

O IMPACTO DO PROGRAMA SAÚDE DA FAMÍLIA SOBRE A SAÚDE DAS CRIANÇAS DA ÁREA RURAL DO BRASIL

Anderson Moreira Aristides dos Santos*
Paulo de Andrade Jacinto**

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 do PSM e OLS, aplicando, respectivamente, análises de sensibilidade proposta por Ichino *et al.* (2008) e Oster (2015). 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 proposta por Ichino *et al.* (2008) indicou resultados robusto para o Norte em todas as faixas, enquanto para região Nordeste a robustez ocorreu para as crianças entre 0 e 3 anos de idade. Já na abordagem de Oster (2015), apenas os resultados da região Norte e das menores faixas de idade foram robustos. Portanto, há 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

Classificação JEL: C21; I10; I18

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 proposed by Ichino *et al.* (2008) and Oster (2015). The results indicates 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, but it is not significant in the Mid-South. For the North region, this impact is independent of age, while in the Northeast it is significant for younger age groups. The sensitivity analysis proposed by Ichino *et al.* (2008) reported robust results for the North in all age groups, while in the Northeast region the robustness only occurred to children between 0 and 3 years old. Based in Oster (2015) approach, only the results of the North region in the smaller age groups were robust. There is confirmation of the hypothesis that this type of policy benefits the individuals of the poorest regions in rural area.

Keywords: Children's health; Brazil's rural area; Treatment Effect; Family Health Program.

Área 11 - Economia Agrícola e do Meio Ambiente

* Doutorando em Economia pelo Programa de Pós-Graduação em Economia do Desenvolvimento da PUC/RS e professor da UFAL (Campus do Sertão).

** Professor Doutor no Programa de Pós-Graduação em Economia do Desenvolvimento da PUC/RS.

1) INTRODUÇÃO

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

A equipe do PSF deve ser composta, no mínimo, por um médico, um enfermeiro, um auxiliar de enfermagem e seis ou mais agentes comunitários de saúde, que ficam responsáveis pelo monitoramento de no máximo 4000 pessoas de uma determinada área. A atuação dessas equipes acontece em unidades básicas de saúde, nas residências e mobilizações na comunidade¹.

Assim, o PSF assume papel importante na atenção básica à saúde no país, sendo que no ano de 1996, este programa estava presente em 4,6% dos municípios do Brasil, em 2007, esse número aumentou para 92,1% (REIS, 2014). E ainda, em 1998 existiam 2000 equipes deste programa com 60.000 agentes comunitários de saúde que cobriam 4% da população do país, essas estatísticas aumentaram respectivamente para 39.000, 265.000 e 62% no ano de 2014 (MACINKO; HARRIS, 2015).

Esse papel chamou a atenção para estudos com intuito de avaliar o seu impacto sobre a taxa de mortalidade infantil no Brasil. Entre eles vale mencionar o de Macinko *et al.* (2006) que utilizaram um painel de dados com efeitos fixos e estados do Brasil, no período 1990-2002. Os autores encontraram que um aumento de 10% na cobertura do PSF esteve associado a uma queda de 4,5% na taxa de mortalidade infantil.

Outro estudo semelhante empregando o mesmo método foi realizado por Macinko *et al.* (2007), porém, a unidade de análise foram as microrregiões do Brasil, no período 1999-2004. Esse trabalho indicou que um aumento de 10% na cobertura do PSF esteve associado a reduções de 0,45% na taxa de mortalidade infantil, 0,6% na mortalidade pós-neonatal e 1% na mortalidade por diarreia. E ainda, este programa não apresentou associação significativa com a taxa de mortalidade neonatal.

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, através de uma base de dados com municípios do Brasil, no período entre 1996 e 2004, mostrando que uma maior cobertura deste programa reduz a taxa de mortalidade infantil. Utilizando essa mesma técnica, sendo que no período de 2000 a 2005, Rasella *et al.* (2010) indicaram que o PSF esteve associado a reduções da mortalidade por diarreia e infecções do trato respiratório.

O estudo de Rocha e Soares (2010) analisou o impacto do PSF sobre as taxas 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 que 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, todas no período 1991-2004. Os principais resultados mostraram que o PSF foi significativamente associado a reduções na mortalidade, sendo que os efeitos foram maiores nas regiões mais pobres (Norte e Nordeste), sobre a mortalidade das crianças, e, das causas perinatais, infecciosas, endócrinas e de doenças respiratórias. E ainda, que este programa apresentou um impacto significativo sobre o aumento na oferta de trabalho, nas taxas de matrículas e nas reduções da fertilidade.

¹ Para mais detalhes ver http://dab.saude.gov.br/portaldab/ape_esf.php.

Especificamente para os municípios da região Nordeste, Duarte e Mesquita (2012) analisaram o impacto do PSF sobre a taxa de mortalidade infantil considerando os graus de cobertura, no período entre 2005 e 2007, utilizando a metodologia de dados em painel e *propensity score* generalizado. As autoras encontram que este programa apresentou um impacto negativo estatisticamente significativo sobre a taxa de mortalidade infantil.

Um trabalho mais recente de 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), através de um painel de efeitos fixos com os municípios do Brasil, no período entre 1998 e 2010. Este artigo mostrou que a associação entre PSF e a PNTMI é mais forte quando a cobertura do PBF aumenta.

Os estudos citados acima, em suas análises mensuram saúde por meio de diferentes taxas de mortalidade, ou seja, fizeram o emprego de dados agregados. O uso de microdados em nível de indivíduos foi realizado no trabalho de Reis (2014) que mensurou o impacto do PSF sobre a saúde das crianças a partir dos dados da PNAD de 2003. A saúde foi mensurada por uma medida subjetiva (saúde reportada) e algumas medidas mais objetivas. A estratégia empírica adotada pelo autor foi comparar a saúde da criança da qual o PSF estava presente no município de sua residência, quando a mesma nasceu (ou 6 meses antes- período pré-natal), mas, que considerando mesmo período de vida, não estava disponível para seu irmão mais velho, considerando indivíduos entre 12 e 84 meses. Os principais resultados do trabalho deste autor indicaram que crianças que viviam em regiões onde o PSF esteve 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 verificaram os fatores socioeconômicos associados à saúde individual (reportada pelos pais) das crianças menores de 6 anos de idade no Brasil rural, com dados da PNAD (2008). A partir da estimação de modelos probit, os autores encontraram que os efeitos marginais do programa 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 acima, 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 (SANTOS *et al.*, 2012; REIS, 2014). E ainda, apesar de em termos teóricos esperar-se que principalmente os indivíduos mais pobres da área rural tenham se beneficiado devido a implantação desse programa, tanto devido à restrição de acesso à serviços de saúde condicionada a questões monetárias, como por eles poderem estar longe da infraestrutura pública de saúde, observa-se que não há estudos que analisam o impacto do PSF sobre saúde das crianças dessa área.

Com base nessas considerações, o objetivo geral deste trabalho é 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), para tanto será utilizado o método de *Propensity Score Matching* (PSM) e dos Mínimos Quadrados Ordinários (OLS), aplicando testes de robustez para os resultados.

Desta forma, este estudo foca na área rural, além de que os dados permitem identificar se o domicílio da criança possuía cadastro no PSF, diferente do artigo do Reis (2014) que fez sua análise para o Brasil como um todo, e dada a disponibilidade de seus dados, mensurou o tratamento através da presença do programa no município, variável que não identifica exatamente se o domicílio possuía cobertura do programa.

Já o estudo de Santos *et al.* (2012) analisou apenas uma associação entre fatores socioeconômicos (dentre eles a do PSF) sobre a saúde das crianças da área rural do Brasil, além de focarem em apenas uma faixa de idade, menores de 6 anos. Sendo que, o presente trabalho visa analisar o impacto desse programa, buscando inferir causalidade, utilizando uma

metodologia que tenta controlar a não aleatoriedade do programa. Embora o controle seja realizado por meio dos fatores observáveis, o mesmo permite aplicar testes de robustez para analisar o quanto o resultado poderia ser afetado por outros fatores que são não observáveis. Com intuito de explorar as possíveis diferenças dos efeitos do PSF, a análise será realizada considerando diferentes faixas de idade das crianças nas regiões do Brasil rural.

Este trabalho está dividido em 4 seções. A segunda seção apresenta a base de dados bem como a estratégia empírica. A terceira seção contém a análise dos resultados. Por fim, a quarta seção faz algumas considerações finais.

2) DADOS E ESTRATÉGIA EMPÍRICA

A fonte de dados que será utilizada neste artigo é a PNAD do ano de 2008. 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. No ano de 2008, o tema saúde foi abordado, apresentando, entre outras questões, uma pergunta que identifica se o domicílio onde o indivíduo reside tem cadastro no Programa Saúde da Família².

A amostra deste trabalho se restringirá as crianças que possuíam até 11 anos de idade (menores de 12 anos), filhos da pessoa de referência da família e residentes na área rural do Brasil (classificação do IBGE do dicionário da PNAD). Todos os cálculos serão 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 serão exploradas para a faixa de idade de 0 a 1 ano (menores de 2 anos). Todos esses cálculos serão realizados para as regiões desse país.

Crianças que viviam em domicílios que tinham cadastro no PSF há pelo um ano serão identificadas como beneficiárias dessa política (tratados), tendo como grupo de comparação as crianças de domicílios que não tinham cadastro nesse programa (grupo de controle).³ Neste caso, crianças com cadastro de menos de 1 ano serão excluídas, tentando garantir que os possíveis benefícios do programa tenham tido tempo necessário para ocorrerem⁴.

Já medidas de saúde das crianças serão mensuradas por: *i*) saúde reportada pelos pais (*dummy* com 1 indicando saúde muito boa, e 0 boa, regular, ruim e muito ruim)⁵; *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 à assistência médica pelo motivo de doença. As duas últimas semanas (anteriores à pesquisa da PNAD) foram o período de referência dos itens *ii* ao *v*.

As análises 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

² 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.

³ 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.

⁴ 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 grande maioria (88,68%) dos cadastrados estava na faixa de cadastro de um ano ou mais.

⁵ Essa classificação foi seguida em outros trabalhos na literatura (NICOLELLA; KASSOUF; BARROS, 2008; SANTOS; TEJADA; EWERLING, 2012), diferente da saúde dos adultos, que em geral é codificada como muito boa e boa comparativamente as demais categorias. Isso se deve entre outros fatores, ao fato de as crianças estarem classificadas em sua ampla maioria nas categorias muito boa e boa.

(IDLER; BENYAMINI, 1997; NORONHA, 2005). Em relação as 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 (2 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, com intuito de trazer mais robustez, alguns resultados com base nessas variáveis também serão explorados.

Na avaliação do impacto do Programa Saúde da Família sobre a saúde, deve ser levado em consideração a possibilidade de um viés de seleção, ou em outras palavras, as crianças que apresentam indicadores socioeconômicos mais precários, e, assim, com maiores possibilidades de possuírem piores níveis de saúde, podem ter prioridade no atendimento do programa. Com base neste argumento, a metodologia utilizada neste trabalho se baseará numa tentativa de eliminação (ou mesmo redução) desse viés através do método *propensity score matching*⁶.

Neste caso, considere um indivíduo i , uma variável de interesse Y (por exemplo, saúde autoavaliada). Existem os indivíduos que participam do programa (tratados, situação $T=1$) 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, necessita-se observar indivíduos antes e depois do tratamento (participar do PSF). Dado que na ampla maioria de estudos de ciências sociais não há como se ter essa situação, deve-se obter um grupo de controle. Desta forma, uma possível saída seria utilizar os indivíduos que não participam do programa. Contudo, haveria um viés que ocorre quando a análise do impacto de um determinado programa leva em consideração apenas o fato de um indivíduo participar ou não do programa. 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 das soluções mais conhecida, feita através de dados observáveis, e, proposta primeiramente pelos autores Rosenbaum e Rubin (1983), refere-se ao *propensity score matching*. A ideia básica por trás desse método é que o viés seria pelo menos reduzido se a comparação for baseada entre os indivíduos tratados (que participam do programa) e os de controle (que não participam do programa) que sejam o mais similar possível.

Os escores de propensão neste caso seria a probabilidade de um indivíduo participar do programa a partir de suas características, dada pela seguinte expressão:

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

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 effect of treatment on the treated*) é apresentado na equação 4:

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

⁶ Para mais detalhes sobre este método ver, por exemplo, Caliendo e Kopeinig (2008).

Contudo, encontrar indivíduos com o mesmo escore de propensão é uma tarefa que não pode ser executável, considerando a natureza contínua da variável, contudo, existem formas de pareamento para contornar esse problema, constituindo assim o método do *propensity score matching*.

Portanto, a primeira etapa para utilização dessa técnica é a realização da estimação de uma regressão multivariada, afim de buscar a probabilidade de um indivíduo participar do programa (pertencer ao grupo dos tratados), onde no presente trabalho, modelos logit serão estimados. O quadro 1 apresenta as variáveis de controle que serão incorporadas nos modelos.

Obtidos os escores de propensão, a segunda etapa é realizar o pareamento, sendo que existem diferentes formas, o presente trabalho empregará as seguintes: *i*) o método do vizinho mais próximo, neste, para cada elemento do grupo dos tratados, n indivíduos do grupo de controle (os valores de n neste trabalho serão 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 0,01, ou seja, esse será o valor base para a distância máxima tolerável entre os escores de propensão dos grupos de comparação; *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 (*average effect of treatment on the treated*) será calculado, sendo que este e outros cálculos deste trabalho serão realizados através do *software* STATA. Ademais, estes métodos serão comparados a estimativas geradas pelo Método dos Mínimos Quadrados Ordinários (OLS), com um modelo OLS que permite interações entre o tratamento e as demais covariadas (FILM- *fully interacted linear matching*)⁷ e o método de regressão linear ajustada ponderada pelo inverso do *propensity score* (IPWRA).

Cabe destacar duas hipóteses fundamentais desse método. A primeira se refere a existência de suporte comum, a mesma estabelece que em cada grupo de controle deve existir um tratamento correspondente, onde temos que $0 < P(X) < 1$. A segunda, a hipótese de independência condicional (CIA-*conditional independence assumption*) 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; KOPEINIG, 2008).

Portanto, a omissão de uma variável importante provocaria um viés no pareamento, apesar disso, não há como mensurá-lo. Contudo, a literatura tem utilizado análises de sensibilidade, para buscar dar robustez aos resultados. O presente artigo aplicará o método sugerido por Ichino, Mealli e Nannicini (2008). A hipótese central para esse teste é que não há independência condicional entre a variável de resultado (*outcome*) e o tratamento dado as variáveis observadas (X). Sendo que, seguindo ideias de trabalhos anteriores⁸, 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:

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

⁷ Para uma aplicação deste método ver Battistin, Nadai e Sianesi (2014)

⁸ Ver Rosenbaum e Rubin (1983), Rosenbaum (2002) e Altonji *et al.* (2005).

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.

O passo seguinte desta análise se refere a caracterizar a distribuição de U considerando quatro parâmetros:

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

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

E assim, dado 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, desta

forma, U é tratado como mais uma variável observada, no cálculo do ATT pelo método do *propensity score matching*. Usando os valores dos parâmetros de sensibilidade, a estimação é repetida muitas vezes (no caso deste trabalho 500) e a estimativa do ATT é obtido através da média dos ATTs da distribuição de U .

No presente trabalho também será aplicada a abordagem sugerida por Oster (2015). Para uma breve explicação, considere novamente variáveis não observadas U . Uma das hipóteses centrais se refere a seleção proporcional (*proportional selection assumptions*) representado pela seguinte equação:

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

Onde $\sigma_{XT} = Cov(X, T)$; $\sigma_{UT} = Cov(U, T)$; $\sigma_X = Var(X)$; $\sigma_U = Var(U)$ 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).

Dada a hipótese de proporcionalidade, seleção igual 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:

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

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 $\hat{\beta}$, quando $\hat{\beta} = 0$, isto nos mostra o grau de seleção sobre as não observáveis relativamente as observáveis que seria suficiente para explicar ausência de impacto do tratamento.

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

Nestes modelos nos quais serão aplicados a abordagem de Oster (2015), além das variáveis do quadro 1, serão incorporadas a saúde da mãe e do pai e efeitos fixos de municípios ao invés de estados. Este último caso tenta controlar esses fatores fixos, como outras políticas públicas, que 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 PSF. Uma abordagem alternativa seria incluir indicadores municipais, no entanto, a PNAD não permite identificar essas áreas geográficas.

3) ANÁLISE DOS RESULTADOS

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

Antes de analisar os resultados do *propensity score matching*, a tabela 2 apresenta a médias de algumas variáveis deste trabalho. Nota-se que para o Brasil, as crianças cadastradas no PSF (há mais de um ano) possuem 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, 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.

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

	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).

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 anos⁹. Observa-se que possuir abastecimento de água e de esgoto adequado e maiores níveis de renda estão associados a menor probabilidade de participação nesse programa, confirmando que em geral os indivíduos mais pobres são os beneficiários deste tipo de política. E ainda, maiores níveis de escolaridade dos pais estão associados a maior probabilidade de ter esse tipo cadastro, isso pode estar relacionado ao argumento de que pais mais educados entendem melhor a importância de aspectos relacionados à saúde. Enquanto 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 relacionados 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 a ser 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, onde a tabela 4 explora estatísticas para dar sustentação a essa condição. Nota-se um bom ajuste após o pareamento com grandes quedas no Pseudo R²; nos valores do viés médio e mediano (alcançando valores baixos). Além de que o teste razão verossimilhança de significância conjunta apresentam-se não significativos, com exceção para a região Nordeste,

⁹ Deixa-se claro que resultados foram gerados para as demais faixas etárias, por questões de espaços não são apresentados aqui. No que se refere as regiões Sudeste, Sul e Centro-Oeste, os resultados deste trabalho apresentaram-se relativamente semelhantes, assim, o Centro-Sul irá ser analisado dessa forma agregada.

contudo, mesmo neste último caso, acredita-se que as reduções nas demais estatística, indicam que este método não deve ser descartado¹⁰.

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

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

Notas: Os modelos contêm todas as variáveis indicadas no quadro 1, que não estão todas apresentadas nessa tabela por questões de espaço.

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	
		Brasil (0 a 5)				Brasil (6 a 11)			
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	
		Norte (0 a 5)				Norte (6 a 11)			
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	
		Nordeste (0 a 5)				Nordeste (6 a 11)			
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	
		Centro-Sul (0 a 5)				Centro-Sul (6 a 11)			
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	

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

Notas: Geradas com base no pareamento kernel.

¹⁰ Gráficos foram gerados para analisar a área de suporte comum, onde houve resultados satisfatórios. Disponíveis aos interessados através de contato com os autores.

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. E ainda, a magnitude do efeito é maior para o grupo mais novo, além de que nele, esses resultados são mais robustos independentemente do método aplicado. Por exemplo, considerando o método do vizinho mais próximo (NN-1), nos tratados a proporção de crianças, desse grupo (0 a 3 anos de idade), com saúde muito boa é 8,8 pontos percentuais (p.p.) maior do que de seu grupo de controle, 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 de que 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.

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	Coef.	0,049***	0,026*	0,022*
	Erro Padrão	(0,019)	(0,015)	(0,013)
N-obs.		2959	4603	5671

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

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 5 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 quadro 1 para os indicadores utilizados no *propensity score*.

Considerando as possíveis diferenças entre as regiões do Brasil, a tabela 6 apresenta cálculos semelhantes aos da tabela anterior, sendo que neste caso para as regiões Norte e Nordeste. Nota-se que para essa primeira região, os resultados são todos significativos a 5 %, considerando 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 a muitas das

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.

Já o Nordeste apresenta um padrão relativamente semelhante ao nacional, onde o impacto do PSF é significativo nas duas menores faixas etárias, sendo que a magnitude é maior para as crianças entre 0 e 3 anos de idade, onde o método do kernel mostra que crianças dessa faixa que foram beneficiadas por esse programa possuíam uma média de saúde muito boa 10,5 p.p. maior do que de seu grupo de comparação. E os coeficientes do OLS se mostraram inferiores aos apresentados nos diferentes pareamentos (6,4 p.p. nessa menor faixa de idade e não significativo para faixa 0 a 5). Ademais, resultados foram explorados para as crianças da área rural do Centro-Sul do país, conforme pode ser visualizado na tabela 7, em geral, eles foram todos não significativos¹¹.

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

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

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

Resultados nesse sentido de importância do PSF nas regiões Norte e Nordeste foram observados em Santos *et al.* (2012) em sua análise de associação do PSF com a saúde infantil na área rural. Já Soares encontraram que este programa apresentou efeito significativo e de larga magnitude nessas duas regiões, onde, por exemplo, na região Norte, municípios que estavam no programa há 8 anos, observaram um impacto de forma a reduzir em 15 a taxa de mortalidade por mil crianças menores de 1 ano, no período 1995-2003. E ainda, estimações para as taxas de mortalidade nessa faixa etária não foram significativas para as regiões Sudeste e Sul, e para mortalidade entre 1 e 4 anos não foram significativas para a região Centro-Oeste.

¹¹ Como observado em Caliendo e Kopeinig (2008) variáveis afetadas pelo tratamento não devem ser incluídas no modelo de *propensity score*. Assim, a 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.

Desta forma, observa-se a importância do PSF, principalmente para as regiões mais pobres do país. Macinko e Harris (2015) defende que um aspecto de relevância desse programa se refere ao seu desenho, como o papel dos agentes comunitários de saúde, estes que ficam responsáveis por realizar visitas às famílias de uma determinada área independente da necessidade ou demanda, realizando atividades de promoção à saúde. Assim, no presente trabalho leva-se como hipótese que este é um fator importante principalmente na área rural do país.

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

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

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

Como explicado na seção anterior, este trabalho irá ter foco na saúde reportada, contudo, resultados serão explorados com base em medidas que apesar de serem mais objetivas, conseguem captar apenas um curto período de tempo. Dessa forma, foram mensurados os ATTs para as regiões Norte e Nordeste de acordo com as demais medidas, a tabela 8 apresenta as variáveis que em geral foram significativas, ou seja, restrições de atividades por diarreia ou vômito e acamados apenas para a região Norte¹². Observa-se que os resultados dessas duas medidas são significativos em nível de 10% para as crianças mais novas das 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 acamados no grupo mais novo dessa tabela o impacto é de 4,8 p.p. (com prevalência igual a 5%). E ainda, os coeficientes gerados pelo OLS são próximos aos de vários pareamentos.

¹² Alguns métodos indicaram significância estatística para o indicador de assistência à saúde devido a doenças.

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

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

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

Esses fatores analisados, nesta última tabela, podem impactar principalmente nos primeiros anos de vida das crianças, contudo, duas dificuldades surgem ao tentar fazer uma análise para grupos de idade ainda menores: quantidade de observações reduzidas e alguns sucessos (ou fracassos) perfeitos nos efeitos fixos de estado¹³. Dessa forma, com intuito de explorar alguns desses casos, este artigo utilizou alguns métodos de pareamento, considerando as variáveis do quadro 1, contudo, sem incluir efeitos fixos de estado, para crianças entre 0 e 1 ano (menores de 2 anos), essas estatísticas são apresentadas na tabela 9. Observa-se resultados significativos e de magnitudes largas. Por exemplo, para o OLS o fato de o domicílio da criança estar cadastrada no PSF aumenta em 7,9 p.p. a probabilidade de ter tido restrições de atividades pelo motivo diarreia ou vômito, onde esse número 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, já que no cálculo do ATT (NN-1), o grupo de controle possuía prevalência condicional de 9,10%, enquanto nos tratados essa estatística era de 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. Sendo que essas causas de mortes incluem problemas durante a gravidez, diarreia e outras doenças intestinais, além de gripe, asma e bronquite, e, que, espera-se que o apoio e informações fornecidas pela presença do PSF sejam 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, com o uso de microdados, Reis (2009) mostrou que a

¹³ Além disso, métodos em que os erros padrões seriam gerados via *bootstrapping* não foram apresentados, já que houve muitas falhas em seu processo de geração.

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.

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

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

Notas: Ver tabela 5 para detalhes das legendas e quadro 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.

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 saúde reportada. A calibragem de fatores não observados U é realizada “imitando” algumas variáveis de controles¹⁴. Esse teste não comporta variáveis contínuas, contudo, segundo os autores, não havendo perda de generalidade. Desta forma, essas foram transformadas em variáveis binárias, onde suas categorias estão representadas entre parênteses nessa tabela (quando a *dummy* assume valor 1). E ainda, esses cálculos estão baseados no método do vizinho mais próximo. Nota-se que os resultados são robustos, independentemente da faixa etária, sendo todos estatisticamente significativos a 10%. No caso das crianças de 0 a 3 anos, a estimativa do ATT tem uma queda bastante acentuada, contudo, esse impacto fica próximo de estimativas geradas pelos outros pareamentos ou mesmo do OLS. Por exemplo, no caso do da simulação com o Bolsa Família, o ATT é de 10,1 p.p., número próximo a 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. Entretanto, para as crianças de 0 a 5 anos, quando considerado a variável água na calibração, o ATT é estatisticamente não significativo.

¹⁴ 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								
No confunder	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								
No confunder	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								
No confunder	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

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

Notas: As primeiras colunas apresentam os valores p_{ij} utilizados para simular a variável *confunder* 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, onde SE é o erro padrão dessa estatística. O comando para STATA utilizado foi o `sensatt` proposto por Nannicini (2007).

Crianças que viviam 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 possuir 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 serão 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 confunder</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 Saúde da Mãe (Boa e Muito Boa)	0,41	0,38	0,32	0,30	1,204	1,504	0,111	0,060
	0,66	0,74	0,74	0,72	1,245	0,973	0,107	0,057
Nordeste- 0 a 5 anos								
<i>No confunder</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 Saúde da Mãe (Boa e Muito Boa)	0,41	0,38	0,31	0,33	0,959	1,318	0,081	0,048
	0,69	0,75	0,73	0,72	1,130	1,067	0,085	0,046

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

Notas: Ver legenda da tabela anterior.

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, independentemente 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 ainda 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 anteriores mostraram que o OLS se mostrou diferente de métodos que controlam de alguma forma efeitos heterogêneos do tratamento. E por fim, 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 totalmente extrema). Observa-se que os resultados para as duas menores faixas de idade (0 a 3 e 0 a 5) são robustas 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 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. E ainda, 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 desta, para explicar ausência de efeito do tratamento. Este fato significa ainda, que os controles possuem 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 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)

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

Notas: Estimções através do método OLS considerando como controles além das variáveis do quadro 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.

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]

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

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

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]

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

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

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, e, foi observado que o PSF possui 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 analisado outros indicadores de saúde, prevalência de diarreia e de acamados se mostraram significativos apenas para região Norte e nas menores faixas de idade. O impacto possui também 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 independentemente 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, onde quando considerado crianças menores de 2 anos, esse programa tem impacto na saúde reportada, na prevalência de diarreia e de acamados. Portanto, há confirmação da hipótese de que indivíduos das regiões mais pobres da área rural são beneficiados por esse tipo de política.

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

REFERÊNCIAS

- ABADIE, A.; 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.; 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.; 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.; SIANESI, B. Misreported schooling, multiple measures and returns to educational qualifications. *Journal of Econometrics*, v. 181, n. 2, p. 136–150, 2014.
- CALIENDO, M.; 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.; 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.
- LEUVEN, 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.
- MACINKO, J. *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.
- MACINKO, J.; 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.; 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.; 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.; 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.; 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.; 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. DOS; TEJADA, C. A. O.; 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: Jorge Abrahão de Castro; Lúcia Modesto. (Orgs.). *Bolsa Família: avanços e desafios*. vol. 2. Brasília: Ipea, 2010.