

Efeitos da Exposição aos Fatores de Risco Comportamentais à Saúde sobre o Atraso Escolar no Brasil*

Aléssio Tony Cavalcanti de Almeida[†]
Ignácio Tavares de Araújo Júnior[‡]

Sumário: 1. Introdução; 2. Função de produção educacional e o papel das condições de saúde; 3. Estratégia empírica; 4. Base de dados; 5. Resultados; 6. Conclusões; Apêndice.

Palavras-chave: Cigarro, Bebida Alcoólica, Sobrepeso, Resultado Escolar.

Códigos JEL: I12, I21, J13, J24.

Este trabalho objetiva 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 em conjunto com técnicas paramétricas e não-paramétricas são utilizados para a estimação do efeito da exposição a esses fatores sobre o indicador de atraso escolar de discentes no 9º ano do ensino fundamental. Os principais resultados da pesquisa indicam 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 pior nível socioeconômico. Esses achados sugerem a importância de políticas públicas que promovam a prevenção desses fatores de risco desde a infância, pois as consequências dessas exposições podem gerar desdobramentos não apenas na saúde mas também no componente educacional do capital humano.

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 (Pesquisa Nacional de Saúde do Escolar–PENSE) 2012 and parametric and nonparametric techniques to estimate the effect of exposure to these factors in the indicator of delay in school progression of students in the 9th grade of elementary school. The main results indicate 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

*Os autores agradecem aos comentários e às sugestões dos pareceristas anônimos que ajudaram a melhorar a presente versão do trabalho. Qualquer erro remanescente é de exclusiva responsabilidade dos autores.

[†]Universidade Federal da Paraíba, Departamento de Economia (UFPB/CCSA). Cidade Universitária, Campus I, João Pessoa, PB, Brasil. CEP 58051-900. Tel. +55 83 3216 7453. Email: alessio@ccsa.ufpb.br

[‡]Universidade Federal da Paraíba, Departamento de Economia (UFPB/CCSA). Cidade Universitária, Campus I, João Pessoa, PB, Brasil. CEP 58051-900. Tel. +55 83 3216 7453. Email: ignacio.tavares@gmail.com



findings of this study suggest the importance of public policies that promote prevention of these risk factors among children, once the exposure to risk factors to health can generate repercussions not only in health but also in the educational component of human capital.

1. INTRODUÇÃO

O capital humano é posto como um elemento essencial para o crescimento e o desenvolvimento socioeconômico de um país (Schultz, 1961; Becker, Murphy & Tamura, 1990; Mankiw, Romer & Weil, 1992; O'Neill, 1995; Foster & Rosenzweig, 1995; Grossman, 2005; Hanushek & Woessmann, 2008; Glewwe, Hanushek, Humpage & Ravina, 2011). 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 programas sociais (Gomes-Neto & Hanushek, 1994; Barros, Mendonça, Santos & Quintaes, 2001; Rios-Neto, César & Riani, 2002; T.M. Soares, 2003; D.C. Machado, 2008; Riani & Rios-Neto, 2008; Sampaio, Sampaio, Mello & Melo, 2011; Ribeiro & Cacciamali, 2012; Almeida, 2014).

Adicionalmente, a literatura internacional vem examinando as consequências no desempenho na escola em razão da exposição dos jovens aos fatores de risco à saúde (Jamison, 1986; Glewwe & Jacoby, 1995; Yamada, Kendix & Yamada, 1996; Crosnoe & Muller, 2004; Kaestner & Grossman, 2009; Kaestner, Grossman & Yarnoff, 2011; Carrell, Hoekstra & West, 2011; Ponzo, 2013; Lindo, Swensen & Waddell, 2013). Como ressaltam Mullahy & Sindelar (1996), Levine, Gustafson & Velenchik (1997), Chaloupka & Warner (1999), Kenkel & Wang (1999), Crosnoe & Muller (2004), Florin, Shults & Stettler (2011) e Cawley & Ruhm (2011), as condições de saúde das crianças e adolescentes, decorrentes de hábitos não saudáveis, possuem impactos diretos no funcionamento individual, com correspondência em fatores: físicos, como limitações funcionais, disfunções no desenvolvimento das capacidades cognitivas, problemas cardiovasculares e respiratórios, maiores incidências de diabetes e outros agravos à saúde; não-físicos, como baixa concentração, problemas psicológicos, maior absenteísmo e menor participação em sala de aula. Por decorrência, seguindo a linha desses autores, a criança exposta a hábitos não saudáveis tende a apresentar menor rendimento na escola.

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 do cigarro e cerca de 40% dos discentes passam no mínimo 5 horas por dia (exclusive feriados e fins de semana) sentados — assistindo televisão, ao computador ou jogando videogame. Esses números revelam que uma quantidade elevada de adolescentes no país adotou ou adota posturas de risco à saúde em idade escolar, sendo relevante examinar as implicações dessas exposições.

É possível encontrar trabalhos aplicados para dados nacionais, como Gomes-Neto, Hanushek, Leite & Frota-Bezzera (1997) e D. C. Machado (2008), que estimam o impacto das condições de saúde dos indivíduos, através de indicadores antropométricos, sobre os resultados escolares. Contudo, ainda são escassos estudos que associam hábitos não saudáveis das crianças e rendimento escolar. 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).

Assim, este trabalho pretende avançar nas discussões dos determinantes do desempenho educacional no Brasil ao avaliar o efeito de comportamentos não saudáveis, representados 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 teoria do capital humano, economia da educação e da saúde, uma vez que as experiências e acúmulos de conhecimentos nessa fase da vida servem de base para as trajetórias econômicas, sociais e de saúde dos indivíduos no longo prazo (Conti, Heckman & Urzua, 2010). O foco na avaliação dos comportamentos em destaque sobre os resultados na escola está baseado em Cawley & Ruhm (2011), que assinalam os fortes efeitos diretos de fumar, beber e sobrepeso na própria saúde.

Os resultados principais são desenvolvidos através de modelos paramétricos (modelo de resposta qualitativa) e não-paramétricos (*propensity score matching* – PSM) em conjunto com informações extraídas dos microdados da PENSE 2012. 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, Krieger & Rosenbaum (1998), Rosenbaum (2002), Wakefield (2003) e Rosenbaum (2010), esta abordagem não exime 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 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, assim como realiza uma série de especificações para testar a robustez dos resultados.

Este artigo está dividido em seis partes, o que inclui esta introdução. A seção 2 apresenta os principais resultados da literatura baseados na função de produção educacional, enfatizando o papel das condições de saúde dos alunos. A seção 3 e a seção 4 tratam, respectivamente, da estratégia empírica e das informações sobre a base de dados e tratamento das variáveis. Por fim, a seção 5 e a seção 6 mostram os principais resultados e conclusões.

2. FUNÇÃO DE PRODUÇÃO EDUCACIONAL E O PAPEL DAS CONDIÇÕES DE SAÚDE

A partir do relatório de Coleman et al. (1966), baseado em uma análise detalhada sobre a igualdade de oportunidades educacionais nas escolas públicas norte-americanas no ano de 1965, existe uma preocupação crescente na literatura especializada em identificar os determinantes do resultado escolar, seja em países desenvolvidos ou em desenvolvimento. A função de produção educacional que baliza esses trabalhos é definida na equação (1):

$$A_i = f(S_i, Q_i, C_i, H_i, I_i), \quad (1)$$

onde: A_i é o desempenho escolar do i -ésimo indivíduo; S é uma variável escalar que mede os anos de estudo; Q representa o vetor de atributos da escola e dos professores; C é o vetor de variáveis individuais do aluno; H é o vetor de características da família; I é o vetor de insumos escolares que estão sob o controle dos pais, tais como a compra de livros e outras necessidades escolares.

Através da função de produção educacional, estudos como Hanushek (1970; 1971; 1986), Eide & Showalter (1998), Rangvid (2007) e Glewwe et al. (2011), em nível internacional, e Barros et al. (2001), Albernaz, Ferreira & Franco (2002), A. F. Machado, Moro, Martins & Rios (2008), Riani & Rios-Neto (2008) e Almeida (2014), em nível nacional, enfatizam o papel do *background* familiar no processo de geração do bom desempenho acadêmico do aluno. Glewwe et al. (2011) fazem uma revisão da literatura nas áreas de educação e economia de 1990 a 2010 nos países em desenvolvimento, tendo como meta investigar as características da escola e do professor no resultado escolar a partir de uma seleção de 79 publicações. Através dessa análise, os autores concluem que as características da escola e do professor possuem pequenos efeitos ou insignificância estatística, levando em conta a qualidade dos modelos estatísticos. Independentemente do agrupamento econômico do país, existe certo consenso acerca da importância do perfil do aluno e da família na produção da educação.



No campo da literatura que sublinha o papel dos atributos de saúde das crianças e adolescentes, que compõem o vetor de características individuais, nos indicadores de realização educacional é recorrente a identificação de uma associação positiva entre boas condições de saúde e medidas de performance educacional (Eide & Showalter, 2011). Os trabalhos apontam os efeitos do status nutricional, mensurado por indicadores antropométricos (Jamison, 1986; Mook & Leslie, 1986; Glewwe & Jacoby, 1995; Gomes-Neto et al., 1997; Crosnoe & Muller, 2004; D. C. Machado, 2008; Kaestner & Grossman, 2009), de fatores comportamentais (Pirie, Murray & Luepker, 1988; Ellickson, Tucker & Klein, 2001; Sigfusdottir, Kristjansson & Allegrante, 2007; Ding, Lehrer, Rosenquist & Audrain-McGovern, 2009; Kristjansson, Sigfusdottir & Allegrante, 2010; Carrell et al., 2011; Ponzio, 2013; Lindo et al., 2013) e outras medidas de saúde (Eide, Showalter & Goldhaber, 2010; Rees & Sabia, 2011; Rees & Sabia, 2014) sobre o nível de proficiência, atraso escolar, conclusão dos estudos e outros indicadores educacionais.

2.1. Condições nutricionais e desempenho escolar

Na área de saúde pública, mais recentemente a obesidade tem sido um dos tópicos mais investigados, reflexo dos riscos gerados sobre o funcionamento físico, morbidade e mortalidade e sua crescente prevalência em todos os grupos etários da população, com destaque para os países em desenvolvimento (Cole, Bellizzi, Flegal & Dietz, 2000; Reilly & Kelly, 2011). Além dos aspectos atrelados estritamente a saúde, o papel social e psicológico do problema da obesidade (preconceito, autoestima, *bullying* etc.) se apresenta como fatores que intensificam a repercussão do excesso de peso na menor aquisição de capital humano por parte dos mais jovens (Crosnoe & Muller, 2004; Ponzio, 2013).

Segundo Reilly & Kelly (2011), crianças e adolescentes obesas enfrentam comorbidades físicas e psicossociais tanto no curto prazo quanto no longo prazo. Baseado nessas premissas, os trabalhos que tratam do papel do status nutricional desse grupo etário mais jovem buscam entender como isso repercute no desempenho escolar. Se por um lado a obesidade pode repercutir em várias dimensões físicas, sociais e econômicas dos indivíduos, a desnutrição infantil observada em países de baixa renda (extremo oposto da obesidade) também tem sido documentada na literatura empírica no que tange os seus impactos na escola¹ (Jamison, 1986; Mook & Leslie, 1986; Glewwe & Jacoby, 1995).

Jamison (1986) estima a relação entre a saúde das crianças chinesas — através de razões normalizadas de altura-idade, peso-idade e peso-altura — e o desempenho na escola para uma amostra com 3.000 estudantes do ensino fundamental em cinco regiões da China. Os seus achados indicam que o baixo status nutricional, mensurado pela altura-idade, atua negativamente na performance estudantil independentemente da região. Nessa mesma direção, o trabalho de Mook & Leslie (1986) é desenvolvido para uma amostra final de 350 crianças entre 5 e 11 anos de idade de escolas primárias da região de Terai no Nepal. As estimativas feitas por mínimos quadrados ordinários (MQO) e por um modelo probit pontuam que a desnutrição induz baixos indicadores escolares (matrícula e promoção na escola). Os autores ressaltam que esforços para melhorar o estado nutricional da criança pode implicar, além dos benefícios para a saúde, em maiores realizações educacionais.

Partindo da hipótese que o atraso escolar são respostas racionais a desnutrição na primeira infância, Glewwe & Jacoby (1995) usam dados do *Ghana Living Standards Survey* (GLSS) 1988–1989 e um modelo probit ordinal com diferentes alternativas de especificação para uma amostra final de 1.757 crianças entre 6 e 15 anos de idade. As estimativas assinalam que a desnutrição na primeira infância causa o ingresso tardio na escola, implicando a pertinência de políticas de intervenção nutricional na primeira infância com vistas na melhoria da saúde e aumento futuro da riqueza desses indivíduos.

A partir dos dados do *National Longitudinal Study of Adolescent Health* (NLSAH) de 2004 e regressão multinível, Crosnoe & Muller (2004) verificam que a obesidade, por ser uma condição de saúde estigmatizada na população norte-americana, relaciona-se de forma negativa com o desempenho acadêmico

¹ Conforme destacam Glewwe & Jacoby (1995), a desnutrição infantil pode prejudicar o crescimento da estatura e reduzir o desenvolvimento mental, bem como diminuir os níveis de motivação e energia da criança. Tais aspectos podem implicar em menor rendimento na escola.

dos alunos do ensino secundário. Esse impacto é ainda maior nas escolas que, por exemplo, possuem altas taxas de alunos participantes de atividades esportivas. Os autores argumentam que o baixo desempenho na escola, motivado por questões de obesidade, tem fortes influências de longo prazo nas relações interpessoais. Utilizando a pesquisa *Icelandic Study* para estudantes com 14 e 15 anos de idade de escolas secundárias da Islândia, Sigfusdottir et al. (2007) revelam que comportamentos saudáveis (índice de massa corporal, dieta nutricional, prática de atividades físicas e fatores de saúde mental) são importantes para melhoria do desempenho acadêmico.

Pelas estimações do modelo empírico de Kaestner & Grossman (2009), baseado nos dados do National Longitudinal Survey of Youth (NLSY) para estudantes entre 14 e 18 anos de idade e nos indicadores de progressão e abandono escolar, uma pior realização acadêmica não pode ser atribuída ao excesso de peso na adolescência. Em contraste, Florin et al. (2011) realçam a importância da percepção do excesso de peso na compreensão dos resultados acadêmicos dos adolescentes, pois essa relação negativa entre o desempenho escolar e a saúde do jovem tem expressivas consequências para o nível de bem-estar do indivíduo quando adulto.

2.2. Consumo de álcool, cigarro e outras substâncias na performance escolar

Esta seção visa apresentar aspectos da literatura nas áreas de saúde e economia sobre a repercussão do consumo de bebida alcoólica e de cigarro por parte de crianças e adolescentes na aquisição de capital humano.

A incidência de consumo de álcool entre os adolescentes ao redor do mundo ainda é elevada, apesar dos esforços empreendidos pelos governos (legislações rígidas, controles publicitários, elevadas alíquotas tributárias etc.) para coibir o uso desse produto entre os jovens (Brasil, 2011). Para Renna (2007), esse tipo de controle nos EUA se deve pelo alto custo social associado ao uso de álcool por parte dos menores de 21 anos de idade, com destaque para os seguintes argumentos: uma das principais causas de morte entre os adolescentes; altos efeitos secundários, medidos pelos danos causados a terceiros e/ou propriedade privada pelos jovens bêbados; consequências negativas para o pleno desenvolvimento das capacidades habilidades cognitivas individuais.

Haja vista à relevância da capacidade cognitiva para a obtenção de uma boa formação educacional, os estudos apontam que a exposição ao álcool pode repercutir negativamente no processo de acumulação do capital humano. Segundo apresentado por Renna (2007), as pesquisas clínicas têm apontado que a bebida alcoólica deteriora o desenvolvimento do cérebro humano, especialmente, para o caso dos adolescentes cujo o cérebro ainda está em processo de desenvolvimento. Adolescentes com histórico de consumo de bebida alcoólica possuem déficits nas habilidades de memória verbal e não-verbal e nas funções visuoespaciais (Brown, Tapert, Granholm & Delis, 2000), assim como os danos cerebrais e neurocognitivos implicam em menor nível de aprendizagem e desenvolvimento intelectual (Zeigler et al., 2005).

Em termos do prisma econômico, Koch & Ribar (2001) argumentam que o consumo de bebidas alcoólicas pelos mais jovens gera efeitos na escolaridade, de modo que essa relação pode ser entendida por meio do arcabouço do modelo de alocação de tempo de Becker (1965), considerando as seguintes hipóteses: o progresso na escola requer investimentos de tempo (por exemplo, frequência nas aulas e estudo) e desembolso de recursos (material escolar, transporte, livros, lanches etc.); o consumo de álcool requer tempo e dinheiro; os jovens têm preferências em relação a estudar, beber e consumir outros bens, tendo contudo restrições orçamentárias e temporais.

Nessa abordagem, o consumo de álcool pode ter um efeito negativo direto sobre a escolaridade caso os recursos individuais que poderiam ser destinados à escola se desviem em direção ao consumo da bebida alcoólica. Adicionalmente, conforme Koch & Ribar (2001), o uso de álcool possui efeitos colaterais na escolaridade, pois tornam o investimento em educação menos produtivos. A perda de produtividade ocorreria, de acordo com Lye & Hirschberg (2010), através de perdas no estoque de saúde (física e mental) do capital humano.



No que tange ao uso de cigarro por parte de crianças e adolescentes e suas implicações na escolaridade, os estudos, como Ellickson et al. (2001), Jeynes (2002) e Martins & Alexandre (2009), convergem em termos gerais para mecanismos similares aos apresentados anteriormente. Como aponta Levine et al. (1997), o cigarro gera consequências graves na saúde dos indivíduos, afetando de forma direta o capital humano. A maior parte dos seus efeitos surgem geralmente mais tarde na vida dos seus usuários, contudo Andreski & Breslau (1993) evidenciam a ocorrência de implicações do consumo desse produto na saúde física e psicológica de usuários jovens.

Na perspectiva clínica, além dos já conhecidos danos à saúde no longo prazo, Jacobsen et al. (2005), Bergen, Martin, Roeger & Allison (2005), Jacobsen, Pugh, Constable, Westerveld & Mencl (2007) e o relatório do U.S. Department of Health and Human Services (2012) destacam como o uso direto do cigarro, abstinência e interrupção do fumo podem repercutir sobre o funcionamento fisiológico e psicológico dos indivíduos mais jovens. Segundo esses estudos, adolescentes usuários ou ex-usuários de cigarro são mais propensos a se defrontarem com:

- deficiências agudas de memória verbal e memória de trabalho após a cessação do uso do tabaco;
- decréscimos crônicos no desempenho cognitivo, sobretudo para aqueles que iniciaram o uso do produto com idade mais precoce;
- maior ansiedade com perdas de humor e concentração quando da abstinência de nicotina;
- problemas respiratórios, crescimento pulmonar prejudicado e início precoce de redução da função pulmonar durante o final da adolescência e início da idade adulta.

Com base nessas disfunções, tal como esperado para os adolescentes expostos ao consumo de bebidas alcoólicas, os fumantes tenderiam a possuir perdas progressivas de rendimento escolar. Conforme o U.S. Department of Health and Human Services (2012), o consumo de cigarros pelos jovens tem consequências adversas imediatas para a saúde, incluindo vício e a aceleração do desenvolvimento de doenças crônicas.

Em termos empíricos, segundo Lye & Hirschberg (2010) e Cawley & Ruhm (2011), os efeitos do consumo de álcool, cigarro e outros tipos de drogas sobre os resultados educacionais têm sido frequentemente documentados na literatura internacional, especialmente as investigações sobre os impactos do uso de bebidas. Koch & Ribar (2001), a partir do painel de dados do NLSY 1979–1990 para pares de irmãos do mesmo sexo, encontram que as consequências do uso de bebidas alcoólicas na juventude sobre a educação é sensível as diferentes especificações econométricas. Eles destacam que o consumo abusivo de álcool diminui os indicadores de escolaridade, mas que o uso moderado não tem um efeito expressivo no desempenho escolar. Outrossim, Chatterji (2006) sugere que o uso de álcool no ensino médio está associado com reduções no nível de escolaridade dos indivíduos. Nesse trabalho, as estimativas são calculadas através dos dados do *National Education Longitudinal Study* (NELS) e de modelos probit.

Ao verificar os resultados de Renna (2007), que utiliza os dados do NLSY de 1979, tem-se que o consumo excessivo de álcool diminui a probabilidade do indivíduo concluir o ensino médio (até 19 anos de idade) em até 5,2% para as mulheres e 14,5% para os homens. Renna (2008) amplia a investigação do trabalho anterior, ao considerar as consequências da utilização de bebidas com teor alcoólico na adolescência sobre um conjunto maior de *outcomes* educacionais, incluindo além da conclusão do ensino médio, indicadores como abandono escolar e conclusão na idade prevista. Através dos dados do NLSY de 1979 e de diferentes modelos estatísticos, como probit ordenado, logit multinomial e probit em dois estágios, o autor confirma a hipótese dos efeitos adversos da bebida alcoólica na realização educacional dos discentes, dada a redução na probabilidade destes alunos alcançarem um maior nível de escolaridade.

Por sua vez, Balsa, Giuliano & French (2011) mostram que o aumento no consumo de álcool resulta em pequenas reduções na qualidade da aprendizagem para estudantes do sexo masculino, enquanto que para as alunas os efeitos não são significativos estatisticamente. Os autores usam dados do NLSAH e modelos de efeitos fixos para analisar o papel do uso de álcool sobre indicadores educacionais quantitativos (dias de aulas perdidos) e qualitativos (dificuldades com concentração na escola, convivência com os

professores ou completar a lição de casa). No caso dos meninos, a redução na performance educacional ocorre em particular no aumento de faltas às aulas e nas dificuldades de realização das tarefas escolares.

Com o objetivo de avaliar o efeito do consumo de álcool sobre o desempenho escolar, Carrell et al. (2011) investigam o papel da lei da idade mínima para beber nos Estados Unidos. Os autores fazem uso de regressão descontínua e os microdados de discentes da Air Force Academy (USAFA) no período de 2000 a 2006, encontrando que a utilização de bebida alcoólica provoca reduções significativas no desempenho acadêmico, especialmente para os alunos com melhor performance. Nessa mesma linha, usando o método de regressão descontínua e os dados de estudantes — que ingressaram no curso com 18 ou 19 anos de idade — da Universidade de Oregon entre 1999 e 2007, Lindo et al. (2013) examinam o efeito do acesso legal às bebidas alcoólicas no desempenho acadêmico. Os resultados centrais da pesquisa indicam que as notas dos discentes são inferiores aos níveis esperados em cerca de 0,03 desvios-padrão, quando se mede o efeito do acesso legal às bebidas. A consequência dessa lei é maior para as mulheres, homens com baixos níveis de habilidades e jovens mais desfavorecidos economicamente.

Outros resultados sobre o uso de substâncias que causam dependência — como cigarro, maconha e *ecstasy* — na realização acadêmica podem ser verificados também na literatura internacional. Através de uma pesquisa de acompanhamento inicial de 7.124 adolescentes de quatro distritos da área metropolitana de Minneapolis-Saint Paul nos Estados Unidos entre os anos de 1979 e 1985, Pirie et al. (1988) identificam que os alunos que abandonam os estudos possuem elevadas taxas de uso frequente do tabaco. Enquanto que a taxa de prevalência de fumantes em idade escolar que estavam matriculados em 1985 era em torno de 16,6% entre os meninos e de 22,5% entre as meninas, essa taxa era superior a 77% entre os indivíduos que haviam abandonado os estudos independentemente do gênero.

Em Yamada et al. (1996), a análise dos hábitos prejudiciais incluem além do consumo de bebidas alcoólicas, o uso de maconha por parte de alunos no ensino médio. Considerando os dados do NLSY e o modelo probit, os autores concluem que existem significantes efeitos adversos da adoção desses comportamentos prejudiciais à saúde nos resultados educacionais, haja vista a redução na probabilidade de conclusão do ensino médio. Para Yamada et al. (1996), a utilização abusiva de substâncias ilícitas, como a maconha, implica em grandes perdas na aquisição de capital humano, sendo altamente recomendado programas preventivos que desencorajem a adoção de hábitos não saudáveis.

Baseado nos dados longitudinais de estudantes da Califórnia e de Oregon e na regressão logística, Ellickson et al. (2001) verificam que o uso prematuro do cigarro por parte dos discentes entre a 7ª série e a 12ª série da educação básica gera maiores dificuldades acadêmicas e tendências de uso regular de substâncias nocivas e comportamentos delinquentes. No comparativo com os não-fumantes, os fumantes prematuros têm ao menos três vezes mais chances de usar de forma regular tabaco e maconha, vender drogas e experimentar gravidez e paternidade precoce, bem como cinco vezes mais chances de abandonar os estudos.

Por meio dos dados do NELS de 1992, os achados de Jeynes (2002) mostram que os padrões de adolescentes acerca do consumo, especialmente, de álcool e cigarro têm impactos negativos e estatisticamente significativos sobre o desempenho dos alunos em testes padronizados de matemática, leitura, ciências e estudos sociais. No que concerne ao uso do cigarro, por exemplo, os alunos fumantes exibem uma nota no cômputo total inferior em 0,15 pontos aos dos alunos que não adotam esse tipo de hábito, tendo em vista os controles socioeconômicos e adoção de outras posturas prejudiciais à saúde. Bergen et al. (2005) encontram, a partir de dados longitudinais para estudantes australianos, que a percepção de insucesso escolar (captada pelo um indicador de performance acadêmica percebida pelo o aluno) que persiste entre 13 anos e 15 anos de idade está significativamente associada com a utilização de álcool, tabaco e maconha entre os adolescentes.

Além do mais, Martins & Alexandre (2009), com dados do National Survey of Drug Use and Health (NSDUH) e Youth Risk Behavior Survey (YRBS) e regressão logística, mostram que os não usuários de drogas — álcool, tabaco, maconha e *ecstasy* — apresentam as melhores notas na escola quando confrontados com os usuários desses itens. De acordo com o citado trabalho, em termos gerais, o uso de *ecstasy* reduz



a realização acadêmica dos estudantes em uma magnitude superior a observada pelo o uso de maconha, álcool e tabaco.

2.3. Literatura Nacional

No Brasil, conforme D. C. Machado (2008), as investigações sobre os efeitos da saúde em resultados escolares estão em fase inicial. Gomes-Neto et al. (1997) introduzem essa discussão, utilizando dados longitudinais do projeto EDURURAL da década de 80, para menos de 400 estudantes de áreas rurais dos estados do Piauí, Ceará e Pernambuco. Apesar dos autores demonstrarem o papel de destaque da boa nutrição na realização acadêmica e no atendimento escolar, eles reconhecem que os resultados devem ser vistos com cautela devido ao restrito desenho amostral.

Nessa linha, D. C. Machado (2008) investiga a relação entre a saúde infanto-juvenil, medida por indicador padronizado de altura-idade, e o ingresso na escola em idade considerada legalmente inadequada. Para isso, a Pesquisa sobre Padrões de Vida (PPV) 1996–1997, realizada pelo o Instituto Brasileiro de Geografia e Estatística (IBGE) para 10 estratos geográficos de localidades das regiões Nordeste e Sudeste do país, é usada concomitantemente com um modelo de resposta binária probit para uma amostra final de 2.712 observações. Esse trabalho ratifica o papel das boas condições de saúde e nutrição sobre a performance escolar.

Observando a pertinência e a escassez de estudos que contemplem a análise da relação entre exposição dos alunos aos fatores de risco à saúde (como sobrepeso, cigarro e bebida alcoólica) e os resultados escolares na literatura nacional, a presente pesquisa pretende contribuir com essa discussão — mesmo sabendo das limitações impostas pelos dados disponíveis no Brasil sobre o tema, que reduzem as possibilidades de uma identificação mais precisa de causa e efeito.

3. 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 no Brasil. Para testar essa hipótese, 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 pelas 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, assim como teste de robustez dos resultados com controles amostrais específicos e diferentes alternativas de especificação dos modelos.

Para Park & Kang (2008) e Jorges, Reinhold & Salm (2011), é preciso atenção na questão da endogeneidade que envolve a medição do efeito causal das relações entre as condições de saúde e a educação, 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, indivíduos mais saudáveis podem priorizar 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. Reforçando esse argumento, Becker & Mulligan (1997) sublinham que a educação tem um efeito causal sobre a saúde, sempre que a escolaridade repercute nas preferências dos agentes em um dado intervalo de tempo. Por isso, os trabalhos que direcionam a análise da saúde sobre a educação em geral restringem a amostra para crianças e adolescentes para amenizar esse tipo de problema (Eide & Showalter, 2011).

Considerando os aspectos mencionados acima, 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 completamente definida, pois o status de saúde da criança precede ao resultado na escola. Portanto, não existe ou é pouco provável 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 efeitos da exposição dos fatores de risco à saúde sobre o desempenho escolar.

É importante destacar que podem existir fatores não-observados específicos da criança (como problemas disciplinares), de seus familiares (como ambientes desestruturados) e da escola (como externalidades negativas) que afetam ambos indicadores de exposição aos fatores de risco à saúde e de desempenho escolar (Koch & Ribar, 2001). Na tentativa de controlar esses fatores, a última parte da seção de resultados reporta diferentes especificações para os modelos logit (estágio I) e PSM (estágio II).

Como destacado na equação (1), as condições de saúde da criança correspondem a um dos inputs da 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 proficiência² e taxas de rendimento escolar³), utiliza-se a distorção idade-série como indicador educacional.⁴

3.1. 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 (2), base para os primeiros resultados desta pesquisa, está descrita a seguir:

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

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 de 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), 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) segue distribuição logística, é possível determinar a probabilidade de que A_i^* seja inferior ou igual a um determinado limiar (j), isto é: $P(A_i^* \leq j) = G(\cdot)$. Assim, a equação (3) é desenvolvida para o caso em que dis_i é enquadrado numa

²Atualmente, 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.

³As taxas de rendimento escolar indicam o fluxo do aluno na escola. As informações de rendimento escolar são sumarizadas por indicadores de aprovação, repetência e abandono escolar.

⁴Esse tipo de estratégia é bastante comum na literatura, que usam medidas referentes ao ingresso tardio na escola e outras medidas de atraso escolar (Glewwe & Jacoby, 1995; Gomes-Neto & Hanushek, 1994; D. C. Machado, 2008; D. C. Machado & Gonzaga, 2007; Ribeiro & Cacciamali, 2012; Rios-Neto et al., 2002).



resposta binária, com: $dis_i = 0$, se $A_i^* \leq 0$; e $dis_i = 1$, se $A_i^* > 0$:

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

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

Os procedimentos adotados para o cálculo do efeito médio da exposição aos fatores de risco sobre a probabilidade de distorção idade-série nessa primeira abordagem estão descritos na equação (4):

$$EMD_k = n^{-1} \sum_{i=1}^n \left[G(\hat{\beta}_0 + \hat{\beta}_k + \mathbf{W}_i \hat{\mathbf{c}}) - G(\hat{\beta}_0 + \mathbf{W}_i \hat{\mathbf{c}}) \right], \quad (4)$$

em que \mathbf{W}_i é o vetor de covariadas com exceção da variável binária de exposição ao fator de risco k .

Como destaca Wooldridge (2010), o cálculo do efeito parcial médio⁵ (EPM) e do efeito parcial na média⁶ (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 através da equação (4) 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 .

3.2. Modelo PSM: efeito médio condicionado ao grupo de controle observável

Esta subseção descreve brevemente a abordagem *propensity score matching* (PSM), usada para o cálculo do efeito médio baseado na identificação de um grupo de controle com similar características observáveis (\mathbf{Z}_k) dos discentes expostos. Este método proposto por Rosenbaum & Rubin (1983) resumem as características de cada unidade em um único score 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. Os modelos empíricos na área da economia da saúde, como Cutler & Glaeser (2005), Park & Kang (2008), Cutler & Lleras-Muney (2010), Rees & Sabia (2011), Jorges et al. (2011) e Rees & Sabia (2014), mostram que a decisão do indivíduo por hábitos não saudáveis é influenciada, especialmente, pelo *background* familiar e atributos individuais e situacionais.

Tendo o suporte desses fatores e da função de produção educacional, a probabilidade condicional de exposição ao fator de risco k condicionado a um vetor de características observáveis é definida por

$$P(H_k = 1 | \mathbf{Z}_k) = G(\mathbf{Z}_k \boldsymbol{\lambda} + \epsilon), \quad (5)$$

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 , composto pelo perfil do aluno, da família, da escola e de covariadas específicas atreladas a decisão de exposição; G representa a função de distribuição acumulada. Supondo que G segue uma distribuição de probabilidade logística, pode-se estimar $P(\mathbf{Z}_k)$ através de um modelo de resposta qualitativa logit.

⁵ $EPM_j = n^{-1} \sum_{i=1}^n [g(\hat{\beta}_0 + \hat{\beta}_1 H_{1i} + \hat{\beta}_2 H_{2i} + \hat{\beta}_3 H_{3i} + \mathbf{X}_i \hat{\mathbf{b}}) b_j]$, sendo $g(\cdot) = \exp(\cdot) / [1 + \exp(\cdot)]^2$.

⁶ $EPPM_j = g(\hat{\beta}_0 + \hat{\beta}_1 \bar{H}_{1i} + \hat{\beta}_2 \bar{H}_{2i} + \hat{\beta}_3 \bar{H}_{3i} + \bar{\mathbf{X}}_i \hat{\mathbf{b}}) b_j$, sendo \bar{H}_{ki} e $\bar{\mathbf{X}}_i$ as médias amostrais das variáveis.

Após a obtenção do escore de propensão $\Pr(Z_k)$, o efeito médio do tratamento sobre o grupo exposto (τ_k) é estimado a partir da equação (6):

$$\tau_k = E_{\Pr(Z_{ki})|H_k=1} \left\{ E \left[dis_i | H_{ki} = 1, P(Z_{ki}) \right] - E \left[dis_i | H_{ki} = 0, P(Z_{ki}) \right] \right\}. \quad (6)$$

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 (6) é preciso utilizar algum algoritmo de pareamento, pois como $\Pr(Z_k)$ 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 Epanechnikov.⁷ 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 seção 5.2 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 $0 < P(H_k = 1 | Z_k) < 1$. Outra importante suposição presume que todas as diferenças relevantes entre os indivíduos expostos e não-expostos são capturadas por Z_k , onde é necessário a garantia, principalmente, da chamada hipótese de independência condicional (CIA): $Y(0), Y(1) \perp H_k | P(Z_k)$, 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, 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, conduzem uma investigação sobre a relação causal entre o tabagismo e o câncer de pulmão. Através da formulação do risco relativo aparente, os autores demonstram 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 indícios 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 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 (7):

$$\frac{P(Z_{ki}) [1 - P(Z_{kj})]}{P(Z_{kj}) [1 - P(Z_{ki})]} = \frac{G(Z_{ki}\lambda + \gamma\epsilon_i)}{G(Z_{kj}\lambda + \gamma\epsilon_j)}. \quad (7)$$

⁷Para maiores detalhes sobre as diferentes técnicas de pareamento, ver Rosenbaum (2010).



Se os indivíduos i e j possuem iguais Z_k , $G(\cdot)$ segue uma distribuição logística e $\epsilon \in (0, 1)$, Rosenbaum (2002) mostra que a equação (7) implica nos seguintes limites para as razões de chances:

$$\Gamma^{-1} \leq \frac{P(Z_{ki})[1 - P(Z_{kj})]}{P(Z_{kj})[1 - P(Z_{ki})]} \leq \Gamma, \quad (8)$$

com $\Gamma = e^Y$.

Essa abordagem também é conhecida como limites de Rosenbaum, já que a equação (8) 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.

4. 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 às 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 final⁸ desta pesquisa 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 da seção de resultados, com exceção para o cálculo específico da equação (6), já que o grupo de controle observável é identificado por *matching* e a diferença média é ponderada pela distribuição do *propensity score*.⁹

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

⁸As variáveis com a maior quantidade de itens com *missing values* foram a escolaridade do pai e da mãe dos discentes. A fim de ilustrar essa perda, do total da amostra original da PENSE, aproximadamente 17% (18.527 observações) e 23% (25.080 observações) dos alunos não informaram o nível de educação da mãe ou do pai. Em conjunto, os *missing values* dessas duas variáveis responderam por uma perda de 28% da amostra. Dado o papel dessas informações para definir o NSE do aluno e a indisponibilidade de outras variáveis na PENSE para captar essa dimensão, apenas as respostas válidas foram preservadas na análise. Ressalta-se que as características médias das variáveis de interesse foram mantidas mesmo após a limpeza do banco de dados.

⁹No 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.

unidades escolares selecionadas e as unidades terciárias de amostragem (UTA) por turmas dessas escolas (IBGE, 2013). O plano amostral da PENSE é complexo e com delineamento misto (diferentes números de estágios para seleção da amostra entre municípios de capitais e fora delas), necessitando ajustar os modelos econométricos, quando for o caso, às especificidades amostrais do banco de dados para evitar problemas de sub ou sobrestimação da variância das estimativas que afetam o processo de inferência (Pessoa, Silva & Duarte, 1997; Silva, Pessoa & Lila, 2002).

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.

4.1. Distorção idade-série

A distorção idade-série é um *outcome* educacional, relacionando-se diretamente com o abandono e a repetência escolar, bem como com o problema de evasão escolar e o ingresso tardio na escola.¹⁰ Para Glewwe & Jacoby (1995), 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. Outrossim, 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 (Glewwe & Jacoby, 1995).

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, o cálculo do indicador de atraso escolar (*dis*) é mostrado na equação (9):

$$dis_{is} = \begin{cases} 1, & \text{se } (I_{is} - I_s^e) \geq 1,5 \\ 0, & \text{c.c.} \end{cases}, \quad (9)$$

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 e mês 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 ($s = 9$). A equação (9) estabelece que crianças com diferenças positivas e maiores que 1,5 ano em $(I_{is} - I_s^e)$ são enquadradas com atraso escolar, de modo que estudantes com $(I_{is} - I_s^e) = 1$ mas que nasceram no segundo semestre do ano escolar não estão classificadas como atrasadas.¹¹ Com base nesse indicador, a Figura 1 exibe a taxa de distorção idade-série, restringida para a amostra final deste estudo (detalhes na seção 4), entre as escolas das redes privada, pública e total por agrupamento regional.

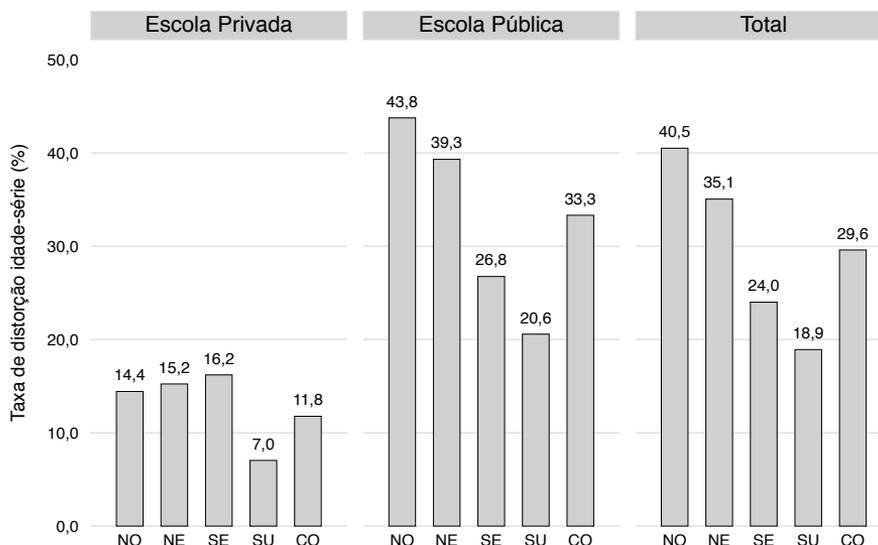
Verificando esses dados, é perceptível que as taxas de atraso escolar das escolas públicas, dado que as escolas privadas respondem por 20% dos alunos na amostra, definem o comportamento do nível de distorção idade-série nas regiões brasileiras. As escolas públicas, de modo geral, têm um escore três vezes pior do que o constatado nas escolas privadas, com exceção da região Sudeste que apresenta uma razão

¹⁰ Além disso, existem evidências na literatura, como em Gomes-Neto & Hanushek (1994), Gomes-Neto et al. (1997) e T. M. Soares (2003), que indicam uma relação negativa entre as estimativas de distorção da idade na escola e o baixo desempenho em testes padronizados.

¹¹ Esse critério é baseado no dicionário de indicadores do INEP (2004), que considera uma margem de um ano de tolerância.



Figura 1. Taxa de distorção idade-série no 9º ano do Ensino Fundamental por dependência administrativa e grandes regiões do Brasil a partir da amostra final deste trabalho.



NO=Norte, NE=Nordeste, SE=Sudeste, SU=Sul, CO=Centro-Oeste

Nota: Amostra composta de 66.844 estudantes do 9º do ensino fundamental com representatividade na população de 1,9 milhão de observações.

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

menor que dois. O destaque fica por parte das regiões Norte e Nordeste que detêm os piores indicadores na rede pública e no total. Contudo, a região Sudeste entre as escolas particulares possui relativamente uma elevada distorção idade-série. Tais fatos sinalizam que o atraso escolar potencialmente se mostra relacionado aos fatores socioeconômicos, tendo em vista as já conhecidas disparidades socioeconômicas entre os estudantes de escolas públicas e privadas e entre as regiões Sul/Sudeste e Norte/Nordeste.

4.2. Indicadores de exposição aos fatores de risco comportamentais

Os tipos de fatores de risco modificáveis à saúde contemplados neste estudo são o uso do cigarro, o consumo de bebidas alcoólicas e o sedentarismo/má alimentação (medido pelo sobrepeso). Em relação ao último indicador, destaca-se que a PENSE apenas possui informações sobre as condições correntes do consumo de alimentos e de práticas de atividades físicas. Por isso, que as informações sobre o índice de massa corporal¹² (*IMC*) são usadas como *proxies* para o estoque de hábitos não saudáveis referentes 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 físicas.¹³ Essa abordagem que obedece a um padrão empírico e específico para cada idade e

¹²O cálculo desse indicador é dado por: $IMC = \text{peso} / (\text{altura}^2)$, com o peso expresso em quilograma (kg) e a altura em metro (m).

¹³Abrevaya & Tang (2010) realçam que as condições nutricionais das crianças são determinadas primariamente por fatores genéticos e pelo ambiente comum (hábitos saudáveis, questões locacionais, *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.

gênero, dentro de intervalos quantílicos definidos, é recomendada pela OMS para crianças e adolescentes (Kuczmarski et al., 2002). Os limites (valores mínimos) do *IMC* que definem os estudantes com sobrepeso para a amostra considerada e o critério de classificação mencionado estão expressos a seguir:¹⁴

Meninas: com 13 anos ou menos, 24,89 kg/m²; com 14 anos, 25,31 kg/m²; com 15 anos, 25,92 kg/m²; com 16 anos, 26,95 kg/m²; com 18 anos, 27,74 kg/m²; com 18 anos ou mais, 29,30 kg/m²;

Meninos: com 13 anos ou menos, 24,77 kg/m²; com 14 anos, 24,80 kg/m²; com 15 anos, 25,39 kg/m²; com 16 anos, 26,37 kg/m²; com 18 anos, 27,46 kg/m²; com 18 anos ou mais, 27,73 kg/m².

Quanto às medidas de exposição ao consumo de álcool e do cigarro, o questionário do estudante aplicado na PENSE dispunha de informações sobre a idade de primeiro contato com tais produtos e a indicação do consumo deles nos últimos 30 dias. Os alunos que declararam 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 acima da idade teórica para o 9º ano do ensino fundamental. Este trabalho usa duas medidas indicativas sobre esses comportamentos, uma relacionada ao tempo de experiência da utilização do produto e outra que combina essa experiência com o uso recente.¹⁵ Assim, esses indicadores estão divididos em:

Definição 1: experiência com o produto com idade igual ou anterior aos 14 anos, doravante denotada apenas por *experiência*;

Definição 2: experiência combinada com o uso nos últimos 30 dias do período da aplicação do questionário da PENSE, denominada a partir de agora de *experiência e uso*.

A Figura 2 exibe a proporção de alunos expostos aos fatores de risco à saúde em todos os casos avaliados, para a amostra final deste estudo. Os dados revelam um número preocupante de crianças expostas ao uso de bebidas alcoólicas com idade inferior ou igual a 14 anos de idade (66% na Definição 1 e 24% na Definição 2). Quanto ao cigarro, 18% dos alunos declaram o consumo do produto com 14 anos ou menos de idade e 4% destes consumiram o produto no último mês. Por sua vez, o excesso de peso contempla aproximadamente 16% dos indivíduos avaliados.¹⁶

4.3. Demais variáveis

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 escolar,¹⁷ indicador de infraestrutura e média do nível socioeconômico dos alunos); fatores locais (capital e grande região geográfica).

Consonante à estratégia empírica adotada, além do vetor X , o vetor de características observáveis Z_k é usado na estimação do *propensity score* para cada fator de exposição k . O vetor Z_k é composto pelo vetor X e por variáveis situacionais relativas ao ambiente familiar e escolar, a saber:

¹⁴Como 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.

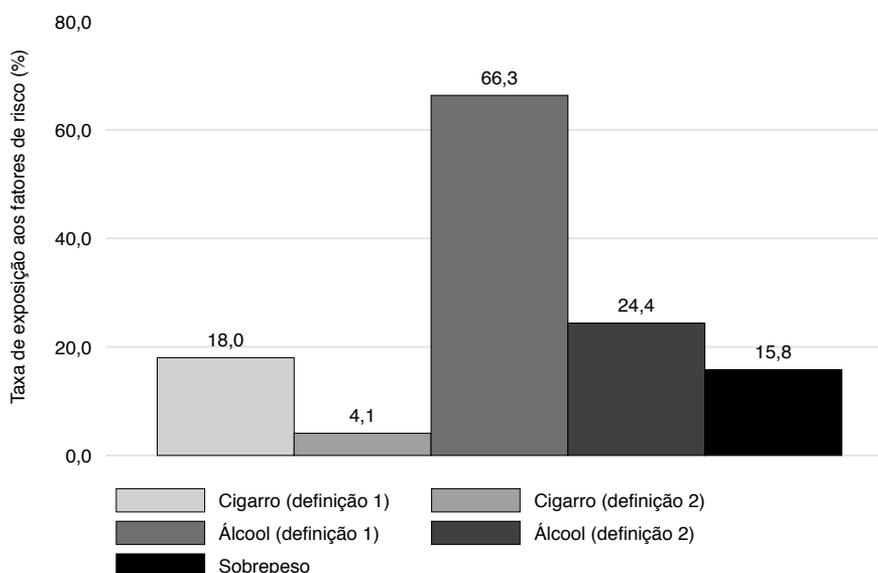
¹⁵Entende-se como uso recente, o consumo do produto nos últimos 30 dias de referência da PENSE.

¹⁶O cálculo dos limites do *IMC* para a definição do excesso de peso é desenvolvido em relação aos 98.080 estudantes que possuíam dados sobre altura e peso, por isso que a frequência relativa de indivíduos acima do quantil 85 é superior a 15%, uma vez que a amostra final possui uma quantidade menor de observações.

¹⁷O conselho escolar presume a participação dos professores, funcionários, pais e alunos no processo decisório da gestão da escola em conjunto com a direção escolar.



Figura 2. Taxa de exposição aos fatores de risco à saúde por parte de estudantes do 9º ano do Ensino Fundamental no Brasil a partir da amostra final deste trabalho.



Nota: Amostra composta de 66.844 estudantes do 9º do ensino fundamental com representatividade na população de 1,9 milhão de observações.

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

- 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.

O nível socioeconômico (NSE) e o indicador de infraestrutura escolar (IEE) 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 para estimar traços latentes, como habilidades cognitivas, qualidade de infraestrutura, nível socioeconômico, etc. (van der 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).

A seguir são apresentadas breves explicações sobre a construção dessas duas variáveis:

NSE representa um indicador que mede o patrimônio e outras condições socioeconômicas dos alunos. Esse indicador está baseado em J.F. Soares & Andrade (2006) e Alves & Soares (2012) que usam o modelo da Teoria da Resposta ao Item (TRI), visto que bases de dados (como a da PENSE) não dispõem de informações sobre o rendimento familiar, bem como o NSE consegue sintetizar em uma única medida uma série de itens associados as condições socioeconômicas. Todos os itens que medem informações diretas e indiretas sobre a escolaridade dos pais e a renda da família são considerados na estimação do NSE. No total, dez itens são incorporados na mensuração do NSE do aluno (descritos com mais detalhes na Tabela A-1 no Apêndice), com destaque para as variáveis de escolaridade dos pais, de patrimônio e de serviços disponíveis na casa da criança. Segundo as informações presentes nessa tabela, a grande parte dos estudantes (65%) só dispõe de um banheiro

em casa, 42% das mães e 47% dos pais nunca estudaram ou não concluíram o ensino fundamental e apenas 10% das casas possuem empregado doméstico. Os parâmetros estimados pela TRI para cada um dos itens pontuam que aqueles relacionados a disponibilidade de computador e acesso à internet em casa são os itens com maior peso de discriminação entre indivíduos com alto e baixo NSE, enquanto a posse de moto é o item que menos distingue. O modelo logístico policotômico de dois parâmetros da TRI (unidimensional e com um sistema de crédito parcial generalizado) é o escolhido para estimação do NSE, devido ao fato dos itens referentes à escolaridade e a quantidade de banheiros terem mais de duas alternativas ordenadas possíveis. O NSE está normalizado¹⁸ para o intervalo entre 0 e 1. Todavia, o valor estimado desse traço latente originalmente se situa entre $-\infty$ e $+\infty$.

IEE considera onze perguntas existentes no questionário da escola sobre a disponibilidade de recursos (equipamentos, ambientes e serviços). Esses itens são: biblioteca, laboratório de informática, computadores em sala de aula, internet na escola, cantina, quadra de esportes, quadra de esporte com cobertura, pista de corrida, piscina e vestiário (ver Tabela A-2 no Apêndice). Os recursos com maior escassez entre as escolas da amostra são pista de corrida (2,6%), piscina (14%) e acesso a computadores em sala de aula (28%), enquanto os mais recorrentes são biblioteca (91%), laboratório de informática (87%) e acesso à internet (86%). Quanto aos itens que mais discriminam as escolas em termos de IEE, em consonância com os parâmetros estimados, destacam-se a existência de quadra de esportes e piscina. Por outro lado, itens que tratam da existência de computadores na escola e cantina são os que menos discriminam. O traço latente para o IEE é calculado usando a versão dicotômica unidimensional da TRI com dois parâmetros, tendo por base Soares-Neto, de Jesus, Karino & de Andrade (2013), no qual o indicador normalizado (transformação idêntica a promovida no NSE) varia de 0 a 1.

5. RESULTADOS

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.

5.1. Exploração inicial: Determinantes da atraso escolar

As estimativas dos determinantes do atraso escolar, controlando para os atributos dos alunos (incluindo os fatores de risco comportamentais) e de suas famílias, as características da escola e os fatores locacionais, estão exibidos na Tabela 1. Constam nessa tabela nove especificações de modelo, que se diferenciam pela inclusão e interação entre a exposição do aluno aos fatores de risco à saúde modificáveis. Inicialmente, os sinais, a magnitude e a significância estatística das variáveis explicativas mostram o mesmo padrão para todas as especificações, onde de forma geral os modelos conseguem classificar corretamente 73% dos casos, isso assumindo um *cut-off* padrão de 50% na probabilidade de atraso escolar.

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 corroboram a literatura empírica

¹⁸A normalização é realizada pela seguinte expressão:

$$\theta_{iN} = \frac{\theta_i - \theta_{\min}}{\theta_{\max} - \theta_{\min}},$$

onde θ_{iN} representa o traço latente normalizado do indivíduo $i \in (0, 1)$ e $\theta_i \in (-\infty, +\infty)$ é o traço latente do indivíduo i .

Tabela 1. Resultados da estimação do modelo logit para a probabilidade de atraso escolar dos estudantes no Brasil 2012.

Covariadas	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Atributos do aluno e da família									
Gênero (homem = 1)	0,5195***	0,5331***	0,5188***	0,5221***	0,5381***	0,5291***	0,5294***	0,5392***	0,5394***
Raça (Branca = 1)	-0,1548***	-0,1605***	-0,1597***	-0,1566***	-0,1568***	-0,1557***	-0,1558***	-0,1553***	-0,1554***
Trabalha	0,5054***	0,5236***	0,5502***	0,5278***	0,4826***	0,4930***	0,4926***	0,4745***	0,4747***
Núcleo familiar biparental	-0,4014***	-0,4142***	-0,4351***	-0,4223***	-0,4011***	-0,3920***	-0,3922***	-0,3963***	-0,3965***
NSE do aluno	-1,4456***	-1,4988***	-1,4506***	-1,4522***	-1,5065***	-1,4744***	-1,4745***	-1,5018***	-1,5028***
Tamanho da família	0,0801***	0,0838***	0,0820***	0,0804***	0,0809***	0,0813***	0,0814***	0,0800***	0,0801***
Atributos da escola									
Total de matrículas (em mil)	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***	-0,2522***	-0,2513***	-0,2536***	-0,2531***	-0,2575***	-0,2574***	-0,2539***	-0,2547***
Índice de infraestrutura (IEE)	-0,7190***	-0,7121***	-0,7072***	-0,7122***	-0,7224***	-0,7183***	-0,7181***	-0,7213***	-0,7219***
NSE da escola	-2,1671***	-2,1781***	-2,1082***	-2,1257***	-2,1595***	-2,2067***	-2,2081***	-2,1679***	-2,1664***
Fatores regionais									
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
Exposição aos fatores de risco									
Experiência: cigarro (def.1)	0,4585***					0,4046***	0,2921***		
Experiência: álcool (def.1)		0,2553***				0,1611*	0,1552*		
Sobrepeso			0,0848*			0,0805*	0,0804*	0,0826*	0,0818*
Experiência e uso: cigarro (def.2)				0,5566***				0,2834***	0,5218***
Experiência e uso: álcool (def.2)					0,5423***			0,5081***	0,5201***
Experiência: cigarro e álcool							0,1208		
Experiência e uso: cigarro e álcool									-0,3017***
Constante									
Constante	1,3905*	1,3098	1,4317*	1,4337*	1,3632*	1,2948	1,2988	1,3451*	1,3423*
Observações	66.844	66.844	66.844	66.844	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	1.897.670	1.897.670	1.897.670	1.897.670
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

Notas: As estimativas são desenvolvidas considerando, além do peso amostral, a estratificação e a UPA do plano amostral.

* *p*-valor < 0,10; ** *p*-valor < 0,05; *** *p*-valor < 0,01.

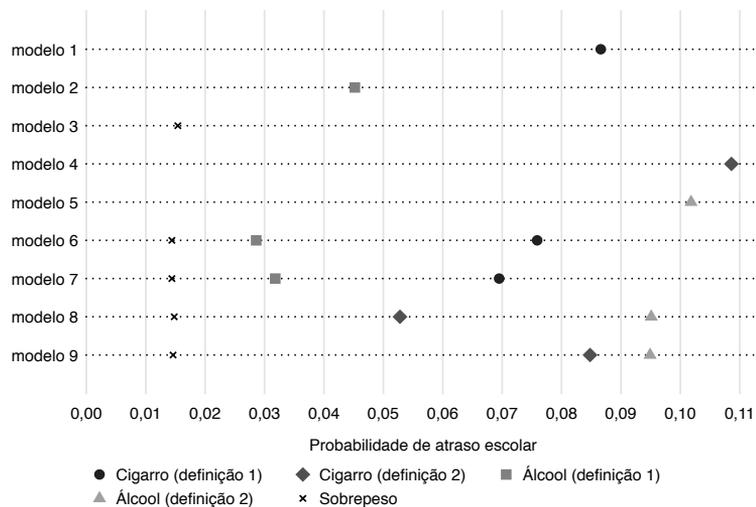
Fonte: Elaboração própria a partir dos microdados da PENSE 2012.

internacional e nacional, que enfatizam o papel do perfil socioeconômico do discente no rendimento escolar. Observando os efeitos marginais¹⁹ na Tabela A-4 do Apêndice estimados para esse vetor de variáveis, o NSE do aluno tem uma importante contribuição na probabilidade de atraso escolar. Um incremento unitário no indicador socioeconômico resulta em uma redução em torno de 26% na probabilidade de estar atrasado na escola.

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. Já o tamanho da escola, ditado pelo total de matrículas, não é estatisticamente significativo. O efeito marginal²⁰ (*Emg*) do IEE é de aproximadamente -13% na probabilidade de atraso escolar e o NSE do conjunto dos alunos da escola, os chamados *peer-effects*, possui uma relação importante com o desempenho do discente ($Emg \approx -38\%$). 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, $-40\% < Emg < -38\%$, e pior infraestrutura, $Emg \approx -33\%$, do que as escolas da rede particular de 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 as altas variâncias dessas estimativas quando se pondera o desenho amostral do *survey* em análise.

Dedicando agora para uma primeira análise da relação entre os fatores de risco comportamentais à saúde e o desempenho na escola, identifica-se uma relação direta entre o baixo indicador educacional do aluno e à exposição aos fatores de risco. Pelos dados da Tabela 1, crianças expostas às bebidas alcoólicas, ao cigarro e ao excesso de peso possuem mais chances de estarem atrasadas na escola. A princípio essa relação corrobora a hipótese dos achados empíricos sobre as condições de saúde física e psicológica vistos na literatura internacional. Para melhor dimensionar os efeitos médios desses hábitos prejudiciais à saúde sobre a performance na escola, a Figura 3 ilustra tais efeitos para cada uma das nove diferentes especificações.

Figura 3. Efeito médio dos fatores de risco comportamentais à saúde — não condicionado ao contrafactual observável.



Nota: Todos os coeficientes são significativos estatisticamente a pelo menos 10%.

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

¹⁹O efeito marginal para variáveis discretas foi calculado pela equação (4) e para as variáveis contínuas pelo o efeito parcial médio (EPM).

²⁰A Tabela A-4 no Apêndice exibe em detalhes as estimativas de efeito marginal de todas as covariadas.



Os alunos expostos ao excesso de peso, ao cigarro e as bebidas alcoólicas, de acordo com o modelo (9) na Figura 3, apresentam respectivamente uma probabilidade, em média, de atraso escolar superior em 1,5%, 8,5% e 9,5% no comparativo com a 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 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 em diferentes níveis socioeconômicos? Assim, a Figura 4 expressa as diferenças na probabilidade de atraso escolar provocado pela exposição aos fatores de risco, mediante 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.

Por meio da Figura 4, 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 4(a), onde estudantes usuários do 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 as bebidas alcoólicas, onde essa diferença de resultado em decorrência desse tipo de consumo diminui gradativamente na medida em que o background socioeconômico melhora — o efeito para o nível mínimo de NSE é de 11,5% e para o maior nível de 6,5%. Em relação ao excesso de peso na Figura 4(c), as diferenças entre as probabilidades de atraso também são superiores nos níveis iniciais de NSE, em que, por exemplo, o efeito médio da exposição é de 1,8%, no primeiro nível, e de 1% no último.

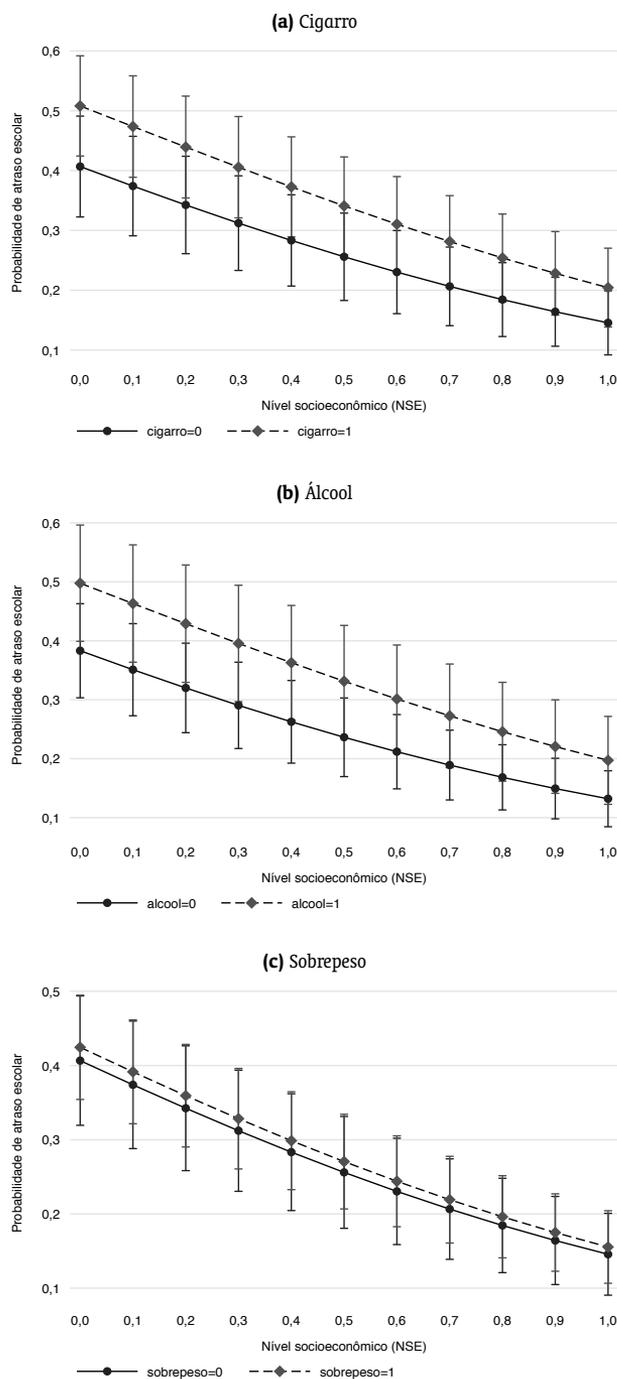
5.2. Efeito médio da exposição usando PSM

Esta subseção explicita os resultados obtidos pela 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 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* $\Pr(Z_k)$ e, em seguida, utilizar algoritmos de pareamento para o cálculo do τ_k . A Tabela 2 exhibe os coeficientes usados na estimação das probabilidades de exposição ao fator de risco à saúde²¹ k (com os números entre parênteses indicando a definição usada para o consumo de álcool e cigarro, em conformidade com a descrição na parte sobre a base de dados).

²¹ 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). A Figura A-1 no Apêndice ilustra que os termos residuais das equações para os determinantes de cada hábito não possuem relações bem definidas, quando se considera o ajustamento das curvas por uma regressão não-paramétrica a partir da abordagem *Linear Locally Estimation*.

Figura 4. Efeito dos fatores de risco comportamentais à saúde por NSE no Brasil 2012 — não condicionado ao contrafactual observável.



Nota: Todos os coeficientes são significativos estatisticamente a pelo menos 1%. Estimativas desenvolvidas através dos parâmetros do modelo (9) na Tabela 1.

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



Tabela 2. Resultados da estimação dos coeficientes para o *propensity score* por tipo de exposição ao fator de risco à saúde da criança no Brasil 2012.

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

Notas: As estimativas são desenvolvidas considerando, além do peso amostral, a estratificação e a UPA do plano amostral.

^a Em consonância com 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.

* p -valor < 0,10; ** p -valor < 0,05; *** p -valor < 0,01.

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

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 as 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 taxa pode ser reflexo 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. Outro indicativo da existência de um área de suporte pode ser visualizada na Figura A-2 no Apêndice, elaborada para cada um dos fatores de risco k em suas diferentes definições. Na citada ilustração, as curvas de densidade do *propensity score* entre os dois grupos para cada um desses fatores têm uma região de sobreposição, ratificando a possibilidade de compatibilização entre as unidades para o processo de pareamento.

As variáveis na Tabela 2 relativas ao núcleo familiar do aluno, trabalho, tamanho da escola, taxa de prevalência do fator de risco na escola e fatores situacionais específicos da família perante ao sobrepeso (captado pela prática alimentar) e uso de bebida alcoólica e cigarro apresentam 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 à 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. Em suma, os sinais dos efeitos das variáveis contextuais das escolas estão condizentes com outros trabalhos, a exemplo de Veronneau, Trempe & Paiva (2014), que apresentam evidências sobre o papel de relevo dos colegas na formação psicossocial da criança e dos efeitos desfavoráveis de amizades com pares desviantes.

As variáveis situacionais da família também são importantes para a tomada de decisão da criança, visto que filhos de pais fumantes são mais propensos à utilização do cigarro. Quanto ao 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 ter esse tipo de postura. Por sua vez, as 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 a 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, em relação ao atraso escolar dos discentes.



Tabela 3. Efeito médio sobre o grupo exposto em relação ao atraso escolar no Brasil 2012, 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

Nota: 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.

* *p*-valor < 0,10; ** *p*-valor < 0,05; *** *p*-valor < 0,01.

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

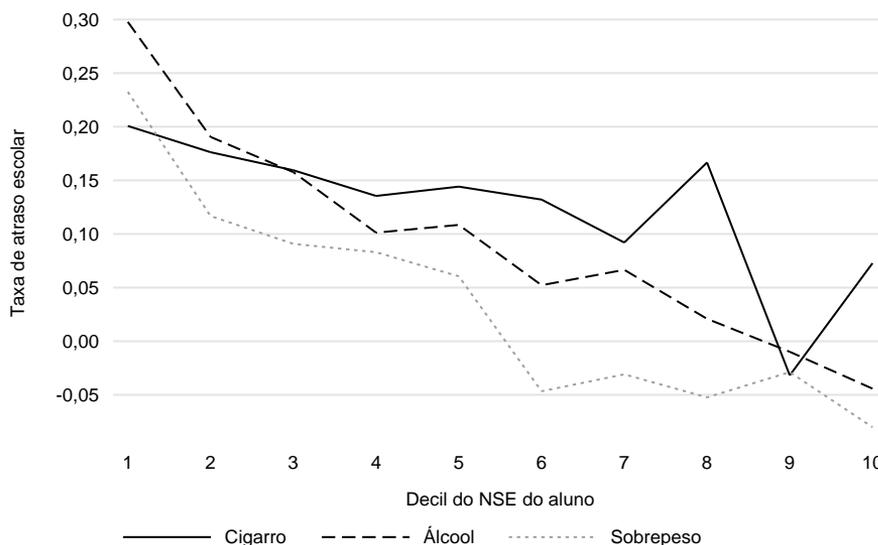
Os resultados na Tabela 3 são claros quanto a direção dos efeitos dos fatores de risco modificáveis sobre o baixo desempenho escolar. As diferenças no resultado educacional entre os grupos exposto e de controle ostentam estimativas estatisticamente significativas à 1%, bem como os sinais dos coeficientes são positivos com as magnitudes praticamente invariantes as diferentes formas de pareamento. Com isso, o efeito para o cigarro é de aproximadamente 8,5% na primeira definição, e superior a 12% na segunda. As estimativas 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%).

Na comparação entre os efeitos calculados diretamente pela estimação da função de produção educacional (seção 5.1) 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 5). 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 é negativamente correlacionado com o *background* socioeconômico.

Assim, as crianças mais pobres expostas aos fatores examinados são mais propensas a um menor resultado escolar. Portanto, as evidências sugerem que programas 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

Figura 5. Efeito médio por decis do nível socioeconômico do aluno computado pela abordagem PSM-V com reposição (CR) no Brasil 2012.



Nota: Efeitos para cigarro e álcool referentes à Definição 2.

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



a relevância da redução da prevalência dos fatores de risco modificáveis na população mais jovem, particularmente 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.

5.3. 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 exhibe 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^+).

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)	Álcool (1)	Álcool (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

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.

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.

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

O valor de Γ reflete a suposição sobre problemas de endogeneidade presentes nas chances de atribuição da exposição motivado por 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 autosseleção endógena do estado de exposição (Rosenbaum, 2002).

Pela Tabela 4, 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 a 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.

No caso do cigarro, por exemplo, os fatores não-observados deveriam ser pelo menos 1,6 vezes maiores nos alunos expostos a esse fator de risco à saúde do que nos alunos não-expostos. Só assim, a probabilidade de exposição ao cigarro poderia ser considerada sobrestimada. 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 *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 trazem indícios sobre o papel de uma postura inadequada à saúde infantil no aumento do atraso escolar.

5.4. Testes de robustez

É válido registrar que apesar das diferentes especificações e modelos estatísticos utilizados nesta pesquisa, os efeitos calculados da exposição aos fatores de riscos comportamentais à saúde ainda são susceptíveis aos problemas de endogeneidade e de má especificação do modelo (viés de variáveis omitidas, tais como questões disciplinares, temperamentais ou de *background* familiar desestruturado). Dessa maneira, como ressaltam Koch & Ribar (2001), se existem características relevantes que não estão sendo controladas nas análises, elas podem levar a uma relação espúria entre a exposição de um fator de risco (como o consumo de bebida alcoólica) e a progressão escolar.

A fim de avaliar a robustez das evidências encontradas, nesta subseção são desenvolvidos testes para verificar se a hipótese de que o mecanismo da exposição via saúde sobre o aumento do atraso escolar pode ser válida ou não. Visto que, variáveis omitidas no prisma comportamental — como indisciplina, delinquência e rebeldia dos jovens — talvez possam ser mais relevantes de explicar essas relações encontradas do que o impacto estritamente via saúde. Essa preocupação torna-se pertinente pela omissão desse controle e, portanto, os coeficientes associados à exposição ao cigarro ou ao álcool podem estar captando possíveis influências do lado socioemocional dos estudantes.



A Tabela 5 apresenta diferentes alternativas de especificações para o modelo de resposta binária logit, que tem como variável dependente a dummy de atraso escolar. Doze especificações são realizadas, com as seguintes características:

- do modelo (1) ao (4), controles para efeitos fixos da turma e da escola;
- do modelo (5) ao (11), testes para atributos comportamentais omitidos, com um modelo irrestrito e restrito para alunos não-expostos, incluindo a prevalência de discentes expostos ao fator de risco à saúde na turma do estudante;
- modelo (12), teste para variáveis de comportamento omitidas, com um modelo que inclui *proxies* para indisciplina e mau comportamento na escola por meio de três indicadores binários em nível do estudante: costume de faltar propositalmente as aulas, envolvimento em brigas na escola e desconhecimento dos pais acerca das atividades que o aluno desenvolve no tempo livre.

Comparando os coeficientes da Tabela 5 com aqueles reportados na Tabela 1, percebe-se que os achados sobre a relação entre a exposição individual aos fatores de risco à saúde e o atraso escolar não se alteram substancialmente. Tais fatos reforçam as hipóteses teóricas e empíricas que estudantes obesos e, sobretudo, que são expostos ao consumo de bebidas alcoólicas e cigarro tendem a apresentar maiores dificuldades na aquisição de capital humano.

Especificamente, as especificações 5, 6 e 7 são realizadas para uma amostra restrita a alunos não-expostos aos fatores de risco. Nesse caso, testa-se a hipótese que estudantes de escolas com maior prevalência de discentes expostos a um dado fator de risco à saúde poderiam apresentar maior distorção idade-série, o que sugeriria a importância da dimensão comportamental frente aos aspectos relativos à saúde individual. Destarte, os coeficientes da prevalência de exposição ao cigarro e ao álcool na turma são insignificantes estatisticamente. Enquanto que a taxa de estudantes com sobrepeso mantém relação negativa com o atraso escolar. Por essas estimativas, a hipótese atrelada ao nível comportamental (como delinquência, rebeldia ou indisciplina) parece não ser o mecanismo que estaria explicando de maneira estrita o maior atraso escolar.

Ao incluir essas prevalências de exposição aos fatores de risco na turma para a amostra final irrestrita (do modelo 8 ao 11), nota-se que tanto o sinal quanto a magnitude dos parâmetros da exposição ao fator de risco à saúde no nível do aluno permanecem praticamente inalterados.

Ademais, a última especificação apresentada na Tabela 5 inclui variáveis explicativas que capturam a indisciplina do aluno. Verifica-se que estudantes com faltas propositais as aulas ou aqueles em que os pais ou responsáveis desconhecem as atividades realizadas pelas crianças no tempo livre possuem maior razão de chances de estarem com distorção idade-série, evidenciando que a indisciplina e o *background* familiar problemático se correlacionam positivamente com o atraso escolar. Assim, mesmo ao controlar por essa dimensão, alunos expostos ao cigarro e a bebida mantêm a tendência de maior atraso na escola. Destaca-se que nessa última especificação, o coeficiente para obesidade não apresentou significância estatística. Com base nesses resultados, nota-se que os fatores socioemocionais dos alunos podem ser uma importante dimensão que define problemas de desempenho escolar, mas que não se pode desconsiderar um possível papel do componente saúde e fisiológico.

Nessa mesma direção, a Tabela 6 reporta especificações alternativas para os resultados do modelo usando o contrafactual observável dos alunos expostos aos fatores de riscos considerados. Nesse caso, elas são desenvolvidas pela combinação de uma regressão ajustada com pesos baseados no *propensity score*, tal como realizado por Hirano & Imbens (2001). Em todas as especificações são incluídas variáveis que capturam o nível de indisciplina do estudante, com diferenças no controle para efeito fixo ou não da turma.

Os resultados apresentados na Tabela 6 mostram que as estimativas calculadas não são sensíveis nem no sinal e nem na magnitude para controle dos efeitos fixos da turma, bem como as variáveis atreladas ao nível de indisciplina dos discentes. Muito embora, as evidências encontradas se mantiveram para os diferentes testes efetuados, elas precisam ainda ser vistas apenas como indicativas de que alunos

Tabela 5. Análise de robustez dos resultados referentes à estimação dos efeitos dos fatores de risco comportamentais à saúde sobre o atraso escolar – Modelo Logit.

Covariadas	(1)	(2)	(3)	(4)	(5)	(6)
Exposição aos fatores de risco						
Cigarro	0,3707***	0,5394***	0,3897***	0,5691***		
Álcool	0,5538***	0,5635***	0,5702***	0,5806***		
Sobrepeso	0,0869***	0,0868***	0,0893***	0,0891***		
Experiência e uso: cigarro e álcool		-0,2160*		-0,2300**		
Controles						
Atributos do aluno e da família	×	×	×	×	×	×
Atributos da escola					×	×
Fatores regionais					×	×
Efeito Fixo da Turma	×	×				
Efeito Fixo da Escola			×	×		
Prevalência de exposição na turma						
Cigarro					0,1316	
Álcool						0,9157
Sobrepeso						
Proxies para indisciplina do aluno						
Faltar propositalmente a aula						
Pais não controlam o tempo livre						
Envolvimento em brigas na escola						
Amostra final total	×	×	×	×		
Amostra restrita aos não-expostos					×	×
Covariadas	(7)	(8)	(9)	(10)	(11)	(12)
Exposição aos fatores de risco						
Cigarro		0,5481***			0,5423***	0,4107***
Álcool			0,4797***		0,4537***	0,4596***
Sobrepeso				0,1059**	0,1002**	0,0712
Experiência e uso: cigarro e álcool					-0,3107***	-0,2765***
Controles						
Atributos do aluno e da família	×	×	×	×	×	×
Atributos da escola	×	×	×	×	×	×
Fatores regionais	×	×	×	×	×	×
Efeito Fixo da Turma						×
Prevalência de exposição na turma						
Cigarro		0,1414			-0,7157	
Álcool			0,9549***		1,0430***	
Sobrepeso	-0,4334**			-0,3683**	-0,5946	
Proxies para indisciplina do aluno						
Faltar propositalmente a aula						0,4372***
Pais não controlam o tempo livre						0,1858***
Envolvimento em brigas na escola						-0,0535
Amostra final total		×	×	×	×	×
Amostra restrita aos não-expostos	×					

Nota: Efeitos para cigarro e álcool referentes à Definição 2.

* p -valor < 0,10; ** p -valor < 0,05; *** p -valor < 0,01.

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

**Tabela 6.** Análise de robustez dos resultados referentes à estimação dos efeitos dos fatores de risco comportamentais à saúde sobre o atraso escolar – Modelo PSM.

Covariadas	(1)	(2)	(3)	(4)	(5)	(6)
Exposição aos fatores de risco						
Cigarro	0,1075***	0,1033***				
Álcool			0,0769***	0,0784***		
Sobrepeso					0,0129***	0,0118***
Controles						
Atributos do aluno e da família	×	×	×	×	×	×
Atributos da escola	×		×		×	
Fatores regionais	×		×		×	
Efeito Fixo da Turma		×		×		×
Proxies para indisciplina do aluno						
Faltar propositalmente a aula	0,0844***	0,0795***	0,0854***	0,0808***	0,0999***	0,0909***
Pais não controlam o tempo livre	0,0115	0,0175**	0,0240***	0,0227***	0,0406***	0,0345***
Envolvimento em brigas na escola	-0,0195*	-0,0101	-0,0065	0,0004	0,0071	0,0112**
Amostra final total	×	×	×	×	×	×

Notas: As estimativas foram calculadas por meio de regressões ponderadas pelo PSM-K, condicionadas a covariadas que pudessem influenciar o atraso escolar, tal como indicadores sobre indisciplina do estudante. Efeitos para cigarro e álcool referentes à Definição 2.

* p -valor < 0,10; ** p -valor < 0,05; *** p -valor < 0,01.

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

expostos aos fatores de risco à saúde são mais propensos a apresentarem problemas de rendimento escolar no Brasil.

6. CONCLUSÕES

Este estudo apresenta direcionamentos empíricos sobre a relação de posturas de risco modificáveis 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 até então destacam basicamente o papel de condições antropométricas sobre o desempenho escolar com amostras menos representativas para o território nacional.

No desenvolvimento da presente pesquisa houve uma preocupação em verificar a robustez das estimativas calculadas, 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 distintos grupos de controle e o nível de incerteza gerado por fatores não-observados capazes de alterar as conclusões sobre a direção dessas estimativas. Os resultados centrais corroboram a hipótese de que crianças expostas aos fatores de risco comportamentais à saúde possuem um maior atraso escolar. Considerando as evidências do presente estudo e a literatura especializada na área, como Renna (2007), D. C. Machado (2008), Ding et al. (2009), Eide et al. (2010) e Lindo et al. (2013), ações que envolvam o enfrentamento das DCNT, no tocante aos fatores de risco modificáveis, podem implicar 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, como reporta a literatura, políticas de promoção a saúde que afetem os hábitos da família também

são relevantes, principalmente os destinados ao enfrentamento dos hábitos não saudáveis praticados em comunidades mais carentes. Pois, as diferenças no resultado escolar estimuladas por tais hábitos ocorrem com mais força entre as crianças com o NSE mais desfavorável. Nessa linha, investimentos na primeira infância podem ser uma alternativa para melhorar as condições de saúde e os rendimentos acadêmicos, já que nessa fase as habilidades cognitivas e sócio-emotivas estão em processo de formação (Conti et al., 2010; HHS, 2012).

Por fim, adverte-se que apesar das evidências encontradas serem sugestivas sobre o papel da exposição aos fatores de risco comportamentais à saúde na performance escolar, a melhor identificação de causa e efeito requer preferencialmente 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 BIBLIOGRÁFICAS

- Abrevaya, J., & Tang, H. (2010). Body mass index in families: Spousal correlation, endogeneity, and intergenerational transmission. *Empirical Economics*, 41(3), 841–864. doi: 10.1007/s00181-010-0403-6
- Albernaz, A., Ferreira, F. & Franco, C. (2002). Qualidade e equidade no ensino fundamental brasileiro. *Pesquisa e Planejamento Econômico*, 32(3), 453–476.
- Almeida, A. T. C. d. (2014). Determinantes dos piores e melhores resultados educacionais dos alunos da rede pública de ensino fundamental no Brasil. *Planejamento e Políticas Públicas*, 42, 147–188. Disponível em: <http://www.ipea.gov.br/ppp/index.php/PPP/article/view/284>
- Alves, M. T. G., & Soares, J. F. (2012). *O nível socioeconômico das escolas de educação básica brasileiras*. Belo Horizonte: UFMG.
- Andreski, P., & Breslau, N. (1993). Smoking and nicotine dependence in young adults: Differences between blacks and whites. *Drug and Alcohol Dependence*, 32(2), 119–125. doi: 10.1016/0376-8716(93)80004-X
- Balsa, A. I., Giuliano, L. M. & French, M. T. (2011). The effects of alcohol use on academic achievement in high school. *Economics of Education Review*, 30(1), 1–15. doi: 10.1016/j.econedurev.2010.06.015
- Barros, R. P. d., Mendonça, R., Santos, D. D. d. & Quintaes, G. (2001). Determinantes do desempenho educacional no Brasil. *Pesquisa e Planejamento Econômico*, 31(1), 1–42. Disponível em: <http://ppe.ipea.gov.br/index.php/ppe/issue/view/18>
- Becker, G. S. (1965). A theory of the allocation of time. *The Economic Journal*, 75(99), 493–517.
- Becker, G. S., & Mulligan, C. B. (1997). The endogenous determination of time preference. *The Quarterly Journal of Economics*, 112(3), 729–758.
- Becker, G. S., Murphy, K. M. & Tamura, R. (1990). Human capital, fertility, and economic growth. *Journal of Political Economy*, 98(5), S12–S37. Disponível em: <http://www.jstor.org/stable/2937630>
- Bergen, H. a., Martin, G., Roeger, L. & Allison, S. (2005). Perceived academic performance and alcohol, tobacco and marijuana use: Longitudinal relationships in young community adolescents. *Addictive behaviors*, 30(8), 1563–1573. doi: 10.1016/j.addbeh.2005.02.012
- Brasil. (2011). *Plano de ações estratégicas para o enfrentamento das doenças crônicas não transmissíveis (DCNT) no Brasil: 2011–2022*. Brasília DF: Ministério da Saúde.
- Brown, S. A., Tapert, S. F., Granholm, E. & Delis, D. C. (2000). Neurocognitive functioning of adolescents: Effects of protracted alcohol use. *Alcoholism, Clinical and Experimental Research*, 24(2), 164–171. doi: 10.1111/j.1530-0277.2000.tb04586.x
- Caliendo, M., & Kopeinig, S. (2005, May). *Some practical guidance for the implementation of propensity score matching* (Discussion Paper N° 1588). Bonn, Germany: Institute for the Study of Labor (IZA). Disponível em: <http://ftp.iza.org/dp1588.pdf>



- Carrell, S. E., Hoekstra, M. & West, J. E. (2011). Does drinking impair college performance? Evidence from a regression discontinuity approach. *Journal of Public Economics*, 95(1-2), 54–62. doi: 10.1016/j.jpubeco.2010.08.008
- Cawley, J., & Ruhm, C. (2011, May). *The economics of risky health behaviors* (Working Paper N° 17081). National Bureau of Economic Research (NBER). doi: 10.3386/w17081
- Chaloupka, F. J., & Warner, K. E. (1999, March). *The economics of smoking* (Working Paper N° 7047). National Bureau of Economic Research (NBER). doi: 10.3386/w7047
- Chatterji, P. (2006). Does alcohol use during high school affect educational attainment? Evidence from the national education longitudinal study. *Economics of Education Review*, 25(5), 482–497. doi: 10.1016/j.econedurev.2005.05.005
- Cole, T. J., Bellizzi, M. C., Flegal, K. M. & Dietz, W. H. (2000). Establishing a standard definition for child overweight and obesity worldwide: International survey. *BMJ*, 320(7244), 1–6. doi: 10.1136/bmj.320.7244.1240
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D. & York, R. L. (1966). *Equality of educational opportunity* (Survey report). Washington DC: National Center for Educational Statistics. Disponível em: <http://files.eric.ed.gov/fulltext/ED012275.pdf>
- Conti, G., Heckman, J. & Urzua, S. (2010). The education-health gradient. *American Economic Review*, 100(2), 234–238. doi: 10.1257/aer.100.2.234
- Cornfield, J., Haenszel, W., Hammond, E. C., Lilienfeld, A. M., Shimkin, M. B. & Wynder, E. L. (1959). Smoking and lung cancer: Recent evidence and a discussion of some questions. *Journal of the National Cancer Institute*, 22(1), 173–203. doi: 10.1093/jje/dyp289
- Crosnoe, R., & Muller, C. (2004). Body mass index, academic achievement, and school context: Examining the educational experiences of adolescents at risk of obesity. *Journal of Health and Social Behavior*, 45(4), 393–407. doi: 10.1177/002214650404500403
- Cutler, D. M., & Glaeser, E. (2005, February). *What explains differences in smoking, drinking and other health-related behaviors* (Working Paper N° 11100). National Bureau of Economic Research (NBER). doi: 10.3386/w11100
- Cutler, D. M., & Lleras-Muney, A. (2010). Understanding differences in health behaviors by education. *Journal of Health Economics*, 29(1), 1–28. doi: 10.1016/j.jhealeco.2009.10.003
- Ding, W., Lehrer, S. E., Rosenquist, J. N. & Audrain-McGovern, J. (2009). The impact of poor health on academic performance: New evidence using genetic markers. *Journal of Health Economics*, 28(3), 578–97. doi: 10.1016/j.jhealeco.2008.11.006
- DiPrete, T. A., & Gangl, M. (2004). Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments. *Sociological Methodology*, 34(1), 271–310.
- Eide, E. R., & Showalter, M. H. (1998). The effect of school quality on student performance: A quantile regression approach. *Economics Letters*, 58, 345–350.
- Eide, E. R., & Showalter, M. H. (2011). Estimating the relation between health and education: What do we know and what do we need to know? *Economics of Education Review*, 30(5), 778–791. doi: 10.1016/j.econedurev.2011.03.009
- Eide, E. R., Showalter, M. H. & Goldhaber, D. D. (2010). The relation between children's health and academic achievement. *Children and Youth Services Review*, 32(2), 231–238. doi: 10.1016/j.childyouth.2009.08.019
- Ellickson, P. L., Tucker, J. S. & Klein, D. J. (2001). High-risk behaviors associated with early smoking: Results from a 5-year follow-up. *Journal of Adolescent Health*, 28(6), 465–473.
- Florin, T. A., Shults, J. & Stettler, N. (2011). Perception of overweight is associated with poor academic performance

- in US adolescents. *The Journal of School Health*, 81(11), 663–670. doi: 10.1111/j.1746-1561.2011.00642.x
- Foster, A. D., & Rosenzweig, M. R. (1995). Learning by doing and learning from others: Human capital and technical change in agriculture. *Journal of Political Economy*, 103(6), 1176–1209. Disponível em: <http://www.jstor.org/stable/2138708>
- Gastwirth, J. L., Krieger, A. M. & Rosenbaum, P. R. (1998). Dual and simultaneous sensitivity analysis for matched pairs. *Biometrika*, 85(4), 907–920.
- Glewwe, P. W., Hanushek, E. A., Humpage, S. D. & Ravina, R. (2011, October). *School resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010* (Working Paper N° 17554). National Bureau of Economic Research (NBER). doi: 10.3386/w17554
- Glewwe, P. W., & Jacoby, H. G. (1995). 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*, 77(1), 156–169.
- Gomes-Neto, J. B., & Hanushek, E. A. (1994). Causes and consequences of grade repetition: Evidence from Brazil. *Economic Development and Cultural Change*, 43(1), 117–148. doi: 10.1086/452138
- Gomes-Neto, J. B., Hanushek, E. A., Leite, R. H. & Frota-Bezerra, R. C. (1997). Health and schooling: Evidence and policy implications for developing countries. *Economics of Education Review*, 16(3), 271–282.
- Grossman, M. (2005, August). *Education and nonmarket outcomes* (Working Paper N° 11582). National Bureau of Economic Research (NBER). doi: 10.3386/w11582
- Hanushek, E. A. (1970). The production of education, teacher quality, and efficiency. In U.S. Office of Education (Ed.), *Do teachers make a difference?* (pp. 79–99). Washington DC: U.S. Department of Health, Education, and Welfare.
- Hanushek, E. A. (1971). Teacher characteristics and gains in student achievement: Estimation using micro data. *The American Economic Review*, 61(2), 280–288. Disponível em: <http://www.jstor.org/stable/10.2307/1817003>
- Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature*, 24(3), 1141–1177.
- Hanushek, E. A., & Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature*, 46(3), 607–668. doi: 10.1257/jel.46.3.607
- HHS (U.S. Department of Health and Human Services). (2012). *Preventing tobacco use among youth and young adults: A report of the Surgeon General* (Report). Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health. Disponível em: <http://www.surgeongeneral.gov/library/reports/preventing-youth-tobacco-use/>
- Hirano, K., & Imbens, G. W. (2001). Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. *Health Services & Outcomes Research Methodology*, 2(3), 259–278.
- IBGE. (2013). *Pesquisa nacional de saúde do escolar 2012*. Rio de Janeiro: Ministério do Planejamento, Orçamento e Gestão – Instituto Brasileiro de Geografia e Estatística. Disponível em: <http://biblioteca.ibge.gov.br/visualizacao/livros/liv64436.pdf>
- INEP. (2004). *Dicionário de indicadores educacionais: Fórmulas de cálculo*. Brasília, DF: MEC/INEP. Disponível em: <http://www.publicacoes.inep.gov.br/portal/download/263>
- Jacobsen, L. K., Krystal, J. H., Mencl, W. E., Westerveld, M., Frost, S. J. & Pugh, K. R. (2005). Effects of smoking and smoking abstinence on cognition in adolescent tobacco smokers. *Biological Psychiatry*, 57(1), 56–66. doi: 10.1016/j.biopsych.2004.10.022
- Jacobsen, L. K., Pugh, K. R., Constable, R. T., Westerveld, M. & Mencl, W. E. (2007). Functional correlates of verbal memory deficits emerging during nicotine withdrawal in abstinent adolescent cannabis users. *Biological*



- Psychiatry*, 61(1), 31–40. doi: 10.1016/j.biopsych.2006.02.014
- Jamison, D. T. (1986). Child malnutrition and school performance in China. *Journal of Development Economics*, 20(2), 299–309. doi: 10.1016/0304-3878(86)90026-X
- Jeynes, W. H. (2002). The relationship between the consumption of various drugs by adolescents and their academic achievement. *The American Journal of Drug and Alcohol Abuse*, 28(1), 15–35. doi: 10.1081/ADA-120001279
- Jurges, H., Reinhold, S. & Salm, M. (2011). Does schooling affect health behavior? Evidence from the educational expansion in Western Germany. *Economics of Education Review*, 30(5), 862–872. doi: 10.1016/j.econedurev.2011.04.002
- Kaestner, R., & Grossman, M. (2009). Effects of weight on children's educational achievement. *Economics of Education Review*, 28(6), 651–661. doi: 10.1016/j.econedurev.2009.03.002
- Kaestner, R., Grossman, M. & Yarnoff, B. (2011). Effects of weight on adolescent educational attainment. In M. Grossman & N. H. Mocan (Eds.), *Economic aspects of obesity* (pp. 283–313). Chicago: University of Chicago Press.
- Kenkel, D. S., & Wang, P. (1999). Are alcoholics in bad jobs? In F. J. Chaloupka (Ed.), *The economic analysis of substance use and abuse: An integration of econometrics and behavioral economic research*. Chicago: University of Chicago Press.
- Koch, S. F., & Ribar, D. C. (2001). A siblings analysis of the effects of alcohol consumption onset on educational attainment. *Contemporary Economic Policy*, 19(2), 162–174.
- Kristjansson, A. L., Sigfusdottir, I. D. & Allegrante, J. P. (2010). Health behavior and academic achievement among adolescents: The relative contribution of dietary habits, physical activity, body mass index, and self-esteem. *Health Education & Behavior*, 37(1), 51–64. doi: 10.1177/1090198107313481
- Kuczumski, R. J., Ogden, C. L., Guo, S. S., Grummer-Strawn, L. M., Flegal, K. M., Mei, Z., ... Johnson, C. L. (2002). 2000 CDC growth charts for the United States: Methods and development. *Vital and Health Statistics*, 11(246), 1–190.
- Levine, P. B., Gustafson, T. A. & Velenchik, A. D. (1997). More bad news for smokers? The effects of cigarette smoking on wages. *Industrial and Labor Relations Review*, 50(3), 493–509.
- Lindo, J. M., Swensen, I. D. & Waddell, G. R. (2013). Alcohol and student performance: Estimating the effect of legal access. *Journal of Health Economics*, 32(1), 22–32. doi: 10.1016/j.jhealeco.2012.09.009
- Lye, J., & Hirschberg, J. (2010). Alcohol consumption and human capital: A retrospective study of the literature. *Journal of Economic Surveys*, 24(2), 309–338. doi: 10.1111/j.1467-6419.2009.00616.x
- Machado, A. F., Moro, S., Martins, L. & Rios, J. (2008). Qualidade do ensino em matemática: Determinantes do desempenho de alunos em escolas públicas estaduais mineiras. *Revista da Anpec*, 9(1), 23–45. Disponível em: https://anpec.org.br/revista/vol9/vol9n1p23_45.pdf
- Machado, D. C. (2008). Efeitos da saúde na idade de entrada à escola. *Pesquisa e Planejamento Econômico*, 38(1), 67–96. Disponível em: <http://ppe.ipea.gov.br/index.php/ppe/article/view/1067>
- Machado, D. C., & Gonzaga, G. (2007). O impacto dos fatores familiares sobre a defasagem idade-série de crianças no Brasil. *Revista Brasileira de Economia*, 61(4), 449–476.
- Mankiw, N. G., Romer, D. & Weil, D. N. (1992). A contribution to the empirics of economic growth. *The Quarterly Journal of Economics*, 107(2), 407–437. doi: 10.2307/2118477
- Martins, S. S., & Alexandre, P. K. (2009). The association of ecstasy use and academic achievement among adolescents in two U.S. national surveys. *Addictive Behaviors*, 34(1), 9–16. doi: 10.1016/j.addbeh.2008.07.022
- Moock, P. R., & Leslie, J. (1986). Childhood malnutrition and schooling in the Terai Region of Nepal. *Journal of Development Economics*, 20, 33–52.
- Mullahy, J., & Sindelar, J. L. (1996). Employment, unemployment, and problem drinking. *Journal of Health Economics*, 15(4), 409–34. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/10164037>

- O'Neill, D. (1995). Education and income growth: Implications for cross-country inequality. *Journal of Political Economy*, 103(6), 1289–1301. Disponível em: <http://www.jstor.org/stable/2138712>
- Park, C., & Kang, C. (2008). Does education induce healthy lifestyle? *Journal of Health Economics*, 27(6), 1516–1531. doi: 10.1016/j.jhealeco.2008.07.005
- Pessoa, D. G. C., Silva, P. L. N. & Duarte, R. P. N. (1997). Análise estatística de dados de pesquisas por amostragem: problemas no uso de pacotes-padrões. *Revista Brasileira de Estatística*, 58(210), 53–75.
- Pirie, P. L., Murray, D. M. & Luepker, R. V. (1988). Smoking prevalence in a cohort of adolescents, including absentees, dropouts, and transfers. *American Journal of Public Health*, 78(2), 176–179.
- Ponzo, M. (2013). Does bullying reduce educational achievement? An evaluation using matching estimators. *Journal of Policy Modeling*, 35(6), 1057–1078. doi: 10.1016/j.jpolmod.2013.06.002
- Rangvid, B. S. (2007). School composition effects in Denmark: Quantile regression evidence from PISA 2000. *Empirical Economics*, 33(2), 359–388. doi: 10.1007/s00181-007-0133-6
- Rees, D. I., & Sabia, J. J. (2011). The effect of migraine headache on educational attainment. *Journal of Human Resources*, 46(2), 317–332. doi: 10.1353/jhr.2011.0019
- Rees, D. I., & Sabia, J. J. (2014). The kid's speech: The effect of stuttering on human capital acquisition. *Economics of Education Review*, 38, 76–88. doi: 10.1016/j.econedurev.2013.07.007
- Reilly, J. J., & Kelly, J. (2011). Long-term impact of overweight and obesity in childhood and adolescence on morbidity and premature mortality in adulthood: Systematic review. *International Journal of Obesity*, 35(7), 891–898. doi: 10.1038/ijo.2010.222
- Renna, F. (2007). The economic cost of teen drinking: Late graduation and lowered earnings. *Health Economics*, 419, 407–419. doi: 10.1002/hec
- Renna, F. (2008). Teens alcohol consumption and schooling. *Economics of Education Review*, 27(1), 69–78. doi: 10.1016/j.econedurev.2006.05.002
- Riani, J. d. L. R., & Rios-Neto, E. L. G. (2008). Background familiar versus perfil escolar do município: Qual possui maior impacto no resultado educacional dos alunos brasileiros. *Revista Brasileira de Estudos de População*, 25(2), 251–269. doi: 10.1590/S0102-30982008000200004
- Ribeiro, R., & Cacciamali, M. C. (2012). Impactos do Programa Bolsa-Família sobre os indicadores educacionais. *Economia*, 13(2), 415–446. Disponível em: http://www.anpec.org.br/revista/vol13/vol13n2p415_446.pdf
- Rios-Neto, E. L. G., César, C. C. & Riani, J. d. L. R. (2002). Estratificação educacional e progressão escolar por série no Brasil. *Pesquisa e Planejamento Econômico*, 32(3), 395–415. Disponível em: <http://ppe.ipea.gov.br/index.php/ppe/article/viewFile/137/72>
- Rosenbaum, P. R. (2002). Attributing to treatment in matched observational studies. *Journal of the American Statistical Association*, 97(457), 183–192.
- Rosenbaum, P. R. (2010). *Design of observational studies*. New York: Springer.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55.
- Sampaio, B., Sampaio, Y., Mello, E. P. G. d. & Melo, A. S. (2011). Desempenho no vestibular, *background* familiar e evasão: Evidências da UFPE. *Economia Aplicada*, 15(2), 287–309. doi: 10.1590/S1413-80502011000200006
- Schultz, T. W. (1961). Investment in human capital. *The American Economic Review*, 51(1), 1–17. Disponível em: <http://www.jstor.org/stable/1818907>
- Sigfusdottir, I. D., Kristjansson, A. L. & Allegrante, J. P. (2007). Health behaviour and academic achievement in Icelandic school children. *Health education research*, 22(1), 70–80. doi: 10.1093/her/cyl044
- Silva, P. L. N., Pessoa, D. G. C. & Lila, M. F. (2002). Análise estatística de dados da pnad: Incorporando a estrutura



- do plano amostral. *Ciência & Saúde Coletiva*, 7(4), 659–670.
- Soares, J. F., & Andrade, R. J. d. (2006). Nível socioeconômico, qualidade e equidade das escolas de Belo Horizonte. *Avaliação e Políticas Públicas em Educação*, 14(40), 107–126.
- Soares, T. M. (2003). Influência do professor e do ambiente em sala de aula sobre a proficiência alcançada pelos alunos avaliados no Simave-2002. *Estudos em Avaliação Educacional*(28), 103–124. doi: 10.18222/ea02820032172
- Soares-Neto, J. J., de Jesus, G. R., Karino, C. A. & de Andrade, D. F. (2013). Uma escala para medir a infraestrutura escolar. *Estudos em Avaliação Educacional*, 24(54), 78–99.
- van der Linden, W. J., & Hambleton, R. K. (Eds.). (2010). *Handbook of modern item response theory* (2ª ed.). New York: Springer.
- Veronneau, M., Trempe, S. & Paiva, A. O. (2014). Risk and protection factors in the peer context: How do other children contribute to the psychosocial adjustment of the adolescent? *Ciência & Saúde Coletiva*, 19(3), 695–705. doi: 10.1590/1413-81232014193.17972013
- Wakefield, J. (2003). Sensitivity analyses for ecological regression. *Biometrics*, 59(1), 9–17.
- Wooldridge, J. W. (2010). *Econometric analysis of cross section and panel data* (2ª ed.). Massachusetts: MIT Press.
- Yamada, T., Kendix, M. & Yamada, T. (1996). The impact of alcohol consumption and marijuana use on high school graduation. *Health Economics*, 5(1), 77–92.
- Zeigler, D. W., Wang, C. C., Yoast, R. A., Dickinson, B. D., McCaffree, M. A., Robinowitz, C. B. & Sterling, M. L. (2005). The neurocognitive effects of alcohol on adolescents and college students. *Preventive Medicine*, 40(1), 23–32. doi: 10.1016/j.ypped.2004.04.044

APÊNDICE.

Tabela A-1. 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	Número de banheiros com chuveiro dentro da casa ^a	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	Faixa de escolaridade da mãe ^a	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	Faixa de escolaridade do pai ^a	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

Notas: Parâmetros estimados pela TRI policotômica ordinal unidimensional: a = inclinação; d = intercepto. Nesse caso o modelo usado foi o de crédito parcial generalizado, obtido através da seguinte equação:

$$Pr_i q(\theta) = \frac{\exp\left(\sum_{j=0}^q Ba_i(\theta - d_{ij})\right)}{\sum_{r=0}^{m_i} \left[\exp\left\{\sum_{j=0}^r Ba_i(\theta - d_{ij})\right\}\right]}$$

onde θ é o traço latente e o indexador i refere-se ao i -ésimo item, com m categorias.

^a Variável policotômica.

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

**Tabela A-2.** 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

Notas: Parâmetros estimados pela TRI dicotômica unidimensional: a = inclinação; d = intercepto. O modelo logístico de dois parâmetros para o caso de resposta binária é expresso por $Pr_i(\theta) = 1/[1 + \exp[-a_i(\theta + d_i)]]$, onde θ é o traço latente e o indexador i refere-se ao i -ésimo item.

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

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
Mãe 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.

Tabela A-4. Efeito marginal de variáveis explicativas discretas (D) e contínuas (C) sobre a probabilidade de atraso escolar no Brasil 2012 – a partir do modelo logit do Estágio I.

Covariadas	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Atributos do aluno e da família									
D. Gênero (homem=1)	0,0937***	0,0965***	0,0942***	0,0945***	0,0966***	0,0954***	0,0954***	0,0967***	0,0968***
D. Raça (Branca=1)	-0,0276***	-0,0288***	-0,0287***	-0,0280***	-0,0279***	-0,0278***	-0,0278***	-0,0276***	-0,0276***
D. Trabalho	0,0968***	0,1009***	0,1066***	0,1018***	0,0918***	0,0942***	0,0941***	0,0901***	0,0901***
D. Núcleo familiar biparental	-0,0740***	-0,0767***	-0,0809***	-0,0782***	-0,0736***	-0,0721***	-0,0722***	-0,0726***	-0,0726***
C. NSE do aluno	-0,2592***	-0,2697***	-0,2617***	-0,2614***	-0,2689***	-0,2641***	-0,2641***	-0,2678***	-0,2680***
C. Tamanho da família	0,0144***	0,0151***	0,0148***	0,0145***	0,0144***	0,0146***	0,0146***	0,0143***	0,0143***
Atributos da escola									
C. Total de matriculas (em mil)	0,0088	0,0092	0,0097	0,0092	0,0095	0,0084	0,0084	0,0091	0,0091
D. Rede pública	-0,0167	-0,0127	-0,0115	-0,0134	-0,0159	-0,0167	-0,0168	-0,0165	-0,0165
D. Conselho Escolar	-0,0478***	-0,0469***	-0,0468***	-0,0472***	-0,0467***	-0,0477***	-0,0477***	-0,0468***	-0,0469***
C. Índice de infraestrutura (IEE)	-0,1289***	-0,1281***	-0,1276***	-0,1282***	-0,1289***	-0,1287***	-0,1286***	-0,1286***	-0,1287***
C. NSE da escola	-0,3886***	-0,3920***	-0,3804***	-0,3826***	-0,3854***	-0,3952***	-0,3955***	-0,3866***	-0,3863***
Fatores regionais									
D. Capital	-0,0606	-0,0630	-0,0638	-0,0626	-0,0629	-0,0603	-0,0603	-0,0621	-0,0620
D. Região Norte	0,0286	0,0285	0,0263	0,0283	0,0315	0,0304	0,0304	0,0329	0,0329
D. Região Nordeste	-0,0207	-0,0221	-0,0224	-0,0207	-0,0216	-0,0197	-0,0197	-0,0198	-0,0198
D. Região Sul	-0,0697	-0,0647	-0,0602	-0,0634	-0,0676	-0,0711	-0,0710	-0,0684	-0,0685
D. Região Centro-Oeste	0,0060	0,0094	0,0110	0,0095	0,0087	0,0063	0,0064	0,0088	0,0089
Exposição aos fatores de risco									
D. Experiência: cigarro (def.1)	0,0866***					0,0759***	0,0695***		
D. Experiência: álcool (def.1)		0,0452***				0,0286	0,0318**		
D. Sobrepena			0,0154**			0,0144*	0,0144*	0,0148**	0,0146**
D. Experiência e uso: cigarro (def.2)				0,1086***				0,0528***	0,0848***
D. Experiência e uso: álcool (def.2)					0,1018***			0,0951***	0,0949***
Observações	66.844	66.844	66.844	66.844	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	1.897.670	1.897.670	1.897.670	1.897.670
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

Notas: As estimativas são desenvolvidas considerando, além do peso amostral, a estratificação e a UPA do plano amostral.

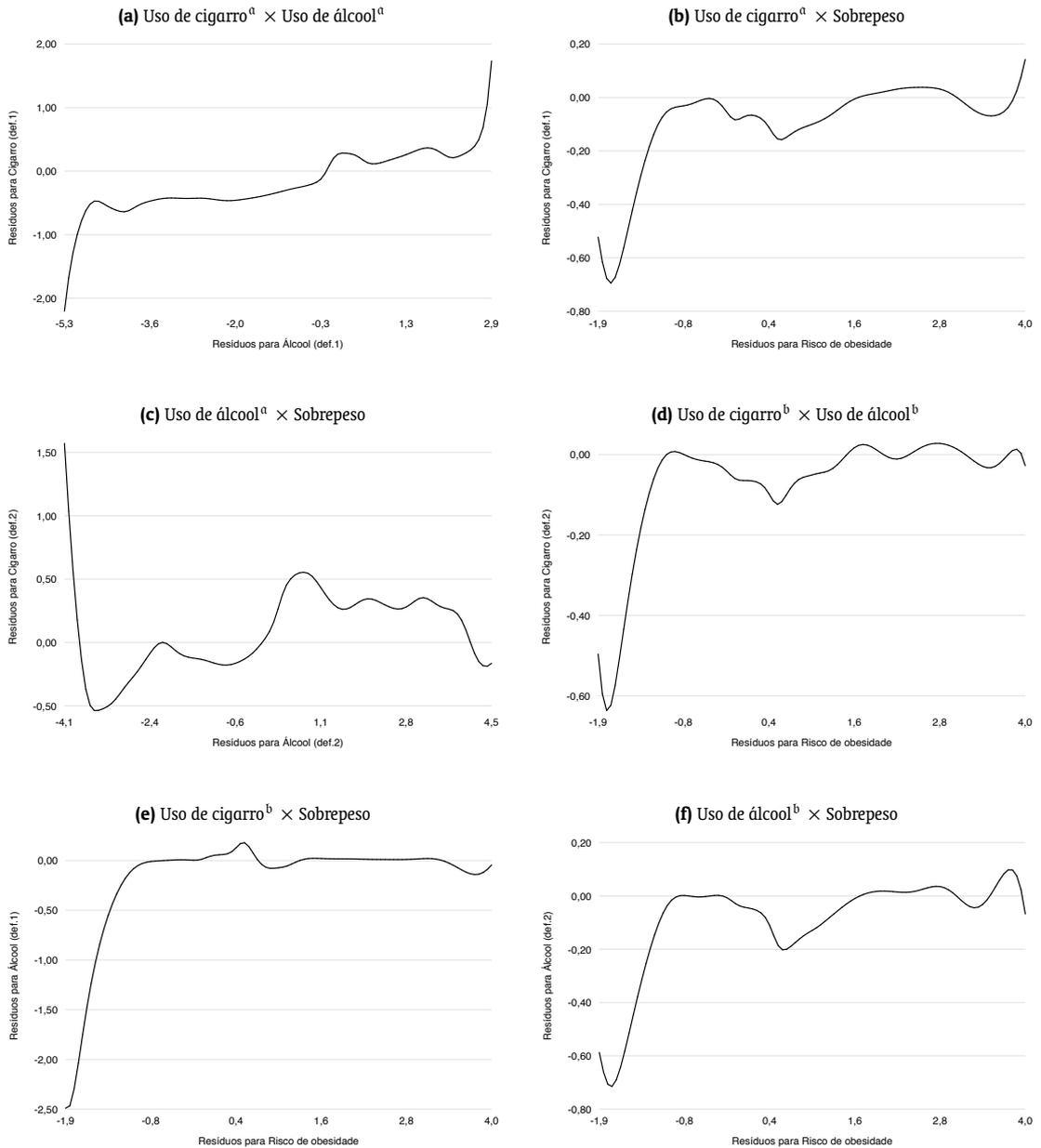
Legenda: D = estimativa calculada pela equação (4); C = estimativa calculada pelo efeito parcial médio (EPM).

* p -valor < 0,10; ** p -valor < 0,05; *** p -valor < 0,01.

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



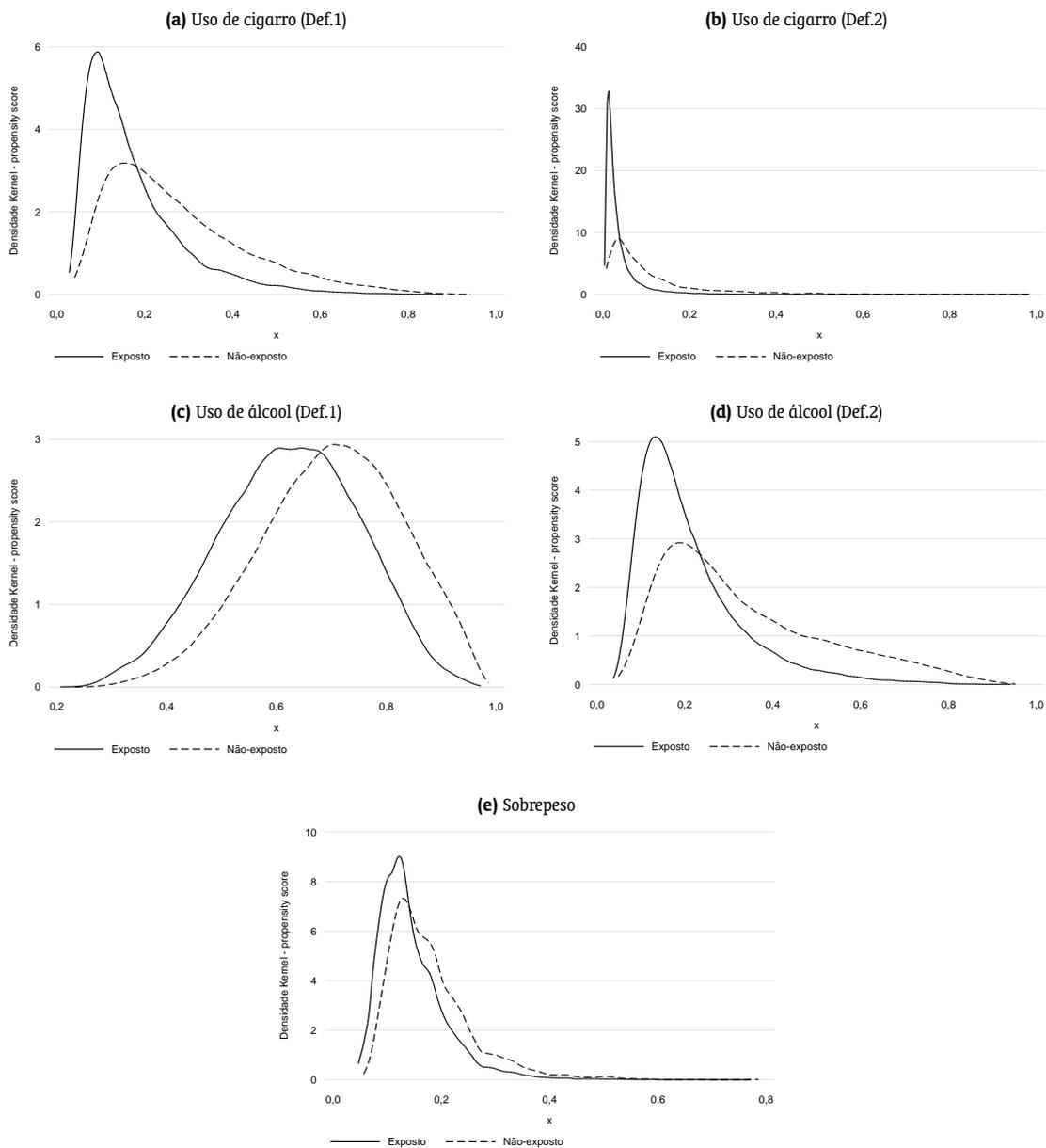
Figura A-1. Curvas não-paramétricas entre os termos residuais das equações do estágio inicial do PSM.



^a Definição 1; ^b Definição 2.

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

Figura A-2. Densidade Kernel do *propensity score* entre expostos e não-expostos por tipo de fator de risco comportamental à saúde.



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