

Estado da publicação: O preprint não foi publicado em outro meio.

Prevalência de helmintíases e desempenho em leitura no 5º ano em 390 municípios brasileiros

Maxwel Adriano Abegg, Kibelle Iamã dos Santos Costa

<https://doi.org/10.1590/SciELOPreprints.18033>

Submetido em: 2026-09-11

Postado em: 2026-09-15 (versão 1)

(AAAA-MM-DD)

Prevalência de helmintíases e desempenho em leitura no 5º ano em 390 municípios brasileiros

Helminth prevalence and fifth-grade reading performance in 390 Brazilian municipalities

Kibelle Iamã dos Santos Costa (<https://orcid.org/0009-0007-7254-8373>); Maxwel Adriano Abegg (<https://orcid.org/0000-0002-0328-1122>)

Programa de Pós-Graduação em Ciências, Tecnologia e Saúde (PPGCTS), Instituto de Ciências Exatas e Tecnologia (ICET), Universidade Federal do Amazonas (UFAM), Itacoatiara, Amazonas, Brasil.

Título curto: Helmintíases e desempenho escolar municipal.

Resumo

A associação entre parasitoses intestinais e desvantagem socioeconômica é amplamente documentada, assim como o impacto negativo da privação material sobre a aprendizagem. Investigamos em que medida as disparidades municipais em leitura associadas à prevalência de quatro helmintíases — *Ascaris lumbricoides*, *Trichuris trichiura*, ancilostomídeos e *Schistosoma mansoni* — persistem depois do ajuste por fatores socioeconômicos e escolares. Neste estudo ecológico, vinculamos o inquérito nacional de parasitoses (2010–2015), que examinou 197.564 escolares em 521 municípios, aos resultados do Sistema de Avaliação da Educação Básica (Saeb) de 2019, nos 390 municípios com pelo menos 100 escolares examinados. A diferença bruta entre os extremos é grande: no quinto de maior prevalência de *A. lumbricoides*, 37,3% dos estudantes alcançaram 200 pontos ou mais em leitura, contra 59,8% no quinto de menor prevalência — lacuna de 22,4 pontos percentuais. O ajuste por renda, escolaridade materna, saneamento e estrutura das escolas atenua essa diferença em 59% a 79%. O que resta, de 4 a 9 pontos percentuais, é modesto e instável: resiste às verificações de sensibilidade, mas diminui quando entram os municípios de menor amostra e deixa de ser significativo na especificação contínua sob a ponderação por variância inversa, que o plano de análise definia como principal. A prevalência acrescenta pouco ao modelo completo, entre 0,4 e 1,2 ponto percentual de R^2 , e as distribuições dos extremos se sobrepõem amplamente; isolada, explica 17% da variação. Os quatro helmintos apresentaram coeficientes contínuos na mesma direção. Como os dados são municipais, nada se pode concluir sobre o aprendizado de cada criança infectada. A prevalência municipal de *A. lumbricoides* funciona, nestes dados, sobretudo como indicador de vulnerabilidade territorial, e pouco acrescenta aos indicadores socioeconômicos já monitorados. As rotinas de análise foram programadas e executadas por modelos de linguagem a partir de solicitações dos autores — sobretudo Claude Opus 5 e ChatGPT 5.6 —, em bibliotecas estatísticas correntes; aos autores coube definir as perguntas,

aprovar as decisões analíticas e conferir os dados de entrada nas fontes oficiais. Os resultados centrais foram reproduzidos por recomputação independente e o modelo espacial, reestimado na biblioteca PySAL; não houve revisão por estatístico independente nem reextração dos dados a partir das fontes primárias, para as quais os autores não têm formação. Dados, código, protocolo e registros de verificação acompanham o trabalho e estão depositados no Zenodo (DOI 10.5281/zenodo.22664954).

Palavras-chave: helmintíase; *Ascaris lumbricoides*; avaliação educacional; estudos ecológicos; determinantes sociais da saúde.

Abstract

The association between intestinal parasitic infections and socioeconomic disadvantage is widely documented, as is the negative impact of material deprivation on learning. We investigated the extent to which municipal disparities in reading associated with the prevalence of four helminthiases — *Ascaris lumbricoides*, *Trichuris trichiura*, hookworms and *Schistosoma mansoni* — persist after adjustment for socioeconomic and school factors. In this ecological study, we linked the national parasitological survey (2010–2015), which examined 197,564 schoolchildren in 521 municipalities, to the 2019 results of the Brazilian national student assessment (Saeb), in the 390 municipalities with at least 100 schoolchildren examined. The crude difference between the extremes is large: in the highest *A. lumbricoides* fifth, 37.3% of students reached 200 reading points or more, against 59.8% in the lowest fifth — a gap of 22.4 percentage points. Adjustment for income, maternal schooling, sanitation and school structure attenuates that difference by 59% to 79%. What remains, 4 to 9 percentage points, is modest and unstable: it withstands the sensitivity checks, but diminishes once municipalities with smaller samples are included and ceases to be significant in the continuous specification under inverse-variance weighting, which the analysis plan defined as the primary scheme. Prevalence adds little to the full model, between 0.4 and 1.2 percentage points of R^2 , and the distributions at the extremes overlap widely; on its own, it explains 17% of the variation. The four helminths showed continuous coefficients in the same direction. Because the data are municipal, nothing can be concluded about the learning of any individual infected child. Municipal *A. lumbricoides* prevalence functions, in these data, primarily as an indicator of territorial vulnerability, and adds little to the socioeconomic indicators already monitored. The analytical routines were programmed and executed by large language models from the authors' prompts — chiefly Claude Opus 5 and ChatGPT 5.6 —, using standard statistical libraries; the authors set the questions, approved the analytical decisions and checked the input data against the official sources. The central results were reproduced by independent recomputation and the spatial model re-estimated in the PySAL library; there was no review by an independent statistician and no re-extraction of the data from the primary sources, checks the authors are not

trained to perform. Data, code, protocol and verification records accompany the work and are deposited in Zenodo (DOI 10.5281/zenodo.22664954).

Keywords: helminthiasis; *Ascaris lumbricoides*; educational measurement; ecological studies; social determinants of health.

Introdução

A relação entre helmintíases intestinais e desenvolvimento cognitivo infantil é objeto de investigação há mais de um século, e o acúmulo de evidências não convergiu para uma resposta única. Os estudos que compararam crianças infectadas e não infectadas identificaram, entre as primeiras, pior desempenho em provas de leitura, memória e raciocínio. A meta-análise de 2018 que reuniu 36 desses trabalhos e 12.920 crianças, contudo, é descrita por seus próprios autores como um conjunto heterogêneo¹, e uma revisão sistemática de 2026 sobre escolares africanos chega à mesma leitura². Quando o problema passou a ser abordado por ensaios de vermifugação em massa — desenho que, em princípio, isolaria o efeito do tratamento —, tampouco se observou benefício médio consistente sobre peso, hemoglobina, cognição ou desempenho escolar^{3,4}. A controvérsia deslocou-se, assim, das comparações individuais para o alcance das intervenções populacionais.

Menos frequentado, um terceiro conjunto de estudos aproxima-se do desenho aqui adotado ao comparar áreas, e não indivíduos. O trabalho de referência explorou a campanha de erradicação da ancilostomíase conduzida no sul dos Estados Unidos na década de 1910 e estimou aumentos expressivos de matrícula, frequência e alfabetização nas áreas de maior infecção antes da intervenção⁵. Uma reanálise posterior, apoiada em séries históricas mais longas, ponderou que a convergência entre áreas já estava em curso e não se acelerou com a campanha⁶. O debate permanece aberto, e a divergência decorre, em boa parte, dos limites inerentes à comparação entre áreas. No Brasil, um estudo com escolares de Minas Gerais relatou pior desempenho no teste de Raven entre crianças com infecção moderada a alta por *A. lumbricoides*⁷, evidência individual restrita a um único contexto local.

Uma hipótese plausível para acomodar esses achados é que a prevalência de infecção opere, em parte, como proxy das vulnerabilidades estruturais do território. O principal obstáculo à aprendizagem não estaria na infecção em si, mas no conjunto de condições que a acompanham: onde há helmintíases, há também, com frequência, esgotamento sanitário inadequado, renda baixa e menor escolaridade materna.

O Brasil dispõe de material adequado para submeter essa hipótese a exame. Entre 2010 e 2015, um inquérito nacional examinou amostras de fezes de 197.564 escolares em 521 municípios de todas as unidades da federação, pelo método de Kato-Katz, com duas lâminas de uma amostra por escolar⁸. Esses dados já foram explorados em relação a água, saneamento e pobreza — com a pobreza identificada como o preditor mais forte⁹ —, a fatores ambientais da

esquistossomose¹⁰ e à modelagem da distribuição dos geo-helminthos no país¹¹. Pelo que identificamos, porém, ainda não haviam sido vinculados a um desfecho de avaliação nacional de aprendizagem. É essa a lacuna que o presente trabalho procura preencher. Perguntamos quanto da diferença de desempenho em leitura entre municípios com maior e menor prevalência permanece após o ajuste por condições socioeconômicas mensuradas, e se o que resta se associa de modo semelhante aos quatro parasitos examinados — *A. lumbricoides*, *T. trichiura*, ancilostomídeos e *S. mansoni*.

Métodos

Desenho e plano. Trata-se de estudo ecológico: a unidade de análise é o município, não a criança, e tudo o que se descreve adiante se refere ao conjunto dos municípios. A exposição foi medida antes do desfecho. Um plano de análise foi redigido e datado antes do primeiro acesso aos dados de desempenho; cada análise está identificada como prevista no plano, secundária ou posterior a ele, e o registro das alterações consta do material suplementar.

Exposição. As prevalências municipais de *A. lumbricoides*, *T. trichiura*, ancilostomídeos — relatados no inquérito como grupo, e não por espécie — e *Schistosoma mansoni* foram transcritas do anexo do relatório⁸ e conferidas com os totais publicados, com correspondência exata: 521 municípios e 197.564 escolares. Cada município foi vinculado ao código de sete dígitos do Instituto Brasileiro de Geografia e Estatística (IBGE). Prevalência designa aqui a proporção de escolares com exame positivo, não a intensidade da infecção. O anexo municipal não informa escola, estrato ou efeito de desenho, de modo que a precisão de cada prevalência foi aproximada pelo modelo binomial; eventual agrupamento da amostragem original não pôde ser incorporado.

O plano definiu *T. trichiura* como exposição primária e *A. lumbricoides* como coprimária, por razão farmacológica: a dose única de albendazol das campanhas escolares tem eficácia elevada contra *A. lumbricoides* e reduzida contra *T. trichiura*¹², de modo que a prevalência desta seria menos atenuada por tratamento prévio. Nos modelos contínuos a prevalência entra como logito empírico — $\ln[(k + 0,5)/(n - k + 0,5)]$, para k positivos entre n examinados —, que acomoda os municípios sem nenhum caso, para os quais o logito convencional seria indefinido. Os municípios foram também divididos, pela prevalência, em cinco grupos de igual tamanho, análise prevista como secundária.

Desfecho. O desfecho provém do Sistema de Avaliação da Educação Básica (Saeb) de 2019, 5º ano do ensino fundamental da rede pública¹³, em dois formatos: a proficiência média municipal em Língua Portuguesa e a proporção de estudantes com ao menos 200 pontos na escala. Esse corte é convenção externa, descritiva¹⁴, e não limiar oficial: o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) não define corte de adequação¹⁵. Como a escolha se deu após o primeiro acesso ao desfecho, as análises que dela dependem foram

repetidas com 175 e 225 pontos. A onda de 2019 é a mais próxima do inquérito, mas os estudantes avaliados não são os examinados parasitologicamente: a prevalência é característica do município no período, não exposição de uma coorte.

Covariáveis. Os modelos incorporam quatro blocos, de quatro fontes públicas. Do Sistema IBGE de Recuperação Automática (SIDRA) vêm esgotamento sanitário, água, população residente e proporção rural¹⁶; do TabNet, do Departamento de Informática do Sistema Único de Saúde (DATASUS), renda domiciliar per capita e escolaridade materna¹⁷. Dos Indicadores Educacionais do Inep vêm docentes com curso superior e estudantes por turma¹⁸; e do Sistema de Vigilância Alimentar e Nutricional (Sisvan), o déficit de estatura entre menores de cinco anos¹⁹. O material suplementar dá, para cada variável, a tabela de origem, o ano e os filtros. O ajuste por elas desconta o que elas próprias explicam, permitindo ver quanto da diferença entre municípios subsiste. A variável do Sisvan pede ressalva: descreve a população acompanhada na atenção primária, não uma amostra representativa do município²⁰. Todos os modelos incluem ainda uma indicadora por unidade da federação, que absorve o que é comum a um estado inteiro e faz com que as comparações se deem sobretudo entre municípios do mesmo estado.

Amostra. Dos 521 municípios do inquérito, 517 tinham resultado divulgado em 2019. Destes, 392 haviam examinado 100 ou mais escolares — limite adotado para evitar prevalências calculadas sobre poucos exames — e nenhum apresentava covariável faltante. Dois foram excluídos por serem o único município elegível de sua unidade da federação, o que impede a comparação intraestadual: Brasília e Arenápolis, em Mato Grosso. Restaram 390 municípios de 25 unidades da federação, com 186.308 escolares examinados.

Ponderação dos municípios. Entre as decisões de modelagem, o peso atribuído a cada município é a que mais afeta os resultados. O plano previa ponderação inversamente proporcional à variância, que privilegia os municípios cuja prevalência foi estimada sobre mais exames. Ocorre que o peso se constrói sobre o logito, cuja imprecisão é máxima onde há poucos positivos — condição do próprio grupo de referência: dos 78 municípios do quinto inferior, 66 não registraram nenhum escolar positivo. Sob a regra prevista, esse estrato recebe 0,6% do peso total, contra 48,1% do quinto superior; sem ponderação, cada quinto responde por 20%. Há ainda uma objeção formal: o peso, em regressão ponderada, corrige a variabilidade do desfecho, e ponderar pela precisão da *exposição* não corrige o erro de medida na variável explicativa. Por isso nenhum esquema é apresentado como o correto, e todas as análises centrais são relatadas sob três: a ponderação prevista, a ausência de ponderação e o peso proporcional à raiz do número de escolares examinados.

Os modelos. Estimamos regressões lineares por mínimos quadrados ponderados em sete etapas, M1 a M7, acrescentando um bloco por etapa. Isso permite acompanhar o coeficiente da exposição conforme os fatores socioeconômicos entram: M1 traz apenas as indicadoras de estado; M2 acrescenta esgotamento e água; M3, renda; M4, escolaridade materna — o

estimando principal do plano; M5, insumos escolares; M6, ruralidade e porte populacional; M7, déficit de estatura. Renda e população ingressam em logaritmo. A unidade da federação entra como efeito fixo, e não como intercepto aleatório: os estados formam conjunto pequeno e fechado, o interesse é a comparação intraestadual, e essa escolha dispensa o pressuposto de independência entre o efeito do estado e as características do município.

Margens de erro e testes múltiplos. As margens de erro são erros-padrão robustos à heterocedasticidade, do tipo HC3; como ele não corrige a dependência entre municípios, relatamos também erros-padrão agrupados por unidade da federação e a inferência espacial descrita adiante. A correção de Holm foi aplicada a três conjuntos: os quatro testes previstos no plano — duas espécies coprimárias por dois domínios de desfecho; o conjunto ampliado de oito, definido após o acesso ao desfecho, que acrescenta ancilostomídeos e *S. mansoni*; e os quatro contrastes entre quintos. Empregamos ainda um controle negativo²¹: como não há via pela qual a infecção influencie a formação dos professores, associação entre prevalência e proporção de docentes com curso superior indicaria confundimento residual.

Semelhança entre municípios vizinhos. Municípios próximos assemelham-se tanto na prevalência quanto no desempenho, e ignorar essa dependência estreita indevidamente as margens de erro. Cada município foi localizado pelas coordenadas de sua sede, com a vizinhança definida de três maneiras: os cinco mais próximos, os dez mais próximos e todos a até 200 km. Para cada uma calculamos o índice de Moran na forma apropriada a resíduos de regressão²², cujo valor esperado na ausência de padrão espacial não é zero — é com ele que o índice observado se compara. Constatada a dependência, a inferência foi refeita por erros-padrão que admitem correlação entre vizinhos²³, sob cortes de 150, 300 e 500 km, e por um modelo de erro espacial.

Verificações de robustez. A dependência de poucos municípios foi examinada por alavancagem, DFBETA e distância de Cook e por três exclusões — de um município por vez, dos dez de maior alavancagem e de uma unidade da federação por vez. Em lugar dos quintos, ajustamos uma curva de forma livre: splines cúbicas naturais com quatro graus de liberdade sobre o logaritmo de um mais a prevalência. No desfecho, deslocamos o corte para 175 e 225 pontos. Na amostra, restringimos aos municípios com 200 ou mais exames e, no sentido oposto, relaxamos o limite de 100 até incluir todos os elegíveis, com os quintos redefinidos em cada recorte. Testamos ainda um desempate alternativo na fronteira dos quintos, a retirada e a substituição do bloco de saneamento e interações da exposição com o contexto socioeconômico; o material suplementar traz cada especificação com seus resultados completos. Permanecem não executadas, e constam do registro de desvios, a exposição pelo componente principal, a exclusão de capitais e do estrato endêmico para esquistossomose, as ondas alternativas do Saeb, o valor E e o modelo com a taxa de participação no Saeb como desfecho.

Processamento e verificação. A análise foi conduzida em Python 3.11.15, com pandas 3.0.2, NumPy 2.4.4, statsmodels 0.14.6, patsy 1.0.2, SciPy 1.17.1 e matplotlib 3.10.9; os mesmos scripts, reexecutados em Python 3.12.13, produziram saídas idênticas. As rotinas foram programadas e executadas com apoio de modelos de linguagem, identificados na declaração adiante; o cálculo coube às bibliotecas nomeadas, com versões fixadas e somas de verificação no material suplementar. Conduzimos quatro verificações. Duas conferem os dados: a reconstrução do fluxo amostral e a reconferência de exposições, desfechos, totais agregados e municípios sorteados contra as fontes originais. Duas conferem o cálculo: uma segunda implementação, independente da biblioteca estatística, que recomputa por álgebra linear os coeficientes e as margens de erro do modelo principal, e a recomputação da correção de multiplicidade fora dessa biblioteca. Elas, somadas a uma avaliação externa do pacote, detectaram cinco divergências, todas em rotinas de verificação ou de diagnóstico e nenhuma nos modelos centrais. Três alteraram valores relatados: os intervalos da reamostragem por estado, os coeficientes por espécie sob a ponderação prevista — adiante já na forma corrigida — e o valor de p do diagnóstico de autocorrelação nos erros filtrados do modelo espacial, que consta apenas do material suplementar. Os coeficientes HC3 dos modelos centrais e os do modelo espacial permanecem inalterados, e as cinco estão descritas, uma a uma, no registro de alterações do material suplementar.

Ética. A análise utiliza exclusivamente bases públicas, agregadas e sem identificação individual, razão pela qual está dispensada de registro e apreciação pelo sistema formado pelos Comitês de Ética em Pesquisa e pela Comissão Nacional de Ética em Pesquisa (CEP/CONEP). A dispensa apoia-se nos incisos II, III e V do artigo 26 da Resolução nº 674, de 6 de maio de 2022, do Conselho Nacional de Saúde²⁴ — informações de acesso público, de domínio público e dados exclusivamente agregados —, conforme o Ofício Circular nº 12/2023 da Conep²⁵. O caso é igualmente compatível com a Resolução CNS nº 510, de 2016²⁶. Os procedimentos éticos e de consentimento da coleta original são os do inquérito de origem⁸.

Resultados

Na amostra analítica, 5,97% dos escolares examinados apresentavam infecção por *A. lumbricoides*, 5,51% por *T. trichiura*, 2,67% por ancilostomídeos e 1,43% por *S. mansoni*. A infecção por *A. lumbricoides* foi detectada em 324 dos 390 municípios; em 66 dos 78 municípios do quinto inferior, nenhum escolar examinado apresentou exame positivo.

Quando a prevalência entra como número

Nessa especificação, prevista no plano, o sinal das estimativas é estável; a precisão, não (Figura 1, painel A). Para *A. lumbricoides*, o modelo ajustado apenas pelo estado indica associação negativa e de magnitude apreciável com a proficiência: $-0,207$. Com a entrada de saneamento, renda e escolaridade materna, o coeficiente cai para $-0,104$ e, no modelo completo, para

−0,080, passando o intervalo a incluir o zero — tudo sob a ponderação prevista no plano. Na escala do Saeb, passar de 1% para 10% de prevalência equivale a cerca de 9 a 11 pontos no primeiro modelo e a cerca de 4 no completo. Sob os outros dois esquemas, o mesmo modelo completo produz coeficientes menores em valor absoluto, porém mais precisos: −0,069 sem ponderação e −0,083 sob a raiz, ambos com intervalo que exclui o zero (Figura S2). Para *T. trichiura*, o padrão se repete: −0,036 sob a ponderação prevista, com intervalo largo, contra −0,069 e −0,071 sob as demais. O resultado descreve-se melhor como sensível à especificação do que como nulo.

O plano previa quatro testes principais — duas espécies por dois domínios de desfecho — no modelo ajustado até a escolaridade materna. Nele, sob a ponderação prevista, o coeficiente de *A. lumbricoides* alcança −0,104 em Língua Portuguesa ($p = 0,039$) e −0,113 em Matemática ($p = 0,031$), enquanto *T. trichiura* não se distingue de zero (Tabela 2, painel A). Aplicado o ajuste de Holm, nenhum dos quatro sobrevive: os valores corrigidos são 0,124 para *A. lumbricoides* e 0,711 para *T. trichiura*. Sem ponderação, ou sob a raiz, os quatro resistem, com valores corrigidos entre 0,0001 e 0,002. No conjunto ampliado de oito testes, aplicado ao modelo completo, repete-se o padrão: nenhum sobrevive sob a regra do plano (menor valor corrigido, 0,270), dois sobrevivem sem ponderação e três sob a raiz.

O sinal não é exclusivo de uma espécie. Os quatro parasitos têm coeficientes negativos em todas as especificações, e o que os distingue é a precisão, que acompanha o número de municípios em que cada um foi detectado — 66 municípios sem qualquer caso para *A. lumbricoides*, 98 para *T. trichiura*, 135 para ancilostomídeos e 261 para *S. mansoni*. No modelo completo sem ponderação, os três primeiros são praticamente equivalentes (−0,069, −0,069 e −0,070), com intervalo que exclui o zero; apenas *S. mansoni* se afasta, com −0,023. Sob a ponderação prevista, o único intervalo que exclui o zero é o dos ancilostomídeos (−0,102 em leitura, $p = 0,037$; −0,111 em Matemática, $p = 0,034$), e nenhum resiste ao ajuste da família de oito, que os leva a 0,270.

A atenuação adicional mais pronunciada ocorreu no passo em que a renda foi incorporada; como renda e infraestrutura são correlacionadas e os blocos entram em ordem fixa, isso descreve a ordem adotada e não estabelece hierarquia causal. No desfecho de controle negativo não se detectou associação: a prevalência não prediz a proporção de docentes com curso superior (−2,1 pontos percentuais; $p = 0,365$). Um único controle negativo nulo não exclui confundimento residual; apenas não o revelou.

A comparação entre extremos de prevalência

Nas análises secundárias, que comparam grupos de municípios, a diferença bruta entre os extremos é ampla (Tabela 1). Do quinto de menor para o de maior prevalência, a proficiência média em Língua Portuguesa cai de 212,3 para 186,4 pontos, e a proporção de crianças com 200

pontos ou mais, de 59,8% para 37,3%. Esses grupos, porém, diferem em praticamente tudo o mais: a renda domiciliar per capita mediana cai de R\$ 544 para R\$ 261, a proporção de mães com oito ou mais anos de estudo, de 59,5% para 40,2%, e o déficit de estatura sobe de 10,4% para 15,2%.

Parte da diferença permanece após o ajuste (Tabela 2, painel B; Figura 1, painel B). No modelo completo, a proporção de crianças com 200 pontos ou mais é 9,15 pontos percentuais menor no quinto superior que no inferior sob a ponderação prevista, 7,25 sem ponderação e 7,72 sob a raiz. Nos três esquemas o intervalo exclui o zero e o valor de p corrigido fica abaixo de 0,05. O quarto quinto apresenta diferenças da mesma ordem, entre 5,07 e 9,91 pontos. Os intermediários são menos estáveis: o terceiro separa-se do primeiro apenas sob a ponderação prevista, e o segundo não se separa em esquema algum. A evidência firme concentra-se, portanto, no quarto e no quinto grupos. A reamostragem por estado, em 2.000 réplicas que preservam as indicadores estaduais, acompanha os intervalos originais: de -7,97 a 3,98 pontos no segundo quinto, de -12,46 a -1,65 no terceiro, de -14,00 a -4,77 no quarto e de -14,40 a -4,91 no quinto. O teste de tendência indica queda de 1,71 ponto percentual por quinto sem ponderação ($p = 0,002$) e de 1,51 sob a ponderação prevista, aí sem significância ($p = 0,233$). Os modelos explicam entre 65,0% e 66,2% da variação entre municípios.

Apesar de detectável, essa diferença residual é pequena diante da variação entre municípios, e as duas distribuições se sobrepõem amplamente: o desempenho do quinto de maior prevalência estende-se de 11,8% a 81,9% de crianças com 200 pontos ou mais, e o do quinto de menor, de 14,0% a 84,7%. Sorteando um município de cada extremo, a probabilidade de o de menor prevalência apresentar melhor desempenho é de 0,85. Isoladamente, a prevalência responde por 17% da variação entre municípios. O ganho de informação é igualmente modesto. Partindo de um modelo com apenas os indicadores de estado, a renda eleva a variação explicada em 7,1 pontos percentuais, a escolaridade materna em 4,3, o saneamento em 3,0 e a prevalência em 1,6; presentes todas as covariáveis, o acréscimo da prevalência eleva-a em 0,4 a 1,2 ponto percentual.

O contraste resiste à maior parte das verificações, detalhadas no material suplementar. Deslocado o corte do desfecho para 175 ou 225 pontos, as diferenças ficam entre -6,06 e -8,70 pontos, com p corrigido entre 0,009 e 0,025 nos nove cenários. A restrição aos 230 municípios com 200 ou mais exames preserva a magnitude, entre -8,57 e -9,78 pontos, e o desempate que mantém juntos os municípios de prevalência idêntica na fronteira dos quintos tampouco altera o quadro, entre -7,45 e -10,08. Quanto à influência, 13 municípios excedem o dobro do limiar usual de alavancagem sob a ponderação prevista, com máximo de 0,56, contra 0,35 sem ponderação e 0,49 sob a raiz. O contraste varia entre -8,39 e -7,06 pontos nos 390 reajustes que excluem um município por vez, sem inversão de sinal, e fixa-se em -8,06 ($p = 0,002$) com a

exclusão simultânea dos dez mais influentes. Nas 25 reestimações que excluem uma unidade da federação por vez, varia entre $-8,90$ e $-6,78$, com p abaixo de $0,01$.

Uma verificação, entretanto, destaca-se e é a que mais qualifica o achado. O limite de 100 escolares examinados define quem ingressa na análise, e os 125 municípios elegíveis por ele excluídos não constituem recorte aleatório. Examinaram menos escolares (mediana de 68 contra 250,5 entre os que passaram o limite), têm prevalência mais baixa ($1,0\%$ contra $2,4\%$), renda mais alta e melhor desempenho ($58,2\%$ contra $50,0\%$ de crianças com 200 pontos ou mais). Com sua inclusão, o contraste se atenua, e a atenuação acompanha o peso que lhes é atribuído. Sob a ponderação prevista, que praticamente os desconsidera, a diferença passa de $-9,15$ para $-7,70$ pontos ($p = 0,009$), e sob a raiz, de $-7,72$ para $-5,77$ ($p = 0,015$). Sem ponderação, que lhes dá o mesmo peso dos demais, ela passa de $-7,25$ para $-3,83$, com intervalo que já inclui o zero ($p = 0,070$). O coeficiente contínuo segue o mesmo padrão. Essa ordenação é a que se esperaria de erro de medida na exposição: quanto maior o peso dado a prevalências estimadas sobre poucos exames, mais as estimativas se aproximam do zero. Ela implica que a magnitude relatada acima é específica da amostra restrita, não propriedade do conjunto dos municípios examinados.

A semelhança entre municípios vizinhos

O desempenho escolar é fortemente estruturado no espaço: sem qualquer ajuste, o índice que mede a semelhança entre vizinhos alcança $+0,59$ para a proporção de crianças com 200 pontos ou mais, contra um valor esperado de $-0,003$. O modelo absorve a maior parte dessa estrutura, mas não toda: nos resíduos do modelo completo o índice cai para $+0,085$ sob a raiz, $+0,083$ sem ponderação e $+0,017$ sob a ponderação prevista. São valores pequenos, mas que se afastam de forma detectável do termo de comparação apropriado, em torno de $-0,05$ nesta amostra (p entre $0,018$ e menos de $0,001$), nas três definições de vizinhança. A consequência sobre a inferência é pequena: com erros-padrão que admitem correlação entre vizinhos, o contraste entre extremos mantém-se com p entre $0,001$ e $0,003$ nos três esquemas e nas três distâncias.

Já o modelo que representa explicitamente essa correlação atenua as estimativas entre 15% e 28% — sem ponderação, o contraste passa de $-7,25$ para $-6,14$ pontos ($p = 0,008$) e o coeficiente contínuo, de $-0,069$ para $-0,050$ ($p = 0,053$), no limite da significância convencional. Esse modelo foi reestimado, para conferência, em biblioteca espacial independente, e as duas implementações coincidem até a terceira casa decimal sob a mesma definição de vizinhança. Adotada a convenção alternativa, que não simetriza o grafo de vizinhos, o contraste vai a $-6,42$ e o coeficiente contínuo a $-0,052$ ($p = 0,044$) — razão a mais para descrever este último como situado no limiar, e não de um lado dele.

Análises exploratórias

Substituída a divisão em quintos por uma curva de forma livre, o resultado é compatível com relação não linear, com diferenças mais negativas nas prevalências elevadas e estabilização no alto da escala (Figura S1). Avaliada nas prevalências medianas dos quintos extremos, a curva produz diferença entre -6,5 e -10,5 pontos percentuais, da mesma ordem dos contrastes por quintos, mas o teste conjunto de seus termos alcança significância em apenas um dos três esquemas ($p = 0,148; 0,061; 0,017$). Não relatamos ponto de mudança: o lugar em que um intervalo deixa de cruzar o zero depende do modelo, da ponderação e do tamanho da amostra, e não estima uma inflexão.

As demais verificações não alteraram o quadro. O contraste pouco se move com a retirada do bloco de saneamento ou sua substituição por indicadores de privação de 2010: -7,38 e -7,13 pontos sob a raiz, contra -7,72. Também não difere por macrorregião (p entre 0,75 e 0,99 no teste das interações; contraste entre -6,85 e -9,60 excluindo uma região por vez). E não depende do contexto socioeconômico: nenhuma das interações com esgotamento, água, escolaridade materna e renda se sustenta nos três esquemas, e as duas metades da amostra definidas pela mediana do esgotamento adequado produzem contrastes semelhantes, entre -6,63 e -8,83 pontos. As duas vias intermediárias não se sustentaram: déficit de estatura e fluxo escolar predizem o desempenho, mas a prevalência não prediz nenhum deles após o ajuste.

Tabela 1. Características dos municípios e desempenho escolar segundo quintos de prevalência de *Ascaris lumbricoides*, e diferença ajustada em relação ao primeiro quinto. Amostra analítica, 390 municípios.

	Q1	Q2	Q3	Q4	Q5
Municípios (n)	78	78	78	78	78
Prevalência de <i>A. lumbricoides</i> , mediana (%)	0,00	0,75	2,42	6,50	16,22
Faixa de prevalência (%)	0,00–0,26	0,26–1,42	1,44–4,21	4,28–10,08	10,10–77,38
Escolares examinados por município, mediana (n)	149,5	296,5	475,0	380,5	208,5
Proficiência em Língua Portuguesa, 5º ano, média (pontos)	212,3	204,2	205,4	197,8	186,4
Estudantes com 200 pontos ou mais, média (%)	59,8	52,1	53,3	46,7	37,3
Renda domiciliar per capita, mediana (R\$ de 2010)	544	367	413	324	261
Mães com 8 ou mais anos de estudo, mediana (%)	59,5	56,6	54,4	53,9	40,2
Esgotamento sanitário adequado, mediana (%)	70,6	46,0	66,9	59,4	44,4
População rural, mediana (%)	22,2	36,7	24,0	31,6	34,7
Déficit de estatura entre crianças registradas no Sisvan, mediana (%)	10,4	11,2	11,2	13,0	15,2
Diferença ajustada em relação a Q1, em pontos percentuais de estudantes com 200 pontos ou mais					
Ponderação prevista no plano de análise	ref.	–3,07 (–8,98 a 2,84)	–6,78 (–11,88 a –1,69)	–9,91 (–15,58 a –4,23)	–9,15 (–15,70 a –2,60)
Sem ponderação	ref.	–2,60 (–6,79 a 1,59)	–4,06 (–7,74 a –0,38)	–5,07 (–9,16 a –0,99)	–7,25 (–11,91 a –2,58)
Ponderação pela raiz do número de examinados	ref.	–1,66 (–6,25 a 2,93)	–4,30 (–8,35 a –0,24)	–6,52 (–10,85 a –2,20)	–7,72 (–12,75 a –2,69)

Os quintos foram definidos pelos postos da prevalência; municípios empatados na fronteira podem cair em quintos adjacentes, e por isso 0,26% aparece em Q1 e em Q2. O desempenho escolar traz a média entre municípios; as demais linhas, a mediana, por serem distribuições assimétricas. Nas três últimas linhas, os parênteses são intervalos de confiança de 95%, não ajustados para multiplicidade; na família dos quatro contrastes, o *p* de Holm fica abaixo de 0,05 para Q4 e Q5 nos três esquemas, e para Q3 apenas sob a ponderação prevista. O ajuste inclui unidade da federação, esgotamento sanitário, abastecimento de água, renda, escolaridade materna, insumos escolares, população rural, porte populacional e déficit de estatura.

Tabela 2. Prevalência como variável contínua: coeficiente da exposição sobre a proficiência em Língua Portuguesa do 5º ano, sob três esquemas de ponderação. A unidade é o desvio-padrão da distribuição nacional entre municípios, por unidade de logito de prevalência — a transformação descrita em Métodos.

	Ponderação prevista no plano	Sem ponderação	Raiz do número de examinados
Estimando principal do plano — modelo ajustado até escolaridade materna (M4)			
<i>A. lumbricoides</i>	-0,104 (-0,204 a -0,005)	-0,086 (-0,135 a -0,038)	-0,100 (-0,152 a -0,048)
<i>T. trichiura</i>	-0,056 (-0,194 a 0,081)	-0,080 (-0,126 a -0,035)	-0,082 (-0,133 a -0,031)
Testes com p de Holm < 0,05 na família original de quatro	nenhum	os quatro	os quatro
Modelo completo (M7)			
<i>A. lumbricoides</i>	-0,080 (-0,201 a 0,041)	-0,069 (-0,124 a -0,013)	-0,083 (-0,142 a -0,024)
<i>T. trichiura</i>	-0,036 (-0,204 a 0,133)	-0,069 (-0,121 a -0,017)	-0,071 (-0,127 a -0,016)
Testes com p de Holm < 0,05 na família ampliada de oito	nenhum	dois de oito	três de oito

Entre parênteses, intervalos de confiança de 95% com erros-padrão HC3, não ajustados para multiplicidade. A família confirmatória do plano tem quatro testes — duas exposições coprimárias por dois domínios de despecho, Língua Portuguesa e Matemática; a tabela mostra só os de Língua Portuguesa, e a linha de resumo considera os quatro. A família ampliada de oito, decidida após o acesso ao despecho, acrescenta ancilostomídeos e *S. mansoni*. O M4 ajusta por unidade da federação, saneamento, água, renda e escolaridade materna; o M7 acrescenta insumos escolares, população rural, porte populacional e déficit de estatura. A sequência completa está na Figura S2.

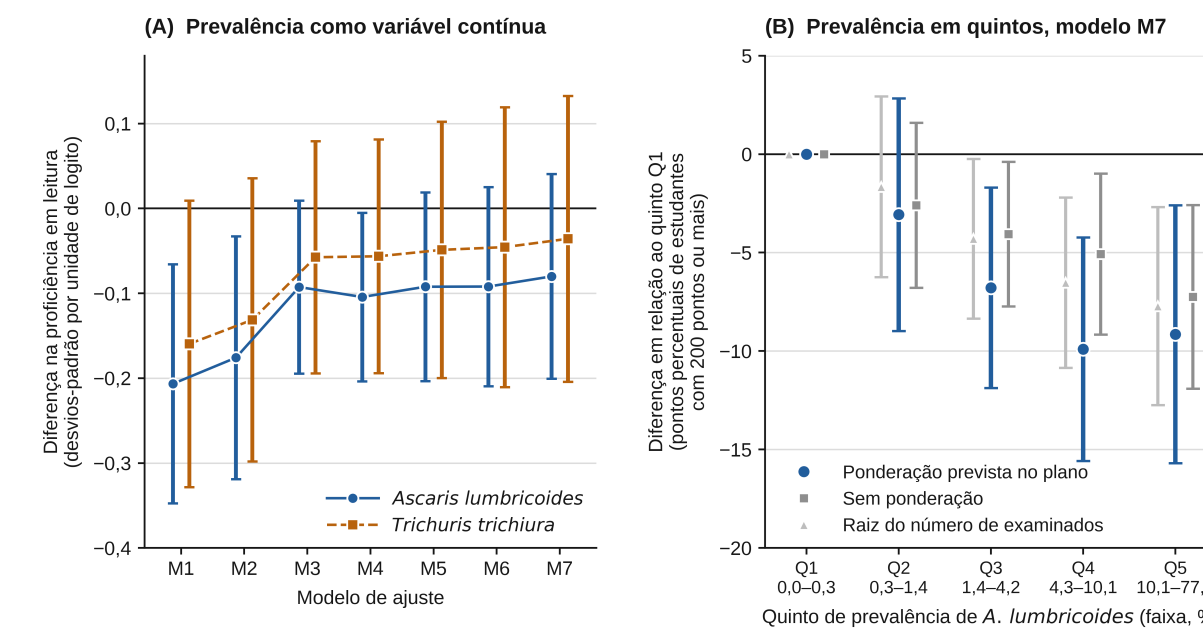


Figura 1. Diferença estimada no desempenho em leitura segundo a prevalência municipal de vermes intestinais. **Painel A:** a prevalência entra como variável contínua na forma prevista no plano (logito), e o despecho é a proficiência média em Língua Portuguesa do 5º ano, em desvios-padrão da distribuição nacional entre municípios; cada ponto é o coeficiente de uma espécie em um dos sete modelos, e as barras são intervalos de confiança de 95%. A sequência acrescenta um bloco de covariáveis por vez:

M1, apenas unidade da federação; M2, mais saneamento e água; M3, mais renda; M4, mais escolaridade materna — o estimando principal do plano; M5, mais insumos escolares; M6, mais população rural e porte populacional; M7, mais déficit de estatura. Usa a ponderação prevista; a mesma sequência sob os três esquemas está na Figura S2. **Painel B**, análise secundária: cada ponto é a diferença estimada no modelo completo M7 entre o quinto indicado e o primeiro, em pontos percentuais de estudantes com 200 pontos ou mais, com as faixas de prevalência sob cada rótulo e as medianas na Tabela 1. Em azul, a ponderação do plano; em cinza, as duas alternativas. O eixo é categórico e os pontos não são ligados por linhas, porque a distância entre quintos no gráfico não corresponde a distância na escala de prevalência. A relação estimada sem cortes consta da Figura S1.

As legendas completas das Figuras S1 (forma da relação estimada sem cortes) e S2 (sensibilidade da especificação contínua à ponderação) acompanham o material suplementar.

Discussão

O gradiente de aprendizagem observado entre municípios com maior e menor prevalência de helmintíases reflete, em sua maior parte, as desigualdades estruturais que acompanham essas infecções, e não o parasitismo isoladamente. Entre três quintos e quatro quintos da diferença bruta são contabilizados por renda, escolaridade materna, saneamento e características das escolas. O que remanesce é pequeno — de quatro a nove pontos percentuais, conforme a amostra e a ponderação. A direção é negativa em todas as especificações, sem uma única inversão, e o contraste sobrevive a mudanças no corte do desfecho, à exclusão dos municípios mais influentes, à exclusão de qualquer estado e à correção para a semelhança entre vizinhos. Ainda assim, a magnitude depende de quem ingressa na amostra, e o ganho de informação sobre um modelo que já contém os indicadores socioeconômicos de rotina é de fração de ponto percentual.

Um segundo aspecto, de natureza instável, impede interpretação mais forte. Na especificação contínua o sinal permanece negativo, mas o intervalo inclui ou exclui o zero conforme a ponderação, e nenhum dos quatro testes previstos no plano sobreviveu ao ajuste para comparações múltiplas sob a regra de peso prevista. A razão é a descrita nos Métodos: essa regra reduz a menos de 1% do peso justamente o grupo de referência — os municípios sem infecção detectada, que são também os de melhor desempenho —, quase suprimindo o termo de comparação. Não se trata de uma especificação correta e outras equivocadas, pois a pergunta se altera ligeiramente em cada uma; relatar as três lado a lado parece-nos preferível a eleger retrospectivamente a mais favorável.

O achado dialoga com o debate sobre a erradicação da ancilostomíase no sul dos Estados Unidos^{5,6}: aqui, como ali, o desenho agregado sustenta a descrição de um gradiente sem decidir a questão causal.

Três aspectos limitam a interpretação da diferença residual. O primeiro é que ela se concentra em análises secundárias e posteriores ao plano. O segundo é que sua magnitude depende do limite de exames adotado para ingresso na amostra, reduzindo-se quando municípios com

prevalências estimadas sobre poucos exames entram com peso igual — comportamento compatível com erro de medida na exposição, que impede tratar os quatro a nove pontos como valor geral. O terceiro é que nenhuma das duas vias intermediárias examinadas se sustentou: déficit de estatura e fluxo escolar predizem o desempenho, mas nenhum deles é predito pela prevalência após o ajuste. No caso do crescimento, o achado é compatível com a literatura que atribui o déficit de estatura a um conjunto amplo de exposições ambientais, e não a um agente isolado²⁷. Anemia e inflamação crônica, as demais vias plausíveis, não puderam ser examinadas por ausência de indicador municipal público.

A escolha de *A. lumbricoides* para as análises por quintos não deve ser lida como especificidade de espécie. O que varia entre os quatro parasitos é a precisão, que acompanha o número de municípios em que cada um foi detectado. Mais municípios no piso da escala significam mais ruído de medida, e ruído dessa natureza desloca as estimativas na direção do zero — comportamento coerente com o Kato-Katz, que deixa passar sobretudo as infecções leves e cuja sensibilidade varia entre espécies²⁸. A ordenação observada é a que esse mecanismo prevê, sem que seja necessário invocar diferença biológica entre as espécies — embora seleção e composição da amostra possam contribuir para o mesmo padrão.

Convém ainda distinguir prevalência de intensidade de infecção. Este estudo mede a proporção de escolares com exame positivo, enquanto a intensidade se refere à carga parasitária — nível em que a revisão com dados individuais discute eventuais ganhos⁴. Caso exista alguma relação com o desempenho, os dados sugerem que ela se concentre em municípios de maior prevalência, sem que se possa inferir dependência da intensidade da infecção em cada criança. A distinção é também programática. Os critérios da Organização Mundial da Saúde para validar a eliminação das geo-helmintíases combinam prevalência de qualquer geo-helminto abaixo de 20% com infecções moderadas ou intensas abaixo de 2%²⁹; na América Latina, os dois indicadores são mapeados separadamente, e os de intensidade são bem mais escassos³⁰.

Daí decorrem as consequências práticas. A indicação de vermifugar permanece governada pelos critérios sanitários vigentes — tratamento anual onde a prevalência basal de qualquer geo-helminto atinge 20% e semestral acima de 50%³¹ —, e não por desempenho escolar. Nenhum valor de prevalência aqui apresentado deve servir como limiar de decisão: os quintos e a curva foram definidos dentro desta amostra e após o exame dos desfechos. E, embora municípios de alta prevalência mereçam prioridade em saneamento e em programas sociais, para a focalização de políticas educacionais os indicadores socioeconômicos já consolidados parecem suficientes.

As limitações decorrem do desenho e não se resolvem com o acréscimo de covariáveis. A mais importante delimita o alcance de tudo o que se afirmou. Os dados permitem observar um panorama coletivo — municípios com alta prevalência têm, em média, menor proporção de

crianças com 200 pontos ou mais —, mas não autorizam afirmar que uma criança infectada leia pior do que um colega não infectado da mesma escola.

Quanto às demais, a amostra do inquérito privilegiou áreas endêmicas de esquistossomose e o limite de 100 exames a estreita ainda mais. A seleção não aleatória dos municípios, a exigência de resultado divulgado no Saeb e esse limite restringem a generalização e podem também afetar a própria associação estimada, como indica a atenuação observada quando o limite é relaxado. Há defasagem temporal: a prevalência foi medida entre 2010 e 2015 e descreve a transmissão do município naquele período, não a exposição individual das crianças avaliadas em 2019. Nesse intervalo, obras de saneamento, campanhas de vermifugação e surtos locais podem tê-la alterado; se de forma independente do desempenho, o efeito esperado é aproximar as estimativas do zero, e se concentrada nos municípios de maior prevalência, o viés pode assumir qualquer direção. Não identificamos série municipal pública sobre a cobertura das campanhas de vermifugação que permitisse tratá-la como covariável, e a taxa municipal de participação no Saeb não foi incorporada à base desta versão, permanecendo entre as análises previstas e não executadas. Os indicadores de saneamento provêm do Censo de 2022, posterior à exposição e ao desfecho, embora sua retirada ou substituição por indicadores de 2010 não altere os resultados. A dependência espacial foi medida e tratada, mas o modelo que a representa deixa o coeficiente contínuo no limite da significância convencional. Por fim, ainda que todas as especificações estejam relatadas, o número delas recomenda cautela na leitura de qualquer estimativa isolada.

A conclusão que o conjunto sustenta diz respeito, portanto, mais ao que a prevalência mede do que ao que ela causa. A diferença bruta de desempenho entre municípios de alta e baixa prevalência de *A. lumbricoides* é grande, e em sua maior parte corresponde à diferença socioeconômica entre esses municípios, mensurada por variáveis que o país já coleta rotineiramente. Remanesce uma parcela que o ajuste não absorve, negativa em todas as especificações, porém pequena, sensível a decisões de amostra e ponderação, e de valor informativo modesto. Decidir se nela há algo além de um marcador exigirá dados individuais e desenho voltado à pergunta causal, que permita comparar crianças infectadas e não infectadas na mesma escola. Enquanto isso, o mapa municipal da infecção segue desigual: um estudo no Piauí descreve persistência focal dessas infecções em áreas empobrecidas, mesmo onde há vermifugação em massa³². E os dados do inquérito, disponibilizados com este trabalho, ainda foram pouco explorados para outras perguntas situadas entre saúde e educação.

Declaração sobre o uso de inteligência artificial

Ferramentas de inteligência artificial foram empregadas em todas as etapas, declaradas conforme o ICMJE³³. Os autores definiram as perguntas e aprovaram o protocolo e as decisões

analíticas; Maxwell Adriano Abegg obteve os dados de entrada e conferiu-os contra os registros oficiais. As rotinas foram programadas e executadas por modelos de linguagem — sobretudo Claude Opus 5 e ChatGPT 5.6 —, em bibliotecas estatísticas correntes, e delas vieram as figuras; os mesmos modelos apoiaram a redação, as citações e a conferência das inferências. O mapeamento de referências recorreu ao Asta, do Allen Institute for AI, e ao Consensus, com cada uma conferida na fonte original; versões, datas e finalidades constam de `FERRAMENTAS\_IA.md`. Os autores revisaram criticamente o manuscrito e as saídas analíticas e assumem responsabilidade integridade e originalidade do trabalho.

Três verificações permanecem por fazer: revisão das escolhas de modelagem — especificação, ponderação, erros-padrão e multiplicidade — por estatístico independente; replicação em outro ambiente, como R ou Stata, pois só o modelo espacial foi reestimado em biblioteca de terceiros; e reextração dos dados. Que vários modelos de linguagem tenham conferido os resultados não supre nenhuma delas: são ferramentas de mesma natureza, sujeitas a modos de falha comuns. O material suplementar, depositado no Zenodo sob o DOI 10.5281/zenodo.22664954, traz dados, código, protocolo datado, registros de verificação e somas de conferência, e convidamos grupos com experiência em regressão ponderada e inferência espacial a reproduzir a análise e reportar divergências.

Declarações

Financiamento. Este estudo não recebeu financiamento específico de agências públicas, comerciais ou sem fins lucrativos.

Conflitos de interesse. Os autores declaram não haver conflitos de interesse.

Contribuições dos autores, segundo a taxonomia CRediT (*Contributor Roles Taxonomy*). Kibelle Iamã dos Santos Costa e Maxwell Adriano Abegg: conceituação, metodologia, redação do rascunho original e redação — revisão e edição. Maxwell Adriano Abegg: curadoria de dados, investigação e supervisão. Os papéis de software, validação, análise formal e visualização não são reivindicados por nenhum dos autores: a programação, a execução das rotinas e a geração das figuras foram feitas com apoio de modelos de linguagem, conforme a declaração acima.

Autor correspondente. Maxwell Adriano Abegg, Instituto de Ciências Exatas e Tecnologia, Universidade Federal do Amazonas (ICET/UFAM), Itacoatiara, Amazonas, Brasil. E-mail: maxabegg@gmail.com. ORCID: <https://orcid.org/0000-0002-0328-1122>.

Figuras. A Figura 1, em PNG, PDF e TIFF de 600 dpi e as Figuras S1 e S2 integram o material suplementar.

Relato. O manuscrito segue a lista de verificação STROBE (*Strengthening the Reporting of Observational Studies in Epidemiology*) para estudos observacionais, adaptada ao desenho ecológico³⁴; a lista preenchida integra o material suplementar.

Declaração de disponibilidade de dados

Todo o conjunto de dados que dá suporte aos resultados deste estudo está disponível publicamente no repositório Zenodo, sob o identificador digital 10.5281/zenodo.22664954, e pode ser acessado em <https://doi.org/10.5281/zenodo.22664954>. O depósito reúne as bases municipais derivadas, a amostra analítica dos 390 municípios, os sete scripts que reproduzem todos os resultados relatados no texto, o plano de análise datado com o registro de suas alterações, os registros de verificação e as somas de verificação SHA-256. Os dados primários que deram origem a essas bases são de acesso público e estão identificados nos Métodos: o Inquérito Nacional de Prevalência da Esquistossomose mansoni e Geo-helminthoses, o Sistema de Avaliação da Educação Básica e os indicadores educacionais do Inep, o IBGE e o DATASUS. Nenhum dado individual identificável foi utilizado.

Referências

1. Pabalan N, Singian E, Tabangay L, Jarjanazi H, Boivin MJ, Ezeamama AE. Soil-transmitted helminth infection, loss of education and cognitive impairment in school-aged children: a systematic review and meta-analysis. *PLoS Negl Trop Dis*. 2018;12(1):e0005523. doi:10.1371/journal.pntd.0005523
2. Adu AM, Simpong DL, Donkor DM, Manu KS, Antwi MH, Kpelle L, et al. Parasitic infections and associated cognitive outcome among school-aged children in Africa: a systematic review. *Brain Behav*. 2026;16(4):e71326. doi:10.1002/brb3.71326
3. Taylor-Robinson DC, Maayan N, Donegan S, Chaplin M, Garner P. Public health deworming programmes for soil-transmitted helminths in children living in endemic areas. *Cochrane Database Syst Rev*. 2019;(9):CD000371. doi:10.1002/14651858.CD000371.pub7
4. Welch VA, Ghogomu E, Hossain A, Riddle A, Gaffey M, Arora P, et al. Mass deworming for improving health and cognition of children in endemic helminth areas: a systematic review and individual participant data network meta-analysis. *Campbell Syst Rev*. 2019;15(4):e1058. doi:10.1002/cl2.1058
5. Bleakley H. Disease and development: evidence from hookworm eradication in the American South. *Q J Econ*. 2007;122(1):73-117.
6. Roodman D. The impacts of hookworm eradication in the American South: a replication study of Bleakley (The Quarterly Journal of Economics, 2007). *J Comments Replications Econ*. 2018;2(2018-3):1-45. doi:10.18718/81781.7
7. Jardim-Botelho A, Raff S, De Ávila Rodrigues R, Hoffman HJ, Diemert DJ, Corrêa-Oliveira R, et al. Hookworm, *Ascaris lumbricoides* infection and polyparasitism associated with poor cognitive performance in Brazilian schoolchildren. *Trop Med Int Health*. 2008;13(8):994-1004. doi:10.1111/j.1365-3156.2008.02103.x
8. Katz N. Inquérito Nacional de Prevalência da Esquistossomose mansoni e Geo-helminthoses. Belo Horizonte: Instituto René Rachou/Fiocruz; 2018. 76 p. Disponível em: <https://arca.fiocruz.br/handle/icict/25662>
9. Mingoti Poague KIH, Mingoti SA, Heller L. Association between water and sanitation and soil-transmitted helminthiasis: analysis of the Brazilian National Survey of Prevalence (2011–2015). *Arch Public Health*. 2021;79:83. doi:10.1186/s13690-021-00602-7
10. Santos MCS, Oliveira GL, Mingoti SA, Heller L. Effect of environmental factors in reducing the prevalence of schistosomiasis in schoolchildren: an analysis of three

- extensive national prevalence surveys in Brazil (1950–2018). *PLoS Negl Trop Dis*. 2023;17(7):e0010804. doi:10.1371/journal.pntd.0010804
11. Chammartin F, Guimarães LH, Scholte RGC, Bavia ME, Utzinger J, Vounatsou P. Spatio-temporal distribution of soil-transmitted helminth infections in Brazil. *Parasit Vectors*. 2014;7:440. doi:10.1186/1756-3305-7-440
12. Adegnika AA, Zinsou JF, Issifou S, Ateba-Ngoa U, Kassa RF, Feugap EN, et al. Randomized, controlled, assessor-blind clinical trial to assess the efficacy of single-versus repeated-dose albendazole to treat *Ascaris lumbricoides*, *Trichuris trichiura*, and hookworm infection. *Antimicrob Agents Chemother*. 2014;58(5):2535-2540. doi:10.1128/AAC.01317-13
13. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Sistema de Avaliação da Educação Básica (Saeb) 2019: resultados. Brasília: Inep; 2020. Disponível em: <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/saeb/resultados>
14. QEDu. Aprendizado adequado: como o QEDu classifica os níveis de proficiência do Saeb. [s.l.]: QEDu. Disponível em: <https://conteudos.qedu.org.br/academia/aprendizado-adequado/>
15. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Perguntas frequentes: Sistema de Avaliação da Educação Básica. Brasília: Inep. Disponível em: <https://www.gov.br/inep/pt-br/aceso-a-informacao/perguntas-frequentes-1/sistema-de-avaliacao-da-educacao-basica-saeb>
16. Instituto Brasileiro de Geografia e Estatística. Sistema IBGE de Recuperação Automática — SIDRA. Rio de Janeiro: IBGE. Tabela 6805, Domicílios particulares permanentes ocupados, por tipo de esgotamento sanitário (Censo Demográfico 2022); tabela 6803, Domicílios particulares permanentes ocupados, por existência de ligação à rede geral de distribuição de água e principal forma de abastecimento de água (Censo Demográfico 2022); tabela 202, População residente, por sexo e situação do domicílio (Censo Demográfico 2010). Disponível em: <https://sidra.ibge.gov.br>
17. Brasil. Ministério da Saúde. Departamento de Informática do SUS. TabNet — informações de saúde. Brasília: DATASUS. Painéis «Trabalho e renda — Censos 1991, 2000 e 2010», «Educação — Censos 1991, 2000 e 2010» e «Nascidos Vivos — desde 1994». Disponível em: <https://datasus.saude.gov.br>
18. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Indicadores educacionais. Brasília: Inep. Disponível em: <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/indicadores-educacionais>
19. Brasil. Ministério da Saúde. Sistema de Vigilância Alimentar e Nutricional (Sisvan): relatórios públicos de estado nutricional. Brasília: Ministério da Saúde. Disponível em: <https://sisaps.saude.gov.br/sisvan/relatoriopublico>
20. Silva NJ, Silva JFM, Carrilho TRB, Pinto EJ, Andrade RCS, Silva SA, et al. Qualidade dos dados antropométricos infantis do Sisvan, Brasil, 2008-2017. *Rev Saúde Pública*. 2023;57:62. doi:10.11606/s1518-8787.2023057004655
21. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology*. 2010;21(3):383-388. doi:10.1097/EDE.0b013e3181d61eeb
22. Cliff AD, Ord JK. Spatial processes: models and applications. London: Pion; 1981. 266 p.
23. Conley TG. GMM estimation with cross sectional dependence. *J Econom*. 1999;92(1):1-45. doi:10.1016/S0304-4076(98)00084-0
24. Brasil. Conselho Nacional de Saúde. Resolução nº 674, de 6 de maio de 2022. Dispõe sobre a tipificação da pesquisa e a tramitação dos protocolos de pesquisa no Sistema CEP/Conep. Diário Oficial da União, Brasília, 2022. Disponível em: <https://www.gov.br/conselho-nacional-de-saude/pt-br/aceso-a-informacao/atos-normativos/resolucoes/2022/resolucao-no-674.pdf/view>
25. Brasil. Comissão Nacional de Ética em Pesquisa. Ofício Circular nº 12/2023/CONEP/SECNS/DGIP/SE/MS, de 27 de julho de 2023. Orientações para a implementação do artigo 26 da Resolução CNS nº 674, de 6 de maio de 2022. Brasília: Conep; 2023.

26. Brasil. Conselho Nacional de Saúde. Resolução nº 510, de 7 de abril de 2016. Diário Oficial da União, Brasília, 24 maio 2016, seção 1, p. 44-46. Disponível em: <https://conselho.saude.gov.br/resolucoes/2016/Reso510.pdf>
27. Harper KM, Mutasa M, Prendergast AJ, Humphrey J, Manges AR. Environmental enteric dysfunction pathways and child stunting: a systematic review. *PLoS Negl Trop Dis*. 2018;12(1):e0006205. doi:10.1371/journal.pntd.0006205
28. Barda B, Cajal P, Villagran E, Cimino R, Juarez M, Krolewiecki A, et al. Mini-FLOTAC, Kato-Katz and McMaster: three methods, one goal; highlights from north Argentina. *Parasit Vectors*. 2014;7:271. doi:10.1186/1756-3305-7-271
29. World Health Organization. Validation of elimination of soil-transmitted helminthiases as a public health problem in children: criteria and procedure. Geneva: WHO; 2026. Disponível em: <https://www.who.int/publications/b/82528>
30. Saboyá MI, Catalá L, Nicholls RS, Ault SK. Update on the mapping of prevalence and intensity of infection for soil-transmitted helminth infections in Latin America and the Caribbean: a call for action. *PLoS Negl Trop Dis*. 2013;7(9):e2419. doi:10.1371/journal.pntd.0002419
31. World Health Organization. Guideline: preventive chemotherapy to control soil-transmitted helminth infections in at-risk population groups. Geneva: WHO; 2017. ISBN 978-92-4-155011-6. Disponível em: <https://www.who.int/publications/i/item/9789241550116>
32. Monteiro KJL, Reis ERC, Nunes BC, Jaeger LH, Calegar DA, Santos JP, et al. Focal persistence of soil-transmitted helminthiases in impoverished areas in the State of Piauí, Northeastern Brazil. *Rev Inst Med Trop São Paulo*. 2018;60:e24. doi:10.1590/s1678-9946201860024
33. International Committee of Medical Journal Editors. Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals. Updated January 2026. Disponível em: <https://www.icmje.org/recommendations/>
34. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet*. 2007;370(9596):1453-1457. doi:10.1016/S0140-6736(07)61602-X

Este preprint foi submetido sob as seguintes condições:

- Os autores declaram que os necessários Termos de Consentimento Livre e Esclarecido de participantes ou pacientes na pesquisa foram obtidos e estão descritos no manuscrito, quando aplicável.
- Os autores declaram que a elaboração do manuscrito seguiu as normas éticas de comunicação científica.
- Os autores declaram que estão cientes que são os únicos responsáveis pelo conteúdo do preprint e que o depósito no SciELO Preprints não significa nenhum compromisso de parte do SciELO, exceto sua preservação e disseminação.
- Os autores declaram que os dados, aplicativos e outros conteúdos subjacentes ao manuscrito estão referenciados.
- O manuscrito depositado está no formato PDF.
- Os autores declaram que a pesquisa que deu origem ao manuscrito seguiu as boas práticas éticas e que as necessárias aprovações de comitês de ética de pesquisa, quando aplicável, estão descritas no manuscrito.
- Os autores declaram que uma vez que um manuscrito é postado no servidor SciELO Preprints, o mesmo só poderá ser retirado mediante pedido à Secretaria Editorial do SciELO Preprints, que afixará um aviso de retratação no seu lugar.
- Os autores concordam que o manuscrito aprovado será disponibilizado sob licença [Creative Commons CC-BY](#).
- O autor submissor declara que as contribuições de todos os autores e declaração de conflito de interesses estão incluídas de maneira explícita e em seções específicas do manuscrito.
- Os autores declaram que o manuscrito não foi depositado e/ou disponibilizado previamente em outro servidor de preprints ou publicado em um periódico.
- Caso o manuscrito esteja em processo de avaliação ou sendo preparado para publicação mas ainda não publicado por um periódico, os autores declaram que receberam autorização do periódico para realizar este depósito.
- O autor submissor declara que todos os autores do manuscrito concordam com a submissão ao SciELO Preprints.