

Mauro Federico Andreu^{1,2}, Ladislao Pablo Diaz Ballve^{1,3}, Daniel Héctor Verdecchia¹, Agustina María Monzón¹, Tatiana Dias de Carvalho¹

O valor de p é adequadamente interpretado pelos profissionais de terapia intensiva? Uma pesquisa *on-line*

*Is the p-value properly interpreted by critical care professionals?
Online survey*

RESUMO

Objetivo: Determinar a prevalência e os fatores de risco para conhecimento insuficiente sobre valores de p entre médicos e terapeutas respiratórios atuantes em terapia intensiva na Argentina.

Métodos: Levantamento transversal *on-line* com 25 questões relativas às características dos participantes, autopercepção e conhecimento sobre valores de p (teoria e prática). Realizaram-se análises de estatística descritiva e regressão logística multivariada.

Resultados: Analisaram-se 376 participantes. Não tinham conhecimento a respeito dos valores de p 237 participantes (63,1%). Segundo análise de regressão logística multivariada, falta de treinamento em metodologia científica (RC ajustadas 2,50; IC95% 1,37 - 4,53; $p = 0,003$) e a quantidade

de leitura (< 6 artigos científicos por ano; RC ajustadas 3,27; IC95% 1,67 - 6,40; $p = 0,001$) foram identificados como independentemente associados com a falta de conhecimento sobre valores de p por parte dos participantes.

Conclusão: A prevalência de conhecimento insuficiente com relação a valores de p entre médicos e terapeutas respiratórios na Argentina foi de 63%. Falta de treinamento em metodologia científica e quantidade de leitura (< 6 artigos científicos por ano) foram identificados como independentemente associados com a falta de conhecimento sobre valores de p por parte dos participantes.

Descritores: Bioestatística; Pesquisa biomédica/estatística & dados numéricos; Interpretação estatística de dados; Testes de hipóteses; Medicina baseada em evidências; Prevalência

1. Licenciatura en Kinesiología y Fisiatría, Departamento de Ciencias de la Salud, Universidad Nacional de La Matanza - San Justo, Buenos Aires, Argentina.

2. Unidade de Terapia Intensiva, Hospital General de Agudos Donacion Santojanni - Buenos Aires, Argentina.

3. Hospital Nacional Profesor Alejandro Posadas - Buenos Aires, Argentina.

INTRODUÇÃO

Para praticar medicina baseada em evidência (MBE), os profissionais de saúde devem confiar em informações clínicas atualizadas.⁽¹⁾ Para lidarem com suas questões clínicas, esses profissionais precisam analisar criticamente o delineamento e o procedimento dos estudos, assim como interpretar seus resultados.⁽²⁾ O teste de significância da hipótese de nulidade (H0), com base nos valores de p – indicadores utilizados para rejeitar ou não a hipótese de nulidade –, é a técnica primária para que se obtenham conclusões dos dados em muitas disciplinas na área da saúde.⁽³⁾

Diversos estudos de levantamento demonstraram que um grande número de profissionais de saúde é incapaz de compreender e interpretar adequadamente resultados estatísticos.⁽⁴⁻⁷⁾ Horton et al. relataram que muitos profissionais de saúde têm maior dificuldade em razão de se relacionarem na literatura médica métodos estatísticos cada vez mais complicados, de tal forma que os eles só são capazes de compreender a análise e a interpretação dos resultados em 21% dos artigos de pesquisa.⁽⁸⁾

Informalmente, o valor de p é a probabilidade sob um modelo estatístico especificado cujo resumo estatístico dos dados (por exemplo, a diferença média entre dois grupos

Conflitos de interesse: Nenhum.

Submetido em 2 de março de 2020

Aceito em 9 de junho de 2020

Autor correspondente:

Mauro Federico Andreu
Ciencias de La Salud Universidad Nacional de la Matanza
Florencio Varela, 1.903 – San Justo
B1754 - Buenos Aires, Argentina
E-mail: mfeandreu@gmail.com

Editor responsável: Leandro Utino Taniguchi

DOI: 10.5935/0103-507X.20210009



comparados na amostra) seria igual ou mais extremo do que o valor observado.⁽⁹⁾ Os equívocos conceituais mais comuns a respeito de valores de p são a falácia da probabilidade inversa, falácia de replicação, falácia da significância clínica ou prática e a falácia do tamanho do efeito.⁽¹⁰⁻¹³⁾

A “falácia da probabilidade inversa” é a falsa crença de que o valor de p indica a probabilidade de a H0 ser verdadeira, considerando-se certos resultados [P (H0/resultados)]. Essencialmente, significa confundir a probabilidade do resultado, assumindo que a hipótese de nulidade é verdadeira [resultados P/H0], com a probabilidade da hipótese de nulidade, considerando certos dados [P H0/resultados].⁽¹⁴⁾

O segundo equívoco é denominado “falácia de replicação”, que é a crença que o valor de p é o grau de replicabilidade do resultado, e seu complemento, 1-p, é frequentemente mal interpretado como a probabilidade de que um resultado será replicado.^(10,13) Isso é, a crença de que os resultados com valor de p de 0,05 significam que, em 95 vezes de cem, os resultados estatisticamente significantes no estudo serão os mesmos em um futuro estudo.⁽¹⁵⁾ Contudo, os valores de p só fornecem pouca informação sobre qual a probabilidade do que ocorra em uma replicação e podem diferir em uma replicação simplesmente por causa da variabilidade da amostra.⁽³⁾

A falsa crença de que o valor de p fornece informação direta sobre o tamanho do efeito é denominada falácia do “tamanho do efeito”.⁽¹⁶⁾ Os pesquisadores creem que quanto menor o valor de p, mais importante o tamanho do efeito.^(12,17)

O último conceito equivocado é denominado falácia da “significância clínica ou prática”, que relaciona a significância estatística com a importância do tamanho do efeito.⁽¹⁰⁾ Um resultado estatisticamente significativo pode, contudo, não ter significância clínica e vice-versa. Assim, a significância dos achados para a prática clínica deve ser descrita por um especialista no assunto e não apresentada unicamente pela estatística.⁽¹¹⁾

Apesar do importante papel desempenhado pela interpretação estatística e a avaliação crítica dos estudos publicados na prática da MBE, não há evidência suficiente relativa ao conhecimento dos profissionais atuantes em terapia intensiva com relação a esse assunto.

O objetivo deste estudo foi determinar a prevalência e os fatores de risco para conhecimento insuficiente a respeito de valores de p entre médicos e terapeutas respiratórios atuantes em terapia intensiva na Argentina.

MÉTODOS

Este foi um levantamento observacional transversal conduzido entre 30 de agosto e 30 de novembro de 2018. Não foi necessário um Termo de Consentimento Livre e

Esclarecido em razão de que a participação foi voluntária e anônima. O protocolo do estudo foi aprovado pelo Comitê de Ética do Hospital Nacional Profesor Alejandro Posadas (312 EmnPeS0/19).

Incluímos em nossa análise profissionais de saúde dos campos de cuidados cardiorrespiratórios. Excluíram-se da análise os profissionais não atuantes na Argentina e os que abandonaram a pesquisa antes da seção B (questão-filtro).

Teste piloto

Antes do estudo, conduziu-se um teste piloto, para avaliar a viabilidade da pesquisa. Para isso, o questionário foi aplicado a 42 profissionais de saúde com registro do tempo demandado para responder às questões. Pedimos também a cada profissional que relatasse se o questionário ou alguma questão específica apresentou dificuldades. Dentre os participantes, 40 (95,2%) responderam que o questionário era claro e compreenderam seu objetivo. Compreenderam todas as questões 37 (88,1%) dos participantes. Três participantes tiveram dificuldades com a questão 19, e dois tiveram dificuldades com a questão 5. Com relação ao grau de dificuldade, sete participantes (16,7%) consideraram que o questionário era muito fácil, cinco (11,9%) o consideraram fácil, 13 (31%) consideraram o questionário moderadamente fácil, 15 (35,7%) o consideraram difícil e dois (2,4%) consideraram o questionário como muito difícil. O tempo mediano para responder ao questionário foi de 6,5 (5 - 8) minutos.

Coleta dos dados

Por conveniência e sem amostra probabilística, os profissionais foram convidados a participar por meio de uma mensagem de *e-mail* e pelas redes sociais. O convite incluía o objetivo do estudo e um *link* para acesso à pesquisa *on-line*, por meio da ferramenta SurveyMonkey® (<https://es.surveymonkey.com/r/valorp>).

Instrumento

O levantamento foi constituído de 25 questões, divididas em três seções (Apêndice 1).

A primeira seção (A) consistiu em 13 questões nominais e ordinais a respeito das características profissionais do participante, como área de atuação, educação acadêmica e experiência relacionada à leitura científica.

A segunda seção (B) consistiu em uma questão nominal dicotômica (sim/não) relativa à autopercepção do participante com relação a seu conhecimento relacionado a valores de p. Se a resposta fosse negativa, o questionário era considerado concluído.

Finalmente, a terceira seção (C) consistiu em 11 questões nominais (verdadeiro/falso/não sei) relativas a valores de

p: seis questões teóricas, quatro questões de interpretação prática e uma questão de definição (Apêndice 1).

As questões na seção C foram aplicadas em ordem aleatória.

Medida primária de desfecho

Falta de conhecimentos relacionados a valores de p foi o desfecho principal do estudo. Os participantes que declararam não ter conhecimento sobre valores de p (uma resposta “não” na questão da seção B) ou aqueles que não atingiram o escore exigido em qualquer das duas categorias (teoria ou prática) foram considerados como “sem conhecimentos sobre valores de p”. Os que abandonaram a pesquisa na seção C sem alcançar o limite exigido em pelo menos uma das duas categorias (teoria ou prática) também foram considerados “sem conhecimentos sobre valores de p”.

Análise estatística

As variáveis categóricas são apresentadas como números e percentagens. As variáveis contínuas com distribuição normal são apresentadas como média e desvio padrão. As variáveis com distribuição não normal são apresentadas como medianas e faixas interquartis. A continuidade da distribuição das variáveis foi avaliada por meio do teste de Kolmogorov-Smirnov.

Foi conduzido um teste quanto a uma diferença nas proporções para comparar as variáveis nominais entre as categorias.

O desfecho principal foi falta de conhecimentos sobre valores de p (teórico e/ou prático). As questões relativas a conhecimentos sobre valores de p foram agrupadas em questões teóricas (15, 16, 19, 20, 21 e 22) e práticas (17, 18, 23 e 24). Considerou-se que o participante tinha conhecimentos teóricos suficientes se respondesse corretamente a pelo menos quatro das seis questões (67%). Considerou-se que o participante tinha conhecimentos práticos suficientes se respondesse corretamente a pelo menos três das quatro questões práticas (75%). Considerou-se que o participante tinha conhecimentos sobre valores de p se o escore exigido fosse atingido para qualquer das categorias.

As associações entre conhecimentos sobre valores de p e outras variáveis foram determinadas por análise univariada. Relataram-se as razões de chance (RC) e seus respectivos intervalos de confiança de 95% (IC95%). Variáveis com valor de $p < 0,15$ foram incluídas no modelo de regressão logística multivariável para identificar as que eram independentemente associadas com conhecimentos sobre valores de p. Utilizou-se método condicional *backward stepwise* (Wald). Considerou-se significativo valor de $p < 0,05$. A análise estatística foi conduzida com utilização do programa IBM *Statistical*

Package for Social Sciences (SPSS) v. 22.0 para Macintosh (IBM Corp., Armonk, NY, Estados Unidos).

RESULTADOS

Obtivemos um total de 896 questionários respondidos; 520 foram excluídos em razão do não cumprimento dos critérios de elegibilidade.

Analisaram-se 376 questionários: 210 (55,9%) dos participantes eram médicos e 166 (44,1%) terapeutas respiratórios. As características da amostra são detalhadas na tabela 1. Apenas 139 (37,0%) participantes responderam satisfatoriamente às questões relativas a valores de p (pelo menos quatro questões teóricas e/ou três questões práticas).

Tabela 1 - Características da amostra

Variáveis	
Idade (anos)	36 (29 - 48)
Sexo masculino	195 (51.9)
Anos desde a graduação	12 (7 - 23)
Profissão/especialização	
Médico/cardiologista	20 (5.3)
Médico/intensivista	100 (26.6)
Médico/pneumologista	90 (23.9)
Terapeuta respiratório	166 (44.1)
Educação em universidade privada	45 (12)
Nível de treinamento completo	
Cursos	166 (44.1)
Residência, bolsa de estudos	246 (65.4)
Especialização ou curso avançado	246 (65.4)
Programa de Mestrado	24 (6.4)
Programa de Doutorado	14 (3.7)
Treinamento em metodologia de pesquisa científica	83 (22.1)
Leitura de seis ou mais artigos científicos por ano	73 (19.4)
Considera o idioma da publicação uma barreira para leitura de artigos científicos	226 (60.1)
Tem autoria de um artigo científico	202 (53.7)

Resultados expressos por mediana (intervalo interquartil) ou n (%).

Duzentos e trinta e sete participantes não compreendiam valores de p (63,1%; IC95% 58,0% - 67,7%). Destes participantes, 47 (12,5%) relataram que não compreendiam valores de p, e 190 (50,5%) relataram que compreendiam valores de p, porém não alcançaram os escores de corte para qualquer das categorias de conhecimento (teórico e prático). Os resultados das seções B e C (questões 14 a 24) estão resumidos na tabela 2.

A autoavaliação dos participantes com relação à “avaliação crítica de um artigo científico” e sua associação com o resultado geral da pesquisa (compreensão ou não compreensão de valores de p) são detalhadas na figura 1 ($p < 0,001$). Só se observaram diferenças

Tabela 2 – Resultados da pesquisa

Você sabe o que é o valor de p?	n = 47/376 (12,5%) participantes responderam “Não” e foram considerados sem conhecimentos sobre os valores de p (fim do questionário)									
	n = 329/376 (87,5%) participantes responderam “Sim” e prosseguiram para as questões sobre valores de p									
Questões sobre valores de p	Questões teóricas					Questões práticas				
	O valor de p é uma probabilidade	Um valor de p não significativo ($p > 0,05$) indica que a hipótese de nulidade é verdadeira	O valor de p indica a probabilidade de que a hipótese de nulidade seja verdadeira em vista dos resultados do estudo	Um valor de p não significativo ($p > 0,05$) indica que devemos aceitar a hipótese de nulidade	Se obtivermos valor de p significativo ($p < 0,05$), devemos rejeitar a hipótese de nulidade	O valor de p obtido ($p = 0,02$) indica a probabilidade de obter resultados similares se o mesmo estudo for repetido com uma amostra similar	Um valor não significativo de p ($p > 0,05$) indica que o efeito do tratamento em análise não é clinicamente importante	Um valor de p não significativo ($p > 0,05$) indica que ambos os tratamentos são similares	Um resultado estatisticamente significativo ($p < 0,05$) indica que o efeito do tratamento sob análise é clinicamente importante	O valor de p observado em nosso estudo foi significativo ($p = 0,02$). Isso confirma que o efeito do tratamento foi maior que o observado em um estudo similar com valor de $p = 0,04$
Correto, n (%)	251 (66,8)	132 (35,1)	68 (18,1)	115 (30,6)	169 (44,9)	80 (21,3)	127 (33,8)	132 (35,1)	102 (27,1)	147 (39,1)
Incorreto, n (%)	59 (15,7)	145(38,6)	171 (45,5)	126 (33,5)	72 (19,1)	154 (41,0)	162 (43,1)	143 (38,0)	154 (41,0)	89 (23,7)
Não sei, n (%)	10 (2,7)	27(7,2)	46 (12,2)	39 (10,4)	32 (8,5)	36 (9,6)	9 (2,4)	17 (4,5)	12 (3,2)	30 (8,0)
Resultado geral	n = 237/376; 63,0% (IC95% 58,0% - 67,7%) não conhecem nem teoria nem prática									
	n = 139/376; 37,0% (IC95% 32,2% - 41,9) conhecem teoria e/ou prática									
	n = 28/376; 7,4% (IC95% 5,2% - 10,5%) conhecem teoria e prática									
	n = 69/376; 18,3% (IC95% 14,7% - 22,5%) conhecem teoria									
	n = 84/376; 22,3% (IC95% 18,4% - 26,8%) conhecem prática									

IC95% - intervalo de confiança de 95%.

entre os participantes que compreendiam valores de p (escores mais elevados) e os demais participantes, assim como entre os que não compreendiam os valores de p (mais baixos escores na escala) e os demais participantes ($p = 0,019$ e $p = 0,005$; respectivamente).

Na questão 25, os participantes deveriam escolher a definição correta do valor de p (item c, “ambas as opções estão corretas”). Apenas 104 de 376 participantes (27,6%) responderam corretamente a esta questão.

Os modelos de regressão logística binária univariada e multivariada se encontram detalhados na tabela 3. Segundo a análise de regressão multivariável, identificou-se falta de treinamento em metodologia científica para pesquisa (RC ajustada - 2,50; IC95% 1,37 - 4,53; $p = 0,003$) e a quantidade de leitura (< seis artigos científicos por ano; RC ajustada 3,27; IC95% 1,67 - 6,40; $p = 0,001$) como independentemente associadas com a falta de conhecimento sobre valores de p.

DISCUSSÃO

Nosso principal achado foi uma elevada prevalência de conhecimento insuficiente sobre valores de p entre médicos e terapeutas respiratórios atuantes em terapia intensiva. Esses achados concordam com os de estudos prévios. Eles revelam que um alto percentual de profissionais de saúde experimentam dificuldades para compreender e interpretar os valores de p.⁽¹⁸⁻²¹⁾

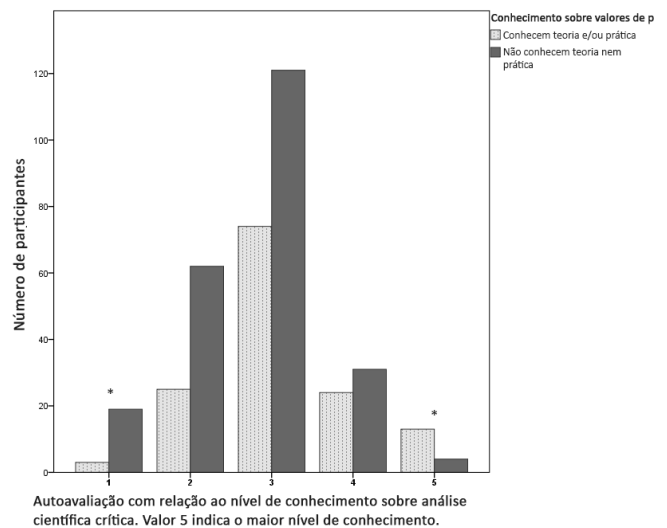


Figura 1 – Escores gerais de conhecimentos sobre valores de p, segundo a autoavaliação dos participantes com relação à análise crítica de artigos científicos. * Significância estatística para a comparação entre os valores de p para “compreender” e “não compreender”.

Segundo levantamento realizado por Badenes-Ribera et al. entre professores espanhóis de psicologia, muitos professores universitários não sabiam como interpretar corretamente os valores de p.⁽²²⁾ Os autores conduziram levantamento semelhante entre estudantes universitários italianos e chilenos e observaram que uma percentagem dos participantes não era capaz de interpretar os valores de p.⁽¹⁴⁾

Tabela 3 - Análise de regressão logística univariada e multivariável

Variáveis	RC (IC95%)	Valor de p	RC ajustada (IC95%)	Valor de p
Idade	1,01 (0,99 - 1,02)	0,26		
Sexo masculino	0,76 (0,50 - 1,16)	0,21		
Anos desde a graduação	1,00 (0,98 - 1,01)	0,99		
Universidade privada	0,66 (0,33 - 1,31)	0,23		
Maior nível concluído de educação após a graduação				
Curso de treinamento	0,88 (0,58 - 1,35)	0,57		
Residência, bolsa de estudo	0,73 (0,47 - 1,15)	0,17		
Especialização ou curso avançado	1,22 (0,79 - 1,89)	0,38		
Pós-graduação Mestrado	1,22 (0,79 - 1,89)	0,18		
Pós-graduação Doutorado	0,42 (0,14 - 1,25)	0,12	0,64 (0,20 - 2,00)	0,44
Falta de treinamento em metodologia de pesquisa científica	2,77 (1,55 - 4,95)	0,001	2,50 (1,37 - 4,53)	0,003
Leitura < seis artigos por ano	3,67 (1,896 - 7,09)	< 0,001	3,27 (1,67 - 6,40)	0,001
Considera o idioma da publicação uma barreira para leitura de artigos científicos	1,43 (0,93 - 2,19)	0,1	1,13 (0,72 - 1,79)	0,58
Baixa autoavaliação com relação à leitura científica	2,09 (1,24 - 3,5)	0,005	1,63 (0,95 - 2,80)	0,073
Foi autor de um artigo científico	0,51 (0,33 - 0,78)	0,002	0,72 (0,45 - 1,15)	0,17

RC - razão de chance; IC95% - intervalo de confiança de 95%.

Msaouel et al. conduziram um levantamento em múltiplas instituições entre residentes médicos gregos a respeito de conceitos básicos em estatística.⁽²⁰⁾ Os resultados demonstraram que muitos dos médicos residentes eram incapazes de interpretar corretamente conceitos que são comumente encontrados na literatura médica. Susarla e Redett também avaliaram o conhecimento, as atitudes e a confiança em relação à bioestatística em população semelhante.⁽²³⁾ Os autores concluíram que os residentes dão um grande grau de importância ao conhecimento em bioestatística, porém têm compreensão apenas pequena a respeito de importantes conceitos estatísticos.

Segundo nosso estudo, encontraram-se dois fatores independentemente associados com a limitação do conhecimento sobre valores de p entre os participantes: a falta de treinamento em metodologia de pesquisa científica e a quantidade de leitura (menor que seis artigos científicos por ano). Esses resultados concordam com a literatura.⁽²⁴⁾

Em nosso estudo, também observamos que ter treinamento em metodologia de pesquisa não impede que os profissionais interpretem incorretamente os valores de p. Assumir que o treinamento previne interpretações incorretas é uma falsa crença que pode se difundir entre os colegas menos experientes ou em treinamento.⁽²⁵⁾

Estudo que avaliou as atitudes e a confiança de médicos residentes com epidemiologia e bioestatística concluiu que ter treinamento em bioestatística e fazer leitura de um número mais elevado de artigos científicos a respeito de estatística e epidemiologia mensalmente se associaram com atitude positiva em relação à bioestatística e aumentaram a confiança em relação a conceitos estatísticos.⁽²³⁾ Semelhantemente, nossos resultados indicam que profissionais que leem mais de seis artigos científicos por ano tiveram níveis maiores de conhecimento sobre valores de p.

A falta de conhecimento sobre valores de p foi mais prevalente com relação ao conhecimento teórico do que ao prático. Isso pode ser devido a que quando os profissionais de saúde leem artigos científicos geralmente não aplicam interpretação não probabilística *sine qua non* dos valores de p. Esses resultados apenas demandam que o leitor aplique rotineiramente a regra de $p < \alpha$. Assim, a interpretação estatística só se baseia no valor de p em comparação com o valor de α .⁽²⁶⁾ Esse pressuposto parece basear-se nos resultados obtidos pela questão a respeito dos valores de p. Embora não tenha havido diferença estatisticamente significativa entre os profissionais que compreendiam os valores de p e os que não o faziam, um maior número de participantes não foi capaz de fornecer definição correta.^(9,27)

A autoavaliação dos participantes com relação à análise crítica deve ser salientada. Os participantes que relataram ter grande habilidade para avaliação crítica (cinco pontos) responderam corretamente ao questionário. Igualmente, os que relataram ter baixa capacidade para análise crítica (um ponto) também mostraram baixos níveis de conhecimento sobre os valores de p. Contudo, é digno de nota que grande percentagem de participantes que relataram alta capacidade para avaliação crítica (três ou quatro pontos) não conseguiu atingir os escores de cortes para conhecimento a respeito dos valores de p. Esse achado pode ser devido a uma contradição existente entre baixo treinamento em estatística e a importância demasiada colocada sobre os valores de p nas publicações médicas.

Os mais comuns conceitos equivocados sobre o valor de p são as “falácias” que podem ameaçar gravemente a correta interpretação dos resultados.⁽¹⁰⁻¹³⁾ Em linha com nossos resultados, Msaouel et al. também observaram que médicos residentes são especialmente inclinados ao viés da falácia do apostador, provocada pela equivocada crença segundo

a qual um evento tem maior probabilidade de ocorrer se já ocorreu previamente e vice-versa. Esse viés pode minar o julgamento clínico e a tomada de decisão médica.⁽²⁰⁾

Os valores de p podem ser equivocadamente interpretados por múltiplos fatores, como os vieses de resultados e de publicação observados na literatura. O viés de resultado é o fenômeno no qual os autores só relatam resultados satisfatórios. Por outro lado, viés de publicação é o fenômeno no qual os periódicos só aceitam artigos com resultados estatisticamente significantes e rejeitam artigos com resultados não significantes.⁽²⁸⁻³¹⁾

Mais de 12% dos participantes relataram não compreender valores de p. Isso provavelmente indica que alguns profissionais não leem artigos científicos. É, portanto, necessário melhorar o treinamento nessa área para assegurar conhecimento de alta qualidade.⁽²⁵⁾ É necessário um treinamento sistemático adequado em bioestatística para eliminar o viés dos profissionais e assegurar que sejam habilidosos na compreensão e comunicação de informações estatísticas.⁽²⁰⁾

Este estudo tem algumas limitações. Primeiramente, os participantes que saíram na seção C foram considerados como “sem de conhecimentos sobre valores de p”. Assim, podemos ter superestimado a prevalência de conhecimento insuficiente, já que esses participantes poderiam ter concluído o questionário e atingido os valores de corte para conhecimentos sobre os valores de p. Semelhantemente, consideramos os participantes que responderam de forma negativa à questão 14 como “sem de conhecimentos sobre valores de p”, sem lhes permitir continuar com as questões na seção C. A razão para excluir tais participantes foi justificada pela consideração de que poderiam resultar em um maior número de desistências e respostas incompletas pela extensão do questionário e a possibilidade de que os

participantes fornecessem respostas aleatórias apenas para concluí-lo. Esse pode ser outro fator a ameaçar a validade da estimativa de “sem de conhecimentos sobre valores de p”.

Em segundo lugar, nós agrupamos arbitrariamente as questões em conhecimentos teóricos e práticos e, também arbitrariamente, determinamos escores de corte para definir falta de conhecimento. Entretanto, mesmo que as questões tenham sido preparadas pelos autores, o conteúdo avaliado em cada uma delas se baseou em estudos prévios.^(14,22) Mais ainda, para evitar respostas aleatórias, acrescentamos a opção “não sei”. Outra limitação deste estudo é o fato da não utilização de um instrumento validado; porém, para minimizar essa limitação, realizamos um teste piloto, no qual 90% dos participantes responderam virtualmente ter compreendido todas as questões.

Este é o primeiro estudo a relatar os conhecimentos sobre valores de p entre médicos e terapeutas respiratórios atuantes em terapia intensiva na Argentina. Segundo os resultados, consideramos que o treinamento em avaliação crítica deve ser incluído nos currículos dos programas de primeiro grau, com especialização em leitura e interpretação de artigos científicos. Mais ainda, os professores na área da saúde devem encorajar seus estudantes a participar de atividades científicas.

CONCLUSÃO

A prevalência geral de conhecimentos insuficientes sobre os valores de p entre médicos e terapeutas respiratórios atuantes em terapia intensiva na Argentina foi de 63%. Encontraram-se dois fatores independentemente associados com a falta de conhecimento sobre valores de p entre os participantes: falta de treinamento em metodologia de pesquisa científica e a quantidade de leitura (menor que seis artigos científicos por ano).

ABSTRACT

Objective: To determine the prevalence of and risk factors for insufficient knowledge related to p-values among critical care physicians and respiratory therapists in Argentina.

Methods: This cross-sectional online survey contained 25 questions about respondents' characteristics, self-perception and p-value knowledge (theory and practice). Descriptive and multivariable logistic regression analyses were conducted.

Results: Three hundred seventy-six respondents were analyzed. Two hundred thirty-seven respondents (63.1%) did not know about p-values. According to the multivariable logistic regression analysis, a lack of training on scientific research methodology (adjusted OR 2.50; 95%CI 1.37 - 4.53; $p = 0.003$) and the amount of reading (< 6 scientific

articles per year; adjusted OR 3.27; 95%CI 1.67 - 6.40; $p = 0.001$) were found to be independently associated with the respondents' lack of p-value knowledge.

Conclusion: The prevalence of insufficient knowledge regarding p-values among critical care physicians and respiratory therapists in Argentina was 63%. A lack of training on scientific research methodology and the amount of reading (< 6 scientific articles per year) were found to be independently associated with the respondents' lack of p-value knowledge.

Keywords: Biostatistics; Biomedical research/statistics & numerical data; Data interpretation, statistical; Hypothesis testing; Evidence-based medicine; Prevalence

REFERÊNCIAS

1. Windish DM, Huot SJ, Green ML. Medicine residents' understanding of the biostatistics and results in the medical literature. *JAMA*. 2007;298(9):1010-22.
2. Finch S, Cumming G, Williams J, Palmer L, Griffith E, Alders C, et al. Reform of statistical inference in psychology: the case of memory & cognition. *Behav Res Methods Instrum Comput*. 2004;36(2):312-24.
3. Cumming G. Replication and p intervals: p values predict the future only vaguely, but confidence intervals do much better. *Perspect Psychol Sci*. 2008;3(4):286-300.
4. Wulff HR, Andersen B, Brandenhoff P, Guttler F. What do doctors know about statistics? *Stat Med*. 1987;6(1):3-10.
5. Berwick DM, Fineberg HV, Weinstein MC. When doctors meet numbers. *Am J Med*. 1981;71(6):991-8.
6. Weiss ST, Samet JM. An assessment of physician knowledge of epidemiology and biostatistics. *J Med Educ*. 1980;55(8):692-7.
7. Andreu MF, Diaz-Ballve LP, Verdecchia DH, Monzón AM, Dias Carvalho T. ¿Qué saben los kinesiólogos argentinos sobre el valor-p? *AJRPT*. 2020;2(1):22-32.
8. Horton NJ, Switzer SS. Statistical methods in the journal. *N Engl J Med*. 2005;353(18):1977-9.
9. Wasserstein RL, Lazar NA. The ASA statement on p-values: context, process, and purpose. *Am Stat*. 2016;70(2):129-33.
10. Nickerson RS. Null hypothesis significance testing: a review of an old and continuing controversy. *Psychol Methods*. 2000;5(2):241-301.
11. Kirk RE. Practical significance: a concept whose time has come. *Educ Psychol Meas*. 1996;56(5):746-59.
12. Cohen J. The earth is round ($p < .05$). *Am Psychol*. 1994;49(12):997-1003.
13. Carver RP. The case against statistical significance testing. *Harv Educ Rev*. 1978;48(3):378-99.
14. Badenes-Ribera L, Frias-Navarro D, Iotti B, Bonilla-Campos A, Longobardi C. Misconceptions of the p-value among Chilean and Italian Academic Psychologists. *Front Psychol*. 2016;7:1247.
15. Fidler F. From statistical significance to effect estimation: statistical reform in psychology, medicine and ecology [thesis]. Melbourne, Australia: Department of History and Philosophy of Science. The University of Melbourne; 2005. [cited 2021 Apr 3]. Available from: https://www.researchgate.net/publication/267403673_From_statistical_significance_to_effect_estimation_Statistical_reform_in_psychology_medicine_and_ecology
16. Gliner JA, Vaske JJ, Morgan GA. Null hypothesis significance testing: Effect size matters. *Hum Dimens Wildl*. 2001;6(4):291-301.
17. Cumming G. Understanding the new statistics: effect sizes, confidence intervals, and meta-analysis. New York, NY: Routledge; 2013.
18. Banks D, Botchway P, Akintorin S, Arcia R, Soyemi K. Pediatric residents' knowledge of epidemiology and statistics. *Int J Med Educ*. 2018;9:323-4.
19. Wellek S. A critical evaluation of the current "p-value controversy". *Biom J*. 2017;59(5):854-72.
20. Msaouel P, Kappos T, Tasoulis A, Apostolopoulos AP, Lekkas I, Tripodaki ES, et al. Assessment of cognitive biases and biostatistics knowledge of medical residents: a multicenter, cross-sectional questionnaire study. *Med Educ Online*. 2014;19:23646.
21. Susarla SM, Redett RJ. Plastic surgery residents' attitudes and understanding of biostatistics: a pilot study. *J Surg Educ*. 2014;71(4):574-9.
22. Badenes-Ribera L, Frias-Navarro D, Monderde-i-Bort H, Pascual-Soler M. Interpretation of the p value: A national survey study in academic psychologists from Spain. *Psicothema*. 2015;27(3):290-5.
23. Susarla SM, Lifchez SD, Losee J, Hultman CS, Redett RJ. Plastic surgery residents' understanding and attitudes toward biostatistics: a national survey. *Ann Plast Surg*. 2016;77(2):231-6.
24. Lecoutre MP, Poitevineau J, Lecoutre B. Even statisticians are not immune to misinterpretations of null hypothesis tests. *Int J Psychol*. 2003;38(1):37-45.
25. Kirk RE. Promoting good statistical practices: some suggestions. *Educ Psychol Meas*. 2001;61(2):213-8.
26. Ben-Zvi D, Garfield J. Statistical literacy, reasoning, and thinking: goals, definitions, and challenges. In: Ben-Zvi D, Garfield J, Editors. *The challenge of developing statistical literacy, reasoning and thinking*. Dordrecht, Netherlands: Springer; 2014. p. 3-15.
27. Greenhalgh T, Howick J, Maskrey N; Evidence Based Medicine Renaissance Group. Evidence based medicine: a movement in crisis? *BMJ*. 2014;348:g3725.
28. Saini P, Loke YK, Gamble C, Altman DG, Williamson PR, Kirkham JJ. Selective reporting bias of harm outcomes within studies: findings from a cohort of systematic reviews. *BMJ*. 2014;349:g6501.
29. Reid EK, Tejani AM, Huan LN, Egan G, O'Sullivan C, Mayhew AD, et al. Managing the incidence of selective reporting bias: a survey of Cochrane review groups. *Syst Rev*. 2015;4:85.
30. Breivik H, Rosseland LA, Stubhaug A. Statistical pearls: importance of effect-size, blinding, randomization, publication bias, and the overestimated p-values. *Scand J Pain*. 2013;4(4):217-9.
31. Ost DE, Seeley EJ, Shojaee S, Yasufuku K. Efforts to limit publication bias and improve quality in the journal: introduction of double-blind peer review. *J Bronchology Interv Pulmonol*. 2019;26(3):143-7.

Apêndice 1 - Questionário.

<p>A. Características dos participantes</p> <p>1. Data de nascimento:</p> <p>2. Sexo:</p> <p>a) Masculino. b) Feminino.</p> <p>3. País de residência:</p> <p>a) Argentina. b) Outro (especificar).</p> <p>4. Ano de graduação:</p> <p>5. Disciplina:</p> <p>a) Medicina. b) Terapia respiratória. c) Enfermagem. d) Outra (especificar).</p> <p>6. Principal área de atuação profissional:</p> <p>7. Sua graduação foi concluída em:</p> <p>a) Universidade pública. b) Universidade privada.</p> <p>8. Indique o nível de educação que você concluiu:</p> <p>a) Cursos de treinamento. b) Residência/bolsa de estudos. c) Estágio. d) Especialização/curso avançado. e) Pós-graduação – Mestrado. f) Pós-graduação – Doutorado. g) Pós-doutorado.</p> <p>9. Você recebeu treinamento em metodologia de pesquisa científica ou análise crítica científica, como curso como complemento curricular?</p> <p>a) Sim. b) Não.</p> <p>10. Quantos artigos científicos você leu aproximadamente no ano passado?</p> <p>a) Eu não li artigos científicos no ano passado. b) 1 a 5. c) 6 a 12. d) Mais de 12.</p> <p>11. Em sua opinião, você considera o idioma uma barreira para ler artigos científicos?</p> <p>a) Sim. b) Não.</p> <p>12. Você participou como autor em um artigo publicado em um periódico científico?</p> <p>a) Sim. b) Não.</p> <p>13. Em uma escala de 5 pontos, qual nível de conhecimento você considera ter a respeito de análise científica crítica (sendo que 5 representa o nível mais alto de conhecimento)?</p>	<p>B. Autopercepção a respeito do seu conhecimento sobre valores de p</p> <p>14. Você sabe o que é o valor de p?</p> <p>a) Sim b) Não (fim do questionário).</p> <p>C. Questões sobre valores de p</p> <p>15. O valor de p é uma probabilidade.</p> <p>a) Verdadeiro. b) Falso. c) Não sei.</p> <p>Imagine que estamos conduzindo um estudo no qual se comparam dois grupos de tratamento e definimos um valor de $p < 0,05$ (erro tipo I ou $\alpha < 5\%$) como “significância estatística”. Com base nesta premissa, assinale se as frases a seguir a respeito da interpretação do valor de p são verdadeiras ou falsas.</p> <p>16. Um valor não significativo de p ($p > 0,05$) indica que a hipótese de nulidade é verdadeira.</p> <p>a) Verdadeiro. b) Falso. c) Não sei.</p> <p>17. Um valor não significativo de p ($p > 0,05$) indica que o efeito do tratamento em análise não é clinicamente importante.</p> <p>a) Verdadeiro. b) Falso. c) Não sei.</p> <p>18. Um valor não significativo de p ($p > 0,05$) indica que ambos os tratamentos são similares.</p> <p>a) Verdadeiro. b) Falso. c) Não sei.</p> <p>19. O valor de p indica a probabilidade de que a hipótese de nulidade seja verdadeira considerando-se os resultados de nosso estudo.</p> <p>a) Verdadeiro. b) Falso. c) Não sei.</p> <p>20. Um valor não significativo de p ($p > 0,05$) indica que devemos aceitar a hipótese de nulidade.</p> <p>a) Verdadeiro. b) Falso. c) Não sei.</p> <p>21. Se obtivermos um valor de p significativo ($p < 0,05$), devemos rejeitar a hipótese de nulidade.</p> <p>a) Verdadeiro. b) Falso. c) Não sei.</p> <p>22. O valor de p obtido ($p = 0,02$) indica a probabilidade de obter resultados similares se o mesmo estudo for repetido com amostra semelhante.</p> <p>a) Verdadeiro. b) Falso. c) Não sei.</p> <p>23. Um resultado estatisticamente significativo ($p < 0,05$) indica que o efeito do tratamento em análise é clinicamente importante.</p> <p>a) Verdadeiro. b) Falso. c) Não sei.</p> <p>24. O valor de p observado em nosso estudo foi significativo ($p = 0,02$). Isso confirma que o efeito do tratamento foi maior do que o observado em um estudo semelhante com valor de $p = 0,04$.</p> <p>a) Verdadeiro. b) Falso. c) Não sei.</p> <p>25. Qual das frases a seguir é a definição do valor de p?</p> <p>a) O valor de p é a probabilidade de se obter um resultado igual ou mais extremo do que o resultado observado. b) O valor de p indica em que grau os dados não são compatíveis com a hipótese de nulidade. c) Ambas as opções estão corretas. d) Não sei.</p>
--	---