1517
TEXTO PARA DISCUSSÃO

ipea

UMA METODOLOGIA PARA DECOMPOR DIFERENÇAS ENTRE DADOS ADMINISTRATIVOS E PESQUISAS AMOSTRAIS, COM APLICAÇÃO PARA O PROGRAMA BOLSA FAMÍLIA E O BENEFÍCIO DE PRESTAÇÃO CONTINUADA NA PNAD

Pedro H. G. Ferreira de Souza

Instituto de Pesquisa Econômica Aplicada

1517

TEXTO PARA DISCUSSÃO

Rio de Janeiro, dezembro de 2010

UMA METODOLOGIA PARA DECOMPOR DIFERENÇAS ENTRE DADOS ADMINISTRATIVOS E PESQUISAS AMOSTRAIS, COM APLICAÇÃO PARA O PROGRAMA BOLSA FAMÍLIA E O BENEFÍCIO DE PRESTAÇÃO CONTINUADA NA PNAD*

Pedro H. G. Ferreira de Souza**

*A autor agradece os excelentes comentários de Rafael Guerreiro Osório e de Sergei Soares.

**Técnico de Planejamento e Pesquisa da Diretoria de Estudos e Políticas Sociais (Disoc) do Ipea.

Governo Federal

Secretaria de Assuntos Estratégicos da Presidência da República
Ministro Samuel Pinheiro Guimarães Neto

ipea **Instituto de Pesquisa Econômica Aplicada**

Fundação pública vinculada à Secretaria de Assuntos Estratégicos da Presidência da República, o Ipea fornece suporte técnico e institucional às ações governamentais – possibilitando a formulação de inúmeras políticas públicas e programas de desenvolvimento brasileiro – e disponibiliza, para a sociedade, pesquisas e estudos realizados por seus técnicos.

Presidente
Marcio Pochmann

Diretor de Desenvolvimento Institucional
Fernando Ferreira

Diretor de Estudos e Relações Econômicas e Políticas Internacionais
Mário Lisboa Theodoro

Diretor de Estudos e Políticas do Estado, das Instituições e da Democracia
José Celso Pereira Cardoso Júnior

Diretor de Estudos e Políticas Macroeconômicas
João Sicsú

Diretora de Estudos e Políticas Regionais, Urbanas e Ambientais
Liana Maria da Frota Carleial

Diretor de Estudos e Políticas Setoriais, de Inovação, Regulação e Infraestrutura
Márcio Wohlers de Almeida

Diretor de Estudos e Políticas Sociais
Jorge Abrahão de Castro

Chefe de Gabinete
Persio Marco Antonio Davison

Assessor-chefe de Imprensa e Comunicação
Daniel Castro

URL: http://www.ipea.gov.br
Ouvidoria: http://www.ipea.gov.br/ouvidoria

Texto para Discussão

Publicação cujo objetivo é divulgar resultados de estudos direta ou indiretamente desenvolvidos pelo Ipea, os quais, por sua relevância, levam informações para profissionais especializados e estabelecem um espaço para sugestões.

As opiniões emitidas nesta publicação são de exclusiva e inteira responsabilidade do(s) autor(es), não exprimindo, necessariamente, o ponto de vista do Instituto de Pesquisa Econômica Aplicada ou da Secretaria de Assuntos Estratégicos da Presidência da República.

É permitida a reprodução deste texto e dos dados nele contidos, desde que citada a fonte. Reproduções para fins comerciais são proibidas.

ISSN 1415-4765

JEL: I38, C13

SUMÁRIO

SINOPSE

ABSTRACT

1 INTRODUÇÃO 7

2 METODOLOGIA 7

3 OS ERROS DE INFERÊNCIA DO PBF NA PNAD 11

4 O ERRO DE INFERÊNCIA DO BPC NA PNAD 15

5 CONCLUSÃO 19

REFERÊNCIAS 20

SINOPSE

Estimativas feitas com base em pesquisas domiciliares amostrais muitas vezes diferem bastante dos dados administrativos. Nas Pesquisas Nacionais por Amostra de Domicílios (PNADs), por exemplo, o número estimado de beneficiários do Programa Bolsa Família (PBF) e do Benefício de Prestação Continuada (BPC) é sempre inferior ao número oficial. O objetivo deste artigo é apresentar uma metodologia simples, baseada nas características do desenho amostral das pesquisas domiciliares, para explicar essa diferença, decompondo-a em três termos: o viés de representatividade (derivado da escolha dos locais, áreas censitárias ou municípios para a pesquisa), o viés de captação (decorrente de problemas de captação nos locais selecionados) e a interação entre ambos. A aplicação dessa metodologia ao PBF e ao BPC mostra que, no primeiro caso, o viés de representatividade explica boa parte do problema: a seleção de municípios pesquisados é responsável por 40% da diferença observada entre os dados oficiais e os da PNAD. No caso do BPC, o viés de representatividade tenderia a agir no sentido oposto. Portanto, o viés de captação é inteiramente responsável pela diferença observada. Além disso, a declaração equivocada do BPC como benefício previdenciário na PNAD parece ocorrer sobretudo no período anterior a 2004 e não explica inteiramente o pequeno número de beneficiários identificados nas PNADs.

ABSTRACT[i]

Estimates based on household surveys often differ considerably from administrative records. In the PNADs, for instance, the estimated number of beneficiaries of the Programa Bolsa Família (PBF) and of the Benefício de Prestação Continuada (BPC) are always lower than the official figures. This paper presents a simple methodology, based on the sampling design of household surveys, to explain these differences, decomposing them in three terms: the representativeness bias (due to selection of locales, census tracts or municipalities for the survey), the selection bias (due to selection issues in the chosen locales) and the interaction among them. The application of this methodology to the PBF and the BPC reveals that, regarding the former, the representativeness bias is accountable for a good part of the problem: the selection of the municipalities to

---

[i]. *The versions in English of the abstracts of this series have not been edited by Ipea's editorial department.*
As versões em língua inglesa das sinopses (abstracts) desta coleção não são objeto de revisão pelo Editorial do Ipea.

be surveyed is responsible for 40% of the difference between official records and the PNAD. In the case of the BPC, the representativeness bias tends to act in the opposite direction. Thus, the selection bias is entirely responsible for the observed difference. Also, the erroneous reporting of the BPC as a Social Security benefit in the PNAD seems to occur mostly in the years prior to 2004 and in any case does not explain by itself the low number of beneficiaries identified in the PNAD.

1 INTRODUÇÃO

Um dos problemas mais comuns na análise de políticas públicas a partir de pesquisas domiciliares amostrais é que as estimativas assim obtidas muitas vezes diferem de forma considerável dos dados administrativos. No Brasil, os casos do Programa Bolsa Família (PBF) e do Benefício de Prestação Continuada (BPC) chamam a atenção: o número de beneficiários de ambos os programas, identificados nas Pesquisas Nacionais por Amostra de Domicílios (PNADs), feitas pelo Instituto Brasileiro de Geografia e Estatística (IBGE), tem sido sistematicamente inferior ao número oficialmente registrado, o que provoca em alguns leitores dúvidas quanto à fidedignidade das informações. O objetivo deste texto é propor uma metodologia simples para explicar essas diferenças com base nas características do desenho amostral das pesquisas domiciliares. Em seguida, aplicaremos a metodologia proposta aos casos do PBF e do BPC na PNAD.

Com isso, não pretendemos pôr em dúvida a robustez das conclusões quase consensuais das análises baseadas em dados da PNAD, até porque não há como verificar em que medida o perfil dos beneficiários não identificados difere daquele dos efetivamente identificados. Esperamos, no entanto, que a metodologia proposta ajude a dirimir algumas das muitas dúvidas que existem sobre o assunto e que possa ser aplicável também em outros casos.

2 METODOLOGIA

Em levantamentos censitários, todas as unidades de um universo de interesse são pesquisadas: por exemplo, todos os indivíduos de um determinado país. Como os custos de tais pesquisas são quase sempre muito elevados, as pesquisas amostrais tendem a ser mais atraentes. No entanto, para serem válidas, é preciso garantir que a amostra selecionada seja representativa da população como um todo. O método mais intuitivo é o de sorteio aleatório simples. No entanto, quando a população é composta por subgrupos muito heterogêneos entre si, amostras estratificadas são vantajosas. Para isso, dividimos a população em estratos relativamente homogêneos e mutuamente exclusivos e, em seguida, selecionamos os casos amostrados dentro de cada estrato.

A PNAD, por exemplo, utiliza uma amostra estratificada que – simplificando muito – termina por dividir o país em três estratos.[1] O primeiro engloba os municípios pertencentes às dez regiões metropolitanas (RMs), que necessariamente são incluídos na amostra. Isso garante que a pesquisa seja representativa para cada RM e, portanto, para o conjunto de RMs. O segundo contém os chamados municípios autorrepresentativos (AR), que, em função do seu porte, também são necessariamente incluídos na amostra. Finalmente, há o estrato de municípios não autorrepresentativos (NAR), que corresponde ao resto do país. Destes, apenas uma pequena fração – sorteada de acordo com probabilidades proporcionais à sua população – é selecionada e considerada representativa de todo o conjunto. Ou seja, os municípios menores têm menor probabilidade de inclusão.[2] Neste trabalho, chamaremos de NAR-I os municípios não autorrepresentativos incluídos na PNAD e de NAR-NI os não autorrepresentativos não incluídos.

Se partirmos do pressuposto de que os registros administrativos estão corretos, então, em qualquer pesquisa amostral estratificada, os eventuais erros de inferência – isto é, diferenças entre o número de casos estimado pela pesquisa e os dados oficiais dos registros administrativos – podem derivar de duas fontes básicas:

1) Viés de representatividade: é possível que o desenho amostral da pesquisa não seja representativo da população como um todo, no que diz respeito ao programa em questão. Por exemplo, a incidência do programa nas áreas selecionadas para a realização das entrevistas pode ser menor do que nas demais áreas e, portanto, a extrapolação dos resultados do primeiro grupo para o segundo inevitavelmente acarretará algum erro de inferência. Na PNAD, este tipo de viés pode ocorrer no estrato de municípios NAR se os municípios incluídos (NAR-I) não forem uma boa amostra dos demais (NAR-NI). No caso do PBF, por exemplo, pode-se especular que, como ele é orientado por cotas municipais e tem forte presença em pequenos municípios do interior do país, é possível que o erro de inferência associado à subestimação do número de beneficiários derive em boa parte deste viés. Esta hipótese será testada na próxima seção.

2) Viés de captação: a amostra é efetivamente representativa de uma determinada população, mas, por algum motivo, a captação do programa é imperfeita, o que faz com que o total estimado seja diferente dos números oficiais. Por um

1. Nos microdados das PNADs os estratos aqui mencionados são chamados de "áreas censitárias".
2. Para mais detalhes sobre o plano amostral da PNAD, ver Silva, Pessoa e Lila (2002).

lado, isso pode ocorrer porque os indivíduos de interesse estão geograficamente concentrados ou são difíceis de serem amostrados: por exemplo, se ocorre um sorteio aleatório dentro de um município, mas o subgrupo que queremos pesquisar está fortemente concentrado em pequenos bairros, é possível que ele esteja subrepresentado em nossa amostra. Por outro lado, este viés também pode ocorrer nos casos em que os indivíduos são de fato sorteados, mas, propositalmente ou não, deixam de informar que participam de determinados programas. O caso oposto também pode ocorrer: propositalmente ou não, talvez influenciados pelo encadeamento e a formulação das perguntas, indivíduos que não participam podem erroneamente se declarar beneficiários de determinados programas. Na PNAD, este tipo de viés pode ocorrer em todos os estratos. Em pesquisas censitárias, também.

O método proposto decompõe os erros de inferência – a diferença entre os dados administrativos e as estimativas das pesquisas amostrais – em três elementos: o viés de representatividade, o viés de captação e a interação entre ambos. O objetivo é de determinar a influência relativa de cada efeito, o que pode ser conseguido a partir da fórmula abaixo:

Erro de inferência = viés de representatividade +
viés de captação +
(1 – taxa captação) * (1 – taxa representatividade) *
RegAdmin

com:

Erro de inferência = RegAdmin – PDA

RegAdmin = número de benefícios nos registros administrativos

PDA = número de benefícios na Pesquisa Domiciliar Amostral; por exemplo, na PNAD

Viés de representatividade = (taxa captação * 1 * RegAdmin) – PDA

Viés de captação = (1 * taxa representatividade * RegAdmin) – PDA

$$\text{Taxa Captação} = \frac{\text{Benefícios } \textit{per capita} \text{ na PDA}}{\text{Benefícios } \textit{per capita} \text{ no RegAdmin}}$$

$$\text{Taxa Representatividade} = \frac{\text{Benefícios } \textit{per capita} \text{ do EstratoPDA no RegAdmin}}{\text{Benefícios } \textit{per capita} \text{ no EstratoReal no RegAdmin}}$$

Taxa repEstratoPDA = os estratos da pesquisa domiciliar em questão; no caso da PNAD, RMs, municípios AR e municípios NAR.

EstratoReal = os estratos "reais"; no caso da PNAD, RMs, municípios AR, municípios NAR-I e municípios NAR-NI.

No caso da PNAD, esta decomposição é particularmente útil porque o desenho amostral atual da pesquisa mantém a seleção de municípios constante entre censos.[3] Em outras palavras, os mesmos 851 municípios selecionados para a PNAD de 2001 continuaram sendo pesquisados em todos os anos, até 2009. Assim, caso a seleção de municípios tenha algum viés em relação a algum programa, este viés será repetido em todas as pesquisas do período se o programa não mudar seu padrão de distribuição espacial. Como os municípios dos estratos RM e AR necessariamente são incluídos, a grande dúvida diz respeito aos municípios do estrato NAR: se houver discrepâncias significativas na incidência do programa entre os NAR incluídos (NAR-I) e os não incluídos (NAR-NI), então os primeiros não serão uma amostra representativa do segundo grupo e, portanto, a estimativa da PNAD será necessariamente distorcida.

A tabela 1 informa alguns dados básicos dos municípios selecionados nas PNADs dos anos 2000. Como nas próximas seções vamos utilizar principalmente a PNAD de 2006 para investigar os erros de inferência do PBF e do BPC, escolhemos esse ano como referência para a tabela.

TABELA 1
Brasil: número de municípios e população por estratos – 2006

Inclusão na PNAD	Estrato	Número de municípios	População (milhões)	%
Sim	RM	173	58,2	31
	AR	134	39,6	21
	NAR-I	544	21,4	11
Não	NAR-NI	4.714	67,6	36
Total		5.565	186,8	100

Fonte: Elaboração própria com base na seleção de municípios da PNAD de 2001-2009 e nas estimativas de população do IBGE para 1ª de julho de 2006, publicadas no *Diário Oficial da União* em 31 de agosto de 2006.

3. Após o Censo de 2010 o IBGE vai reformular suas pesquisas, com a implementação do Sistema Integrado de Pesquisas Domiciliares (SIPD), que prevê a existência de um novo plano amostral – denominado Amostra Mestra – que será compartilhado por todas as pesquisas. Para mais informações, ver Freitas et al. (2007) e Coordenação de Trabalho e Rendimento/IBGE (2008).

3 OS ERROS DE INFERÊNCIA DO PBF NA PNAD

A tabela 2 mostra como os erros de inferência do PBF são um problema crônico nas PNADs, ainda que sua intensidade apresente variações. Isto vai na contramão das expectativas de alguns autores, que imaginavam que os erros seriam temporários.[4] Rocha (2008) e Soares *et al.* (2006), por exemplo, argumentaram que, como o período 2003-2004 foi de grandes mudanças institucionais e de unificação dos programas existentes sob a égide do PBF, o erro de inferência provavelmente decorreria da confusão dos entrevistados, que estariam declarando o programa a que pertenciam originalmente, fenômeno que poderia ser superado com o tempo e a consolidação do programa.

TABELA 2
Brasil: famílias de beneficiárias do PBF de acordo com os registros administrativos e a PNAD – 2006-2008
(Em milhões de benefícios)

Mês/ano	a. Registro administrativo[1]	b. PNAD	Método[2]	b/a (%)
Setembro/2006	11,018	8,142	Suplemento	74
Setembro/2006	11,018	9,421	Valor típico	86
Setembro/2007	10,961	7,198	Valor típico	66
Setembro/2008	10,843	8,924	Valor típico	82

Fonte: Elaboração própria com base em dados da Secretaria de Avaliação e Gestão da Informação do Ministério do Desenvolvimento Social e Combate à Fome (Sagi/MDS) e microdados das PNADs de 2006-2008.

Nota: [1]Os dados da Sagi/MDS não especificam se os números incluem ou não os benefícios temporariamente bloqueados. Caso estes estejam contemplados, então a diferença entre o dado oficial e a PNAD seria menor do que a da tabela. De qualquer maneira, pode-se especular que: *i*) os benefícios bloqueados tendem a ser residuais, representando um pequeno percentual do total; e *ii*) é possível que eles se distribuam de forma mais ou menos aleatória entre os estratos da PNAD. Assim, sua retirada provavelmente não influenciaria em nada as principais conclusões obtidas mais adiante.
[2]Adotou-se aqui uma versão simplificada do método dos valores típicos: para cada ano, todos os "outros rendimentos" individuais abaixo do valor máximo do PBF foram atribuídos ao programa. O método dos valores típicos foi desenvolvido e validado sobretudo por Barros (2006) e Barros, Carvalho e Franco (2006).

Uma hipótese plausível é a de que o erro de inferência que leva à subestimação dos beneficiários resultaria da inexistência de perguntas específicas sobre o PBF no corpo básico da PNAD, o que, naturalmente, teria menor influência nos anos em que foi incluído o suplemento especial. Só que, mesmo no suplemento de 2006, o erro permanece significativo, o que ensejou uma nova hipótese específica para esse ano: em setembro de 2006 – mês em que foram realizadas as entrevistas da PNAD – cerca de 1,5 milhão de cartões do PBF talvez ainda estivessem nos Correios e, portanto, seus titulares não teriam recebido de fato o benefício (SOARES; RIBAS; SOARES, 2009).

4. As PNADs de 2004 e 2006 contaram com suplementos especiais sobre programas de transferência de renda que assinalavam se algum morador do domicílio recebia PBF, BPC ou outros programas. O valor recebido, todavia, não era discriminado, estando misturado às demais fontes residuais de renda declaradas como "outros rendimentos" (variável v1273). Uma metodologia para desagregar os rendimentos desses programas a partir da combinação do suplemento com esta variável foi desenvolvida por Soares *et al.* (2006).

Outras explicações possíveis referem-se à natureza da projeção populacional da PNAD, que tornaria a pesquisa menos confiável para a obtenção de números absolutos (SOARES; RIBAS; SOARES, 2009) e à sua amostra nucleada, que, por definição, só capta bem populações razoavelmente espalhadas, mas não grupos geograficamente concentrados, que tenderiam a estar sub-representados (SOARES *et al.*, 2006).

Na decomposição proposta acima, todos estes motivos são agregados no que chamamos de viés de captação. A rigor, não há como testá-los diretamente com os dados disponíveis. Pode-se especular que a diminuição do erro nos anos com o suplemento sugere que a introdução de perguntas específicas reduziria o problema; entretanto, só se observa um efeito forte em 2007. Não há como saber de antemão se esse foi um ano atípico ou se a consolidação e a visibilidade adquiridas pelo programa posteriormente contribuíram para contrabalançar a ausência de perguntas específicas.

A tabela 3 mostra as informações necessárias para o cálculo da decomposição para a PNAD de 2006, ano em que, graças ao suplemento especial, a incerteza acerca da identificação dos beneficiários é menor.[5] Vemos, neste caso, que a concentração geográfica dos beneficiários provavelmente também desempenha um papel relevante, uma vez que a captação dos beneficiários é pior nos estratos RM e AR, compostos por municípios maiores, mais ricos e mais heterogêneos.

TABELA 3
Brasil: benefícios do PBF nos registros administrativos e na PNAD, por estratos – 2006
(Em milhões)

Estrato		Reg Admin: # de PBFs (milhões)	a. Benefícios *per capita*	PNAD: # de PBFs (milhões)[1]	b. Benefícios *per capita*	b/a
RM		1,954	0,034	1,249	0,022	0,639
AR		1,629	0,041	1,150	0,029	0,706
NAR	NAR-I	1,487	0,070	1,381	0,065	0,928
	NAR-NI	5,948	0,089	4,363	0,065	0,734
Total		11,018	0,059	8,142	0,044	0,739

Fonte: Elaboração própria com base em dados da Sagi/MDS e da PNAD de 2006.
Nota: [1]Obviamente, não é possível, apenas com base na PNAD, separar a população dos NAR-I e dos NAR-NI. As proporções de cada grupo foram obtidas a partir das estimativas de população do IBGE para 1º de julho de 2006, publicadas no *Diário Oficial da União* em 31 de agosto daquele ano.

O método proposto deve ser aplicado aos dados de cada estrato listado na tabela acima, sendo que, no caso de NAR-NI, a taxa de representatividade é a mesma de

5. Afinal, o método dos valores típicos, embora pareça muito robusto, envolve, por definição, certo nível de incerteza quanto à origem dos rendimentos atribuídos ao PBF.

NAR-I (0,928) e a taxa de captação é a razão entre os valores *per capita* dos registros administrativos de NAR-I e NAR-NI (0,070/0,089 = 0,790). Naturalmente, o viés de representatividade só existe para o estrato NAR-NI: supõe-se que, caso este estrato estivesse presente na PNAD, sua taxa de captação seria a mesma do estrato NAR-I. Em outras palavras, o viés de representatividade supõe o cálculo contrafactual do número de beneficiários que seriam identificados caso houvesse um estrato NAR-NI na PNAD e a taxa de captação fosse de cerca de 93% do número real de beneficiários do estrato.

A tabela 4 mostra os resultados da decomposição. O erro de inferência é de 2,875 milhões de beneficiários: são pouco mais de 11 milhões nos dados administrativos e cerca de 8,15 milhões na PNAD. Cerca de 40% desta diferença pode ser atribuída unicamente ao viés de representatividade, decorrente do desenho amostral da PNAD, e outros 3% derivam da interação entre o viés de representatividade e o de captação.

Ou seja, mesmo com todo o viés de captação existente, caso a amostra da PNAD fosse efetivamente representativa da distribuição territorial do PBF, o número de beneficiários identificados aumentaria em 1,16 milhão, chegando a 9,30 milhões de famílias. Analogamente, se a captação do PBF na PNAD fosse perfeita para o desenho amostral atual, o número de famílias beneficiárias aumentaria em 1,63 milhão, atingindo um teto de 9,77 milhões, número ainda inferior aos 11,02 milhões oficialmente registrados.

TABELA 4
Brasil: decomposição do erro de inferência – 2006
(Em milhões)

Estrato		Erro de inferência	I. Viés de captação	II. Viés de representatividade	III. Interação	Total (I+II+III)
RM		0,705	0,705	0	0	0,705
AR		0,479	0,479	0	0	0,479
NAR	NAR-I	0,107	0,107	0	0	0,107
	NAR-NI	1,585	0,337	1,158	0,089	1,585
Total		2,875	1,628	1,158	0,089	2,875
%		100	57	40	3	100

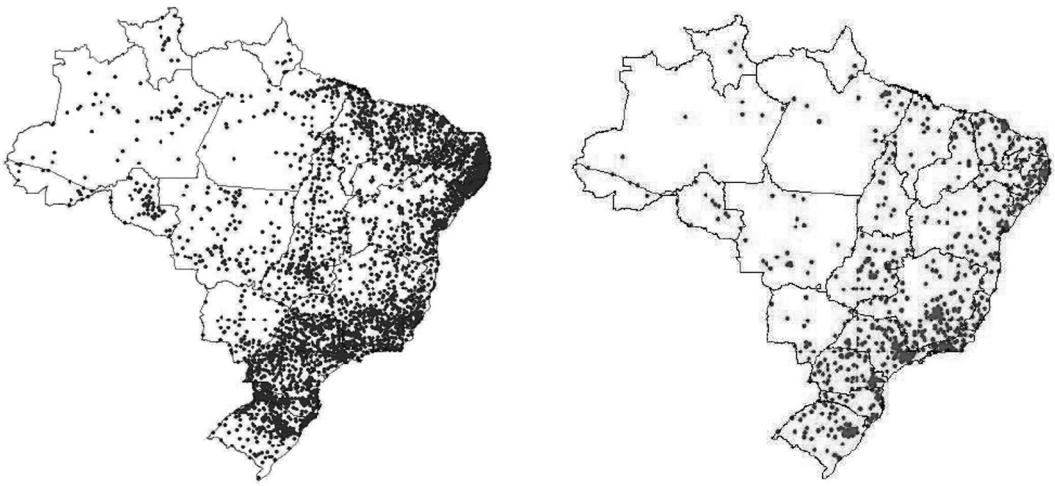
Fonte: Elaboração própria com base em dados da Sagi/MDS e da PNAD de 2006.

Os resultados mostram os limites da PNAD e do seu desenho amostral quando o objetivo é analisar programas com forte penetração em pequenos municípios do interior do país. Trata-se de um obstáculo impossível de ser contornado por métodos

de identificação mais refinados. Dado o viés introduzido pelo desenho amostral, seria espantoso se a PNAD reproduzisse os números oficiais.

Não temos como saber, pelo menos até que estejam disponíveis os resultados do Censo de 2010, em que medida este viés afeta as principais conclusões derivadas das PNADs. Felizmente, o novo desenho amostral das PNADs Contínuas a partir de 2011 (ver nota 4) tem potencial para minimizar esses problemas, em razão do maior número de setores censitários e municípios incluídos: serão 12.800 setores em 3.328 municípios, contra 7.816 setores em 851 municípios nas PNADs de 2001-2009 (COORDENAÇÃO DE TRABALHO E RENDIMENTO/IBGE, 2008). A diferença pode ser percebida visualmente na figura 1: a Amostra Mestra das novas PNADs Contínuas que irão a campo depois do Censo de 2010 inclui setores censitários de um número maior de municípios, com muito mais capilaridade pelo interior do país.

FIGURA 1
Municípios com setores censitários selecionados para a PNAD Contínua (esquerda) e para as PNADS dos anos 2000 (direita)



Fonte: Coordenação de Trabalho e Rendimento/IBGE (2008).

Por enquanto, os únicos dados disponíveis baseados na nova Amostra Mestra são os da POF de 2008-2009, que, no entanto, visitou apenas um subconjunto formado por 4.696 setores censitários em 1.757 municípios e, surpreendentemente, apresentou resultados quase idênticos aos da PNAD de 2008: na nova POF, é possível identificar

cerca de 8 milhões de famílias beneficiárias do PBF, contra quase 11 milhões de famílias oficialmente registradas. De qualquer modo, é preciso esperar os resultados das PNADs Contínuas para podermos avaliar melhor o efeito de tais mudanças no desenho amostral.

4 OS ERROS DE INFERÊNCIA DO BPC NA PNAD

Assim como ocorre com o PBF, o número de beneficiários do BPC estimado pela PNAD tem sido cronicamente inferior ao dos dados administrativos. Em boa medida, isso surpreende, pois este benefício é teoricamente mais fácil de captar tanto por pagar um valor fixo único e de maior vulto – 1 salário mínimo (SM) – quanto por ser talvez menos afetado por questões amostrais ou geográficas, uma vez que se trata de um direito individual, assegurado pela Constituição Federal de 1988 e não limitado por cotas municipais. No entanto, vemos na tabela 5 que a identificação dos beneficiários é extremamente problemática no período anterior a 2004, ano do primeiro suplemento especial. Há uma melhora sensível entre 2004 e 2006, mas o erro de inferência relativo volta a aumentar em 2007 e 2008. Mais uma vez, a fonte mais confiável parece ser o suplemento da PNAD de 2006.

TABELA 5
Brasil: número de benefícios do BPC de acordo com o registro administrativo e a PNAD – 2006
(Em milhões)

Ano	a. Registro administrativo[1]	b. PNAD	Método	b/a (%)
1996	0,346	0,028	Valor típico	8
1997	0,646	0,062	Valor típico	10
1998	0,848	0,113	Valor típico	13
1999	1,033	0,240	Valor típico	23
2001	1,339	0,175	Valor típico	13
2002	1,561	0,381	Valor típico	24
2003	1,701	0,289	Valor típico	17
2004	1,984	0,925	Valor típico	45
2005	2,208	1,198	Valor típico	53
2006	2,430	1,588	Valor típico	64
2007	2,618	1,298	Valor típico	48
2008	2,862	1,474	Valor típico	50
2004	1,984	0,901	Suplemento	45
2006	2,430	1,318	Suplemento	54

Fonte: Elaboração própria com base em dados do *Anuário Estatístico de Previdência Social* (Aeps/Infologo) e das PNADs de 1996-2008.

Nota: [1]Para 1996-2003 os dados se referem ao mês de dezembro; no período 2004-2008 a informação é para o mês de setembro (mês da realização das PNADs).

A maioria dos autores que enfrentaram o problema do BPC propôs uma mesma hipótese: dado o caráter e o valor do benefício, muitos beneficiários provavelmente declaram as rendas recebidas como se fossem aposentadorias ou pensões, captadas separadamente na PNAD (ver SOARES *et al*, 2006; BARROS; CARVALHO; FRANCO, 2006; ROCHA, 2008). Dada a popularidade dessa hipótese, faremos, além da decomposição proposta, também uma rápida análise exploratória relacionada a ela.

A tabela 6 reforça a ideia de que o BPC não é afetado pelo mesmo viés amostral do PBF. Com efeito, a incidência do benefício no estrato NAR-I é até maior do que no estrato NAR-NI, o que implica que o benefício tenderia a estar superestimado na PNAD, caso sua captação fosse perfeita: se isso ocorresse, o número de beneficiários na PNAD seria maior do que o dos registros administrativos.

TABELA 6
Brasil: benefícios do BPC nos registros administrativos e na PNAD, por estratos – 2006
(Em milhões)

Estratos		Reg Admin: # de BPCs (milhões)	a. Benefícios *per capita*	PNAD: # de BPCs (milhões)[1]	b. Benefícios *per capita*	b/a
RM		0,636	0,011	0,210	0,004	33
AR		0,531	0,013	0,240	0,006	45
NAR	NAR-I	0,333	0,016	0,183	0,009	55
	NAR-NI	0,931	0,014	0,580	0,009	62
Total		2,430	0,013	1,213	0,007	50

Fonte: Elaboração própria com base em dados do Aeps/Infologo e da PNAD de 2006.

Nota: [1]Ver nota da tabela 3.

A tabela 7 apresenta a decomposição proposta, que confirma os indícios anteriores: no caso do BPC, o viés introduzido pelo desenho amostral (viés de representatividade) tenderia a provocar uma superestimação do número de beneficiários e, portanto, a diferença entre registros administrativos e PNAD pode ser inteiramente atribuída ao viés de captação. Se a captação da PNAD fosse perfeita, o número de BPCs identificados seria de 2,550 milhões, 5% a mais do que os 2,430 milhões oficialmente pagos.

Uma vez descartada a influência do desenho amostral, resta examinar a possibilidade de o BPC ser declarado como rendimento de aposentadorias ou pensões. Infelizmente, não pudemos dispor da distribuição municipal dos dados previdenciários, o que possibilitaria testar a influência do desenho amostral da PNAD sobre estes benefícios. De qualquer maneira, a comparação entre os registros administrativos e

a PNAD já permite uma primeira observação: como o BPC cresceu muito na última década, é de se esperar que a razão entre o número de benefícios previdenciários da PNAD e o dos dados administrativos tenha aumentado ao longo do tempo, caso os beneficiários efetivamente estejam declarando o BPC como aposentadorias ou pensões.

TABELA 7
Brasil: decomposição do erro de inferência – 2006

Estratos		Erro de inferência	I. Viés de captação	II. Viés de representatividade	III. Interação	Total (I+II+III)
RM		0,426	0,426	0	0	0,426
AR		0,291	0,291	0	0	0,291
NAR	NAR-I	0,149	0,149	0	0	0,149
	NAR-NI	0,352	0,471	–0,659	–0,054	0,352
Total		1,218	1,337	–0,659	–0,054	1,218
%		100	110	–5	–4	100

Fonte: Elaboração própria com base em dados da Sagi/MDS e da PNAD de 2006.

A comparação direta pode ser encontrada na tabela 8. Restrições quanto à disponibilidade de dados limitam o exercício aos anos 2001-2008. De imediato, percebe-se um comportamento não homogêneo ao longo de todo o período: entre 2001 e 2003 o BPC expandiu-se, com um aumento de 27% no número de benefícios, e a razão entre os benefícios previdenciários na PNAD e nos registros administrativos subiu de 0,99 para 1,11, o que constitui um bom indício de que, neste período, há um grau razoável de declarações equivocadas do BPC na PNAD.[6] Contudo, no período seguinte o programa continuou sua expansão – aumento de 72% no número de beneficiários entre 2003 e 2008 – mas a razão entre benefícios previdenciários na PNAD e nos dados administrativos diminuiu de 1,11 para 1,04.[7] Ou seja, há razões para crer que a declaração errônea do BPC como aposentadoria ou pensão é um fenômeno bastante real no período anterior ao primeiro suplemento, mas com força declinante no período subsequente, provavelmente em função do melhor treinamento dos entrevistadores.

Se supusermos que a PNAD deveria ser capaz de estimar o número exato de benefícios previdenciários de 1 SM, teríamos então um excesso de cerca de 600 mil benefícios em 2008. Se supusermos uma taxa de captação igual à de 2001 (0,99), o excesso aumentaria para 760 mil benefícios, ou 5% do total de benefícios identificados

6. No mesmo período, o número de benefícios previdenciários iguais a 1SM aumentou 4%.
7. Entre 2003 e 2008 o número de benefícios previdenciários iguais a 1 SM cresceu 22%.

na PNAD. Não temos como saber se esses benefícios são, de fato, BPCs. Mesmo que o sejam, entretanto, ainda há razões para suspeitar que a não captação – isto é, indivíduos que recebem o BPC e não são entrevistados ou, caso o sejam, não o declaram nem como "outros rendimentos" nem como aposentadoria ou pensão – permanece como um problema relevante.

TABELA 8
Brasil: benefícios assistenciais e previdenciários equivalentes a 1 SM de acordo com registros administrativos e a PNAD – 2006
(Em milhões)

Ano	Registro administrativo[1]		PNAD		%		
	A BPC	B Previdência (1 SM)[2]	C BPC[3]	D Previdência (1 SM)	C/A	D/B	(C+D)/(A+B)
2001	1,339	10,649	0,175	10,497	13	99	89
2002	1,561	10,881	0,381	11,551	24	106	96
2003	1,701	11,049	0,289	12,303	17	111	99
2004	2,061	11,401	0,925	12,196	45	107	97
2005	2,277	11,926	1,198	12,942	53	109	100
2006	2,477	12,472	1,588	13,168	64	106	99
2007	2,681	12,942	1,298	13,518	48	104	95
2008	2,934	13,431	1,474	14,001	50	104	95

Fonte: Elaboração própria com base em dados do Aeps/Infologo e das PNADs de 2001-2008.

Notas: [1]Dados referentes ao mês de dezembro de cada ano; disto decorre a diferença, para 2004-2008, com relação à tabela 5.

[2]Inclui apenas as seguintes rubricas do Aeps: "42-Ap. tempo contrib LOPS; 46-Ap. tempo contrib especial; Outras Ap. tempo contrib; Ap. idade; Ap. invalidez; pensões por morte; aposentadorias acidentárias; pensões acidentárias; pensões mensais vitalícias; EPU - Ap. TC anistiados; EPU - Ap. invalidez; EPU - pensão por morte; EPU - pensões mensais vitalícias".

[3]Identificado pelo método dos valores típicos.

A conta é simples: a PNAD 2008 identifica metade dos 2,934 milhões de BPCs existentes. Pelo seu desenho amostral, como vimos, caso a captação fosse perfeita o número da PNAD deveria ser ainda superior ao oficial. Mantendo a cobertura de 2001, vimos que há um excesso de 760 mil benefícios da previdência. Mesmo que todos esses sejam BPCs, ainda estariam faltando outros 700 mil benefícios deste programa, ou 24% do total dos dados administrativos. Levando em consideração o fato de que a captação perfeita da PNAD aparentemente superestimaria o número de benefícios em 5%, restariam ainda 850 mil benefícios não declarados nem como "outros rendimentos" nem como aposentadorias ou pensões.

Em outras palavras, mesmo no caso limite de que não haja influência do desenho amostral da PNAD na identificação dos benefícios previdenciários e de que todo o "excesso" do número de aposentadorias e pensões de 1 SM na PNAD resulte de

declarações errôneas de beneficiários do BPC, ainda assim uma parte substantiva dos benefícios deste último programa permanece não identificada, parte maior até mesmo do que o número de benefícios declarados equivocadamente como aposentadorias ou pensões (850 *versus* 760 mil).

5 CONCLUSÃO

A análise de políticas públicas com base em pesquisas domiciliares oferece muitas vantagens, mas uma dificuldade normalmente encontrada é a de que nem sempre os números assim estimados são compatíveis com os dados dos registros administrativos. Neste trabalho, desenvolvemos uma metodologia simples para explicar estas diferenças, destacando a importância de possíveis vieses introduzidos pelo desenho amostral dessas pesquisas.

No caso brasileiro, o erro de inferência do número de benefícios do PBF e do BPC nas PNADs tem intrigado pesquisadores e leigos. Para ajudar a explicar esse fenômeno, a metodologia proposta foi então aplicada aos dois programas.

No caso do PBF, concluímos que o desenho amostral da PNAD introduz um viés significativo, responsável por 40% da diferença entre os dados oficiais e os da própria PNAD. Isso ocorre em função da inclusão apenas de um conjunto restrito de municípios – sorteados com probabilidades proporcionais à sua população – para representar todos aqueles que não pertencem a RMs ou ao estrato autorrepresentativo. Além disso, embora não tenhamos realizado testes diretos, encontramos alguns indícios de que a ausência de perguntas específicas sobre o PBF no corpo básico da PNAD e a concentração geográfica de alguns grupos de beneficiários provavelmente também são fatores relevantes.

Quanto ao BPC, um programa que, ao contrário do PBF, independe de cotas municipais, constatamos que o desenho amostral exerce uma influência oposta: caso a captação da PNAD fosse perfeita, o número de benefícios identificados tenderia a ser maior do que o registrado oficialmente.

Como a maioria dos autores convergiu para uma mesma explicação para o erro de inferência, resolvemos também examinar rapidamente a hipótese de que os beneficiários do BPC declaram erroneamente os rendimentos do programa como aposentadorias ou pensões.

Apesar de não ser possível chegar a conclusões definitivas, as evidências sugerem que este fenômeno ocorreu sobretudo no período anterior a 2004, data do primeiro suplemento especial sobre programas de transferência de renda. Nos anos mais recentes, essa tendência aparentemente se enfraqueceu bastante, sem desaparecer por completo. De qualquer maneira, nossas estimativas indicam que, mesmo se atribuirmos todo o "excesso" de benefícios previdenciários de 1 SM na PNAD (760 mil) ao BPC, ainda assim não daríamos conta de um número expressivo de benefícios deste programa (850 mil). Seja como for, há bons motivos para crer que o erro de inferência na estimação do número de beneficiários do BPC e também do PBF poderia ser atenuado caso houvesse perguntas específicas sobre estes programas no questionário principal da PNAD.

Finalmente, não é possível, com os dados disponíveis, concluir em que medida estes erros do PBF e do BPC afetam as análises baseadas na PNAD. Se o perfil dos beneficiários que não são identificados não for radicalmente diferente dos beneficiários identificados, pouca coisa mudaria. O que se pode dizer é que a magnitude dos erros de inferência aconselha certa cautela no uso da PNAD para analisar tanto o BPC quanto o PBF, e chama a atenção para os limites do desenho amostral da PNAD quando o objetivo é estudar programas com forte presença nos menores municípios brasileiros.

REFERÊNCIAS

BARROS, R. P. A efetividade do salário mínimo em comparação à do Programa Bolsa Família como instrumento de redução da Pobreza e da desigualdade. *In*: BARROS, R. P.; FOGUEL, M. N.; ULYSSEA, G. (Org.). **Desigualdade de renda no Brasil**: uma análise da queda recente. Brasília: Ipea, 2006. v. 2.

BARROS, R. P.; CARVALHO, M.; FRANCO, S. O papel das transferências públicas na queda recente da desigualdade de renda brasileira. *In*: BARROS, R. P.; FOGUEL, M. N.; ULYSSEA, G. (Org.). **Desigualdade de renda no Brasil**: uma análise da queda recente. Brasília: Ipea, 2006. v. 2.

COORDENAÇÃO DE TRABALHO E RENDA/IBGE. O Sistema Integrado de Pesquisas Domiciliares (SIPD). ENCONTRO NACIONAL DE ESTUDOS POPULACIONAIS, 16., 2008. **Anais...** Associação Brasileira de Estudos Populacionais (Abep), 2008. Disponível em: <http://www.abep.nepo.unicamp.br/encontro2008/docsPDF/ABEP2008_1903.pdf>

FREITAS, M. P. S. *et al.* **Amostra mestra para o Sistema Integrado de Pesquisas Domiciliares**. Rio de Janeiro: IBGE, 2007 (Textos para Discussão, n. 23).

ROCHA, S. Transferências de renda federais: focalização e impactos sobre pobreza e desigualdade. **Revista de Economia Contemporânea**, v. 12, n. 1, p. 67-97, jan./abr. 2008.

SILVA, P. L. N.; PESSOA, D. G. C.; LILA, M. F. Análise estatística de dados da PNAD: incorporando a estrutura do plano amostral. **Ciência & Saúde Coletiva**, v. 7, n. 4, p. 659-670, 2002.

SOARES, F. V. *et al*. Programas de transferência de renda no Brasil: impactos sobre a desigualdade. *In*: BARROS, R. P.; FOGUEL, M. N.; ULYSSEA, G. (Org.). **Desigualdade de renda no Brasil**: uma análise da queda recente. Brasília: Ipea, 2006. v. 2.

SOARES, S. S. D.; RIBAS, R. P.; SOARES, F. V. **Focalização e cobertura do Programa Bolsa-Família**: qual o significado dos 11 milhões de famílias? Brasília: Ipea, 2009 (Texto para Discussão, n. 1.396).

© Instituto de Pesquisa Econômica Aplicada – **ipea** 2010

EDITORIAL

Coordenação
Cláudio Passos de Oliveira

Supervisão
Andrea Bossle de Abreu

Revisão
Eliezer Moreira
Elisabete de Carvalho Soares
Fabiana da Silva Matos
Gilson Baptista Soares
Lucia Duarte Moreira
Míriam Nunes da Fonseca

Editoração
Roberto das Chagas Campos
Aeromilson Mesquita
Camila Guimarães Simas
Carlos Henrique Santos Vianna
Maria Hosana Carneiro Cunha

Capa
Luís Cláudio Cardoso da Silva

Projeto Gráfico
Renato Rodrigues Bueno

Livraria do Ipea

SBS – Quadra 1 – Bloco J – Ed. BNDES, Térreo.
70076-900 – Brasília – DF
Fone: (61) 3315-5336

Correio eletrônico: livraria@ipea.gov.br
Tiragem: 500 exemplares

Ipea – Instituto de Pesquisa Econômica Aplicada

Secretaria de Assuntos Estratégicos da Presidência da República