



Faculdade de Economia,
Administração e Contabilidade
de Ribeirão Preto
Universidade de São Paulo

Texto para Discussão

Série Economia

TD-E 02 / 2011

**O efeito da saúde sobre o
desempenho escolar**

Prof. Me. Daniel de Araújo Roland



Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto
Universidade de São Paulo

Universidade de São Paulo
Faculdade de Economia, Administração e Contabilidade
de Ribeirão Preto

Reitor da Universidade de São Paulo
João Grandino Rodas

Diretor da FEA-RP/USP
Sigismundo Bialoskorski Neto

Chefe do Departamento de Administração
Marcos Fava Neves

Chefe do Departamento de Contabilidade
Adriana Maria Procópio de Araújo

Chefe do Departamento de Economia
Walter Belluzzo Junior

CONSELHO EDITORIAL

Comissão de Pesquisa da FEA-RP/USP

Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto
Avenida dos Bandeirantes, 3900
14049-905 Ribeirão Preto - SP

A série TEXTO PARA DISCUSSÃO tem como objetivo divulgar: i) resultados de trabalhos em desenvolvimento na FEA-RP/USP; ii) trabalhos de pesquisadores de outras instituições considerados de relevância dadas as linhas de pesquisa da instituição. Veja o site da Comissão de Pesquisa em www.cpq.fearp.usp.br.
Informações: e-mail: cpq@fearp.usp.br

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE DE
RIBEIRÃO PRETO
DEPARTAMENTO DE ECONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

DANIEL DE ARAÚJO ROLAND

O efeito da saúde sobre o desempenho escolar

Orientador: Prof. Dr. Luiz Guilherme
Dácar da Silva Scorzafave

RIBEIRÃO PRETO
2011

Prof. Dr. João Grandino Rodas
Reitor da Universidade de São Paulo

Prof. Dr. Sigismundo Bialoskorski Neto
Diretor da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto

Prof. Dr. Walter Belluzzo Júnior
Chefe do Departamento de Economia

DANIEL DE ARAÚJO ROLAND

O efeito da saúde sobre o desempenho escolar

Dissertação apresentada ao Programa de Pós-Graduação em Economia – Área: Economia Aplicada da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo como requisito para obtenção do título de Mestre em Ciências.

Orientador: Prof. Dr. Luiz Guilherme Dácar da Silva Scorzafave

RIBEIRÃO PRETO
2011

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

FICHA CATALOGRÁFICA

Roland, Daniel de Araújo

O efeito da saúde sobre o desempenho escolar. Ribeirão Preto, 2011.

68 p. : il. ; 30cm

Dissertação de Mestrado, apresentada à Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto/USP. Área de concentração: Economia Aplicada.

Orientador: Scorzafave, Luiz Guilherme Dácar da Silva.

1. Educação. 2. Saúde. 3. Desempenho escolar.

FOLHA DE APROVAÇÃO

Nome: ROLAND, Daniel de Araújo
Título: O efeito da saúde sobre o desempenho escolar

Dissertação apresentada à Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo para a obtenção do título de Mestre em Ciências.

Área de concentração: Economia Aplicada

Aprovado em: _____

BANCA EXAMINADORA

Prof. Dr. Luiz Guilherme Dácar da Silva Scorzafave Instituição: FEARP/USP

Julgamento: _____ Assinatura: _____

Prof. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

*Dedico esta dissertação à minha
família, em especial aos meus pais
pelo seu incansável esforço em
proporcionar felicidade e educação
intelectual e moral aos filhos.*

AGRADECIMENTOS

Existem muitas pessoas, não citadas aqui, que contribuíram direta ou indiretamente para a realização deste trabalho. Pessoas fundamentais na minha formação acadêmica, que contribuíram para formar a pessoa que sou hoje. Fica aqui o meu agradecimento a todas essas pessoas.

Agradeço aos meus pais por todo o apoio que sempre me deram em todos os momentos da minha vida, por sempre incentivar os estudos e por tudo que fizeram e fazem por mim e pelo meu irmão. Serei eternamente grato aos dois. Agradeço também aos meus avôs, já falecidos, mas que até hoje servem de exemplo como pessoas batalhadoras, íntegras, honestas e de bom coração. Agradeço também, é claro, às minhas avós pela sua ternura e dedicação à família.

Agradeço à Fapeam – Fundação de Amparo à Pesquisa do Estado do Amazonas – pelo apoio financeiro a este trabalho.

Agradeço a todo o pessoal da USP, dos funcionários e professores aos colegas de turma e amigos de outros departamentos. Aos membros do grupo de pesquisa Observatório da Educação, tanto os alunos quanto os professores participaram diretamente da realização desse trabalho.

Aos amigos que fiz durante o mestrado, fica o meu agradecimento. Alisson, João e Eduardo, que compartilharam centenas de horas de estudo lado a lado. André e Maurício, colegas de mestrado e amigos para todas as horas. Agradeço também à Gabriela, que desde o início do mestrado estendeu a mão em amizade e tornou a adaptação em Ribeirão Preto bem mais fácil para todos, além de ter me ajudado bastante em momentos difíceis.

Agradeço à Marina, a qual conheço há 3 anos mas que estreitamos muito a amizade durante o mestrado. Compartilhamos muitos momentos bons, outros tantos ruins, mas sua amizade e companheirismo certamente foram essenciais para suportar as exigências do mestrado.

Finalmente, agradeço a pessoa que mais me ajudou neste trabalho. Meu orientador, Luiz Guilherme Scorzafave, um excelente orientador, professor e pesquisador. Agradeço por seu esforço, atenção e dedicação a este trabalho sem o qual teria sido impossível chegar até aqui.

RESUMO

ROLAND, D. A. **O efeito da saúde sobre o desempenho escolar**. 2011. 68 f. Dissertação (Mestrado) – Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2011.

O objetivo deste estudo é analisar o impacto da saúde sobre o desempenho escolar de alunos da quarta série do ensino fundamental em todo o Brasil no ano de 2005 e 2007. No Brasil, a carência de bases de dados contendo simultaneamente indicadores de saúde e de desempenho escolar dificulta as pesquisas sobre o tema. Como alternativa, este estudo utiliza dados agregados referentes à oferta de serviços de saúde no município e microrregião, provenientes do DATASUS e da Pesquisa Assistência Médico-Sanitária, como *proxy* da saúde dos alunos da quarta série nos municípios. Os dados para o desempenho escolar dos alunos são provenientes da Prova Brasil 2005 e 2007. Dessa forma, a unidade de observação passa a ser a microrregião, e não o aluno. A hipótese a ser testada é se a oferta de serviços de saúde na microrregião afeta o desempenho médio dos alunos dessa microrregião. Utiliza-se inicialmente uma análise de regressão linear contendo as variáveis indicadoras de saúde como regressores em uma *cross-section* para 2005. Em seguida, a análise é feita em painel, com observações de 2005 e 2007. Os resultados indicam que há impacto positivo, mas apenas em alguns indicadores e, mesmo assim, o impacto é pequeno. Outra abordagem é a utilização de surtos de dengue em municípios para determinar se existe impacto negativo no desempenho escolar dos alunos de municípios com surto. Para estimação nessa abordagem foi utilizado *matching*. Os resultados dessa análise indicam impacto negativo, porém sensível à escolha das variáveis e de algoritmos de estimação.

Palavras-chaves: Saúde do aluno. Desempenho escolar. Prova Brasil. DATASUS. Matching.

ABSTRACT

ROLAND, D. A. **The effect of health on school proficiency.** 2011. 68 f. Dissertação (Mestrado) – Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2011.

The objective of this study is to analyze the impact of health on the school proficiency of students in the fourth grade of elementary school in Brazil in the years 2005 and 2007. In Brazil, the lack of data containing simultaneously indicators of health and school proficiency makes it harder to research on the subject. As an alternative, this study utilizes aggregated data, like the offer of health services in cities, taken from DATASUS and “Pesquisa Assistência Médico-Sanitária”, as a proxy for the health of students of the fourth grade in the cities. The data for school proficiency of students comes from “Prova Brasil” 2005 and 2007. That way, the observation unit becomes the city, not the student. The hypothesis to be tested is whether or not the offer of health services in a city can affect the school proficiency. Initially a linear regression is utilized containing indicators of health variables as regressors in a cross-section in 2005. Next, the analysis is made in panel data, with observations of 2005 and 2007. The results show that there is positive impact, but only in a few indicators of health and, even then, the impact is small. Another approach was to utilize “dengue” outbreaks in cities as a way to determine whether there was a negative impact in the school proficiency of students that resided in a city with an outbreak. Matching was used for the estimation in this approach. The results of this analysis indicate a negative impact, but sensible to the choice of variables and estimation algorithms.

Key-words: Student’s health. School proficiency. Prova Brasil. DATASUS. Matching.

LISTA DE TABELAS, GRÁFICOS E FIGURAS

Tabela 1 - Estatísticas descritivas de características do aluno (matemática - 4ª série)	34
Tabela 2 - Estatísticas descritivas de características do aluno (português - 4ª série)	35
Tabela 3 - Estatísticas descritivas do <i>background</i> familiar do aluno (matemática - 4ª série)	36
Tabela 4 - Estatísticas descritivas do <i>background</i> familiar do aluno (português - 4ª série)	37
Tabela 5 - Estatísticas descritivas da escola (matemática - 4ª série)	39
Tabela 6 - Estatísticas descritivas da escola (português - 4ª série)	40
Tabela 7 - Desempenho médio dos municípios, com e sem informações de variável de saúde, nas provas de matemática e português.	41
Tabela 8 - Distribuição dos municípios com e sem surtos de dengue em 2005	44
Gráfico 1 - Média das taxas de dengue dos municípios com surto e demais municípios	44
Tabela 9 - Impacto dos indicadores de saúde no desempenho escolar (matemática)	51
Tabela 10 - Impacto dos indicadores de saúde no desempenho escolar (português)	53
Figura 1 - Correlação na prova de matemática com a proporção de nascidos vivos com 4 ou mais consultas pré-natais	54
Figura 2 - Correlação na prova de português com a proporção de nascidos vivos com 4 ou mais consultas pré-natais	54
Figura 3 - Correlação na prova de matemática com a proporção de nascidos vivos com 7 ou mais consultas pré-natais	54
Figura 4 - Correlação na prova de português com a proporção de nascidos vivos com 7 ou mais consultas pré-natais	54
Tabela 11 - Impacto dos indicadores de saúde no desempenho escolar, estimadores de efeitos fixos (matemática, 4ª série)	55
Tabela 12 - Impacto dos indicadores de saúde no desempenho escolar, estimadores de efeitos aleatórios (matemática, 4ª série)	56
Tabela 13 - Impacto dos indicadores de saúde no desempenho escolar, estimadores de efeitos fixos (português, 4ª série)	57
Tabela 14 - Impacto dos indicadores de saúde no desempenho escolar, estimadores de efeitos aleatórios (português, 4ª série)	57
Tabela 15 - Efeito do surto de dengue em 2005 na Prova Brasil - matemática	58
Figura 5 - Histograma do <i>propensity score</i> calculado de acordo com a especificação 1	59

Tabela 16 - Testes de média após o cálculo do <i>propensity score</i> com variáveis da especificação 1 e pareamento com algoritmo de <i>kernel</i> , janela de 0,001	60
Tabela 17 - Efeito do surto de dengue em 2005 na Prova Brasil 2005 - português	61
Tabela 1A - Impacto dos indicadores de saúde no desempenho escolar, estimadores MQO agrupados (matemática - 4ª série)	68
Tabela 2A - Impacto dos indicadores de saúde no desempenho escolar, estimadores MQO agrupados (português - 4ª série)	68

SUMÁRIO

INTRODUÇÃO	13
CAPÍTULO 1 - REVISÃO DA LITERATURA	18
1.1 Educação	19
1.2 Saúde e produtividade	22
2 CAPÍTULO 2 - DADOS	30
2.1 Dados - Indicadores e qualidade da infra-estrutura de saúde em 2005	31
2.1.1 Análise descritiva dos dados da Prova Brasil 2005.....	33
2.2 Variáveis adicionais utilizadas na formação do painel 2005-2007	41
2.3 Municípios com surto de dengue em 2005.....	42
2.3.1 Análise descritiva dos dados de surto de dengue em 2005.....	43
3 CAPÍTULO 3 - METODOLOGIA	45
3.1 Indicadores de saúde e qualidade da infra-estrutura de saúde em <i>cross-section</i>	45
3.2 Métodos de estimação em painel.....	46
3.3 <i>Matching</i> para municípios com surto de dengue em 2005	48
4 CAPÍTULO 4 - RESULTADOS	50
4.1 Resultados das regressões em <i>cross-section</i>	51
4.2 Resultado das regressões em painel	54
4.3 Resultado das regressões de <i>matching</i>	57
5 CAPÍTULO 5 - CONSIDERAÇÕES FINAIS	62
REFERÊNCIAS BIBLIOGRÁFICAS	65
APÊNDICE	68

Introdução

Educação e saúde são dois assuntos que ocupam um lugar importante na formulação de políticas públicas. Pelo lado da educação, a disponibilidade de escolas e universidades, assim como sua qualidade, são importantes determinantes do nível de desenvolvimento de um país. De acordo com diversos estudos¹, a formação de capital humano é apontada como um dos principais fatores capazes de explicar as diferenças observadas com relação ao crescimento dos países. A educação, portanto, deve ocupar um lugar fundamental na agenda de pesquisa dos determinantes do desenvolvimento econômico.

O Brasil passou por um processo de universalização do ensino fundamental no país, com notável avanço na década de 1990. A busca pela melhora nos índices de acumulação de capital humano teve significativo impulso com a implementação do FUNDEF (Fundo de Desenvolvimento do Ensino Fundamental e de Valorização do Magistério). Como o repasse de verbas públicas adicionais passou a ser calculado de acordo com o número de alunos matriculados, o programa incentivou os municípios a colocar na escola crianças que antes não a freqüentavam. Com isso, em 2001, observou-se que 97% das crianças entre 7 e 14 anos estavam matriculadas na escola.

Por outro lado, os indicadores qualitativos do sistema educacional brasileiro não apresentaram melhoras significativas. O país realiza anualmente exames nacionais para avaliar a proficiência dos alunos em diferentes séries tanto do ensino fundamental quanto do ensino médio. Um dos exames, o SAEB (Sistema Nacional de Avaliação da Educação Básica), em 2003, evidenciou a má qualidade do ensino brasileiro. Nos resultados da 4ª série, na prova de Língua Portuguesa, apenas 4,8% dos alunos obtiveram notas consideradas ‘adequadas’ para a série; 39,7% tiveram notas ‘intermediárias’, 36,7% ‘críticas’ e 18,7% ‘muito críticas’, ou seja, mais da metade tiveram notas críticas ou muito críticas. Já os

¹ Mankiw, Romer e Weil (1992), Romer (2006)

resultados do PISA de 2003 (*Programme for International Student Assessment*) mostram que os estudantes brasileiros tiveram o pior desempenho relativamente aos outros 40 países participantes, com uma nota média de 356 pontos em matemática (área principal da avaliação em 2003), frente a uma média de 496 pontos para os países membros da OCDE (Organization for Economic Cooperation and Development).

Apesar do cenário desanimador, desde 2000, o primeiro ano em que o Brasil participou do PISA, o país apresenta evolução positiva no seu desempenho. Em 2000, o desempenho médio dos alunos brasileiros foi de 396 pontos em literatura, 375 em ciências e 334 em matemática. Em 2009, o desempenho médio foi de 412 pontos em literatura, 405 pontos em ciências e 386 em matemática. O país apresentou a terceira melhor evolução entre os países participantes, mas ainda está abaixo da média dos países da OCDE. O Brasil está 81 pontos atrás em literatura, 96 pontos em ciências e 116 pontos em matemática. Dos 65 países participantes, o país ocupou a 53ª posição em 2009.

Além da educação, a saúde também é apresentada como uma forma de capital humano. Grossman (1972) foi o primeiro a modelar a demanda por saúde e a apontar o estado de saúde de um indivíduo como determinante do número de dias saudáveis disponíveis para atividades produtivas no mercado de trabalho e na vida pessoal.

Essa importância da educação e saúde, em alguma medida, se traduz em investimentos que os governos realizam nessas áreas. No Brasil, segundo dados do Instituto Nacional de Estudos e Pesquisas Educacionais (INEP), em 2006 foram gastos em educação R\$ 104,4 bilhões nas esferas Federal, Estadual e Municipal. Em 2008, de acordo com um relatório do Ministério da Saúde, do total disponibilizado para aplicação na área de saúde, foram utilizados, na execução de programas e ações sobre sua responsabilidade, recursos na ordem de R\$ 54 bilhões, o que representou aproximadamente 8,2% da arrecadação da União². Os

² Fonte: Receita Federal e Ministério da Saúde.

gastos com saúde e educação são executados por esferas separadas do governo. No caso do governo federal os gastos são realizados pelo Ministério da Saúde e pelo Ministério da Educação, respectivamente. Apesar de as políticas públicas de saúde e educação serem formuladas por ministérios diferentes, os efeitos de uma extravasam para a área de atuação da outra. Por exemplo, programas destinados à oferta de merenda escolar são formulados pelo Ministério da Educação, mas obviamente tem impacto positivo na saúde. O Ministério da Saúde, por sua vez, ao tentar evitar a transmissão de doenças como a dengue e o vírus HIV, realiza campanhas de conscientização, o que tem efeito positivo na educação da população.

A relação entre educação e saúde vai além dos efeitos de transbordamento de políticas destacados acima. O próprio capital humano individual, constituído pela saúde e educação de um indivíduo, compartilha dessa relação, como sugerido por Grossman (2006). O autor comenta que saúde e educação interagem entre si afetando a produção, o custo e a utilidade um do outro, e até mesmo a dinâmica em que as duas formas de capital humano são obtidas e acumuladas, como por exemplo o período da vida em que isso ocorre, pode interferir nessas relações. Este raciocínio fica claro no caso de duas pessoas idosas, onde uma pessoa recebeu educação sobre os benefícios de uma dieta sadia e balanceada quando jovem e a outra pessoa recebe a mesma educação já na meia-idade. Nesse caso, o nível de saúde do primeiro indivíduo tende a ser melhor que o do segundo indivíduo.

Analisando a literatura internacional, pode-se argumentar que a saúde de um indivíduo tem uma correlação positiva com sua educação (Grossman e Kaestner, 1997). Três maneiras distintas, mas não mutuamente exclusivas, de causalidade podem ocorrer. No primeiro caso, como já explicado por Kenkel (1991), maiores níveis de escolaridade podem ser traduzidos em uma saúde melhor, pois pessoas com melhor educação conseguem ser mais eficientes na “produção” de saúde, obtendo mais facilmente informações sobre prevenção de doenças e utilizando essas informações de forma mais eficaz.

O segundo caso ocorre quando não há de fato uma relação de causalidade direta entre saúde e educação, mas sim uma outra variável, que afeta na mesma direção tanto a saúde quanto a educação do indivíduo. Por exemplo, a escolaridade dos pais pode afetar tanto a saúde quanto a escolaridade de seus filhos. Assim, pais com maiores níveis de escolaridade entendem melhor a necessidade de as crianças estarem bem nutridas e terem uma boa educação e tomam atitudes que afetam positivamente tanto a saúde como a educação dos filhos.

O terceiro caso possível, é aquele em que a saúde afeta positivamente o desempenho escolar. Este caso parece ser especialmente relevante quando o indivíduo se trata de uma criança. Diversas pesquisas³ procuram estimar os parâmetros de uma função de produção educacional onde a variável dependente é associada a alguma medida relacionada com o sucesso escolar, geralmente traduzida em proficiência acadêmica ou frequência escolar. Esta variável dependente é função de um vetor, ou mais comumente um escalar, indicativo da saúde da criança. O interesse das pesquisas está em estimar o coeficiente associado a essa variável, que indica qual é o impacto da saúde no sucesso escolar, controlando-se pelos demais fatores que influenciam este sucesso, como por exemplo, características pessoais da criança, características provenientes de seu ambiente familiar, de sua comunidade e da escola.

Embora a literatura internacional se defronte com uma série de dificuldades de estimação, os resultados parecem apontar para uma relação de causalidade positiva de saúde para desempenho escolar. No Brasil, pelo levantamento da literatura feito até agora, apesar de extensa pesquisa sobre a relação de escolaridade e o trabalho infantil e seus efeitos futuros na saúde e escolaridade das crianças, não há muitos trabalhos que tratam diretamente do efeito da saúde no desempenho escolar de crianças.

No sentido de preencher essa lacuna, o objetivo desta pesquisa é responder a pergunta:

³ Glewwe e Jacoby (1995), Alderman et alli (2001), Glewwe, Jacoby e King (2001), Bobonis, Miguel e Sharma (2006).

Qual o impacto da saúde no desempenho escolar de crianças brasileiras da quarta série do ensino fundamental? Idealmente, para que tal questão fosse respondida, seria necessário o uso de uma base de dados com indicadores de saúde e de desempenho escolar individualizada por aluno. Entretanto, não há uma variável relevante de saúde associada ao aluno em nenhuma base do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), que é responsável pela avaliação do sistema educacional brasileiro com intuito de subsidiar a formulação de políticas públicas para área educacional. Para isso, o INEP realiza exames nacionais de proficiência acompanhados de questionários sócio-econômicos. Algumas secretarias estaduais de educação também realizam seus próprios exames externos de proficiência, mas tampouco foi encontrada nos questionários desses exames qualquer variável indicadora das condições de saúde do aluno.

Para contornar esse problema, foram utilizadas duas diferentes abordagens, ambas utilizando dados de indicadores de saúde agregados em nível municipal e de microrregiões. Na primeira delas, para capturar a situação de saúde das crianças, foram utilizadas algumas *proxies* constituídas por (i) oferta de serviços de saúde nos municípios e microrregiões, como leitos disponíveis e número de estabelecimentos públicos de saúde e (ii) qualidade dos indicadores de saúde nos municípios e microrregiões, como coeficiente de mortalidade infantil e percentual de bebês nascidos vivos abaixo do peso. Estes indicadores são dados provenientes do DATASUS e da Pesquisa de Assistência Médico-Sanitária de 2005 do IBGE.

Usando esse tipo de dado, foram estimados dois tipos de modelos. Inicialmente, as variáveis de desempenho escolar foram regredidas contra as *proxies* de saúde em uma cross-section para 2005. Em seguida, estimou-se um modelo com um painel de dados de microrregiões e dois períodos de tempo (2005 e 2007). Nesse caso, foram usadas como *proxies* de saúde apenas as variáveis de saúde do DATASUS. Os resultados das estimações indicam que mesmo depois de controladas diversas variáveis, há um pequeno efeito positivo e

significante de alguns indicadores de saúde sobre a proficiência escolar, em especial o que indica o percentual de nascidos vivos dado que a mãe realizou 4 ou mais consultas pré-natais. Entretanto, para a maior parte dos indicadores de saúde utilizados, os resultados foram não significativos.

A segunda abordagem usada para contornar a ausência de dados individualizados acerca do estado de saúde dos alunos foi a identificação de municípios que tiveram surto de dengue no ano de 2005. A idéia é que nas localidades onde houve surto de dengue, os alunos tenderiam a faltar mais às aulas, prejudicando assim o seu desempenho escolar. Dessa forma, a hipótese a ser testada é avaliar se o surto de dengue em um município afeta negativamente o desempenho médio dos alunos neste município. Neste caso, a estimação foi feita através do método de pareamento por escore de propensão. Os resultados, embora não sejam robustos sob todas as especificações do modelo de pareamento, indicam que um surto de dengue no ano de 2005 diminuiu em cerca de 8 pontos o desempenho médio dos alunos dos municípios afetados, o que corresponde a cerca de 20% do desvio padrão da nota na escala do SAEB.

Capítulo 1 - Revisão da Literatura

Esta seção tem como objetivo discutir e apresentar os principais trabalhos e o desenvolvimento da literatura de Economia da Educação, destacando a relação entre saúde e produtividade de um indivíduo. Será abordado tanto o caso dos adultos, onde a produtividade é medida em termos salariais e de eficiência no trabalho, como o caso das crianças, onde a produtividade se traduz mais claramente em desempenho escolar.

Estudos sobre os efeitos do capital humano em atividades produtivas fora do mercado de trabalho foram iniciados na década de 60, após o reconhecimento por parte dos economistas de que as decisões e resultados obtidos no âmbito pessoal e familiar poderiam ser modeladas adequadamente com ferramentas usualmente utilizadas na área. Dentre os

primeiros estudos feitos, ganha destaque os realizados por Gary Becker, onde é estabelecida a importância e a dinâmica da produtividade em áreas fora do mercado de trabalho, como o âmbito familiar.

Dessa forma, ganharam importância na literatura econômica os estudos sobre a produção de *commodities*. Essa denominação foi cunhada por Becker (1965) com o intuito de descrever a utilização de bens e serviços adquiridos, mais o tempo do indivíduo, em atividades que entram diretamente em sua função de utilidade. Atividades de lazer como passear com a família pelo parque e praticar atividades esportivas, por exemplo, requerem meio de locomoção até o local, serviço de manutenção da área de lazer e saúde adequada para o desempenho de certas atividades, entre outros insumos, além de tempo.

Becker também sugeriu a hipótese de que um número maior de anos de escolaridade aumenta a eficiência na produção de *commodities*. Com isso, a importância da educação ganhou maiores proporções, uma vez que tanto no mercado de trabalho quando em outros aspectos da vida pessoal de um indivíduo, a educação possui um impacto positivo.

Sendo assim, o restante deste capítulo divide-se em duas partes. A primeira delas aborda brevemente o desenvolvimento da literatura sobre Economia da Educação, enquanto que a segunda apresenta o desenvolvimento de estudos relacionados à saúde e sua ligação com a produtividade de um indivíduo, seja no efeito sobre salários ou no efeito sobre desempenho escolar.

1.1 Educação

Estudos econômicos na área de educação tiveram como marco inicial o trabalho de Coleman et al. (1966), conhecido como *Coleman Report*. O trabalho realizou uma extensa pesquisa sobre educação nos Estados Unidos reunindo dados até então escassos e de difícil acesso. O estudo contou com a ajuda de cerca de vinte mil professores em todo o país, em

quatro mil escolas, com 645 mil alunos. Ele teve como foco identificar: i) qual o nível de segregação em escolas públicas, ii) se havia equidade da qualidade de ensino entre as escolas, iii) qual o desempenho dos alunos, medido em testes padronizados, e iv) qual era a relação entre o desempenho dos alunos e o tipo de escola que freqüentavam. O estudo forneceu uma visão geral e muito ampla do sistema educacional americano, mas talvez seu principal resultado tenha sido indicar que o *background* socioeconômico dos alunos, ou seja, características referentes ao ambiente familiar do aluno, responde por 80% do seu desempenho escolar. Características da escola, como sua infra-estrutura, responderia por menos de 2% do desempenho dos alunos, enquanto que professores contribuem com menos de 4%.

O trabalho de Coleman também desenvolveu o modelo que passou a ser utilizado na maior parte dos estudos posteriores sobre educação, com uma função de produção educacional do tipo $y = F(c, m, g, p, s)$, onde o desempenho escolar é função das características do aluno, das características de sua família, de seus colegas de classe, do professor do aluno e de outras características escolares, respectivamente. Outros estudos subsequentes utilizam variáveis diferentes de acordo com o interesse da pesquisa. Por exemplo, no lugar de desempenho escolar como variável explicada, alguns estudos utilizam a frequência escolar ou anos de escolaridade⁴ enquanto outros estudos adicionam variáveis explicativas mais específicas, como as características da comunidade em que o aluno reside, ou um indicador do seu nível de saúde ou o grau de exposição à segregação racial⁵ ou violência.

Alguns estudos, ainda na década de 60 e 70, contestaram os resultados apresentados no *Coleman Report*, indicando erros de estimação que poderiam invalidar os resultados, porém a grande importância do *background* familiar frente a outros aspectos relacionados ao

⁴ Ver Alderman et alli (2001) Arendt (2005). Para o caso brasileiro, ver Barros, Mendonça e Santos (1999). O estudo serviu como um marco para a compreensão do cenário educacional brasileiro.

⁵ Ver Rumberger e Willms (1992) e Bifulco e Ladd (2006). Para dados brasileiros, ver Soares e Alves (2003).

aluno parece ser um resultado robusto e é frequentemente encontrado em outros estudos na literatura.

No caso particular do Brasil, o estudo de Albernaz, Ferreira e Franco (2002) utilizou dados do SAEB 1999 para alunos da oitava série do ensino fundamental e indicou que 80% da variância do desempenho médio entre escolas se deve ao *background* sócio-econômico dos alunos.

Mesmo admitindo-se a hipótese da reduzida importância de outros fatores como infraestrutura da escola e características dos professores, uma série de estudos procurou estimar qual o real impacto desses e de diversos outros fatores no desempenho escolar dos alunos. Esse interesse em estudar tais componentes pode ser explicado pelo fato de que esses insumos estão sob o controle direto dos formuladores de políticas públicas. Portanto, se determinado equipamento em uma escola ou se determinado treinamento para professor apresentar um impacto positivo e significativo na nota do aluno, é possível fazer as mudanças necessárias no curto prazo para que o desempenho médio dos alunos melhore.

Um dos estudos que procura analisar outros fatores importantes além do *background* familiar é o de Felício e Fernandes (2005), que utilizaram dados do SAEB 2001 no Estado de São Paulo, com alunos da quarta série do ensino fundamental e que teve como um de seus objetivos avaliar qual o papel da escola no desempenho dos alunos. A preocupação dos autores foi em demonstrar que a relação entre desempenho escolar e a qualidade da escola pode ter sido subestimada em estudos anteriores, o que daria uma importância menor para este fator com relação ao ambiente sócio-econômico familiar. Para contornar esse problema, as *proxies* utilizadas para indicar a qualidade do ensino nas escolas foram mais cuidadosamente selecionadas. Ainda assim, os resultados indicam que nos testes de matemática, a escola pode ser responsável por uma diferença entre 8,7% e 34,4% nas notas. Nos testes de português a diferença pode chegar a 28,4%, sendo o restante explicado pelo

background familiar dos alunos.

Como dito anteriormente, embora o ambiente sócio-econômico familiar seja responsável por uma parte considerável do desempenho dos alunos, o fato de que outros fatores estarem mais facilmente sob o controle dos formuladores de políticas públicas faz com que análises de diferentes fatores e sua importância nas notas sejam de interesse da sociedade. Sendo assim, uma vez que a oferta de serviços de saúde e de programas de acompanhamento da saúde de crianças está sob controle do Estado, descobrir qual é exatamente a importância da saúde dos alunos no seu desempenho escolar pode ajudar a entender melhor como elevar a qualidade do ensino brasileiro. Afinal, saúde e educação são os dois principais componentes do capital humano, e como existem relações entre esses dois componentes, tentar entendê-las é essencial. Na seção seguinte, procura-se descrever quais as formas e mecanismos pelo qual a saúde pode afetar o desempenho de um indivíduo e o desempenho escolar de uma criança. A partir da compreensão da maneira como saúde afeta educação, se torna mais fácil a formulação de políticas públicas com o intuito de elevar o nível de saúde das crianças.

1.2 Saúde e produtividade

Durante a década de 60 e meados da década de 70, já existia a idéia de que a saúde era, assim como a educação, uma forma de capital humano (Mushkin 1962, Becker 1964 e Fuchs 1966). O trabalho pioneiro de Grossman (1972) contribuiu com um modelo original para a demanda por saúde e que serviu de base para diversos estudos nos anos subsequentes. O autor argumenta que o capital humano associado à saúde funciona de um modo diferente com relação a outras formas de capital humano. O estudo partiu da hipótese de que a saúde impactava diretamente apenas nos dias saudáveis disponíveis para atividades produtivas tanto no mercado de trabalho, gerando renda, quanto na vida pessoal, por exemplo, tendo mais

disposição para atividades que geram bem-estar, o que resultaria em impactos positivos na sua função de utilidade. A hipótese de que a saúde pode afetar a produtividade e o salário de um trabalhador, que é exatamente o que ocorre com a educação, embora seja admitida pelo autor, não é explorada por ele nesse estudo.

Esta hipótese, não explorada por Grossman, passou a ter respaldo na comunidade acadêmica somente anos depois. Strauss (1986) analisou uma amostra de famílias agricultoras em Serra Leoa e chegou à conclusão de que uma ingestão adequada de calorias tinha correlação positiva com a produtividade das famílias. Deolalikar (1988), utilizando dados em painel para uma região rural no sul da Índia, chegou à conclusão de que os indicadores para saúde possuem alta correlação com a produtividade das fazendas e com o salário dos indivíduos.

É necessário frisar o fato de que existem diversos indicadores de saúde na literatura sobre o tema, cada um mostrando um aspecto da saúde de um indivíduo. Na maioria dos estudos, devido à dificuldade em se obter dados sobre saúde, utilizam-se poucos indicadores, como o número de calorias e níveis de nutrientes ingeridos, ou o índice de massa corpórea, calculado com o peso e altura de um indivíduo. Esses diferentes indicadores de saúde indicam diferentes aspectos da saúde, que é mais ampla do que qualquer indicador é capaz de descrever. No caso de crianças, por exemplo, o número de dias, no presente mês, em que há ocorrência de diarreia e peso abaixo do recomendado indica apenas o estado de saúde atual da criança, enquanto que sua altura é um indicador do seu quadro de saúde em anos anteriores, uma vez que níveis baixos de saúde por tempo prolongado prejudicam o crescimento. Porém, a altura não diz muito sobre o presente nível de saúde da criança.

Essa distinção entre os indicadores de saúde é importante. Para entender melhor como o “estoque” de saúde e o nível atual de saúde, ou “fluxo” de saúde, funcionam, suponha uma função de produção qualquer de um indivíduo adulto, $Y_t = f(H_t, H_{t-1}, \alpha, T, E)$. Considerando

Y sua produção, que é função do seu nível de saúde atual (H_t), do seu nível de saúde no passado ou seu estoque de saúde (H_{t-1}), de sua habilidade inata (α), das condições do ambiente de trabalho (T) e de sua escolaridade (E).

O nível de saúde atual em um dado dia, H_t , poderia nos indicar se o indivíduo está com dor de cabeça, mal-estar ou com azia. Já seu estoque de saúde, H_{t-1} , poderia indicar se o indivíduo possui dor crônica na coluna ou algum outro tipo de desconforto crônico, além de simplesmente indicar o estado de saúde no período passado.

Entender qual é o impacto de diferentes indicadores de saúde em determinada variável pode estabelecer mais claramente o caminho pelo qual as políticas públicas podem ser formuladas. Tomar uma aspirina para controlar uma dor de cabeça certamente é menos complexo do que uma intervenção cirúrgica para amenizar dores na coluna. Da mesma forma, se tratando de crianças, políticas públicas que se concentram em resultados de curto prazo na saúde das crianças são mais facilmente implementadas, e seus resultados mais facilmente observados, do que aquelas políticas em que o foco é no desenvolvimento cognitivo adequado das crianças. Resolver casos de diarreias, desnutrição e mortes por falta de higienização pós-parto é relativamente mais fácil do que assegurar o crescimento saudável das crianças, um processo que demoraria anos.

De acordo com Thomas e Strauss (1997), a dificuldade em se obter dados contendo variáveis indicadoras de saúde, tanto de curto quanto de longo prazo, e medidas de produtividade ou salário fez com que estudos sobre o tema fossem raros até a segunda metade da década de 90. Além disso, a maior parte dos estudos se concentrava na estimação da correlação entre saúde e produtividade ou salário, dando pouca importância para a direção de causalidade. Era ignorado o fato de que a renda pode afetar indiretamente a saúde, mudando o comportamento do indivíduo, que passa a consumir uma dieta diferente, por exemplo, e isto pode vir a afetar seu salário.

Thomas e Strauss (1997) argumentam também que, em estudos anteriores, a utilização de poucas variáveis indicadoras de saúde e o foco em analisar o setor rural limitava a validade das pesquisas. Os autores procuraram corrigir essas deficiências analisando dados de áreas urbanas no Sul e Nordeste do Brasil e utilizando quatro medidas indicadoras de saúde: altura, índice de massa corpórea, ingestão *per capita* de calorias e de proteínas. Os resultados mostram que todos os indicadores de saúde impactam positivamente no salário de homens e mulheres no mercado de trabalho.

Os estudos comentados até agora nesta seção demonstraram a relação entre saúde e produtividade do indivíduo adulto, onde diferenciais de produtividade podem ser observados na diferença salarial. Mas como mencionado anteriormente, a produtividade em uma criança pode ser mais facilmente medido através de medidas de sucesso escolar, em especial o desempenho escolar. Sobre o efeito da saúde na educação, Glewwe e Miguel (2008) descrevem claramente a importância dos níveis de saúde, em diferentes períodos da vida, no desempenho escolar dos alunos. Os autores “dividem” a vida do aluno em dois períodos, onde a variável explicada, o desempenho acadêmico, é medido no segundo período, mas o nível de saúde no primeiro e no segundo período entram separadas como variáveis explicativas.

Ainda nessa vertente de pesquisa que trabalha com a relação entre saúde e educação, outros estudos exploraram a interação entre indivíduos em uma família, mais especificamente o efeito que os pais exercem em seus filhos. Os pais, responsáveis por seus filhos, possuem papel fundamental na formação de capital humano das crianças. Decisões sobre os níveis de educação e saúde das crianças recaem essencialmente sobre os pais, uma vez que crianças não possuem condições financeiras e desenvolvimento intelectual suficiente para escolher seus próprios níveis de educação e saúde.

Kassouf e Senauer (1996) ao analisar a influência dos pais na determinação do capital humano familiar, sugerem a existência de correlação positiva entre educação dos pais e a

saúde e educação da criança, onde pais com maior educação possuem filhos com melhor estado de saúde e melhor educação. Glewwe (1999) observa que a escolaridade da mãe é mais importante na determinação da saúde dos filhos do que a escolaridade dos pais.

A influência da educação dos pais na saúde dos filhos não é óbvia, nem tão direta como pode parecer. Além disso, conhecer as formas como as variáveis se relacionam entre si é importante para que as estimações dos efeitos sejam feitas de modo não viesado e os resultados apresentados sejam precisos.

Existem três hipóteses para explicar como a educação dos pais pode afetar a saúde dos seus filhos. A primeira delas está relacionada à educação formal, mais especificamente à produção familiar do insumo saúde, e pode ser dividida em duas: pessoas mais formalmente educadas obtêm informações mais detalhadas sobre prevenção, diagnóstico e tratamento de doenças. Nesse aspecto, Glewwe (1999) estudou no Marrocos a relação entre escolaridade das mães e a saúde dos filhos e descobriu que as habilidades de leitura e álgebra obtidas na escola ajudavam as mães a interpretar melhor as informações de rótulos de remédios assim como identificar e tratar doenças mais eficientemente. A outra relação de educação formal e saúde indica que pessoas mais formalmente educadas possuem maior capacidade de buscar informações e transformá-las em conhecimento para melhoria da saúde de seus filhos (Thomas, Strauss e Henriques, 1991).

A segunda hipótese está relacionada ao ambiente social, ou seja, à aquisição, no mercado, de insumos para a produção de saúde. Assim, pais mais educados conseguem tirar maior vantagem das clínicas locais e programas de nutrição (Thomas, Strauss e Henriques, 1991), bem como possuem maior facilidade de mudar de métodos tradicionais para métodos mais modernos de diagnóstico e tratamento (Glewwe, 1999).

A última hipótese é a do efeito indireto da renda, como apontado por Thomas e Strauss (1997). Um dos fatores que determina o investimento em saúde dentro da família é a renda.

Nota-se que pais mais educados possuem maior salário e maior renda, de forma que parte do efeito do aumento da educação sobre a saúde advém indiretamente do aumento da renda (Kassouf e Senauer, 1996; Thomas, Strauss e Henriques, 1991).

Como já mencionado, é importante ter conhecimento dessas hipóteses e de como as variáveis se relacionam. Dependendo dessas relações, estimações por Mínimos Quadrados Ordinários de uma função de produção educacional podem ser viesadas de três maneiras: viés de simultaneidade, onde uma variável afeta ao mesmo tempo saúde e educação, que é o caso da renda familiar, por exemplo; viés de erros de medida, que ocorre quando a saúde da criança não é propriamente medida; e viés causado pela correlação da variável indicadora de saúde com algum termo presente no erro (Behrman e Lavy, 1997).

Muitos estudos, como apontam Glewwe e Miguel (2008), procuram contornar o problema de viés causado por correlação entre a variável indicadora de saúde e o erro, assim como o viés causado por erro de medida, utilizando variáveis instrumentais. Um exemplo é o estudo de Ding *et al.* (2009), que utiliza genes de estudantes nos Estados Unidos como variáveis predeterminadas, pois estes genes foram identificados na pesquisa do genoma humano como determinantes de doenças que resultam em má qualidade de vida, como depressão, déficit de atenção/hiperatividade e obesidade e, uma vez que os genes são gerados na concepção, não há nenhuma correlação destes com variáveis de escolha, como por exemplo a escola em que os indivíduos irão estudar ou a comunidade em que irão residir. Levando isso em conta, os genes tornaram-se variáveis instrumentais adequadas na estimação do impacto da saúde no desempenho escolar, que é o objeto de estudo dos autores. Eles alertam, no entanto, para a possível correlação entre os genes utilizados no estudo e o desempenho escolar futuro, apesar de que os resultados demonstram ser robustos diante da possível violação de exogeneidade. As conclusões são de que obesidade e depressão possuem impacto negativo significativo no desempenho escolar.

Para controlar o viés causado pela correlação de variáveis não observáveis, presentes no erro, com a variável indicadora de saúde, Behrman e Lavy (1997) utilizaram-se da técnica de efeitos-fixos em uma *cross-section*. O objetivo do estudo era averiguar o impacto da saúde, medido pela altura, no desempenho escolar de alunos de Gana. Para controlar variáveis não observáveis da família e da comunidade em que as crianças residiam, os autores selecionaram uma amostra onde cada aluno possuía pelo menos um irmão. Assim, foi possível obter mais de uma observação da família e da comunidade e controlar para efeitos não observáveis que possuíam correlação com o desempenho escolar.

O estudo anterior conseguiu controlar o efeito de variáveis não observáveis da família e da comunidade, mas pelo fato de sua base de dados ser retirada de uma *cross-section*, os autores foram incapazes de controlar para variáveis não observáveis do indivíduo. Glewwe e Miguel (2008) argumentam que esse problema pode ser contornado através da utilização de dados em painel. Se o pesquisador possui dados com duas observações do mesmo indivíduo, variáveis não observáveis fixas no tempo podem ser diferenciadas e, portanto, não precisariam ser efetivamente medidas. Ainda assim, os autores alertam para as suposições necessárias para que as estimações sejam precisas, por exemplo, a de que essas variáveis não observáveis sejam, de fato, fixas, e de que elas não sejam correlacionadas com variáveis que mudam no tempo.

Como se pode ver, a literatura internacional teve avanço considerável nas estimações do efeito da saúde no desempenho escolar. Para dados brasileiros, a literatura parece não ter avançado muito no tema. Talvez isso seja explicado pelo fato de que não existe um número expressivo de bases de dados contendo simultaneamente uma variável indicadora de desempenho escolar e outra variável indicadora de saúde, como é encontrado nos diversos estudos internacionais já apresentados nessa seção.

Uma exceção é o estudo de Hanushek *et al* (1997), realizado com dados da área rural

do Ceará. Alunos da segunda série foram consultados em 1985 e novamente em 1987, quando supostamente deveriam estar na quarta série, permitindo a possibilidade da criação de um painel, ainda que com poucas observações (378 alunos). Informações foram coletadas sobre a família dos alunos, sua altura, peso, capacidade visual e índice de nutrição, assim como seu desempenho escolar. Os autores se concentraram em três medidas de sucesso escolar: i) evasão escolar, ii) probabilidade de aprovação à série seguinte e iii) desempenho escolar. Nas três medidas de sucesso a saúde parece ter impacto positivo, em especial a capacidade visual dos alunos.

Infelizmente, o estudo de Hanushek *et al.* (1997) não foi acompanhado de novos estudos, tampouco houve a realização de novas coletas de dados permitindo estudar os efeitos da saúde no desempenho escolar. Para contornar o problema da falta de uma base de dados adequada, o presente estudo utilizará o conjunto de serviços de saúde e procedimentos médicos disponíveis em uma microrregião como *proxy* da saúde das crianças que ali residem. Isso impossibilita a avaliação do impacto individual da saúde de um aluno no seu desempenho escolar, tornando necessário trabalhar com a média do desempenho dos alunos de uma determinada região. Essa prática é comum nos estudos sobre criminalidade. Hartung e Pessoa (2007) utilizam dados de virtualmente todos os municípios no Estado de São Paulo e demonstram que os resultados já estabelecidos na literatura para dados individuais, por exemplo a indicação de que filhos de mães adolescentes e solteiras tem maior probabilidade de cometerem crimes, se mantêm quando são utilizados dados agregados por município, isto é, o percentual de mães adolescentes e solteiras em uma determinada cidade pode explicar em parte o índice de criminalidade naquela cidade.

Este estudo tentará, portanto, avaliar a hipótese de que a saúde afeta a educação das crianças brasileiras e, através de uma maneira inovadora, espera-se preencher uma lacuna pouco explorada na literatura nacional que poderá abrir espaço para novas discussões e

futuros estudos relacionando a saúde e o desempenho escolar de crianças brasileiras.

Capítulo 2 – Dados

Esta seção está dividida em três partes. A primeira parte contém informações sobre as fontes dos dados de 2005 utilizados para regressões de indicadores de saúde e qualidade da infra-estrutura como *proxies* de saúde. As variáveis de controle também são analisadas nesta seção. Em seguida há uma breve subseção com análises descritivas destas variáveis.

A segunda parte traz uma breve descrição da estrutura do painel de dados. Nesse caso, utilizou-se as Provas Brasil, os Censos Escolares e informações do DATASUS para 2005 e 2007. A pesquisa Assistência Médico-Sanitária não foi utilizada, pois sua última realização se deu em 2005 apenas, não havendo dados para 2007.

A terceira parte desta seção traz informações sobre as fontes dos dados que indicam surtos de dengue ocorridos em 2005, assim como explicações sobre o tratamento dos dados.

Nas regressões em cross-section, os dados que formam o vetor do desempenho escolar são provenientes da Prova Brasil 2005, que é um exame nacional de proficiência em matemática e língua portuguesa dos alunos das escolas públicas de ensino fundamental em área urbana. Todos os estudantes na quarta e oitava série de escolas com mais de 20 alunos fazem o teste. Os dados da prova permitem avaliar tanto o indivíduo quanto a média do desempenho de sua escola, município, região e unidade da Federação. Além da prova de proficiência, foi aplicado um questionário sócio-econômico para captar características do aluno, do seu professor, dos seus pais e da sua residência. Esse questionário foi utilizado para compor a matriz de variáveis explicativas, que contém informações do aluno, de seu *background* familiar e características de seu professor, assim como informações sobre a escola que o aluno frequenta.

Para complementar as informações sobre a escola, foi utilizado o Censo Escolar 2005.

Realizado anualmente, o censo coleta dados sobre estabelecimentos, matrículas, funções docentes e infra-estrutura de escolas públicas de ensino infantil, fundamental e médio em todo o Brasil.

2.1 Dados – Indicadores e qualidade da infra-estrutura de saúde em 2005

Apesar de todas as informações estarem disponíveis por município, este trabalho utilizará os dados agregados por microrregião, segundo o IBGE. Espera-se que assim seja controlado o efeito de “migração” dos residentes de municípios pequenos para municípios maiores com maior e melhor infra-estrutura de saúde em busca de tratamento médico.

As informações que formam o vetor de interesse, a “saúde média” de cada microrregião, são provenientes de duas bases de dados. A primeira delas é a pesquisa Assistência Médico Sanitária 2005 (AMS), que foi realizada pelo IBGE e descreve a oferta de serviços de saúde e a capacidade instalada no Brasil. Esta pesquisa tem caráter censitário com informação sobre os estabelecimentos de saúde por município e microrregião com regime ambulatorial, internação ou de serviço de apoio a diagnósticos e terapia. Com o apoio do Ministério da Saúde, essa pesquisa foi realizada em 1999, 2002 e 2005, possuindo assim a evolução dos estabelecimentos de saúde. A pesquisa de 2005 coletou informações de mais de 77 mil estabelecimentos de saúde em todo o território nacional. Infelizmente, o IBGE disponibilizou os microdados por município apenas para algumas das informações contidas na pesquisa, o que restringe a capacidade de avaliação de todos os indicadores de saúde.

A segunda fonte de informações é proveniente do DATASUS, que é a base de dados do Sistema Único de Saúde do Ministério da Saúde, criada para armazenar as informações de saúde para controle, gestão e planejamento. Esse sistema possui uma grande quantidade de informação que vai desde o financiamento até o tipo de procedimento realizado. Os dados podem ser desagregados por município, levando em consideração o local de residência do

indivíduo atendido. Têm-se informações ambulatoriais fornecidas pelo sistema de informação ambulatorial (SIA) e informações hospitalares fornecidas pela autorização de internação hospitalar (AIH).

Para formulação das tabelas com estatísticas descritivas na seção seguinte, foi utilizada uma base contendo apenas informações do Censo Escolar 2005 e da Prova Brasil 2005, desagregados por aluno. Dessa forma, as médias, desvios-padrão e demais informações nas tabelas 1 a 6 referem-se aos alunos, e não aos municípios ou microrregiões.

Para a elaboração das estatísticas descritivas foram criadas variáveis *dummy* para indicar a ocorrência de determinada característica, como por exemplo o aluno ser do sexo feminino, ou então ter 11 anos de idade. Isto se deve ao fato de que os dados seriam posteriormente agrupados por município, e em seguida em microrregiões, gerando então o percentual da amostra com determinada característica. Além da criação de variáveis *dummy*, também foram eliminadas da amostra as observações que não contavam com informações sobre gênero, escolaridade dos pais, idade e outras informações referentes ao aluno e seu *background* familiar. A amostra inicial contava com 3.392.880 alunos da quarta e oitava série do ensino fundamental, distribuídos em 5.387 municípios e a amostra final contou com 1.596.687 alunos da quarta série em 4.947 municípios de todo o Brasil, agregados em 554 microrregiões.

Posteriormente, as informações das quatro fontes de dados, DATASUS, Pesquisa AMS, Censo Escolar e Prova Brasil, foram reunidas em uma única base de dados para o ano de 2005. A partir dessa base de dados agregada por microrregião foi possível realizar regressões do desempenho escolar contra diversas medidas de saúde e diferentes grupos de controle.

2.1.1 Análise Descritiva dos dados da Prova Brasil 2005

A nota média dos alunos da quarta série na prova de matemática foi 183,64 pontos, com um desvio padrão de 39,4 pontos, com pontuação mínima de 78,09 pontos e máxima de 330,65. Os 5% piores alunos, representados na coluna “P5”, tiraram menos de 122,5 pontos, em uma escala de 0 a 500, enquanto que os 5% melhores, representados na coluna “P95” tiraram mais de 253,20 pontos.

As notas de português foram mais baixas, com média de 176,53 pontos e desvio padrão de 41,25 pontos. Observa-se também, em comparação com as notas de matemática, uma redução tanto das notas mínimas quanto da máxima, 74,84 e 324,62 pontos, respectivamente. Os 5% piores alunos tiveram nota inferior a 112,88 pontos e os 5% melhores fizeram mais que 248,93 pontos. As tabela 1 e 2, a seguir, apresenta estes dados e outras informações referentes a características dos alunos e de suas famílias. Os valores de pontuação máxima e mínima foram suprimidos das tabelas pois apresentavam pouca variação entre os diferentes perfis listados.

Foram feitos testes de diferença de médias e a maior parte foi significativa a 1%, indicando que de fato havia diferença entre as médias. As exceções, não significantes até mesmo a 10%, foram a diferença entre médias de cor branca contra indígenas, que não é significativa, assim como a diferença das médias de alunos com 8 e 14 anos e também entre 13 e 15 anos. Testes individuais para cada variável também foram feitos, onde cada média da variável binária foi testada contra a média do restante da amostra. Estes testes indicaram diferença significativa, ou seja, todas as variáveis tinham média diferente do restante da amostra em conjunto.

Tabela 1 – Estatísticas descritivas de características do aluno (matemática - 4ª série)

Variável		(%)	Média	Desv. Padrão	P5	P95
Brasil		100	183,64	39,4	122,5	253,2
Sexo	Masc	50,53	183,98	40,7	120,43	255,29
	Fem	49,47	183,30	38,04	124,73	250,58
Cor	Branco	35,85	189,47	41,40	125,13	261,48
	Pardo	46,98	182,81	37,97	123,48	249,33
	Preto	9,96	170,69	35,43	115,04	233,33
	Amarelo	2,88	178,29	39,86	116,72	249,43
	Indígena	4,33	177,94	36,80	120,01	242,83
Idade	8 anos	1,02	165,01	38,80	106,02	233,78
	9 anos	4,68	185,66	39,49	124,00	254,49
	10 anos	47,10	192,02	39,92	129,71	260,87
	11 anos	26,59	181,25	38,10	122,08	248,80
	12 anos	10,67	169,51	33,50	115,99	227,07
	13 anos	5,17	166,95	33,42	113,79	224,64
	14 anos	2,58	165,37	34,15	111,79	224,63
	15 anos ou mais	2,19	166,76	37,95	110,23	236,38
Já foi reprovado?	Nunca	69,96	189,80	40,03	127,26	259,11
	Uma vez	21,73	169,52	33,72	115,62	227,79
	Mais de uma vez	8,31	168,78	34,07	114,35	227,65

Fonte: Elaboração própria com base nos dados da Prova Brasil 2005.

As médias dos diferentes perfis de alunos na prova de português foram em sua maior parte significativas, exceto pela diferença entre alunos de 8 e 13 anos e também de alunos de 14 e 15 anos.

É possível notar que a média para alunos do sexo masculino e feminino é parecida na prova de matemática, porém as meninas têm desempenho superior em cerca de 9 pontos na prova de português. Tanto na prova de matemática quanto na de português os alunos com 9 e 10 anos têm desempenho superior à média. Alunos com 11 anos têm desempenho próximo da média e alunos com 12 anos ou mais têm desempenho inferior. Os alunos de cor branca, que representam 35,85% da amostra, têm desempenho superior aos demais grupos étnicos. O fato do aluno não ter reprovação também indica um melhor desempenho, comparado com alunos

que já foram reprovados.

Tabela 2 – Estatísticas descritivas de características do aluno (português - 4ª série)

Variável		(%)	Média	Desv. Padrão	P5	P95
Brasil		100	176,53	41,25	112,88	248,92
Sexo	Masc	50,53	172,14	40,94	110,26	244,83
	Fem	49,47	181,02	41,08	116,45	235,15
Cor	Branco	35,85	182,31	43,04	115,2	257,3
	Pardo	46,98	175,83	39,92	113,64	245,51
	Preto	9,96	162,78	37,29	106,96	228,86
	Amarelo	2,88	170,01	41,27	108,19	243,34
	Indígena	4,33	172,3	39,56	110,96	241,46
Idade	8 anos	1,02	157,46	39,65	101,63	229,52
	9 anos	4,68	180,00	40,73	115,51	250,07
	10 anos	47,1	185,95	41,47	119,87	257,40
	11 anos	26,59	174,08	40,13	112,67	245,11
	12 anos	10,67	160,40	34,80	107,13	220,59
	13 anos	5,17	157,00	34,42	105,03	216,70
	14 anos	2,58	155,10	34,70	104,00	216,32
	15 anos	2,19	155,17	37,69	102,39	224,80
Reprovado	Nao	69,96	183,70	41,76	117,44	255,89
	uma vez	21,73	160,29	34,68	107,09	220,152
	mais de uma vez	8,31	158,61	34,66	105,74	218,57

Fonte: Elaboração própria com base nos dados da Prova Brasil 2005.

As tabelas 3 e 4 trazem informações referentes às características da residência e da família dos alunos. A variável “água” mostra, por exemplo, que os alunos cuja residência possui água encanada (91,29%) possuem um desempenho médio de 185,06 pontos na prova de matemática, enquanto que aqueles que não possuem (8,71%) têm uma média de 168,88 pontos. Alunos cuja residência não possui energia elétrica representam 2,95% da amostra e têm um desempenho médio de 159,9 pontos, bem abaixo da média da amostra total, 183,64.

É de se notar, tanto na prova de português quanto na de matemática, o fato de que alunos cuja residência possui mais de três quartos têm um desempenho inferior, na média, em relação aos alunos cuja residência possui três quartos.

Tabela 3 – Estatísticas descritivas do *background* familiar do aluno (matemática, - 4ª série)

Variável		(%)	Média	Desv.Padrão	P5	P95
Água	Sim	91,29	185,06	39,51	123,72	254,45
	Nao	8,71	168,88	35,03	113,77	230,45
Energia	Sim	97,05	184,37	39,32	123,39	253,64
	Não	2,95	159,9	34,51	106,79	221,07
TV	Mais de uma	37,71	191,58	40,63	127,83	261,77
	Uma	55,10	180,10	37,95	121,15	247,18
	Não tem	7,19	169,18	35,63	113,53	232,02
Rádio	Mais de um	34,43	190,33	40,57	126,81	260,39
	Um	56,88	180,86	38,46	121,23	248,97
	Não tem	8,69	175,41	37,04	117,39	240,76
Banheiro	Mais de um	24,41	189,71	41,25	125,44	261,58
	Um	72,87	182,40	38,53	122,55	250,27
	Não tem	2,72	162,62	35,13	108,34	224,63
Número De Quartos	Quatro ou mais	13,57	182,73	39,55	121,57	252,56
	Três	31,36	187,38	39,93	125,77	257,62
	Dois	39,46	184,56	38,88	124,33	253,15
	Um	14,31	175,59	38,16	115,83	243,23
	Nenhum	1,30	163,99	35,93	108,36	227,05
Carro	Mais de um	13,14	184,69	41,28	120,83	257,74
	Um	32,56	188,21	41,00	124,15	259,45
	Nenhum	54,3	180,66	37,65	121,97	246,88
Mora com	Pai e mãe	66,23	185,29	39,79	123,76	255,25
	Só com a mãe	19,89	182,55	38,80	121,83	250,64
	Só com o pai	3,44	175,47	38,92	114,14	243,98
	Outros	10,44	177,99	37,26	119,72	243,66
Escolaridade da mãe	Ensino Superior	10,77	190,73	41,27	127,23	263,48
	Ensino Médio	26,64	194,45	41,39	129,35	265,43
	Ensino Fundamental	13,15	183,82	39,54	122,2	253,25
	Até 4a série do Ens. Fund. Nunca estudou ou não completou a 4a série do Ens. Fund.	13,55	181,66	38,11	122,03	248,74
		8,80	173,40	37,28	115,30	239,29
	Aluno não sabe	27,09	181,88	38,20	122,62	249,37
Escolaridade do pai	Ensino Superior	9,94	189,93	40,66	127,41	261,72
	Ensino Médio	20,48	193,03	41,95	127,2	265,37
	Ensino Fundamental	11,99	183,60	39,62	121,8	253,34
	Até 4a série do Ens. Fund. Nunca estudou ou não completou a 4a série do Ens. Fund.	11,09	182,36	38,76	121,8	250,59
		9,14	173,64	37,31	115,41	239,57
	Aluno não sabe	37,36	182,71	38,31	123,25	250,13

Fonte: Elaboração própria com base nos dados da Prova Brasil 2005.

Também destaca-se o fato de que alunos cuja família possui mais de um carro e cuja mãe ou pai possui ensino superior têm desempenho inferior em relação aos alunos cuja família possui um carro e o pai ou a mãe estudou apenas até completar o Ensino Médio. Os testes de diferença de médias parecem validar o resultado, demonstrando a um nível de 1% que as médias são de fato diferentes. Isto parece ir de encontro a outros resultados na literatura, que apontam que a relação positiva entre escolaridade dos pais e desempenho escolar.

Outras variáveis, no entanto, estão de acordo com o que é comumente encontrado em outros estudos na literatura. Variáveis que possivelmente indicam alguma relação com a renda da família, como o número de rádios na residência, assim como número de banheiros e de aparelhos televisores, têm relação positiva com o desempenho. Como exemplo, pode-se ressaltar que alunos cuja residência possui mais de um aparelho televisor, que representam 37,71% da amostra, têm um desempenho médio de 191,58 pontos na prova de matemática e de 184,65 pontos na prova de português, enquanto alunos que moram em residências com apenas um aparelho televisor (55,1%) têm desempenho médio de 180,1 pontos em matemática e 173,13 pontos em português. Alunos que não possuem aparelho televisor (7,19%) têm o pior desempenho médio, de apenas 160 pontos. Quando o aluno mora com o pai e a mãe, seu desempenho médio é superior à média, enquanto que se o aluno mora somente com o pai, seu desempenho é inferior até mesmo aos alunos que moram com outras pessoas.

Tabela 4 – Estatísticas descritivas do *background* familiar do aluno (português - 4ª série, continua)

Variavel		(%)	Média	Desv.Padrão	P5	P95
Água	Sim	91,29	178,11	41,39	113,87	250,5
	Nao	8,71	160,02	35,79	106,06	222,84
Energia	Sim	97,05	177,37	41,15	113,66	249,46

	Não	2,95	149,02	34,28	100,6	211,03
TV	Mais de uma	37,71	184,65	42,55	117,17	258,2
	Uma	55,1	173,13	39,81	111,84	242,99
	Não tem	7,19	160	36,25	105,77	223,75
Rádio	Mais de um	34,43	183,34	42,54	116,28	257,2
	Um	56,88	173,75	40,23	111,86	244,58
	Não tem	8,69	167,71	38,81	108,57	236,21
Banheiro	Mais de um	24,41	182,5	42,93	115,2	257,25
	Um	72,87	175,4	40,44	112,93	246,35
	Não tem	2,72	153,22	36,09	101,59	218,32
Número De Quartos	Quatro ou mais	13,57	174,15	41,02	111,58	246,47
	Três	31,36	179,79	41,49	115,34	252,33
	Dois	39,46	178,1	40,93	114,54	249,72
	Um	14,31	169,16	40,54	108,16	241,27
	Nenhum	1,30	156,43	38,16	102,27	225,3
Carro	Mais de um	13,14	175,81	42,55	111,23	251,17
	Um	32,56	180,91	42,72	114,34	255,34
	Nenhum	54,30	174,08	39,79	112,49	243,83
Mora com	Pai e mãe	66,23	178,31	41,55	114	251,16
	Só com a mãe	19,89	175,54	40,96	112,24	247,57
	Só com o pai	3,44	166,9	40,00	106,61	237,65
	Outros	10,44	170,28	39,09	110,26	238,83
Escolaridade da mãe	Ensino Superior	10,77	184,37	43,00	117,29	259,91
	Ensino Médio	26,64	188,80	43,65	118,88	263,66
	Ensino Fundamental	13,15	176,57	41,34	112,67	248,91
	Até 4a série do Ens. Fund. Nunca estudou ou não completou a 4a série do Ens. Fund.	13,55	173,88	39,63	112,61	243,35
		8,80	164,46	38,23	107,07	231,86
	Aluno não sabe	27,09	175,24	40,13	112,91	245,31
Escolaridade do pai	Ensino Superior	9,94	183,64	42,46	116,96	258,24
	Ensino Médio	20,48	187,00	44,18	117,14	263,34
	Ensino Fundamental	11,99	176,22	41,34	112,54	248,8
	Até 4a série do Ens. Fund. Nunca estudou ou não completou a 4a série do Ens. Fund.	11,09	174,28	40,11	112,33	244,63
		9,14	164,61	38,24	107,21	232,33
	Aluno não sabe	37,36	176,2	40,36	113,41	246,7

Fonte: Elaboração própria com base nos dados da Prova Brasil 2005.

As tabelas 5 e 6 apresentam informações referentes à estrutura das escolas cujos

alunos foram submetidos ao exame da Prova Brasil 2005. Escolas que possuem laboratório de informática representam 34,37% da amostra e seus alunos possuem um desempenho superior em cerca de 6 pontos na prova de matemática e 7 pontos na prova de português. Se a escola não possui esgoto tratado, o desempenho dos alunos fica mais de 10 pontos abaixo da média da prova de matemática, que é 183,64, e da média da prova de português, que é de 176,53.

Tabela 5 – Estatísticas descritivas da escola (matemática - 4ª série)

Variável		(%)	Média	desv. padrão	P5	P95
Possui laboratório de informática	Sim	34,37	187,77	40,23	124,64	257,70
	Não	65,63	181,48	38,79	121,55	250,30
Possui laboratório de ciências	Sim	12,08	189,93	39,73	126,82	257,90
	Não	87,92	182,78	39,28	121,99	252,30
Possui biblioteca	Sim	47,13	186,34	39,38	125,04	255,00
	Não	52,87	181,25	39,27	120,53	251,00
Possui sala de Leitura	Sim	25,14	185,83	40,64	122,13	257,00
	Não	74,86	182,91	38,96	122,61	251,70
Possui internet	Sim	62,13	184,17	40,09	121,88	254,60
	Não	37,87	182,79	38,24	123,48	250,20
Usam computador na sala de aula	Sim	28,14	188,83	40,39	125,20	258,70
	Não	71,86	181,62	38,83	121,58	250,50
Tem esgoto Tratado	Sim	99,88	183,66	39,41	122,51	253,20
	Não	0,12	171,93	36,61	113,54	237,10
Serve merenda Escolar	Sim	99,72	183,66	39,41	122,50	253,20
	Não	0,28	179,69	37,59	122,12	246,90
Possui sala de tv e video	Sim	35,81	187,08	39,86	124,70	256,50
	Não	64,19	181,73	39,02	121,40	250,90

Fonte: Elaboração própria com base nos dados da Prova Brasil 2005 e Censo Escolar 2005.

Pode-se perceber que a infra-estrutura da escola, indicada pela existência de laboratórios de informática e ciências, acesso à internet e biblioteca, entre outros, é importante, e a ausência dessa infra-estrutura indica desempenho dos alunos abaixo da média, tanto na prova de matemática quanto de português.

Tabela 6 – Estatísticas descritivas da escola (português - 4ª série)

Variável		(%)	Média	Desv. Padrão	P5	P95
Possui laboratório de informática	Sim	34,37	181,12	42,22	114,7	254,3
	Não	65,63	174,13	40,52	112,04	245,5
Possui laboratório de ciências	Sim	12,08	183,22	42,00	116,57	255,8
	Não	87,92	175,61	41,06	112,46	247,9
Possui biblioteca	Sim	47,13	178,98	41,05	114,9	250,6
	Não	52,87	174,35	41,3	111,35	247,4
Possui sala de Leitura	Sim	25,14	179,59	42,74	112,99	254
	Não	74,86	175,51	40,68	112,85	247
Possui internet	Sim	62,13	177,14	42,02	112,39	250,8
	Não	37,87	175,54	39,93	113,74	245,4
Usam computador na sala de aula	Sim	28,14	181,9	42,24	115,27	255
	Não	71,86	174,23	40,66	112,08	246,1
Tem esgoto Tratado	Sim	99,88	176,55	41,25	112,9	249
	Não	0,12	163,93	39,52	106,49	233,8
Serve merenda Escolar	Sim	99,72	176,54	41,25	112,88	249
	Não	0,28	172,73	39,36	113,19	241,1
Possui sala de tv e video	Sim	35,81	180,3	41,85	114,79	253,1
	Não	64,19	174,43	40,76	111,97	246,2

Fonte: Elaboração própria com base nos dados da Prova Brasil 2005 e Censo Escolar 2005.

Na tabela 7 é possível observar que 4.193 municípios possuem informação sobre coeficiente de mortalidade infantil enquanto 754 municípios não possuem. A média do desempenho escolar nos municípios com essa informação é 177,65 pontos em matemática e 168,66 pontos em português. Já os 754 municípios restantes possuem a média de 184,33 pontos em matemática e 173,78 pontos em português. A diferença de média é estatisticamente significativa em quase toda a tabela, as exceções são as médias de matemática e português para a variável de percentual de nascidos vivos abaixo do peso e as médias de português para a variável indicadora do número de estabelecimentos públicos de saúde no município.

Tabela 7 – Desempenho médio dos municípios, com e sem informações de variável de saúde, nas provas de matemática e português.

Variável		Observações	Média de matemática	Média de Português
Percentual de nascidos vivos com quatro ou mais consultas pré-natais	Sim	4943	178,659	169,458
	Não	4	187,472	180,89
Coeficiente de Mortalidade Infantil	Sim	4193	177,647	168,659
	Não	754	184,334	173,962
Percentual de nascidos vivos com sete ou mais consultas pré-natais	Sim	4944	178,671	169,472
	Não	3	171,369	161,757
Percentual de nascidos vivos abaixo do peso	Sim	4868	178,648*	169,717*
	Não	79	179,793*	169,463*
Estabelecimentos públicos de saúde	Sim	4935	178,658	169,4*
	Não	12	182,145	169,468*
Estabelecimentos públicos com internação	Sim	1980	173,172	172,772
	Não	2967	182,333	164,515
Estabelecimentos de emergência com pediatria	Sim	2232	181,637	173,017
	Não	2715	176,224	166,549
Estabelecimentos com apoio à terapia e diagnose	Sim	2294	175,412	165,538
	Não	2653	182,429	174,012

* Diferença de médias não significativas

Fonte: Elaboração própria com base nos dados do DATASUS, Pesquisa AMS e Prova Brasil 2005.

2.2 Variáveis adicionais utilizadas na formação do painel 2005-2007

A Prova Brasil 2007 contou com a participação de 4.109.265 alunos, sendo 2.310.302 alunos da quarta série e 1.798.963 alunos da oitava série do ensino fundamental em 5.553 municípios. A amostra final⁶ para o ano de 2007 teve 820.775 alunos da quarta série distribuídos em 4.994 municípios.

Também foram utilizados dados do Censo Escolar 2007 e informações do DATASUS em 2007. A amostra final utilizada no painel teve 4.959 municípios, agregados em 554 microrregiões. As características do Censo Escolar 2007 e das informações do DATASUS são as mesmas daquelas apresentadas anteriormente na descrição das variáveis em 2005, exceto

⁶ Há diferenças no desempenho escolar dos alunos da amostra inicial e a final. Os alunos da amostra inicial tiveram média de 195,67 em português e 212,17 em matemática enquanto que os alunos da amostra final tiveram desempenho médio de 175,43 e 193,43, respectivamente.

pela ausência das variáveis indicadoras de acesso à água encanada e energia elétrica na residência do aluno e se a escola possui sala de leitura ou possui computadores nas salas de aula.

2.3 Municípios com surto de dengue em 2005

O DATASUS também disponibiliza para acesso público, através do Sistema de Informação de Agravos de Notificação (SINAN), informações sobre doenças que devem ser obrigatoriamente notificadas pelos hospitais e postos de saúde. Dessa base de dados foi possível retirar dados, agregados por município, referentes a casos de dengue dos anos de 2001 a 2007.

Dessa forma foi possível avaliar a evolução dos casos de dengue nos municípios e identificar aqueles que mantinham certo nível de pessoas infectadas a cada ano, mas que em 2005 apresentaram um salto, aumentando em mais de 10 vezes a incidência da doença na população com relação ao nível do ano anterior. Assim, foram coletados dados para dengue na faixa etária de 10 a 14 anos, idade mais comum aos alunos da quarta série do ensino fundamental. Municípios com menos de 5 mil habitantes foram excluídos da amostra pois pequenas variações no número absoluto de casos de dengue poderiam provocar grandes variações na proporção de casos entre 2004 e 2005.

Outras características dos municípios, que servem como variáveis de controle, também foram retiradas do portal eletrônico do DATASUS. Foram obtidas informações sobre a qualidade da coleta de lixo e o percentual de alfabetizados nos municípios e também das instalações sanitárias e o abastecimento de água nas residências.

O Ministério de Desenvolvimento Social e Combate à Fome disponibiliza em seu portal eletrônico a Matriz de Informação Social, diversas informações sócio-demográficas. Desta base de dados foram retirados os gastos per capita dos municípios com assistência

social, percentual de pobres e indigentes no município, assim como o número de indivíduos atendidos pelo programa Bolsa Alimentação.

Todas as variáveis de controle mencionadas foram convertidas em taxas por mil habitantes para que fosse possível fazer comparações entre os municípios independente do tamanho da população. Assim, as variáveis referentes ao número de pessoas com coleta de lixo, acesso à rede de esgoto e rede de água canalizada foram transformadas em taxas por mil habitantes. As variáveis indicadoras de alfabetização e pobreza/indigência, que já estavam em percentual da população, e a variável indicadora de gasto per capita dos municípios com assistência social não foram convertidas pois já estavam controladas pela população.

Devido à falta de observações de casos de dengue em 2004 e/ou 2005 em alguns municípios, dos 4.687 municípios na base de dados, a amostra trabalhada contou com 1.524 municípios sendo que 78 destes tiveram surto de dengue em sua população, conforme o critério adotado. A sub amostra na faixa etária de 10 a 14 anos contou com 418 municípios e, destes, 43 municípios tiveram surtos de dengue nessa faixa etária de acordo com o critério adotado.

2.3.1 Análise Descritiva dos dados de surto de dengue em 2005

A tabela 8 mostra a distribuição dos municípios que tiveram surto de dengue em 2005 no Brasil e nas regiões do país, divididos entre municípios com surto de dengue e demais municípios. A maior parte dos dados coletados sobre dengue no SINAN são provenientes da região Nordeste. A região Sul apresenta pouca representatividade, o que pode ser explicado pelo fato de que dificilmente há proliferação de mosquitos da dengue em áreas em que a temperatura chega com frequência a menos de 17 graus Celsius.

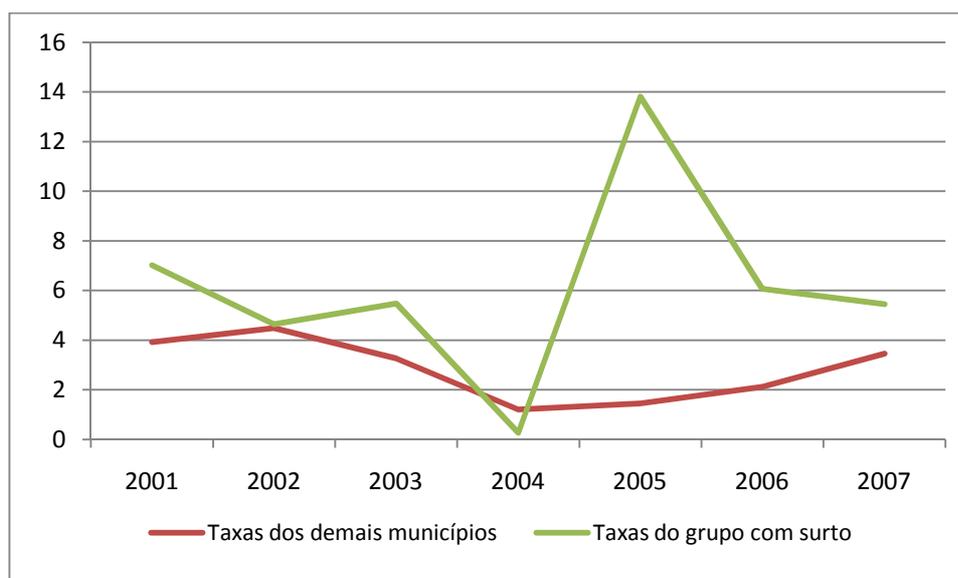
Tabela 8 – Distribuição dos municípios com e sem surtos de dengue em 2005

	População			Faixa etária 10-14 anos		
	Demais	Surto	Total	Demais	Surto	Total
Brasil	1446	78	1524	375	43	418
Norte	219	6	225	84	2	86
Nordeste	764	56	820	196	32	228
Centro-Oeste	162	7	169	42	6	48
Sudeste	270	9	279	51	3	54
Sul	30	0	30	1	0	1

Fonte: Elaboração própria com base nos dados do DATASUS em 2007

O gráfico 1 mostra a média das taxas de dengue (casos de dengue por mil habitantes) dos 1.524 municípios da amostra, separados em grupo de municípios com surto de dengue e outro com os demais municípios, com 78 e 1.446 municípios em cada grupo, respectivamente. É possível ver que, embora a média das taxas nos grupos não siga o mesmo padrão de comportamento, a diferença entre os grupos é pequena, não ultrapassando a margem de quatro pontos entre um e outro.

Gráfico 1 – Média das taxas de dengue dos municípios com surto e demais municípios



Fonte: Elaboração própria com base no DATASUS.

Em 2004, tanto o grupo de municípios com surto quanto os demais municípios

apresentaram médias parecidas, porém em 2005 o primeiro grupo apresentou um salto de mais de 11 pontos na taxa, enquanto o grupo com os demais municípios variou positivamente menos de um ponto. Isso indica a grande diferença entre os grupos nesse período e que não foi observada em outros anos.

Capítulo 3 - Metodologia

Assim como o capítulo anterior, este se divide em três partes. A primeira descreve a metodologia usada para a estimação dos modelos com dados em cross section para o ano de 2005, em que as variáveis independentes de interesse são indicadores de saúde e da qualidade da infra-estrutura de saúde das microrregiões. A segunda parte traz uma breve revisão sobre métodos de estimação com dados em painel e os diferentes tipos de estimadores utilizados neste trabalho. A terceira parte descreve a metodologia utilizada nas regressões em que a variável independente de interesse passa a ser o fato de um município ter tido ou não um surto de dengue no ano de 2005.

3.1 Indicadores de saúde e qualidade da infra-estrutura de saúde em cross-section

No primeiro método utilizado para atingir o objetivo da pesquisa foi feito o cruzamento, em uma *cross-section*, de uma série de bases de dados do ano de 2005.

Com os dados da Prova Brasil fornecendo um indicador do desempenho escolar dos alunos e características do aluno, sua família e residência, dados do Censo Escolar fornecendo características da escola e os dados do DATASUS e da pesquisa Assistência Médico Sanitária servindo como *proxy* para a variável “saúde média das crianças”, foi possível verificar qual é a relação entre a oferta de serviços e procedimentos de saúde realizados na microrregião *i* e o desempenho escolar dos alunos da 4ª série desse município. Assim estima-se, por mínimos quadrados ordinários, a seguinte equação:

$$y_i = \beta_0 + \beta_1'X_i + \beta_2'Z_i + \varepsilon_i$$

Onde y_i é a média do desempenho escolar das crianças da 4ª série na microrregião i ; X_i é um conjunto de variáveis indicando a média para cada microrregião de diversas características dos pais, do aluno, dos professores e da escola que podem influenciar a saúde e a educação. Z_i é a variável indicadora de saúde de interesse. A variável ε_i é a média do termo aleatório.

A variável de desempenho escolar foi regredida contra cada indicador de saúde individualmente. Em seguida foram adicionados grupos de controle, sucessivamente.

Como mencionado na descrição dos dados utilizados no estudo, foram criadas *dummies* para indicar características dos alunos, seu *background* familiar e características de suas escolas. Dessa forma foi possível gerar a análise descritiva dos dados. Para estimar as regressões, as variáveis *dummy* foram reunidas e passaram a indicar o percentual de indivíduos, por microrregião, com determinada característica.

3.2 Métodos de estimação em painel

Conforme mencionado anteriormente, estudos realizados com uma base de dados em *cross-section* podem apresentar problemas de estimação devido a potenciais problemas com variáveis não observáveis.

De acordo com Wooldridge (2002), a principal motivação para utilização de dados em painel é exatamente resolver o problema de viés causado por variável não observável. Se a variável não observável for constante no tempo, então ela é chamada de efeito não-observável. Seguindo Wooldridge (2002), podemos escrever um modelo básico de efeitos não-observáveis da seguinte forma:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it} \quad t = 1, 2, \dots, T$$

Onde \mathbf{x}_{it} é $1 \times K$, $\boldsymbol{\beta}$ é o vetor de parâmetros a serem estimados, c_i é o efeito não-

observável, constante no tempo, e u_{it} é o erro idiossincrático. A principal questão a ser tratada na metodologia de painel é determinar se o efeito não-observável é aleatório, ou seja, se ele é uma variável aleatória, ou então se é efeito fixo e deve ser estimado como um parâmetro a ser estimado para cada observação i .

Diversos estimadores podem ser utilizados para estimar o vetor β utilizando dados em painel. O primeiro deles utilizado neste trabalho, o estimador de mínimos quadrados ordinários agrupados (*pooled*), é consistente se o erro idiossincrático e o efeito não-observável não são correlacionados com as variáveis explicativas, ou seja, $E(\mathbf{x}'_{it}u_{it}) = \mathbf{0}$ e $E(\mathbf{x}'_{it}c_i) = \mathbf{0}$, para $t = 1, 2, \dots, T$.

O segundo estimador utilizado no trabalho, o de efeitos aleatórios, necessita de algumas suposições adicionais para que seja consistente e não viesado. Além de ortogonalidade, é necessária exogeneidade estrita entre o efeito não-observável e as variáveis explicativas: (a) $E(u_{it}|\mathbf{x}_i, c_i) = 0, t = 1, \dots, T$. E também (b) $E(c_i|\mathbf{x}_i) = 0$, onde $\mathbf{x}_i \equiv (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iT})$. Como o estimador de efeitos aleatórios é o estimador de mínimos quadrados generalizados factíveis, também é necessária a condição de rank, $\text{Rank } E(\mathbf{X}'_i\mathbf{\Omega}^{-1}\mathbf{X}_i) = K$, onde $\mathbf{\Omega}$ é a matriz de variância.

Finalmente, é necessário também: (c) $E(\mathbf{u}_i\mathbf{u}'_i|\mathbf{x}_i, c_i) = \sigma_u^2\mathbf{I}_T$ e (d) $E(c_i^2|\mathbf{x}_i) = \sigma_c^2$. Essas condições implicam que os erros idiossincráticos são serialmente não correlacionados entre si e supõem homocedasticidade do efeito não observável, c_i . O estimador de efeitos aleatórios é dado por:

$$\hat{\beta}_{EA} = \left(\sum_{i=1}^N \mathbf{X}'_i \hat{\mathbf{\Omega}}^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^N \mathbf{X}'_i \hat{\mathbf{\Omega}}^{-1} \mathbf{y}_i \right)$$

O terceiro estimador utilizado neste trabalho para dados em painel, o estimador de efeitos fixos, requer menos suposições do que os anteriores. É necessário apenas que seja respeitada a exogeneidade estrita da variável explicativa condicionada no efeito não

observável: $E(u_{it}|x_i, c_i) = 0, t = 1, 2, \dots, T$. O estimador de efeitos fixos pode estimar consistentemente efeitos parciais na presença de variáveis não observáveis constantes no tempo, sendo mais robusto do que a análise por efeitos aleatórios. No entanto, há um porém: não é possível fazer a distinção entre efeito não observável constante no tempo e variáveis explicativas constantes no tempo. Isso se deve à transformação *within*, que faz a primeira diferença da equação $y_{it} = x_{it}\beta + c_i + u_{it}, t = 1, \dots, T$.

Seja $\bar{x}_i = T^{-1} \sum_{t=1}^T x_{it}$ e $\bar{y}_i = T^{-1} \sum_{t=1}^T y_{it}$, o estimador de efeitos fixos, ou estimador *within*, é dado por:

$$\hat{\beta}_{EF} = \left(\sum_{i=1}^N \sum_{t=1}^T (x_{it} - \bar{x}_i)' (x_{it} - \bar{x}_i) \right)^{-1} \left(\sum_{i=1}^N \sum_{t=1}^T (x_{it} - \bar{x}_i)' (y_{it} - \bar{y}_i) \right)$$

Como é possível ver na fórmula acima, para identificar o modelo, o estimador de efeitos fixos usa apenas a variação dentro de cada unidade de observação ao longo do tempo.

3.3 Matching para municípios com surto de dengue em 2005

Na segunda abordagem desta pesquisa foi utilizado o método de *matching* para avaliar qual é o efeito que um surto de dengue em um município tem sobre o desempenho escolar médio daquele município. A ideia por trás desse método é parear observações que não recebem tratamento (grupo controle) com observações que recebem tratamento (grupo tratado). No presente trabalho o tratamento é o surto de dengue. Municípios que tiveram surto de dengue em 2005 são comparados a um grupo de municípios com características semelhantes, exceto pelo fato de que o primeiro grupo teve surto de dengue, mas o segundo grupo de municípios não teve.

Uma vez que seria difícil encontrar municípios exatamente com as mesmas características, i.e. ter valores muito próximos para cada variável de controle, é feita uma aproximação disso através da estimação da probabilidade de que um município tenha surto de

dengue, dada as variáveis de controle. O *propensity score matching* seria, então, a probabilidade de que um município tenha tido surto de dengue em 2005. Em vez de procurar municípios com valores semelhantes para diversas variáveis, passa-se a procurar municípios com probabilidades semelhantes de terem apresentado surto de dengue em 2005, o que é um processo mais simples.

Para que a comparação seja possível devemos encontrar municípios com probabilidades semelhantes. E por isso, na segunda etapa, é feito o *matching*, que é o processo de parear os indivíduos no grupo de tratamento e controle cuja probabilidade de receber tratamento seja parecida. Dessa forma, quando compararmos os dois grupos (municípios com e sem surto de dengue) com base em características observáveis semelhantes, poderemos atribuir as diferenças de desempenho escolar que surgirem entre os dois grupos apenas ao efeito do tratamento (ter tido surto de dengue em 2005)

Feita a estimação do *propensity score*, que é feito pela função *logit*, o próximo passo é escolher o algoritmo do *matching*. De acordo com Lee (2005), o efeito do tratamento sobre os tratados (ATT) é a média ponderada das diferenças entre o produto médio do grupo dos tratados e dos não tratados, e é conhecido como o algoritmo mais simples para estimar o efeito do tratamento médio sobre os tratados.

$$\text{É dado por } \Delta^M = \frac{1}{N_t} \sum_{i \in \{D=1\}} \left[y_{1i} - \sum_j w_{ij} y_{0j} \right] \quad 0 < w_{ij} \leq 1 \text{ e } \{D=1\} \text{ tratados}$$

Estimadores de *matching* distintos são criados pela variação de w_{ij} . Ou seja, pela definição dos vizinhos, pelo modo como se lida com o suporte comum e pelos pesos dados aos vizinhos. Sendo assim, diferentes tipos de algoritmos foram utilizados, conforme os dados apresentados.

Um dos algoritmos utilizados foi o *matching* por estratificação, o qual consiste em dividir o *propensity score* em intervalos nos quais as unidades tratadas e não tratadas

possuam, na média, o mesmo *propensity score*. Conforme dito anteriormente, ATT é a média ponderada das diferenças entre o produto médio do grupo dos tratados e dos não tratados.

Chamemos de b os blocos definidos sobre os intervalos do *propensity score*. Então, o efeito

do tratamento no b -ésimo bloco é: $ATT_b^s = \frac{1}{N_b^T} \sum_{i \in I(b)} y_{1i} - \frac{1}{N_b^c} \sum_{j \in I(b)} y_{0j}$. Com isso, o efeito do

$$\text{tratamento é: } ATT^s = \sum_{b=1}^B ATT_b^s \left[\frac{\sum_{i \in I(b)} D_i}{\sum D_i} \right].$$

Também foi utilizado o algoritmo de *kernel*, dado por $w_{ij} = \frac{K\left(\frac{x_j - x_i}{h}\right)}{\sum_j^{N_{ci}} K\left(\frac{x_j - x_i}{h}\right)}$, onde k

é um *kernel*. Este é um estimador não paramétrico que usa médias ponderadas de todos os indivíduos no grupo controle para construir o produto contra-factual. Para este algoritmo, é necessário escolher a função *kernel* e as bandas de variações. Mas as bandas de variações possuem um *trade-off*. Quanto maior a banda menor a variação e maior será o viés. Então a escolha da banda é um compromisso entre variância pequena e estimativas não viesadas das verdadeiras funções de densidade.

O terceiro e último algoritmo usado foi o chamado *matching* de vizinhos próximos, o qual é definido como $A_i(x) = \{j \mid \min_j \|x_i - x_j\|\}$, com um e dois vizinhos, ambos com e sem reposição. Nesse algoritmo, um município que participou do tratamento é comparado apenas com um ou dois municípios que possuem *propensity score* mais próximo. Em conjunto com esse algoritmo é possível utilizar uma janela limitadora de “distância” entre os *propensity scores*. Caso a distância seja maior que o limite, o vizinho não é válido para comparação.

Capítulo 4 – Resultados

Este capítulo divide-se em três partes. Na primeira parte estão os resultados das

regressões em *cross-section* para o ano de 2005. Na segunda parte estão os resultados para regressões em painel com observações do ano 2005 e 2007. Finalmente, na terceira parte, encontram-se os resultados das regressões de *matching* para surtos de casos de dengue em 2005.

4.1 Resultados das regressões em cross-section

As tabelas 9 e 10 apresentam o impacto das variáveis indicadoras de saúde no desempenho escolar dos alunos na prova de matemática e português em microrregiões. Na primeira coluna, a variável indicadora de saúde foi a única variável explicativa na regressão. Na segunda coluna adicionam-se as variáveis de controle I, que são as variáveis referentes ao gênero do aluno, sua cor e se já foi reprovado uma ou mais de uma vez. Estas informações referentes ao aluno foram apresentadas nas tabelas 1 e 2. Também foram adicionados controles referentes ao percentual de pobreza e indigência na microrregião.

Em seguida, na terceira coluna, são adicionadas mais variáveis referentes ao *background* familiar do aluno, como as características físicas de sua residência e informações sobre a escolaridade da mãe e do pai, que foram apresentadas nas tabelas 3 e 4. Estas variáveis, em conjunto com o grupo de variáveis de controle I, formam o grupo de variáveis de controle II. Finalmente, na quarta coluna de coeficientes, adicionaram-se as características da escola onde o aluno estudava no ano em que prestou o exame da Prova Brasil.

Tabela 9 – Impacto dos indicadores de saúde no desempenho escolar (matemática - 4ª série, continua)

Variável	Sem controle	Controle I	Controle II	Controle III
Percentual de nascidos vivos com quatro ou mais consultas pré-natais	0,687***	-0,008	-0,042	-0,045
Percentual de nascidos vivos com sete ou mais consultas pré-natais	0,497***	0,046**	0,013	0,004
Percentual de nascidos vivos	4,527***	0,993***	-0,147	-0,035

abaixo do peso				
Coefficiente de	-0,941***	0,107**	-0,036	-0,033
Mortalidade infantil				
Estabelecimentos públicos	-14,863***	-0,698	-2,131*	-2,073*
de saúde				
Estabelecimentos públicos	17,321	11,518**	0,077	-1,219
com internação				
Estabelecimentos de	23,652	11,633	8,666	0,038
emergência com pediatria				
Estabelecimentos com apoio	87,008***	13,562***	-1,910	-4,382
à terapia e diagnose				

Obs.: * Significativo a 10% ** Significativo a 5% *** Significativo a 1%

Fonte: Elaboração própria com base nos dados da Prova Brasil 2005, Censo Escolar 2005, DATASUS e Pesquisa Assistência Médico Sanitária 2005.

Quanto maior o número de mães que fazem quatro ou mais consultas pré-natais, maior o desempenho escolar. Isso se traduz no indicador “percentual de nascidos vivos com quatro ou mais consultas pré-natais” e “percentual de nascidos vivos com sete ou mais consultas pré-natais”. Esse índice é calculado com o número de nascidos vivos que realizaram quatro ou mais consultas pré-natais e sete ou mais consultas pré-natais, respectivamente, dividido pelo número total de nascidos vivos no ano de 2005.

Sem controle, os coeficientes apresentam impacto positivo no desempenho escolar quando entram na regressão sem controle. Para cada ponto percentual a mais de nascidos vivos de mães com quatro ou mais consultas pré-natais, o desempenho na prova de matemática aumenta em média 0,69 pontos. Para sete ou mais consultas pré-natais, o aumento é de quase 0,5 pontos. Quando é adicionado o primeiro grupo de variáveis de controle, que são as características do aluno, apenas o coeficiente da variável indicadora do percentual de nascidos vivos com sete ou mais consultas pré-natais permanece significativa, porém com impacto reduzido a 0,046 pontos. Adicionando o segundo grupo de controle, todos os coeficientes passam a ser não significantes. O mesmo padrão é observado nas provas de matemática e de português.

Tabela 10 – Impacto dos indicadores de saúde no desempenho escolar (português - 4ª série)

Variável	Sem controle	Controle I	Controle II	Controle III
Proporção de nascidos vivos com quatro ou mais consultas pré-natais	0,679***	-0,023	0,036	-0,034
Proporção de nascidos vivos com sete ou mais consultas pré-natais	0,501***	0,039*	0,003	-0,006
Percentual de nascidos vivos abaixo do peso	4,673***	0,952***	-0,200	-0,092
Coefficiente de Mortalidade infantil	-1,002***	0,048	-0,037	-0,034
Estabelecimentos públicos de saúde p/ mil hab.	-16,651***	-1,890	-3,266***	-3,640***
Estabelecimentos públicos com internação p/ mil hab.	7,725***	5,749	-1,868	-2,505
Estabelecimentos de emergência com pediatria p/ mil habitantes	5,366	-0,955	2,253	-6,217
Estabelecimentos com apoio à terapia e diagnose p/ mil habitantes	79,598***	4,580	-5,928	-8,670**

Obs.: * Significativo a 10% ** Significativo a 5% *** Significativo a 1%

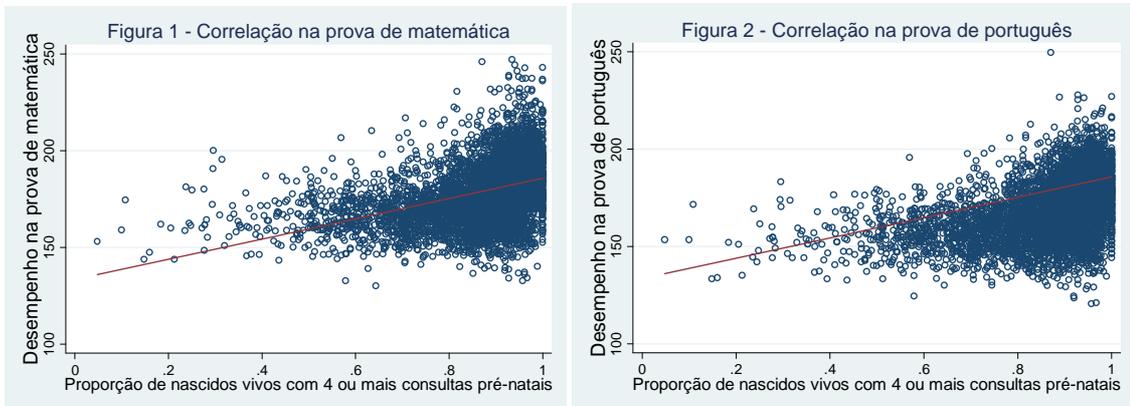
Fonte: Elaboração própria com base nos dados da Prova Brasil 2005, Censo Escolar 2005. DATASUS e Pesquisa Assistência Médico-Sanitária 2005.

Alguns coeficientes, no entanto, não apresentaram o impacto esperado, inclusive até mesmo o sentido do impacto não era esperado. De acordo com as regressões, a proporção de nascidos vivos abaixo do peso tem impacto positivo e significante no desempenho dos alunos, e, além disso, o número de estabelecimentos públicos de saúde por mil habitantes tem impacto negativo.

Os demais coeficientes apresentados na tabela servem como visão geral para os resultados de outros indicadores. Mais doze indicadores de saúde foram testados, mas os padrões apresentados foram pouco animadores. A maior parte é não significativa, alguns deles até mesmo sem a adição de controle.

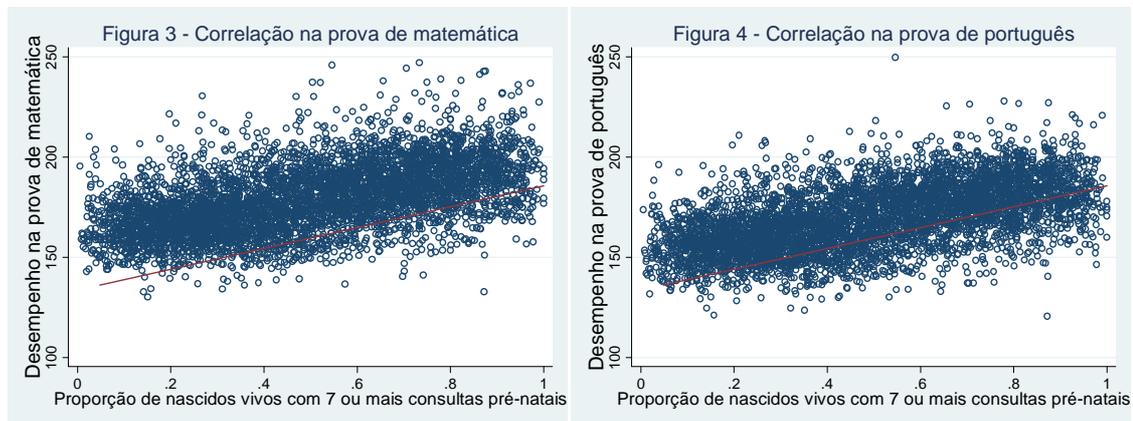
As figuras 1 e 2 traçam a correlação entre o desempenho escolar e a proporção de nascidos vivos com quatro ou mais consultas pré-natais. Na prova de matemática a correlação

foi de 0,403 e na prova de português foi de 0,417. A linha traçada indica a melhor predição linear para os dados.



Fonte: Elaboração própria com base nos dados do DATASUS e Prova Brasil 2005

As figuras 3 e 4 traçam a correlação entre o desempenho escolar e a proporção de nascidos vivos com sete ou mais consultas pré-natais. Pode-se perceber que as observações estão mais dispersas ao longo do eixo x, contendo o valor da variável indicadora de saúde. Com isso, a reta de predição linear se ajusta melhor aos dados. A correlação na prova de matemática foi de 0,544 e na prova de português foi de 0,562



Fonte: Elaboração própria com base nos dados do DATASUS e Prova Brasil 2005

4.2 Resultados das regressões em painel

Como a literatura sobre métodos de estimação em painel costuma dar ênfase à escolha de estimadores entre o de efeito fixo e de efeito aleatório, as tabelas a seguir apresentam

somente estes dois⁷. A tabela 9 traz os resultados das regressões com estimadores de efeito fixo em painel com observações de microrregiões no ano de 2005 e 2007 para a prova de matemática.

Na primeira coluna de estimadores, a variável indicadora de saúde foi regredida sem nenhum controle. Na segunda, foram utilizados controles referentes à média das características dos alunos na microrregião. Na terceira adicionaram-se controles de *background* familiar e na quarta, adicionaram-se controles das escolas.

Testes de Hausman foram feitos para avaliar qual estimador seria o mais apropriado, mas os resultados foram inconclusivos. O estimador de efeitos fixos ou o estimador de efeitos aleatórios eram indicados como mais apropriados dependendo das especificações dos modelos. Pelo fato de existirem somente duas observações no tempo, a confiabilidade da eficácia do teste de Hausman fica prejudicada.

Tabela 11 – Impacto dos indicadores de saúde no desempenho escolar, estimadores de efeitos fixos (matemática - 4ª série)

Variável	Efeitos Fixos			
	I	II	III	IV
Percentual de nascidos vivos cujas mães tiveram 7 ou mais consultas pré-natais	0,281***	0,236***	0,118***	0,090***
Número de óbitos infantis por 10 mil habitantes	-7,080***	-3,088***	0,695*	0,766*
Coefficiente de Mortalidade Infantil	0,867***	0,686***	0,128***	0,078***
Número de óbitos neonatais por 10 mil habitantes	1,149***	0,584***	0,101**	0,087**
Coefficiente de Mortalidade Neonatal tardia	0,066**	0,028*	-0,006**	-0,006***

Obs.: *Significativo a 10%, **Significativo a 5%, ***Significativo a 1%

Fonte: Elaboração própria com base nos dados da Prova Brasil, Censo Escolar e Datasus, anos 2005 e 2007.

Analisando as tabelas 11 e 12, que apresentam os resultados dos estimadores de efeitos fixos e estimadores aleatórios, respectivamente, é possível ver que o coeficiente da variável que indica o percentual de nascidos vivos cujas mães tiveram sete ou mais consultas pré-natais foi significativo a 1% em todas as regressões e os coeficientes indicaram impacto semelhante, independente do tipo de estimador escolhido. Nos modelos mais completos, com

⁷ As tabelas com estimador de mínimos quadrados ordinários agrupados (pooled ols) encontram-se no apêndice.

todas as variáveis de controle eram incluídas, esse impacto era de 0,09 pontos no modelo de efeitos fixos e 0,07 no de efeitos aleatórios. Ou seja, um aumento de 1 p.p. no indicador de saúde está associado a um aumento de menos de um décimo de ponto na escala da Prova Brasil, m efeito muito pequeno.

Apesar de o efeito ser pequeno, isso indica que um cuidado maior das gestantes, ou dos médicos com as gestantes, que vai além do número de consultas pré-natais, têm impacto positivo no desempenho escolar.

Tabela 12 – Impacto dos indicadores de saúde no desempenho escolar, estimadores de efeitos aleatórios (matemática - 4ª série)

Variável	Efeitos Aleatórios			
	I	II	III	IV
Percentual de nascidos vivos cujas mães tiveram 7 ou mais consultas pré-natais	0,295***	0,232***	0,087***	0,070***
Número de óbitos infantis por 10 mil habitantes	-7,665***	-2,665***	-0,482**	-0,365
Coefficiente de Mortalidade Infantil	0,828***	0,604***	0,109***	0,072**
Número de óbitos neonatais por 10 mil habitantes	0,918***	0,450***	0,092**	0,061
Coefficiente de Mortalidade Neonatal tardia	0,052**	0,022	-0,008***	-0,008***

Obs.: *Significativo a 10%, **Significativo a 5%, ***Significativo a 1%

Fonte: Elaboração própria com base nos dados da Prova Brasil, Censo Escolar e Datasus, anos 2005 e 2007.

Há um impacto significativo também do número de óbitos infantis por 10 mil habitantes, embora o coeficiente seja mais sensível de acordo com as variáveis escolhidas. Os coeficientes estimados por efeitos fixos acabam mudando de sinal quando são adicionados mais controles. Além disso, o restante das variáveis tem impacto no sentido contrário ao esperado como, por exemplo, o coeficiente de mortalidade infantil e o número de óbitos neonatais tardios por 10 mil habitantes, que seria de se esperar terem impacto negativo.

As tabelas 13 e 14 mostram os resultados das regressões para a prova de português. Novamente é possível perceber que a variável indicadora do percentual de nascidos vivos cujas mães tiveram sete ou mais consultas pré-natais tem coeficiente positivo e significativo, porém não em todas as regressões.

Tabela 13 – Impacto dos indicadores de saúde no desempenho escolar, estimadores de efeitos fixos (português - 4ª série)

Variável	Efeitos Fixos			
	I	II	III	IV
Percentual de nascidos vivos cujas mães tiveram 7 ou mais consultas pré-natais	0,125***	0,085***	0,052***	0,030
Número de óbitos infantis por 10 mil habitantes	-3,180***	-1,122***	0,390	0,480
Coefficiente de Mortalidade Infantil	0,391***	0,252***	0,059*	0,033
Número de óbitos neonatais por 10 mil habitantes	0,528***	0,214***	0,051	0,376
Coefficiente de Mortalidade Neonatal tardia	0,025	0,005	-0,005***	-0,005**

Obs.: *Significativo a 10%, ** Significativo a 5%, *** Significativo a 1%

Fonte: Elaboração própria com base nos dados da Prova Brasil, Censo Escolar e Datasus, anos 2005 e 2007.

Tabela 14 – Impacto dos indicadores de saúde no desempenho escolar, estimadores de efeitos aleatórios (português - 4ª série)

Variável	Efeitos Aleatórios			
	I	II	III	IV
Percentual de nascidos vivos cujas mães tiveram 7 ou mais consultas pré-natais	0,138***	0,081***	0,019***	0,009
Número de óbitos infantis por 10 mil habitantes	-5,328***	-1,800***	-0,416*	-0,319
Coefficiente de Mortalidade Infantil	0,373***	0,193***	0,055	0,048
Número de óbitos neonatais por 10 mil habitantes	0,457***	0,136***	0,029	0,009
Coefficiente de Mortalidade Neonatal tardia	0,025	0,006	-0,007***	-0,006***

Obs.: *Significativo a 10%, ** Significativo a 5%, *** Significativo a 1%

Fonte: Elaboração própria com base nos dados da Prova Brasil, Censo Escolar e Datasus, anos 2005 e 2007.

A quarta coluna das tabelas 13 e 14, que contêm todas as variáveis de controle utilizadas no trabalho, não apresentam coeficientes significantes com exceção do coeficiente de mortalidade neonatal tardia, porém o efeito é pequeno. Para cada ponto a mais no coeficiente, o desempenho médio diminui em 0,005 pontos e 0,006 pontos de acordo com os estimadores de efeitos fixos e aleatórios, respectivamente.

Ou seja, o resultado das regressões em painel apresenta-se altamente sensível a escolha da *proxy* de saúde, sendo que em todos os casos, o efeito em termos de nota é extremamente baixo, mesmo nos casos em que foi estatisticamente significativo.

4.3 Resultados das regressões de matching

Foram testadas diferentes especificações para a realização do *propensity score*

matching. Para cada especificação, foram mescladas diferentes combinações das variáveis de controle: percentual de pobres no município, latitude do município, estabelecimentos públicos de saúde por mil habitantes, taxa de internações por doença diarreica aguda em menores de 5 anos, a maneira como o lixo é coletado nos municípios, por exemplo, se ele é jogado em terreno baldio, se é incinerado ou se é coletado por serviço municipal. Também foram incluídas variáveis que descreviam como era o abastecimento de água nas residências e como era o tratamento de dejetos sólidos na propriedade. A variável independente de interesse também mudou de acordo com a especificação, foram testados surtos de dengue na população e também surtos de dengue somente na faixa etária de 10 a 14 anos.

A primeira coluna na tabela 15, NN(1), que traz os resultados do impacto de um surto de dengue na prova de matemática, foi estimada com o algoritmo de *nearest neighbour*, com um vizinho, sem reposição e sem *caliper*. Nenhum dos resultados foi significativo, inclusive quando testado com combinações entre um e dois vizinhos, com e sem reposição e com *calipers* entre 0,00001 a 0,01. A segunda e terceira coluna da tabela foram estimadas com algoritmo de *kernel* gaussiano.

Tabela 15 – Efeito do surto de dengue em 2005 na Prova Brasil – matemática.

	População			Faixa etária 10-14 anos		
	NN(1)	Kernel 0,01	Kernel 0,001	NN(1)	Kernel 0,01	Kernel 0,001
Especificação 1	-2,180	-10,510***	-4,492***	0,530	-8,777***	-3,683
Especificação 2	-1,947	-9,999***	-3,329*	0,561	-7,996***	1,265***
Especificação 3	-1,502	-10,214***	-3,705**	4,721	-8,236***	-2,432***
Especificação 4	-1,111	-10,219***	-3,856***	-1,634	-7,986***	-3,044***

Fonte: Elaboração própria com base nos dados da Prova Brasil, DATASUS e Matriz de Informação Social.

Obs.: Controles para regressão do probit (todas especificações contam com latitude, dummies para região e acesso a esgoto tratado, por mil habitantes):

Especificação 1: número de pessoas atendidas por coleta de lixo, por mil habitantes, gasto per capita do município em assistência social e taxa de internação por doença diarreica aguda em menores de 5 anos.

Especificação 2: número de pessoas atendidas por coleta de lixo, taxa de internação por doença diarreica aguda em menores de 5 anos e número de pessoas, por mil habitantes, que recebiam Bolsa Alimentação.

Especificação 3: número de pessoas, por mil habitantes, que jogam lixo em lagos, rios, lagoas ou no mar.

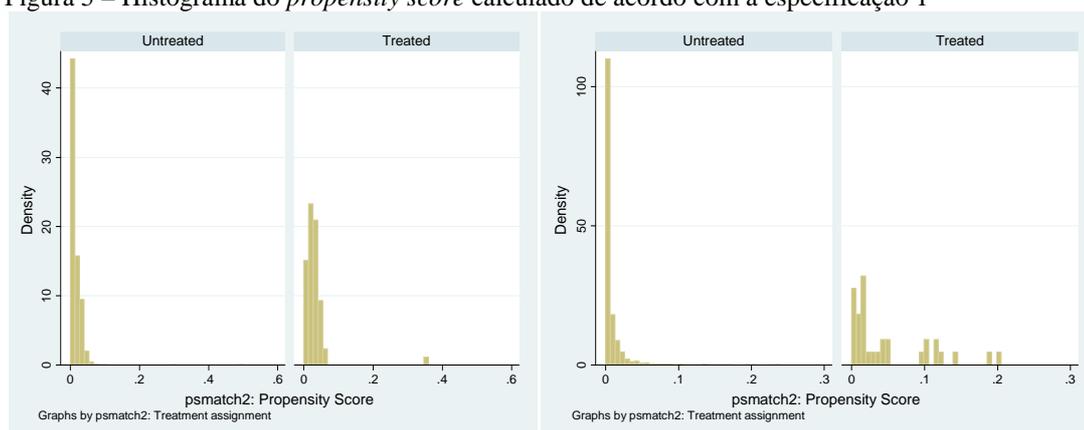
Especificação 4: número de pessoas com acesso à rede de água canalizada, por mil habitantes.

Obs.: * Significativo a 10%; ** Significativo a 5%; *** Significativo a 1%

Com os algoritmos de *kernel*, obteve-se resultados significantes em quase todas as especificações, com magnitude variando de 3 a 10 pontos negativos. Ou seja, a incidência de um surto de dengue estaria associada a uma diminuição dessa magnitude no desempenho dos alunos em matemática.

É necessário dizer que a qualidade do suporte comum do *propensity score* não foi tão boa quanto se esperava. A distribuição do suporte comum em todas as especificações ficou limitada entre 0,0 a 0,4, com concentração de observações na faixa 0,001 a 0,01, o que significa dizer que na maior parte da amostra, apenas municípios com probabilidade baixa de terem surto foram pareados com sucesso. A figura 5 apresenta o histograma do *propensity score* calculado de acordo com as variáveis da especificação 1, discriminadas na observação da tabela 15. No lado esquerdo da figura encontra-se o histograma levando em consideração surtos de dengue na população em geral e no lado direito da figura encontra-se o histograma com surtos de dengue na faixa etária 10-14 anos.

Figura 5 – Histograma do *propensity score* calculado de acordo com a especificação 1



Fonte: Elaboração própria com base nos dados da Prova Brasil, DATASUS e Matriz de Informação Social

Após o cálculo do *propensity score* e do pareamento, foram realizados testes para verificar se as médias das variáveis explicativas nos grupos de municípios com surto e no grupo dos demais municípios eram semelhantes antes e depois do pareamento. Esperava-se

que antes do pareamento as médias das variáveis fossem diferentes entre os grupos e fossem semelhantes após o pareamento. Os testes indicaram que essa hipótese se confirmou, embora em alguns casos antes mesmo do pareamento as médias das variáveis já eram semelhantes entre os grupos. A tabela 16 apresenta, como exemplo, o teste das variáveis na especificação 1 e a utilização do algoritmo de *kernel* com janela 0,001 para pareamento das observações após o cálculo do *propensity score* para surtos de dengue na população. As médias dos grupos de municípios que tiveram surto em 2005 são relatadas na tabela sob a coluna “tratamento”. As médias dos demais municípios estão sob a coluna “controle”.

Tabela 16 – Teste de média após o cálculo de *propensity score* com variáveis da especificação 1 e pareamento com algoritmo de *kernel*, janela de 0,001.

Variável	Amostra	Média		Teste-t p> t
		Tratamento	Controle	
Latitude	não pareada	-10,138	-16,481	0,000
	pareada	-10,138	-12,366	0,051
Taxa de coleta de lixo	não pareada	409,12	491,920	0,026
	pareada	409,12	436,000	0,591
Taxa de acesso a esgoto	não pareada	192,11	236,490	0,235
	pareada	192,11	202,370	0,833
Taxa de internação por doença diarreica em menores de 5 anos	não pareada	31,572	22,613	0,001
	pareada	31,572	27,600	0,335
Número de estabelecimentos públicos de saúde (por mil habitantes)	não pareada	0,677	0,468	0,000
	pareada	0,677	0,684	0,978
Gasto per capita com assistência social	não pareada	2751,0	3096,9	0,408
	pareada	2751,0	2938,5	0,704
Dummy para região Norte	não pareada	0,097	0,093	0,702
	pareada	0,097	0,087	0,851
Dummy para região Nordeste	não pareada	0,661	0,313	0,0
	pareada	0,661	0,548	0,202
Dummy para região Sudeste	não pareada	0,145	0,310	0,005
	pareada	0,145	0,205	0,386
Dummy para região Centro-Oeste	não pareada	0,097	0,091	0,867
	pareada	0,097	0,096	0,992

Fonte: Elaboração própria com base nos dados da Prova Brasil, DATASUS e Matriz de Informação Social.

Com exceção da variável latitude, todas as variáveis apresentaram diferença de média não significativa após o pareamento. Outros testes foram conduzidos em todas as

especificações de variáveis, com surtos na população e na faixa etária 10 a 14 anos e com os algoritmos apresentados na tabela 15 e os resultados são semelhantes. As variáveis que apresentavam diferenças de médias entre os grupos de municípios com surto e os demais, após o pareamento não apresentavam diferenças significativas. Isso significa dizer que, antes do pareamento, os grupos se diferenciavam nas variáveis e após o pareamento os grupos tornaram-se mais parecidos, exceto pelo fato de um grupo ter tido surto de dengue e o outro não.

A tabela 17 traz os resultados para o efeito dos surtos de dengue na prova de português. O comportamento dos estimadores é semelhante ao apresentado na tabela anterior com regressões para matemática. Os coeficientes estimados para *nearest neighbours* não foram significativos. Os coeficientes de *kernel* com janela 0,01 são ligeiramente menores em comparação e quando a janela é diminuída para 0,001 quase todos eles perdem a significância.

Tabela 17 - Efeito do surto de dengue em 2005 na Prova Brasil 2005 – português

	População			Faixa etária 10-14 anos		
	NN(1)	Kernel 0,01	Kernel 0,001	NN(1)	Kernel 0,01	Kernel 0,001
Especificação 1	-2,920	-9,557**	-3,826*	4,163	-6,165**	-1,760
Especificação 2	-2,056	-9,233**	-2,718	2,783	-5,511**	-1,209
Especificação 3	-0,991	-9,209**	-2,872	-5,060	-5,516**	-0,263
Especificação 4	-0,566	-9,212**	-3,031	-0,919	-5,305**	-0,213

Fonte: Elaboração própria com base nos dados da Prova Brasil 2005 e DATASUS.

* Significativo a 10% ** Significativo a 1%

As especificações apresentadas aqui são apenas uma parte das especificações testadas para este trabalho. Todas as outras apresentaram resultados semelhantes, sendo sensíveis principalmente aos algoritmos utilizados. Destaca-se o fato de que, independente das variáveis utilizadas, os algoritmos de *kernel* apresentaram resultado consistente com janela de 0,01 em todas as especificações testadas.

Capítulo 5 - Considerações Finais

Este trabalho teve como objetivo identificar qual o efeito da saúde no desempenho escolar em português e matemática de crianças da quarta série do ensino fundamental (atual quinto ano). Como não existe base de dados contendo variáveis indicadoras de saúde e desempenho escolar simultaneamente para indivíduos, utilizou-se como estratégia a agregação de informações dos municípios e microrregiões. Dessa forma, foi possível avaliar se aspectos de saúde de um município ou microrregião tem impacto no desempenho escolar médio dos alunos. A unidade de observação passou a ser, então, o município ou microrregião e não mais o indivíduo.

Nesse sentido, três abordagens diferentes foram testadas. A primeira abordagem foi a regressão do desempenho escolar contra variáveis indicadoras de qualidade de saúde e infraestrutura de saúde em 554 microrregiões. Todas as informações são referentes ao ano 2005. Os dados de desempenho escolar são provenientes da Prova Brasil 2005 e as variáveis de saúde são provenientes do DATASUS e da pesquisa Assistência Médico-Sanitária. O Censo Escolar 2005 e o questionário da Prova Brasil forneceram variáveis de controle para características dos alunos, de suas famílias e escolas.

As regressões em *cross-section* não indicaram resultados consistentes com resultados encontrados na literatura. A maior parte dos coeficientes não apresentou efeito e mesmo aqueles que apresentaram eram sensíveis às variáveis de controle utilizadas na regressão.

A segunda abordagem contou com informações de um painel com observações dos anos 2005 e 2007 para estimar o efeito da saúde no desempenho escolar. Os dados novamente são provenientes da Prova Brasil, Censo Escolar e DATASUS. Não existem dados para 2007 e, portanto, as variáveis de saúde utilizadas limitavam-se à indicadores de qualidade da saúde nas microrregiões. Nessa abordagem, foram encontrados resultados positivos para o percentual de nascidos vivos cujas mães tiveram sete ou mais consultas pré-natais. Na prova

de matemática, mesmo adicionando diversas variáveis de controle, para cada ponto percentual a mais da variável havia um acréscimo no desempenho escolar de 0,09 pontos de acordo com o estimador de efeitos fixos e 0,07 pontos pelo estimador de efeitos aleatórios. Na prova de português, os efeitos foram modestos, 0,03 e 0,009 respectivamente.

A terceira abordagem foi a utilização da incidência de dengue nos municípios, em 2005, para descobrir se havia algum impacto no desempenho escolar. As variáveis de saúde são do DATASUS, as de desempenho escolar da Prova Brasil e as variáveis de controle são do DATASUS e da Matriz de Informação Social do Ministério do Desenvolvimento Social e Combate à Fome. Os resultados foram sensíveis às especificações das regressões, mas de maneira geral os algoritmos de *kernel* utilizados no *matching* apresentaram resultados consistentes.

Quando a população analisada foi o total de habitantes nos municípios, aqueles que tiveram surto de dengue em 2005 apresentaram desempenho médio na prova de matemática inferior com relação aos municípios que não tiveram surto de dengue. Dependendo da janela do *kernel*, essa diferença foi de 10,5 a 3 pontos aproximadamente. Quando a população de análise se limitou a crianças de 10 a 14 anos, a diferença foi de 8,7 a 3 pontos. No caso de português, as diferenças de desempenho permaneceram, sendo de 9,5 a 2,7 pontos na população total e de 5,3 a 0,2 pontos, respectivamente, embora os coeficientes não foram significantes em todas as regressões.

A dificuldade em se obter resultados positivos robustos às especificações e métodos utilizados aponta a fragilidade das bases de dados utilizadas. Embora no presente trabalho existam alguns indícios de que medidas de saúde agregadas têm algum efeito positivo no desempenho escolar dos alunos, essa hipótese não pôde ser confirmada. Reforça-se assim a necessidade de continuar explorando o assunto em busca de novos métodos para testar essa hipótese assim como a necessidade de contar com bases de dados mais adequadas para

análise, que é facilmente obtida em outros países.

Referências Bibliográficas

- ALBERNAZ, A.; FERREIRA, F. H. G.; FRANCO, C. Qualidade e equidade na Educação Fundamental Brasileiro. **PPE**, v. 33 No.3. 2002.
- ALDERMAN, H.; BEHRMAN, J.; LAVY, V.; MENON, R. “Child health and school enrollment”. **Journal of Human Resources**, 36 (1), 185–205, 2001.
- ARENDET, J. N. Does education cause better health? A panel data analysis using school reforms for identification. **Economics of Education Review**, vol. 24(2), pg. 149-160, Abril 2005.
- BARROS, R. P. de, MENDONÇA, R., SANTOS, D. **Determinantes do desempenho Educacional no Brasil**. IPEA, mimeo, 1999.
- BECKER, G. S. **Human Capital**. New York: Columbia University Press (for NBER), 1964.
- BECKER, G. S. A Theory of the Allocation of Time. **The Economic Journal**, vol. 75, n. 299, p. 493-517, 1965.
- BEHRMAN, J.R.; LAVY, V. **Child health and schooling achievement: association, causality and household allocations**. CARESS Working Papers 97-23. University of Pennsylvania, 1997.
- BIFULCO, R.; LADD, H. F. School choice, racial segregation and test-score gaps: evidence from North Carolina’s charter school program. **Journal of Policy Analysis and Management**, v. 26 Issue 1, p. 31-56, Boston, 2006.
- BOBONIS, G.; MIGUEL, E.; SHARMA, C. P. “Iron deficiency, anemia and school participation”. **Journal of Human Resources**, 41 (4), 692–721, 2006.
- COLEMAN, J. S. et al. **Equality of educational Opportunity**. U.S. Government printing office. Washington, DC, 1966.
- DEOLALIKAR, A. Nutrition and labor productivity in agriculture: Estimates for rural south India, **Review of Economics and Statistics** 70, 406-413, 1988.
- DING, W. *et alli*. The impact of poor health on academic performance: New evidence using genetic markers. **Journal of Health Economics** 28, 578–597, 2009.
- FELÍCIO, F.; FERNANDES, R. O efeito da qualidade da escola sobre o desempenho escolar: uma avaliação do ensino fundamental no Estado de São Paulo. *Anais do XXXIII Encontro Nacional de Economia*, Natal-RN, Brasil. ANPEC, 2005.
- FUCHS, V. R. “**Some Economic Aspects of Mortality in the United States.**” Mimeo. New York: NBER, 1966.
- GLEWWE, P. Why does mother’s schooling raise child health in developing countries?

Evidence from Morocco. **The Journal of Human Resources**, Madison, v.34, n.1, p.124-159, Jan. 1999.

GLEWWE, P.; JACOBY, H. “An economic analysis of delayed primary school enrollment in a low income country: The role of early childhood nutrition”. **Review of Economic Statistics**, 77 (1), 156–169, 1995.

GLEWWE, P.; JACOBY, H.; KING, E. “Early childhood nutrition and academic achievement: A longitudinal analysis”. **Journal of Public Economics**, 81 (3), 345–368, 2001.

GLEWWE, P.; MIGUEL, E. A. *The Impact of Child Health and Nutrition on Education in Less Developed Countries*. In: SCHULTZ, T. P.; STRAUSS, J. A. (Ed.) **Handbook of Development Economics**, Vol. 4, cap. 56, p. 3561-3606, 2008.

GROSSMAN, M. On the concept of health capital and demand for health. **Journal of Political Economy**, Chicago, v. 80, n. 2, p. 235-255, July 1972.

GROSSMAN, M. KAESTNER. The effect of education on health. In: J.R.Behrman; N.Stacey, ed., **The social benefits of education**. University of Michigan Press: Ann Arbor, 300p, 1997.

GROSSMAN, M. *Education and nonmarket outcomes*. In: HANUSHEK, E.; WELCH, F. (Ed.) **Handbook of the Economics of Education**, cap. 10, p.577-633, 2006.

HANUSHEK, A. *et alli*. Health and schooling: Evidence and policy implications for developing countries. **Economics of Education Review**, vol. 16(3), p. 271-282, Junho 1997.

HARTUNG, G. C.; PESSOA, S. **Fatores demográficos como determinantes da criminalidade**. XXXV Encontro Nacional de Economia, 2007. Recife, Pernambuco. Anais do XXXV Encontro Nacional de Economia.

KASSOUF, A.L.; SENAUER, B. Direct and indirect effects of parental education on malnutrition among children in Brazil: a full income approach. **Economic Development and Cultural Change**, Chicago, v.44, n.4, p.817-838, Jul. 1996.

KENKEL, D.S. Health behavior, health knowledge, and schooling. **Journal of Political Economy**. v. 99, p. 287-305. 1991.

MANKIW, G.; ROMER, D.; WEIL, D. The contribution to the empiric growth. **The Quarterly Journal of Economics**, v. 107, p. 407-37, 1992.

MUSHKIN, Selma J. “Health as an Investment”. **Journal of Political Economy**. 70, no. 2, suppl. p. 129-157, Outubro, 1962

ROMER, D. *Advanced macroeconomics*. New York, McGraw-Hill, 3 ed. 2006.

RUMBERGER, R. W.; WILLMS, J. D. The impact of racial and ethnic segregation on the achievement gap in California high schools. **Educational Evaluations and Policy Analysis**, v. 14, n. 4, p. 377-396, 1992.

SOARES, J. F.; ALVES, M. T. G. Desigualdades raciais no sistema brasileiro de educação básica. *Educação e Pesquisa*, v. 9, n. 1, p. 147-165, 2003.

STRAUSS, J. "Does Better Nutrition Raise Farm Productivity?" **Journal of Political Economy**. vol. 94 (2), p. 297-320, 1986.

THOMAS, D.; STRAUSS, J. Health and Wage: Evidence on men and women in urban Brazil. **Journal of Econometrics**. vol. 77 p. 159-185, 1997.

THOMAS, D.; STRAUSS J; HENRIQUES, M.H. How does mother's education affect child height. **The Journal of Human Resources**, Madison, v.26, n.2, p.183-211, Spring, 1991.

WOOLDRIDGE, J. M. **Econometric analysis of cross section and panel data**. Cambridge: The MIT Press, 2002.

Apêndice

Tabela 1A – Impacto dos indicadores de saúde no desempenho escolar, estimadores MQO agrupados (matemática - 4ª série)

Variável	MQO agrupados			
	I	II	III	IV
Percentual de nascidos vivos cujas mães tiveram 7 ou mais consultas pré-natais	0,370***	0,239***	0,094***	0,093***
Número de óbitos infantis por 10 mil habitantes	-7,751***	-2,494***	-0,723***	0,061***
Coefficiente de Mortalidade Infantil	0,575***	0,563***	0,010**	0,088***
Número de óbitos neonatais por 10 mil habitantes	0,482***	0,392***	0,087***	0,066
Coefficiente de Mortalidade Neonatal tardia	0,037***	0,018	-0,007**	-0,007*

Obs.: *Significativo a 10%, ** Significativo a 5%, *** Significativo a 1%

Fonte: Elaboração própria com base nos dados da Prova Brasil, Censo Escolar e Datasus, anos 2005 e 2007.

Tabela 2A – Impacto dos indicadores de saúde no desempenho escolar, estimadores MQO agrupados (português - 4ª série)

Variável	MQO agrupados			
	I	II	III	IV
Percentual de nascidos vivos cujas mães tiveram 7 ou mais consultas pré-natais	0,224***	0,087***	0,024	0,029*
Número de óbitos infantis por 10 mil habitantes	-6,749***	-1,794***	-0,569***	0,470**
Coefficiente de Mortalidade Infantil	0,170***	0,169***	0,073	0,076*
Número de óbitos neonatais por 10 mil habitantes	0,076	0,073	0,017	0,009
Coefficiente de Mortalidade Neonatal tardia	0,023***	0,005	-0,006**	-0,005*

Obs.: *Significativo a 10%, ** Significativo a 5%, *** Significativo a 1%

Fonte: Elaboração própria com base nos dados da Prova Brasil, Censo Escolar e Datasus, anos 2005 e 2007.