

Tamanho do efeito em estudos observacionais na área de Saúde Bucal Coletiva: importância, cálculo e interpretação

Size effect in observational studies in Public Oral Health: importance, calculation and interpretation

Flávia Martão Flório (<https://orcid.org/0000-0001-7742-0255>)¹

Luciane Zanin (<https://orcid.org/0000-0003-0218-9313>)¹

Leônidas Marinho dos Santos Júnior (<https://orcid.org/0000-0002-4685-906X>)¹

Marcelo de Castro Meneghim (<https://orcid.org/0000-0003-2673-3627>)²

Gláucia Maria Bovi Ambrosano (<https://orcid.org/0000-0001-5950-6370>)²

Abstract *The objective of this study was to analyze the scientific literature in public oral health regarding calculation, presentation, and discussion of the effect size in observational studies. The scientific literature (2015 to 2019) was analyzed regarding: a) general information (journal and guidelines to authors, number of variables and outcomes), b) objective and consistency with sample calculation presentation; c) effect size (presentation, measure used and consistency with data discussion and conclusion). A total of 123 articles from 66 journals were analyzed. Most articles analyzed presented a single outcome (74%) and did not mention sample size calculation (69.9%). Among those who did, 70.3% showed consistency between sample calculation used and the objective. Only 3.3% of articles mentioned the term effect size and 24.4% did not consider that in the discussion of results, despite showing effect size calculation. Logistic regression was the most commonly used statistical methodology (98.4%) and Odds Ratio was the most commonly used effect size measure (94.3%), although it was not cited and discussed as an effect size measure in most studies (96.7%). It could be concluded that most researchers restrict the discussion of their results only to the statistical significance found in associations under study.*

Key words *Statistical data interpretation, Observational study, Bias*

Resumo *O objetivo deste estudo foi analisar a literatura científica da área de saúde bucal coletiva quanto ao cálculo, apresentação e discussão do tamanho do efeito em estudos observacionais. A literatura científica na área (2015 a 2019) foi analisada quanto: a) informações gerais (periódico e diretrizes aos autores, número de variáveis e desfechos), b) objetivo e coerência com o cálculo amostral apresentado; c) tamanho do efeito (apresentação, medida utilizada e coerência com a discussão dos dados e conclusão). Foram analisados 123 artigos, de 66 periódicos. A maioria dos artigos avaliados apresenta um único desfecho (74%) e não menciona a realização de cálculo amostral (69,9%). Dentre os que realizaram, para 70,3% havia coerência entre o cálculo amostral utilizado e o objetivo. Apenas 3,3% dos artigos mencionam o termo tamanho do efeito e 24,4% não o consideram na discussão dos resultados, apesar de terem calculado. A regressão logística foi a metodologia estatística mais utilizada (98,4%) e o Odds Ratio a medida de tamanho do efeito mais utilizada (94,3%), embora não tenha sido citada e discutida como uma medida de tamanho do efeito na maioria dos estudos (96,7%). Os pesquisadores, em sua maioria, restringiram a discussão dos resultados apenas à significância estatística encontrada nas associações testadas.*

Palavras-chave *Interpretação estatística de dados, Estudo observacional, Viés*

¹ Departamento de Saúde Coletiva, Faculdade São Leopoldo Mandic.

R. José Rocha Junqueira
13, Swift. 13045-755
Campinas SP Brasil.
flaviaflorio@yahoo.com

² Faculdade de Odontologia de Piracicaba, Universidade Estadual de Campinas.
Piracicaba SP Brasil.

Introdução

O tamanho do efeito é uma medida descritiva que permite a discussão dos resultados em termos de magnitude do efeito da intervenção ou do fator de estudo¹, sendo recomendado que este valor seja reportado e interpretado pelos pesquisadores em seus artigos científicos².

Analisados em conjunto, o tamanho do efeito e a significância estatística permitem que a significância real seja avaliada sem um possível efeito enganoso do tamanho amostral^{3,4}, que pode ocorrer quando apenas a significância estatística é levada em consideração⁵. Dessa forma é possível descrever e analisar os efeitos observados já que efeitos grandes, mas não estatisticamente significativos, sugerem que as pesquisas futuras necessitam de maior poder do teste (maior tamanho da amostra), enquanto efeitos pequenos, mas significativos devido ao grande tamanho amostral, devem ser levados em consideração e discutidos, evitando a supervalorização do efeito observado³.

O nível de significância ainda domina a preferência dos pesquisadores ao discutir os dados encontrados, mesmo não sendo novo o debate sobre a necessidade de sua adequada interpretação⁶ já que o significado real e as interpretações isoladas do p-valor podem vir acompanhados por erros de interpretação⁷. Em estudos observacionais, muito mais do que em ensaios randomizados, viés e confusão podem suprimir a premissa de que há apenas 5% de probabilidade de que o efeito observado seja visto por acaso quando na realidade não há efeito, já que por definição, neste tipo de estudo, não há uma intervenção e a exposição pode não ser a única explicação potencial para as diferenças observadas nos resultados⁵.

Os testes de hipóteses são aplicados para que se controle as probabilidades de erros ao se rejeitar ou não uma hipótese. Mas, quando analisados isoladamente, os resultados destes testes apenas informam a probabilidade do resultado encontrado ser do acaso e, é frequente que resultados com valores de probabilidade mais baixos (por exemplo, $p < 0,001$) sejam erroneamente interpretados como tendo um efeito mais forte do que aqueles com valores de p mais elevados (por exemplo, $p < 0,05$)⁸. A determinação da magnitude do efeito de interesse e a precisão da estimativa da magnitude desse efeito⁹ são aspectos fundamentais a serem considerados para a ponderação da importância clínica ou prática dos resultados, devendo-se para isso considerar a análise dos tamanhos dos efeitos e os intervalos de confiança^{4,9}.

Cohen apresenta e classifica os tamanhos de efeitos para diversas metodologias estatísticas^{10,11}, sendo comumente apresentado como a diferença média padronizada (d de Cohen ou g de Hedges) ou como a força da associação (r de Pearson) entre dois grupos ou variáveis¹². Cohen^{10,11} também forneceu diretrizes para a interpretação desses valores, com base na noção de que um efeito médio deve ser perceptível a olho nu de um observador cuidadoso: valores de 0,20; 0,50 e 0,80 para d de Cohen e g de Hedges, e 0,10; 0,30 e 0,50 para o coeficiente de correlação, são comumente considerados, respectivamente, como indicativos de efeitos pequenos, médios e grandes, que representam a manifestação do fenômeno avaliado na população.

O tamanho do efeito depende do resultado obtido e da população de interesse e por isso sugere-se que a classificação da distribuição dos tamanhos de efeito deva ser analisada em cada uma das áreas de estudo¹².

Na área da saúde bucal coletiva, com grande frequência, as investigações buscam identificar associação entre fatores de risco ou de proteção para doenças ou medidas clínicas. Nesse caso, as medidas que quantificam a magnitude dessa associação normalmente são expressas pelo *Odds Ratio* (OR), razão de prevalência (RP) ou risco relativo (RR), a depender do delineamento do estudo e do tipo das variáveis estudadas¹³, sendo as duas primeiras medidas indicadas para estudos observacionais transversais, com o OR também indicado em estudos do tipo caso controle e RR, indicado para os estudos longitudinais. Essas medidas são consideradas estatísticas de tamanho de efeito não padronizadas, pois indicam a direção e a força da associação entre as variáveis de exposição e o desfecho.

Para o OR, que é o índice de tamanho do efeito mais utilizado para demonstrar aumento ou diminuição na chance de doença em estudos epidemiológicos, os autores determinaram que, para uma taxa de doença de 1% no grupo não exposto, os limites de referência que refletem uma “associação fraca” (d de Cohen=0,20); uma “associação moderada” (d de Cohen=0,50) ou uma “associação forte” (d de Cohen=0,80) são os OR de 1,68, 3,47 e 6,71, respectivamente. Considerando-se uma taxa de doença de 5% em pessoas não expostas, os limites de referência correspondentes são 1,52, 2,74 e 4,72⁸.

Dessa forma, o objetivo do presente estudo foi analisar e discutir um recorte da literatura científica específica da área de saúde bucal coletiva quanto ao cálculo, apresentação e discussão

do tamanho do efeito nos resultados de estudos observacionais. Além disso o estudo teve como objetivo detalhar os cálculos e a interpretação de medidas de tamanho do efeito que podem ser utilizados em artigos da área.

Métodos

Tipo de estudo e considerações éticas

Estudo observacional, retrospectivo, com discussão teórica. Por tratar-se de estudo com dados coletados de bases de domínio público, não houve necessidade de avaliação ética.

Estratégia de busca, seleção dos periódicos e estudos

Em janeiro de 2020, foi realizada busca nas bases de dados eletrônicas considerando o período de janeiro de 2015 a dezembro de 2019. As buscas ocorreram considerando artigos publicados com acesso aberto e gratuito, no MEDLINE via PubMed, utilizando os termos MeSH (*Medical Subject Headings*): (oral health) OR (dentistry) AND (logistic models) AND (analysis regression) AND free full text[sb] AND “last 5 years”[PDat])). Foram incluídos todos os estudos observacionais encontrados.

Variáveis do estudo

Duas examinadoras calibradas realizaram a busca dos artigos e por consenso, auxiliadas em casos de dúvidas ou discordâncias por uma terceira examinadora, coletaram e analisaram as seguintes informações nos artigos selecionados:

Informações dos artigos: periódico; ano de publicação.

Sobre o estudo: tipo de estudo; objetivo do estudo; tamanho da amostra; número de variáveis; detalhamento dos desfechos; instrumentos utilizados na coleta dos dados; presença ou não do cálculo estatístico da amostra; parâmetros utilizados para o cálculo do tamanho da amostra; coerência do cálculo da amostra com o objetivo da pesquisa; metodologia estatística utilizada, cita o termo tamanho de efeito?; apresentação do tamanho do efeito e, se sim: qual medida foi apresentada, qual o valor do tamanho do efeito mínimo significativo, se o valor mínimo significativo foi médio ou grande e não significativo, esse achado foi discutido?, se o valor mínimo

significativo foi pequeno e significativo, foi discutido?; considerou na conclusão o tamanho do efeito encontrado?

Cálculo e classificação do tamanho do efeito

As medidas de tamanho de efeito encontradas nos artigos na área foram detalhadas quanto a seus conceitos, cálculos e as interpretações.

Tamanho de efeito nas normas das revistas na área

Foi realizada a busca nas normas das revistas responsáveis pela publicação de três ou mais artigos selecionados para a presente pesquisa, buscando-se nas diretrizes aos autores a presença de recomendação de apresentação de tamanho de efeito.

Resultados

Descrição dos estudos

Foram incluídos no estudo 123 artigos, de 66 periódicos, sendo 9,8% (12) publicados em 2015; 17,1% (21) em 2016; 30,1% (37) em 2017; 25,2% (31) em 2018 e 17,9% (22) em 2019.

Na Tabela 1 é apresentado um resumo das principais características dos artigos analisados. Nota-se que a maioria dos estudos selecionados apresentava um único desfecho (74%), não relatou o cálculo amostral (69,9%) e dentre os que relataram, em 70,3% deles, havia coerência entre o cálculo amostral e o objetivo do estudo. Para os artigos em que essas coerências não foram observadas, nota-se um erro comum de calcular a amostra para o objetivo de estimar prevalências em estudos com objetivo de medir associação.

Na Tabela 2 são apresentadas as metodologias estatísticas utilizadas nos trabalhos e a forma de apresentação dos resultados. Nota-se que a regressão logística foi a metodologia estatística mais utilizada e os tamanhos dos efeitos das associações foram representados nos artigos principalmente pelo *Odds ratio*, que por sua vez, teve magnitudes pequenas e pouco discutida na maioria dos artigos. Nota-se ainda que apenas 3,3% dos artigos mencionam o termo tamanho do efeito e 24,4% não consideram, apesar de terem calculado, o tamanho do efeito na discussão dos resultados.

Tabela 1. Características dos artigos analisados (janeiro de 2015 a dezembro de 2019, Base MEDLINE - via PubMed).

Característica	Categorias	Frequências	%
Número de desfechos	1	91	74,0
	2	25	20,3
	3	3	2,4
	4 ou mais	4	3,3
Apresentação de cálculo da amostra	Não	86	69,9
	Sim	37	30,1
Coerência do cálculo da amostra com a pergunta e o delineamento	Não	11	29,7
	Sim	26	70,3

Fonte: Autores.

Tabela 2. Características metodológicas das pesquisas na área de Saúde Coletiva (janeiro de 2015 a dezembro de 2019, Base MEDLINE - via PubMed).

Característica	Categorias	Frequências	%
Metodologias estatísticas utilizadas	Regressão linear	1	0,8
	Regressão Logística	121	98,4
	Regressão de Poisson	1	0,8
Tamanhos de efeito apresentados	Odds ratio	116	94,3
	Risco relativo	2	1,6
	Razão de prevalência	1	0,8
	Não apresenta	4	3,2
OR mínimo significativo (tamanho de efeito) nos casos em que são apresentados OR	Até 1,5	54	55,1
	1,5-2,0	16	16,3
	2,0-2,5	12	12,2
	2,5-3,0	7	7,1
	>3,0	9	9,2
Quando o OR mínimo significativo foi médio ou grande, foi discutido	Não	3	33,3
	Sim	6	66,7
Quando o OR mínimo significativo foi pequeno foi discutido	Não	12	44,4
	Sim	15	55,6
Fala em tamanho de efeito	Não	119	96,7
	Sim	4	3,3
Levou em consideração o tamanho do efeito para a conclusão?	Não	30	24,4
	Sim	93	75,6

Fonte: Autores.

Tamanho de efeito nas normas das revistas na área

Na Tabela 3 são apresentados os resultados da busca nas normas das revistas quanto à presença de recomendação aos autores para a apresentação de tamanho de efeito em seus manuscritos. Nota-se que juntas publicaram 50,3% da produção avaliada e apenas 2 dos 10 periódicos listados fazem menção, nas diretrizes aos autores, sobre a apresentação do tamanho do efeito.

Análise dos tamanhos de efeito apresentados

Como forma de apresentar os tamanhos de efeito utilizados nos artigos da área, detalha-se conceitos de *Odds ratio* (OR), risco relativo (RR) e razão de prevalência (RP) e os seus cálculos com base em dados simulados.

Tabela 3. Periódicos com mais artigos avaliados e recomendações sobre apresentação de tamanho do efeito, segundo diretrizes aos autores. (janeiro de 2015 a dezembro de 2019, Base MEDLINE - via PubMed).

Periódico	Frequência	%	Recomendação
Braz Oral Res	16	13,0	Não constava
Cien Saude Colet	11	8,9	Não constava
PLoS One	9	7,3	<ul style="list-style-type: none"> • Reporting of statistical methods: If appropriate, provide sample sizes, along with a description of how they were determined. <i>If a sample size calculation was performed, specify the inputs for power, effect size and alpha.</i> Where relevant, report the number of independent replications for each experiment. • Reporting of statistical results (Reporting parameters): Test statistics (F/t/r) and associated degrees of freedom should be provided. <i>Effect sizes and confidence intervals should be reported where appropriate.</i> If percentages are provided, the numerator and denominator should also be given.
Int J Environ Res Public Health	7	5,7	Não constava
Int J Equity Health	4	3,3	Não constava
Aust Dent J.	3	2,4	Não constava
BMC Public Health	3	2,4	Não constava
BMJ Open	3	2,4	<ul style="list-style-type: none"> • Results: main results with (for quantitative studies) 95% confidence intervals and, where appropriate, the exact level of statistical significance and the number need to treat/harm. Whenever possible, state absolute rather than relative risks
J Appl Oral Sci.	3	2,4	Não constava
J Formosan Med Assoc.	3	2,4	Não constava

Fonte: Autores.

Odds ratio (OR)

Os OR com os respectivos intervalos de confiança podem ser estimados a partir dos coeficientes dos modelos de regressão logística.

Para exemplificar o cálculo e facilitar a interpretação da medida foram utilizados dados simulados, apresentados na Tabela 4. Simulou-se dois estudos transversais para avaliar a associação entre o consumo de bebidas adoçadas e a experiência de cárie em crianças, com resultados semelhantes, mas tamanhos de amostra diferentes, tendo sido utilizada a análise de regressão logística para estimar os OR.

Considerando-se os dados da simulação 1, observa-se que o tamanho da amostra utilizado foi de 64. Apesar do OR ser de 2,15, o intervalo de confiança é amplo devido ao pequeno tamanho da amostra (IC95%: 0,66-6,95) e a associação não foi estatisticamente significativa ($p=0,3211$). Foi então simulado o resultado do mesmo estudo (Simulação 2), porém com tamanho maior da amostra ($n=632$). Observa-se que os resultados foram semelhantes, ou seja, o OR foi de 2,11, mas

com IC95% de 1,44-3,08 e nesse caso a associação foi estatisticamente significativa ($p=0,0001$).

Nota-se que nos dois casos, o OR é próximo a dois, mas dependendo do tamanho da amostra há alteração na amplitude do intervalo de confiança e na significância estatística. Na simulação 2 observa-se que as crianças que consumiam bebidas adoçadas apresentavam 2,11 (IC95%: 1,44-3,08) vezes mais chance de apresentar experiência de cárie. Para compreender o que representa essa chance significativa, nota-se que no grupo de crianças que não consumiam bebidas adoçadas foram observados 172 escolares que tinham experiência de cárie, portanto, a chance de apresentar experiência de cárie nesse grupo é de $172/75=2,29$. Da mesma forma, a chance de apresentar experiência de cárie em crianças que consomem bebidas adoçadas é de $319/66=4,83$. A razão entre essas duas chances ($4,83/2,29$) resulta no *Odds ratio* (2,11).

Quando o OR é significativamente maior que um, a categoria estudada apresenta mais chance do evento do que a categoria de referência.

Razão de prevalência (RP)

As RP com os respectivos intervalos de confiança podem ser estimadas a partir dos modelos de regressão Binomial Negativa e Poisson.

Na Tabela 4 são também apresentados os resultados da simulação 2, calculando esta medida de associação em substituição ao OR. Nota-se que no grupo das crianças com experiência de cárie, a prevalência de escolares que consumiam bebidas adoçadas é 1,19 vez maior do que no grupo das crianças sem experiência de cárie. No grupo de crianças com experiência de cárie, a prevalência de crianças que não consomem bebidas adoçadas é de 69,6% e de 82,9% para as que consomem. Calculando a razão entre as duas prevalências (82,9%/69,6%) chega-se na razão de prevalência (1,19). Quanto maior o afastamento da RP em relação a RP=1 (tanto para mais como para menos), maior é o tamanho do efeito para essa variável.

Quando a RP é significativamente maior que um, a categoria estudada apresenta maior prevalência do evento do que a categoria de referência.

Risco relativo (RR)

Essa medida de associação só pode ser calculada em estudos longitudinais do tipo coorte¹³ e, portanto, representa o risco relativo de desenvolver o desfecho nos expostos em relação aos não expostos. Os RR com os respectivos intervalos de confiança podem ser estimados a partir dos modelos de regressão Binomial Negativa e Poisson.

Enquanto a já citada RP é a razão entre duas prevalências, o RR é a razão entre duas incidências. Como exemplo, em um estudo simulado (Tabela 5), avaliou-se o impacto da experiência de cárie na qualidade de vida relacionada a saúde

bucal. A partir da análise de regressão binomial negativa foram estimados os RR. Para a experiência de cárie, o RR foi de 1,50 (IC95%: 1,04-2,17), $p=0,0204$. Nesse caso a interpretação é que a presença de cárie é associada a 50% de aumento no impacto da saúde bucal na qualidade de vida. Da mesma forma que o OR e a RP, quanto maior o afastamento do RR em relação ao RR=1 (tanto para mais como para menos), maior é o tamanho do efeito para essa variável.

Nota-se que no grupo sem experiência de cárie o risco de ter pior qualidade de vida é de 30/100=30%. Já no grupo com experiência de cárie o risco de ter pior qualidade de vida é de 45/100=45%. Então o risco relativo=45%/30%=1,5. Ou seja, as pessoas com experiência de cárie têm 1,5 vez mais risco de ter pior qualidade de vida.

Discussão

O presente estudo reforça o fato de que apesar da literatura da área de estatística expor há muito tempo a necessidade e a importância da apresentação e discussão do tamanho de efeito nos artigos, observou-se que apenas 3,3% dos artigos avaliados mencionaram em seus textos o termo tamanho do efeito e 24,4% não consideraram o tamanho do efeito na discussão dos resultados, apesar de o terem calculado. Desde 1925 Fisher propunha que os pesquisadores acrescentassem a taxa de correlação ou o η (eta) à significância da análise de variância (ANOVA), ou seja, o tamanho do efeito, representando a força da associação entre as variáveis independentes e dependentes¹⁴.

Tabela 4. Exemplo do uso do odds ratio (OR) ou razão de prevalência (RP) na análise da associação entre o consumo de bebidas adoçadas e a experiência de cárie em crianças (dados simulados).

Consumo de bebidas adoçadas	Frequência	Experiência de cárie		(IC95%)	p-valor
		Não	Sim		
Odds ratio - Simulação 1					
Não	25	8 (30,0%)	17 (68,8%)	Referência	
Sim	39	7 (16,9%)	32 (81,8%)	2,15 (0,66-6,95)	0,3211
Odds ratio - Simulação 2					
Não	247	75 (30,4%)	172 (69,6%)	Referência	
Sim	385	66 (17,1%)	319 (82,9%)	2,11 (1,44-3,08)	0,0001
Razão de prevalência - Simulação 2					
Não	247	75 (30,4%)	172 (69,6%)	Referência	
Sim	385	66 (17,1%)	319 (82,9%)	1,19 (1,08-1,31)	<0,0001

Fonte: Autores.

Tabela 5. Uso do Risco Relativo (RR) na análise da associação entre a experiência de cárie e a qualidade de vida relacionada à saúde bucal (dados simulados).

Experiência de cárie	Frequência	Qualidade de vida		RP (IC95%)	p-valor
		Pior	Melhor		
Não	100	30 (30,0%)	70 (70,0%)	Referência	
Sim	100	45 (45,0%)	55 (55,0%)	1,50 (1,04–2,17)	0,0204

Fonte: Autores.

Ainda que de forma mais lenta que o necessário, tem havido uma pressão por parte das revistas científicas sobre os pesquisadores para que os tamanhos dos efeitos sejam relatados e interpretados nos artigos^{3,15}. Entre as revistas analisadas no recorte da literatura realizado neste estudo, selecionou-se aquelas que publicaram três ou mais dos estudos incluídos, que juntas contabilizaram mais de 50% dos artigos selecionados, e verificou-se que apenas 20% delas sugeriam explicitamente em suas diretrizes aos autores, que o tamanho do efeito fosse relatado nos artigos. Este achado está de acordo com estudo anterior que identificou que apenas uma pequena parcela de periódicos de diversas áreas recomendava explicitamente nas normas aos autores o cálculo da magnitude do efeito¹⁵.

Há muita confusão na literatura sobre a definição correta de tamanho do efeito que por vezes vem sendo utilizado de forma incorreta. Kelley e Preacher¹⁶ propõem uma definição para o tamanho do efeito e discutem a partir de três particularidades (dimensão, medida/índice e valor). Segundo esses autores, o tamanho de efeito pode ser apresentado com uma estatística que estima a magnitude do efeito (por exemplo o coeficiente de correlação=0,3) ou com uma interpretação qualitativa dessa estatística (correlação mediana), que deve levar em consideração a aplicabilidade prática do achado. Ainda segundo os autores, o tamanho do efeito está frequentemente vinculado à ideia de significância substantiva (por exemplo, importância prática, clínica, médica ou gerencial), que pode ser entendida como o grau em que as partes interessadas (cientistas, profissionais, políticos, gerentes, consumidores, decisão fabricantes, público em geral etc.) considerariam uma descoberta importante e digna de atenção e possivelmente de ação.

Neste contexto, a utilização exclusiva do nível de significância para analisar e discutir os achados não é suficiente^{4,5} já que este apenas informa se o resultado da pesquisa é devido ao efeito analisado ou ao acaso (variabilidade da amostra). A

significância prática informa se os resultados são úteis no mundo real e é analisada pelo tamanho do efeito encontrado, sendo fundamental chamar a atenção dos pesquisadores para a necessidade de comunicar os tamanhos de efeito em suas publicações^{10,11}. Além disso, tamanhos de efeito previamente observados podem servir de base para o cálculo do poder, para a estimativa do tamanho amostral adequado em estudos posteriores^{1,3,17}, para a compreensão dos resultados do estudo no contexto de pesquisas anteriores, além de facilitar a incorporação de seus resultados em meta-análises futuras, muito relevantes como método padrão de revisão quantitativa em biologia⁹.

Segundo Kirk¹⁴ a magnitude do efeito pode ser classificada em três categorias: a) medida da força das associações, b) medida do tamanho do efeito (tipicamente diferença padronizada entre as médias), c) outras medidas.

A maioria dos artigos que apresenta e/ou discute tamanho de efeito utiliza ANOVA, teste t e calcula o efeito de acordo com Cohen^{10,11}, mas como verificado no presente estudo, essas metodologias estatísticas são pouco utilizadas em artigos na área de Saúde Bucal Coletiva e muito pouco se fala em tamanho de efeito quando se utiliza análise de regressão logística, metodologia estatística utilizada em 98,4% dos artigos avaliados.

No presente estudo foi observado que os artigos apresentam o tamanho do efeito pela medida da força das associações entre as variáveis, já que em 94,3% dos artigos selecionados foi verificado a apresentação do *Odds ratio* (OR), corroborando ao explicitado por Chen *et al.*⁸ que relataram que este provavelmente seja o índice de tamanho de efeito mais amplamente utilizado em estudos epidemiológicos por refletir as chances de um resultado bem-sucedido ou desejado no grupo de intervenção em relação às chances de um resultado similar no grupo controle¹⁵.

Breaugh¹⁸ destaca alguns conceitos errados sobre estimativas de tamanho de efeito e introduz uma série de medidas de tamanho de efeito que, segundo o autor, dependendo do contexto

da pesquisa e do público, podem melhor comunicar a importância da relação entre duas variáveis. No caso de variáveis dicotômicas, há uma limitação do uso do phi como uma medida do tamanho do efeito (ϕ é uma medida comumente utilizada como tamanho do efeito em análises de tabelas de contingência 2 x 2) porque a sua possível amplitude é afetada pela distribuição da variável. Segundo o autor em determinadas áreas como a medicina, é comum que uma taxa de risco seja reportada como uma medida de tamanho de efeito. E, neste contexto, muitos estatísticos têm sugerido reportar o OR como uma medida do efeito, em vez da taxa de risco ou o coeficiente phi, como verificado nos artigos avaliados no presente estudo.

Uma propriedade desejável de um OR é que seu possível intervalo de valores não é influenciado pelas distribuições marginais das variáveis. Foi observado no presente estudo que os artigos avaliados apresentaram na grande maioria (96,7%) o OR como uma medida do tamanho do efeito, mas 24,4% não levaram em consideração este valor no momento da discussão dos resultados e conclusão do trabalho o que leva à compreensão de que os autores têm baseado a discussão e a conclusão de seus trabalhos apenas nos p-valores. Nos artigos avaliados, 8,9% concluem com base em associação significativa sem mencionar que o tamanho do efeito era pequeno. Além disso observou-se 33,3% dos artigos concluem não significância na associação, sem mencionar que o OR foi médio ou grande, ou seja, provavelmente a amostra foi pequena no estudo e outros estudos precisam ser realizados com amostra maiores. Enfatiza-se, portanto, que os autores têm que levar em consideração e em conjunto, essas duas informações importantes, ou seja, o p-valor e o tamanho do efeito, no caso o grau de associação (OR).

Chen *et al.*⁸ apresentam uma classificação do OR em pequeno, médio e grande de acordo com as probabilidades que estão sendo comparadas e Durlak¹⁵ apresenta um guia para a seleção, cálculo e interpretação dos tamanhos de efeito. Nesse estudo são discutidos diferentes tipos de tamanhos de efeito comumente usados.

Ferguson¹⁹ recomenda tamanhos de efeito de razão de chances pequenos, médios e grandes de 2,0; 3,0 e 4,0, mas recomenda cautela em seu uso, pois não são “ancorados” ao coeficiente de correlação de Pearson. Embora muitos apontaram problemas com ϕ como medida de associação e incentivarem o uso de OR como uma alternativa, recomendações de tamanho do efeito para OR

não existem em geral. Os autores demonstram a relação entre o ϕ e o OR e recomendam tamanhos dos efeitos da razão de chances, derivados do trabalho de Cohen. Para uma alocação 1:1 (*allocation ratio*), OR de 1,22; 1,86 e 3,00 correspondem a tamanhos de efeito pequenos, médios e grandes.

Assim, o tamanho de efeito (significância substantiva) complementa a significância estatística e uma medida não substitui a outra, devendo ser analisadas de forma complementar, para que seja dado um passo em direção à veracidade científica. Ialongo²⁰ apresenta uma introdução e um guia para o leitor interessado no uso da estimativa de tamanho de efeito e ressalta que a evidência pode ser quantificada pelos testes de hipóteses, que representam a probabilidade (ou p-valor) pelo qual é provável que se considere a observação moldada pelo acaso (a chamada “hipótese nula”) e não pelo fenômeno (a chamada “hipótese alternativa”). O tamanho no qual o p-valor é considerado pequeno ou suficiente para excluir o efeito do acaso corresponde à significância estatística. Então, quando o pesquisador chega a um resultado não significativo, devem ser consideradas duas possibilidades: a primeira é que não há o fenômeno e se está observando apenas o efeito do acaso, e a segunda é que o fenômeno existe, mas seu efeito é pequeno e confundido com o efeito do acaso.

É na segunda possibilidade que se coloca a questão da importância de apresentar o fenômeno quando este realmente existe, quantificando-o pelo cálculo do tamanho do efeito, ou seja, o quão grande (ou pequeno) é o efeito esperado produzido pelo fenômeno em relação à observação através da qual pretende-se detectá-lo. Por esse motivo, os pesquisadores devem ser incentivados a apresentar o tamanho do efeito em seu trabalho, particularmente o relatando sempre que o p-valor for mencionado.

Dentre as limitações do presente estudo, ressalta-se que foi realizado um recorte da literatura para a contextualização do tema e as frequências apresentadas aplicam-se apenas a esse recorte. Apesar disso, os resultados aqui apresentados permitiram a realização da discussão teórica sobre o tema oportunizando a compreensão de que o relato e discussão do tamanho do efeito nos estudos deve ser feita como rotina e que revisores e editores de periódicos científicos devem atentar-se ao seu relato e apropriada discussão.

Conclui-se no presente estudo que os pesquisadores, em sua maioria, restringiram a discussão de seus resultados apenas à significância

estatística encontrada nas associações testadas e os periódicos não indicam explicitamente a necessidade de apresentar a magnitude dos efeitos, bem como de considerá-la na discussão dos resultados e conclusão do estudo.

Colaboradores

FM Flório, GMB Ambrosano e MC Meneghim idealizaram o estudo. FM Flório e L Zanin conduziram o estudo e a redação do manuscrito. FM Flório, L Zanin, LM Santos Junior, GMB Ambrosano e MC Meneghim contribuíram com a revisão final e crítica do manuscrito.

Referências

1. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Front Psychol* 2013; 4:863.
2. Wilkinson L, Task Force on Statistical Inference. Statistical methods in psychology journals: Guidelines and explanations. *Am Psychol* 1999; 54:594-604.
3. Lindenau JDR, Guimarães LSP. Calculando o tamanho de efeito no SPSS. *Rev HCPA* 2012; 32(3):363-381.
4. Espírito Santo H, Daniel F. Calcular e apresentar tamanhos do efeito em trabalhos científicos (1): As limitações do $p < 0,05$ na análise de diferenças de médias de dois grupos. *Rev Port Invest Comport Soc* 2015; 1(1):3-16.
5. Schuemie MJ, Ryan PB, DuMouchel W, Suchard MA, Madigan D. Interpreting observational studies: why empirical calibration is needed to correct p-values. *Stat Med* 2014; 33(2):209-218.
6. Baker M. Statisticians issue warning over misuse of P values. *Nature* 2016; 531(7593):151.
7. Gigerenzer G. Statistical Rituals: The Replication Delusion and How We Got There. *Adv Methods Pract Psychol Sci* 2018; 1(2):198-218.
8. Chen H, Cohen P, Chen S. How Big is a Big Odds Ratio? Interpreting the Magnitudes of Odds Ratios in Epidemiological Studies. *Commun Stat Simul Comput* 2010; 39(4):860-864.
9. Nakagawa S, Cuthill IC. Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biol Rev Camb Philos Soc* 2007; 82(4):591-605.
10. Cohen J. *Statistical power analysis for the behavioral sciences*. 2ª ed. Mahwah: Lawrence Erlbaum Associates; 1988.
11. Cohen J. A power primer. *Psychol Bull* 1992; 112:155-159.
12. Brydges CR. Effect Size Guidelines, Sample Size Calculations, and Statistical Power in Gerontology. *Innov Aging* 2019; 3(4):igz036.
13. Papaléo CLM. *Estimação de risco relativo e razão de prevalência com defecho binário*. Porto Alegre: Universidade Federal do Rio Grande do Sul; 2009.
14. Kirk RE. Practical significance: A concept whose time has come. *Edu Psychol Measuram* 1996; 56:746-759.
15. Durlak JA. How to select, calculate, and interpret effect sizes. *J Pediatr Psychol* 2009; 34(9):917-928.
16. Kelley K, Preacher KJ. On Effect Size. *Psychol Methods* 2012; 17(2):137-152.
17. Olivier J, Bell ML. Effect sizes for 2x2 contingency tables. *PLoS One* 2013; 8(3):e58777.
18. Breaugh JA. Effect Size Estimation: Factors to Consider and Mistakes to Avoid. *Journal of Management* 2003; 29(1) 79-97.
19. Ferguson CJ. An effect size primer: A guide for clinicians and researchers. *Prof Psychol Res Pract* 2009; 40(5):532-538.
20. Ialongo C. Understanding the effect size and its measures. *Biochem Med (Zagreb)* 2016; 26(2):150-163.

Artigo apresentado em 25/04/2022

Aprovado em 12/08/2022

Versão final apresentada em 14/08/2022

Editores-chefes: Romeu Gomes, Antônio Augusto Moura da Silva