

O Impacto do Programa Saúde da Família Sobre a Saúde das Crianças da Área Rural do Brasil^{1,2}

Anderson Moreira Aristides dos Santos³ e Paulo de Andrade Jacinto⁴

Resumo: Este trabalho buscou analisar o impacto do Programa Saúde da Família (PSF) sobre a saúde das crianças (menores de 12 anos) da área rural do Brasil com base nos dados da PNAD (2008), utilizando os métodos *Propensity Score Matching* e Mínimos Quadrados Ordinários, aplicando análises de sensibilidade. Os resultados indicaram que o PSF possui um impacto estatisticamente significativo e de magnitude elevada sobre a saúde das crianças da área rural do Brasil. Contudo, ele está concentrado nas regiões Norte e Nordeste, sendo não significativo no Centro-Sul do País. No caso da região Norte, esse impacto independe da idade, enquanto na região Nordeste ele é significativo para as menores faixas de idade. A análise de sensibilidade indicou robustez dos resultados, principalmente para crianças mais novas. Portanto, há a confirmação da hipótese de que indivíduos das regiões mais pobres da área rural são beneficiados por esse tipo de política.

Palavras-chaves: Saúde da criança, Brasil rural, efeito do tratamento, Programa Saúde da Família.

Abstract: *This paper aims to analyze the impact of Family Health Program (PSF) on the children's health (under 12 years old) in Brazil's rural area, using data from the PNAD (2008) and the PSM and OLS methods, applying sensitivity analysis. The results indicate that the PSF has a statistically significant and high magnitude impact on the children's health in Brazil's rural area. However, it is concentrated in the North and Northeast, and it is not significant in the Mid-South. For the Northern region, this impact is independent of age, while in the Northeast it is significant for younger age groups. The sensitivity analysis show robustness of the results especially for younger children. There is confirmation of the hypothesis that this type of policy benefits the individuals of the poorest regions in rural area.*

Key-words: *Children's health, Brazil's rural area, treatment effect, Family Health Program.*

DOI - <http://dx.doi.org/10.1590/1234-56781806-94790550202> **Classificação JEL:** C21, I10, I18.

-
1. Data de submissão: 19 de agosto de 2015. Data de aceite: 19 de fevereiro de 2017.
 2. Este trabalho é baseado na tese do primeiro autor, que agradece ao fomento da Capes em seu doutoramento na PUC/RS.
 3. Universidade Federal de Alagoas. Maceió, Alagoas, Brasil. E-mail: anderson_moreira_aristides@hotmail.com
 4. Universidade Federal do Paraná. Curitiba, Paraná, Brasil. E-mail: paulo.jacinto@ufpr.br

1. Introdução

O Programa Saúde da Família (PSF), denominado atualmente como Estratégia Saúde da Família, foi implantado pelo Ministério da Saúde em 1994, com o objetivo de garantir acesso aos cuidados primários de saúde, com base nos princípios do SUS de universalidade, integralidade, equidade e participação social.

A equipe do PSF deveria ser composta de, no mínimo, um médico, um enfermeiro, um auxiliar de enfermagem e seis ou mais agentes comunitários de saúde, que seriam responsáveis pelo monitoramento de, no máximo, 4.000 pessoas de uma determinada área. A atuação dessas equipes aconteceria em unidades básicas de saúde, nas residências e com mobilizações nas comunidades⁵.

Em 1996, o programa estava presente em 4,6% dos municípios do Brasil; esse número aumentou para 92,1% em 2007 (REIS, 2014). Em 1998, existiam 2.000 equipes deste programa com 60.000 agentes comunitários de saúde que cobriam 4% da população do País. Em 2014, essas estatísticas aumentaram, respectivamente, para 39.000, 265.000 e 62% (MACINKO e HARRIS, 2015).

Essas informações evidenciam o importante papel que o PSF assumiu na atenção básica à saúde no Brasil, despertando o interesse para estudos com o intuito de avaliar o seu impacto sobre a taxa de mortalidade infantil no Brasil. Um exemplo disso é o estudo de Macinko *et al.* (2006) que, a partir de um painel de dados com efeitos fixos para estados do Brasil, no período de 1990-2002, mostrou que um aumento de 10% na cobertura do PSF estava associado a uma queda de 4,5% na taxa de mortalidade infantil.

Outro estudo de Macinko *et al.* (2007), considerando como unidade de análise as microrregiões do Brasil para o período 1999-2004, gerou evidências de que um aumento de 10% na cobertura do PSF estava associado a reduções de 0,45% na taxa de mortalidade infantil, 0,6% na mortalidade pós-neonatal e de 1% na mortalidade por diarreia. Porém, ao analisar a taxa de mortalidade neonatal, este estudo não apresentou associação significativa.

A importância do PSF também foi observada em Aquino *et al.* (2009), através de uma análise com o método de dados em painel com resposta binomial negativa e efeitos fixos para os municípios do Brasil no período entre 1996 e 2004. Os resultados evidenciaram que uma cobertura maior deste programa reduziu a taxa de mortalidade infantil. Empregando essa mesma técnica para o período de 2000 a 2005, Rasella *et al.* (2010)

5. Para mais detalhes, ver <http://dab.saude.gov.br/portaldab/ape_esf.php>.

mostraram que o PSF esteve associado à redução da mortalidade por diarreia e infecções do trato respiratório.

Em um estudo ainda mais amplo, Rocha e Soares (2010) analisaram o impacto do PSF sobre a taxa de mortalidade, considerando diferentes causas e faixas etárias, além dos efeitos indiretos na escolaridade, trabalho infantil, oferta de trabalho de adultos e fertilidade. Diferente dos trabalhos anteriores, os quais utilizaram a cobertura do programa, variável que pode ser potencialmente endógena, os autores aplicaram o método de diferença em diferenças, mensurando o tratamento através da presença do programa no município há “j” anos (j variando de 1 a 8). Essa análise contemplou os macrodados de municípios, na análise de impactos diretos, e microdados nos demais casos, todos no período 1991-2004. Os resultados mostraram que o PSF foi significativamente associado à redução na mortalidade, sendo que os efeitos foram maiores nas regiões mais pobres (Norte e Nordeste), sobre a mortalidade das crianças e nas causas perinatais, infecciosas, endócrinas e de doenças respiratórias. Este programa também apresentou impacto significativo sobre o aumento na oferta de trabalho, nas taxas de matrículas e nas reduções da fertilidade.

Em um trabalho mais recente, Guanais (2013) explorou o impacto da combinação da cobertura do PSF e do Programa Bolsa Família (PBF) sobre a taxa de mortalidade infantil pós-neonatal (PNTMI), a partir de um painel de dados com efeitos fixos para os municípios do Brasil, no período de 1998 a 2010. As evidências mostraram que a associação entre PSF e a PNTMI é mais forte quando a cobertura do PBF aumenta.

Um aspecto importante nos estudos relatados acima está relacionado ao uso de taxas de mortalidade, ou seja, nesses artigos foram empregados dados agregados. Para o Brasil, o uso de microdados, em nível de indivíduos, foi realizado no trabalho de Reis (2014) para mensurar o impacto do PSF sobre a saúde das crianças, a partir dos dados da Pesquisa Nacional por Amostras de Domicílios (PNAD) de 2003. A saúde foi mensurada por uma

medida subjetiva (saúde reportada) e algumas medidas mais objetivas, como, por exemplo, a prevalência de diarreia ou vômito. A estratégia empírica adotada foi comparar a saúde da criança quando nasceu (ou seis meses antes, tempo esse que corresponde ao período pré-natal), levando em conta a presença do PSF no município de sua residência, mas que, considerando o mesmo período de vida, não estava disponível para seu irmão mais velho, em uma amostra com crianças com idade entre 12 e 84 meses. Os resultados indicaram que crianças que viviam em regiões onde o PSF estava disponível durante o período pré-natal e na primeira infância são mais saudáveis do que crianças que não tinham o programa disponível durante o mesmo período de suas vidas.

Outro estudo empregando dados, em nível de indivíduos, foi realizado por Santos *et al.* (2012), que usaram em sua análise os fatores socioeconômicos associados à saúde individual (reportada pelos pais) das crianças menores de seis anos de idade, do Brasil rural, com dados da PNAD (2008). A partir da estimação de modelos não lineares, evidenciou-se que os efeitos marginais do PSF são significativos apenas para as regiões Norte e Nordeste, e que esses efeitos são maiores para os mais pobres e no caso de maior escolaridade das mães.

As evidências descritas sugerem que o PSF tem um impacto positivo importante sobre a saúde das crianças. Contudo, ainda é escasso o número de estudos que empregaram microdados. Espera-se em termos teóricos que, principalmente, os indivíduos mais pobres da área rural tenham se beneficiado da implantação desse programa, tanto devido à restrição de acesso a serviços de saúde para a população mais carente, como também por eles poderem residir longe da infraestrutura pública de saúde. No entanto, observa-se que não há estudos que analisam o impacto do PSF sobre saúde das crianças dessa área.

O PSF apresenta um grau de cobertura significativo na área rural. Segundo dados da PNAD (2008), 52% das crianças (menores de 12 anos) da área urbana residiam em domicílios que tinham

cadastro no PSF; essa estatística era majoritariamente superior na área rural, alcançando mais do que 69%.

A partir dessas considerações, este artigo tem como objetivo analisar o impacto do PSF sobre a saúde das crianças (menores de 12 anos) da área rural do Brasil com base nos dados da PNAD (2008). Para tanto, serão utilizados o método de *Propensity Score Matching* (PSM) e dos Mínimos Quadrados Ordinários (OLS) com a aplicação de testes de robustez para os resultados.

O presente estudo difere dos demais acima citados na medida em que foca na área rural. Além disso, os dados permitem identificar se o domicílio da criança possuía cadastro no PSF, o que é diferente da proposta de identificação usada em Reis (2014), que fez sua análise para o Brasil como um todo e mensurou o tratamento através da presença do programa no município, variável que não identifica exatamente se o domicílio possuía cobertura do programa. Em relação ao estudo de Santos *et al.* (2012), a diferença reside na análise do impacto desse programa, buscando inferir causalidade ao utilizar uma metodologia que tenta controlar a não aleatoriedade do programa e testes de robustez, considerando diferentes faixas de idade das crianças nas regiões do Brasil rural.

Além dessa breve introdução, o presente estudo está dividido em três seções. A segunda seção apresenta a base de dados e a estratégia empírica. A terceira seção contém a análise dos resultados e, por fim, a quarta seção faz algumas considerações finais.

2. Dados e estratégia empírica

A PNAD 2008 foi a fonte de dados utilizada neste artigo. A PNAD abrange todo o território brasileiro e apresenta, anualmente (exceção dos anos de Censo), características gerais da população, informações de educação, trabalho, renda, entre outras. Além disso, essa pesquisa periodicamente contém suplementos com temas específicos. Em 2008, o tema saúde foi abordado,

apresentando, entre outras questões, uma pergunta que identifica se o domicílio onde o indivíduo reside tem cadastro no PSF⁶.

A amostra deste trabalho é formada pelas crianças que tinham até 11 anos de idade (menores de 12 anos), filhos da pessoa de referência da família e residentes na área rural do Brasil (seguindo classificação do dicionário da PNAD). Todos os cálculos foram baseados nas seguintes faixas de idade: 0 a 3 (menores de 4 anos); 0 a 5 (menores de 6 anos) e 6 a 11 anos (menores de 12 anos). Além disso, algumas estimativas foram exploradas para a faixa de idade de 0 a 1 ano (menores de 2 anos). Os cálculos foram realizados para todas as regiões do País.

As crianças que viviam em domicílios que tinham cadastro no PSF há pelo um ano foram identificadas como beneficiárias dessa política (tratados), tendo como grupo de comparação as crianças de domicílios que não tinham cadastro nesse programa (grupo de controle).⁷ Neste caso, crianças com cadastro menor do que 1 ano foram excluídas, tentando garantir que os possíveis benefícios do programa tenham tido tempo necessário para ocorrerem.⁸

As medidas de saúde das crianças foram mensuradas por: i) saúde reportada pelos pais (*dummy* com 1 indicando saúde muito boa e 0 boa, regular, ruim e muito ruim); ii) prevalência de restrições de atividades; iii) prevalência de acamados; iv) prevalência de acamados pelo motivo de diarreia ou vômito e v) prevalência de procura

6. A Pesquisa Nacional de Saúde (PNS-2013) também apresenta essa questão; contudo, os dados disponíveis neste momento não contêm algumas variáveis de controle importantes, tal como renda. Além disso, nessa pesquisa não há como controlar efeitos fixos em nível de município.

7. Reis (2014) defende que a presença do programa no período pré-natal apresenta impacto importante na saúde da criança. Contudo, neste trabalho não há como criar uma variável para contemplar essa ideia, já que na PNAD, a última faixa para o programa é um ano ou mais.

8. Em geral, os resultados foram semelhantes ao considerar como tratados qualquer criança que residia em um domicílio cadastrado no PSF (independente do tempo). Esses resultados e outros que forem comentados e não apresentados aqui estão disponíveis aos interessados via contato com os autores. Ademais, a maioria (88,68%) dos cadastrados estava na faixa de cadastro de um ano ou mais.

à assistência médica pelo motivo de doença, considerando como período para os itens ii ao v , as duas semanas anteriores à pesquisa da PNAD.

A codificação da variável saúde reportada está baseada em outros trabalhos na literatura (NICOLELLA *et al.*, 2008; SANTOS *et al.*, 2012), diferente da saúde dos adultos, que em geral é codificada como muito boa e boa comparativamente às demais categorias. Isso se deve entre outros fatores, ao fato de as crianças estarem classificadas em sua ampla maioria (90%) nas categorias muito boa e boa.

As análises dos resultados deste trabalho se basearão, principalmente, nessa primeira medida. Apesar da limitação deste indicador, devido ao mesmo ser caracterizado por sua subjetividade, ele pode ser visto como amplo, podendo estar relacionado a vários tipos de morbidade e mortalidade (IDLER e BENYAMINI, 1997; NORONHA, 2005). Em relação às demais medidas, elas apresentam a clara vantagem de serem mensurações mais objetivas, a partir do relato de a criança ter apresentado algum problema de saúde que a leve a característica específica de cada indicador. Contudo, elas se restringem a um curto período de tempo (duas semanas), captando, desta forma, apenas um curto prazo e podem não ser capazes de mensurar diferentes morbidades, além de que a última medida depende da oferta de saúde. De qualquer forma, alguns resultados com base nessas variáveis também foram explorados com intuito de trazer mais robustez à análise.

Na avaliação do impacto do PSF sobre a saúde, devemos considerar a possibilidade de um viés de seleção, ou, em outras palavras, as crianças que apresentam indicadores socioeconômicos mais precários e, assim, com maior possibilidade de ter piores níveis de saúde, podem ter prioridade no atendimento do programa. Levando-se em conta este argumento, a metodologia utilizada neste trabalho se baseou numa tentativa de eliminação (ou mesmo redução) desse viés, através do método *propensity score matching*⁹.

Neste caso, considere um indivíduo i , uma variável de interesse Y (por exemplo, saúde auto-avaliada). Existem os indivíduos que participam do programa (tratados, $T = 1$) e os que não participam (controle, $T = 0$). Considerando a variável de interesse, o resultado do tratamento para cada indivíduo pode ser expresso por:

$$\pi_i = Y_i = TY_{1i} + (1 - T)Y_{0i} \quad (1)$$

Sendo que o impacto médio do programa para os beneficiários pode ser representado por:

$$\pi_i = E(Y_{1i} | T = 1) - E(Y_{0i} | T = 1) \quad (2)$$

Nota-se que o ideal seria observar os indivíduos nas situações com o tratamento (participando do PSF) e sem o tratamento. Dado que na ampla maioria de estudos de ciências sociais não há como se ter essa situação, deve-se observar um grupo de controle. Uma possível saída seria utilizar as informações de saúde dos indivíduos que não são tratados (os que não estão cadastrados no PSF); contudo, poderia haver um viés, já que as conclusões podem ser influenciadas pelo efeito de variáveis importantes pré-programa, podendo levar o pesquisador a conclusões equivocadas, ou seja, haveria distinções entre $E(Y_{0i} | T = 1)$ e $E(Y_{0i} | T = 0)$, influenciando, assim, o resultado.

Para contornar este problema, uma solução bastante conhecida, feita através de dados observáveis, e proposta inicialmente pelos autores Rosenbaum e Rubin (1983), refere-se ao PSM. A ideia básica por trás desse método é que o viés seria pelo menos reduzido se a comparação for baseada entre os indivíduos tratados e os de controle (não tratados) que sejam o mais similar possível.

Assim, o escore de propensão é apresentado na expressão (3), esta que representa a probabilidade de um indivíduo participar do programa a partir de suas características.

$$P(x_i) = \text{Probabilidade}(T_i = 1 | x_i) \quad (3)$$

Obtido o escore de propensão, por exemplo, através de uma distribuição de probabilidade logística, representado por $p(x_i)$, o efeito médio do tratamento sobre os tratados (ATT – *average*

9. Para mais detalhes sobre este método ver, por exemplo, Caliendo e Kopeinig (2008).

effect of treatment on the treated) pode ser representado pela seguinte expressão:

$$ATT = E\{E[Y_{1i} | T_i = 1, P(x_i)] - E[Y_{0i} | T_i = 0, P(x_i)] | T_i = 1\} \quad (4)$$

Primeiramente, estimou-se um modelo de regressão multivariada, buscando a mensuração da probabilidade de um indivíduo participar do programa PSF. A Tabela 1 apresenta as variáveis de controle incorporadas aos modelos desta etapa que foram estimados pelo método *logit*.

Contudo, encontrar indivíduos com o mesmo escore de propensão é uma tarefa que não pode ser executável, considerando a natureza contínua da variável, havendo formas de pareamento para contornar essa questão. O presente trabalho empregou os seguintes: i) método do vizinho mais próximo, nesse para cada elemento do

grupo dos tratados, n indivíduos do grupo de controle (os valores de n neste trabalho são 5 e 1) são selecionados considerando que sejam o mais próximo possível com base no valor do escore de propensão; ii) vizinhos mais próximos sem substituição, quando ocorre esse tipo de pareamento, cada observação de controle só é utilizada uma vez na comparação com tratados, diferente das formas do item anterior, em que cada observação do grupo de controle pode ser utilizada mais de uma vez na comparação com as observações dos tratados; iii) *radius*, neste caso cada observação do grupo de tratados é pareada com as observações de controle dentro de um raio específico, sendo que este trabalho combinará este método com um *caliper* de 0,1%, ou seja, esse será o valor base para a distância máxima tolerável entre os escores de propensão dos grupos de comparação;

Tabela 1. Variáveis independentes do modelo de probabilidade de participação no PSF

Variáveis ou Grupo	Explicação
Indicadores Individuais	
Idade	Idade da criança em meses
Sexo	<i>Dummy</i> com 1 indicando que a criança é do sexo feminino
Cor	<i>Dummy</i> com 1 indicando que a cor da criança é branca
Plano de Saúde	<i>Dummy</i> com 1 indicando que a criança possui plano de saúde
Indicadores Domiciliares	
Total Pessoas (Outras)	Número de pessoas com idade maior ou igual a 12 anos de idade que residem no domicílio da criança
Total de Crianças	Número de crianças de 0 a 11 anos de idade (menores de 12) que residem no domicílio da criança
Água	<i>Dummy</i> com 1 indicando que o domicílio onde a criança reside possui água canalizada ligada à rede geral ou através de poço ou nascente
Esgoto	<i>Dummy</i> com 1 indicando que o domicílio onde a criança reside possui banheiro com escoadouro conectado à rede coletora de esgoto ou pluvial ou a uma fossa séptica ligada ou não a uma rede coletora
Lixo	<i>Dummy</i> com 1 indicando que há coleta de lixo de forma direta ou indireta
Índice de posse	Índice obtido através do método de análise dos componentes principais, considerando a posse no domicílio dos seguintes indicadores: energia elétrica; paredes e telhados adequados; geladeira ou freezer; televisão; dvd; telefone; fogão; máquina de lavar; computador; carro ou moto.
Renda Domiciliar <i>per capita</i>	Renda domiciliar <i>per capita</i> (em logaritmo natural)
Bolsa Família	<i>Dummy</i> onde 1 indica que no domicílio da criança alguém recebe benefício do Programa Bolsa Família. Indicador obtido indiretamente, segundo proposta feita por Soares <i>et al.</i> (2010)
Indicadores dos pais	
Idade da Mãe	Idade da mãe em anos
Idade do Pai	Idade do pai em anos
Escolaridade do pai	Escolaridade do pai em anos
Escolaridade da mãe	Escolaridade da mãe em anos
Pai Fuma	<i>Dummy</i> onde 1 indica que o pai fuma ou já fumou diariamente
Mãe Fuma	<i>Dummy</i> onde 1 indica que a mãe fuma ou já fumou diariamente
Indicador das UFs	Efeitos Fixos para as Unidades da Federação (<i>Dummies</i>)

Fonte: Elaboração própria.

iv) Kernel, neste método cada observação dos tratados é combinada com várias observações do grupo de controle, com pesos inversamente proporcionais entre a distância desses dois grupos.

Com base nesses métodos, o ATT foi calculado. Todas as mensurações foram realizadas através do auxílio do software STATA. Ademais, estas estimativas serão comparadas às geradas pelo Método dos Mínimos Quadrados Ordinários (OLS), por um modelo OLS, o qual permite interações entre o tratamento e as demais covariadas (FILM – *fully interacted linear matching*)¹⁰ e a do método de regressão linear ajustada ponderada pelo inverso do *propensity score* (IPWRA).

Dois hipóteses fundamentais desse método podem ser destacadas. A primeira se refere à existência de suporte comum, que estabelece que em cada grupo de controle deve existir um tratamento correspondente, em que temos $0 < P(X) < 1$. A segunda, a hipótese de independência condicional (CIA), significa que, dado o controle de um conjunto de variáveis observadas (X) que não são afetadas pelo tratamento, os resultados potenciais são independentes do tratamento atribuído (participação no programa). A implicação desta hipótese é que todas as variáveis, as que afetam o tratamento (o programa PSF) e/ou o resultado (saúde das crianças), teriam que ser controladas no modelo (CALIENDO e KOPEINIG, 2008). A omissão de uma variável importante provocaria um viés no pareamento, não havendo como mensurá-lo. Entretanto, a literatura tem utilizado análises de sensibilidade para buscar dar robustez aos resultados.

O presente artigo aplicou o método sugerido por Ichino *et al.* (2008). A hipótese central para esse teste é que não haja independência condicional entre a variável de resultado (*outcome*) e o tratamento, condicional às variáveis observadas (X). Seguindo ideias de trabalhos anteriores,¹¹ os autores observam que a CIA deve ser assegurada dada uma variável binária não observada U e as

observadas (X). E, assim, se U fosse observado, o ATT poderia ser mensurado, dado que:

$$E(Y_0 | T = 1, X, U) = E(Y_0 | T = 0, X, U) \quad (5)$$

O passo seguinte desta análise caracteriza a distribuição de U com base em quatro parâmetros:

$$p_{ij} = \Pr(U = 1 | T = i, Y = j) = \Pr(U = 1 | T = i, Y = j, X) \quad (6)$$

Em que $i, j \in \{0, 1\}$, dada a probabilidade que $U = 1$ para cada um dos quatro grupos definido pela categoria do tratamento e do resultado (*outcome*).

Considerando os valores dos parâmetros p_{ij} , um valor de U é atribuído para cada indivíduo, considerando que pertença a alguma das quatro categorias (dada as combinações de ij), e U é tratado como mais uma variável observada no cálculo do ATT pelo método do PSM. Usando os valores dos parâmetros de sensibilidade, a estimação é repetida muitas vezes (no caso deste trabalho, 500) e a estimativa do ATT é obtida através da média dos ATTs da distribuição de U .

A abordagem sugerida por Oster (2015) também foi utilizada nesse estudo para dar robustez às análises. Para uma breve explicação, considere novamente variáveis não observadas U . Uma das hipóteses centrais se refere à seleção proporcional (*proportional selection assumptions*), representada pela seguinte equação:

$$\delta \frac{\sigma_{XT}}{\sigma_X} = \frac{\sigma_{UT}}{\sigma_U} \quad (7)$$

Em que $\sigma_{XT} = Cov(X, T)$; $\sigma_{UT} = Cov(U, T)$; $\sigma_X = Var(X)$; $\sigma_U = Var(U)$ e δ é o coeficiente de proporcionalidade.

Considere ainda, três modelos de regressão: um que tem como variável independente apenas o tratamento (PSF), com $\hat{\beta}$ e \hat{R} representando, respectivamente, o coeficiente estimado do tratamento e o R^2 dessa regressão; $\tilde{\beta}$ e \tilde{R} representam essas mesmas estatísticas sendo que para um modelo que inclui todas as variáveis independentes observadas (T e X); enquanto, β e R_{max} representam as estatísticas para um modelo que tem como variáveis independentes o tratamento (T), as observadas (X) e as não observadas (U).

10. Para uma aplicação deste método, ver Battistin, Nadai e Sianesi (2014).

11. Ver Rosenbaum e Rubin (1983), Rosenbaum (2002) e Altonji *et al.* (2005).

Dada a hipótese de proporcionalidade, uma seleção igual (entre as observáveis e as não observáveis) corresponde ao caso em que $\delta = 1$, com $\delta \neq 1$, a expressão abaixo refere-se a uma boa aproximação para o viés, quando δ está relativamente bastante próximo de 1:

$$\beta^* = \tilde{\beta} - \frac{\delta(\hat{\beta} - \tilde{\beta})(R_{max} - \tilde{R})}{\tilde{R} - \hat{R}} \quad (8)$$

E, assim, há a possibilidade da mensuração de um valor aproximado para $\hat{\delta}$ (equação 9) que é o grau de seleção para o qual β é igual a um valor alvo, $\tilde{\beta}$ quando $\hat{\beta} = 0$, isto nos mostra o grau de seleção sobre as não observáveis relativamente às observáveis, que seria suficiente para explicar uma ausência de impacto do tratamento.

$$\hat{\delta} \approx \frac{(\hat{\beta} - \tilde{\beta})(\tilde{R} - \hat{R})}{(\tilde{\beta} - \hat{\beta})(R_{max} - \tilde{R})} \quad (9)$$

Para a abordagem de Oster (2015), além das variáveis descritas na Tabela 1, a saúde da mãe e do pai e efeitos fixos de municípios¹² (no lugar de estados) foram incorporadas aos modelos. Este último caso tenta controlar fatores fixos municipais, como outras políticas públicas, as quais podem ser correlacionadas com o PSF. Contudo, algumas dificuldades podem surgir devido aos seguintes fatores: número alto de parâmetros a serem estimados (perda de graus de liberdade); pequena quantidade de crianças em alguns municípios e possível alta colinearidade com o PSF¹³. Uma abordagem alternativa seria incluir indicadores municipais; no entanto, a PNAD não permite identificar o nome dessas áreas geográficas para imputar essas informações.

12. Efeitos fixos para a Unidade Primária de Amostragem (UPA) que, no caso da PNAD, possui correspondência com os municípios.

13. No caso do PSM, a inclusão de efeitos fixos de municípios é tecnicamente inviável, ocorrendo erros no algoritmo devido à forte violação da hipótese de suporte comum.

3. Análise dos resultados

3.1. Análise Descritiva, participação no PSF e ajustes dos modelos

Antes de analisar os resultados do PSM, a Tabela 2 apresenta as médias de algumas variáveis deste trabalho. No que se refere às regiões Sudeste, Sul e Centro-Oeste, os principais resultados deste trabalho apresentaram-se relativamente semelhantes, assim, o Centro-Sul irá ser analisado dessa forma agregada. Nota-se que, para o Brasil, as crianças cadastradas no PSF (há mais de um ano) têm, em média, piores condições socioeconômicas (exceção da saúde) do que as não cadastradas. Por exemplo, entre os tratados, a renda média domiciliar *per capita* é de R\$ 159 e, para os não tratados, R\$ 211, ou seja, esse segundo caso é 33% maior do que o primeiro. Quando comparadas as estatísticas desses dois grupos para Norte e Nordeste, as diferenças são menores. Em todos os casos, a proporção de crianças cadastradas no Programa Bolsa Família é maior entre os tratados. Com base na amostra deste trabalho, pode-se destacar também que a cobertura do PSF é bastante superior no Nordeste, aproximadamente 80% das crianças da área rural dessa região possuíam cadastro no programa com tempo superior a um ano.

De acordo com modelos *logit*, a Tabela 3 apresenta a probabilidade de participação no PSF (sempre referindo-se a um tempo de cadastro superior a um ano) para o Brasil e regiões, nas faixas de idade de 0 a 5 e 6 a 11 anos.¹⁴ Observa-se que possuir abastecimento de água e de esgoto adequado e maiores níveis de renda são fatores associados a uma menor probabilidade de participação nesse programa, confirmando que, em geral, os indivíduos mais pobres são os beneficiários deste tipo de política. Maiores níveis de escolaridade dos pais estão associados a uma maior probabilidade de a criança ter esse tipo de cadastro

14. Deixa-se claro que resultados foram gerados para as demais faixas etárias e, por questões de espaço, não são apresentados aqui.

Tabela 2. Média das variáveis para tratados e não tratados, Brasil e regiões, crianças de 0 a 11 anos de idade

PSF	Brasil		Norte		Nordeste		Centro-Sul	
	Sim	Não	Sim	Não	Sim	Não	Sim	Não
Saúde Reportada	0,303	0,263	0,221	0,138	0,293	0,247	0,370	0,359
Acamados	0,026	0,028	0,027	0,040	0,028	0,030	0,023	0,020
Diarreia	0,012	0,013	0,011	0,029	0,012	0,008	0,012	0,006
Água	0,525	0,651	0,400	0,405	0,373	0,463	0,901	0,936
Esgoto	0,171	0,262	0,169	0,240	0,112	0,162	0,288	0,339
Escolaridade do Pai	3,577	4,145	3,810	3,730	2,747	2,759	5,094	5,294
Escolaridade da Mãe	4,576	5,069	5,090	4,568	3,853	3,913	5,717	6,132
Pai Fuma	0,440	0,450	0,442	0,487	0,437	0,399	0,444	0,456
Mãe Fuma	0,230	0,241	0,208	0,269	0,244	0,215	0,217	0,239
Renda	159,027	210,871	163,269	174,111	115,097	136,696	244,009	282,232
Bolsa Família	0,590	0,401	0,530	0,376	0,705	0,625	0,397	0,278
PSF	0,657		0,511		0,798		0,553	
Observações	7395	3867	1199	1147	4124	1044	2072	1676

Fonte: Cálculos do autor com base nos dados da PNAD (2008).

Tabela 3. Modelo logit de probabilidade de participação no PSF para o Brasil e regiões e faixas de idade

Variáveis	Brasil		Norte		Nordeste		Centro-Sul	
	0 a 5	6 a 11	0 a 5	6 a 11	0 a 5	6 a 11	0 a 5	6 a 11
Água	-0,177* (0,097)	-0,153* (0,086)	0,028 (0,198)	0,198 (0,182)	-0,380*** (0,138)	-0,271** (0,124)	-0,011 (0,230)	-0,099 (0,196)
Esgoto	-0,158 (0,100)	-0,182** (0,088)	-0,508** (0,231)	-0,674*** (0,218)	0,050 (0,181)	-0,446*** (0,159)	-0,146 (0,148)	0,139 (0,126)
Escolaridade do Pai	0,026** (0,013)	0,037*** (0,011)	0,018 (0,027)	0,060** (0,024)	0,010 (0,021)	0,031 (0,021)	0,045** (0,020)	0,036** (0,017)
Escolaridade da Mãe	0,035*** (0,012)	0,019* (0,011)	0,075*** (0,026)	0,082*** (0,023)	0,021 (0,021)	-0,001 (0,019)	0,033* (0,020)	-0,006 (0,017)
Renda	-0,175*** (0,056)	-0,177*** (0,052)	-0,349*** (0,117)	-0,259** (0,112)	0,026 (0,090)	-0,086 (0,084)	-0,249** (0,098)	-0,168* (0,087)
Bolsa Família	0,386*** (0,081)	0,459*** (0,071)	0,502*** (0,156)	0,606*** (0,140)	0,264** (0,129)	0,487*** (0,123)	0,441*** (0,146)	0,299** (0,117)
N	4.603	5.671	2.121	2.569	997	1.164	1.485	1.938

Notas: Os modelos contêm todas as variáveis indicadas na Tabela 1, que não estão apresentadas nessa tabela por questões de espaço.

Fonte: Cálculos do autor com base nos dados da PNAD (2008).

no PSF, o que pode estar relacionado ao argumento de que pais mais educados entendem melhor a importância de aspectos relacionados à saúde. O fato de ser beneficiário do Programa Bolsa Família, variável mensurada pela *proxy* sugerida por Soares *et al.* (2010), aumenta a probabilidade de o domicílio da criança possuir esse tipo de cadastro no PSF. O PBF exige condicionalidades relacionadas à saúde das crianças e das gestantes, o que pode afetar positivamente a procura por serviços de saúde, criando sinergias entre esses programas.

No método explorado neste trabalho, um aspecto importante que deve ocorrer é que, após o pareamento, não haja diferenças nas características dos indivíduos tratados e não tratados (modelo balanceado).¹⁵ A Tabela 4 apresenta as estatísticas para dar sustentação a essa condição. Nota-se um bom ajuste após o pareamento com grandes quedas no Pseudo R^2 e nos valores do viés médio e

15. Gráficos foram gerados para analisar a área de suporte comum, em que houve resultados satisfatórios. Eles estão disponíveis aos interessados através de contato com os autores.

Tabela 4. Balance antes e depois do pareamento para o Brasil e regiões e faixas de idade

	Pseudo R ²	Valor-p do χ^2	Viés Médio	Viés Mediano	Pseudo R ²	Valor-p do χ^2	Viés Médio	Viés Mediano
	<i>Brasil (0 a 5)</i>				<i>Brasil (6 a 11)</i>			
Não Pareada	0,17	0,00	13,5	13,3	0,149	0,000	14,1	14,3
Pareada	0,01	0,18	2,0	1,2	0,005	0,146	1,8	0,9
	<i>Norte (0 a 5)</i>				<i>Norte (6 a 11)</i>			
Não Pareada	0,14	0,00	13,2	9,9	0,16	0,00	15,6	13,0
Pareada	0,01	0,88	3,8	2,9	0,01	0,85	3,4	2,4
	<i>Nordeste (0 a 5)</i>				<i>Nordeste (6 a 11)</i>			
Não Pareada	0,10	0,00	13,3	10,0	0,10	0,00	12,1	8,7
Pareada	0,02	0,00	5,1	3,6	0,01	0,00	3,2	1,8
	<i>Centro-Sul (0 a 5)</i>				<i>Centro-Sul (6 a 11)</i>			
Não Pareada	0,114	0,00	12,1	9,6	0,10	0,00	12,3	9,4
Pareada	0,01	0,96	3,2	2,1	0,01	0,55	3,6	2,4

Notas: Geradas com base no pareamento kernel.

Fonte: Cálculos do autor com base nos dados da PNAD (2008).

mediano (alcançando valores baixos). O resultado para o teste razão verossimilhança de significância conjunta apresenta-se não significativo, com exceção para a região Nordeste. Contudo, mesmo no caso desta última região, houve importantes reduções nas demais estatísticas (viés e Pseudo R²), além de que alguns testes adicionais mostraram balanceamento da amostra.¹⁶

3.2. Análise do impacto do PSF

A Tabela 5 apresenta a estimativa do impacto do PSF sobre a saúde das crianças (reportadas pelos pais) da área rural do Brasil, considerando os diferentes métodos e faixas de idade. Observa-se que, em geral, os resultados são significativos a 10%, para as menores faixas de idade e não significativos em todos os métodos para as crianças de 6 a 11 anos de idade. A magnitude do efeito é maior para o grupo mais novo, além de que, nele, esses resultados são mais robustos independente do método aplicado. Por exemplo, considerando o método do vizinho mais próximo (NN-1) e essas crianças mais novas (0 a 3 anos), nos tratados a proporção de crianças com saúde

muito boa é 8,8 pontos percentuais (p.p.) maior do que de seu grupo de controle (crianças que residem em domicílios que não têm cadastro no PSF), sendo essa a estimativa do ATT; já o efeito para a segunda faixa (0 a 5 anos) é de 4,8 p.p. Além disso, nota-se que o coeficiente gerado pelo OLS é inferior aos demais casos. Deve-se destacar que este último método é sensível a problemas de erro de especificação, além do fato de que ele não consegue captar efeitos heterogêneos do tratamento. Quando analisado o OLS-FILM, estimador semelhante, mas que permite interações do tratamento com as demais covariadas, coeficientes próximos aos de alguns pareamentos podem ser visualizados.

Considerando as possíveis diferenças entre as regiões do Brasil, a Tabela 6 apresenta os cálculos para as regiões Norte e Nordeste. Nota-se que, para essa primeira região, os resultados são todos significativos em nível de significância de 5% para os diferentes métodos e as três faixas de idade. Além disso, diferente dos resultados encontrados para o Brasil, os coeficientes do OLS ficam bem próximos aos de muitas estimativas do ATT de diferentes pareamentos. Como exemplo, na faixa etária entre 0 e 5 anos, o OLS apresenta um efeito de 9,1 p.p., enquanto o ATT do kernel é de 9,7 p.p.

16. Aplicamos também o método proposto por Imai e Ratkovic (2014). Os resultados indicaram aceitação da hipótese de modelo balanceado, inclusive para a região Nordeste.

Tabela 5. Efeito do PSF sobre a saúde das crianças, reportada pelos pais, da área rural do Brasil, segundo diferentes métodos e faixas de idade

Método	Estatística	0 a 3	0 a 5	6 a 11
Kernel	ATT	0,074***	0,033	0,010
	Erro padrão	(0,022)	(0,021)	(0,017)
NN(1)	ATT	0,088**	0,048*	0,005
	Erro Padrão	(0,035)	(0,028)	(0,024)
NN(5)	ATT	0,088***	0,052**	0,010
	Erro Padrão	(0,023)	(0,023)	(0,020)
NN(1) SR	ATT	0,047**	0,029*	0,021
	Erro Padrão	(0,021)	(0,016)	(0,014)
Radius	ATT	0,078***	0,042**	0,007
	Erro Padrão	(0,023)	(0,020)	(0,018)
IPWRA	ATT	0,095***	0,056***	0,017
	Erro Padrão	(0,021)	(0,018)	(0,017)
OLS-FILM	ATT	0,086***	0,044**	0,016
	Erro Padrão	(0,028)	(0,021)	(0,018)
OLS	ATT	0,049***	0,026*	0,022*
	Erro Padrão	(0,019)	(0,015)	(0,013)
	N-obs.	2959	4603	5671

Notas: Kernel – Tipo Epanechnikov com o valor da janela (*bandwidth*) de 0,06 e utilizando suporte comum; NN(1) – método do vizinho mais próximo com reposição; NN(5) – método com os cinco vizinhos mais próximos com reposição; NN(1) SR – método do vizinho mais próximo sem reposição; Radius – Caliper de 0,1% utilizando suporte comum; IPWRA – regressão ajustada, ponderada pelo inverso da probabilidade; OLS-FILM – Mínimos Quadrados Ordinários com interações do tratamento; OLS – Mínimos Quadrado Ordinários. Para kernel, NN(1)SR e Radius erros padrões gerados por *bootstrap* com 200 replicações e uso do comando *psmatch2* no Stata de Leuven e Sianesi (2003). Para os métodos NN(1) e NN(5) erros padrões gerados com o comando *teffect* do Stata, considerando as contribuições do artigo de Abadie e Imbens (2012). IPWRA foi calculado pelo *teffect* com erros padrões robustos. Erros Padrões robustos também para o OLS. Ver Tabela 1 para os indicadores utilizados *no propensity score*.

Fonte: Cálculos do autor com base nos dados da PNAD (2008).

Tabela 6. Efeito do PSF sobre a saúde das crianças, reportada pelos pais, da área rural das regiões Norte e Nordeste, segundo diferentes métodos e faixas de idade

		Norte			Nordeste		
		0 a 3	0 a 5	6 a 11	0 a 3	0 a 5	6 a 11
Kernel	ATT	0,106***	0,097***	0,079***	0,105***	0,063**	0,003
	Erro padrão	(0,038)	(0,030)	(0,025)	(0,037)	(0,031)	(0,029)
NN(1)	ATT	0,174***	0,097**	0,093**	0,108**	0,097**	-0,023
	Erro Padrão	(0,044)	(0,042)	(0,039)	(0,054)	(0,045)	(0,041)
NN(5)	ATT	0,124***	0,103***	0,088***	0,110**	0,067*	0,004
	Erro Padrão	(0,042)	(0,035)	(0,034)	(0,047)	(0,039)	(0,031)
NN(1) SR	ATT	0,129***	0,104***	0,068***	0,023	-0,009	-0,040
	Erro Padrão	(0,031)	(0,025)	(0,022)	(0,036)	(0,031)	(0,026)
Radius	ATT	0,100**	0,080**	0,071***	0,108***	0,069**	-0,001
	Erro Padrão	(0,043)	(0,031)	(0,027)	(0,040)	(0,033)	(0,030)
IPWRA	ATT	0,116***	0,111***	0,078***	0,126***	0,077***	0,012
	Erro Padrão	(0,034)	(0,027)	(0,024)	(0,033)	(0,028)	(0,027)
OLS-FILM	ATT	0,108***	0,098***	0,084***	0,109**	0,063*	0,007
	Erro Padrão	(0,037)	(0,031)	(0,030)	(0,046)	(0,035)	(0,029)
OLS	ATT	0,095***	0,091***	0,072***	0,064**	0,019	0,010
	Erro Padrão	(0,032)	(0,025)	(0,024)	(0,031)	(0,025)	(0,021)
	N-obs.	655	997	1164	1371	2121	2569

Notas: Ver Tabela 5 para detalhes das legendas e Tabela 1 para os indicadores utilizados *no propensity score*.

Fonte: Cálculos do autor com base nos dados da PNAD (2008).

Os resultados da região Nordeste apresentam um padrão relativamente semelhante ao nacional. Observa-se que o impacto do PSF é significativo nas duas menores faixas etárias, sendo que a magnitude é maior para as crianças mais novas. Por exemplo, o método do kernel mostra que as crianças de 0 a 3 anos, beneficiárias desse programa, têm média de saúde muito boa, 10,5 p.p. maior do que de seu grupo de comparação. Os coeficientes do OLS são inferiores aos apresentados nos diferentes pareamentos (6,4 p.p. nessa menor faixa de idade e não significativo para faixa de 0 a 5 anos). Resultados para as crianças da área rural do Centro-Sul do País podem ser visualizados na Tabela 7. Em geral, eles são todos não significativos.¹⁷

Resultados semelhantes aos encontrados aqui, em relação à importância do PSF nas regiões Norte e Nordeste, foram observados em Santos *et al.* (2012), contudo, apenas em uma análise de associação do PSF com a saúde infantil na área rural. Já Rocha e Soares (2010) encontraram efeito significativo e de larga magnitude desse programa nessas duas regiões; por exemplo, na região Norte, municípios que estavam no programa há oito anos tiveram redução de 15 óbitos na taxa de mortalidade, por mil crianças menores de 1 ano, no período entre 1995 e 2003. Além disso, os impactos sobre as taxas de mortalidade nessa faixa etária não foram significativos nas regiões Sudeste e Sul; já para a mortalidade entre 1 e 4 anos não houve efeito significativo para a região Centro-Oeste.

Observa-se, dessa forma, a importância do PSF, principalmente para as regiões mais pobres do País. Macinko e Harris (2015) argumentam que um dos aspectos importantes desse programa se refere ao seu desenho, como o papel dos agentes comunitários de saúde, que ficam responsáveis por realizar visitas às famílias de uma deter-

minada área, independente da necessidade ou demanda, realizando atividades de promoção à saúde.

O foco desse estudo reside na saúde reportada. Contudo, resultados foram explorados com base em medidas mais objetivas, que conseguem captar um curto período de tempo. Os ATTs para esses casos foram estimados para as regiões Norte e Nordeste. A Tabela 8 apresenta os casos que em geral foram significativos, ou seja, restrições de atividades por diarreia ou vômito e prevalência de acamados na região Norte.¹⁸ Observa-se que os resultados dessas duas medidas são significativos em nível de 10% para as crianças mais novas (duas menores faixas), com algumas exceções, dependendo do método do pareamento. Além disso, as magnitudes são grandes, já que, por exemplo, o método kernel indica que, para as crianças entre 0 e 5 anos, o impacto do PSF sobre diarreia e vômito é de -3 p.p., sendo que a prevalência para esse indicador é de 3,3% (média não condicional). Para prevalência de acamados, no grupo mais novo dessa tabela, o impacto é de -4,8 p.p. (com prevalência igual a 5%). Os coeficientes gerados pelo OLS são próximos aos de vários pareamentos.

Esses fatores analisados na última tabela podem impactar, principalmente, nos primeiros anos de vida das crianças; contudo, duas dificuldades surgem ao tentar fazer uma análise para os grupos de idade ainda menores: quantidade de observações reduzidas e alguns sucessos (ou fracassos) perfeitos nos efeitos fixos de estado.¹⁹ Dessa forma, com intuito de explorar alguns desses casos, este artigo utilizou alguns métodos de pareamento, considerando as variáveis da Tabela 1, mas sem incluir os efeitos fixos de estado, para crianças entre 0 e 1 ano (menores de 2 anos); essas estatísticas são apresentadas na Tabela 9. Observam-se resultados significativos e de magnitudes largas. Por exemplo, para o OLS,

17. Como observado em Caliendo e Kopeinig (2008), variáveis afetadas pelo tratamento não devem ser incluídas no modelo de *propensity score*. Assim, as variáveis saúde do pai e da mãe não foram utilizadas para gerar os resultados apresentados nestas tabelas; entretanto, foram realizados testes com essas variáveis e os resultados, em geral, não mudaram.

18. Alguns métodos indicaram significância estatística para o indicador de assistência à saúde devido a doenças.

19. Além disso, métodos em que os erros-padrão seriam gerados via *bootstrapping* não foram apresentados, já que houve muitas falhas em seu processo de geração.

Tabela 7. Efeito do PSF sobre a saúde das crianças, reportada pelos pais, da área rural do Centro-Sul, segundo diferentes métodos e faixas de idade

		0 a 3	0 a 5	6 a 11
Kernel	ATT	-0,016	-0,039	-0,027
	Erro padrão	(0,040)	(0,030)	(0,027)
NN(1)	ATT	-0,005	-0,074	-0,035
	Erro Padrão	(0,064)	(0,046)	(0,037)
NN(5)	ATT	-0,023	-0,054	-0,045
	Erro Padrão	(0,045)	(0,035)	(0,030)
NN(1) SR	ATT	-0,012	-0,003	-0,008
	Erro Padrão	(0,031)	(0,029)	(0,024)
Radius	ATT	0,017	-0,058*	-0,030
	Erro Padrão	(0,042)	(0,034)	(0,028)
IPWRA	ATT	-0,042	-0,043	-0,011
	Erro Padrão	(0,038)	(0,031)	(0,025)
OLS-FILM	ATT	-0,033	-0,031	-0,018
	Erro Padrão	(0,040)	(0,031)	(0,027)
OLS	ATT	-0,028	-0,018	-0,005
	Erro Padrão	(0,032)	(0,026)	(0,022)
N-obs.		1012	1610	2052

Notas: Ver Tabela 5 para detalhes das legendas e Tabela 1 para os indicadores utilizados no propensity score.

Fonte: Cálculos do autor com base nos dados da PNAD (2008).

Tabela 8. Efeito do PSF sobre indicadores de saúde das crianças, diarreia ou vômito e acamados, da área rural da região Norte, segundo diferentes métodos e faixas de idade

		Diarreia ou Vômito			Acamados		
		0 a 3	0 a 5	6 a 11	0 a 3	0 a 5	6 a 11
Kernel	ATT	-0,025	-0,030*	-0,003	-0,048*	-0,037**	-0,002
	Erro padrão	(0,019)	(0,017)	(0,006)	(0,025)	(0,019)	(0,014)
NN(1)	ATT	-0,029*	-0,017	-0,002	-0,048*	-0,010	-0,002
	Erro Padrão	(0,017)	(0,017)	(0,010)	(0,028)	(0,020)	(0,016)
NN(5)	ATT	-0,034*	-0,036**	-0,003	-0,057**	-0,057*	0,002
	Erro Padrão	(0,018)	(0,015)	(0,009)	(0,022)	(0,031)	(0,013)
NN(1) SR	ATT	-0,035**	-0,033***	-0,005	-0,038**	-0,033**	0,000
	Erro Padrão	(0,015)	(0,012)	(0,006)	(0,017)	(0,013)	(0,008)
Radius	ATT	-0,025	-0,028*	-0,005	-0,046*	-0,037*	0,001
	Erro Padrão	(0,019)	(0,016)	(0,007)	(0,026)	(0,019)	(0,013)
IPWRA	ATT	-0,030	-0,031*	-0,003	-0,046*	-0,038*	-0,004
	Erro Padrão	(0,020)	(0,017)	(0,006)	0,024	0,020	(0,013)
OLS- FILM	ATT	-0,034*	-0,037**	-0,003	-0,053**	-0,046**	-0,007
	Erro Padrão	(0,019)	(0,015)	(0,009)	(0,022)	(0,018)	(0,012)
OLS	ATT	-0,027*	-0,024*	-0,004	-0,041**	-0,037**	0,001
	Erro Padrão	(0,016)	(0,012)	(0,007)	(0,019)	(0,015)	(0,010)
N-obs.		655	997	1164	655	997	1164

Notas: Ver Tabela 5 para detalhes das legendas e Tabela 1 para os indicadores utilizados no propensity score.

Fonte: Cálculos do autor com base nos dados da PNAD (2008).

Tabela 9. Efeito do PSF sobre indicadores de saúde das crianças, diarreia ou vômito e acamados, da área rural da região Norte, segundo diferentes métodos e crianças de 0 a 1 ano

	Diarreia ou vômito	Acamados	Saúde Autoavaliada
NN(1)			
ATT	-0,077**	-0,077*	0,140*
Erro Padrão	(0,037)	(0,042)	(0,074)
NN(5)			
ATT	-0,077***	-0,073**	0,150**
Erro Padrão	(0,028)	(0,031)	(0,060)
IPWRA			
ATT	-0,087***	-0,083***	0,170***
Erro Padrão	(0,029)	(0,031)	(0,051)
OLS-FILM			
ATT	-0,079***	-0,078***	0,170***
Erro Padrão	(0,029)	(0,027)	(0,053)
OLS			
ATT	-0,079***	-0,049*	0,136***
Erro Padrão	(0,027)	(0,029)	(0,051)
N-obs.	298	298	298

Notas: Ver Tabela 5 para detalhes das legendas e Tabela 1 para os indicadores utilizados *no propensity score*. Contudo, diferente dos casos anteriores, os efeitos fixos de estados não foram utilizados para o cálculo dos modelos desta tabela.

Fonte: Cálculos do autor com base nos dados da PNAD (2008).

o fato de o domicílio da criança estar cadastrado no PSF reduz em 7,9 p.p. a probabilidade de ter tido restrições de atividades pelo motivo diarreia ou vômito, o que é bastante próximo ao gerado, por exemplo, pelo método do vizinho mais próximo, sendo que a prevalência dessa variável é de 5,36% (média não condicional). Nota-se que esse número é expressivo: no cálculo do ATT (NN-1), os grupos controles e tratados possuem, respectivamente, prevalência condicional de 9,10% e 1,40%.

Em sua análise com macrodados, Rocha e Soares (2010) observaram que os maiores impactos do PSF sobre a taxa de mortalidade infantil estão associadas com a mortalidade por afecções perinatais, doenças infecciosas e doenças respiratórias. As causas de mortes incluem: problemas durante a gravidez, diarreia e outras doenças intestinais, além de gripe, asma e bronquite. Os autores defendem que o apoio e informações fornecidas pela presença do PSF são eficazes principalmente nesse grupo de doenças. Macinko *et al.* (2007) e Rasella, Aquino e Barreto (2010) também observaram uma associação entre PSF e reduções na mortalidade por diarreia. Enquanto que, com

o uso de microdados, Reis (2009) mostrou que a presença do PSF no período pré-natal tem um impacto de redução em 0,7 p.p. na prevalência de restrições de atividades pelo motivo diarreia ou vômito e 1,8 p.p. na prevalência de acamados.

3.3. Análise de robustez dos resultados

A Tabela 10 apresenta os resultados de relaxamento da CIA através da análise de sensibilidade proposta por Ichino *et al.* (2008) para a região Norte e a variável saúde reportada. A calibragem de fatores não observados U é realizada “imitando” algumas variáveis de controles.²⁰ Esse teste não comporta variáveis contínuas, porque, segundo os autores que propuseram esse método, este fato não acarreta em perda de generalidade. Desta forma, os indicadores contínuos utilizados para calibragem foram transformados em variáveis binárias e suas categorias estão representadas entre parênteses nessa tabela (quando a *dummy* assume valor 1). Ainda, esses cálculos

20. Os resultados focam em algumas das principais variáveis. Os resultados também foram robustos aos demais casos.

Tabela 10. Análise de sensibilidade de Ichino *et al.* (2008), saúde reportada, região Norte

	p11	p10	p01	p00	Γ	Λ	ATT	SE
Norte – 0 a 3 anos								
<i>No confounder</i>	0,00	0,00	0,00	0,00	–	–	0,173	0,044
Crianças (≤ 2)	0,67	0,48	0,53	0,54	1,548	0,987	0,125	0,053
Água	0,48	0,38	0,52	0,34	2,621	1,226	0,119	0,054
Escolaridade da Mãe (≤ 4)	0,35	0,49	0,30	0,55	0,45	0,52	0,116	0,055
Renda (≤ 140)	0,49	0,66	0,40	0,62	0,364	1,162	0,129	0,052
Bolsa Família	0,44	0,42	0,46	0,26	3,08	1,957	0,101	0,058
Saúde da Mãe (Boa e Muito Boa)	0,74	0,71	0,74	0,71	1,380	1,037	0,129	0,053
Norte – 0 a 5 anos								
<i>No confounder</i>	0,00	0,00	0,00	0,00	–	–	0,098	0,036
Crianças ≤ 2	0,61	0,47	0,59	0,50	1,583	0,966	0,099	0,043
Água	0,42	0,51	0,38	0,55	0,519	0,872	0,099	0,043
Escolaridade da Mãe (≤ 4)	0,17	0,19	0,22	0,24	0,949	0,743	0,098	0,043
Renda (≤ 140)	0,54	0,65	0,42	0,62	0,442	1,137	0,101	0,044
Bolsa Família	0,48	0,43	0,41	0,28	1,886	1,916	0,096	0,046
Saúde da Mãe (Boa e Muito Boa)	0,77	0,71	0,77	0,70	1,586	1,117	0,097	0,044
Norte – 6 a 11 anos								
<i>No confounder</i>	0,00	0,00	0,00	0,00	–	–	0,094	0,039
Crianças ≤ 2	0,48	0,47	0,51	0,51	1,098	0,875	0,084	0,041
Água	0,38	0,41	0,50	0,41	1,533	0,940	0,085	0,041
Escolaridade da Mãe (≤ 4)	0,59	0,54	0,50	0,67	0,494	0,670	0,075	0,043
Renda (≤ 140)	0,27	0,21	0,43	0,28	2,113	0,682	0,088	0,041
Bolsa Família	0,46	0,45	0,60	0,49	1,616	0,803	0,085	0,041
Saúde da Mãe (Boa e Muito Boa)	0,67	0,66	0,42	0,62	0,413	1,377	0,089	0,041

Notas: As primeiras colunas apresentam os valores p_j utilizados para simular a variável *confounder* binária (U) para cada caso. Γ são razões de chances, para os não tratados, da variável de interesse (*outcome*) considerando o efeito de U e controlando as covariadas observadas (X). Λ representa as razões de chances do tratamento (T), considerando o efeito de U e controlando as covariadas observadas (X). *ATT* foi gerado com base no método dos vizinhos mais próximos (NN-1), através da incorporação de U como covariada, e a estatística representa a média do efeito num processo repetido 500 vezes, em que SE é o erro-padrão dessa estatística. O comando para STATA utilizado foi o *sensatt*, proposto por Nannicini (2007).

Fonte: Cálculos do autor com base nos dados da PNAD (2008).

estão baseados no método do vizinho mais próximo. Nota-se que os resultados são robustos, independente da faixa etária, sendo todos estatisticamente significativos a 10%. No caso das crianças de 0 a 3 anos, a estimativa do ATT tem queda bastante acentuada; entretanto, esse impacto fica próximo de estimativas geradas pelos outros pareamentos ou mesmo do OLS. Por exemplo, no caso da simulação com o Bolsa Família, o ATT é de 10,1 p.p., número próximo à estimativa gerada pelo ATT do Radius (10 p.p.). Nas demais faixas, visualiza-se variações bem menores do ATT, principalmente nas crianças entre 0 e 5 anos de idade.

Nos resultados para a região Nordeste (Tabela 11), nota-se que, entre as crianças de 0 a 3 anos, os resultados são todos estatisticamente significativos e o ATT não varia muito. Porém, para crian-

ças de 0 a 5 anos, quando considerada a variável água na calibração, o ATT é estatisticamente não significativo.

Crianças que residem em municípios que implantaram políticas públicas de saúde (além do PSF) ou outras que afetam esse indicador, ou em termos gerais que possuíam melhor eficiência em sua gestão que afetasse a qualidade de vida da população, podem ter saúde melhor do que as que viviam em lugares que não tiveram essas condições, ou em outras palavras, fatores não observados fixos nos municípios que podem ser correlacionados com o PSF poderiam estar afetando o resultado. Visando controlar essa questão, na abordagem de Oster (2015), efeitos fixos em nível municipal foram incorporados, além da saúde dos pais.

Tabela 11. Análise de sensibilidade de Ichino *et al.* (2008), saúde reportada, região Nordeste

	p11	p10	p01	p00	Γ	Λ	ATT	SE
Nordeste – 0 a 3 anos								
<i>No confounder</i>	0,00	0,00	0,00	0,00	–	–	0,108	0,054
Crianças ≤2	0,68	0,61	0,70	0,63	1,610	0,931	0,108	0,057
Água	0,36	0,35	0,44	0,51	0,753	0,560	0,107	0,061
Escolaridade da Mãe (≤4)	0,49	0,57	0,61	0,50	1,839	1,105	0,109	0,057
Renda (≤140)	0,16	0,19	0,12	0,13	0,979	1,511	0,106	0,058
Bolsa Família	0,41	0,38	0,32	0,30	1,204	1,504	0,111	0,060
Saúde da Mãe (Boa e Muito Boa)	0,66	0,74	0,74	0,72	1,245	0,973	0,107	0,057
Nordeste – 0 a 5 anos								
<i>No confounder</i>	0,00	0,00	0,00	0,00	–	–	0,085	0,042
Crianças ≤2	0,63	0,58	0,69	0,62	1,422	0,825	0,084	0,048
Água	0,34	0,36	0,41	0,49	0,736	0,632	0,077	0,049
Escolaridade da Mãe (≤4)	0,51	0,60	0,64	0,56	1,488	0,984	0,086	0,046
Renda (≤140)	0,19	0,19	0,19	0,15	1,403	1,234	0,082	0,046
Bolsa Família	0,41	0,38	0,31	0,33	0,959	1,318	0,081	0,048
Saúde da Mãe (Boa e Muito Boa)	0,69	0,75	0,73	0,72	1,130	1,067	0,085	0,046

Notas: Ver legenda da tabela anterior.

Fonte: Cálculos do autor com base nos dados da PNAD (2008).

Dessa forma, a Tabela 12 explora essas novas estimações pelo método OLS, para as regiões Norte e Nordeste, considerando as diferentes faixas de idade. Observa-se que os resultados são robustos para a região Norte, independente da faixa etária, sendo que o efeito em termos de magnitude é maior para as crianças mais novas. Por exemplo, na faixa etária entre 0 e 3 anos, o fato de ter sido tratado aumenta a probabilidade de ter saúde muito boa em 8,2 p.p, resultado bastante expressivo, considerando que a média desse indicador é de 20%. Já para a região Nordeste, os resultados passaram a ser não significativos independente da faixa etária. Cabe ressaltar que, para essa última região, os resultados de tabelas ante-

riores indicaram que o OLS se mostrou diferente de métodos que controlam de alguma forma efeitos heterogêneos do tratamento. Ademais, lembra-se que, quando aplicado o PSM e o teste de sensibilidade de Ichino *et al.* (2008), os resultados dessa região foram robustos, principalmente para o grupo de crianças mais novas (0 a 3 anos).

A Tabela 13 apresenta os resultados da abordagem de Oster (2015) com base nos modelos da Tabela 12. O R^2 máximo de um modelo com todas as variáveis (observadas mais as não observadas) não é conhecido; assim, deve-se atribuir valores ao mesmo. No caso deste trabalho serão utilizados 0,6; 0,7; 0,8; 0,9 e 1,0 (este último valor refere-se a uma situação extrema).

Tabela 12. Efeito do PSF sobre a saúde das crianças, reportada pelos pais, da área rural da região Norte, com a inclusão de efeitos fixos municipais e saúde dos pais

		Norte			Nordeste		
		0 a 3	0 a 5	6 a 11	0 a 3	0 a 5	6 a 11
OLS	Coefficiente	0,082**	0,077***	0,045*	-0,030	-0,045	0,013
	Erro padrão	(0,037)	(0,030)	(0,027)	(0,035)	(0,029)	(0,222)

Notas: Estimções através do método OLS, considerando como controles, além das variáveis da Tabela 1, saúde do pai e saúde mãe (*dummy* com 1 indicando saúde muito boa e boa) e, em vez de efeitos fixos de estados, efeitos fixos de municípios.

Fonte: Cálculos do autor com base nos dados da PNAD (2008).

Tabela 13. Abordagem de Oster (2015) com base nos modelos da Tabela 12

R máximo	0,6	0,7	0,8	0,9	1,00
Norte 0 a 3					
δ para $\beta = 0$	4,05	2,77	2,11	1,70	1,42
Conj. Id. ($\delta = 1$)	[0,065; 0,082]	[0,055; 0,082]	[0,045; 0,082]	[0,035; 0,082]	[0,030; 0,082]
Norte 0 a 5					
δ para $\beta = 0$	4,31	3,18	2,52	2,08	1,77
Conj. Id. ($\delta = 1$)	[0,061; 0,077]	[0,055; 0,077]	[0,048; 0,077]	[0,042; 0,077]	[0,035; 0,077]
Norte 6 a 11					
δ para $\beta = 0$	1,28	0,99	0,81	0,68	0,59
Conj. Id. ($\delta = 1$)	[0,010; 0,045]	[0,000; 0,045]	[-0,011; 0,045]	[-0,021; 0,045]	[-0,032; 0,045]

Notas: Nesta abordagem foi utilizado o comando `psacalc` do STATA criado pela própria Oster.

Fonte: Cálculos do autor com base nos dados da PNAD (2008).

Observa-se que os resultados para as duas menores faixas de idade (0 a 3 e 0 a 5) são robustos a esse teste, apesar de os resultados para as crianças entre 6 e 11 anos não apresentarem essa robustez ($\delta < 1$). Por exemplo, no caso das crianças de 0 a 5 anos, o valor de δ igual a 1,77 significa que fatores não observáveis teriam que ser 1,77 mais forte do que o efeito dos fatores observáveis para explicar uma ausência de efeito do PSF. Seguindo uma ideia do artigo de Altonji *et al.* (2005), Oster (2015) indica que o valor de δ igual a 1 é um *cutoff* heurístico, no qual fatores observáveis seriam ao menos tão importantes quanto os não observáveis. Para esse mesmo grupo (crianças de 0 a 5 anos), o limite inferior da estimativa do ATT (para $\delta = 1$) é igual a 3,5 p.p.

A Tabela 14 apresenta a abordagem de Oster (2015) com base nos modelos OLS da Tabela 8, ou seja, explorando resultados das crianças entre 0 e 1 ano, considerando saúde reportada, diarreia e acamados. Observa-se que todos os valores de δ são menores que 0; isso indica que a inclusão de controles nos modelos, em vez de mover o efeito (beta) em direção a zero, faz o sentido oposto e, desta forma, as não observáveis teriam que possuir uma relação diferente dessa, para explicar ausência de efeito do tratamento. Este fato significa, ainda, que os controles têm correlações opostas com o tratamento e a variável de resultado. Ademais, percebe-se valores elevados de δ (desconsiderando o sinal negativo) para os casos diarreia e saúde reportada.

Tabela 14. Abordagem de Oster (2015) com base nos modelos da Tabela 9 (OLS)

Diarreia ou Vômito					
δ para $\beta = 0$	-3,70	-3,19	-2,80	-2,49	-2,24
Conj. Id. ($\delta = 1$)	[-0,090; -0,079]	[-0,098; -0,079]	[-0,102; -0,079]	[-0,105; -0,079]	[-0,108; -0,079]
Acamados					
δ para $\beta = 0$	-0,51	-0,44	-0,38	-0,33	-0,30
Conj. Id. ($\delta = 1$)	[-0,123; -0,049]	[-0,137; -0,049]	[-0,152; -0,049]	[-0,166; -0,049]	[-0,181; -0,049]
Saúde Autoavaliada					
δ para $\beta = 0$	-12,49	-10,60	-9,21	-8,14	-7,30
Conj. Id. ($\delta = 1$)	[0,136; 0,145]	[0,136; 0,147]	[0,136; 0,149]	[0,136; 0,151]	[0,136; 0,153]

Notas: Nesta abordagem foi utilizado o comando `psacalc` do STATA criado pela própria Oster.

Fonte: Cálculos do autor com base nos dados da PNAD (2008).

4. Considerações finais

Este trabalho buscou analisar o impacto do Programa Saúde da Família sobre a saúde das crianças (menores de 12 anos) da área rural do Brasil com base nos dados da PNAD (2008), utilizando os métodos do *Propensity Score Matching* (PSM) e dos Mínimos Quadrados Ordinários (OLS), aplicando, respectivamente, análises de sensibilidade proposta por Ichino *et al.* (2008) e Oster (2015).

A análise principal deste trabalho se baseou na saúde reportada. Foi observado que o PSF tem um impacto sobre a saúde das crianças da área rural do Brasil estatisticamente significativo e de magnitude elevada. Contudo, esse impacto está concentrado nas regiões Norte e Nordeste, sendo não significativo no Centro-Sul do País. No caso da região Norte, esse impacto independe da idade, enquanto na região Nordeste ele é significativo para as menores faixas de idade, principalmente entre as crianças de 0 a 3 anos de idade (menores de 4 anos).

Quando analisados outros indicadores de saúde, prevalência de diarreia e de acamados, os resultados se mostraram significativos apenas para região Norte e nas menores faixas de idade. O impacto também tem magnitude elevada, principalmente entre as crianças com idade entre 0 e 1 ano (menores de 2 anos de idade).

A abordagem de sensibilidade proposta por Ichino *et al.* (2008) indicou que os resultados da saúde reportada para região Norte são todos robustos, independente dos grupos de idade. Para região Nordeste, os resultados que apresentaram maior robustez foram da faixa de idade de 0 a 3 anos. Já na abordagem de Oster (2015), apenas os resultados da região Norte e das crianças dos grupos mais novos foram robustos; por exemplo, para as crianças menores de 2 anos, esse programa tem impacto na saúde reportada, na prevalência de diarreia e de acamados. Portanto, há a confirmação da hipótese de que os indivíduos das regiões mais pobres da área rural são beneficiados por esse tipo de política.

Em linhas gerais, os resultados apresentados buscaram avançar em relação à literatura

existente em duas direções. A primeira buscou apresentar uma análise dos impactos do PSF sobre a saúde das crianças na área rural, tendo em vista sua expressiva cobertura nessa área. A segunda residiu na estratégia de identificação, empregando dados que permitiram identificar se o domicílio que a criança reside possuía cadastro no PSF. Trata-se de uma estratégia diferente da utilizada por Reis (2014), que mensurou o tratamento através do programa no município, podendo ter incorrido em um erro, já que a variável não identificava exatamente se o domicílio possuía o programa.

Com relação à metodologia utilizada, com vistas a inferir causalidade, ficou evidenciado que o programa é mais eficaz com as crianças que residem nas regiões mais pobres do Brasil. Trata-se de um resultado importante na medida em que mostra a efetividade de uma política pública na área de saúde e diante de um quadro de escassez de recursos. Tal informação contribui para pensar em um contínuo aprimoramento da aplicação dos recursos públicos com vistas a um bem-estar social.

Contudo, deve-se destacar que esta análise contemplou apenas a disponibilidade de cadastro superior a um ano no programa. Porém, diferentes lugares que têm o programa podem conter estruturas diferentes, tanto em quantidade como em qualidade das equipes do Programa Saúde da Família. Portanto, principalmente com maior cobertura do PSF, análises que consigam contemplar esses aspectos podem ser importantes para pesquisas futuras.

5. Referências

- ABADIE, A. e IMBENS, G. W. *Matching on the Estimated Propensity Score*. NBER Working Paper: National Bureau of Economic Research, Inc., 2012.
- ALTONJI, J. G., ELDER, T. E. e TABER, C. Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. *Journal of Political Economy*, v. 113, n. 1, p. 151-184, 2005.
- AQUINO, R., DE OLIVEIRA, N. F. e BARRETO, M. L. Impact of the family health program on infant

- mortality in Brazilian municipalities. *American Journal of Public Health*, v. 99, n. 1, p. 87-93, 2009.
- BATTISTIN, E., DE NADAI, M. e SIANESI, B. Misreported schooling, multiple measures and returns to educational qualifications. *Journal of Econometrics*, v. 181, n. 2, p. 136-150, 2014.
- CALIENDO, M. e KOPEINIG, S. Some Practical Guidance for the Implementation of Propensity Score Matching. *Journal of Economic Surveys*, v. 22, n. 1, p. 31-72, 2008.
- GUANAIS, F. C. The combined effects of the expansion of primary health care and conditional cash transfers on infant mortality in Brazil, 1998-2010. *American Journal of Public Health*, v. 103, n. 11, p. 2000-2006, 2013.
- ICHINO, A., MEALLI, F. e NANNICINI, T. From temporary help jobs to permanent employment: what can we learn from matching estimators and their sensitivity? *Journal of Applied Econometrics*, v. 23, n. 3, p. 305-327, 2008.
- IMAI, K. e RATKOVIC, M. Covariate balancing propensity score. *Journal of the Royal Statistical Society, Series B* 76, p. 243-263, 2014.
- LEUVEN, E. e SIANESI, B. *PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing*. [S.l.]: Boston College Department of Economics, 2003.
- MACINKO, J. *et al.* Evaluation of the impact of the Family Health Program on infant mortality in Brazil, 1990-2002. *Journal of Epidemiology and Community Health*, v. 60, n. 1, p. 13-19, 2006.
- _____. *et al.* Going to scale with community-based primary care: an analysis of the family health program and infant mortality in Brazil, 1999-2004. *Social Science & Medicine*, v. 65, n. 10, p. 2070-2080, 2007.
- _____. e HARRIS, M. J. Brazil's Family Health Strategy — Delivering Community-Based Primary Care in a Universal Health System. *New England Journal of Medicine*, v. 372, n. 23, p. 2177-2181, 2015.
- NANNICINI, T. Simulation-based sensitivity analysis for matching estimators. *Stata Journal*, v. 7, n. 3, p. 334-350, 2007.
- NICOLELLA, A. C., KASSOUF, A. L. e BARROS, A. L. M. DE. O impacto do trabalho infantil no setor agrícola sobre a saúde. *Revista de Economia e Sociologia Rural*, v. 46, n. 3, p. 673-701, 2008.
- NORONHA, K. V. M. D. S. *A Relação entre o Estado de Saúde e a Desigualdade de Renda no Brasil*. 2005. 187 CEDEPLAR/FACE, UFMG, Minas Gerais.
- OSTER, E. *Unobservable Selection and Coefficient Stability: Theory and Evidence*. [S.l.]: Brown University, 2015.
- RASELLA, D., AQUINO, R. e BARRETO, M. L. Reducing childhood mortality from diarrhea and lower respiratory tract infections in Brazil. *Pediatrics*, v. 126, n. 3, p. e534-540, 2010.
- REIS, M. Public primary health care and children's health in Brazil: evidence from siblings. *Journal of Population Economics*, v. 27, n. 2, p. 421-445, 18 ago. 2014.
- ROCHA, R. e SOARES, R. R. Evaluating the impact of community-based health interventions: evidence from Brazil's Family Health Program. *Health Economics*, v. 19 Suppl, p. 126-158, 2010.
- ROSENBAUM, P. R. *Observational studies*. Springer-Verlag, 2002.
- ROSENBAUM, P. R. e RUBIN, D. B. Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome. *Journal of the Royal Statistical Society. Series B (Methodological)*, v. 45, n. 2, p. 212-218, 1983.
- ROSENBAUM, P. R. e RUBIN, D. B. The central role of the propensity score in observational studies for causal effects. *Biometrika*, v. 70, n. 1, p. 41-55, 1983.
- SANTOS, A. M. A., TEJADA, C. A. O. e EWERLING, F. Os determinantes socioeconômicos do estado de saúde das crianças do Brasil rural. *Revista de Economia e Sociologia Rural*, v. 50, n. 3, p. 473-492, 2012.
- SOARES, S. *et al.* Os impactos do benefício do programa Bolsa Família sobre a desigualdade e a pobreza. In: CASTRO, J. A.; MODESTO, L. (Orgs.). *Bolsa Família: avanços e desafios*. Vol. 2. Brasília: Ipea, 2010.

