

Uma metodologia para explicar diferenças entre dados administrativos e pesquisas amostrais, com aplicação para o Bolsa Família e o Benefício de Prestação Continuada na PNAD*

Pedro Herculano Guimarães Ferreira de Souza**

Estimativas feitas com base em pesquisas domiciliares amostrais muitas vezes diferem bastante dos dados administrativos. Nas PNADs, o número estimado de beneficiários do Programa Bolsa Família (PBF) e do Benefício de Prestação Continuada (BPC) é sempre bem 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.

Palavras-chave: Políticas públicas. Pesquisas amostrais. Dados administrativos.

* O autor agradece os excelentes comentários de Rafael Guerreiro Osório e Sergei Soares e aos dois pareceristas anônimos.

** Mestre em sociologia pelo Instituto Universitário de Pesquisas do Rio de Janeiro (IUPERJ), doutorando em sociologia pela Universidade de Brasília (UnB). Técnico de Pesquisa e Planejamento da Diretoria de Estudos Sociais (Disoc) do Instituto de Pesquisa Econômica Aplicada – Ipea (pedro.ferreira@ipea.gov.br; pedrosouza@gmail.com).

Introdução

Um dos problemas mais comuns na análise de políticas públicas a partir de pesquisas domiciliares amostrais é que as estimativas assim obtidas – diretamente ou por *proxies* – quase sempre diferem dos dados administrativos oficiais. Em boa medida, tais diferenças são esperadas, devido à natureza e aos objetivos de cada fonte de informações, uma vez que registros administrativos são feitos para possibilitar a gestão de programas, auxiliando a tomada de decisões e o controle de sua execução,¹ enquanto pesquisas domiciliares costumam ter finalidades muito mais gerais e um desenho amostral que procura ser representativo de uma população ampla. No caso da Pesquisa Nacional por Amostra de Domicílios (PNAD) – uma das mais importantes pesquisas domiciliares do país, realizada anualmente há mais de três décadas, com representatividade nacional² –, por exemplo, são levantadas informações sobre migração, escolaridade, fecundidade, trabalho, entre outras.

Há casos, no entanto, em que tais diferenças entre registros oficiais e pesquisas domiciliares são de magnitude considerável e se mostram persistentes ao longo do tempo, o que inevitavelmente impõe problemas aos pesquisadores. O primeiro deles, naturalmente, é o de descobrir as causas da divergência, uma vez que elas podem introduzir vieses substantivos nos resultados. Infelizmente, até aqui, basicamente inexistem trabalhos que se dedicam a uma investigação sistemática dessas causas.

No Brasil, há dois casos que chamam a atenção neste sentido: os números de beneficiários do Programa Bolsa Família (PBF) e do Benefício de Prestação Continuada (BPC) estimados – por aproximação – nas PNADs têm sido sistematicamente inferiores aos resultados oficialmente registrados. Para o PBF, o número de famílias beneficiárias na PNAD tem sido em média entre 20% e 25% menor do que o oficial, enquanto no caso do BPC a discrepância é ainda maior, uma vez que o número de beneficiários da PNAD tem sido cerca de metade do oficial (ver, por exemplo, SOARES et al., 2006). Em valores absolutos, isso implica um déficit de cerca de três milhões de famílias no PBF e de 1,5 milhão de beneficiários do BPC.³

O problema ganha contornos ainda maiores quando se considera que o PBF e o BPC são os dois principais programas de assistência social no país, com impactos substanciais sobre a pobreza e a desigualdade (SOARES et al., 2010; HOFFMANN, 2010), com gastos somados de quase R\$ 35 bilhões em 2010, o que representou quase 1% do PIB. Além disso,

¹ Sobre a evolução recente dos registros administrativos no Brasil e seu potencial uso em pesquisas acadêmicas, ver Senra (1996) e Ferreira (2008).

² A PNAD teve sua primeira edição em 1967 e, depois de um período de muitas mudanças – inclusive com a interrupção da pesquisa para a realização do Estudo Nacional sobre Despesa Familiar (Endef), em 1974 e 1975 –, assumiu seu formato definitivo a partir de 1976. Desde então, houve apenas a continuação da expansão da cobertura geográfica para algumas áreas residuais, o que se completou em 2004 com a inclusão também das áreas rurais da Região Norte. Para uma recapitulação da origem e dos primeiros anos da PNAD, ver Martine et al. (1988).

³ Para mais detalhes sobre a gênese, evolução e consolidação do PBF e do BPC, ver, respectivamente, Sposati (2010) e Ipea (2009).

ambos os programas contam com registros administrativos eficientes e grande transparência, publicamente disponíveis e atualizados com presteza.⁴

O objetivo deste texto, portanto, é propor uma metodologia nova e simples para identificar uma possível e importante fonte de tais diferenças: o desenho amostral das pesquisas domiciliares. Em seguida, a metodologia proposta será aplicada aos casos do PBF e do BPC na PNAD. Os resultados mostram que, para o PBF, uma fatia considerável da diferença entre PNAD e dados administrativos decorre da seleção de municípios do desenho amostral da pesquisa, reforçando a necessidade de cautela quando o objetivo é usar a PNAD para estudar programas com forte presença nos menores municípios brasileiros. Para o BPC, por outro lado, não há influência do desenho amostral, mas sim problemas na captação de beneficiários dentro dos estratos sorteados.

Com isso, não se pretende 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. Espera-se, no entanto, que a metodologia proposta ajude a dirimir algumas das muitas dúvidas existentes sobre o assunto e que possa ser aplicável também em outros casos.

Metodologia

As diferenças verificadas entre as estimações de pesquisas amostrais e os dados administrativos podem decorrer de muitas causas específicas aos programas em questão. Com efeito, como será visto nas próximas seções, as hipóteses levantadas para o caso do PBF e do BPC na PNAD, em geral, estão intimamente ligadas a detalhes institucionais e operacionais dos programas. Contudo, não se pode descartar *ex ante* a possibilidade de existência de causas mais gerais, comuns a todas as estimativas feitas com base em uma mesma pesquisa. Os trabalhos existentes dedicam pouca atenção a essa possibilidade, enquanto a metodologia proposta aqui faz o oposto, uma vez que tenta quantificar o quanto da diferença observada decorre do próprio desenho amostral das pesquisas.

As origens do problema são fáceis de entender. Em levantamentos censitários, todas as unidades de um universo de interesse são pesquisadas: por exemplo, todos os indivíduos de determinado país. Como os custos de tais levantamentos 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 total. O método mais intuitivo é o de sorteio aleatório simples. Mas, quando a população é composta por subgrupos muito heterogêneos entre si, amostras estratificadas são vantajosas. Para isso, divide-se a população em estratos relativamente homogêneos e mutuamente exclusivos e, em seguida, selecionam-se os casos amostrados dentro de cada estrato.

⁴ A Matriz de Informações Sociais, disponível no *site* da Secretaria de Avaliação e Gestão da Informação (Sagi), do Ministério do Desenvolvimento Social (MDS), apresenta dados municipalizados do PBF e do BPC, entre outros programas, desde janeiro de 2004: <http://aplicacoes.mds.gov.br/sagi/mi2007/tabelas/mi_social.php>.

A PNAD, por exemplo, utiliza uma amostra estratificada que – simplificando – termina por dividir o país em três estratos ou “áreas censitárias”. O primeiro engloba os municípios pertencentes a dez regiões metropolitanas (RMs), que necessariamente são incluídos na amostra: RMs de Belém, Belo Horizonte, Curitiba, Distrito Federal, Fortaleza, Porto Alegre, Recife, Rio de Janeiro, Salvador e São Paulo. 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 correspondem ao restante do país. Destes, apenas uma pequena fração é selecionada e considerada representativa de todo o conjunto. Ou seja, os municípios menores têm menor probabilidade de inclusão. Neste trabalho, chamaremos de NAR-I os municípios não autorrepresentativos incluídos na PNAD e de NAR-NI os não autorrepresentativos não incluídos. O sorteio dos municípios NAR é feito de acordo com probabilidades proporcionais à sua população, ou seja, os municípios maiores têm maior probabilidade de inclusão (para mais detalhes, ver SILVA et al., 2002).

O pressuposto básico é de que os registros administrativos estão corretos, o que é bastante razoável nos casos em que refletem diretamente a folha de pagamento dos programas, como no PBF e BPC. 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:

- *viés de representatividade* – é possível que o desenho amostral da pesquisa não seja representativo do total da população no que diz respeito ao programa em questão. Afinal, a PNAD não é desenhada para ser especificamente representativa para o PBF, o BPC ou qualquer outro programa governamental específico.⁵ 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 apenas no estrato de municípios não autorrepresentativos (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;
- *viés de captação* – a amostra é efetivamente representativa de 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 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 se pretende pesquisar está fortemente concentrado em pequenos bairros, é provável que ele esteja sub-representado na amostra. Por outro lado, este viés também pode ocorrer nos casos em que os indivíduos são de fato sorteados, mas, propositalmente

⁵ Sobre o plano amostral das PNADs, ver Silva et al. (2002).

ou não, não informam 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. Finalmente, há um tipo de erro de captação que decorre especificamente do fato de que, como o questionário básico da PNAD não pergunta diretamente sobre a participação no PBF e no BPC, a identificação dos beneficiários se dá por inferência a partir dos valores declarados nos “outros rendimentos”. No entanto, os anos com os suplementos especiais sobre transferência de renda (2004 e 2006) não revelam padrões muito discrepantes dos demais e o método de identificação via valores típicos parece ser bastante robusto (BARROS et al., 2006; SOARES et al., 2006). Para os propósitos deste artigo, esse último tipo de erro é menos importante.

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.

Para quantificar cada um desses elementos, no entanto, é preciso calcular primeiro as taxas de representatividade e de captação. A primeira é calculada apenas com base nos registros administrativos, sendo a razão entre o número de benefícios *per capita* nos estratos constitutivos da pesquisa domiciliar e aquele constante nos estratos “reais”:

$$TxRpr_{\text{estratoREAL}} = \frac{BenefPC(RA_{\text{estratoPD}})}{BenefPC(RA_{\text{estratoREAL}})}$$

Para ilustrar, no caso da PNAD existem três estratos na pesquisa (RM, AR e NAR) e quatro estratos “reais” (RM, AR, NAR-I e NAR-NI), uma vez que um grupo de municípios (NAR-NI) não faz parte da pesquisa, sendo “representado” por outro grupo (NAR-I). Assim, a taxa de representatividade para o estrato “real” NAR-NI é dada pela razão observada nos registros administrativos entre o número de benefícios *per capita* do estrato NAR-I e do estrato NAR-NI:

$$TxRpr_{\text{NAR-NI}} = \frac{BenefPC(RA_{\text{NAR-I}})}{BenefPC(RA_{\text{NAR-NI}})}$$

Já a taxa de captação é calculada a partir da razão entre o número de benefícios *per capita* na pesquisa domiciliar e nos registros administrativos para aquele estrato:

$$TxCap_{\text{estratoREAL}} = \frac{BenefPC(PD_{\text{estratoPD}})}{BenefPC(RA_{\text{estratoPD}})}$$

Na PNAD, por exemplo, a taxa de captação para o estrato “real” NAR-NI é calculada por:

$$TxCap_{\text{NAR-NI}} = \frac{BenefPC(PD_{\text{NAR-I}})}{BenefPC(RA_{\text{NAR-I}})}$$

A partir das taxas de representatividade e captação e dos valores observados nos registros administrativos, é possível obter o número de beneficiários estimado pela pesquisa para um dado estrato “real”:

$$Benef(PNAD_{\text{estratoREAL}}) = TxRpr_{\text{estratoREAL}} * TxCap_{\text{estratoREAL}} * Benef(RA_{\text{estratoREAL}})$$

Obter o número dos benefícios do estrato “real” na PNAD, de acordo com a equação acima, pode parecer contraintuitivo, uma vez que o estrato NAR-NI não está incluído na pesquisa, que conta apenas com o estrato NAR-I. Neste caso – e também em algumas das

tabelas aqui apresentadas, nas quais indicamos o número estimado de benefícios na PNAD para os estratos NAR-I e NAR-NI –, a solução é bem simples: como a PNAD supõe que os municípios NAR são representativos tanto dos NAR-I quanto dos NAR-NI, basta distribuir os números da pesquisa proporcionalmente à população dos dois estratos, obtida a partir das estimativas de população publicadas anualmente pelo IBGE. Por exemplo, em 2006, os dados do IBGE indicam que os municípios do estrato NAR-I e aqueles do estrato NAR-NI representavam, respectivamente, 24% e 76% do total do estrato NAR. Assim, os números da PNAD para o estrato NAR foram sempre divididos proporcionalmente.

A partir das taxas de representatividade e de captação, é possível quantificar – em números absolutos – o viés de representatividade e o de captação, além do efeito composto deles:

$$Erro_{estratoREAL} = VRpr_{estratoREAL} + VCap_{estratoREAL} + (1 - TxRpr_{estratoREAL}) * (1 - TxCap_{estratoREAL}) * RA_{estratoREAL}$$

Onde:

$$VRpr_{estratoREAL} = (1 * TxCap_{estratoREAL} * RA_{estratoREAL}) - PD_{estratoPD}$$

$$VCap_{estratoREAL} = (TxRpr_{estratoREAL} * 1 * RA_{estratoREAL}) - PD_{estratoPD}$$

Em outras palavras, $VRpr_{estratoREAL}$ é a diferença entre o número de beneficiários que seria obtido caso só houvesse problemas de captação e o número estimado pela pesquisa domiciliar; $VCap_{estratoREAL}$ é a diferença entre o número de beneficiários que seria obtido caso só houvesse problemas de representatividade e o número estimado pela pesquisa domiciliar;⁶ $(1 - TxRpr_{estratoREAL}) * (1 - TxCap_{estratoREAL}) * RA_{estratoREAL}$ é a interação entre ambos os efeitos, correspondendo ao erro adicional decorrente da existência de problemas de captação em locais onde há problemas de representatividade.

No caso da PNAD, esta decomposição é particularmente útil porque o desenho amostral atual da pesquisa mantém a seleção de municípios constante entre Censos. Em outras palavras, os mesmos 851 municípios selecionados para a PNAD 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 na PNAD 2006.

⁶ O termo “1” na fórmula dos vieses de representatividade e captação foi incluído apenas para deixar claro que cada um é calculado como se apenas o outro problema existisse: por exemplo, no caso do viés de representatividade, o número “1” indica que a taxa de representatividade é igual a 1 e, portanto, não há viés possível. Caso o termo fosse substituído pelo seu valor real, a expressão entre parênteses retornaria, nos dois casos, o número de benefícios estimado pela pesquisa domiciliar e, portanto, cada viés seria igual a 0 após a subtração.

TABELA 1
Número de municípios e população, segundo estratos
Brasil – 2006

Inclusão na PNAD	Estratos	Número de municípios	População (em milhões)	%
Sim	RM	173	58,2	31,0
	AR	134	39,6	21,0
	NAR-I	544	21,4	11,0
Não	NAR-NI	4.714	67,6	36,0
Total		5.565	186,8	100,0

Fonte: IBGE. Pesquisa Nacional por Amostra de Domicílios – PNAD. Estimativas de população do IBGE para 1º de julho de 2006 (*Diário Oficial da União*, 31/08/2006).

Os erros de inferência do PBF na PNAD

A identificação dos benefícios e dos beneficiários do Programa Bolsa Família, na PNAD, não pode ser feita diretamente, uma vez que o questionário básico da pesquisa não conta com nenhuma pergunta específica sobre o programa. Em 2004 e 2006, no entanto, a pesquisa contou com suplementos especiais sobre programas de transferência de renda que perguntavam explicitamente se a família era beneficiária do PBF. O valor recebido, todavia, não era discriminado: nestes, assim como nos outros anos, as transferências declaradas do PBF foram registradas na rubrica “outros rendimentos” (variável v1273), uma categoria residual que em tese também abarca rendimentos de juros, dividendos e outros.⁷

Assim, para usar informações relativas ao PBF em outros anos que não 2004 e 2006 e, mesmo nesses anos, caso seja desejável separar os rendimentos decorrentes do programa, é preciso recorrer a alguma metodologia de desagregação. A escolha óbvia é a de contar como renda do PBF os valores declarados em “outros rendimentos”, que correspondem ou se aproximam dos valores efetivamente pagos pelo programa. Isso é possível porque o número de valores é razoavelmente pequeno: basta fazer as combinações possíveis entre o benefício fixo (R\$ 68 no final de 2010), os benefícios variáveis para crianças (R\$ 22 por criança no final de 2010, com limite de até três por família) e os benefícios variáveis para jovens e adolescentes (R\$ 33 por jovem no final de 2010, com limite de até dois por família). Esse é o chamado método dos “valores típicos”, desenvolvido e validado por Barros (2006) e Barros et al. (2006).

Já Soares et al. (2006) propõem uma variação deste método para os anos em que houve suplemento. Neste caso, os autores consideraram que, entre as famílias que declararam receber PBF, uma vez retirado o BPC (ver adiante) dos “outros rendimentos”, a parcela remanescente, até um valor máximo de um salário mínimo, seria considerada renda do PBF.

Outra variação do método dos “valores típicos”, finalmente, foi empregada por Soares et al. (2010): neste caso, como os autores trabalhavam com anos em que não houve

⁷ Duas excelentes análises sobre a forma de captação dos rendimentos nas PNADs de 1980 e 1990 estão em, respectivamente, Médici (1988) e Rocha (2003).

suplemento, a renda dos “outros rendimentos” inferior ao teto do programa a cada ano foi considerada proveniente do PBF.

Apesar da variedade de abordagens, tanto com o suplemento quanto com as diversas variações do método de valores típicos, o número de famílias beneficiárias do PBF na PNAD é sempre bastante inferior ao dos registros administrativos, como se vê na Tabela 2. Salvo para 2007, ano que foi surpreendentemente atípico, os dados oficiais são em média 20% maiores do que na PNAD. Trata-se de um problema crônico, que vai na contramão das expectativas iniciais de alguns autores, que imaginaram que os erros seriam temporários. 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
Benefícios do PBF, por fonte de informações
Brasil – 2006-2009

Período	Registros administrativos (em milhões) (a)	PNAD (em milhões) (b)	Método (1)	(b)/(a) (%)
Set./2006	11,018	8,142	Suplemento	74,0
Set./2006	11,018	9,421	Valor típico	86,0
Set./2007	10,961	7,198	Valor típico	66,0
Set./2008	10,843	8,924	Valor típico	82,0
Set./2009	11,994	9,352	Valor típico	78,0

Fonte: Ministério do Desenvolvimento Social e Combate à Fome/Secretaria de Avaliação e Gestão da Informação – MDS/SAGI; IBGE. Microdados da Pesquisa Nacional por Amostra de Domicílios – PNAD.

(1) 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 et al. (2006).

Dada a continuidade do problema, novas hipóteses foram levantadas na literatura sobre o PBF. Uma delas 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 este 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, os seus titulares não recebiam de fato o benefício (SOARES et al., 2009).

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 et al., 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 este foi um ano atípico ou se a consolidação e 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 2006, ano em que, graças ao suplemento especial, a incerteza acerca da identificação dos beneficiários é menor.⁸ Observa-se, neste caso, que a concentração geográfica dos beneficiários provavelmente também desempenha 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
Benefícios do PBF, por fonte de informações, segundo estratos de municípios
Brasil – 2006

Estratos de municípios	Registros administrativos		PNAD		(b)/(a)	
	Benefícios (em milhões)	Benefícios <i>per capita</i> (a)	Benefícios (em milhões)	Benefícios <i>per capita</i> (b)		
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)	1,487	0,070	1,381	0,065	0,928
	NAR-NI (1)	5,948	0,089	4,363	0,065	0,734
Total	11,018	0,059	8,142	0,044	0,739	

Fonte: Ministério do Desenvolvimento Social e Combate à Fome/Secretaria de Avaliação e Gestão da Informação – MDS/SAGI; IBGE. Pesquisa Nacional por Amostra de Domicílios – PNAD 2006.

(1) As proporções dos grupos NAR-I e NAR-NI 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 de 2006.

O método proposto deve ser aplicado aos dados de cada estrato listado na Tabela 3, sendo que, no caso de NAR-NI, a taxa de representatividade é a mesma de 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.

⁸ 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.

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 – que pode decorrer de inúmeras causas que, infelizmente, não são diretamente testáveis –, 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 9,77 milhões, número ainda bastante inferior aos 11,02 milhões oficialmente registrados.

TABELA 4
Decomposição do erro de inferência do PBF na PNAD, segundo estratos de municípios
Brasil – 2006

Em milhões					
Estratos	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	0	0,107
	NAR-NI	1,585	0,337	1,158	0,089
Total N. abs.	2,875	1,628	1,158	0,089	2,875
%	100,0	57,0	40,0	3,0	100,0

Fonte: Ministério do Desenvolvimento Social e Combate à Fome/Secretaria de Avaliação e Gestão da Informação – MDS/SAGI/; IBGE. Pesquisa Nacional por Amostra de Domicílios – PNAD 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.

Finalmente, não deve causar espanto o fato de que a maior parte da subestimação decorra do viés de captação. Em primeiro lugar, como já foi dito, o questionário básico da PNAD não pergunta especificamente sobre a participação no PBF. Além disso, outras pesquisas domiciliares do IBGE, que teoricamente são menos suscetíveis ao viés de representatividade e perguntam diretamente sobre o programa, como o Censo Demográfico e a POF, também subestimam o número de beneficiários do PBF: no Censo 2010, é possível identificar 9 milhões de famílias; na POF 2008-2009, cerca de 8 milhões. Ou seja, nos dois casos, há uma disparidade de algo entre 2 e 3 milhões de famílias em relação aos registros administrativos. Esses números são até maiores do que os do viés de captação na PNAD, mas não é aconselhável compará-los diretamente, pois as

três pesquisas apresentam inúmeras diferenças de amostra, temas, questionário e até treinamento dos entrevistadores.

De qualquer modo, é de esperar que o viés de representatividade seja bastante atenuado nas novas PNADs Contínuas, cujos resultados devem começar a ser divulgados até 2014, afinal, seu plano amostral já terá como base, assim como a POF 2008-2009, a “Amostra Mestra” do IBGE. Por exemplo, a previsão é de que as PNADs Contínuas visitem mais 3.300 municípios, contra apenas 851 daquelas realizadas de 2001 a 2009 (FREITAS et al., 2007; COORDENAÇÃO DE TRABALHO E RENDIMENTO/IBGE, 2008).

Por enquanto, o que se pode dizer é que não há indícios de que o viés de representatividade afete as principais conclusões sobre o PBF derivadas das PNADs, mas é preciso esperar os resultados das PNADs Contínuas para podermos avaliar melhor o efeito de tais mudanças no desenho amostral.

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. Por um lado, isso surpreende, pois tal benefício é teoricamente mais fácil de captar tanto por pagar um valor fixo único e de maior vulto – um salário mínimo – 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. Por outro lado, no entanto, é preciso levar em conta que o menor número de beneficiários do programa torna sua estimação menos precisa em uma pesquisa como a PNAD, cujos objetivos são bem mais amplos: como consta na documentação da própria pesquisa, quanto menor o número absoluto estimado, maior o coeficiente de variação em termos percentuais, o que indica maior incerteza na estimativa.

Na Tabela 5, observa-se 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 2006.

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 (SOARES et al., 2006; BARROS et al., 2006; ROCHA, 2008). Dada a popularidade desta 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, significando que o benefício tenderia a estar sobre-estimado 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 5
Benefícios do BPC, por fonte de informações
Brasil – 1996-2009

Anos	Registro administrativo (em milhões) (a) (1)	PNAD (em milhões) (b)	Método	(b)/(a) (%)
1996	0,346	0,028	Valor típico	8,0
1997	0,646	0,062	Valor típico	10,0
1998	0,848	0,113	Valor típico	13,0
1999	1,033	0,240	Valor típico	23,0
2001	1,339	0,175	Valor típico	13,0
2002	1,561	0,381	Valor típico	24,0
2003	1,701	0,289	Valor típico	17,0
2004	1,984	0,925	Valor típico	45,0
2005	2,208	1,198	Valor típico	53,0
2006	2,430	1,588	Valor típico	64,0
2007	2,618	1,298	Valor típico	48,0
2008	2,862	1,474	Valor típico	50,0
2009	3,085	1,570	Valor típico	51,0
2004	1,984	0,901	Suplemento	45,0
2006	2,430	1,318	Suplemento	54,0

Fonte: Ministério da Previdência Social. Anuário Estatístico de Previdência Social (AEPS/Infologo); IBGE. Pesquisa Nacional por Amostra de Domicílios – PNAD 1996-2008.

(1) Para 1996-2003, os dados referem-se ao mês de dezembro; no período 2004-2009 a informação é para o mês de setembro (mês da realização das PNADs).

TABELA 6
Benefícios do BPC, por fonte de informações, segundo estratos de municípios
Brasil – 2006

Estratos de municípios	Registros administrativos		PNAD		(b)/(a) (%)	
	Benefícios (em milhões)	Benefícios per capita (a)	Benefícios (em milhões) (1)	Benefícios per capita (b)		
RM	0,636	0,011	0,210	0,004	33,0	
AR	0,531	0,013	0,240	0,006	45,0	
NAR	NAR-I	0,333	0,016	0,183	0,009	55,0
	NAR-NI	0,931	0,014	0,580	0,009	62,0
Total	2,430	0,013	1,213	0,007	50,0	

Fonte: Ministério da Previdência Social. Anuário Estatístico de Previdência Social (AEPS/Infologo); IBGE. Pesquisa Nacional por Amostra de Domicílios – PNAD 2006.

(1) As proporções dos grupos NAR-I e NAR-NI 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 de 2006.

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 sobre-estimação 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 BPC identificado seria de 2,550 milhões, 5% a mais do que os 2,430 milhões oficialmente pagos.

TABELA 7
Decomposição do erro de inferência do BPC na PNAD, segundo estratos de municípios
Brasil – 2006

Em milhões					
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	0,149
	NAR-NI	0,352	0,471	-0,659	-0,054
Total N. abs.	1,218	1,337	-0,659	-0,054	1,218
%	100,0	110,0	-5,0	-4,0	100,0

Fonte: Ministério da Previdência Social. Anuário Estatístico de Previdência Social (AEPS/Infologo); IBGE. Pesquisa Nacional por Amostra de Domicílios – PNAD 2006.

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 registros administrativos tenha aumentado ao longo do tempo, caso os beneficiários efetivamente estejam declarando o BPC como aposentadorias ou pensões.

A comparação direta pode ser encontrada na Tabela 8. Restrições quanto à disponibilidade de dados limitam o exercício entre 2001 e 2008. De imediato, percebe-se um comportamento não homogêneo ao longo de todo o período: de 2001 a 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 indicício de que, neste período, há um grau razoável de declarações equivocadas do BPC na PNAD.⁹ 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 1,04.¹⁰ 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.

⁹ No mesmo período, o número de benefícios previdenciários iguais a um salário mínimo aumentou 4%.

¹⁰ Entre 2003 e 2008, o número de benefícios previdenciários iguais a um salário mínimo cresceu 22%.

TABELA 8
Benefícios assistenciais e previdenciários equivalentes a um salário mínimo, por fonte de informações
Brasil – 2001-2009

Anos	Registros administrativos (em milhões) (1)		PNAD (em milhões)		%		
	A	B	C	D	C/A	D/B	(C+D)/ (A+B)
	BPC	Previdência um SM	BPC (2)	Previdência um SM			
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
2009	3,167	14,026	1,570	14,800	50	106	95

Fonte: Ministério da Previdência Social. Anuário Estatístico de Previdência Social (AEPS/Infologo); IBGE. Pesquisa Nacional por Amostra de Domicílios – PNAD 2001-2008.

(1) Dados do mês de dezembro de cada ano; disto decorre a diferença, para 2004-2009, em relação à Tabela 5.

(2) Identificado pelo método dos valores típicos.

Se supusermos que a PNAD deveria ser capaz de estimar o número exato de benefícios previdenciários de um salário mínimo, 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 na PNAD. Não temos como saber se estes benefícios são, de fato, BPCs. Mesmo em caso afirmativo, ainda há razões para suspeitar que a não captação – isto é, indivíduos que recebem o BPC e não são entrevistados ou não o declaram nem como “outros rendimentos” nem como aposentadoria ou pensão – permanece sendo um problema relevante.

A conta é simples: a PNAD 2008 identifica metade dos 2,934 milhões de BPCs existentes. Pelo seu desenho amostral, como visto, caso a captação fosse perfeita, o número da PNAD deveria ser ainda superior ao oficial. Mantendo a cobertura de 2001, observa-se 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. Considerando-se o fato de que a captação perfeita da PNAD aparentemente sobre-estimaria 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 um salário mínimo 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 mil, contra 760 mil).

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, apresentou-se uma metodologia simples para explicar parte destas diferenças, destacando a importância de possíveis vieses introduzidos pelo desenho amostral destas pesquisas.

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

No caso do PBF, é possível concluir 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 regiões metropolitanas ou ao estrato autorrepresentativo.

Quanto ao BPC, um programa que, ao contrário do PBF, independe de cotas municipais, constatou-se que o desenho amostral exerce 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, examinou-se também 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, as estimativas indicam que, mesmo atribuindo todo o “excesso” de benefícios previdenciários de um salário mínimo na PNAD (760 mil) ao BPC, ainda assim faltaria um número expressivo de benefícios deste programa (850 mil). Seja como for, é possível que o erro de inferência na estimação do número de beneficiários do BPC e do PBF fosse atenuado caso houvesse perguntas específicas sobre estes programas no questionário principal da PNAD. Nesse sentido, o ideal seria realizar testes-piloto ou pesquisas cognitivas para avaliar tal hipótese.

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 daqueles 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 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.; 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. (Orgs.). **Desigualdade de renda no Brasil: uma análise da queda recente**. Brasília: Ipea, v. 2, 2006, p. 41-86.
- COORDENAÇÃO DE TRABALHO E RENDA/IBGE. O Sistema Integrado de Pesquisas Domiciliares – SIPD. In: XVI ENCONTRO NACIONAL DE ESTUDOS POPULACIONAIS. **Anais...** Abep, 2008. Disponível em <http://www.abep.nepo.unicamp.br/encontro2008/docsPDF/ABEP2008_1903.pdf>. Acesso em: 12 ago. 2010.
- FERREIRA, F. P. M. Registros administrativos como fonte de dados estatísticos. **Informática Pública**, ano 10, n. 1, p. 81-93, 2008.
- FREITAS, M. P. S.; LILA, M. F.; AZEVEDO, R. V.; ANTONACI, G. A. **Amostra Mestra para o Sistema Integrado de Pesquisas Domiciliares**. Rio de Janeiro: IBGE, 2007 (Textos para discussão, n. 23). Disponível em: <http://www.ibge.gov.br/home/estatistica/indicadores/sipd/texto_discussao_23.pdf>. Acesso em: 12 ago. 2010.
- HOFFMANN, R. The evolution of income distribution in Brazil: what promotes and what restricts the decline in inequality. In: CONFERENCE A COMPARATIVE ANALYSIS OF GROWTH AND DEVELOPMENT: ARGENTINA AND BRAZIL. University of Illinois, 22-23 April 2010.
- MARTINE, G.; CAMARANO, A. A.; FLETCHER, P.; NEUPERT, R. A PNAD: notas para uma avaliação. In: SAWYER, D. (Org.). **PNADs em foco: anos 80**. Belo Horizonte: Abep, 1988, p. 281-305.
- MÉDICI, A. C. A mensuração da subjetividade: notas sobre a variável renda nas PNADs. In: SAWYER, D. (Org.). **PNADs em foco: anos 80**. Belo Horizonte: Abep, 1988, p. 121-151.
- ROCHA, S. A investigação da renda nas pesquisas domiciliares. **Economia e Sociedade**, v. 12, n. 2, julho/dezembro, p. 205-224, 2003.
- _____. Transferências de renda federais: focalização e impactos sobre pobreza e desigualdade. **Revista de Economia Contemporânea**, v. 12, n. 1, janeiro/abril, p. 67-97, 2008.
- SENRA, N. C. A questão dos registros administrativos vis-à-vis a geração de estatísticas. **Revista Brasileira de Estudos de População**, v. 13, n. 2, p. 199-205, 1996.
- 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.
- SPOSATI, S. Bolsa Família: um programa com futuro(s). In: CASTRO, J. A.; MODESTO, L. **Bolsa Família 2003-2010: avanços e desafios**. Brasília: Ipea, v. 2, 2010, p. 273-306.
- SOARES, F. V.; SOARES, S. S. D.; MEDEIROS, M.; OSORIO, R. G. Programas de transferência de renda no Brasil: impactos sobre a desigualdade. In: BARROS, R. P.; FOGUEL, M. N.; ULYSSEA, G. (Orgs.). **Desigualdade de renda no Brasil: uma análise da queda recente**. Brasília: Ipea, v. 2, 2006, p. 87-129.
- 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. 1396). Disponível em: <http://www.ipea.gov.br/sites/000/2/publicacoes/tds/td_1396.pdf>. Acesso em: 12 ago. 2010.
- SOARES, S. S. D.; OSORIO, R. G.; SOUZA, P. H. G. F.; SILVEIRA, F. G. Os impactos do benefício do Programa Bolsa Família sobre a desigualdade e a pobreza. In: CASTRO, J. A.; MODESTO, L. **Bolsa Família 2003-2010: avanços e desafios**. Brasília: Ipea, p. 25-52, 2010.

Abstract

A methodology for explaining differences between administrative data and findings from sample-based surveys regarding the Family Allowance Program and the Continuous Cash Benefit Program

Estimates based on household surveys often differ considerably from administrative records. In the National Household Sample Surveys (PNADs) of the Brazilian Statistics Department, the estimated number of beneficiaries of the Family Allowance (Bolsa Família) Program (PBF), and of beneficiaries of the Continuous Cash Benefit (Benefício de Prestação Continuada) Program (BPC) is always lower than official figures. This paper presents a simple methodology, based on the sampling design of household surveys, to explain these differences, by decomposing them into three terms: the representativeness bias (derived from the choice of places, census tracts or municipalities for the survey); the data collection bias (derived from data collection problems at the chosen sites); and the interaction between them. The application of this methodology to the present cases shows that the representativeness bias of the PBF accounts for 40% of the difference between official records and the PNAD. For the BPC, the representativeness bias tends to act in the opposite direction. That is, in other words, the data collection bias is entirely responsible for the observed difference. Also, the erroneous reporting of the BPC as a Social Security benefit on the PNAD seems to have occurred mostly in the years prior to 2004, and does not entirely explain the low number of beneficiaries identified in the PNADs.

Keywords: Public policies. Sample-based surveys. Administrative data.

Resumen

Una metodología para explicar diferencias entre datos administrativos y encuestas por muestreo, con aplicación para el Bolsa Familia y el Beneficio de Prestación Continuada en la PNAD

Estimaciones realizadas en base a encuestas de hogares por muestreo muchas veces difieren bastante de los datos administrativos. En la Pesquisa Nacional por Amostra de Domicílios - PNAD, el número estimado de beneficiarios del Programa Bolsa Familia (PBF) y del Beneficio de Prestación Continuada (BPC) siempre es bastante inferior al número oficial. El objetivo de este artículo es presentar una metodología simple, basada en las características del diseño por muestreo de las encuestas de hogares, para explicar esta diferencia, descomponiéndola en tres aspectos: el factor de representatividad (derivado de la elección de los sitios, áreas censitarias o municipios para la encuesta); el factor de captación (resultante de problemas de captación en los sitios seleccionados); y la interacción entre ambos. La aplicación de dicha metodología al PBF y al BPC muestra que, en el primer caso, el sesgo de representatividad explica una gran parte del problema: la selección de municipios investigados es la responsable por un 40% de la diferencia observada entre los datos oficiales y los de la PNAD. En el caso del BPC, el sesgo de representatividad tendería a actuar en el sentido opuesto. Por lo tanto, el sesgo de captación es totalmente responsable por la diferencia observada. Además, la declaración equivocada del BPC como beneficio de la previsión social en la PNAD parece ocurrir sobre todo en el periodo anterior a 2004 y no explica por completo el pequeño número de beneficiarios identificados en las PNADs.

Palabras clave: Políticas públicas. Encuestas por muestreo. Datos administrativos.

Recebido para publicação em 05/03/2011

Aceito para publicação em 16/06/2011

