

Efeitos da exposição aos fatores de risco comportamentais à saúde sobre o atraso escolar no Brasil

Aléssio Tony Cavalcanti de Almeida
Programa de Pós-Graduação em Economia
Universidade Federal da Paraíba

Ignácio Tavares de Araújo Júnior
Programa de Pós-Graduação em Economia
Universidade Federal da Paraíba

Resumo

O objetivo deste trabalho é avaliar o papel da exposição de alunos aos fatores de risco comportamentais à saúde – cigarro, bebida alcoólica e excesso de peso – no resultado educacional no Brasil. Para tanto, os microdados da Pesquisa Nacional de Saúde do Escolar (PENSE) 2012, do Instituto Brasileiro de Geografia e Estatística (IBGE), para estudantes no 9º ano do ensino fundamental e técnicas paramétricas e não-paramétricas são utilizados para a estimação do efeito da exposição a esses fatores sobre o indicador de atraso escolar, medido pela distorção idade-série. Com base em uma amostra final de 67 mil discentes, os resultados centrais da pesquisa confirmam a hipótese de que a exposição aos fatores de risco se relaciona de forma direta com o atraso escolar, sendo os efeitos mais intensos para os estudantes com menor nível socioeconômico. Esses achados ratificam a importância de políticas públicas que promovam a prevenção desses fatores de risco desde a infância, pois as consequências dessas exposições geram desdobramentos não apenas na saúde mas também no componente educacional do capital humano, interferindo a trajetória socioeconômica de longo prazo.

Palavras-chave: Fatores de risco, cigarro, bebida alcoólica, sobrepeso, resultado escolar.

Abstract

This study aims to evaluate the role of exposure of students to behavioral risk factors to health– smoking, alcohol and overweight – in educational attainment in Brazil. We use microdata from the National Survey of School Health (PENSE) 2012, provided by Brazilian Institute of Geography and Statistics (IBGE), for students in the 9th grade of elementary school and parametric techniques and nonparametric to estimate the effect of exposure to these factors in the indicator of delay in school progression, measured by age-grade distortion. Based on a final sample of 67,000 students, the central results confirm the hypothesis that exposure to risk factors has direct effect on delay in school progression. Furthermore, these effects are more intense for students with lower socioeconomic level. Then, the findings of this study ratify the importance of public policies that promote prevention of these risk factors among children, once the exposure to risk factors to health generates repercussions not only in health but also in the educational component of human capital, impacting the long-term socioeconomic development.

Keywords: Risk factors, smoking, alcoholic drink, overweight, educational attainment.

JEL classification: I12, I21, J13, J24.

1 Introdução

O capital humano é posto como um elemento essencial para o crescimento e o desenvolvimento socioeconômico de um país. Devido à importância da educação no processo de formação do capital humano, existem pesquisas específicas para a compreensão dos fatores associados ao desempenho escolar dos indivíduos. Desde meados do século XX, em especial a partir do relatório de Coleman *et al.* (1966), que essa temática vem sendo objeto de interesse crescente para a área acadêmica e para os *policy makers*. Para o Brasil, a grande parte das pesquisas sobre os determinantes de indicadores educacionais examinam o papel das condições socioeconômicas do discente, das características dos insumos escolares e dos efeitos de programas de transferência de renda (GOMES-NETO; HANUSHEK, 1994; BARROS *et al.*, 2001; ALBERNAZ *et al.*, 2002; RIOS-NETO *et al.*, 2002; SOARES, 2003; MACHADO; GONZAGA, 2007; MACHADO, 2008; RIANI; RIOS-NETO, 2008; RIBEIRO; CACCIAMALI, 2012).

Mais recentemente, a literatura internacional desta área vem destacando a necessidade de examinar os efeitos da exposição aos fatores de risco a saúde dos indivíduos nos segmentos jovens da população e as consequências atreladas aos *outcomes* na escola (JAMISON, 1986; GLEWWE; JACOBY, 1995; CROSNOE; MULLER, 2004; KENKEL *et al.*, 2006; DING *et al.*, 2009; EIDE *et al.*, 2010; FLORIN *et al.*, 2011; PONZO, 2013). Como ressalta Mullahy & Sindelar (1996), Levine *et al.* (1997), Chaloupka & Warner (1999), Kenkel & Wang (1999), Crosnoe & Muller (2004) e Florin *et al.* (2011), as condições de saúde dos indivíduos, decorrentes das posturas de risco mencionadas, possuem impactos diretos no funcionamento individual, com correspondência em fatores: físicos, como limitações funcionais, problemas cardiovasculares e respiratórios, e maiores incidências de diabetes ou de outros agravos à saúde; e, não-físicos, com ênfase na baixa concentração, maior absentismo e problemas psicológicos. Consequentemente, a criança exposta ao excesso de peso, ao uso de cigarro e ao consumo de bebida alcoólica tem menor rendimento na escola.

É possível identificar trabalhos aplicados para dados nacionais, como Gomes-Neto *et al.* (1997) e Machado (2008), que relacionam a saúde dos indivíduos, por meio de indicadores antropométricos, e resultados escolares. Contudo, ainda são escassos estudos no país que verificam, mais precisamente, o papel de hábitos não saudáveis na criança e a repercussão na escola. A preocupação com os chamados fatores de risco modificáveis, faz parte inclusive de um recente plano de ações estratégicas, lançado pelo governo brasileiro, para o enfrentamento das Doenças Crônicas Não-Transmissíveis (DCNT), com metas de redução do crescimento das DCNT no período de 2011 a 2022 (BRASIL, 2011).

A Pesquisa Nacional de Saúde do Escolar (PENSE) de 2012, para alunos do último ano do ensino fundamental no Brasil, mostra que 67% dos estudantes já consumiram bebidas alcoólicas, 20% foram expostos ao uso de cigarro e cerca de 40% dos discentes passam no mínimo 5 horas por dia (exclusive feriados e fins de semana) sentado – assistindo televisão, no computador ou jogando videogame. Esses números revelam que uma quantidade elevada de indivíduos no país adotou ou adota posturas de risco em idade escolar, sendo relevante identificar as implicações dessas exposições. Assim, este trabalho pretende avançar nas discussões dos determinantes do desempenho educacional no Brasil ao avaliar o efeito, em especial, de comportamentos não saudáveis, medidos pela exposição ao cigarro, à bebida alcoólica e aos problemas de sedentarismo e má alimentação (captado pelo excesso de peso), sobre a distorção idade-série de estudantes no 9º ano do ensino fundamental. Esta abordagem possibilita a integração das literaturas de desenvolvimento do capital humano, economia da educação e da saúde, uma vez que essa fase de vida serve de base para as trajetórias econômicas, sociais e de saúde dos indivíduos no longo prazo (CONTI *et al.*, 2010).

Os resultados principais da pesquisa são desenvolvidos através de modelos paramétricos (modelo de resposta qualitativa) e não-paramétricos (*propensity score matching* - PSM), com o objetivo de calcular o efeito da exposição de cada uma das posturas não saudáveis no indicador de atraso escolar. Em relação ao PSM, os efeitos médios são desenvolvidos por meio da comparação dos resultados do grupo exposto em relação ao auferido pelo grupo de controle com idênticas características observáveis. Como salientam Gastwirth *et al.* (1998), Rosenbaum (2002), Wakefield (2003) e Rosenbaum (2010), esta abordagem não exige a influência dos fatores não-observados nos resultados, sendo importante verificar a sensibilidade deles a presença de variáveis omitidas, já que os estimadores do PSM podem ser tendenciosos quando a variável indicadora da exposição ao tratamento é endógena. Sobre esse ponto, o presente trabalho usa a estratégia de

estimação de DiPrete & Gangl (2004), com base nos limites de Rosenbaum, para analisar até que ponto as conclusões do estudo podem se alterar em resposta da não inclusão de características não-observadas dos indivíduos.

A amostra final desta pesquisa é composta por cerca de 70 mil estudantes, informações extraídas dos microdados da PENSE 2012 para discentes do 9º ano do ensino fundamental de escolas públicas e particulares do Brasil, com representatividade amostral para todo o território nacional. Segundo o IBGE (2013), a PENSE foi elaborada a partir de convênio entre o IBGE e o Ministério da Saúde, tendo o suporte do Ministério da Educação.

Este artigo está dividido em cinco partes, o que inclui esta introdução. A Seção 2 e a Seção 3 tratam, respectivamente, da estratégia empírica e das informações sobre a base de dados e tratamento das variáveis. A Seção 4 mostram os resultados da pesquisa. Por fim, as conclusões são apresentadas na Seção 5.

2 Estratégia empírica

A hipótese suscitada por esta pesquisa diz respeito ao papel de comportamentos de risco à saúde na infância e na primeira adolescência sobre o desempenho escolar. A adoção de hábitos não saudáveis tendem a afetar a concentração nos estudos e a menor participação em sala de aula, bem como os alunos são mais susceptíveis a faltarem aulas devido à baixa imunidade e outras alterações no metabolismo que dificultam o desenvolvimento da criança e geram mais probabilidades de incidência de doenças (CROSNOE; MULLER, 2004; REILLY; KELLY, 2011).

Park & Kang (2008) e Jorges *et al.* (2011) apontam que para captar o efeito causal das relações entre as condições de saúde e a educação é preciso atenção na questão de endogeneidade, pois indivíduos com mais anos de estudo podem optar pela adoção de um estilo de vida mais saudável, condizentes com as preferências intertemporais dos agentes entre o consumo presente e futuro. Por outro lado, é possível que indivíduos mais saudáveis priorizem os resultados da função de produção da educação. Nessa situação, o problema de endogeneidade é melhor definido quando os agentes já apresentam maturidade para formulação das preferências intertemporais. Com base em Becker & Mulligan (1997), a educação tem um efeito causal sobre a saúde, sempre que a escolaridade repercute nas preferências dos agentes em um dado intervalo de tempo. Muito embora, possam existir fatores não-observados específicos da criança e de seus familiares que afetem ambos indicadores de saúde e de escola, trabalhos como Glewwe & Jacoby (1995) argumentam que estudos restritos para crianças podem amenizar esse tipo de problema.

Desse modo, a estratégia adotada para mensurar os efeitos pretendidos está baseada em um desenho amostral direcionado a alunos do ensino fundamental que foram expostos aos fatores de risco comportamentais no período antes ou equivalente a pré-adolescência (faixa etária entre 11 e 14 anos de idade). Nessa fase é suposto que a simultaneidade entre educação e condições de saúde ainda não está maturada, pois o status de saúde da criança precede ao resultado na escola. Portanto, não existe a relação de causalidade reversa, mas sim a direção de que problemas de saúde, seja em termos de condições físicas ou psicológicas, impactam no desempenho do aluno. Ademais, a taxa de desconto sobre o valor futuro da saúde ainda pode ser tida como indefinida, visto que a criança não dispõe de maturidade e informações suficientes para formular suas expectativas quanto à maximização intertemporal da sua utilidade, condizente com o estoque de saúde futuro. As escolhas de risco nessa fase de vida é influenciada por fatores exógenos aos alunos, como o ambiente familiar e os efeitos dos pares (CROSNOE; MULLER, 2004). Mesmo admitindo essas hipóteses, análises de sensibilidade são desenvolvidas para verificar se as características não-observadas dos alunos poderiam alterar os resultados dos parâmetros estimados.

As condições de saúde da criança correspondem a um dos importantes *inputs* função de produção educacional, estando presente no vetor de características individuais do aluno. Como no Brasil não existem microdados disponíveis pelos órgãos estatísticos oficiais que reúnam em conjunto informações detalhadas sobre saúde do indivíduo e resultados na escola (nível de proficiência¹ e taxas de rendimento escolar²), foi preciso dispor de uma base que fornecesse informações sobre a saúde do escolar e que fosse possível

¹ Atualmente, o Brasil dispõe de um conjunto de testes padronizados aplicados para os estudantes da educação básica, como a Provinha Brasil, a Prova Brasil e o ENEM. Contudo, nenhum desses exames apresentam dados sobre comportamentos não saudáveis dos alunos e outras medidas de saúde.

² As taxas de rendimento escolar indicam o fluxo do aluno na escola. As informações de rendimento escolar são sumarizadas

identificar o desempenho educacional do aluno. Considerando a PENSE 2012 do Instituto Brasileiro de Geografia e Estatística (IBGE), que aplica um questionário para identificar as condições de saúde e os hábitos dos alunos, utiliza-se a distorção idade-série como indicador educacional.

As estimações dos parâmetros de interesse obedecem três estágios: I) estimação da função de distorção idade-série para exploração inicial dos resultados e primeiras estimativas da associação dos hábitos não saudáveis com o *outcome* escolar; II) cálculo do efeito médio da exposição aos fatores de risco por meio de diferentes técnicas de pareamento, com o intuito de verificar se a direção dos resultados apontados no estágio I são mantidos; III) análise da sensibilidade das estimativas do estágio anterior através dos limites de Rosenbaum para quantificar o tamanho de possíveis interferências de variáveis não-observadas.

2.1 Modelo logit: efeito médio não condicionado ao contrafactual observável

O modelo empírico desta parte se baseia na função de produção educacional, assumindo que o desempenho escolar é função dos insumos relativos às características dos alunos, do *background* familiar, da escola e dos fatores locais. Nos atributos específicos dos estudantes, as estimativas das exposições aos fatores de risco comportamentais à saúde são os parâmetros de interesse. A Equação 1, base para os primeiros resultados desta pesquisa, está descrita a seguir:

$$A_i^* = f(\beta_0 + \beta_1 H_{1i} + \beta_2 H_{2i} + \beta_3 H_{3i} + \mathbf{X}_i \mathbf{b} + u_i) \quad (1)$$

Onde: A_i^* é a variável latente relacionada ao desempenho escolar do *i-ésimo* indivíduo; H_1, H_2 e H_3 são variáveis binárias que refletem a exposição aos fatores de risco pelo o indivíduo i , respectivamente, uso de cigarro, álcool e excesso de peso; \mathbf{X}_i é um vetor que contém as demais variáveis explanatórias que representam outras características específicas do aluno (raça, gênero), da família (anos de estudos da mãe e do pai, nível socioeconômico, estrutura do núcleo familiar, tamanho da família em termos de membros), da escola (recursos de infraestrutura, dependência administrativa) e fatores locais (região de localização da escola).

Supondo que a variável qualitativa distorção idade-série (dis_i), que reflete a variável não-observada do desempenho escolar, segue distribuição logística é possível determinar a probabilidade de que A_i^* seja inferior ou igual a um determinado limiar (j), isto é: $Pr(A_i^* \leq j) = G(\cdot)$. Assim, a Equação 2 é desenvolvida para o caso em que dis_i é enquadrado numa resposta binária, com: $dis_i = 0$, se $A_i^* \leq 0$; e $dis_i = 1$, se $A_i^* > 0$.

$$Pr(dis_i = 1) = G(\beta_0 + \beta_1 H_{1i} + \beta_2 H_{2i} + \beta_3 H_{3i} + \mathbf{X}_i \mathbf{b} + u_i) \quad (2)$$

Onde: $G(\Theta) = 1/[1 + \exp(-\Theta)]$.

A Equação 3 descreve os procedimentos adotados para o cálculo do efeito médio da exposição aos fatores de risco sobre a probabilidade de distorção idade-série, com base nessa primeira abordagem.

$$EMD_k = n^{-1} \sum_{i=1}^n [G(\hat{\beta}_0 + \hat{\beta}_k + \mathbf{W}_i \hat{\mathbf{c}}) - G(\hat{\beta}_0 + \mathbf{W}_i \hat{\mathbf{c}})] \quad (3)$$

Onde: \mathbf{W}_i é o vetor de covariadas com exceção da variável binária de exposição ao fato de risco k .

Como destaca Wooldridge (2010), o cálculo do efeito parcial médio³ (*EPM*) e do efeito parcial na média⁴ (*EPPM*) não fazem sentido para variáveis explicativas discretas, como é a situação de H_1, H_2 e H_3 , que medem a exposição aos fatores de risco comportamentais, por isso a necessidade de estimar o efeito dessas covariadas de interesse por meio da Equação 3 que expressa a média de alteração na probabilidade dos alunos estarem com atraso escolar quando se compara à situação de não-exposição ao fator de risco H_k .

2.2 Modelo PSM: efeito médio condicionado ao grupo de controle observável

A Equação 3 não pondera a existência de fatores omitidos que podem influenciar a decisão de exposição a um dado fator de risco H_k , o que reduz a confiabilidade do efeito calculado em torno da diferença entre as condições de exposição e não-exposição. Na tentativa de minimizar esse problema, fez-se o uso da

por indicadores de aprovação, repetência e abandono escolar.

³ $EPM_j = n^{-1} \sum_{i=1}^n [g(\hat{\beta}_0 + \hat{\beta}_1 H_{1i} + \hat{\beta}_2 H_{2i} + \hat{\beta}_3 H_{3i} + \mathbf{X}_i \hat{\mathbf{b}}) b_j]$, sendo $g(\cdot) = \exp(\cdot)/[1 + \exp(\cdot)]^2$.

⁴ $EPPM_j = g(\hat{\beta}_0 + \hat{\beta}_1 \bar{H}_{1i} + \hat{\beta}_2 \bar{H}_{2i} + \hat{\beta}_3 \bar{H}_{3i} + \bar{\mathbf{X}}_i \hat{\mathbf{b}}) b_j$, sendo $\bar{\Theta}$ a média amostral da variável.

abordagem *propensity score matching* (PSM), em que as estimativas do efeito médio são realizadas por meio da identificação de um grupo de controle com similar características observáveis (\mathbf{Z}_k) dos discentes expostos. Este método foi proposto por Rosenbaum & Rubin (1983), visando resumir as características de cada unidade em um único escore que viabiliza o pareamento.

As dificuldades para medir o efeito causal de uma determinada covariada diz respeito ao problema do contrafactual, dado que só é possível observar apenas um dos resultados potenciais $Y_i(1)$, devido à indisponibilidade da informação sobre o resultado do indivíduo tratado caso ele não fosse exposto, isto é, o seu resultado contrafactual é não observável, $Y_i(0)$. Como essa informação é um *missing value*, torna-se fundamental a construção do contrafactual do grupo exposto aos fatores de risco, denotado como grupo de controle – indivíduos não-expostos, mas que guardam características similares ao grupo tratado.

O primeiro estágio do PSM requer a estimação da probabilidade do aluno ser exposto ao fator de risco k , sendo necessário dispor de fatores que influenciam essa decisão, bem como covariadas associadas com a variável de resultado. Levando em conta modelos empíricos na área da economia da saúde, como Cutler & Glaeser (2005), Park & Kang (2008), Cutler & Lleras-Muney (2010) e Jorges *et al.* (2011), que estudam os fatores determinantes da adoção de comportamentos não saudáveis (uso de tabaco, álcool e/ou drogas ilícitas, condições nutricionais, entre outros), nota-se que a decisão do indivíduo por tais hábitos é influenciada, especialmente, pelos atributos individuais, *background* familiar e situacionais.

Tendo o suporte desses fatores e da função de produção educacional, as variáveis sobre o perfil do aluno, da família, da escola e de covariadas específicas atreladas a decisão de exposição são escolhidas para a estimação do *propensity score*. Assim, a probabilidade condicional de exposição ao fator de risco k condicionado a um vetor de características observáveis é definida pela Equação 4.

$$Pr(H_k = 1 | \mathbf{Z}_k) = G(\mathbf{Z}_k \boldsymbol{\lambda} + \epsilon) \quad (4)$$

Onde: H_k é o indicador de exposição ao fator de risco saúde k ; \mathbf{Z}_k é o vetor de características observáveis para o fator de risco k ; G representa a função de distribuição acumulada. Supondo que G segue uma distribuição de probabilidade logística, pode-se estimar $Pr(\mathbf{Z}_k)$ por meio de um modelo de resposta qualitativa logit.

Após a obtenção do *propensity score* $Pr(\mathbf{Z}_k)$, o efeito médio do tratamento sobre o grupo exposto (τ_k) é estimado a partir da Equação 5 a seguir:

$$\tau_k = E_{Pr(\mathbf{Z}_{ki})|H_k=1} \{ E[dis_i | H_{ki} = 1, Pr(\mathbf{Z}_{ki})] - E[dis_i | H_{ki} = 0, Pr(\mathbf{Z}_{ki})] \} \quad (5)$$

O efeito médio de exposição, estimado pela abordagem PSM, é determinado pela diferença média nos resultados sobre o suporte comum, devidamente ponderados pela distribuição do *propensity score* dos participantes (CALIENDO; KOPEINIG, 2005). Para o cálculo da Equação 5 é preciso utilizar algum algoritmo de pareamento, pois como $Pr(\mathbf{Z}_k)$ é uma variável contínua as chances de identificar duas unidades com igual valor desse escore são ínfimas. Os métodos de pareamentos usados são os seguintes: r-vizinhos mais próximo (PSM-V), com e sem reposição, onde r refere-se a quantidade de vizinhos – nesta pesquisa os valores para r são 1 e 10; raio (PSM-R); e, Kernel (PSM-K), sendo usada a função kernel de Epanechnikov⁵. Conforme sugerido por Rosenbaum (2010), os cálculos do efeito médio por PSM devem ser feitos por mais de uma técnica de pareamento, para analisar a manutenção dos sinais e magnitudes das estimativas. Assim, os resultados apresentados na Subseção 4.2 são desenvolvidos para nove diferentes formas de pareamento, onde no PSM-R e para uma parte do PSM-V são usados um nível máximo de tolerância (*caliper*) de 0,1% da distância do *propensity score*.

Duas suposições centrais são requeridas para o cálculo de τ_k . A primeira diz respeito a hipótese de existência do suporte comum ou sobreposição, no qual $0 < Pr(H_k = 1 | \mathbf{Z}_k) < 1$. Outra importante suposição presume que todas as diferenças relevantes entre os indivíduos expostos e não-expostos são capturadas por \mathbf{Z}_k , onde é necessário a garantia, principalmente, da chamada hipótese de independência condicional (CIA): $Y(0), Y(1) \perp H_k | Pr(\mathbf{Z}_k)$, isto é, os fatores não-observados não afetam a decisão de exposição.

De acordo com Rosenbaum (2002), a hipótese CIA do PSM é altamente restritiva, pois admite que todas as variáveis que influenciam a decisão de exposição e os *outcomes* estão imputadas no modelo. Porém,

⁵ Para maiores detalhes sobre as diferentes técnicas de pareamento, ver Rosenbaum (2010).

a existência de fatores omitidos podem interferir na probabilidade de exposição da criança a um determinado fator de risco. A questão levantada por essa literatura concerne na medição da sensibilidade das estimativas em relação às características não-observadas. Trabalhos como [Gastwirth et al. \(1998\)](#) propõem análises de sensibilidade para examinar o tamanho do viés omitido na atribuição da variável indicadora de exposição que poderiam modificar as conclusões dos efeitos.

[Cornfield et al. \(1959\)](#), um dos pioneiros na análise de sensibilidade, conduz uma investigação sobre a relação causal entre o tabagismo e o câncer de pulmão. Através da formulação do risco relativo aparente, os autores indicam que um fator não-observado deveria ser 9 vezes mais frequente entre os fumantes do que nos não-fumantes para o desenvolvimento do câncer, concluindo que não se pode rejeitar o fato de que o tabagismo é um hábito ofensivo a saúde e a longevidade. Nessa direção, [Gastwirth et al. \(1998\)](#), [Rosenbaum \(2002\)](#), [Caliendo & Kopeinig \(2005\)](#) e [Rosenbaum \(2010\)](#) sugerem a análise de sensibilidade para as estimativas do PSM para prover evidências sobre o grau de dependência dos resultados a hipótese (CIA) não-testável, tendo em vista a lógica do risco relativo aparente.

Se existe um viés omitido, dois indivíduos com idênticas características observáveis \mathbf{Z}_k têm diferentes chances de serem expostos ao fator de risco k , pois o efeito de ϵ (termo de erro estocástico) captado por γ é diferente de zero. Nessas condições, a razão de chances de dois indivíduos pareados serem expostos é expressa pela Equação 6.

$$\frac{Pr(\mathbf{Z}_{ki})[1 - Pr(\mathbf{Z}_{kj})]}{Pr(\mathbf{Z}_{kj})[1 - Pr(\mathbf{Z}_{ki})]} = \frac{G(\mathbf{Z}_{ki}\lambda + \gamma\epsilon_i)}{G(\mathbf{Z}_{kj}\lambda + \gamma\epsilon_j)} \quad (6)$$

Se os indivíduos i e j possuem iguais \mathbf{Z}_k , bem como supondo que $G(\cdot)$ segue uma distribuição logística e que $\epsilon \in (0, 1)$, [Rosenbaum \(2002\)](#) mostra que a Equação 6 implica nos seguintes limites para as razões de chances:

$$\Gamma^{-1} \leq \frac{Pr(\mathbf{Z}_{ki})[1 - Pr(\mathbf{Z}_{kj})]}{Pr(\mathbf{Z}_{kj})[1 - Pr(\mathbf{Z}_{ki})]} \leq \Gamma \quad (7)$$

Onde: $\Gamma = e^\gamma$

Essa abordagem, que permite a análise de sensibilidade para as estimativas do PSM, também é conhecida como limites de Rosenbaum, já que a Equação 7 informa o quanto as mudanças nos valores de Γ alteram a inferência acerca do efeito. Portanto, o valor de Γ mede o grau de afastamento que uma estimativa feita por PSM está livre de fatores não-observados.

Neste estudo, a citada medida possibilita identificar o tamanho da influência das variáveis não-observadas sobre a probabilidade de exposição aos fatores de risco, permitindo investigar em que medida as variáveis omitidas podem interferir na robustez dos resultados encontrados. Por fim, a estatística de teste MH é a mais indicada para testar a hipótese do viés do efeito de exposição, visto as peculiaridades da variável de resultado que é binária ([CALIENDO; KOPEINIG, 2005](#)).

3 Base de dados

Para a execução desse estudo são utilizados os microdados da PENSE 2012, que trazem informações com abrangência nacional sobre fatores comportamentais de risco e de proteção à saúde dos estudantes. O público-alvo desse *survey* é formado por alunos do 9º ano do ensino fundamental do turno diurno de escolas públicas e particulares com no mínimo 15 alunos. Segundo o IBGE (2013), a escolha do 9º ano do ensino fundamental é realizada devido ao mínimo de escolarização necessária para responder ao questionário autoaplicável e com base na faixa etária de referência (entre 13 e 15 anos de idade) recomendada pela Organização Mundial da Saúde - OMS (*World Health Organization* - WHO) para fins de comparabilidade com pesquisas internacionais sobre o tema.

Os dados da PENSE 2012 são estruturados a partir de questionários direcionados aos alunos, que possuem informações sobre atributos do discente – com o diferencial de perguntas dirigidas as condições de saúde – e de seus familiares, e às características da escola. Essa base de dados inicialmente é composta por 109.104 estudantes do último ano do ensino fundamental pertencentes à 2.842 escolas das redes pública (municipal, estadual e federal) e privada. Após uma série de recortes amostrais e exclusão dos *missing*

values, a amostra final desta pesquisa dispõe de aproximadamente 67 mil alunos dispostos em 2.835 unidades escolares. O desenho amostral da PENSE é feito por meio da estratificação do território nacional, no qual cada um dos 27 municípios das capitais e do Distrito Federal formam um estrato geográfico, e os demais municípios agrupados em suas regiões de origem compõem outros cinco estratos geográficos. Os municípios das capitais têm uma amostra de escolas selecionadas (unidades primárias de amostragem - UPA) e, em seguida, as turmas dessas escolas são escolhidas (unidades secundárias de amostragem - USA). Para os municípios fora da capital, as UPA são compostas por agrupamentos de municípios, as USA por unidades escolares selecionadas e as unidades terciárias de amostragem (UTA) por turmas dessas escolas (IBGE, 2013).

Para facilitar a descrição das variáveis, as próximas três subseções exibem o detalhamento, respectivamente, da variável de resultado (atraso escolar), das variáveis referentes à exposição aos fatores de risco comportamentais à saúde e das demais covariadas.

3.1 Distorção idade-série

A distorção idade-série é um *outcome* educacional, pois está relacionada diretamente com o abandono e a repetência escolar, bem como com o problema de evasão escolar e o ingresso tardio na escola⁶. Glewwe & Jacoby (1995) apontam que o atraso escolar contradiz a teoria do capital humano, pelo fato dos custos de oportunidade do tempo da criança serem mais baixos na infância, sendo o ingresso tardio na escola mais custoso para o indivíduo e seus familiares. Nessa mesma direção, os problemas de progressão e evasão da criança na escola geram desdobramentos no valor futuro da renda esperada pelo o indivíduo, decorrentes da ampliação do *trade-off* entre as decisões de estudo e de trabalho para pessoas com atraso escolar, afetando decisivamente a formação educacional. Quando um indivíduo com 18 anos de idade ainda está cursando o último ano do ensino fundamental, a conclusão dessa etapa de ensino e a sequência para outra (ensino médio e ensino superior) possuem altos custos de oportunidade, principalmente para as pessoas mais pobres que precisam contribuir com a renda familiar.

A construção da variável distorção idade-série (*dis*), doravante tratada como atraso escolar, parte do princípio que um sistema educacional seriado, como o brasileiro, há uma adequação teórica entre a série e a idade do discente. Atualmente, a idade recomendada para ingresso no ensino fundamental no Brasil é de 6 anos de idade, como prediz a Lei nº 11.274/2006, que alterou os artigos 29, 30, 32 e 87 da Lei nº 9.394/1996. Sabendo que essa etapa de ensino tem duração total de 9 anos, é possível presumir a idade adequada do estudante em cada série. Logo, a Equação 8 mostra o cálculo do indicador de atraso escolar (*dis*).

$$dis_{is} = \begin{cases} 1, & \text{se } (I_{is} - I_s^e) \geq 2 \text{ ou } [(I_{is} - I_s^e) = 1 \text{ para } m_i \leq 6] \\ 0, & \text{c.c.} \end{cases} \quad (8)$$

Onde: $dis_{is} = 1$ indica a existência de atraso escolar, $dis_{is} = 0$ indica que o indivíduo i está frequentando a série s na idade adequada; I_{is} é a idade em anos do aluno i , obtida pelo ano de nascimento, que frequenta a série s ; I_s^e é a idade recomendada para a série s , que é 14 anos para o último ano do ensino fundamental ($s = 9$); m_i é o mês de nascimento do aluno i . A Equação 8 estabelece que crianças com diferenças positivas em $(I_{is} - I_s^e)$ são enquadrados com atraso escolar, com exceção dos estudantes que apresentam a diferença de um ano idade mas que nasceram no segundo semestre do ano escolar⁷.

3.2 Indicadores de exposição a fatores de risco comportamentais

Este estudo intenciona verificar a associação da exposição a fatores de risco comportamentais à saúde por parte do aluno e o desempenho na escola. Os tipos de exposição contemplam uso de cigarro, consumo de bebidas alcoólicas e sedentarismo/má alimentação (medido pelo sobrepeso).

Os microdados da PENSE 2012 possuem informações sobre a frequência do consumo de diferentes grupos alimentares e práticas de atividades físicas. Contudo, tais dados referem-se às condições correntes.

⁶ Além disso, existem evidências na literatura, como em Gomes-Neto & Hanushek (1994), Gomes-Neto *et al.* (1997) e Soares (2003), que indicam uma relação negativa entre as estimativas de distorção da idade na escola e o baixo desempenho em testes padronizados

⁷ Esse critério é baseado no dicionário de indicadores do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (2004), que considera uma margem de um ano, no qual a tolerância adotada neste presente trabalho é um pouco mais restrita.

Por isso, o uso de informações sobre o índice de massa corporal⁸ (IMC) do aluno com classificação específica por idade e sexo, como uma *proxy* para o estoque de hábitos não saudáveis em relação à alimentação e práticas de atividades físicas. Crianças com valor de IMC acima do 85º quantil da distribuição do IMC específica por idade-gênero são denotadas como estudantes com excesso de peso.

As condições nutricionais, ditadas pelo IMC, captam o estoque passado dos investimentos em saúde, dentre eles a adoção de hábitos de consumo não saudáveis (dieta inadequada) e falta de práticas de atividades físicas⁹. Essa abordagem que obedece a um padrão empírico e específico para cada idade e gênero, dentro de intervalos quantílicos definidos, é recomendada pela OMS para crianças e adolescentes (KUCZMARSKI *et al.*, 2002). Em conformidade com a amostra considerada e esse critério de classificação, os limites (valores mínimos) de IMC que definem os estudantes com sobrepeso estão expressos a seguir¹⁰:

- Meninas: com 13 anos ou menos, 24,89 kg/m^2 ; com 14 anos, 25,31 kg/m^2 ; com 15 anos, 25,92 kg/m^2 ; com 16 anos, 26,95 kg/m^2 ; com 18 anos, 27,74 kg/m^2 ; com 18 anos ou mais, 29,30 kg/m^2 ;
- Meninos: com 13 anos ou menos, 24,77 kg/m^2 ; com 14 anos, 24,80 kg/m^2 ; com 15 anos, 25,39 kg/m^2 ; com 16 anos, 26,37 kg/m^2 ; com 18 anos, 27,46 kg/m^2 ; com 18 anos ou mais, 27,73 kg/m^2 .

Quanto às medidas da exposição ao consumo de álcool e do cigarro estavam disponíveis no questionário do estudante informações sobre a idade que o mesmo teve o primeiro contato com tais produtos, bem como indicadores sobre o consumo desses bens nos últimos 30 dias. Os alunos que declararam ter tido o primeiro contato com bebida alcoólica ou com cigarro com mais de 14 anos de idade foram desconsiderados da amostra, pois a primeira exposição teria ocorrido em uma faixa etária que já estaria acima da idade teórica para o 9º ano do ensino fundamental. Este trabalho usa duas medidas indicativas sobre esses comportamentos, uma relacionada à experiência com o uso do produto e outra que combina a experiência com o uso corrente. Assim, esses indicadores estão divididos em: definição 1 – experiência com o produto com idade igual ou anterior aos 14 anos; e, definição 2 – experiência com o produto (com idade igual ou inferior a 14 anos) combinado com o uso corrente (nos últimos 30 dias do período da aplicação do questionário da PENSE).

3.3 Demais variáveis

Consonante à estratégia empírica adotada, dois vetores específicos de variáveis de controle se fazem necessários. O primeiro é representado pelo vetor X utilizado para uma exploração inicial dos resultados, conjuntamente com as variáveis indicadoras de exposição aos fatores de risco. Por sua vez, o vetor de características observáveis Z_k usado na estimação do *propensity score* para cada fator de exposição k é composto pelo vetor X e por variáveis situacionais relativas à tais fatores no ambiente familiar e escolar. Essas variáveis situacionais incorporam as seguintes informações: na família – indicativos de pais fumantes, de reação da família quanto ao uso de bebida alcoólica pela criança e de promoção de hábitos alimentares saudáveis (consumo de frutas e legumes); na escola – taxa de prevalência de exposição ao fator de risco k entre os alunos. As estatísticas descritivas dessas variáveis estão detalhadas na Tabela A.3 no Apêndice.

Considerando a função de produção educacional, o vetor X é formado por: atributos da criança (gênero, raça, trabalho infantil) e de seus familiares (núcleo familiar, nível socioeconômico, quantidade de membros na família); características da escola (total de salas de aula, dependência administrativa pública, existência de conselho escolar¹¹, indicador de infraestrutura e média do nível socioeconômico dos alunos); fatores locais (capital e grande região geográfica).

⁸ O cálculo desse indicador é dado por: $IMC = peso/(altura^2)$, com o peso expresso em quilograma (kg) e a altura em metro (m).

⁹ Abrevaya & Tang (2010) realçam que as condições nutricionais das crianças são determinadas primariamente por fatores genéticos e pelo ambiente comum (hábitos saudáveis, questões locais, *peer-effects*, entre outros). No entanto, para eles a taxa de crescimento de indivíduos com excesso de peso nos últimos anos indica que o fator ambiente comum vem assumindo um maior grau de importância.

¹⁰ Como à definição do excesso de peso utiliza a distribuição por quantis do IMC, as crianças com idade inferior a 13 anos e superior a 18 anos tiveram que ser avaliadas em conjunto, respectivamente, com alunos de 13 anos de idade e de 18 anos de idade, devido à insuficiência amostral.

¹¹ O conselho escolar presume a participação dos professores, funcionários, pais e alunos no processo decisório da gestão da escola em conjunto com a direção escolar.

No que tange às variáveis de nível socioeconômico (NSE) e ao indicador de infraestrutura escolar (IEE), os mesmos são estimados através da Teoria da Resposta ao Item (TRI), abordagem estatística que ajusta um conjunto de respostas conforme as propriedades de cada item avaliado com o intuito de estimar traços latentes, como habilidades cognitivas, qualidade de infraestrutura, nível socioeconômico etc. (LINDEN; HAMBLETON, 2010). O traço latente do NSE e do IEE captura, respectivamente, o nível socioeconômico do estudante e a qualidade dos recursos estruturais (serviços e infraestrutura) da escola. Para a estimação do traço latente é usado um algoritmo Bayesiano, conhecido como estimador *expected a posteriori* (EAP). As Tabelas A.1 e A.2 no Apêndice descrevem, respectivamente, os itens que compõem os indicadores IEE e NSE.

4 Resultados

Para responder aos objetivos pretendidos desta pesquisa, os resultados encontram-se divididos em três partes: estimação da função de distorção idade-série para exploração inicial dos resultados e primeiras indicações da associação dos hábitos não saudáveis com o *outcome* escolar; cálculo do efeito médio da exposição por meio de técnicas de pareamento; análise da sensibilidade das estimativas do PSM aos fatores não-observados.

4.1 Exploração inicial: Determinantes da atraso escolar

Os resultados estimados da probabilidade do atraso escolar em função dos atributos dos alunos (incluindo os fatores de risco comportamentais) e de suas famílias, das características da escola e dos fatores locacionais estão exibidos na Tabela 1. Constam nesta tabela nove especificações de modelo, que se diferenciam pela inclusão e interação entre a exposição do aluno aos fatores de risco comportamentais à saúde.

Inicialmente, os sinais, a magnitude e a significância estatística das variáveis explicativas mostram o mesmo comportamento para todas as especificações, onde de forma geral os modelos conseguem classificar corretamente 73% dos casos, isso assumindo um *cut-off* padrão de 50%. A amostra final é composta por 66.844 estudantes do ensino fundamental, que representativamente correspondem a 1,9 milhão de discentes da população, em 32 estratos (27 capitais e 5 regiões compostas pelos demais municípios) e em 1.473 unidades primárias de amostragem (UPA). Essas informações do desenho amostral ajustam as estimativas calculadas ao longo não só desta parte, mas de todos os resultados do corrente estudo, com exceção para o cálculo específico da Equação 5, já que o grupo de controle observável é identificado por *matching* e a diferença média é ponderada pela distribuição do *propensity score*¹².

Quanto ao primeiro grupo de variáveis, atributos gerais do aluno e da família, pontua-se que as variáveis referentes ao gênero, trabalho e a quantidade de membros da família elevam as chances de atraso escolar, ao passo que alunos de cor branca, com maior nível socioeconômico (NSE) e que co-residem com ambos os pais tendem a ter um melhor desempenho. Esses indicativos mostram consonância com a literatura empírica internacional e nacional, que enfatizam o papel das variáveis contextuais do perfil socioeconômico do discente no rendimento escolar. Observando os efeitos marginais expressos para esse vetor de variáveis, o NSE do aluno tem uma importante contribuição na probabilidade de atraso escolar, no qual incrementos no indicador socioeconômico resultam em uma redução em 26% em tais probabilidades.

As unidades escolares com melhores atributos em termos de gestão escolar democrática (reflexo da existência de conselho escolar), de qualidade de infraestrutura e de melhor *background* socioeconômico de seus alunos reduzem as chances do discente apresentar distorção idade-série, enquanto que o tamanho da escola, ditado pelo total de matrículas, não é estatisticamente significativo. Assim, o efeito marginal do IEE é de aproximadamente -13% na probabilidade de atraso escolar e, mais uma vez, o efeito do NSE, sendo que o do conjunto dos alunos da escola (os chamados *peer-effects*), possuem uma relação importante com o desempenho do discente ($\approx -38\%$). O controle para características do aluno e da escola retiram os efeitos da rede de dependência administrativa, porque as escolas públicas brasileiras possuem alunos com menos *background* socioeconômico (-38%) e pior infraestrutura (-33%) do que as escolas da rede particular de

¹² No cálculo do *propensity score* é feita a inclusão do desenho amostral, com a proposta de melhor inferir as variáveis que mantêm contatos com a postura de risco da criança e dispor de parâmetros para o modelo com mais proximidade das estimativas da população.

Tabela 1: Resultados da estimação do modelo logit para a probabilidade de atraso escolar

Covariadas	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Gênero (homem=1)	0,5195 ⁱ	0,5331 ⁱ	0,5188 ⁱ	0,5221 ⁱ	0,5381 ⁱ	0,5291 ⁱ	0,5294 ⁱ	0,5392 ⁱ	0,5394 ⁱ
Raça (Branca=1)	-0,1548 ⁱ	-0,1605 ⁱ	-0,1597 ⁱ	-0,1566 ⁱ	-0,1568 ⁱ	-0,1557 ⁱ	-0,1558 ⁱ	-0,1553 ⁱ	-0,1554 ⁱ
Trabalha	0,5054 ⁱ	0,5236 ⁱ	0,5502 ⁱ	0,5278 ⁱ	0,4826 ⁱ	0,4930 ⁱ	0,4926 ⁱ	0,4745 ⁱ	0,4747 ⁱ
Núcleo familiar biparental	-0,4014 ⁱ	-0,4142 ⁱ	-0,4351 ⁱ	-0,4223 ⁱ	-0,4011 ⁱ	-0,3920 ⁱ	-0,3922 ⁱ	-0,3963 ⁱ	-0,3965 ⁱ
NSE do aluno	-1,4456 ⁱ	-1,4988 ⁱ	-1,4506 ⁱ	-1,4522 ⁱ	-1,5065 ⁱ	-1,4744 ⁱ	-1,4745 ⁱ	-1,5018 ⁱ	-1,5028 ⁱ
Tamanho da família	0,0801 ⁱ	0,0838 ⁱ	0,0820 ⁱ	0,0804 ⁱ	0,0809 ⁱ	0,0813 ⁱ	0,0814 ⁱ	0,0800 ⁱ	0,0801 ⁱ
Total de matrículas	0,0488	0,0509	0,0539	0,0511	0,0535	0,0467	0,0468	0,0513	0,0509
Rede pública	-0,0921	-0,0698	-0,0634	-0,0741	-0,0883	-0,0923	-0,0927	-0,0916	-0,0919
Conselho Escolar	-0,2578 ⁱ	-0,2522 ⁱ	-0,2513 ⁱ	-0,2536 ⁱ	-0,2531 ⁱ	-0,2575 ⁱ	-0,2574 ⁱ	-0,2539 ⁱ	-0,2547 ⁱ
Índice de infraestrutura (IEE)	-0,7190 ⁱ	-0,7121 ⁱ	-0,7072 ⁱ	-0,7122 ⁱ	-0,7224 ⁱ	-0,7183 ⁱ	-0,7181 ⁱ	-0,7213 ⁱ	-0,7219 ⁱ
NSE da escola	-2,1671 ⁱ	-2,1781 ⁱ	-2,1082 ⁱ	-2,1257 ⁱ	-2,1595 ⁱ	-2,2067 ⁱ	-2,2081 ⁱ	-2,1679 ⁱ	-2,1664 ⁱ
Capital	-0,3283	-0,3398	-0,3433	-0,3374	-0,3422	-0,3269	-0,327	-0,3384	-0,3379
Região Norte	0,1508	0,15	0,1384	0,1491	0,167	0,1602	0,1603	0,174	0,174
Região Nordeste	-0,114	-0,1219	-0,1237	-0,1141	-0,1201	-0,1086	-0,1087	-0,1102	-0,11
Região Sul	-0,4072	-0,3748	-0,3469	-0,3673	-0,3959	-0,4164	-0,416	-0,4021	-0,4027
Região Centro-Oeste	0,0321	0,0503	0,0587	0,0509	0,0468	0,0341	0,0346	0,0477	0,0478
Cigarro (def.1)	0,4585 ⁱ					0,4046 ⁱ	0,2921 ⁱ		
Álcool (def.1)		0,2553 ⁱ				0,1611 ^t	0,1552 ^t		
Sobrepeso			0,0848 ^t			0,0805 ^t	0,0804 ^t	0,0826 ^t	0,0818 ^t
Cigarro (def.2)				0,5566 ⁱ				0,2834 ⁱ	0,5218 ⁱ
Álcool (def.2)					0,5423 ⁱ			0,5081 ⁱ	0,5201 ⁱ
Cigarro e álcool (def.1)							0,1208		
Cigarro e álcool (def.2)									-0,3017 ⁱ
Constante	1,3905 ^t	1,3098	1,4317 ^t	1,4337 ^t	1,3632 ^t	1,2948	1,2988	1,3451 ^t	1,3423 ^t
Observações	66.844	66.844	66.844	66.844	66.844	66.844	66.844	66.844	66.844
Estratos	32	32	32	32	32	32	32	32	32
UPA	1,473	1,473	1,473	1,473	1,473	1,473	1,473	1,473	1,473
Count-R2 (%)	73,30	73,21	73,09	73,24	73,39	73,28	73,30	73,47	73,45

Fonte: Elaboração própria a partir dos microdados da PENSE 2012.

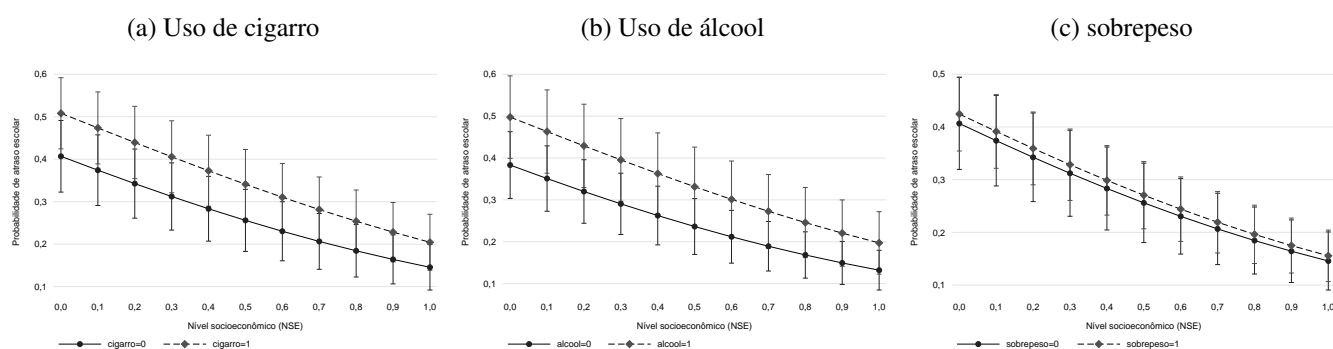
Nota: ⁱ representa coeficientes significativos a pelo menos 5% e ^t a pelo menos 10%; . As estimativas são desenvolvidas considerando, além do peso amostral (a amostra expandida corresponde a uma população de 1.897.670 estudantes), a estratificação e a UPA do plano amostral.

ensino. No que tange aos fatores atrelados às características locais da unidade escolar na Tabela 1, nenhum dos coeficientes se mostram estatisticamente significativos, dada principalmente às altas variâncias dessas estimativas quando se pondera o desenho amostral do *survey* em análise.

Os resultados da Tabela 1 indicam que alunos que são expostos ao excesso de peso, ao cigarro e as bebidas alcoólicas, de acordo com o modelo (9), apresentam respectivamente uma probabilidade, em média, de atraso escolar superior em 1,5%, 8,5% e 9,5% quando comparados à situação de não-exposição. Ao longo de cada uma das especificações disponíveis, os efeitos apresentam pequenas alterações de magnitudes (em especial para definição 1 da exposição ao álcool e definição 2 da exposição ao cigarro), mas os sinais dos coeficientes permanecem mantidos. Os efeitos médios variam: de 7,0% a 8,7% na definição 1 e de 5,3% a 10,9% na definição 2, para o uso de cigarro; de 2,9% a 4,5% na primeira definição e de 9,5% a 10,2% na segunda, para o consumo de bebida alcoólica; de 1,4% a 1,5%, para o excesso de peso. No tocante ao NSE dos alunos, uma pergunta de interesse para as políticas públicas que pode ser colocada neste ponto é: como as diferenças nos resultados escolares de alunos expostos e não-expostos aos fatores de risco à saúde se comportam para diferentes níveis socioeconômicos? Assim, a Figura 1 expressa as diferenças na probabilidade de atraso escolar provocado pela exposição aos fatores de risco, meditando diferentes níveis socioeconômicos. As estimativas são elaboradas com base nos resultados da última especificação do modelo logit dispostos na Tabela 1, com o NSE variando do seu valor mínimo (0) ao seu valor máximo (1) por meio de incrementos de 0,10 unidades na citada variável.

Conforme os resultados dessa ilustração, as diferenças nas probabilidades de atraso escolar entre os expostos e não-expostos são maiores para os alunos mais pobres. Essa constatação pode ser inicialmente visualizada na Figura 1a, onde estudantes expostos ao uso de cigarro com menor nível de NSE têm uma diferença na probabilidade de distorção idade-série de 10%, enquanto que para os indivíduos com maior NSE essa diferença cai para 5,9%. Em igual direção ocorre com o uso de álcool, onde essa diferença de resultado em decorrência desse tipo de consumo diminui gradativamente com a maior padrão de *background*

Figura 1: Efeito dos fatores de risco comportamentais à saúde por NSE – não condicionado ao contrafactual observável



Fonte: Elaboração própria a partir dos microdados da PENSE 2012.

Nota: Todos os valores são significativos estatisticamente à 1%. Estimativas desenvolvidas através dos parâmetros do modelo (9) na Tabela 1.

socioeconômico – o efeito para o grau mínimo de NSE é de 11,5% e para o maior nível de 6,5%. Em relação ao excesso de peso na Figura 1c, as diferenças entre as probabilidades de atraso também são superiores nos níveis iniciais de NSE, onde, por exemplo, o efeito médio da exposição é de 1,8%, no primeiro nível, e de 1% no último.

4.2 Efeito médio da exposição usando PSM

Esta subseção apresenta os resultados obtidos a partir da abordagem PSM, para verificar se os efeitos médios da exposição são mantidos. Os primeiros achados na subseção anterior sinalizam que os alunos expostos aos fatores de risco à saúde tendem a apresentar uma pior *performance* educacional. Especificamente a essas estimativas, a preocupação inicial em termos da inferência dos resultados diz respeito a garantia da ortogonalidade entre a variável indicadora da exposição e os fatores omitidos. Nesse cenário, é recomendável a consideração de outra abordagem estatística que possibilite outra forma de comparabilidade entre as possíveis diferenças nos resultados educacionais provocadas pelos fatores de risco.

Pensando nas características observáveis que podem afetar a decisão de exposição, esse instrumental estatístico visa identificar um grupo de estudantes não-expostos ao fator de risco k que guardem idênticos atributos \mathbf{Z}_k , de modo que a diferença nos resultados escolares possam ser atribuídas à exposição ao fator de risco. Para a construção do efeito médio sobre o grupo exposto ao fator k , denotado como τ_k , é preciso inicialmente estimar o *propensity score* $Pr(\mathbf{Z}_k)$ e, em seguida, utilizar algoritmos de pareamento para o cálculo do τ_k . A Tabela 2 exibe os coeficientes usados na estimação das probabilidades de exposição ao fator de risco à saúde¹³ k (com os números entre parênteses indicando a definição usada para o consumo de álcool e cigarro, conforme apresentada na descrição da base de dados).

Analisando os indicadores de grau de ajustamento dos modelos estimados na Tabela 2, a proporção de casos corretamente previstos é de no mínimo 67,85% para o modelo que traz como variável dependente a decisão de exposição ao consumo de álcool (definição 1), apontando que em todos os modelos as classificações corretas são bem superiores às ocorrências classificadas erroneamente. Pela taxa de previsão correta para indivíduos não-expostos ($H_k = 0$) e expostos ($H_k = 1$), os modelos conseguem fazer previsões de estudantes para cada um dos cenários de exposição, independentemente da distribuição da frequência amostral de cada uma das variáveis dependentes, que em geral são muito concentradas em $H_k = 0$. Entretanto, a pequena taxa de classificações corretas de $H_k = 1$ no modelo para excesso de peso alerta para possíveis problemas de especificações (visto que o questionário da PENSE não dispõe, por exemplo, de questões acerca das condições nutricionais dos pais, sendo fatores omitidos da modelagem). Quanto ao modelo cigarro (2), essa

¹³ A escolha por cada hábito comportamental de risco é modelada de forma independente, pois a relação entre as diferentes exposições não são altas ou estatisticamente diferente de zero, tal como encontrada em outros estudos como Cutler & Glaeser (2005).

Tabela 2: Resultados da estimação dos coeficientes para o *propensity score* por tipo de exposição ao fator de risco à saúde da criança

Covariadas	Cigarro(1)	Cigarro(2)	Álcool(1)	Álcool(2)	Sobrepeso
Gênero (homem=1)	0,0199	-0,1318	-0,2755***	-0,1899***	-0,0161
Raça (Branca=1)	-0,0813**	-0,1998***	0,0268	-0,026	-0,0017
Trabalha	0,6392***	0,8592***	0,5302***	0,6431***	0,0593
Núcleo familiar biparental	-0,4570***	-0,5101***	-0,3853***	-0,3221***	-0,012
NSE do aluno	-0,1469	-0,0177	0,7808***	0,3957**	-0,107
Tamanho da família	0,0233**	0,0745***	-0,0233***	0,0226**	0,0054
Pai fumante	0,3716***	0,3458***			
Mãe fumante	0,3766***	0,3009***			
Reação da família ao álcool			-1,0998***	-1,3870***	
Frequência de consumo: legumes					0,0056
Frequência de consumo: frutas					-0,0119**
Tamanho da escola	0,0237	0,1037***	0,0696**	0,0501***	0,0502***
Rede pública	0,0097	0,0901	-0,045	-0,0564	-0,0093
Conselho Escolar	-0,0716***	-0,0684	0,0042	0,0298	0,0142
Índice de infraestrutura (IEE)	-0,0244	-0,3697***	-0,1477	-0,026	-0,0208
NSE da escola	0,6864***	1,2920***	1,1731***	0,0742	0,1362
Prevalência de exposição na escola ^a	5,9415***	14,3350***	2,8594***	5,2736***	6,9877***
Capital	0,0093	-0,0277	-0,0433	0,0474*	0,0131
Região Norte	-0,0842*	-0,2639**	-0,1027	-0,0482	-0,0506**
Região Nordeste	-0,0671	-0,3272***	-0,0269	0,0675	-0,0469***
Região Sul	0,0536***	0,1715***	0,3415***	0,0571	0,0045
Região Centro-Oeste	0,001	-0,0325	0,0395	0,0051	-0,0248
Constante	-2,9883***	-4,7274***	0,6090***	-1,5817***	-2,9053***
Observações	66.844	66.844	66.844	66.844	66.844
População	1.897.670	1.897.670	1.897.670	1.897.670	1.897.670
Estratos	32	32	32	32	32
UPA	1.473	1.473	1.473	1.473	1.473
Count-R2 (%)	82,16	95,85	67,85	77,28	84,19
Taxa de suporte (%)	47,94	64,24	30,49	37,47	57,75
Previsão correta de $H_k = 0$ (%)	97,89	99,83	17,75	96,26	99,86
Previsão correta de $H_k = 1$ (%)	9,69	2,05	92,75	19,83	0,60

Fonte: Elaboração própria a partir dos microdados da PENSE 2012.

Nota: *p-valor<0,1; **p-valor<0,05; ***p-valor<0,01. As estimativas são desenvolvidas considerando, além do peso amostral, a estratificação e a UPA do plano amostral.

^a Conforme o fator de risco considerado. Por exemplo: no modelo referente ao excesso de peso, essa variável reflete a taxa de alunos na escola enquadrados nessa definição.

taxa pode ser tida como uma decorrência da altíssima concentração de indivíduos não-expostos na amostra (96%). Através da taxa de suporte, que expõe a proporção de indivíduos classificados como expostos pelo modelo mas que na verdade não adotaram posturas de risco, nota-se que, para o *cut-off* padrão de 50%, todos os modelos possuem ocorrências de indivíduos candidatos ao pareamento. Essa interseção entre estudantes não-expostos que poderiam ser classificados como expostos e indivíduos corretamente demarcados como $H_k = 1$ sugere que uma das condições necessárias para o PSM está sendo respeitada, que é a hipótese de suporte comum.

Os coeficientes das covariadas na Tabela 2 revelam que as variáveis relativas ao núcleo familiar do aluno, trabalho, tamanho da escola, taxa de prevalência do fator de risco na escola e os fatores situacionais específicos da família perante ao sobrepeso (captado pela prática alimentar) e uso de bebida alcoólica e cigarro possuem praticamente a mesma direção de efeitos sobre a probabilidade de exposição k . No vetor do contexto escolar, as escolas de maior porte em termos de matrículas geram uma repercussão positiva na decisão do aluno assumir posturas de risco a saúde, em que tal fato pode estar refletindo uma maior influência dos colegas de escola, devido as chances de maior heterogeneidade de comportamentos não-saudáveis entre os alunos. Em igual sentido mas com magnitude maior, as escolas detentoras de maiores taxas de prevalência de alunos expostos dispõem de um cenário de maior incentivo para um determinado estudante ser influenciado pelos seus colegas, não é por acaso que esse *peer-effect* exibe um elevado efeito marginal na probabilidade do discente ser exposto a uma escolha prejudicial a sua saúde. Dentro desse prisma, esses direcionamentos das variáveis contextuais das escolas estão condizentes com outros trabalhos, tais como Veronneau *et al.* (2014), que alvitram o papel de relevo dos colegas na formação psicossocial da criança e dos efeitos desfavoráveis de amizades com pares desviantes.

Tabela 3: Efeito médio sobre o grupo exposto conforme o grupo de controle observável por tipo de exposição e diferentes técnicas de pareamento

Modelo	Cigarro(1)	Cigarro(2)	Álcool(1)	Álcool(2)	Sobrepeso
PSM-V (CR)					
Efeito médio	0,0855***	0,1258***	0,0411***	0,0900***	0,0237***
Erro-padrão	0,00800	0,01303	0,00494	0,00641	0,00695
PSM-V (CR, SC)					
Efeito médio	0,0854***	0,1259***	0,0417***	0,0901***	0,0237***
Erro-padrão	0,00665	0,01372	0,00523	0,00654	0,00696
PSM-V (CR, Caliper)					
Efeito médio	0,0850***	0,1255***	0,0418***	0,0908***	0,0239***
Erro-padrão	0,00585	0,01430	0,00506	0,00681	0,00627
PSM-V (CR, Caliper, SC)					
Efeito médio	0,0850***	0,1255***	0,0419***	0,0908***	0,0239***
Erro-padrão	0,00691	0,01319	0,00469	0,00653	0,00782
PSM-V (SR)					
Efeito médio	0,0844***	0,1240***	0,0404***	0,0935***	0,0239***
Erro-padrão	0,00646	0,01224	0,00386	0,00512	0,00654
PSM-V (SR, Caliper)					
Efeito médio	0,0850***	0,1244***	0,0358***	0,0872***	0,0244***
Erro-padrão	0,00563	0,01187	0,00406	0,00544	0,00625
PSM-V10 (SR, Caliper)					
Efeito médio	0,0836***	0,1272***	0,0395***	0,0851***	0,0168***
Erro-padrão	0,00519	0,01059	0,00488	0,00577	0,00534
PSM-R					
Efeito médio	0,0847***	0,1283***	0,0413***	0,0865***	0,0158***
Erro-padrão	0,00504	0,00986	0,00443	0,00486	0,00512
PSM-K					
Efeito médio	0,0869***	0,1366***	0,0423***	0,0872***	0,0159***
Erro-padrão	0,00496	0,00963	0,00421	0,00477	0,00505

Fonte: Elaboração própria a partir dos microdados da PENSE 2012.

Nota: *p-valor<0,1; **p-valor<0,05; ***p-valor<0,01. Erro-padrão calculado através de *bootstrap* com 100 reposições. No pareamento realizado por PSM-K o valor da janela (*bandwidth*) usado foi de 0,06.

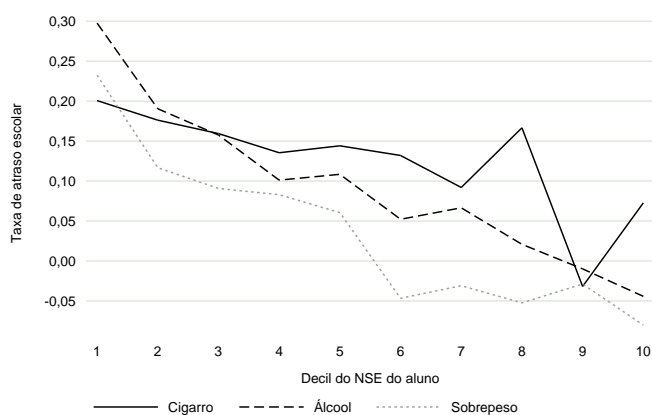
Legenda: CR=com reposição; SR=sem reposição; SC=suporte comum; *Caliper*, nível de tolerância, de 0,1%. O SC impõe um suporte comum que desconsidera observações de tratamento com *propensity score* maior que o valor máximo do escore dos controles ou menor do que o valor mínimo.

As variáveis situacionais da família também são importantes para a tomada de decisão da criança, visto que filhos de pais fumantes são mais propensos a serem expostos ao uso de cigarro. No que respeita o consumo de bebidas alcoólicas, as famílias que reagem de forma muito contrária a adoção desse comportamento pela criança reduzem em 19%, para o modelo álcool (1), e 28%, para álcool (2), a probabilidade do aluno adotar esse tipo de postura. Por sua vez, a variável do contexto familiar do sobrepeso evidencia que boas práticas alimentares promovidas pela família diminuem as chances do discente obter essa classificação, pois as crianças que consomem com mais frequência frutas dispõem de menos chances de terem excesso de peso. Logo, seja no ambiente familiar ou na escola, a decisão da criança se expor a um fator de risco, sobretudo, ao álcool e ao cigarro, é diretamente proporcional aos fatores contextuais vivenciados por ela, o que por si só tais aspectos geram implicações de políticas.

Após a estimação do *propensity score* e identificação do grupo de controle com similares características observadas, a Tabela 3 contém os resultados do efeito médio sobre o grupo exposto baseado em nove formas de pareamento, ditadas pelos PSM-V e suas variantes, PSM-R e PSM-K, sobre o atraso escolar.

Os resultados na Tabela 3 são claros quanto a direção dos efeitos dos fatores de risco sobre o baixo desempenho escolar. As diferenças no resultado educacional entre os grupos exposto e de controle ostentam estimativas estatisticamente significativas à 1%, bem como os sinais dos coeficientes são positivos com as magnitudes praticamente invariantes as diferentes formas de pareamento. Com isso, o efeito para o uso de cigarro é de, aproximadamente, 8,5% quando se avalia a primeira definição, e de mais de 12% na segunda. Os impactos para o consumo de bebida alcoólica e o sobrepeso seguem essa mesma tendência apontada pelos indivíduos expostos ao cigarro, só que com efeitos ligeiramente menores para a bebida alcoólica (beirando os 4% na definição 1 e 9% na definição 2) e com efeitos de magnitude ainda menores para o excesso de peso (variando de 1,6% a 2,4%).

Figura 2: Efeito médio por decis do nível socioeconômico do aluno computado pela abordagem PSM-V com reposição (CR)



Fonte: Elaboração própria a partir dos microdados da PENSE 2012.

Na comparação entre os efeitos calculados diretamente pela estimação da função de produção educacional (subseção 4.1) e os mensurados por PSM – mesmo presumindo diferentes controles nas duas estratégias de estimação –, o papel dos fatores de risco à saúde da criança sobre o resultado na escola permanece, inclusive quando se verifica a tendência do efeito médio (τ_k) em diferentes níveis socioeconômicos dos discentes (ver Figura 2). Nessa figura, desenvolvida por decis do NSE do aluno para os efeitos médios calculados para o modelo PSM-V (CR), sublinha-se que o efeito médio da exposição do fator k exibe uma tendência de relação negativa com o *background* socioeconômico.

Assim, as crianças mais pobres que são expostas apresentam mais riscos de terem seu resultado escolar diminuído. Políticas de enfrentamento a essas práticas não saudáveis poderiam beneficiar de forma mais intensa os grupos com níveis de renda mais baixos, sendo não apenas uma ação de promoção a saúde mas também uma política direcionada a justiça social, com possíveis atenuantes da desigualdade de renda. Na literatura econômica, Conti *et al.* (2010) destacam a relação benéfica entre a saúde da criança, resultados na escola e desdobramentos na trajetória interpessoal de longo prazo.

Do ponto de vista das políticas sobre o enfrentamento das DCNT no Brasil, existe uma preocupação dos *policy makers* com os fatores de risco modificáveis, dado, por exemplo, o plano de ações estratégicas para o enfrentamento das DCNT (BRASIL, 2011). Nesse cenário, os achados desta pesquisa ratificam a relevância da redução da prevalência dos fatores de risco modificáveis na população mais jovem, com destaque para os que dispõem de menor NSE. Ações de tal natureza podem resultar, inclusive, em ganhos significativos na *performance* educacional dos indivíduos.

Os achados dessa etapa da pesquisa partem de uma hipótese muito restritiva (CIA), cuja a presença de variáveis omitidas na especificação do modelo *propensity score* pode resultar em um viés na estimação dos efeitos médios sobre o grupo exposto. Segundo Rosenbaum (2010), como essa hipótese não pode ser testada diretamente, é preciso fazer pelo menos uma análise de sensibilidade desses estimadores.

4.3 Análise de sensibilidade: qual o tamanho do viés de fatores não-observados?

A análise de sensibilidade não exclui a possibilidade de que fatores não-observados existam, mas aponta o quão danoso eles são para os resultados. Com base nessa análise a Tabela 4 exibe os chamados limites de Rosenbaum para cada uma das diferentes especificações de pareamento e tipo de exposição investigado. As informações presentes na tabela exibem o primeiro valor de Γ com pelo menos 10% de significância estatística sob a hipótese de sobrestimação e sua respectiva probabilidade de não rejeição da hipótese nula (p^+).

O valor de Γ reflete a suposição sobre problemas de endogeneidade presentes nas chances de atribuição da exposição motivado por a uma covariada não-observada. Para cada modelo foi calculado um intervalo de níveis críticos com suas probabilidades de não-rejeição da hipótese de sobrestimação dos resultados. Os valores presentes na tabela indicam o limite mínimo com não-rejeição da hipótese nula em que o efeito do

tratamento sofre um distúrbio decorrente da auto-seleção endógena do estado de exposição (ROSENBAUM, 2002).

Tabela 4: Análise de sensibilidade aos fatores não-observados por tipo de exposição e diferentes técnicas de pareamento nos estimadores por PSM

Modelo	Limites	Cigarro(1)	Cigarro(2)	Alcool(1)	Alcool(2)	Sobrepeso
PSM-V (CR)	Γ	1,45	1,60	1,15	1,60	1,10
	p^+	0,14	0,13	0,10	0,42	0,26
PSM-V (CR, SC)	Γ	1,45	1,60	1,15	1,60	1,10
	p^+	0,14	0,13	0,20	0,42	0,26
PSM-V (CR, Caliper)	Γ	1,45	1,60	1,15	1,60	1,10
	p^+	0,16	0,13	0,22	0,42	0,24
PSM-V (CR, Caliper, SC)	Γ	1,45	1,60	1,15	1,60	1,10
	p^+	0,16	0,13	0,22	0,42	0,24
PSM-V (SR)	Γ	1,45	1,60	1,20	1,50	1,10
	p^+	0,45	0,15	0,28	0,22	0,26
PSM-V (SR, Caliper)	Γ	1,45	1,60	1,20	1,45	1,10
	p^+	0,40	0,15	0,37	0,14	0,24
PSM-V10 (SR, Caliper)	Γ	1,45	1,75	1,10	1,70	1,10
	p^+	0,14	0,11	0,23	0,42	0,40
PSM-R	Γ	1,70	2,00	1,10	1,70	1,10
	p^+	0,16	0,23	0,25	0,31	0,48
PSM-K	Γ	1,70	1,95	1,10	1,70	1,10
	p^+	0,13	0,11	0,26	0,33	0,42

Fonte: Elaboração própria a partir dos microdados da PENSE 2012.

Nota: Γ é o primeiro valor com significância estatística do intervalo para a suposição de sobrestimação; p^+ refere-se a probabilidade de não rejeição da hipótese de sobrestimação no efeito da exposição.

A Tabela 4 mostra que a interferência provocada pelos fatores não-observados varia consideravelmente entre os indicativos de exposição. Os efeitos médios do excesso de peso sobre o atraso escolar são os menos robustos à possível presença de viés de seleção, já que o seu nível crítico de Γ é igual a 1,10 para todos os casos, isto é, o viés omitido capaz de gerar questionamentos desses resultados ocorre quando fatores não incluídos no modelo respondem por uma diferença de 10% na razão de chances de atribuição da exposição de excesso de peso entre os grupos de estudantes exposto e não-exposto. Para os demais fatores de risco, o nível crítico de Γ tem um valor maior, indicando uma menor sensibilidade dessas estimativas, pois seria requerido na pior das hipóteses um viés oculto de 1,40 vezes no cigarro (1), de 1,60 no cigarro (2) e de 1,45 no álcool (2) para tornar espúria a conclusão dos efeitos sobre o indicador educacional. Para o álcool (1), o nível crítico varia de 1,10 a 1,20, conforme a técnica de pareamento adotada.

Se os alunos expostos ao uso do cigarro têm, por exemplo, 1,6 vezes mais chances de estar atrasado na escola do que uma criança não-fumante, logo os fatores não-observados devem ser pelo menos 1,6 vezes maiores nos alunos expostos a esse fator de risco. Só assim, os efeitos médios para essa variável poderiam ser considerados sobrestimados caso isso de fato ocorresse. A título de exemplo, uma variável não-observada para resultar uma mudança desse tipo deveria ter um coeficiente maior do que variáveis relativas ao contexto familiar situacional estabelecida por um pai ou por uma mãe fumante, que de acordo com o modelo referente ao cigarro (1) na Tabela 2 exercem uma contribuição de aproximadamente 1,4 vezes na razão de chances da criança ter experiência com o produto.

DiPrete & Gangl (2004) frisam que os limites de Rosenbaum impõem os piores cenários dos resultados, transmitindo informações sobre o grau de incerteza contido nos estimadores obtidos por meio de *matching*. Nas estimativas deste estudo, tal análise evidencia que, até para os efeitos mais sensíveis aos fatores não-observados, existe ainda uma margem para que os resultados não possam ser enquadrados como viesados. Até porque, um Γ de 1,10 não implica que os efeitos da exposição não sejam positivos sobre o atraso escolar, mas que o intervalo de confiança para o supracitado efeito inclua zero se uma variável não-observada provoca uma diferença em 1,10 vezes na razão de chances de atribuição da exposição entre as unidades expostas e de controle. Logo, os achados empíricos, discutidos ao longo da seção de resultados, trazem indícios sobre o papel de uma postura inadequada à saúde infantil no aumento do atraso escolar.

5 Conclusões

Este estudo apresenta direcionamentos empíricos sobre a relação de posturas de risco à saúde na fase escolar e o desempenho educacional do indivíduo. Como ressaltado, esta pesquisa faz um incremento na literatura nacional, pois os trabalhos de complementariedade entre saúde e educação no país ainda não são muito explorados, bem como eles, *Gomes-Neto et al. (1997)* e *Machado (2008)*, até então destacam basicamente o papel de condições antropométricas sobre o desempenho escolar com amostras menos representativas para o território nacional.

Para o desenvolvimento dos efeitos pretendidos houve uma preocupação constante em verificar a robustez dos efeitos calculados, por isso a utilização de recortes amostrais específicos, diferentes métodos estatísticos e técnicas de pareamento e análises de sensibilidade para identificar a manutenção dos resultados para diferentes grupos de controle e o nível de incerteza gerado por fatores não-observados que fossem capazes de alterar as conclusões sobre a direção dessas estimativas.

Os resultados centrais corroboram a hipótese de que crianças expostas a fatores de risco à saúde, como sobrepeso, uso de cigarro e consumo de bebidas alcoólicas, possuem um maior atraso escolar. Conforme *Glewwe & Jacoby (1995)*, o atraso escolar envolve altos custos de oportunidade para o indivíduo e sua família, sobretudo para os pertencentes as classes econômicas menos favorecidas. Sendo assim, ações que envolvam o enfrentamento das DCNT, no tocante aos fatores de risco modificáveis, implicam não apenas melhorias na saúde física e psicológica, mas também numa melhoria do rendimento escolar.

A promoção de programas educativos na escola para mostrar os malefícios desses fatores de risco à saúde e ao resultado escolar seria valioso para os estudantes, a julgar pela forte influência dos *peer-effects* no ambiente escolar sobre a probabilidade da criança decidir adotar posturas de risco, dado que altas taxas de prevalência de colegas da escola expostos repercutem diretamente na decisão individual. Ademais, medidas que atuem na direção da família são relevantes, em que programas de promoção a saúde desenvolvidos pelo Sistema Único de Saúde (SUS) pudessem ter um papel mais efetivo no enfrentamento dos hábitos não saudáveis praticados em comunidades mais carentes. Pois, a diferença no resultado escolar estimuladas por tais hábitos ocorrem com mais força no extremo negativo da distribuição do NSE do aluno. Nessa linha, *Conti et al. (2010)* assinalam que investimentos na primeira infância podem ser uma alternativa para melhorar as condições de saúde e os seus rendimentos acadêmicos, já que nessa fase habilidades cognitivas e sócio-emotivas estão em processo de formação.

Por fim, adverte-se que os resultados discutidos neste presente trabalho são apenas indicativos do papel da exposição aos fatores de risco comportamentais à saúde na *performance* na escola, devido às limitações inerentes a uma base de dados estática. A melhor identificação de causa e efeito requer necessariamente pesquisas longitudinais, que acompanhem a mesma criança ao longo da vida escolar, inclusive verificando seus indicadores de saúde física e psicológica, para assim investigar com maior robustez o impacto da exposição aos fatores em destaque.

Referências

- ABREVAYA, J.; TANG, H. Body mass index in families: spousal correlation, endogeneity, and intergenerational transmission. *Empirical Economics*, v. 41, n. 3, p. 841–864, out. 2010. ISSN 0377-7332.
- ALBERNAZ, A.; FERREIRA, F.; FRANCO, C. Qualidade e Equidade no Ensino Fundamental Brasileiro. *Pesquisa e Planejamento Econômico*, v. 32, n. 3, p. 453–476, 2002.
- ALVES, M. T. G.; SOARES, J. F. *O nível socioeconômico das escolas de educação básica brasileiras Apoio : O nível socioeconômico das escolas de educação básica brasileiras*. Belo Horizonte: UFMG, 2012. 1–57 p.

- BARROS, R. P.; MENDONCA, R.; SANTOS, D. D.; QUINTAES, G. Determinantes do desempenho educacional no Brasil. *Pesquisa e Planejamento Econômico*, v. 31, n. 1, p. 1–42, 2001.
- BECKER, G. S. ; MULLIGAN, C. B. . The Endogenous Determination of Time Preference. *The Quarterly Journal of Economics*, v. 112, n. 3, p. 729–758, 1997.
- BRASIL. *Plano de Ações Estratégicas para o enfrentamento das doenças crônicas não transmissíveis (DCNT) no Brasil: 2011-2022*. Brasília-DF: Ministério da Saúde, 2011. ISBN 9788533418318.
- CALIENDO, M.; KOPEINIG, S. Some Practical Guidance for the Implementation of Propensity Score Matching. *IZA Working Paper n. 1588*, p. 1–32, 2005.
- CHALOUPKA, F.; WARNER, K. E. The economics of smoking. *NBER Working Paper n. 7047*, p. 1–70, 1999.
- COLEMAN, J. S.; HOPKINS, J.; CAMPBELL, E. Q.; HOBSON, C. J. *Equality of educational opportunity*. Washington, D.C.: U.S. Government Printing Office, 1966. 1–737 p.
- CONTI, G.; HECKMAN, J.; URZUA, S. The Education-Health Gradient. *American Economic Review*, v. 100, n. 2, p. 234–238, maio 2010. ISSN 0002-8282.
- CORNFIELD, J.; HAENSZEL, W.; HAMMOND, E. C.; LILIENFELD, A. M.; SHIMKIN, M. B.; WYNDER, E. L. Smoking and Lung Cancer: Recent Evidence and a Discussion of Some Questions. *Journal of the National Cancer Institute*, v. 22, n. 1, p. 173–203, out. 1959. ISSN 1464-3685.
- CROSNOE, R.; MULLER, C. Body Mass Index, Academic Achievement, and School Context: Examining the Educational Experiences of Adolescents at Risk of Obesity. *Journal of Health and Social Behavior*, v. 45, n. 4, p. 393–407, dez. 2004. ISSN 0022-1465.
- CUTLER, D.; GLAESER, E. What explains differences in smoking, drinking and other health-related behaviors? *NBER Working paper series n. 11100*, p. 1–20, 2005.
- CUTLER, D. M.; LLERAS-MUNEY, A. Understanding differences in health behaviors by education. *Journal of health economics*, v. 29, n. 1, p. 1–28, jan. 2010. ISSN 1879-1646.
- DING, W.; LEHRER, S. F.; ROSENQUIST, J. N.; AUDRAIN-MCGOVERN, J. The impact of poor health on academic performance: New evidence using genetic markers. *Journal of health economics*, v. 28, n. 3, p. 578–97, maio 2009. ISSN 0167-6296.
- DIPRETE, T. A.; GANGL, M. Assessing Bias in the Estimation of Causal Effects : Rosenbaum Bounds on Matching Estimators and Instrumental Variables Estimation with Imperfect Instruments. *Sociological Methodology*, v. 34, n. 1, p. 271–310, 2004.
- EIDE, E. R.; SHOWALTER, M. H.; GOLDHABER, D. D. The relation between children’s health and academic achievement. *Children and Youth Services Review*, Elsevier Ltd, v. 32, n. 2, p. 231–238, fev. 2010. ISSN 01907409.
- FLORIN, T. a.; SHULTS, J.; STETTLER, N. Perception of overweight is associated with poor academic performance in US adolescents. *The Journal of school health*, v. 81, n. 11, p. 663–70, nov. 2011. ISSN 1746-1561.
- GASTWIRTH, J. L.; KRIEGER, A. M.; ROSENBAUM, P. R. Dual and simultaneous sensitivity analysis for matched pairs. *Biometrika*, v. 85, n. 4, p. 907–920, 1998.
- GLEWWE, P.; JACOBY, H. G. An economic analysis of delayed primary school enrollment in a low income country: the role of early childhood nutrition. *The review of Economics and Statistics*, v. 77, n. 1, p. 156–169, 1995.

- GOMES-NETO, J. B.; HANUSHEK, E. A. Causes and Consequences of Grade Repetition: Evidence from Brazil. *Economic Development and Cultural Change*, v. 43, n. 1, p. 117–148, 1994.
- GOMES-NETO, J. B.; HANUSHEK, E. A.; LEITE, R. H.; FROTA-BEZZERA, R. C. Health and Schooling: Evidence and Policy Implications for Developing Countries. *Economics of Education Review*, v. 16, n. 3, p. 271–282, 1997.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. *Dicionário de indicadores educacionais: fórmulas de cálculo*. Brasília-DF: MEC/INEP, 2004. 1–28 p.
- Instuto Brasileiro de Geografia e Estatística. *Pesquisa Nacional de Saúde do Escolar*. Rio de Janeiro: IBGE, 2013. 1–256 p. ISBN 9788524042898.
- JAMISON, D. T. Child Malnutrition and School Performance in China. *Journal of Development Economics*, v. 20, p. 299–309, 1986.
- JURGES, H.; REINHOLD, S.; SALM, M. Does schooling affect health behavior? Evidence from the educational expansion in Western Germany. *Economics of Education Review*, v. 30, n. 5, p. 862–872, out. 2011. ISSN 02727757.
- KENKEL, D.; LILLARD, D.; MATHIOS, A. The roles of high school completion and GED receipt in smoking and obesity. *NBER Working paper series n. 11990*, p. 1–47, 2006.
- KENKEL, D. S.; WANG, P. Are Alcoholics in Bad Jobs? In: CHALOUPKA, F. J. (Ed.). *The Economic Analysis of Substance Use and Abuse: An Integration of Econometrics and Behavioral Economic Research*. Chicago: University of Chicago Press, 1999.
- KUCZMARSKI, R. J.; OGDEN, C. L.; GUO, S. S. 2000 CDC growth charts for the United States: Methods and development. *National Center for Health Statistics: Vital Health Stat*, v. 11, n. 246, p. 191, 2002.
- LEVINE, P. B.; GUSTAFSON, T. A.; VELENCHIK, A. D. More bad news for smokers? The effects of cigarette smoking on wages. *Industrial and Labor Relations Review*, v. 50, n. 3, p. 493–509, 1997.
- LINDEN, W. J. van der; HAMBLETON, R. K. (Ed.). *Handbook of Modern Item Response Theory*. 2. ed. New York: Springer, 2010. 1–510 p.
- MACHADO, D. C. Efeitos da saúde na idade de entrada à escola. *Pesquisa e Planejamento Econômico*, v. 38, n. 1, p. 67–96, 2008.
- MACHADO, D. C.; GONZAGA, G. O Impacto dos Fatores Familiares Sobre a Defasagem Idade-Série de Crianças no Brasil. *Revista Brasileira de Economia*, v. 61, n. 4, p. 449–476, 2007.
- MULLAHY, J.; SINDELAR, J. L. Employment, unemployment, and problem drinking. *Journal of health economics*, v. 15, n. 4, p. 409–34, ago. 1996. ISSN 0167-6296. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/10164037>>.
- PARK, C.; KANG, C. Does education induce healthy lifestyle? *Journal of health economics*, v. 27, n. 6, p. 1516–31, dez. 2008. ISSN 0167-6296.
- PONZO, M. Does bullying reduce educational achievement? An evaluation using matching estimators. *Journal of Policy Modeling*, v. 35, n. 6, p. 1057–1078, nov. 2013. ISSN 01618938.
- REILLY, J. J.; KELLY, J. Long-term impact of overweight and obesity in childhood and adolescence on morbidity and premature mortality in adulthood: systematic review. *International journal of obesity*, Nature Publishing Group, v. 35, n. 7, p. 891–8, jul. 2011. ISSN 1476-5497.

- RIANI, J. d. L. R.; RIOS-NETO, E. L. G. Background familiar versus perfil escolar do município: qual possui maior impacto no resultado educacional dos alunos brasileiros. *Revista Brasileira de Estudos Populacionais*, v. 25, n. 2, p. 251–269, 2008.
- RIBEIRO, R.; CACCIAMALI, M. C. Impactos do Programa Bolsa-Família Sobre os Indicadores Educacionais. *Economia*, v. 13, n. 2, p. 415–446, 2012.
- RIOS-NETO, E. L. G.; CESAR, C. C.; RIANI, J. d. L. R. Estratificação educacional e progressão escolar por série no Brasil. *Pesquisa e Planejamento Econômico*, v. 32, n. 3, p. 395–416, 2002.
- ROSENBAUM, P. R. Attributing to Treatment in Matched Observational Studies. *Journal of the American Statistical Association*, v. 97, n. 457, p. 183–192, 2002.
- ROSENBAUM, P. R. *Design of observational studies*. New York: Springer, 2010. 1–382 p. ISBN 9781441912121.
- ROSENBAUM, P. R.; RUBIN, D. B. The central role of the propensity score in observational studies for causal effects. *Biometrika*, v. 70, n. 1, p. 41–55, 1983.
- SOARES, J. F.; ANDRADE, R. J. de. Nível socioeconômico, qualidade e equidade das escolas de Belo Horizonte. *Avaliação e Políticas Públicas em Educação*, v. 14, n. 40, p. 107–126, 2006.
- SOARES-NETO, J. J.; JESUS, G. R. de; KARINO, C. A.; ANDRADE, D. F. de. Uma escala para medir a infraestrutura escolar. *Estudos em Avaliação Educacional*, v. 24, n. 54, p. 78–99, 2013.
- SOARES, T. M. Influência do professor e do ambiente em sala de aula sobre a proficiência alcançada pelos alunos avaliados no Simave-2002. *Estudos em Avaliação Educacional*, n. 28, p. 103–124, 2003.
- VERONNEAU, M.-H.; TREMPE, S.-C.; PAIVA, A. O. Risk and protection factors in the peer context: how do other children contribute to the psychosocial adjustment of the adolescent? *Ciência & Saúde Coletiva*, v. 19, n. 3, p. 695–705, mar. 2014. ISSN 1413-8123.
- WAKEFIELD, J. Sensitivity analyses for ecological regression. *Biometrics*, v. 59, n. 1, p. 9–17, 2003.
- WOOLDRIDGE, J. W. *Econometric Analysis of Cross Section and Panel Data*. 2. ed. Massachusetts: MIT Press, 2010. 1–1064 p.

Apêndice

Tabela A.1: Descrição dos itens utilizados para a elaboração do indicador de infraestrutura da escola (IEE)

Item	Descrição do item	Escala	Prop.	a	d
1	Existência de biblioteca na escola	1, se sim; 0, c.c.	0,908	1,275	2,481
2	Existência de laboratório de informática	1, se sim; 0, c.c.	0,871	1,044	2,188
3	Acesso direto à computadores da escola na sala de aula	1, se sim; 0, c.c.	0,279	0,430	-1,234
4	Acesso à internet através de equipamentos da escola	1, se sim; 0, c.c.	0,859	1,061	1,335
5	Existência de sala para os recursos de mídia/comunicação	1, se sim; 0, c.c.	0,683	1,008	0,512
6	Existência de cantina	1, se sim; 0, c.c.	0,552	0,959	-0,069
7	Existência de quadra de esportes	1, se sim; 0, c.c.	0,830	5,705	4,590
8	Existência de quadra de esporte coberta	1, se sim; 0, c.c.	0,745	3,309	0,699
9	Existência de pista para corrida/atletismo na escola	1, se sim; 0, c.c.	0,026	1,400	-4,502
10	Existência de piscina em condições de uso	1, se sim; 0, c.c.	0,140	1,943	-3,338
11	Existência de vestiário em condições de uso para os alunos	1, se sim; 0, c.c.	0,358	1,839	-1,387

Fonte: Elaboração própria a partir dos microdados da PENSE 2012, tendo por base Soares-Neto *et al.* (2013).

Nota: Parâmetros estimados pela TRI dicotômica unidimensional: a =inclinação; d =intercepto. O modelo logístico de dois parâmetros para o caso de resposta binária foi utilizado para o cálculo do IEE.

Tabela A.2: Descrição dos itens utilizados para a elaboração do Nível Socioeconômico (NSE)

Item	Descrição do item	Escala	Prop.	a	d
1	Existe telefone fixo (convencional)	1, se sim; 0, c.c.	0,528	1,193	0,017
2	O aluno tem celular	1, se sim; 0, c.c.	0,850	0,742	2,076
3	Possui computador em casa	1, se sim; 0, c.c.	0,660	6,371	3,160
4	Acesso à internet em casa	1, se sim; 0, c.c.	0,604	4,890	1,897
5	Algum morador na residência possui carro	1, se sim; 0, c.c.	0,502	1,613	0,211
6	Algum morador na residência possui moto	1, se sim; 0, c.c.	0,297	-0,112	-0,688
7	<i>Número de banheiros com chuveiro dentro da casa</i>	De 0 a 4		1,180	
	Sem banheiro	0	0,043	0,000	0,000
	Apenas um banheiro	1	0,649	1,000	3,436
	Dois banheiros	2	0,228	2,000	2,615
	Três banheiros	3	0,055	3,000	0,615
	Quatro banheiros ou mais	4	0,025	4,000	-1,612
8	Empregado doméstico na casa cinco ou mais dias por semana	1, se sim; 0, c.c.	1,480	1,465	-2,502
9	<i>Faixa de escolaridade da mãe</i>	De 0 a 4		0,749	
	Nunca estudou ou ensino fundamental incompleto	0	0,423	0,000	0,000
	Ensino fundamental completo ou médio incompleto	1	0,181	1,000	-0,463
	Ensino médio completo	2	0,234	2,000	-0,174
	Ensino superior incompleto	3	0,054	3,000	-1,818
	Ensino superior completo ou mais anos de estudo	4	0,109	4,000	-1,702
10	<i>Faixa de escolaridade do pai</i>	De 0 a 4		0,724	
	Nunca estudou ou ensino fundamental incompleto	0	0,468	0,000	0,000
	Ensino fundamental completo ou médio incompleto	1	0,176	1,000	-0,653
	Ensino médio completo	2	0,207	2,000	-0,470
	Ensino superior incompleto	3	0,045	3,000	-2,151
	Ensino superior completo ou mais anos de estudo	4	0,105	4,000	-1,899

Fonte: Elaboração própria a partir dos microdados da PENSE 2012, tendo por base Soares & Andrade (2006) e Alves & Soares (2012).

Nota: Parâmetros estimados pela TRI policotômica ordinal unidimensional: a =inclinação; d =intercepto. Nesse caso o modelo usado foi o de crédito parcial generalizado.

Tabela A.3: Estatística descritivas das variáveis

Variável	N	P	Média	DP	Mín.	Máx.
Gênero (homem=1)	66.844	1.897.670	0,480	0,500	0,000	1,000
Raça (Branca=1)	66.844	1.897.670	0,386	0,487	0,000	1,000
Trabalha	66.844	1.897.670	0,129	0,336	0,000	1,000
Experiência: cigarro	66.844	1.897.670	0,180	0,384	0,000	1,000
Experiência e uso: cigarro	66.844	1.897.670	0,041	0,198	0,000	1,000
Experiência: álcool	66.844	1.897.670	0,663	0,473	0,000	1,000
Experiência e uso: álcool	66.844	1.897.670	0,244	0,429	0,000	1,000
Sobrepeso	66.844	1.897.670	0,158	0,365	0,000	1,000
Núcleo familiar biparental	66.844	1.897.670	0,679	0,467	0,000	1,000
NSE do aluno	66.844	1.897.670	0,458	0,199	0,000	1,000
Tamanho da família	66.844	1.897.670	4,583	1,614	1,000	30,000
Pai fumante	66.844	1.897.670	0,139	0,346	0,000	1,000
Mã fumante	66.844	1.897.670	0,085	0,279	0,000	1,000
Reação contrária dos pais ao uso do álcool	66.844	1.897.670	0,906	0,292	0,000	1,000
Frequência de consumo: legumes	66.844	1.897.670	3,397	2,590	0,000	7,000
Frequência de consumo: frutas	66.844	1.897.670	3,100	2,495	0,000	7,000
Tamanho da escola (matrículas em mil)	66.844	1.897.670	0,983	0,568	0,067	8,300
Rede pública	66.844	1.897.670	0,802	0,399	0,000	1,000
Conselho Escolar	66.844	1.897.670	0,867	0,340	0,000	1,000
Índice de infraestrutura (IEE)	66.844	1.897.670	0,480	0,196	0,000	1,000
NSE da escola	66.844	1.897.670	0,455	0,145	0,105	0,941
<i>Prevalência de exposição na escola</i>						
Experiência: cigarro (def.1)	66.844	1.897.670	0,194	0,107	0,000	0,714
Experiência e uso: cigarro (def.2)	66.844	1.897.670	0,050	0,049	0,000	0,538
Experiência: álcool (def.1)	66.844	1.897.670	0,669	0,132	0,050	1,000
Experiência e uso: álcool (def.2)	66.844	1.897.670	0,259	0,106	0,000	0,722
Sobrepeso	66.844	1.897.670	0,156	0,068	0,000	0,600
Capital	66.844	1.897.670	0,780	0,415	0,000	1,000
Região Norte	66.844	1.897.670	0,082	0,275	0,000	1,000
Região Nordeste	66.844	1.897.670	0,257	0,437	0,000	1,000
Região Sul	66.844	1.897.670	0,157	0,364	0,000	1,000
Região Centro-Oeste	66.844	1.897.670	0,080	0,271	0,000	1,000

Fonte: Elaboração própria a partir dos microdados da PENSE 2012.