

ASPECTOS CONCEITUAIS E METODOLÓGICOS DA AVALIAÇÃO DE POLÍTICAS E PROGRAMAS SOCIAIS

Marília Ramos*

O foco deste artigo são os aspectos conceituais e principalmente metodológicos relacionados com a avaliação de políticas e programas sociais. São apresentados o desenho de uma pesquisa de avaliação; os procedimentos (econômicos) necessários para que uma política possa ser objetivamente avaliada e alguns dos recursos metodológicos e estatísticos necessários para que possamos ter os resultados mais fidedignos e válidos possíveis. São descritos e explicados os principais aspectos a serem considerados no processo de avaliação de uma política ou programa social: conhecer em detalhe o programa que se quer avaliar, saber sobre seus objetivos, o formato (em termos de desenho de pesquisa, se foi um experimento ou não) e possuir uma vasta gama de informações sobre o programa, bem como sobre os atingidos por ele. Também devem constar no banco de dados características (heterogeneidade do universo a ser analisado) daqueles não atingidos pelo programa. Outras variáveis necessárias no banco dizem respeito aos aspectos (características dos pesquisados) que influenciam a participação no programa, mas que não afetam diretamente os resultados dele. Outro aspecto seria a identificação de possíveis fontes de vieses: o fato de os grupos (atingidos e não-atingidos pela política) não serem comparáveis. Neste caso, a seleção cuidadosa do grupo comparativo (não-atingido pela política) pode eliminar aquele viés, e esta seleção pressupõe a escolha do grupo comparativo (contrafactual) com a mesma distribuição de características observáveis que o grupo tratado (atingido pela política). Dentre os métodos para lidar com o viés da falta de comparabilidade dos grupos, três soluções são apresentadas, através de exemplos concretos, neste artigo: a técnica da diferença da diferença, a técnica do escore propensão para pareamento (*propensity score matching*) e, por fim, o procedimento que utiliza variáveis instrumentais.

Palavras-chave: Avaliação; Políticas Públicas; Técnicas de Avaliação; Resultados de Políticas.

CONCEPTUAL AND METHODOLOGICAL ASPECTS RELATED TO THE POLICY AND PROGRAM EVALUATION

The subject of this article is the set of methodological and conceptual aspects related to the evaluation of social policies and programs. The design of a evaluation research as well as the objective procedures to evaluate the results from a social policy or program are presented. Specifically a set of criteria to evaluate a social policy are presented: the knowledge about the policy/program, the knowledge about its goals, the policy design (if it is an experimental design or not) and the need to have a broad data set about the information related to the subjects involved and not involved in the policy/program. Another set of important variables that should be included in the data set are the characteristics of the subjects that affect the participation in the policy/program but do not affect the results. Another aspect presented in the article is the fact that the groups (involved and not by the policy/program) should be comparable. It is necessary to choose a comparable control group to avoid bias. Several techniques to avoid or to lead

* Professora e Pesquisadora do Programa de Mestrado e Doutorado em Desenvolvimento Regional da Universidade de Santa Cruz do Sul (Unisc).

with bias due to the group (control and treatment) selection are presented: the technique of difference of difference, the propensity score matching and the use of instrumental variables.

Key words: Policy Evaluation; Policy Results; Techniques of Evaluation.

ASPECTOS CONCEPTUALES Y METODOLÓGICOS RELACIONADOS CON LA EVALUACIÓN DE POLÍTICAS Y PROGRAMAS SOCIALES

El foco de este artículo se concentra en los aspectos conceptuales y principalmente metodológicos relacionados con la evaluación de políticas y programas sociales. Se presentan aquí: el diseño de una investigación de evaluación, los procedimientos econométricos necesarios para que una política pueda ser objetivamente evaluada, así como algunos recursos metodológicos y estadísticos necesarios para obtener los resultados más fieles y válidos posibles. Son descritos y explicados los principales aspectos que deben ser considerados en el proceso de evaluación de una política o programa social, a saber: conocer en detalle el programa que se quiere evaluar, sus objetivos, el formato (en términos de si diseño de investigación fue un experimento o no) y poseer una vasta gama de informaciones sobre el programa, así como sobre los individuos afectados por el. También deben constar en el banco de datos las características de los no afectados por el programa. Otras variables necesarias en el banco se relacionan con los aspectos que afectan la participación en el programa, pero que no afectan directamente sus resultados. Otro aspecto incluye la identificación de posibles fuentes de tendenciosidad, tal como que el hecho de que los grupos afectados y no afectados por la política no sean comparables. En este caso, la selección cuidadosa del grupo comparativo (no afectado por la política) puede eliminar aquella tendenciosidad y esta selección presupone la escogencia del grupo comparativo con la misma distribución de características observables del grupo tratado (afectado por la política). Entre los métodos para tratar la tendenciosidad de falta de comparabilidad entre los grupos, se presentan tres soluciones en este artículo: la técnica de la llamada "diferencia en la diferencia", la técnica del "escor de propensión para pareamiento (*propensity score matching*) y el procedimiento que utiliza variables instrumentales.

Palabras-clave: Evaluación de Políticas; Resultados de Políticas; Tecnicas de Evaluación.

ASPECTS CONCEPTUELS ET ASPECTS MÉTHODOLOGIQUES LIÉS À L'ÉVALUATION DE POLITIQUES ET DES PROGRAMMES SOCIAUX

Les grandes lignes de cet article concernent les aspects conceptuels et principalement les aspects méthodologiques liés à l'évaluation de politiques et des programmes sociaux. On présente un schéma d'une recherche d'évaluation: quelles sont les procédures (économétriques) nécessaires pour qu'une politique puisse être évalué objectivement et quelles sont certaines des ressources méthodologiques et statistiques qui sont nécessaires pour avoir des résultats les plus représentatifs et valables possibles. L'article aborde les principaux aspects à prendre en compte dans le processus d'évaluation d'une politique ou d'un programme social, qui sont : connaître en détail le programme évalué, leurs objectifs, leur format (en termes de dessin de recherche, s'il a été soumis à un traitement expérimental ou non) et obtenir une vaste gamme d'informations sur le programme ainsi que sur les personnes touchés par ce programme. Il faut aussi préciser les caractéristiques (hétérogénéité de l'univers) de ceux qui ne sont pas touchés par le programme dans la banque de données. D'autres variables qui doivent être également prises en

compte dans la banque sont les aspects (caractéristiques des recherchés) qui influencent la participation dans le programme, mais qui ne touchent pas directement les résultats du programme. Un autre aspect est l'identification de possibles difficultés: les données concernant les groupes (touchés et non touchés par la politique) peuvent ne pas être comparables. Dans ce cas, l'adoption d'un choix attentif du groupe comparatif (non touché par la politique) permet d'éliminer cette difficulté, et cette sélection suppose un choix du groupe comparatif (contrefactuel) avec la même distribution de caractéristiques observables que le groupe traité (touché par la politique). Parmi les méthodes pour faire face au problème de l'absence de comparabilité des groupes, on signale particulièrement dans cet article: la technique de la différence dans les différences, la technique de score de propension (propensity score matching) et finalement, la méthode des variables instrumentales.

Mots-clés: L'évaluation de Politiques; Programmes Sociaux; Résultats de Politiques.

1 INTRODUÇÃO

O presente artigo visa discutir aspectos conceituais e principalmente metodológicos relacionados com a avaliação de políticas e programas sociais. Num primeiro momento são expostas questões pertinentes aos objetivos e à necessidade de se fazer a avaliação, enfatizando-se em seguida um conjunto de procedimentos metodológicos relacionados com as técnicas do processo de avaliação.

Cabe destacar que não visamos aqui julgar ou apontar estratégias de avaliação de políticas ou programas sociais, no sentido de definir a melhor técnica ou a mais correta. Visamos basicamente apontar e discutir as condições necessárias para a aplicação de técnicas de avaliação, bem como os pontos fortes e as limitações de algumas delas. Ressaltamos que o foco será a avaliação dos resultados da aplicação de uma política ou programa social, em termos de um ou muitos indicadores, sendo que não é o escopo deste artigo discutir aspectos da avaliação que priorizam a opinião e o juízo dos atingidos por aquelas formas de intervenção na realidade social. Não negamos a importância de ouvirmos aqueles que são diretamente afetados, apenas tomamos como foco os aspectos concernentes à avaliação técnica de resultados.

É importante deixar claro que estamos tratando da *avaliação formal*, a qual consiste do exame sistemático de certos objetos, com base em procedimentos científicos de coleta e análise de informação sobre conteúdo, estrutura, processo, resultados e/ou impactos de políticas, programas, projetos ou quaisquer intervenções planejadas na realidade (RUA, 2000). As definições de avaliação são muitas, mas um aspecto consensual é a sua característica de atribuição de valor. A decisão de aplicar recursos em uma ação pública sugere o reconhecimento do valor de seus objetivos pela sociedade; assim, sua avaliação deve "verificar o cumprimento de objetivos e validar continuamente o valor social incorporado ao cumprimento desses objetivos" (MOKATE, 2002).

Além de ser útil para medir a eficácia da gestão pública, a avaliação também contribui para a busca e obtenção de ganhos das ações governamentais em termos

de satisfação dos usuários e de legitimidade social e política. Por essas e outras razões, tem sido ressaltada a importância dos processos de avaliação para a reforma das políticas públicas, modernização e democratização da gestão pública.

Visando complementar os estudos existentes sobre avaliação, mostramos neste artigo em que consiste um desenho de pesquisa de avaliação; quais os procedimentos necessários para que uma política possa ser objetivamente avaliada; e quais são alguns dos recursos metodológicos e estatísticos necessários para que possamos chegar aos resultados mais fidedignos e válidos possíveis.

2 O PROBLEMA DA AVALIAÇÃO E AS SOLUÇÕES ALTERNATIVAS: A QUESTÃO DA NECESSIDADE DE COMPARAÇÃO

Quando pensamos em avaliar uma determinada política ou um programa social específico, devemos ter claro que as condições para a avaliação já devem estar presentes desde o processo de formulação da referida política/programa. Isto porque precisamos ter informações sobre as condições dos grupos (ou regiões, ou instituições, ou qualquer outra unidade de análise) antes de a política/programa ter sido implantado(a). Sem esta informação se torna praticamente inviável pensarmos em qualquer tipo de avaliação.

Assim, tomamos a possibilidade de acessarmos informações referentes ao resultado pretendido com as referidas políticas, antes de as mesmas terem sido implementadas e após sua realização, como um dos requisitos primeiros para a possibilidade de avaliarmos alguma política ou programa social. Estes dois momentos no tempo (o antes e o depois) são essenciais, pois sem eles não temos como identificar possíveis mudanças que possam ter ocorrido em função daquela intervenção.

Outro pré-requisito crucial em desenhos de avaliação diz respeito à necessidade de um grupo comparativo. Aqui entramos numa das mais complexas discussões presentes no processo de delineamento de uma pesquisa de avaliação. Basicamente, como o leitor já deve ter se dado conta, o desenho de uma pesquisa de avaliação se baseia nas diretrizes de uma pesquisa de caráter experimental,¹ onde informações antes e depois do tratamento são necessárias, bem como a necessidade de pelo menos dois grupos: o que recebe o tratamento (grupo experimental ou, no nosso caso, que foi atingido pela política/programa social) e grupo de controle (aquele que não recebe ou, no caso, não foi atingido pela política/programa social).

1. Este desenho é muito mais usado nas ciências naturais e se constitui num delineamento de pesquisa, no qual o pesquisador deseja verificar uma relação de causa e efeito, em que X seria um tratamento qualquer e Y o efeito do mesmo. Para tanto o pesquisador seleciona aleatoriamente elementos que receberão o tratamento e outros que não. Nas ciências humanas podemos trabalhar com a mesma lógica, porque normalmente não temos como selecionar aleatoriamente aqueles que recebem algum tratamento e aqueles que não, pois isso já deve ser dado naturalmente. Exemplo: não podemos submeter um grupo de jovens ao uso de drogas e outro não, para ver o impacto do uso da mesma na agressividade, por isso precisamos encontrar na sociedade jovens usuários e outros não usuários.

Idealmente falando, se queremos apontar causas, o mais correto seria que pudéssemos comparar o mesmo grupo (ou sujeitos, instituições, regiões) fazendo parte de uma política/programa social e ele mesmo sem ter feito parte. Isto é, o ideal seria ter na pesquisa o sujeito atingido e seu clone exato. Como isso não é possível, um dos pontos mais delicados do desenho de avaliação diz respeito à tentativa de encontrarmos o grupo comparativo ou, como se chama na literatura sobre avaliação, o *contrafactual*.

Cabe retomar aqui de forma breve a explicação metodológica para a necessidade de um grupo comparativo, para deixar claro por que não podemos trabalhar somente com informações referentes ao grupo tratado (que é alvo de uma determinada política ou programa social) antes e depois do tratamento.

Especificamente, quando estamos tentando verificar o impacto de uma política/programa social, estamos lidando com relações de causa e efeito. O que queremos é poder identificar o impacto daquele programa/política em alguma dimensão da vida econômica, social ou política (resultados esperados). É por isso que a lógica por trás das pesquisas de avaliação está inserida na lógica dos experimentos, ainda que nem sempre a aplicação políticas/programas sociais seja feita de forma experimental.

Nesse caso retomamos a discussão, presente na literatura na área metodológica, sobre a necessidade do grupo de controle. Se analisarmos somente o grupo de tratamento (no nosso caso os atingidos pela política/programa social), antes e depois teremos muita dificuldade de separar o impacto daquele programa/política do efeito de outros aspectos que interferem no resultado que está sendo investigado. Este aspecto diz respeito à questão da validade interna da pesquisa. Especificamente devemos isolar, controlar, tudo aquilo que poderia afetar o resultado, mas que não diz respeito ao tratamento que estamos querendo avaliar. Por exemplo, se dizemos que um curso de redação melhora as habilidades de redação dos alunos iniciantes na universidade, devemos demonstrar que não foram outras causas através do tempo que melhoraram tal habilidade, como por exemplo, o fato de os alunos terem se tornado mais velhos e maduros.

Podemos citar aqui algumas variáveis que podem afetar os resultados da aplicação de algum tratamento e que podem atrapalhar nossa avaliação se não as considerarmos:

1) *História* – Existem experiências únicas que os sujeitos têm entre os momentos da pesquisa (antes e após o tratamento). Estas experiências podem afetar as respostas dos sujeitos.

2) *Maturação* – Mudanças que ocorrem naturalmente, dada a passagem do tempo. Por exemplo, quanto mais tempo leva um estudo, mais chances existem de os sujeitos envolvidos ficarem cansados ou chateados, mais ou menos motivados. Ou os sujeitos podem ter envelhecido e por isso suas respostas mudaram.

3) *Teste* – Muitos experimentos realizam um pré-teste com os sujeitos para saberem em que nível eles estão. Uma consequência é que esse pré-teste pode contaminar as respostas futuras ao teste após o tratamento recebido.

4) *Instrumentação* – Mudar os métodos de medida (ou a maneira de administrá-los) pode afetar o que está sendo medido. Por exemplo: se sujeitos humanos são os observadores, pode ser que o julgamento do observador tenha mudado através do tempo e não a performance dos pesquisados.

5) *Regressão em direção à média* – Quando os sujeitos participantes de um estudo são escolhidos porque apresentam escores em alguma variável muito altos ou muito baixos. Ao retestar esses sujeitos, as médias daqueles com altos escores somente poderão baixar, e as daqueles com baixos tenderão a subir. O escore maior pode apenas diminuir e o escore mais baixo, apenas aumentar.

6) *Seleção* – Os grupos devem ser comparáveis. O que garante isso nos experimentos é a aleatoriedade (nem sempre possível).

7) *Mortalidade* – Quando sujeitos saem do estudo. Se, por exemplo, um grupo de comparação tem um alto nível de sujeitos que saíram do experimento ou morreram, em comparação com o outro grupo, as diferenças observadas entre os grupos são questionáveis. Isto é muito comum em estudos através do tempo.

Assim sendo o grupo de controle, ou no caso das pesquisas de avaliação, o contrafactual, é essencial nas pesquisas sobre causas e efeitos (no nosso caso, impacto).

Experimentos sociais se constituem no método mais adequado para estimarmos o impacto de programas ou políticas sociais. O desenho experimental aplicado às ciências sociais resolve o problema da avaliação pelo processo aleatório de designação de quem receberá ou não a intervenção (política). A partir de um conjunto de potenciais participantes, alguns são aleatoriamente selecionados para receber o tratamento, grupo de tratamento e outros para não receber, grupo de controle. Como exemplo, podemos citar programas que visam qualificar profissionais em situação de desemprego ou subemprego. Assim, do conjunto de pessoas nessa situação, sorteia-se (aleatoriamente) quem vai participar de um curso de qualificação, por exemplo, e quem não vai.

O resultado para o grupo de controle (no nosso exemplo, grupo de desempregados que não participou do curso de qualificação) é o contrafactual; este permite identificar o resultado no estado não tratado. O elemento-chave é a aleatoriedade das unidades (no nosso exemplo o sorteio) dentro e fora do tratamento, de forma que os resultados se tornam independentes da seleção ao tratamento. Isto é, a seleção para receber ou não o tratamento é independente de qualquer característica específica, observada ou não, que os grupos possam ter, e isso garante o bloqueio de possíveis vieses de seleção quando comparamos o resultado observado entre o grupo de tratamento e o grupo de controle.

Entretanto, nem sempre é possível realizar experimentos sociais, por uma série de razões, tais como os custos políticos e monetários, a incapacidade de realizar experimentos com um desenho universal de políticas (em que a política ou programa não tem um público alvo, mas vale para todos) e porque o uso de grupos de controle pode levantar discussões éticas.²

Um dos maiores desafios na tentativa de se estimar intervenções de políticas sociais, apresentados na literatura sobre avaliação, é medir o resultado de interesse utilizando-se um contrafactual (grupo controle).

Dado que os possíveis resultados não podem ser observados através de uma única unidade de observação (por exemplo, um mesmo sujeito ter participado e não ter participado), a essência da estratégia de identificação seria identificar os resultados num contrafactual (grupo controle). Os experimentos em que um grupo de unidades elegíveis (indivíduos, domicílios, localidades etc.) são aleatoriamente excluídas do tratamento ou intervenção, proporcionam a estimativa mais clara do resultado no contrafactual e têm sido a referência para se avaliar políticas de intervenção social. Experimentos aleatórios (onde grupos tratamento e controle são definidos aleatoriamente) não são uma panaceia; eles podem prover um estimador do impacto bastante consistente.

Para ilustrarmos a importância da aleatoriedade na seleção dos grupos na pesquisa de avaliação podemos citar o seguinte exemplo: suponhamos que queremos avaliar o impacto de programas de qualificação profissional (por exemplo, cursos profissionalizantes) na empregabilidade. Se os participantes do curso não forem selecionados aleatoriamente entre um grupo de desempregados, mas decidirem por *n* razões procurar o referido curso, pode ser que aqueles que procuraram sejam pessoas com características especiais, as quais afetariam sua capacidade de obter um emprego e aí não conseguiremos isolar o efeito específico do curso na probabilidade de aquisição do emprego. Assim sendo, a aleatoriedade garantiria que teríamos, entre aqueles que participaram do curso e aqueles que não participaram, indivíduos com características equivalentes.

Contudo, uma limitação dos experimentos é que eles podem ser muito custosos para serem implementados e também podem privar certos grupos de receber certos benefícios, o que traz à tona questões de ordem ética. Entretanto, quando são aplicados corretamente, o consenso entre os pesquisadores é de que este método produz estimadores mais acurados para se avaliar o impacto de programas.

Quando os experimentos não são possíveis, os pesquisadores têm de se basear em métodos não experimentais para lidar com problemas de viés de seleção na estimativa

2. Como no exemplo citado na nota 1 relativo à pesquisa sobre o efeito do uso de drogas na agressividade, em que não podemos submeter jovens ao uso de drogas (grupo tratamento) e outros não (grupo controle).

dos impactos dos programas. Muitos modelos estatísticos e econométricos têm sido desenvolvidos para controlar variáveis de confundimento (efeitos misturados) e aspectos relacionados com a seletividade. Estas técnicas requerem a imposição de certos pressupostos que não são testáveis, embora muitas de suas implicações possam ser ou não viáveis com dados disponíveis. Métodos não experimentais podem produzir vieses de autoseleção por causa das diferenças nas fontes e qualidade dos dados.

Infelizmente, os pesquisadores raramente têm a oportunidade de “tornar aleatórias” variáveis como o desempenho educacional, a imigração ou salário mínimo (SM), por exemplo. Muitos pesquisadores precisam confiar em estudos que geralmente falham em produzir a mesma força de evidência de um experimento aleatório (ANGRIST; KRUEGER, 1999). Alguns estudos, baseados em métodos quase-experimentais, trazem dúvidas quanto à credibilidade da pesquisa empírica na economia. Isto ocorre porque é praticamente impossível controlar adequadamente todas as variáveis relevantes.

A partir destas considerações, um aspecto importante seria verificar, quando possível, se métodos não experimentais são bons substitutos para os experimentos aleatórios.

3 SELEÇÃO ALEATÓRIA DOS GRUPOS DE TRATAMENTO (RECEBEDOR DA POLÍTICA OU PROGRAMA SOCIAL) E DE CONTROLE

Quando a aplicação de uma determinada política/programa social diz respeito a um projeto onde os autores da política definem quem participa e quem não participa, isto é, quando os executores decidem, tal como em experimentos clássicos, quem será o alvo e quem não, o critério primeiro para garantir a construção de um grupo de controle (contrafactual) o mais parecido possível com o grupo de tratamento deve ser a aleatoriedade.

A análise causal, portanto, no caso do impacto de uma política ou programa social, está inserida em um problema que, segundo Moffitt (2003), é baseado na noção fundamental do contrafactual de uma unidade de análise, como o indivíduo, o estado ou o país. Esta noção torna-se clara ao considerar a seguinte situação: o indivíduo i tem dois possíveis resultados, $Y1i$ e $Y0i$, em que $Y1i$ é o resultado se o indivíduo experimenta um evento particular ou toma uma determinada decisão, e $Y0i$ é o resultado caso não experimente o evento, mantendo o resto constante. A diferença entre os dois ($Y1i - Y0i$) é o efeito causal do evento ou da ação. Somente um dos resultados pode ser observado, já que o indivíduo não pode fazer duas coisas ao mesmo tempo. O resultado do grupo que não recebeu o tratamento é chamado de contrafactual (MOFFITT, 2003).

Autores como Angrist e Krueger (1998) concordam que o maior desafio da pesquisa empírica, e no nosso caso na avaliação de uma política/programa social,

envolve afirmações sobre respostas contrafactuais (*what if*). Diferentes esferas, como o indivíduo ou o governo, gostariam de saber qual resposta poderia ser observada se a variável fosse manipulada de outra maneira. Na prática, não é sempre claro definir o mundo contrafactual.

Na prática, o pesquisador pode aprender mais sobre os resultados contrafactuais se ele considerar o experimento aleatório. Em muitas pesquisas, principalmente na área médica, acredita-se que a melhor evidência sobre o resultado contrafactual é observada a partir de um experimento aleatório. A falta deste é uma das principais razões por que a pesquisa econométrica sempre convence menos do que a pesquisa em outras ciências mais experimentais (ANGRIST; KRUEGER, 1999). Outros autores, como Rosenzweig e Wolpin (2000), também salientam a importância da realização de experimentos aleatórios, que, junto com outras suposições adicionais, produzem estimativas de interesse geral na economia.

Assim, conforme já mencionado, existem algumas técnicas específicas, que contribuem para que, na impossibilidade da seleção aleatória daqueles que serão atingidos por uma política ou programa social, possamos garantir a comparabilidade. Apresentamos a seguir algumas destas técnicas.

4 O PROCEDIMENTO DA DIFERENÇA DA DIFERENÇA

A estratégia chamada de “diferença da diferença” utiliza dados em painel (através do tempo), aplicados para um conjunto de grupos, nos casos em que alguns desses estão expostos a variáveis causais de interesse e outros não. Um importante componente desta estratégia é a escolha de variáveis para fazer comparações e responder a questões contrafactuais (ANGRIST; KRUEGER, 1999).

Esta técnica se utiliza de informações sobre os participantes e não-participantes, coletadas antes da aplicação da política ou programa social, comparadas com aquelas mesmas informações coletadas após a aplicação da intervenção. Basicamente se subtraem as duas diferenças (antes e depois, e tratamento e controle) ou se usa, numa análise de regressão, uma variável dicotômica para participantes (=1) e outra para o momento posterior à intervenção (= 1). Da interação entre as duas temos a variável que nos dará o efeito do programa.

Tomamos como exemplo um programa de educação remediatória que foi aplicado entre 1994 e 2001 na Índia (Mumbai e Vadadora), chamado Programa *Balsakhi*. Visava melhorar a performance de alguns estudantes com dificuldades em matemática e nas habilidades para leitura. Para tanto foram selecionadas aleatoriamente, entre escolas e estudantes com dificuldades, aquelas que iriam participar do programa e aquelas que não iriam. O Programa *Balsakhi* se constituía da contratação de tutores (os *balsakhis*) que iriam fornecer, às escolas e crianças selecionadas, aulas extras de matemática e leitura.

A equação a seguir ilustra, matematicamente, como podemos identificar o impacto do programa exemplificado nas notas das crianças nas escolas participantes em comparação com aquelas escolas não-participantes. Tomemos a equação:

$$Y = \beta_0 + \beta_1 + \beta_2 + \beta_3 \text{ onde:}$$

Y = escore nos testes após;

β_1 = ano posterior ao programa (=1);

β_2 = participou do programa (=1); e

β_3 = interação ano * participou.

Para identificarmos a “diferença da diferença” substituímos os valores nas variáveis dicotômicas como mostra o esquema a seguir. O valor estimado do Y para os quatro grupos representados nas linhas e colunas da tabela 1 podem ser calculados pela soma dos β s, notando que vários deles são iguais a zero.

Podemos observar na tabela 1 que o β_3 indica a diferença da diferença. Concretamente quando colocamos as informações do projeto exemplificado da tabela 1 numa planilha estatística e efetuamos a análise de regressão, obtemos os resultados da tabela 2:

TABELA 1

Esquema teórico da técnica da diferença da diferença

	Participante	Não-participante	Diferença
Antes	$\beta_0 + \beta_2$	β_0	β_2
Após	$\beta_0 + \beta_1 + \beta_2 + \beta_3$	$\beta_0 + \beta_1$	$\beta_2 + \beta_3$
Diferença	$\beta_1 + \beta_3$	β_1	β_3

Fonte: Elaboração da autora.

TABELA 2

Resultados da análise de regressão que utiliza a técnica da diferença da diferença

Fonte	soma quadrados	gl	MS	Número de obs. = 24208		
Modelo	2796447,76	3	932149,255	F(3,24204) = 1986,70		
Resíduo	11356418,4	24204	469,195934	Prob > F = 0,0000		
Total	14152866,2	24207	584,660063	R-quadrado = 0,1976		
				R-quadrado ajustado = 0,1975		
				Raiz MSE = 21,661		
Teste	Coef.	erro-padrão	t	P> t	[95% Intervalo de conf]	
após	18,1956500	0,3976920	45,75	0,000	7,4161500	18,97515
participante	0,7492239	0,3892857	1,92	0,054	-0,0138002	1,512248
partic*após	5,5745390	0,5571308	10,01	0,000	4,4825280	6,666549
_cons	25,4056900	0,2778870	91,42	0,000	24,8610200	25,95037

Fonte: Tabela obtida pela análise de regressão realizada no *software* Stata versão 9.1. Dados oriundos do Programa *Balsakhi*, fornecidos pelo professor Doutor Chandler Stolp da Universidade do Texas.

Se substituirmos e somarmos os valores encontrados nos coeficientes de regressão calculados e fornecidos pelo *software*, obteremos o esquema da tabela 3:

TABELA 3

Resultados estimados da análise de regressão que utiliza a técnica da diferença da diferença

	Participantes	Não-participantes	Diferença
Antes	26,16	25,41	0,75
Após	49,92	43,61	6,32
Diferença	23,76	18,20	5,57

Fonte: Elaboração da autora, com base nos resultados da análise de regressão apresentados na tabela 2.

Note-se que os valores encontrados dizem respeito aos resultados do cálculo da estimativa de Y . Por exemplo: o valor 26,16 foi obtido somando-se os coeficientes β_0 e β_2 , porque os demais ficam igual a zero, já que neste caso estamos lidando com o período antes ($\beta_1 = 0$) e assim sendo o coeficiente β_3 fica também igual a zero ($0 * 1$).

Podemos observar que, levando-se em conta que as escolas participantes e não-participantes foram aleatoriamente selecionadas, esperávamos que as diferenças entre participantes e não-participantes, antes da realização do programa, não fossem estatisticamente significativas como ilustra o coeficiente β_2 . Entretanto, a diferença apareceu como marginalmente significativa ($P = 0,054$), o que indica que havia diferenças entre os grupos antes de o programa ter sido implantado. Cabe destacar que sempre que encontrarmos diferenças significativas entre os grupos antes de o programa ser implantado, devemos pensar em acrescentar controles em nossos modelos ou nos utilizarmos de técnicas de pareamento tais como score de propensão para pareamento, o qual será discutido na seção seguinte, para tornarmos os grupos comparáveis.

A diferença entre os resultados dos testes antes e depois do programa no grupo das escolas participantes é 23,76 e se manifestou estatisticamente significativa ($P = 0,000$). A diferença entre o momento antes do programa e depois no grupo das escolas não-participantes é 18,20. A diferença entre participantes e não-participantes após o programa é 6,32 e estatisticamente significativa. Isto acontece na medida em que, apesar de as médias das notas dos alunos de ambos os grupos terem crescido após o programa, o crescimento foi maior no grupo das escolas participantes.

Entretanto, para verificarmos o impacto do programa, temos de olhar para a diferença da diferença, que foi igual a 5,57 e é estatisticamente significativa, indicando que o programa teve um impacto nas notas dos alunos das escolas participantes.

Cabem aqui duas observações finais relacionadas com a técnica da diferença da diferença. Uma delas é que esta técnica pode ser substituída por um simples teste

de diferença de médias entre participantes e não-participantes, onde a variável testada (dependente) seria a diferença de médias entre os dois momentos no tempo entre os participantes e não-participantes. Se fizermos isso e compararmos os resultados, chegaremos a um valor idêntico obtido no teste t e o valor do coeficiente que indica a diferença da diferença (no exemplo anterior o β_3). É pertinente salientar que somente obteremos estes mesmos resultados nas diferenças em ambos os procedimentos quando tivermos exatamente os mesmos sujeitos (grupos) antes e depois do programa. No caso do exemplo do Projeto *Balsakhi* os mesmos estudantes e escolas em ambos os grupos (participantes e não-participantes) foram observados antes e depois.

5 ESCORE DE PROPENSÃO PARA PAREAMENTO (*PROPENSITY SCORE MATCHING*)

Esta técnica visa encontrar um grupo de unidades comparáveis entre não-participantes e participantes de alguma intervenção. Basicamente ela é utilizada quando os grupos (participantes e não-participantes de um programa ou política social) não foram selecionados aleatoriamente e por isso pode haver algum viés nos resultados, dado haver possibilidade de incomparabilidade entre os grupos.

Esta técnica consiste em identificar unidades não-tratadas que sejam similares às unidades tratadas e comparar as médias no resultado procurado entre estes dois grupos para identificar o impacto do tratamento (programa). A técnica do pareamento baseado no escore de propensão a participar considera que a seleção se dá por características observáveis.

Os procedimentos de pareamento são implementados utilizando-se um escore balanceado, computado a partir de um modelo de regressão logística onde a variável dependente seria 1 = participou e 0 = não participou. Especificamente usamos o logaritmo das chances de participar (log odds). Estimamos um modelo *logit* utilizando todas as variáveis observadas como preditoras para obtermos a probabilidade predita e computar a razão das chances (log odds – ratio) para cada observação na amostra do grupo de controle e do grupo de tratamento.

Basicamente o procedimento desta técnica testa se os escores de propensão a participar, tendo como variáveis independentes algumas características observadas, são estatisticamente os mesmos no grupo de controle e no de tratamento. Se não forem, o processo continua, com a retirada de unidades que estão mais distantes, até que tenhamos um equilíbrio nas observações. Uma vez que haja certo equilíbrio, isto é, que as unidades estejam comparáveis, podemos efetuar o teste de diferença de médias, da nossa variável que se refere ao resultado do programa que estamos querendo estimar, entre o grupo de controle e o de tratamento.

Um conceito importante dentro desta técnica é a ideia de um suporte comum, que se constitui na região onde o equilíbrio dos escores de propensão, entre os grupos de controle e de tratamento, se apresenta. Esta região de suporte comum é obtida a partir do descarte daqueles casos que estiverem muito abaixo ou muito acima da média dos escores. Entretanto, este procedimento de descarte dos casos que estiverem fora da região de suporte comum pode causar alguns problemas, como: bons pares podem ser perdidos próximo aos limites da região de suporte comum e a exclusão de casos pode mudar os parâmetros a serem estimados (redução do tamanho da amostra).

Pelo fato de esta técnica se basear nas características observáveis, ela apresenta certa limitação no que diz respeito às características não observadas que podem estar na base de processos de seletividade. Contudo, a literatura sobre esta técnica (HECKMAN; ISHIMURA; TODD, 1998) enfatiza que uma estratégia de avaliação que consegue controlar rigorosamente pelas características observadas e que consegue informações sobre o grupo de controle e o de tratamento de forma semelhante pode permitir estimadores confiáveis do impacto de um programa.

Como exemplo do uso do escore de propensão para pareamento, podemos citar a pesquisa de avaliação do Programa Nacional de Suporte ao Trabalho, desenvolvido em meados dos anos 1970 nos Estados Unidos para oferecer experiência de trabalho a trabalhadores que estivessem enfrentando problemas econômicos e sociais. Aqueles aleatoriamente designados para participar do programa tiveram treinamento em restaurantes e na construção civil. As informações relacionadas com o período anterior ao programa (ganhos, educação, estado civil e raça) foram medidas através de enquetes e de arquivos da administração pública. Em que pese os candidatos terem sido aleatoriamente selecionados, o fato de a avaliação ter sido feita num período de dois anos fez com que aqueles que entraram cedo no programa pudessem ter características diferentes daqueles que entraram mais tarde. Além disso não havia informação sobre se os participantes e não-participantes haviam estado no mercado de trabalho antes do programa, e assim a variável de interesse, a renda em 1978, poderia ser afetada caso tivessem sido empregados antes do programa. Então, havia a necessidade do procedimento de pareamento, a partir das características observadas antes do tratamento.

O que exatamente este procedimento faz é calcular a propensão (probabilidade) a participar baseado em características observáveis, após o que é feito o cálculo do valor estimado (y) na equação para todos os sujeitos (participantes e não-participantes), isto é, com base nos coeficientes da regressão logística somam-se as características de cada indivíduo (multiplicando-se pelo valor dos coeficientes da regressão) e cada indivíduo terá um valor de y diferente (um escore). Ressaltamos que para cada indivíduo no grupo participante do programa o pesquisador quer encontrar o indivíduo no grupo de controle (não-participante) que apresenta o

escore de propensão mais próximo (medido pela diferença absoluta entre os escores). Este procedimento se chama “vizinho mais próximo” (*nearest neighbor*).

Após este procedimento não há mais necessidade de uma análise de regressão com variáveis de controle, basta fazer um teste de diferença de médias do resultado que queremos observar.

Apresentamos a seguir a tabela 4, a partir do *output* fornecido pelo *software* estatístico Stata, do cálculo do escore de propensão e do resultado da diferença nos ganhos entre o grupo de tratamento e o de controle na avaliação do impacto do Programa Nacional de Suporte ao Trabalho nos ganhos dos participantes em 1978.

TABELA 4

Resultados da análise de regressão logística que estima o escore de propensão para pareamento

Log da probabilidade = -292.12772					
Número de obs. = 445					
LR qui2(10) = 19.94					
Prob > qui2 = 0.0298					
Pseudo R2 = 0.0330					
Participação	Coef.	Erro-padrão	z	P> z	[95% Intervalo conf]
ida	0.0082756	0.0145206	0.57	0.569	-0.0201842 0.0367355
educ	-0.0828202	0.0723030	-1.15	0.252	-0.2245314 0.0588910
negro	-0.2215734	0.3684377	-0.60	0.548	-0.9436980 0.5005513
hispanic	-0.8556853	0.5127832	-1.67	0.095	-1.8607220 0.1493513
casado	0.1960350	0.2783724	0.70	0.481	-0.3495648 0.7416348
s/curso sup	-0.8980952	0.3145636	-2.86	0.004	-1.5146290 -0.2815618
renda 1974	-0.0000447	0.0000301	-1.48	0.138	-0.0001037 0.0000143
taxa des 1974	-0.1927101	0.3764940	-0.51	0.609	-0.9306248 0.5452045
renda 1975	0.0000292	0.0000479	0.61	0.541	-0.0000646 0.0001231
taxa des.1975	-0.3368532	0.3213219	-1.05	0.294	-0.9666325 0.2929262
_cons	1.6222680	1.1019260	1.47	0.141	-0.5374665 3.7820020
Método de pareamento = kernel Metric = pscore					
Variável amostra			Tratado	Controle	Diferença
Renda após o Programa	Não-pareados		6349.14537	4554.80229	1794.3430
Pareados			6336.15652	4443.92743	1892.2291

Fonte: Tabela obtida pela análise de regressão realizada no *software* Stata versão 9.1. Dados oriundos do Programa Nacional de Suporte ao Trabalho dos Estados Unidos, fornecidos pelo professor Doutor Chandler Stolp da Universidade do Texas.

Observamos que as diferenças nas rendas entre o grupo de tratamento e o de controle, após o tratamento, são bem menores quando não utilizamos o procedimento do pareamento. Especificamente esta diferença é de US\$ 1.794,34 antes do pareamento e de US\$ 1.892,22 após o pareamento.

Observando a tabela 5 podemos ver as diferenças nas variáveis observadas entre o grupo de tratamento e o de controle sem o pareamento e com o pareamento. Podemos observar que em todas as variáveis observadas, após o pareamento, não existe diferença estatisticamente significativa entre o grupo de controle e o de tratamento.

TABELA 5

Características antes do programa, grupo de controle e de tratamento

Variáveis	Amostra	Médias				t-test	
		Tratado	Controle	% viés	Viés	t	p> t
T	Não-pareados	1	0
	Pareados	1	0
Idade	Não-pareados	25.816	25.054	10.7		1.12	0.265
	Pareados	25.809	25.75	0.8	92.3	0.08	0.934
Educ	Não-pareados	10.346	10.088	14.1		1.50	0.135
	Pareados	10.328	10.392	-3.5	75.1	-0.34	0.731
Negro	Não-pareados	0.84324	0.82692	4.4		0.45	0.649
	Pareados	0.85246	0.86475	-3.3	24.7	-0.36	0.720
hisp	Não-pareados	0.05946	0.10769	-17.5		-1.78	0.076
	Pareados	0.06011	0.05464	2.0	88.7	0.24	0.811
Casado	Não-pareados	0.18919	0.15385	9.4		0.98	0.327
	Pareados	0.19126	0.18852	0.7	92.3	0.07	0.944
s/curso sup	Não-pareados	0.70811	0.83462	-30.4		-3.22	0.001
	Pareados	0.71585	0.70628	2.3	92.4	0.21	0.830
Renda 1974	Não-pareados	2095.6	2107	-0.2		-0.02	0.982
	Pareados	2081.5	1899.1	3.4	-1493.1	0.39	0.697
u74	Não-pareados	0.70811	0.75	-9.4		-0.98	0.326
	Pareados	0.71038	0.73361	-5.2	44.6	-0.53	0.599
Renda 1975	Não-pareados	1532.1	1266.9	8.4		0.87	0.382
	Pareados	1488.3	1542.7	-1.7	79.5	-0.17	0.867
u75	Não-pareados	0.6	0.68462	-17.7		-1.85	0.065
	Pareados	0.60656	0.63525	-6.0	66.1	-0.60	0.548

Fonte: Tabela obtida pelo teste de propensão para pareamento realizado no *software* Stata versão 9.1. Dados oriundos do Programa Nacional de Suporte ao Trabalho dos Estados Unidos, fornecidos pelo professor Doutor Chandler Stolp da Universidade do Texas.

6 VARIÁVEIS INSTRUMENTAIS

É sempre desejável procurar situações, quando trabalhamos com análise de regressão múltipla, nas quais seja razoável supor que as variáveis omitidas (aquelas não incluídas no modelo) não sejam correlacionadas com as variáveis de interesse. Tais situações podem surgir se o pesquisador supõe aleatoriedade ou em alguma situação próxima desta. Na economia, estas situações não são facilmente percebidas, o que

torna alguns estudiosos céticos ou pessimistas em relação à pesquisa econômica empírica (ANGRIST; KRUEGER, 1999).

O problema apontado pode ser explicado da seguinte forma: a variável-resposta, Y_p , é afetada por características observáveis e não-observáveis (imensuráveis). O problema de endogeneidade (variáveis omitidas correlacionadas com as incluídas no modelo) surge porque o evento (ou ação) é também afetado por características imensuráveis (ou seja, por variáveis não incluídas no modelo de regressão). Isto ocorre porque não é possível explicar totalmente os determinantes ou os motivos que levam um indivíduo a tomar uma decisão ou ter alguma experiência, enquanto outros não (por exemplo, alguns fumam, estudam, se casam, enquanto outros indivíduos não). Se estas características não-observáveis são correlacionadas com as observáveis que afetam a variável-resposta, uma correlação espúria será estabelecida entre o evento (ou a ação) e a variável-resposta Y_i (MOFFITT, 2003), causando o viés de confundimento.³

Comumente a associação entre duas variáveis pode estar sendo afetada por características que confundem a existência de uma possível relação causal. A associação positiva entre educação e renda é um exemplo disto. É observado, por exemplo, que indivíduos com mais educação parecem ter outras características, como pais mais ricos ou maior habilidade. A estratégia mais comum de identificação na educação (e na economia em geral) é a tentativa de reduzir o viés que surge a partir de comparações simples (ou ingênuas). Para tanto, são utilizadas regressões para controlar variáveis que são confundidas com a educação. Por exemplo, a habilidade do indivíduo pode ser um dos principais determinantes do desempenho educacional. O controle desta variável, assim como de outras, pode diminuir a magnitude do coeficiente de educação na regressão, implicando que parte da associação entre educação e renda pode ser atribuída a outras variáveis e não só à educação (ANGRIST; KRUEGER, 1999).

Algumas estratégias de identificação do problema de confundimento são baseadas em variáveis instrumentais. Estas podem ser tratadas como uma variação exógena para a aproximação de um teste aleatório. O método que utiliza variáveis instrumentais pode eliminar o viés ao encontrar uma variável que seja correlacionada com a variável explicativa (variável independente) e não com a variável-resposta (ANGRIST; KRUEGER, 1999).

Da mesma forma, Gujarati (2000) confirma que a variável instrumental é uma *proxy* para a variável explicativa (variável independente), que, embora deva ser altamente correlacionada com esta variável, não pode ser relacionada com a

3. Aqui, entende-se confundimento a partir do fato de atribuímos o efeito de uma variável noutra, quando na realidade existe uma terceira que afeta ambas.

perturbação estocástica (termo do erro, o qual envolve variáveis explicativas não incluídas no modelo).

A maior dificuldade a respeito das variáveis instrumentais é encontrar uma que seja realmente válida, apesar das vantagens relacionadas à utilização destas variáveis. Como exemplo concreto do uso de variáveis instrumentais, podemos citar a pesquisa sobre o impacto da Reforma da Previdência Rural Brasileira,⁴ onde houve um aumento na renda dos aposentados, uma diminuição da idade para se tornar elegível e a possibilidade da presença de mais de um pensionista no domicílio, nos arranjos domiciliares (composição dos domicílios).

Especificamente, tal pesquisa toma como hipótese que o acréscimo de renda gerado por aquela reforma contribui para novas configurações na maneira como os domicílios se organizam, isto é, por exemplo, filhos principalmente podem voltar a morar com os pais idosos, dado que estes últimos, após a reforma, tiveram um acréscimo de renda. Para tanto, poderíamos pensar num modelo onde a diferença na renda domiciliar entre o momento anterior e posterior à reforma seria uma variável a afetar os arranjos domiciliares. Contudo, esta relação pode ser recíproca (causando viés). Isto é, não podemos saber se arranjos domiciliares diferenciados ou com maior número de residentes é que levaram ao incremento da renda ou se o incremento da renda é que levou aos novos arranjos. Este é um caso típico onde devemos utilizar uma variável instrumental, em vez da renda domiciliar diretamente.

No exemplo citado devemos pensar num conjunto de características causadas pela reforma (tais como idade para elegibilidade, sexo, local de moradia rural), as quais servem como *proxy* para a renda domiciliar. Assim, estimamos esta variável instrumental e a utilizamos no nosso modelo principal em vez de utilizarmos diretamente a variável renda.

7 CONSIDERAÇÕES FINAIS

O objetivo deste trabalho foi tentar demonstrar que uma avaliação rigorosa do impacto de programas e políticas sociais é mais difícil que se possa imaginar num primeiro momento.

Simplificadamente, a título de fechamento deste artigo, podemos resumir os principais aspectos a serem considerados no processo de avaliação de uma política ou programa social.

O primeiro seria conhecer em detalhe o programa que se quer avaliar, saber sobre seus objetivos, o formato (em termos de desenho de pesquisa, se foi um

4. Tal pesquisa vem sendo desenvolvida pela autora através de financiamento do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo Edital de Ciências Humanas de 2007.

experimento ou não), e possuir uma vasta gama de dados (variáveis quantitativas e qualitativas, numa planilha de dados que possam ser analisados estatisticamente) sobre o programa bem como sobre os atingidos por ele, com resultados que possam demarcar o alcance ou não dos objetivos originalmente propostos com referido programa. Também deve constar no banco de dados características (heterogeneidade do universo a ser analisado) do universo de atingidos e daqueles não-atingidos pelo programa. Outras variáveis necessárias no banco, para permitir uma avaliação minimamente coerente, dizem respeito aos aspectos (características dos pesquisados) que influenciam a participação no programa, mas que não afetam diretamente os resultados daquele. Estes aspectos são extremamente relevantes naquelas ocasiões em que não temos um desenho de programa com distribuição aleatória dos sujeitos (para o grupo que participa e aquele que não participa do programa), o que normalmente acontece. Convém salientar que falamos aqui em sujeitos, mas a unidade de análise numa pesquisa de avaliação pode ser formada por domicílios, regiões, municípios e outras, dependendo do tipo de programa.

Após a identificação dos aspectos apontados parte-se para a identificação de possíveis fontes de vieses. Estas podem se originar de duas fontes. Uma do fato de os grupos não serem comparáveis (diferirem muito em termos de certas características observáveis). Neste caso a seleção cuidadosa do grupo comparativo (que não participou do programa) pode eliminar aquele viés, e esta seleção pressupõe a escolha do grupo comparativo (contrafactual) com a mesma distribuição de características observáveis que o grupo tratado (atingido pelo programa).

A outra fonte reside nas diferenças entre grupo tratado e de controle em termos de características não-observáveis (viés de seleção). Esta fonte de viés surge quando, para dados valores de algumas características dos grupos, existe uma relação sistemática entre participação no programa e os resultados na ausência do mesmo. Em outras palavras, existem variáveis não-observadas que conjuntamente influenciam os resultados que estamos tentando checar e a probabilidade de participar do programa, condicional com outras características observadas.

Exemplo: quando queremos avaliar um programa como o Bolsa Família no Brasil, sabemos que para dele participarem as famílias devem ter uma renda domiciliar máxima X (no caso 1/2 SM por pessoa no domicílio). Então, quando queremos verificar se a participação no programa gerou algum impacto na frequência das crianças à escola, montamos uma equação na qual as variáveis que explicam a frequência à escola podem ser as mesmas que explicam a probabilidade de ser elegível para integrar o programa. No caso, a renda seria um exemplo: ela afeta a probabilidade de participar e ao mesmo tempo a frequência das crianças à escola, já que se supõe que famílias mais pobres tendem a enviar seus filhos com menos frequência à escola porque eles têm de trabalhar.

Vários métodos existem para resolver aqueles vieses. O problema essencial que estes métodos enfatizam é o fato de não podermos observar os resultados nos participantes se eles não tivessem participado do programa. Um grupo comparativo deve então ser utilizado para se identificar o que teria acontecido sem o programa.

A primeira alternativa para lidar com este problema seria o processo de aleatoriedade na distribuição dos grupos, o que garante a comparabilidade dos mesmos, mas que quase sempre não é o que acontece na formulação e aplicação de políticas e programas sociais. Normalmente eles têm grupos-alvo, isto é, grupos mais necessitados, num primeiro momento, e isto geralmente parece mais ético (priorizar aqueles realmente mais carentes).

Dentre os métodos para lidar com o viés da falta de comparabilidade dos grupos, apresentamos neste artigo três soluções, as quais não são mágicas e nem sempre viáveis, mas podem atenuar o problema da falta de aleatoriedade na distribuição dos grupos.

Apresentamos a técnica da diferença da diferença. Neste procedimento o avaliador compara o resultado a ser avaliado no grupo de tratamento (que recebeu o programa) e no de controle antes (primeira diferença) e depois do programa ter ocorrido (segunda diferença) controlando-se por características (variáveis de controle) que possam se diferenciar entre os grupos. Para isso ambos os grupos devem ter sido submetidos à aplicação de questionários idênticos (exatamente as mesmas variáveis, antes e depois).

Num segundo momento apresentamos a técnica do escore propensão para pareamento (*propensity score matching*). Neste método o avaliador tenta pegar um grupo de comparação (controle) ideal dentro de uma grande *survey* (enquete). Este grupo comparativo é pareado com o de tratamento com base na probabilidade predita para a participação baseada num conjunto de variáveis explicativas (obtida por uma regressão logística). Cabe destacar que o grupo de comparação deve ser selecionado dentro do mesmo contexto em que o grupo tratamento (para garantir a comparabilidade), e também é necessário que tenhamos as mesmas informações (variáveis) coletadas para ambos os grupos com uso de um mesmo instrumento de pesquisa (questionário).

Por fim destacamos o procedimento que utiliza variáveis instrumentais. Tais variáveis, como já dissemos, consistem naquelas que influenciam a participação no programa, mas não os resultados do mesmo (aqueles que queremos avaliar). Se tais variáveis existirem, então elas identificam a fonte de variação exógena nos resultados atribuídos ao programa.

Como conclusão, salientamos que neste artigo apresentamos alguns dos principais conceitos, princípios e técnicas necessários no processo de avaliação de um

programa ou política social. Destacamos a necessidade de que haja um conjunto de informações imprescindíveis, um banco de dados completo sobre os participantes e não-participantes, antes e após o programa ter sido implementado, e também precisamos conhecer em detalhe os objetivos do programa. Sem estas informações preliminares nenhuma das técnicas aqui apresentadas é passível de ser utilizada com rigor. Isto quer dizer que todo o programa ou política social, ao ser idealizado, deve vir acompanhado de uma série de informações necessárias para uma posterior avaliação após a implementação. Se estes requisitos não forem obedecidos, os procedimentos técnicos apresentados neste artigo talvez não possam ser aplicados.

Contudo, temos consciência de que nem sempre a realidade é tão perfeita e muito menos que em contextos de programas aplicados em países menos desenvolvidos seja possível obter informações desejadas. Mas cabe ao pesquisador avaliador saber lidar com as informações disponíveis e ser honesto ao apresentar seus resultados, destacando os possíveis limites da abrangência da avaliação que foi possível fazer. Temos ciência, também, de que os conhecimentos sobre avaliação, bem como a abertura e disponibilidade dos idealizadores e aplicadores de políticas e programas sociais nem sempre chegam próximo do ideal. Questões de outras ordens (ideológicas, políticas ou mesmo financeiras), as quais transcendem às questões metodológicas aqui apresentadas, podem trazer grandes obstáculos à realização de uma avaliação realmente séria e baseada em critérios isentos e científicos.

REFERÊNCIAS

- ANGRIST, J.; KRUEGER, A. Empirical strategies in labor economics. In: ASHENFELTER, O.; CARD, D. (Eds.). *The handbook of labor economics*. v. 3A, Chapter 23. Amsterdam: North-Holland, 1999 (Handbooks in Economics, n. 5).
- GUJARATI, D. N. *Econometria básica*. 3. ed. São Paulo: Makron Books, 2000. 846p.
- HECKMAN, J.; ISHIMURA, H.; TODD, P. Matching as an econometric evaluation estimator. *Review of Economic Studies*, v. 65, p. 261-294, 1998.
- MARTELETO, L. The role of demographic and family change on children's schooling: evidence from Brazil. 2001. 204 f. Tese (Doutorado em Demografia) - The University of Michigan, 2001.
- MOFFITT, R. *Remarks on the analysis of casual relationships in population research*. Johns Hopkins University, 2003. Disponível em <<http://www.econ.jhu.edu/People/Moffitt/causal.pdf>>. Acesso em: 16 dez. 2006.
- MOKATE, K. M. Convirtiendo el "monstruo" en aliado: la evaluación como herramienta de la gerencia social. *Revista do Serviço Público*, v. 8, n. 1, p. 91-136, 2002.
- ROSENZWEIG, M. R.; WOLPIN, K. I. Natural "natural experiments" in economics. *Journal of Economics Literature*, v. 38, n. 4, p. 827-874, Dec. 2000.
- RUA, M. G. *Avaliação de políticas, programas e projetos*: notas introdutórias. 2000. Mimeografado.
- POVERTY ACTION LAB. Remedying education: evidence from two randomized experiments in India. Disponível em: <www.povertyactionlab.org/papers>.