



"ENVIO DE DADOS FALTANTES POR AUTORES DE PESQUISAS INCLUÍDAS EM REVISÃO SISTEMÁTICA E FATORES ASSOCIADOS: ESTUDO TRANSVERSAL”

Natália Dutra dos Reis, Carolina Müller Ferreira, Marcus Tolentino Silva, Taís Freire Galvão
Faculdade de Ciências Farmacêuticas – Universidade Estadual de Campinas

I. OBJETIVOS

O objetivo principal foi analisar a frequência de envio de dados faltantes em revisão sistemática e os fatores associados.

Foram estabelecidos também objetivos específicos: estimar os motivos para ausência de resposta e envio dos dados solicitados aos autores; e também avaliar a efetividade do contato com autores e processos envolvidos para obtenção dos dados faltantes.

II. DESCRIÇÃO

A presente pesquisa trata-se de estudo transversal realizado com estudos identificados em revisão sistemática prévia sobre prevalência de obesidade infantil no Brasil (Registro no PROSPERO: CRD42018091713). Foram elegíveis aqueles estudos que não apresentaram os dados necessários para inclusão na revisão sistemática ou que o apresentaram de forma incompleta. O desfecho primário foi a frequência de envio de dados faltantes nos estudos, aferido por envio dos dados solicitados através de contatos com autores/coautores correspondente do estudo. As variáveis independentes foram os demais dados extraídos na revisão sistemática, como as características do estudo. Outras variáveis geradas durante o contato com os autores também foram analisadas, por exemplo: as características do autor contatado, tempo de resposta, entre outras.

Para obter os dados, foram enviados e-mails para autores e coautores, em três tentativas mínimas, com intervalo aproximado de 10 dias entre cada contato. Um modelo de e-mail padrão para a solicitação dos dados foi elaborado, contendo tabelas padronizadas para preenchimento dos dados nos critérios estabelecidos na revisão sistemática. Também foi oferecida a opção de envio do banco de dados anonimizado. Os e-mails foram enviados a partir de uma conta Gmail criada especificamente para a presente pesquisa. Os envios dos e-mails e suas respostas foram compilados em planilha Excel, de modo a padronizar as tentativas de contato e os resultados de cada tentativa. Os autores foram contatados durante 108 semanas (abril 2018 – maio 2020).



No caso de publicações duplicadas, foi considerado o ano da primeira publicação. Uma das publicações foi escolhida arbitrariamente como publicação principal e foi registrado o gênero do autor correspondente, ou do pesquisador responsável pelo envio dos dados. Criamos variáveis *dummy* para a análise de sobrevivência. Número de publicações, autores contatados e e-mails enviados foram dicotomizados em 1 ou 2 ou mais. Caso tenha sido publicado na forma de tese/dissertação e artigo, ou se o e-mail foi enviado para servidor pessoal e institucional, foi considerada a segunda categoria. O número de tentativas de contato não contabilizou as respostas dos autores. O tempo foi medido pela diferença em dias entre a data da primeira tentativa de contato e o recebimento da resposta ou finalização das tentativas. Atribuímos *post-hoc* que uma mensagem inicial levava 15 minutos de trabalho e os e-mails de acompanhamento, 5 minutos. O tempo de trabalho gasto no envio de e-mails foi calculado pela soma de cada tipo de e-mail enviado multiplicado pelo tempo, em minutos.

Descrevemos as frequências absolutas e relativas das variáveis, estratificando-as pelo sucesso no recebimento dos dados e pelo tempo médio e desvio padrão, em dias. As frequências de resposta foram descritas pelo tempo de trabalho, em minutos, e as diferenças na distribuição por tipo de resposta foram testadas pelo teste de *Wilcoxon*. A análise estatística comparou o sucesso no recebimento dos dados solicitados de acordo com variáveis independentes. Estimamos a razão de risco (HR) e o intervalo de confiança de 95% (IC) para o sucesso em cada variável em comparação com a categoria de referência, com modelos de riscos proporcionais de *Cox* ponderados. A significância estatística foi assumida se o valor de $p < 0,05$. As curvas de eventos cumulativos foram estimadas com o método de *Kaplan-Meier*, considerando o sucesso (obtenção dos dados) como resultado de sobrevivência e o teste de *log-rank* foi usado para comparar os efeitos da variável estatisticamente significativa. Nenhum ajuste para testes múltiplos foi feito. Todas as análises foram realizadas com o programa estatístico *Stata*, versão 14.2 (*StataCorp, College Station, Texas, EUA*).

O projeto foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Estadual de Campinas pelo Parecer Consolidado nº. 3.671.280/2019, Certificado de Apresentação de Apreciação Ética no. 18850919.6.0000.5404. O Termo de Consentimento Livre e Esclarecido não foi considerado necessário.

III. RESULTADOS

Dos 9.394 registros encontrados, 274 estudos potencialmente elegíveis para a revisão sistemática necessitavam de mais dados, dos quais 110 eram duplicados, resultando em 164 estudos únicos contatados. Foram enviados 272 e-mails de solicitação de dados (iniciais) e 398 mensagens de acompanhamento, consumindo 6.070 minutos no total (mediana 30, intervalo interquartil: 25, 45; 101,2 horas úteis).



Dos 164 estudos, autores de 51 nos responderam com os dados solicitados preenchidos na tabela ou no banco de dados anonimizado (31,1%; IC 95% = [23,9, 38,3%]), dos quais 1 foi enviado em disquete pelo correio e em 8 estudos os autores estavam inacessíveis, porque os e-mails falharam e não haviam endereços de e-mail alternativos disponíveis (Tabela 1). Uma pesquisadora estava inacessível porque já havia falecido, mas os coautores nos enviaram os dados necessários. O tempo médio para receber uma resposta ou desistir do contato foi de $89,6 \pm 113,8$ dias (mediana 36,0, intervalo interquartil: 17,5, 142,5). A maioria dos estudos foram conduzidos em 2010-2017 (31,1%), publicados em 2014-2019 (29,9%), como artigos científicos (78,7%). Maior frequência de sucesso (obtenção dos dados) e menor tempo para receber resposta ou finalizar o contato foi observado em pesquisas publicadas em 2014-2019 (49,0%; $73,2 \pm 98,6$ dias). Os estudos contatados até 3 vezes tiveram menor duração em dias para recebermos resposta ou finalizarmos o contato ($47,7 \pm 73,3$) (Tabela 1).

110 autores dos estudos responderam a nossa mensagem (67,1%), mas a maioria não enviou os dados solicitados devido à indisponibilidade do conjunto de dados da pesquisa, incluindo 2 pesquisadores que se aposentaram desde a publicação da pesquisa e não tiveram mais acesso aos dados (14,0%), nenhum motivo informado após alguma interação inicial (13,4%), envio de documento que não continha os dados solicitados, como o artigo completo ou outro manuscrito (3,7%), recusou-se a enviar dados para a pesquisa (3,1%), ou o estudo foi percebido como ilegível após explicações dos autores (1,8%). Das 54 não respostas, 46 (28,1%) não responderam após pelo menos três tentativas de contato a um ou mais autores (Tabela 2). Não houve diferença no tempo de trabalho entre contato respondido e não respondido ($p = 0,374$).

Estudos publicados como artigos científicos (HR = 3.01, IC 95% = [1.18, 7.70]) e com mais de uma publicação (HR = 2.00, IC 95% = [1.14, 3.51]) tiveram uma probabilidade maior de enviar os dados solicitados para a revisão sistemática em comparação com estudos disponíveis apenas na forma de teses ou dissertações e em apenas uma publicação, respectivamente. A maior taxa de sucesso foi observada se o e-mail contatado fosse pessoal em vez de institucional (HR = 7.55, IC 95% = [2.67, 21.34]) e se houvesse até 3 tentativas de contato em vez de mais contatos (HR = 4.79, IC 95% = [2.46, 9.29]) (Figura 1).

IV. CONCLUSÃO

Solicitar dados para uma revisão sistemática consome tempo e é eficaz em apenas três em cada dez estudos. Para metade dos estudos, demorou-se mais de um mês para obtermos dados ou finalizarmos o contato, levando duas horas de trabalho para um caso de sucesso. O maior sucesso foi observado em até três tentativas de contato, contatando o e-mail pessoal de autores, em estudos publicados na forma de artigo científico, e em duas ou mais publicações.



Tabela 1. Características dos estudos com dados solicitados, dados enviados e tempo para recebimento ou finalização do contato (média e desvio padrão [DP]), em dias (n = 164).

Variáveis	Total		Sucesso		Tempo	
	N	%	N	%	Média	±DP
Ano da pesquisa						
1986–2003	33	20.1	5	15.2	61.5	100.7
2004–2006	36	22.0	11	30.6	127.8	131.8
2007–2009	38	23.2	10	26.3	96.8	107.5
2010–2017	51	31.1	24	47.1	79.7	111.1
Não informado	6	3.7	1	16.7	52.5	89.5
Ano de publicação						
1996–2007	37	22.6	8	21.6	78.2	113.0
2008–2010	42	25.6	10	23.8	112.6	115.5
2011–2013	36	22.0	9	25.0	96.7	130.4
2014–2019	49	29.9	24	49.0	73.2	98.6
Tipo de publicação						
Artigo	129	78.7	40	31.0	66.1	91.0
Tese ou dissertação	24	14.6	5	20.8	166.0	135.8
Artigo e tese ou dissertação	11	6.7	6	54.6	198.0	168.8
Número de publicações						
1	120	73.2	24	20.0	70.9	98.8
2	24	14.6	13	54.2	169.2	141.5
3–14	20	12.2	14	70.0	106.3	123.0
Tipo de instituição						
Public	132	80.5	12	37.5	75.0	102.9
Private	32	19.5	39	29.6	93.1	116.4
Gênero do autor correspondente						
Feminino	119	72.6	37	31.1	92.9	116.1
Masculino	45	27.4	14	31.1	80.7	108.3
Número de autores contatados						
1	104	63.4	31	29.8	66.1	97.3
2	52	31.7	16	30.8	120.6	129.9
3–8	8	4.9	4	50.0	193.0	107.9
Servidor de e-mail						
Pessoal	112	68.3	46	41.1	79.1	100.8
Institucional	26	15.9	5	19.2	131.8	159.0
Ambos	26	15.9	-	-	92.5	107.9
Número de e-mails enviados						
1–3	97	59.2	37	38.1	47.7	73.3
4–6	43	26.2	7	16.3	149.2	138.4



7-26	16	9.8	7	43.8	227.6	81.1
Falharam na entrega	8	4.9	-	-	-	-

Tabela 2. Respostas dos dados solicitados aos estudos por tempo de trabalho, em minutos (média e desvio padrão [DP], n = 164).

Desfecho	Total		Tempo	
	N	%	Média	±DP
Respondeu				
Dados enviados pelos autores	51	31.1	38.4	35.0
Conjunto de dados indisponível	23	14.0	38.0	13.0
Parou de responder	22	13.4	40.0	20.0
Enviou outro manuscrito sem o dado	6	3.7	30.0	17.0
Recusou a enviar os dados	5	3.1	30.0	11.7
Inelegibilidade atribuída após o contato	3	1.8	30.0	18.0
Não respondeu				
Sem resposta após 3 tentativas	46	28.1	38.5	14.3
E-mail falhou em chegar ao destinatário	8	4.9	20.6	7.8
Total	164	100.0	37.0	23.3

Figura 1. Taxa de sucesso de acordo com o tempo em dias de recebimento dos dados solicitados por tipo de publicação, tipo de servidor de e-mail, número de publicações e número de tentativas de contato.

