

ESTUDO TÉCNICO

N.º 08/2015

Avaliação de impacto dos beneficiários do Programa Bolsa Família matriculados no Pronatec Bolsa Formação: um estudo CASO-CONTROLE

MDS

SAGI

MINISTÉRIO DO DESENVOLVIMENTO SOCIAL E COMBATE À FOME

SECRETARIA DE AVALIAÇÃO E GESTÃO DA INFORMAÇÃO

Estudo Técnico

Nº 08/2015

Avaliação de impacto dos beneficiários do Programa Bolsa Família matriculados no Pronatec
Bolsa Formação: um estudo CASO-CONTROLE.

Técnico responsável

Júlio César Gomes Fonseca

Flávio Sousa da Vitória

Francisca de Fátima de Araújo Lucena

Yuri César Silva

Revisão

Paulo de Martino Jannuzzi

Estudos Técnicos SAGI é uma publicação da Secretaria de Avaliação e Gestão da Informação (SAGI) criada para sistematizar notas técnicas, estudos exploratórios, produtos e manuais técnicos, relatórios de consultoria e reflexões analíticas produzidas na secretaria, que tratam de temas de interesse específico do Ministério de Desenvolvimento Social e Combate à Fome (MDS) para subsidiar, direta ou indiretamente, o ciclo de diagnóstico, formulação, monitoramento e avaliação das suas políticas, programas e ações.

O principal público a que se destinam os Estudos são os técnicos e gestores das políticas e programas do MDS na esfera federal, estadual e municipal. Nesta perspectiva, são textos técnico-científicos aplicados com escopo e dimensão adequados à sua apropriação ao Ciclo de Políticas, caracterizando-se pela objetividade, foco específico e tempestividade de sua produção.

Futuramente, podem vir a se transformar em artigos para publicação na Cadernos de Estudos, Revista Brasileira de Monitoramento e Avaliação (RBMA) ou outra revista técnica-científica, para alcançar públicos mais abrangentes.

Palavras-chave: *Inclusão produtiva; avaliação de impacto; mercado de trabalho formal, estudo retrospectivo, estudo caso-controle, algoritmo.*

Unidade Responsável

Secretaria de Avaliação e Gestão da Informação

Esplanada dos Ministérios | Bloco A | Sala 307

CEP: 70.054-906 Brasília | DF

Fone: 61 2030-1501 | Fax: 2030-1529

www.mds.gov.br/sagi

Secretário de Avaliação e Gestão da Informação

Paulo de Martino Jannuzzi

Secretária Adjunta

Paula Montagner

APRESENTAÇÃO

O presente estudo técnico apresenta os resultados da aplicação do estudo do tipo caso-controle em registros administrativos do Cadastro Único para Programas Sociais (CadÚnico), a fim de analisar o impacto de políticas públicas. Para tanto, o presente texto divide-se nas seguintes partes: introdução, metodologia, seleção *caso-controle*, tratando casos órfãos, base de dados, perfis *caso-controle*, análise dos resultados e considerações finais.

1. Introdução

A avaliação de programas públicos é imprescindível para o desenvolvimento e a adaptação contínua das formas e instrumentos de ação pública (FREY, 2000). Nessa perspectiva, a avaliação sistemática dos programas públicos de qualificação profissional, em particular os realizados pelos estudos sobre a inserção dos beneficiários do Programa Bolsa Família (PBF) no mercado de trabalho formal, especificamente aqueles que recebem incentivos patrocinados pelo Programa Nacional de Acesso ao Ensino Técnico e Emprego (Pronatec), modalidade Bolsa Formação – Plano Brasil Sem Miséria (BSM) motivou o presente estudo, que apresenta uma aplicação da metodologia de estudos do tipo *Caso-Controle*.

Os estudos dessa natureza são amplamente utilizados na literatura epidemiológica clássica (KATZ, 2006), onde estudos clínicos com delineamento caso-controle são utilizados para validar os impactos de novos tratamentos em relação aos convencionalmente conhecidos e utilizados (SCHLESSELMAN, 1982). A natureza técnica dos estudos *caso-controle* justifica sua escolha e aplicação no presente estudo, dada sua legitimidade técnica em quantificar e mensurar as associações de interesse, entre as características avaliadas. A garantia de semelhança dos grupos considerados permite identificar impactos da ação proposta, por meio da verificação de mudanças na ocorrência do desfecho de interesse, mediante fatores estudados.

Neste contexto, os esforços tratados e apresentados no Caderno de Estudos “Inclusão Produtiva Urbana: o que fez o Pronatec/Bolsa Formação entre 2011 e 2014” representa o ponto de partida para a proposição da metodologia em questão. Os resultados apresentados no referido caderno delinearam inicialmente o grupo dos *casos* e os elementos de decisão para a escolha do grupo dos *controles*. A montagem dos grupos de controle ocorreu a partir da seleção aleatória, por meio de algoritmo posteriormente descrito, que permitiu parear, para cada indivíduo definido inicialmente como par, um indivíduo com características semelhantes,

o qual foi nominado controle. A escolha dos controles foi realizada a partir de um grupo de seis características observadas nos *casos*, as quais serão mencionadas oportunamente.

Posterior à aplicação dos critérios de escolha de casos e controles foi construída a base de dados, para a qual os primeiros levantamentos apresentam o perfil dos grupos de estudo (*caso e controle*), permitindo validar o algoritmo de seleção dos grupos de controle a partir da observação da similaridade entre as seis características dos indivíduos de ambos os grupos.

Em seguida, apresentam-se análises dos fatores que corroboram as hipóteses levantadas nos estudos anteriores, em relação aos impactos obtidos a partir do conjunto de iniciativas de cursos técnicos ofertados segundo demandas territoriais focalizadas, direcionadas aos beneficiários do PBF, junto às unidades participantes do Serviço Nacional de Aprendizagem, denominado Sistema “S” (SENAI, SENAC, SENAT e SENAR) e de redes públicas (Rede Federal de Educação Profissional, Científica e Tecnológica etc).

O presente estudo considerou inicialmente o total de 2.490.120 registros de indivíduos matriculados no Pronatec Bolsa Formação¹, para os quais foi utilizado o recorte de indivíduos com 18 a 64 anos encontrados no Cadastro Único do mês de abril de 2014. Esse recorte representou 61,75% do total inicial e é composto por 738.133 indivíduos apenas inscritos no CadÚnico e 799.397 beneficiários do PBF, sendo o total (1.537.530) desses dois grupos considerado os *casos* do presente estudo. A trajetória trilhada pelos egressos do sistema é o evento de interesse denominado sucesso no estudo, sendo que a trajetória dos *casos* será comparada à dos *controles* para avaliar o impacto das ações de formação.

¹ As fontes de dados que compõem a base inicial considerada neste estudo estão descritas no artigo “Contribuição do Programa Nacional de Acesso ao Ensino Técnico e Emprego – Bolsa Formação para formalização do emprego e dos microempreendedores individuais: análise dos participantes beneficiários do Bolsa Família e inscritos no Cadastro Único” do Caderno de Estudos número 24 do MDS/SAGI.

2. Metodologia

A metodologia utilizada para a seleção dos indivíduos do grupo *controle* foi definida a partir da ação inicial de marcação de todos os indivíduos que pertenciam à base do Cadastro Único do mês de abril de 2014 e não pertencessem ao grupo dos 1.537.530 indivíduos tratados no estudo anterior.

Como segundo critério, os *controles* deveriam possuir informações declaradas nas seis variáveis utilizadas como camadas de similaridade, com as seguintes categorias de informação, quando for o caso:

- Código do município;
- Código da microrregião;
- Sexo: feminino e masculino;
- Escolaridade: classificação elaborada segundo orientação do Departamento de Cadastro Único (DECAU), vinculado à Secretaria Nacional de Renda e Cidadania (SENARC).
 - Nunca frequentou escola (NF_ESC)
 - Ensino Fundamental Incompleto (EF_INC)
 - Ensino Fundamental Completo (EF_COM)
 - Ensino Médio Incompleto (EM_INC)
 - Ensino Médio Completo (EM_COM)
 - Ensino Superior Incompleto (ES_INC)
 - Ensino Superior Completo (ES_COM)
 - Sem Informação (SI_ESC)
- Faixa Etária: idade agrupada por faixas etárias.
 - 18 a 29 anos (18-29)
 - 30 a 49 anos (30-49)
 - 50 a 64 anos (50-64)
- Faixa de Atualização Cadastral: classificação elaborada visando estabelecer a sincronicidade das informações fornecidas pelos usuários do Cadastro Único. Observa-se que mesmo com todos os esforços para a redução do tempo de atualização das informações, ainda há bastante trabalho de conscientização a ser feito no sentido de se ter um cadastro com menor tempo de atualização de dados entre as informações coletadas, minimizando assim os riscos de discrepâncias. Sendo assim, as seguintes faixas de período de atualização foram consideradas.

- 0 a 12 meses (0a12)
- 13 a 24 meses (13a24)
- 25 a 36 meses (25a36)
- 37 a 48 meses (37a48)
- 49 ou mais (49mais)

3. Seleção Caso-Controle

Se houve informação disponível para as combinações possíveis, ao final, as seis camadas de similaridade teriam mais de 588 milhões de combinações para a seleção de controles. O processamento desse quantitativo levaria dias para ser realizado, caso não fosse estabelecido critério de seleção, em conjunto com um algoritmo para otimizar a mesma. Para o estabelecimento dos critérios de seleção foi definida uma estrutura de indexação que permitisse reconhecer a existência e disponibilidade de determinado indivíduo de *controle* na base de dados.

A necessidade de seleção dos *controles* conduziu à construção de um algoritmo otimizado que permitisse a indexação do conjunto de características definidos por camadas, possibilitando determinar quantos e quais indivíduos do conjunto já haviam sido selecionados. A regra básica seria fixar o município do caso e selecionar um indivíduo controle no mesmo município, obrigatoriamente, para “*parear*” com o caso. As demais características que definem as camadas precisavam também ser consideradas no processo de seleção dos controles, tratando-se da escolha aleatória, para cada caso, dos respectivos controles, sempre que os mesmos existissem em quantidade maior ou igual no conjunto de características das diversas variáveis constituintes das camadas. A combinação de camadas foi denominada *layer*, sendo utilizada como chave de indexação do indivíduo, o que possibilitou a marcação dos controles selecionados, evitando assim a duplicidade de seleção.

Observou-se no processo de determinação dos controles, várias combinações de casos onde não existiam controles suficientes na base de dados, representando assim combinações incompletas ou, simplesmente, inexistência total do perfil de caso dentro do universo dos controles disponíveis. Para exemplificar, considere a situação apresentada abaixo, na qual o *layer* do *caso* identificado na base de dados não possuía nenhuma ocorrência equivalente nas respectivas camadas de *controle*, ainda que houvesse 500 registros de casos disponíveis na base de dados de controle:

- **Código do Município:** 1100015

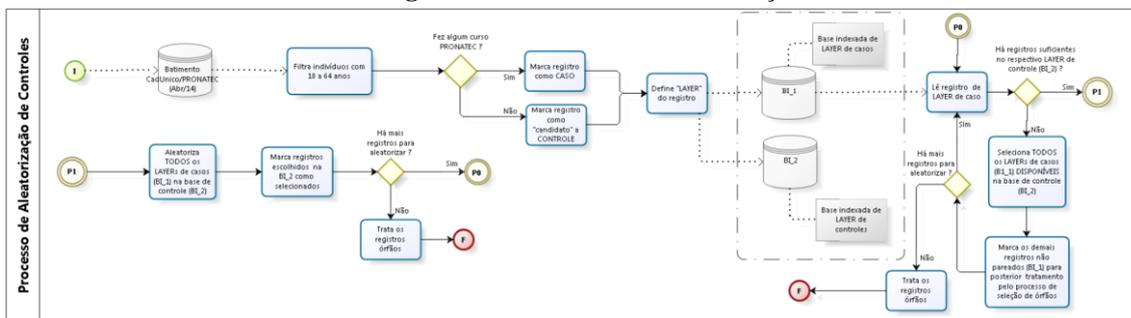
- **Código da MicroRegião:** 11006
- **Sexo:** F
- **Escolaridade:** EF_INC
- **Faixa Etária:** 50-64
- **Faixa de Atualização:** 49mais

Nesse caso, os 500 registros de *casos* passam a ser considerados *casos órfãos completos*. Situações como esta se tornam tanto mais comuns, quanto maior o número de combinações de características definidoras das camadas, aqui denominadas *layers*.

Outro cenário possível ocorreu quando um determinado quantitativo de *layers* existia para os *casos*, porém, sem registros suficientes nos *controles* disponíveis com as mesmas características. Os casos nessa situação foram denominados como *casos órfãos incompletos*.

O fluxograma 1 mostra, com detalhes, o processo de seleção de casos e controles. Os pontos sinalizados com a indicação “F” (destaque em vermelho) marcam o início do fluxograma de seleção de controle para casos órfãos tratados a seguir, constantes do Apêndice 1.

Fluxograma 1 – Processo de aleatorização dos controles.



4. Tratando Casos Órfãos

A ocorrência de órfãos conduziu à necessidade de criação de um algoritmo que fosse capaz de “adotar” tais casos particulares, de forma a manter as características mais próximas possíveis do órfão em questão, além de desconsiderar qualquer possibilidade de duplicidade de seleção. Assim, além da organização indexada, tornou-se crucial o desenho de um fluxograma que permitisse a compreensão de cada etapa de seleção, que apesar de ter seus critérios determinísticos nos momentos de adoção, mantinha ainda as características de seleção aleatória gerais, inerentes à proposta de se criar um estudo de caso-controle.

Destaca-se que a escolaridade foi tomada como ponto de partida, considerando-se a mudança nessa característica para o nível imediatamente acima. Por exemplo, se o caso para o qual se busca um controle tem ensino fundamental ou médio completo altera-se a escolaridade para

um nível acima. Nesse caso o critério para adoção do controle passa de ensino fundamental completo para ensino médio incompleto. Em outro exemplo, para adoção do controle cujo caso possui escolaridade ensino médio completo a camada ensino superior incompleto passa a ser considerada, conservando-se as demais características do *layer* intactas. Procede-se, então, a uma nova busca por controle para o caso em observação. Os demais casos de escolaridade só serão alterados em situações onde a hierarquia de seleção atingir seu último nível de procura dentro do município e ainda assim não for encontrado um controle disponível com as mesmas características.

A escolha do critério de escolaridade como ponto de partida do algoritmo foi motivada pelo fato do tema em questão estar fortemente vinculado a esta variável, por tratar-se de qualificação profissional, para a qual existem critérios de escolaridade mínima de acesso. Entende-se que, quanto mais fidedigno for o elemento pareado neste item, menor será o viés de seleção e, conseqüentemente, melhores serão os resultados analíticos.

Na avaliação dos casos órfãos, a cada “troca” de informação no *layer* era testado se haviam controles disponíveis e, em caso afirmativo, a busca era encerrada e o algoritmo seguia para o próximo caso². A etapa seguinte avaliou a faixa de atualização e, nessa camada, a troca de informações ocorria de forma hierárquica, sempre procurando por um controle que tinha atualizado suas informações no CadÚnico em um intervalo de tempo inferior ao caso considerado. Dessa forma, garantiu-se que a informação dada pelo controle fosse ainda mais atual do que a do próprio caso pareado³.

As demais etapas foram realizadas de forma sistemática, sempre retornando às características originais do caso cada vez que uma tentativa de se encontrar um controle com uma camada alterada é realizada. Este critério garante que o controle pareado tenha, no máximo, uma camada diferente do caso, aquela alterada para permitira a adoção.

A última etapa foi realizada apenas em casos muito raros⁴ onde nenhuma das tentativas anteriores conseguiu parear um caso, sendo realizada a troca do município como último recurso de localização e pareamento do indivíduo do grupo controle. Na tentativa de reduzir um pouco o viés desta seleção, que por si só acentua diferenças que vão além do nível individual, optou-se para que o município a ser escolhido aleatoriamente pertencesse à mesma microrregião, segundo classificação definida pelo IBGE.

² No fluxograma apresentado no quadro 1 o símbolo “F” indica fim da seleção de um controle, ou seja, atribui-se aquele controle ao caso sendo pareado, marca-se o controle como “*não disponível*” e passa-se a ler o próximo caso.

³ Salvo os casos em que a faixa de atualização era a mesma nos dois grupos.

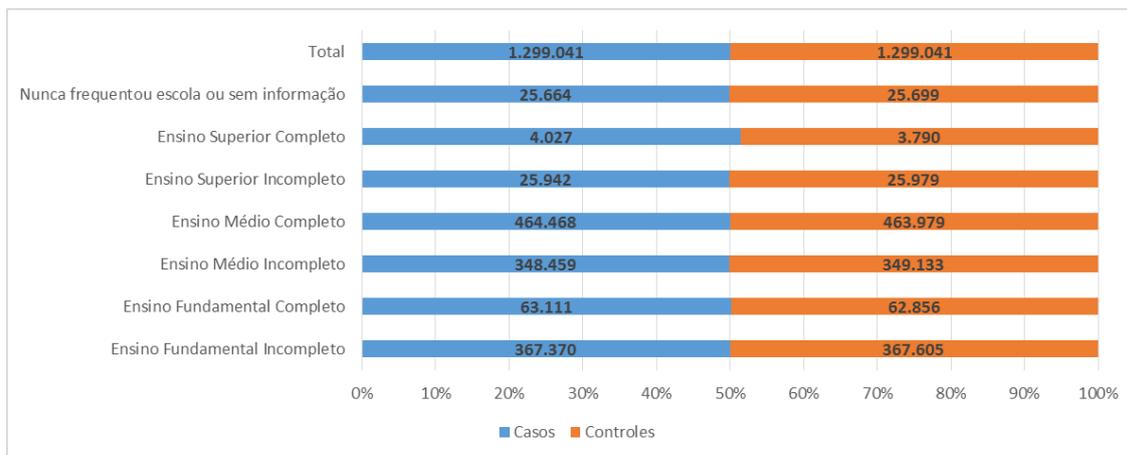
⁴ Apenas oito dos 1.299.041 casos do estudo apresentaram este cenário.

5. Base de dados

Após o pareamento dos casos e controles pelo método de aleatorização de seleção, com posterior verificação de ocorrências de órfãos, foi observada uma diferença entre o quantitativo de casos do estudo original (1.537.530) e os finais considerados no presente estudo. A diferença foi devida a ausência de informações nas seis camadas de informação utilizadas na metodologia de localização de controles, para os registros do CadÚnico. Dessa maneira, os 308.489 registros de casos que não possuíam as informações completas nas seis camadas no Cadastro Único foram excluídos do estudo, perfazendo ao final 1.229.041 casos, sendo aleatorizado o mesmo quantitativo de controles.

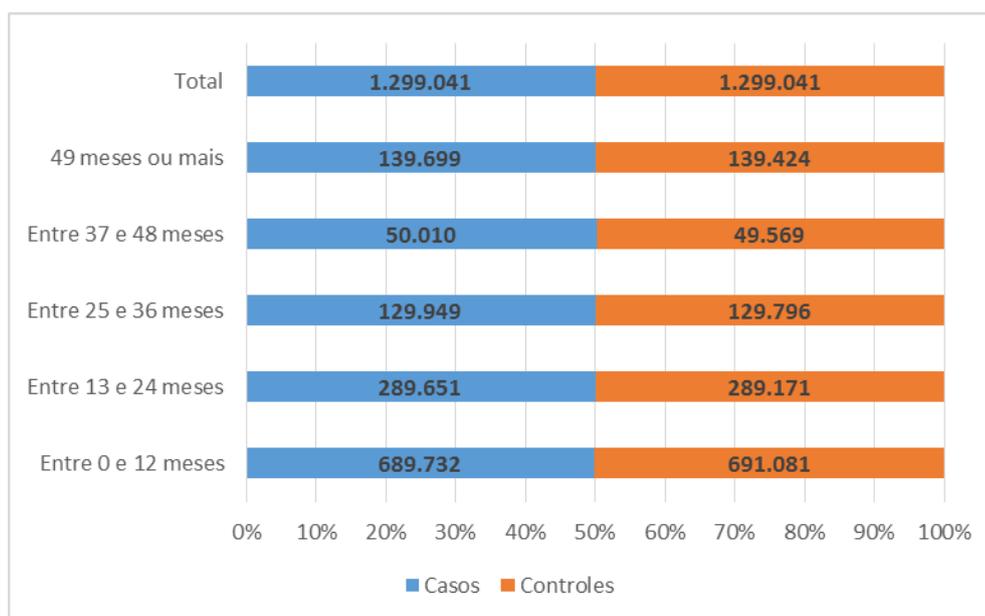
O Gráfico 1 apresenta os quantitativos selecionados para cada uma das faixas de escolaridade, onde se observa a semelhanças dos quantitativos em cada faixa. Observa-se a paridade da quantidade de registros nos grupos caso e controle, evidenciando o desenho balanceado do estudo.

Gráfico 1 – Distribuição da escolaridade entre os casos-controle.



Fonte: Base de dados final de estudo (Cadastro Único de Abril de 2014)

Gráfico 2 – Distribuição da faixa de atualização entre os casos-controle.



Fonte: Base de dados final de estudo (Cadastro Único de Abril de 2014)

A validação do método empregado foi realizada antes mesmo das primeiras análises de perfil, quando foi constatado que, dos 1.299.041 casos de estudo, 99,54% dos controles pareados foram aleatorizados sem a necessidade de aplicação do algoritmo de seleção de órfãos.

O fluxo do processo de construção da base está resumido abaixo e, embora não tenha a intenção de descrever os detalhes mais técnicos⁵ da criação das rotinas utilizadas, descreve o processo de forma geral, desde a determinação dos casos, até a entrega da base de dados final para estudo. Em termos gerais, o processo de criação da base de dados para análise do estudo em questão foi realizado nas seguintes etapas:

1. Seleção de indivíduos de 18 a 64 anos na base de dados utilizada na análise dos participantes beneficiários do Bolsa Família e inscritos no Cadastro Único. Total de *casos*: **1.537.530**.
Total de registros: **49.154.684**.
2. Normalização das variáveis utilizadas como camadas na criação dos *layers*, em geral agrupando as mesmas em faixas menores de observações.
3. Marcação dos indivíduos que fizeram matrícula em algum curso Pronatec, na base filtrada pela faixa etária especificada, como *casos*, e os demais registros como

⁵ Todas as rotinas de seleção de controles foram escritas em *Python* e utilizaram como insumo, procedimentos que vêm sendo testados pelo Departamento de Monitoramento como solução para pareamentos de grandes bases de dados, bem como para a indexação de arquivos texto.

controles. Nesta etapa foram excluídos os registros que não continham informações para todas as camadas e gerados os *layers* de casos e controles.

Total de *casos*: **1.299.041**.

Total de *controles*: **47.855.643**.

4. Estudo quantitativo dos *layers* gerados. Esta fase permitiu um mapeamento dos universos de casos órfãos, para os quais era necessária a criação do algoritmo de identificação e pareamento dos mesmos.
5. Definição dos critérios de seleção de registros e criação do fluxograma a ser utilizado como referência na criação das rotinas de seleção aleatória de controles.
6. Indexação dos registros a serem pareados por meio da criação de arquivos indexados cuja chave de acesso foram os *layers*. Os arquivos foram separados segundo as combinações das faixas de escolaridade e faixas de atualização. Nesta fase foram criadas estruturas que permitiriam o *bloqueio* de registros, na medida em que fossem sendo pareados.

Total de arquivos indexados com *casos*: **40**.

Total de arquivos indexados com *controles*: **40**.

7. Criação das rotinas que processariam os dados segundo os critérios de seleção definidos na etapa 5.
8. Tratamento de todos os *layers* não-órfãos. Nessa fase, os controles pareados foram marcados como *bloqueados*, assim como todos os casos órfãos.

Total de registros: **1.293.065**.

9. Tratamento de todos os *layers* órfãos. Nessa etapa foi concluído o restante da seleção de controles para o estudo.

Total de registros: **5.976**.

10. Tratamento de algumas datas para efeito de cálculo dos estudos realizados e normalização das variáveis da base de dados.

Total de registros: **2.598.082**.

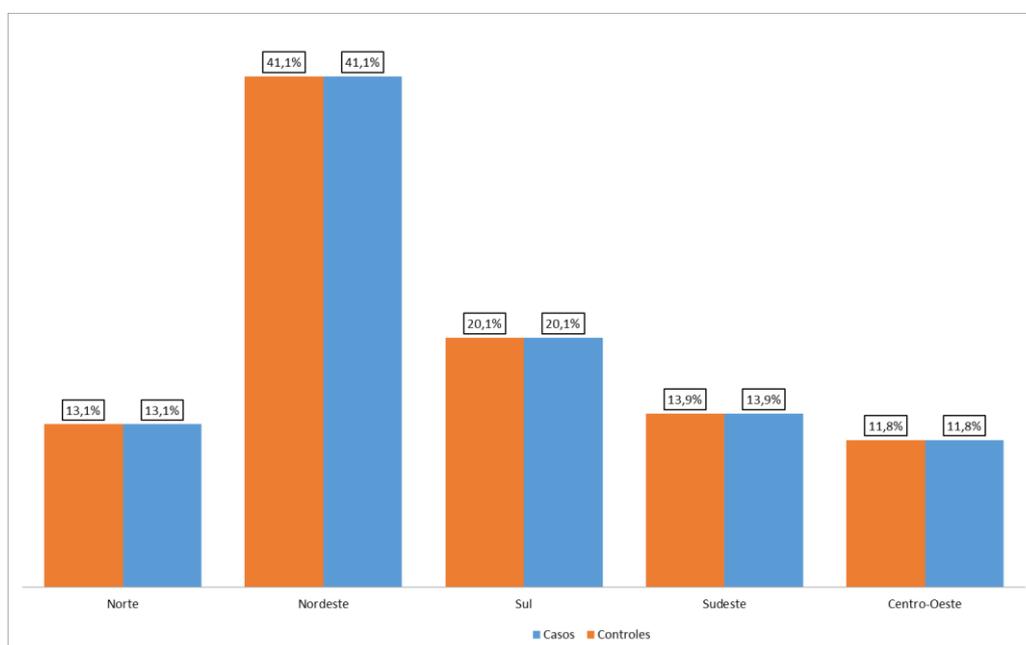
Para melhor compreensão dos passos de seleção dos órfãos o fluxograma completo está disponível no Apêndice 1.

A base de dados com indivíduos pareados foi integrada ao arquivo de estudo anterior (SOUSA et al. 2014), em que constavam os dados do SISTEC, RAIS 2012 e 2013, CAGED 2014.

6. Perfis Caso-Controle

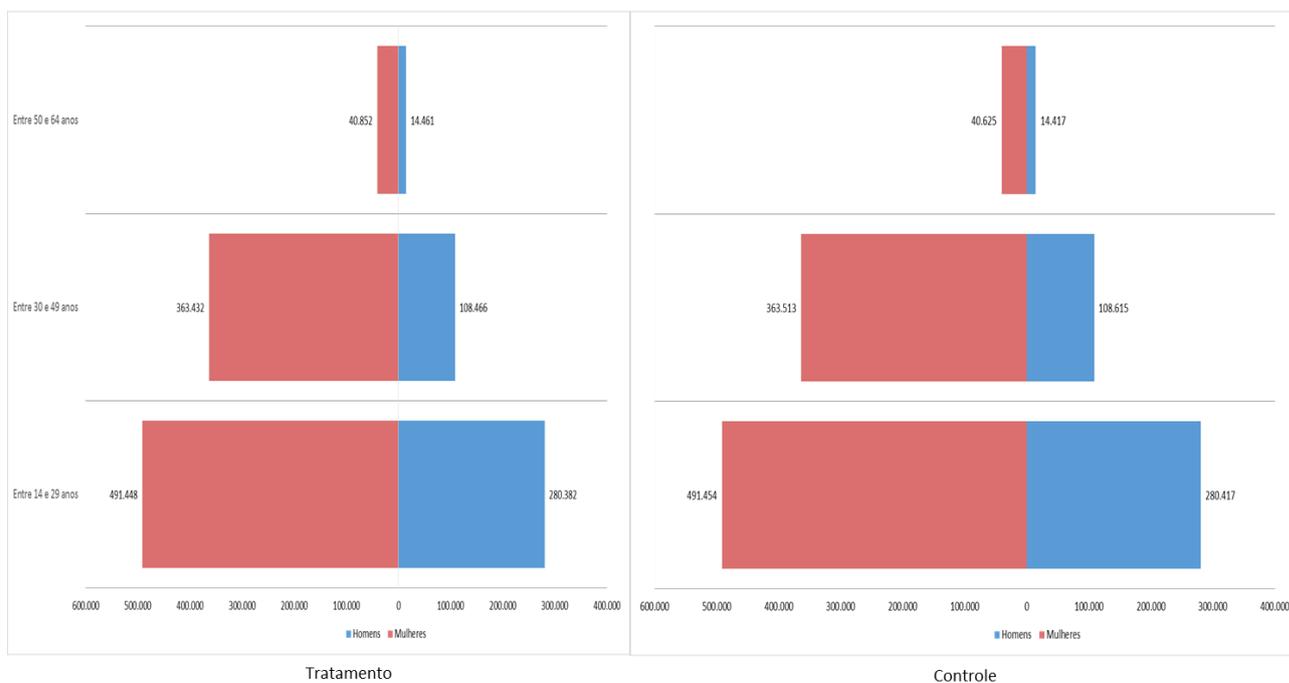
No processo de escolha do quantitativo de matrículas para o presente estudo optou-se pela paridade entre casos e controles, representando um desenho balanceado de estudo. Destaca-se a reprodução da ocorrência de paridade também nas categorias das variáveis observadas. No que se refere às comparações regionais observa-se a predominância no quantitativo de matrículas na região nordeste (41,1%) e sul (20,1%). As regiões norte e sudeste contam com pouco mais de 13% do total de matrículas no Pronatec em ambos os grupos. A região centro-oeste apresentou 11,8% do total de matrículas realizadas em 2014.

Gráfico 3 – Distribuição das matrículas segundo grupos e regiões. Brasil, 2014.



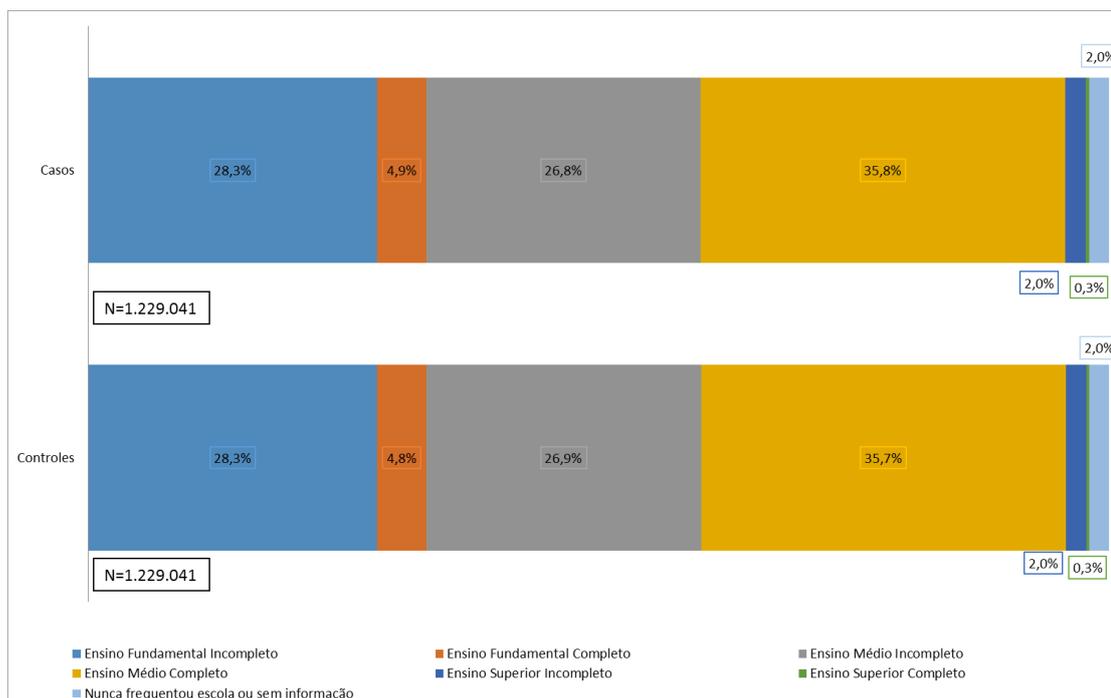
Considerando a distribuição etária dos alunos matriculados no Pronatec a pirâmide etária (Gráfico 4) mostra maior quantitativo de mulheres em todas as faixas etárias, em ambos os grupos (caso e controle). A faixa de 14 a 29 anos concentra ainda o maior quantitativo de matrículas, em ambos os sexos.

Gráfico 4 – Pirâmide etária dos alunos matriculados no Pronatec por grupo.
Brasil, 2014.



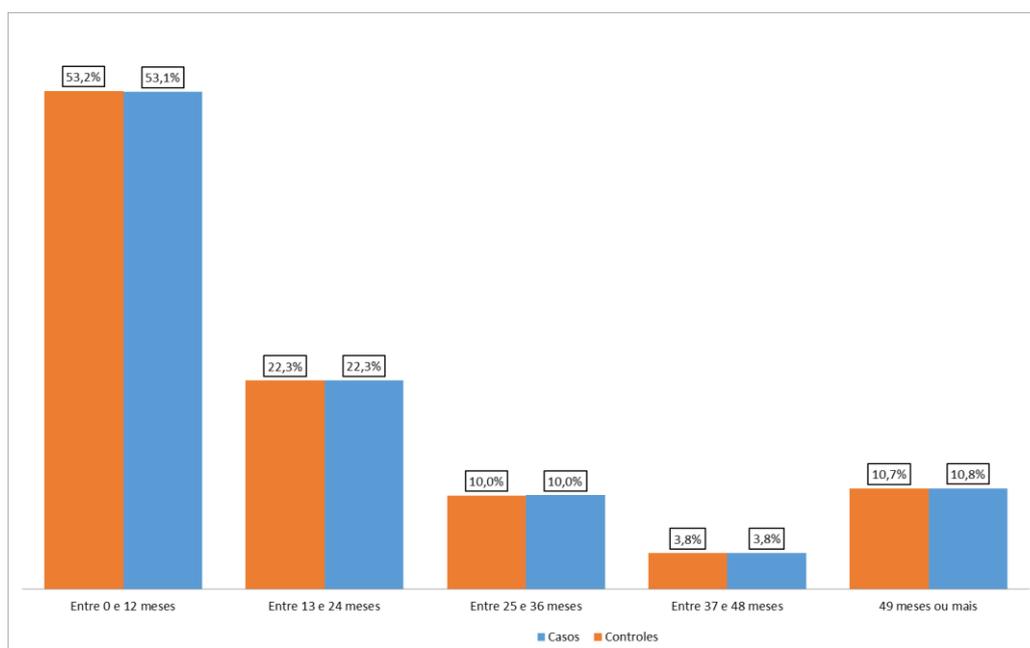
A distribuição da escolaridade dos alunos do Pronatec concentra-se nas fases iniciais da educação básica, tanto para o grupo de alunos classificados como casos, quanto para os controles, totalizando 60% de matrículas de alunos com escolaridade até ensino médio incompleto. Pouco mais de 35% dos alunos, de ambos os grupos, possuem ensino médio completo (Gráfico 5).

Gráfico 5 – Distribuição das matrículas segundo grupos e escolaridade. Brasil, 2014.



Ao destacar o aspecto da homogeneidade entre os indivíduos do grupo caso e controle a característica das faixas de tempo de atualização representa um aspecto importante a ser considerado na escolha dos controles do estudo. Dessa maneira, a maior parte dos alunos apresenta tempo de atualização dos dados do Cadastro Único de até 12 meses, ultrapassando ainda os 75% de alunos com tempo de atualização até 24 meses (Gráfico 6), que representa o prazo regulamentar para atualização cadastral, segundo regras do programa Bolsa Família.

Gráfico 6 – Distribuição das matrículas segundo grupos e faixas de atualização. Brasil, 2014.



7. Análise dos Resultados

A observação quanto à dinâmica dos vínculos de emprego formal dos indivíduos participantes foram apresentados em forma de resultados no presente estudo.

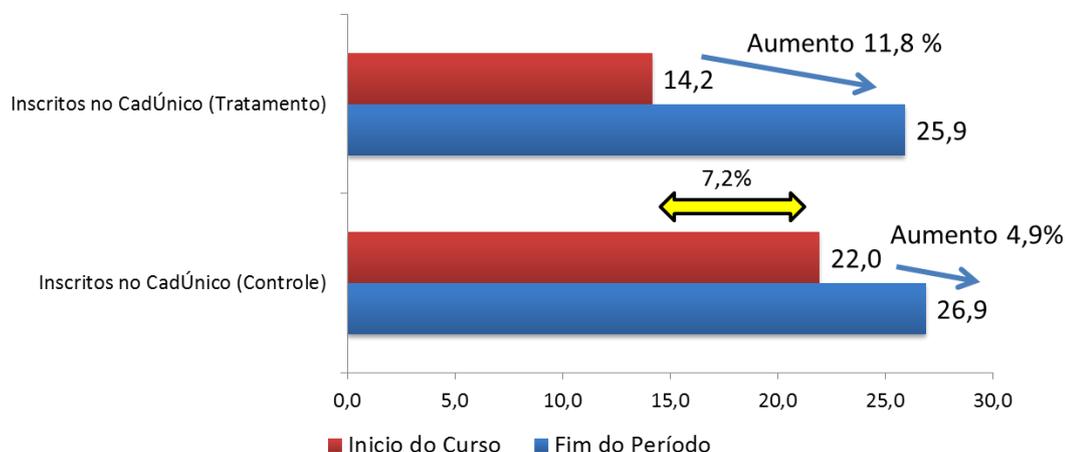
Os resultados mostram a mobilidade de vínculo empregatício formal dos indivíduos entre o início e o final do curso de formação do Pronatec para casos e controles. Observa-se que o percentual de indivíduos realizando curso do Pronatec (Tratamento ou Caso) e com vínculo formal no início do curso era de 14,2%, enquanto para o grupo que não estava cursando (Controle), o percentual era de 22% de inserção no mercado formal. Os resultados mostram ainda que, mesmo havendo maior inserção de indivíduos-Controle do que dos Casos no mercado formal no início do curso, o percentual inserido no mercado formal, ao final do período de observação, é bastante semelhante para Casos (25,9%) e Controles (25,9%). Fica evidenciado, dessa maneira, que a realização do curso do Pronatec beneficia quem o realiza, fazendo com que se equalize a taxa de formalização no mercado de trabalho entre os dois grupos (Tabela 1, Gráfico 7).

Tabela 1 – Distribuição dos indivíduos de 18 a 64 anos de idade registrados no CadÚnico segundo grupo de estudo e existência de vínculo empregatício formal ou formalização como microempreendedor individual no mês de início do curso e último mês do período estudado – Brasil, 2011 a 2014.

Status dos participantes	Participantes no setor formal					Participantes setor formal(%)		
	Total	Início do período	Fim do período	Varição absoluta	Varição relativa	Início do período	Fim do período	Varição p.p.
CadÚnico (Controle)	1.299.041	284.639	348.485	63.846	22,4	22,0%	26,9%	4,9%
CadÚnico (Tratamento)	1.299.041	183.684	336.261	152.577	83,1	14,2%	25,9%	11,8%

Nota: o “Início do período” refere-se ao início do curso para o grupo tratamento, o qual foi também atribuído ao grupo controle, a partir do caso pareado.

Gráfico 7 – Distribuição dos indivíduos de 18 a 64 anos de idade registrados no CadÚnico segundo grupo de estudo e existência de vínculo empregatício formal ou formalização como microempreendedor individual no mês de início do curso e último mês do período estudado – Brasil, 2011 a 2014.



Nota: o “Início do período” refere-se ao início do curso para o grupo tratamento, o qual foi também atribuído ao grupo controle, a partir do caso pareado.

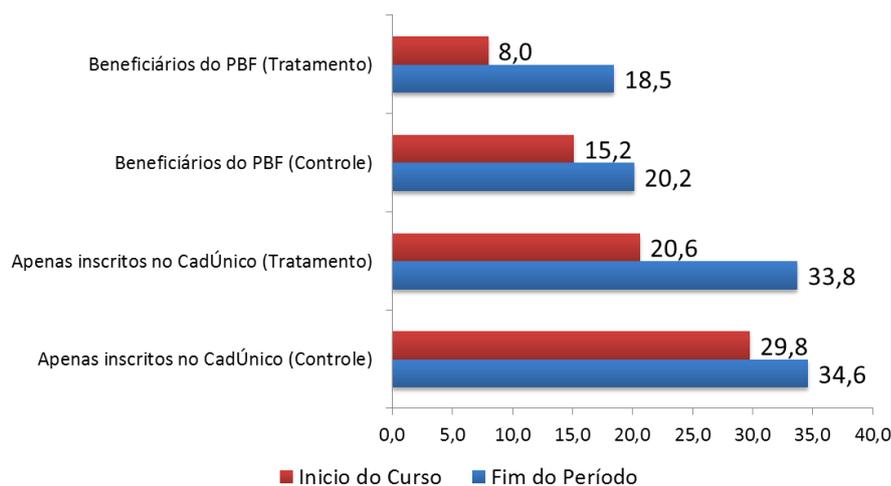
Quando se analisa os resultados para Beneficiários do Bolsa Família os efeitos do Pronatec parecem se potencializar. Tanto para os inscritos no CadÚnico (mas não beneficiário), quanto para o grupo de beneficiários do Programa Bolsa Família (PBF), as taxas de formalização são menores no início do período, para aqueles que fizeram o Pronatec (em relação ao grupo Controle). Ao final do período, contudo, o percentual de indivíduos com vínculo formal é semelhante entre os dois grupos, caso e controle, seja entre beneficiários ou não (Tabela 2 e Gráfico 8).

Tabela 2 – Distribuição dos indivíduos registrados no CadÚnico de 18 a 64 anos de idade segundo indicação de PBF, matrícula no Pronatec Bolsa Formação, grupo de estudo e existência de vínculo empregatício formal ou formalização como microempreendedor individual no mês de início do curso e último mês do período estudado – Brasil, 2011 a 2014.

Status dos participantes	Participantes no setor formal					Participantes setor formal(%)		
	Total	Início do período	Fim do período	Varição absoluta	Varição relativa	Início do período	Fim do período	Varição p.p.
PBF (Controle)	693.879	104.916	139.499	34.583	33,0	15,2%	20,2%	5,0
PBF (Tratamento)	665.034	53.280	122.782	69.502	130,4	8,0%	18,5%	10,5
Não-PBF (Controle)	605.162	179.723	208.986	29.263	16,3	29,8%	34,6%	4,8
Não-PBF (Tratamento)	634.007	130.404	213.479	83.075	63,7	20,6%	33,8%	13,1

Nota: o “Início do período” refere-se ao início do curso para o grupo tratamento, o qual foi também atribuído ao grupo controle, a partir do caso pareado.

Gráfico 8 – Distribuição dos indivíduos registrados no CadÚnico de 18 a 64 anos de idade segundo indicação de PBF, matrícula no Pronatec Bolsa Formação, grupo de estudo e existência de vínculo empregatício formal ou formalização como microempreendedor individual no mês de início do curso e último mês do período estudado – Brasil, 2011 a 2014.



Nota: o “Início do período” refere-se ao início do curso para o grupo tratamento, o qual foi também atribuído ao grupo controle, a partir do caso pareado.

Em todas as regiões se observa o efeito do Pronatec na “equalização” das taxas de formalização ao final do período, tal como descrito anteriormente. No Nordeste, região com maior contingente de beneficiários do Programa Bolsa Família, o efeito é ainda maior: o contingente de indivíduos que chegou ao final do período com vínculo formal dobrou entre os que realizaram curso do Pronatec na região (passou de 50 mil para 101 mil). Se o programa não tivesse chegado aos pequenos e médios municípios do Semiárido, seus efeitos de “equalização” seriam certamente menores.

Tabela 3 – Distribuição dos indivíduos de 18 a 64 anos de idade registrados no CadÚnico segundo região brasileira e grupo de estudo e existência de vínculo empregatício formal ou formalização como microempreendedor individual no mês de início do curso e último mês do período estudado – Brasil, 2011 a 2014.

Região	Status dos participantes	Participantes no setor formal					Participantes setor formal(%)		
		Total	Início do período	Fim do período	Varição absoluta	Varição relativa	Início do período	Fim do período	Varição p.p.
Norte	CadÚnico (Controle)	170.000	28.801	35.360	6.559	22,8	16,9%	20,8%	3,9
	CadÚnico (Tratamento)	170.000	18.153	32.589	14.436	79,5	10,7%	19,2%	8,5
Nordeste	CadÚnico (Controle)	533.827	94.951	115.749	20.798	21,9	17,8%	21,7%	3,9
	CadÚnico (Tratamento)	533.827	50.330	101.559	51.229	101,8	9,4%	19,0%	9,6
Sudeste	CadÚnico (Controle)	260.526	72.713	87.121	14.408	19,8	27,9%	33,4%	5,5
	CadÚnico (Tratamento)	260.526	46.979	85.852	38.873	82,7	18,0%	33,0%	14,9
Sul	CadÚnico (Controle)	181.088	52.164	63.846	11.682	22,4	28,8%	35,3%	6,5
	CadÚnico (Tratamento)	181.088	42.070	69.544	27.474	65,3	23,2%	38,4%	15,2
Centro-Oeste	CadÚnico (Controle)	153.600	36.010	46.409	10.399	28,9	23,4%	30,2%	6,8
	CadÚnico (Tratamento)	153.600	26.152	46.717	20.565	78,6	17,0%	30,4%	13,4

Nota: o “Início do período” refere-se ao início do curso para o grupo tratamento, o qual foi também atribuído ao grupo controle, a partir do caso pareado.

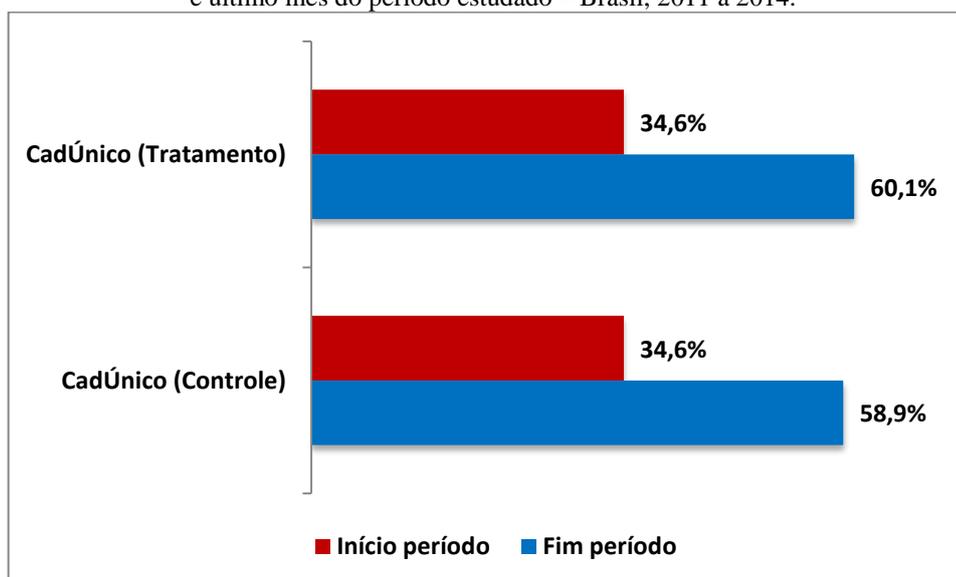
Na avaliação dos efeitos do Pronatec para os dois grupos (controle e tratamento) considerando-se indivíduos pareados já inseridos no mercado formal antes do início do curso/período, observa-se um efeito marginalmente superior para o grupo de participantes dos cursos do Pronatec (Tabela 5 e Gráfico 9). Observa-se que ambos os grupos, partindo com mesmo percentual de indivíduos no mercado formal no início do período (34,6%), chegam ao final do período com taxas de formalização maiores, mas ligeiramente maiores entre os participantes do Pronatec (60%) em relação aos não-participantes (58,9%).

Tabela 4 – Distribuição dos indivíduos registrados no CadÚnico de 18 a 64 anos de idade segundo grupo de estudo e existência de vínculo empregatício formal ou formalização como microempreendedor individual, para indivíduos pareados no início do período, já inseridos no mercado de trabalho formal, no mês de início do curso e último mês do período estudado – Brasil, 2011 a 2014.

Status dos participantes	Participantes no setor formal					Participantes setor formal(%)		
	Total	Início do período	Fim do período	Varição absoluta	Varição relativa	Início do período	Fim do período	Varição p.p.
CadÚnico (Controle)	167.964	58.056	98.853	40.797	70,3	34,6%	58,9%	24,3
CadÚnico (Tratamento)	167.964	58.056	100.867	42.811	73,7	34,6%	60,1%	25,5

Nota: o “Início do período” refere-se ao início do curso para o grupo tratamento, o qual foi também atribuído ao grupo controle, a partir do caso pareado.

Gráfico 9 – Distribuição dos indivíduos registrados no CadÚnico de 18 a 64 anos de idade segundo grupo de estudo e existência de vínculo empregatício formal ou formalização como microempreendedor individual, para indivíduos pareados no início do período, já inseridos no mercado de trabalho formal, no mês de início do curso e último mês do período estudado – Brasil, 2011 a 2014.



Nota: o “Início do período” refere-se ao início do curso para o grupo tratamento, o qual foi também atribuído ao grupo controle, a partir do caso pareado.

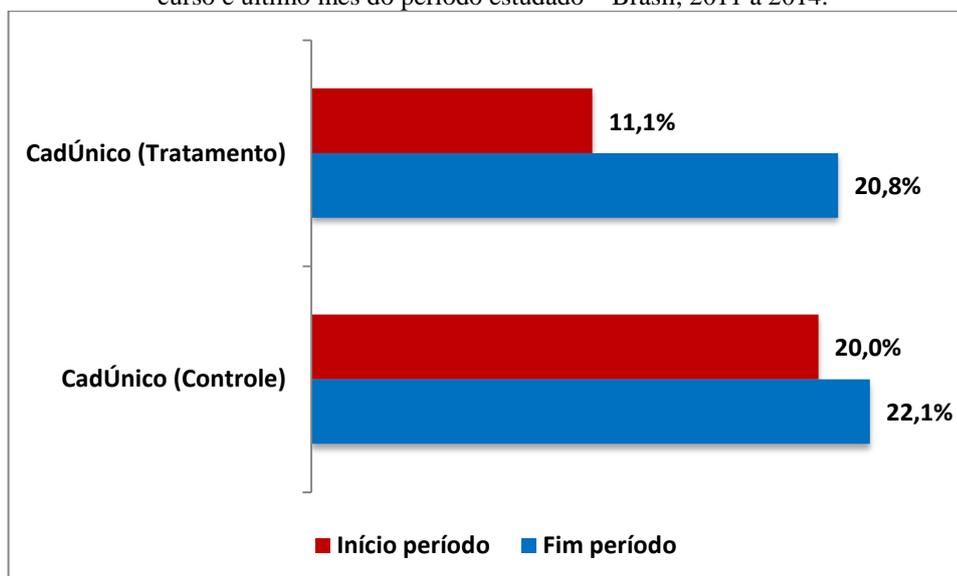
Considerando a perspectiva complementar de vinculação, ou seja, aqueles pares cujo caso-controle que não estavam ambos vinculados no início do período, o grupo de tratamento apresenta variação percentual superior ao grupo controle (Tabela 5). Enquanto os Controles apresentam variação relativa de 10,2%, no grupo Tratamento o percentual atinge 87,4% (Tabela 5 e Gráfico 10).

Tabela 5 – Distribuição dos indivíduos registrados no CadÚnico de 18 a 64 anos de idade segundo grupo de estudo e existência de vínculo empregatício formal ou formalização como microempreendedor individual, para indivíduos NÃO-pareados no início do período, já inseridos no mercado de trabalho formal, no mês de início do curso e último mês do período estudado – Brasil, 2011 a 2014.

Status dos participantes	Participantes no setor formal					Participantes setor formal(%)		
	Total	Início do período	Fim do período	Variação absoluta	Variação relativa	Início do período	Fim do período	Variação p.p.
CadÚnico (Controle)	1.131.077	226.583	249.632	23.049	10,2	20,0%	22,1%	2,0
CadÚnico (Tratamento)	1.131.077	125.628	235.394	109.766	87,4	11,1%	20,8%	9,7

Nota: o “Início do período” refere-se ao início do curso para o grupo tratamento, o qual foi também atribuído ao grupo controle, a partir do caso pareado.

Gráfico 10 – Distribuição dos indivíduos registrados no CadÚnico de 18 a 64 anos de idade segundo grupo de estudo e existência de vínculo empregatício formal ou formalização como microempreendedor individual, para indivíduos NÃO-pareados no início do período, já inseridos no mercado de trabalho formal, no mês de início do curso e último mês do período estudado – Brasil, 2011 a 2014.



Nota: o “Início do período” refere-se ao início do curso para o grupo tratamento, o qual foi também atribuído ao grupo controle, a partir do caso pareado.

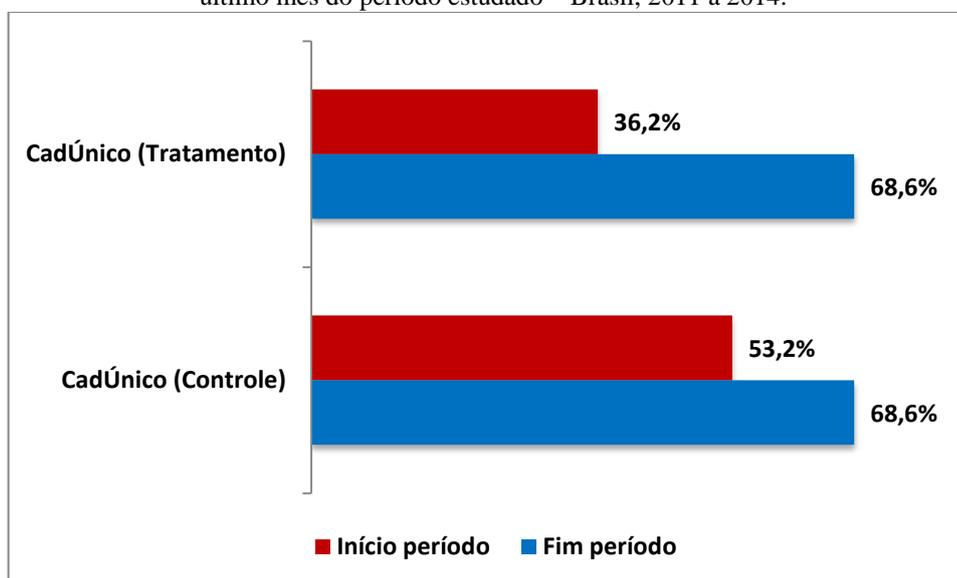
Avaliar o Pronatec considerando a perspectiva dos pares (caso-controle) inseridos no mercado formal ao final do período revela que, nesse caso, a evolução do quantitativo de inserção é maior para os indivíduos que fizeram curso do Pronatec (Tabela 6 e Gráfico 11). Apesar do baixo de inserção no início do período para contingente do grupo de casos, o grupo conseguiu inserção ao final do período, perfazendo uma variação relativa de quase 90%, contra os cerca de 30% do grupo controle.

Tabela 6 – Distribuição dos indivíduos registrados no CadÚnico de 18 a 64 anos de idade segundo grupo de estudo e existência de vínculo empregatício formal ou formalização como microempreendedor individual, para indivíduos pareados no final do período, inseridos no mercado de trabalho formal, no mês de início do curso e último mês do período estudado – Brasil, 2011 a 2014.

Status dos participantes	Participantes no setor formal					Participantes setor formal(%)		
	Total	Início do período	Fim do período	Varição absoluta	Varição relativa	Início do período	Fim do período	Varição p.p.
CadÚnico (Controle)	166.951	88.832	114.606	25.774	29,0	53,2%	68,6%	15,4
CadÚnico (Tratamento)	166.951	60.449	114.606	54.157	89,6	36,2%	68,6%	32,4

Nota: o “Início do período” refere-se ao início do curso para o grupo tratamento, o qual foi também atribuído ao grupo controle, a partir do caso pareado.

Gráfico 11 – Distribuição dos indivíduos registrados no CadÚnico de 18 a 64 anos de idade segundo grupo de estudo e existência de vínculo empregatício formal ou formalização como microempreendedor individual, para indivíduos pareados no final do período, inseridos no mercado de trabalho formal, no mês de início do curso e último mês do período estudado – Brasil, 2011 a 2014.



Nota: o “Início do período” refere-se ao início do curso para o grupo tratamento, o qual foi também atribuído ao grupo controle, a partir do caso pareado.

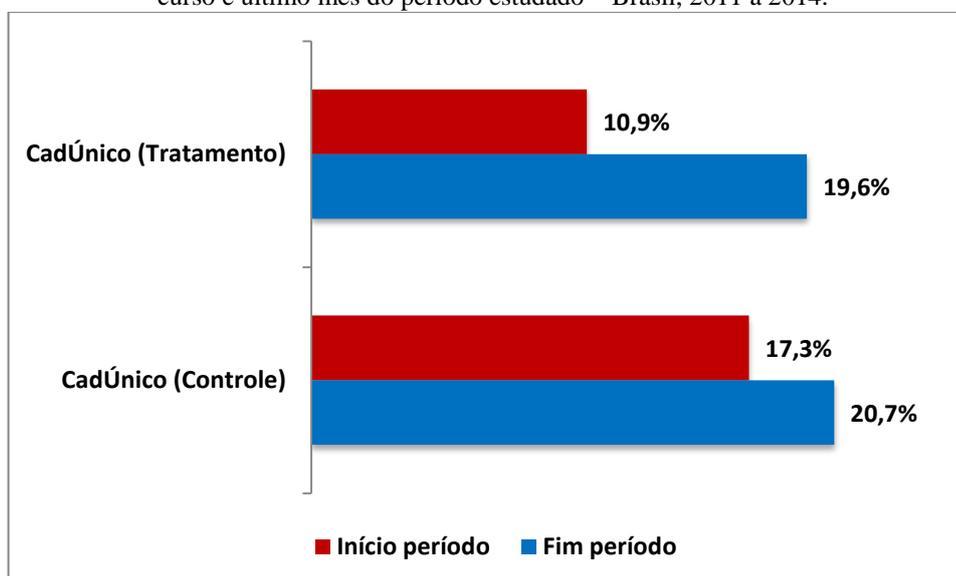
A observação dos complementares dos indivíduos pareados ao final do período, que representam os pares cujo caso-controle que não estavam ambos vinculados ao final período, revela ainda a efetivação de vínculos de trabalho formal superiores para os que cursaram Pronatec, em comparação àqueles que não cursaram (Tabela 7 e Gráfico 12).

Tabela 7 – Distribuição dos indivíduos registrados no CadÚnico de 18 a 64 anos de idade segundo grupo de estudo e existência de vínculo empregatício formal ou formalização como microempreendedor individual, para indivíduos NÃO-pareados no final do período, inseridos no mercado de trabalho formal, no mês de início do curso e último mês do período estudado – Brasil, 2011 a 2014.

Status dos participantes	Participantes no setor formal					Participantes setor formal(%)		
	Total	Início do período	Fim do período	Varição absoluta	Varição relativa	Início do período	Fim do período	Varição p.p.
CadÚnico (Controle)	1.132.090	195.807	233.879	38.072	19,4	17,3%	20,7%	3,4
CadÚnico (Tratamento)	1.132.090	123.235	221.655	98.420	79,9	10,9%	19,6%	8,7

Nota: o “Início do período” refere-se ao início do curso para o grupo tratamento, o qual foi também atribuído ao grupo controle, a partir do caso pareado.

Gráfico 12 – Distribuição dos indivíduos registrados no CadÚnico de 18 a 64 anos de idade segundo grupo de estudo e existência de vínculo empregatício formal ou formalização como microempreendedor individual, para indivíduos NÃO-pareados no final do período, inseridos no mercado de trabalho formal, no mês de início do curso e último mês do período estudado – Brasil, 2011 a 2014.



Nota: o “Início do período” refere-se ao início do curso para o grupo tratamento, o qual foi também atribuído ao grupo controle, a partir do caso pareado.

Estes resultados mostram claramente o impacto do Pronatec considerando o desfecho do Tratamento como a mudança de status (de não inserção para inserção formal no mercado de trabalho), ou mesmo como a permanência no vínculo formal ao final do período. Isso, não é pouco, considerando a elevada rotatividade do mercado de trabalho brasileiro.

Considerações finais

Os resultados obtidos no presente estudo fornecem evidências adicionais acerca do impacto do Pronatec no aumento da empregabilidade formal dos participantes, já antecipado pelos resultados apresentados em Sousa et al (2015). Os impactos são particularmente intensos entre beneficiários do Bolsa Família e de participantes residentes no Nordeste. Não houvesse o Pronatec, as chances de formalização dos seus participantes seriam certamente menores. Não tivesse o Pronatec alcançado os pequenos municípios do interior do Brasil, em especial no Nordeste e Norte, beneficiários do Programa Bolsa Família teriam tido menor possibilidade de inserção no mercado formal.

É o que apontam as evidências obtidas a partir de análises mais robustas, com pareamento de indivíduos em um estudo *quasi*-experimental, considerando como impacto a mudança de status do vínculo laboral ou a permanência no mercado formal ao final do período. Dos 2,5 milhões de participantes do Pronatec Bolsa Formação, objeto do estudo anterior, identificou-se 1,5 milhão que estavam no CadÚnico. A partir de um algoritmo de seleção aleatória foram selecionados pares de indivíduos com características similares (município, microrregião, sexo, escolaridade, faixa etária e faixa de atualização cadastral). Após os procedimentos e exclusões necessárias, a base de dados final passou a contar com 2.458.082 indivíduos, sendo 1.229.041 casos com seus respectivos controles.

O estudo aqui apresentado mostra que os participantes conseguem igualar as diferenças de inserção no mercado formal existentes no início do período. Os efeitos do programa para beneficiários do Programa Bolsa Família são ainda mais intensos, ainda mais quando são comparadas sob o aspecto regional, revelando que a região nordeste do país apresenta os maiores percentuais de inserção de indivíduos que cursaram Pronatec.

Embora as evidências da análise descritiva apresentada já apontem para a robustez dos resultados da contribuição do Pronatec na inserção dos indivíduos no mercado de trabalho formal, é possível avançar em análises estatísticas específicas que avaliem os efeitos do programa, considerando a mudança de status de inserção ao longo do período em análise.

Referências bibliográficas

FREY, K. **Políticas públicas: um debate conceitual e reflexões referentes à prática da análise de políticas públicas no Brasil**. Planejamento e Políticas Públicas, Brasília, n. 21, p. 211-259, jun. 2000.

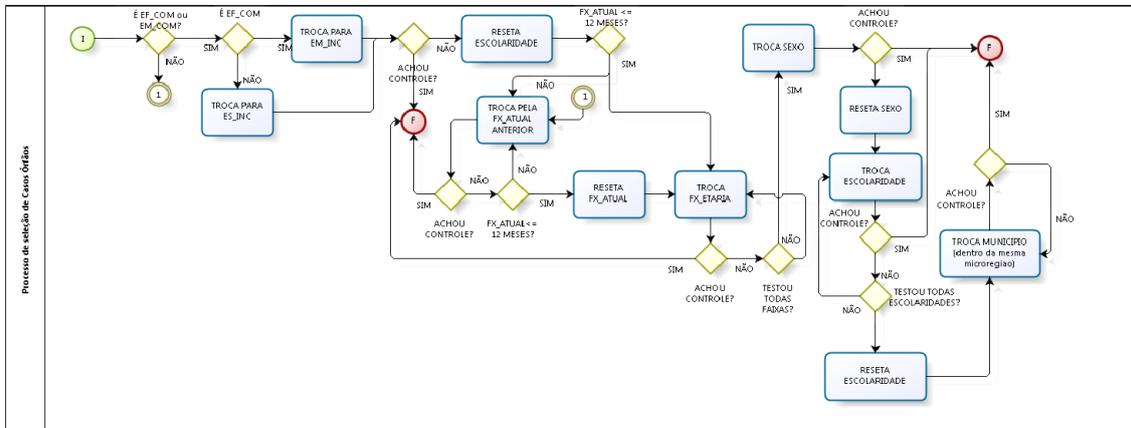
KATZ, MH. **Study Design and Statistical Analysis: A Practical Guide for Clinicians**. England: Cambridge University Press, 2006.

SCHLESSELMAN, JJ. **Case-control studies**. EUA: Oxford University Press, Inc, 1982.

SOUSA, MF, SILVA, YC, JANNUZZI, PM. **Contribuição do Programa Nacional de Acesso ao Ensino Técnico e Emprego – Bolsa Formação para formalização do emprego e dos microempreendedores individuais: análise dos participantes beneficiários do Bolsa Família e inscritos no Cadastro Único**. Cadernos de Estudos Desenvolvimento Social em Debate – N. 24 (2015). Brasília, DF. Ministério do Desenvolvimento Social e Combate à Fome; Secretaria de Avaliação e Gestão da Informação, 2005, p. 156-179.

APÊNDICE

Fluxograma 2 – Processo de seleção de casos órfãos.



Fonte: Dados produzidos pelo Departamento de Monitoramento - SAGI