971, outros em 1972, 1973, 1974 e 1975²⁶. No entanto, o denominador pessoa-tempo tem uma desvantagem importante, pois assume que a probabilidade de doença durante o período do estudo é constante, de modo que 10 pessoas seguidas por um ano são iguais a uma pessoa seguidas por 10 anos. Como o risco de muitas doenças crônicas aumenta com a idade, essa suposição muitas vezes não é válida.

Um estudo de doença cardíaca cuja taxa de incidência é de 2,5 por 1.000 pessoas-ano, é facilmente compreendido pelos epidemiologistas, mas não pela maioria dos demais profissionais, por se tratar de um jargão epidemiológico. Para converter esse jargão em algo mais compreensível, basta substituir “pessoa-ano” por “pessoa por ano”. Relatar os resultados como 2,5 novos casos de doença cardíaca por 1.000 pessoas por ano soa melhor e transmite a sensação da taxa de incidência como um processo dinâmico, indicando a velocidade com que novos casos de doença ocorrem na população.

QUANDO UTILIZAR ODDS RATIO (OR), RISCO RELATIVO (RR), RAZÃO DE PREVALÊNCIAS (RP) OU HAZARD RATIO (HR)?

Odds ratio (OR), risco relativo (RR), razão de prevalências (RP) e *hazard ratio* (HR) são medidas

de associação ou de efeito, conforme exista ou não suposição de causalidade entre o fator em estudo e o desfecho¹⁰. Todas se caracterizam por serem obtidas a partir da razão entre duas medidas de odds, de incidência ou de prevalência.

Para entender melhor as diferenças e os motivos pelos quais se deve utilizar uma ou outra medida de efeito, considere a notação apresentada no Quadro 2, da distribuição de uma variável exposição em relação a presença de certa doença (desfecho).

Quadro 2: Frequência de doentes entre expostos ou não.

	Doentes	Não doentes
Expostos	a	b
Não expostos	c	d

- O *odds ratio* (OR), também conhecido como razão de *odds* ou razão de chances, é calculado, principalmente, em estudos do tipo caso-controle, embora também possa ser calculado em estudos transversais ou longitudinais, quando o desfecho for raro (até 2% de incidência/prevalência, aproximadamente)²⁶. No caso de estudos caso-controle, o *odds ratio* calculado é a razão entre a chance de exposição (razão entre expostos e não expostos) dos doentes em relação a dos não doentes²⁶, conforme a Expressão 5, também podendo ser estimado via Regressão Logística.

$$OR = \frac{\text{Odds de exposição entre os doentes}}{\text{Odds de exposição entre os não doentes}} = \frac{a/c}{b/d}$$

(Expressão 5)

- O risco relativo (RR) é um termo amplo para designar razões de taxas (*rate ratio*) ou razões de risco (*risk ratio*), utilizado em estudos longitudinais. É definido como a razão entre o risco de desenvolver o desfecho nos expostos e o risco de desenvolver o desfecho nos não expostos a partir das taxas de incidência, conforme a Expressão 6 e pode ser estimado via Regressão de Poisson com variâncias robustas ou modelo logbinomial^{2,27}.

$$RR = \frac{\text{Incidência de doença entre expostos}}{\text{Incidência de doença entre não exposto}} = \frac{a/(a+b)}{c/(c+d)}$$

(Expressão 6)

- A razão de prevalências (RP), como o próprio nome diz, é obtida pela razão entre a prevalência

da doença nos expostos e a prevalência da doença nos não expostos, conforme Expressão 7. Seu uso é restrito aos estudos transversais, mas também pode ser estimado via Regressão de Poisson com variâncias robustas ou modelo logbinomial.

$$RP = \frac{\text{Prevalência de doença entre expostos}}{\text{Prevalência de doença entre não exposto}} = \frac{a/(a+b)}{c/(c+d)}$$

(Expressão 7)

- O *hazard ratio* (HR), ou razão de riscos proporcionais, é calculado quando o desfecho de interesse é o tempo até o evento ocorrer. Desta forma, a razão de riscos proporcionais é obtida dividindo o risco (*hazard*) de desenvolver a doença entre os expostos () pelo risco entre os não expostos (), ambos em função do tempo, conforme Expressão 8. No entanto, seu cálculo é feito a partir da regressão de Cox².

$$HR = \frac{h_1(t)}{h_0(t)}$$

(Expressão 8)

É comum a confusão entre as medidas, assim como seu uso incorreto. Cabe salientar que risco relativo não é sinônimo de odds ratio (nem de RP ou HR), apesar de muitas vezes ser utilizado como tal de forma inadequada. Para decidir qual medida deve ser utilizada, deve-se considerar o delineamento do estudo e a análise estatística empregada. Para mais detalhes, aconselha-se a leitura do capítulo 4 (páginas 67-72, mais especificamente) de Rothman et al.¹⁰.

QUAL A DIFERENÇA ENTRE SIGNIFICÂNCIA ESTATÍSTICA E RELEVÂNCIA CLÍNICA?

Essa pergunta é bastante interessante, pois estes conceitos são, por vezes, confundidos, quando na realidade não têm o mesmo significado e nem estão necessariamente relacionados.

A significância estatística é um resultado relacionado a um teste de hipóteses. Um resultado estatisticamente significativo é aquele onde, em um teste de hipóteses, se rejeitou a hipótese nula. É geralmente representada pelo valor de p, conforme visto no segundo artigo desta série, em Hirakata et al. (2019)²⁷.

Conforme o título dos artigos de Alderson²⁰ e Altman e Bland¹⁹ "A ausência de evidência não é evidência de ausência" (em tradução literal). Ou seja, a ausência de significância estatística não significa que o efeito de interesse não existe. E mesmo

quando há significância estatística, não significa que o achado é útil ou relevante para determinar a etiologia da doença, isto é, não prova causalidade. Além disso, a significância estatística não implica na validade externa do achado, ou seja, que ele pode ser generalizado.

Page²⁸ define relevância clínica como a menor mudança no desfecho que é considerada clinicamente importante, tem algum valor para o clínico ou para o paciente, e/ou resultaria numa mudança de conduta para o caso do paciente. Em pesquisa, a relevância clínica é medida pela estimativa pontual e sua incerteza dada pelo intervalo de confiança. Assim, a avaliação da significância estatística é apenas uma parte da avaliação da relevância clínica.

Para ilustrar a diferença entre relevância clínica e significância estatística, considere o exemplo apresentado na Figura 3, onde a linha pontilhada representa uma específica relevância clínica, uma mudança a partir da qual se considera o achado importante clinicamente. Na hipótese nula, temos a ausência de efeito (por exemplo, resultado no grupo de pacientes tratados e no grupo placebo sendo iguais). Cinco estudos são realizados e cinco intervalos de

confiança são estimados. Estatisticamente, aqueles intervalos de confiança que não cruzam o eixo da hipótese nula são os resultados significativos (a, b, c). Já no contexto clínico, o que supera a linha pontilhada é considerado um resultado importante na prática (a); os que não atingem esse eixo não são resultados relevantes (c) ou têm até mesmo efeito contrário ao esperado (e); e os intervalos que cruzam o eixo da relevância clínica (b, d) mostram resultados dúbios, que podem ser importantes ou não. Nos estudos com resultados dúbios ou com estimativas muito amplas (b, d), pode-se concluir que o tamanho amostral não permitiu obter um intervalo suficientemente preciso para obter conclusões mais definitivas, o que pode ser evitado com um cálculo cuidadoso do tamanho de amostra antes de se iniciar a pesquisa. Lembrando que a medida de relevância clínica é um dos principais ingredientes para o cálculo do tamanho da amostra. Já o estudo que não atinge a relevância clínica, porém é estatisticamente significativo (c), exibe a situação em que há perigo de se afirmar que um efeito é importante apenas pela sua significância estatística.

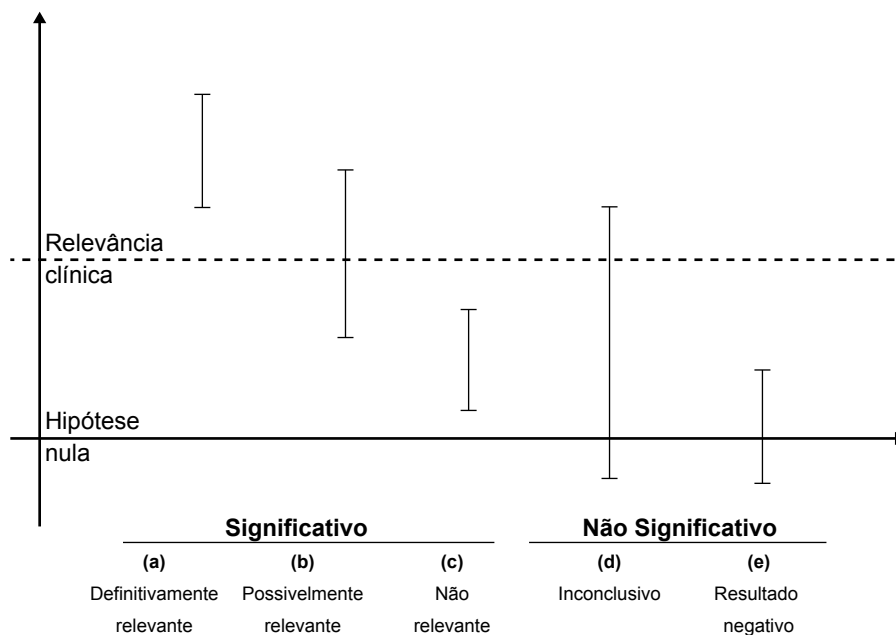


Figura 3: Interpretação conjunta de significância estatística e relevância clínica.

Por fim, é importante enfatizar que a relevância clínica deve ser incorporada na pesquisa ainda na etapa do planejamento, garantido o cálculo apropriado para o tamanho da amostra. E, principalmente, deve ser considerada antes de iniciar o processo de tomada de decisão, para interpretação correta dos achados.

COMO DEFINIR O TIPO DE AMOSTRA: ALEATÓRIA OU POR CONVENIÊNCIA?

Ao desenvolver um estudo sobre determinada população, normalmente não há recursos suficientes para coletar informações e dados de todos os indivíduos, apenas de parte da população, uma amostra.

Assim, a ideia é coletar dados de uma pequena parte de um grande grupo e aprender algo sobre esse grupo maior. A técnica utilizada para se obter as unidades (sujeitos, pacientes, etc) que formarão essa pequena parte se chama amostragem.

Existem dois grandes tipos de amostragem: as probabilísticas e as não-probabilísticas. Nas probabilísticas, algum mecanismo de sorteio é utilizado, por isso a amostra produzida pode ser considerada aleatória. Nada impede que este sorteio seja feito com probabilidades diferentes para cada unidade, ainda assim a amostra permanece aleatória. Já nas amostragens não-probabilísticas, não se utiliza sorteio e sim algum método onde o pesquisador, de certa forma, influenciará nas unidades selecionadas.

A amostra por conveniência é um tipo de amostragem não-probabilística, onde o pesquisador escolhe convenientemente as unidades para participar da amostra. É importante notar que, se um pesquisador define que os pacientes que participarão da amostra serão todos aqueles que vierem a um específico serviço de um hospital em determinado período, por exemplo, mesmo não sabendo de antemão quais pacientes virão, ainda é uma amostra de conveniência (não-aleatória). A ideia é que qualquer definição conveniente que não utiliza sorteio restringe a participação de algumas unidades que não respeitam aquela definição, mas pertencem à população em estudo (por exemplo, pacientes que não utilizam aquele hospital, ou pacientes que não compareceram no hospital no período do estudo, por estarem muito mal ou muito bem), descaracterizando a amostragem aleatória.

Aconselha-se, sempre que possível, a utilização de amostragens probabilísticas, pois as análises estatísticas (intervalos de confiança, testes de hipóteses, etc) são desenvolvidas considerando essa suposição. A teoria estatística é baseada na probabilidade dos resultados acontecerem, quando a amostra não é aleatória não é possível calcular essa probabilidade, o que impossibilita calcular os erros das estimativas. Na prática o que acontece é o uso das técnicas estatísticas mesmo em amostras não probabilísticas. Bolfarine e Bussab²⁹ discutem estes e outros conceitos utilizados em amostragem.

QUANDO UTILIZAR RANDOMIZAÇÃO EM BLOCOS?

A randomização, ou alocação aleatória, é um método para selecionar as unidades (indivíduos, pacientes, etc) elegíveis para o estudo e dividi-los, aleatoriamente, em grupos, de tal forma que o resultado dos grupos não sejam afetados pelas características dos participantes⁸. Não é necessário que os grupos sejam divididos exatamente com o mesmo número de participantes, mas sim que ao final da coleta o tamanho amostral calculado para cada grupo seja atingido. Quando for de interesse manter um certo equilíbrio no número de participantes em cada grupo durante todo o período de coleta, para não chegar no meio do período, por exemplo, com um número muito maior de participantes em um dos grupos, a randomização pode ser feita em blocos.

A randomização em blocos se caracteriza pela formação de blocos de indivíduos, dentro dos quais há a aleatorização (sorteio de quem irá participar de cada grupo do estudo). É utilizada com o objetivo de reduzir viés e tornar os braços mais homogêneos, principalmente em estudos onde o tamanho da amostra não é muito grande^{30,31}. Por exemplo, suponha que o tamanho de amostra total de um estudo seja de 300 indivíduos, sendo 150 para cada braço. A randomização em blocos de 30, por exemplo, garante que serão alocados 15 participantes para cada braço entre os primeiros 30 indivíduos incluídos, depois os próximos 30 da mesma forma e assim por diante, até completar 300. É importante destacar que quanto maior for o tamanho do bloco, menor é a chance do pesquisador deduzir qual o braço que o próximo paciente irá participar. O ideal é que o mínimo de pessoas saibam como a aleatorização foi feita e que os blocos sejam de tamanhos aleatórios, para dificultar a possibilidade de previsibilidade. Bland et al.³² coloca que blocos de 20 ou mais são preferíveis, pois as seqüências são praticamente impossíveis de adivinhar.

Conflitos de Interesse

Os autores declaram não haver conflito de interesse.

REFERÊNCIAS

- Mancuso ACB, Castro SMJ, Guimarães LSP, Leotti VB, Hirakata VN, Camey SA. Estatística Descritiva: perguntas que você sempre quis fazer, mas nunca teve coragem. *Clin Biomed Res*. 2019;38(4):414-8.
- Szklo M, Nieto FJ. *Epidemiology: beyond the basics*. Burlington: Jones & Bartlett Learning; 2004.
- Lorenzi DRS, Madi JM. Sífilis congênita como indicador de assistência pré-natal. *Rev Bras Ginecol Obstet* [Internet]. 2001 [citado 2019 Jul 26];23(10):647-52. Disponível em: <https://www.nescon.medicina.ufmg.br/biblioteca/imagem/0452.pdf>
- Cadavid D, Mellion M, Hupperts R, Edwards KR, Calabresi PA, Druľović J, et al. Safety and efficacy of opicinumab in patients with relapsing multiple sclerosis (SYNERGY): a randomised, placebo-controlled, phase 2 trial. *Lancet Neurol*. 2019; 18(9):845-56.
- Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman TB, editores. *Delineando a pesquisa clínica: uma abordagem epidemiológica*. Porto Alegre: Artmed; 2003.

6. Collett D. *Modelling survival data in medical research*. Boca Raton: Chapman & Hall/CRC; 2003.
7. Nunes LN, Camey SA, Guimarães LSP, Mancuso ACB, Hirakata VN. Os principais delineamentos na Epidemiologia. *Clin Biomed Res* [Internet]. 2013 [citado 2019 Out 18];33(2):178-83. Disponível em: <https://seer.ufrgs.br/hcpa/article/view/42338>
8. Mancuso ACB, Camey SA, Nunes LN, Hirakata VN, Guimarães LSP. Os principais delineamentos na Epidemiologia: ensaios clínicos. *Clin Biomed Res* [Internet]. 2013 [citado 2019 Out 18];33(3/4):286-94. Disponível em: <https://seer.ufrgs.br/hcpa/article/view/44253/28281>
9. Guimarães LSP, Hirakata VN, Camey SA, Nunes LN, Mancuso ACB. Os principais delineamentos na Epidemiologia: ensaios clínicos (parte II) *Clin Biomed Res* [Internet]. 2013 [citado 2019 Jul 23];33(3/4):295-302. Disponível em: <https://seer.ufrgs.br/hcpa/article/view/44657/28397>
10. Rothman K, Greenland S, Lash T. *Epidemiologia moderna*. 3. ed. Porto Alegre: Artmed; 2011.
11. Guyatt G, Rennie D, Meade MO, Cook DJ. *Users' guides to the medical literature: essentials of evidence-based clinical practice*. 2. ed. New York: McGraw Hill Professional; 2008.
12. Schünemann H, Brożek J, Guyatt G, Oxman A. *Grade handbook: introduction to grade handbook* [Internet]. 2013 [citado 2019 Out 18]. Disponível em: https://gdt.gradepro.org/app/handbook/handbook.html#fnt_ref1
13. Brasil. Ministério da Saúde. Secretaria de Ciência, Tecnologia e Insumos Estratégicos. Departamento de Ciência e Tecnologia. *Diretrizes metodológicas: Sistema GRADE – manual de graduação da qualidade da* evidência e força de recomendação para tomada de decisão em saúde [Internet]. Brasília: Ministério da Saúde; 2014 [citado 2019 Out 18]. Disponível em: http://bvsm.s.saude.gov.br/bvsm/publicacoes/diretrizes_metodologicas_sistema_grade.pdf
14. Camanho GL. Editorial: nível de evidência. *Rev Bras Ortop*. 2009;44(6):1-2.
15. Galvão CM. Editorial: níveis de evidência. *Acta Paul Enferm*. 2006;19(2):5.
16. Agência Nacional de Vigilância Sanitária (Anvisa). *Guia para desfechos para estudos clínicos de medicamentos oncológicos*. Brasília: Anvisa; 2015.
17. Oxford centre for evidence-based medicine – levels of evidence (March 2009) [Internet]. *CEBM*. 2009 [citado 2019 Jul 25]. Disponível em: <https://www.cebm.net/2009/06/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/>
18. Pinto VF. Estudos clínicos de não-inferioridade: fundamentos e controvérsias. *J Vasc Bras*. 2010;9(3):145-51.
19. Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ*. 1995;311(7003):485.
20. Alderson P. Absence of evidence is not evidence of absence. *BMJ*. 2004;328(7438):476-7.
21. Agência Nacional de Vigilância Sanitária. *Regularização de produtos – equivalência farmacêutica: análise de estudos* [Internet]. 2009 [citado 2019 citado Ago 15]. Disponível em: <http://portal.anvisa.gov.br/registros-e-autorizacoes/medicamentos/equivalencia-farmacologica/analise-de-estudos>
22. Gupta SK. Intention-to-treat concept: a review. *Perspect Clin Res*. 2011;2(3):109-12.
23. Shah PB. Intention-to-treat and per-protocol analysis. *CMAJ*. 2011;183(6):696.
24. Obermeier DM. *Intention-To-Treat (ITT) vs. Per Protocol (PP) analysis: what to choose* [Internet]. Clinfo.EU. 2018 [citado 2019 Ago 20]. Disponível em: <https://www.clinfo.eu/itt-vs-pp/>
25. U.S. Department of Health and Human Services. Centers for Disease Control and Prevention. *Principles of Epidemiology in Public Health Practice* [Internet]. 3. ed. Atlanta: CDC; 2011 [citado 2019 Jul 24]. Disponível em: <https://www.cdc.gov/csels/dsepd/ss1978/index.html>
26. Last JM. *A Dictionary of epidemiology*. 4. ed. New York: Oxford University Press; 2001.
27. Hirakata VN, Mancuso ACB, Castro SMJ. Teste de hipóteses: perguntas que você sempre quis fazer, mas nunca teve coragem. *Clin Biomed Res* [Internet]. 2019 [citado 2019 Ago 23];39(2):181-5. Disponível em: <https://seer.ufrgs.br/hcpa/article/view/93649/pdf>
28. Page P. Beyond statistical significance: clinical interpretation of rehabilitation research literature. *Int J Sports Phys Ther*. 2014;9(5):726-36.
29. Bolfarine H, Bussab WO. *Elementos de amostragem*. São Paulo: Blucher; 2005.
30. Sedgwick P. Block randomisation. *BMJ*. 2011;343(d7139):1-2.
31. Efrid J. Blocked randomization with randomly selected block sizes. *Int J Environ Res Public Health*. 2011;8(1):15-20.
32. Bland JM, Butland BK, Peacock JL, Poloniecki J, Reid F, Sedgwick P. *Statistics Guide for Research Grant Applicants* [Internet]. London: St George's Hospital Medical School; 2012 [citado 2019 Ago 27]. Disponível em: <https://www-users.york.ac.uk/~mb55/guide/guide14.pdf>

Recebido: 11 set, 2019

Aceito: 13 set, 2019