

## Avaliação de Políticas Públicas e Programas Sociais: aspectos conceituais e metodológicos\*

Marília Ramos<sup>†</sup>

**Resumo:** 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; quais os procedimentos (econométricos) necessários para que uma política possa ser objetivamente avaliada e quais são alguns dos recursos metodológicos e estatísticos que são necessários para termos 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, quais sejam: 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 do programa. Outro aspecto seria a identificação de possíveis fontes de vieses: o fato dos 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 na diferença, a técnica do escore propensão para pareamento (*propensity score matching*) e por fim apresentamos o procedimento que utiliza variáveis instrumentais.

---

\* Trabalho apresentado no XVI Encontro Nacional de Estudos Populacionais, ABEP, realizado em Caxambú- MG – Brasil, de 29 de setembro a 03 de outubro de 2008

<sup>†</sup> PhD em Sociologia pela Purdue University, EUA. Pós-doutora em Métodos Quantitativos e Avaliação de Políticas Públicas pela Universidade do Texas, EUA. Professora e Pesquisadora do Programa de Mestrado e Doutorado em Desenvolvimento Regional da Universidade de Santa Cruz do Sul.

## **Introdução**

O presente artigo visa discutir aspectos conceituais e principalmente metodológicos relacionados com a avaliação de políticas e programas sociais. Num primeiro momento são discutidas questões relacionadas com os objetivos e a necessidade de avaliação, logo em seguida é enfatizado um conjunto de procedimentos metodológicos relacionados com 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 aqui 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 foco deste artigo a discussão sobre uma parte da avaliação que prioriza a opinião e avaliação dos atingidos por aquelas formas de intervenção na realidade social. Não negamos a importância de ouvirmos aqueles diretamente atingidos, apenas tomamos como foco aspectos concernentes à avaliação técnica de resultados.

É importante deixar claro, que estamos tratando da avaliação formal, a qual consiste no exame sistemático de certos objetos, baseado em procedimentos científicos de coleta e análise de informação sobre o 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, sendo 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 é decisiva também para o processo de aprendizagem institucional e 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.

No Brasil, a importância da avaliação das políticas públicas é reconhecida em documentos oficiais e científicos, mas esse reconhecimento formal ainda não se traduz, em algumas regiões do Brasil, em processos de avaliação sistemáticos e consistentes que subsidiem a gestão pública (Hartz et Pouvourville, 1998). Especificamente observa-se a proliferação de pesquisas, denominadas de avaliação, as quais não se utilizam de técnicas objetivas de avaliação e ficam restritas à opinião dos sujeitos participantes, ou seja, a visão subjetiva dos atingidos diretamente pela referida política. Assim sendo, visando complementar os estudos que já existem sobre avaliação, apresentamos 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 que são necessários para possamos ter os resultados mais fidedignos e válidos possíveis.

### **1-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 da política/programa ter sido implantado(a). Sem esta informação se torna praticamente inviável pensarmos em qualquer tipo de avaliação.

Assim sendo tomamos a possibilidade de acessarmos informações, referentes ao resultado que estamos buscando com as referidas políticas, antes das 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.

Um 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 podido se dar conta, o desenho de uma pesquisa de avaliação se baseia nas diretrizes de uma pesquisa de caráter experimental, onde informações antes e depois do tratamento são necessárias, bem como a necessidade de pelo menos 2 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).

Idealmente falando, 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 encontramos o grupo comparativo ou, como se chama na literatura sobre avaliação, o contra-factual.

Cabe retomar aqui de forma breve a explicação metodológica para a necessidade de um grupo comparativo, para deixar claro porque não podemos somente trabalhar 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.

Neste caso retomamos a discussão, presente na literatura na área metodológica sobre a necessidade do grupo controle. Se analisamos 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 o 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 deles 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 existem mais chances dos sujeitos envolvidos ficarem cansados ou chateados, mais ou menos motivados. Ou 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 este 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 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 eles apresentam escores em alguma variável, muito altos ou muito baixos. Ao re-testar estes sujeitos as médias daqueles com altos escores somente poderão baixar e 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 fora 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 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 tratamento e outros para não receber, grupo controle. O resultado para o grupo controle é o contrafactual; este permite identificar o resultado no estado não tratado. O elemento chave é a aleatoriedade das unidades 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 grupo tratamento e grupo controle.

Entretanto, os experimentos sociais não são sempre possíveis de serem realizados por uma série de razões tais como os custos políticos e monetários, a inabilidade de realizar experimentos em desenho universais de políticas (onde 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.

Consequentemente, o teste da fidedignidade de um não experimento é uma questão central na literatura sobre avaliação. Métodos não experimentais tipicamente

permitem identificar os impactos impondo pressupostos não testáveis; experimentos aleatórios, quando possíveis, podem ser usados para se acessar a validade daqueles pressupostos e a performance de técnicas alternativas de avaliação.

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 contra-factual.

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. Os experimentos sociais, onde o um grupo de unidades elegíveis (indivíduos, domicílios, localidades etc) são aleatoriamente excluídos 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 não são uma panacéia; eles podem prover um estimador do impacto bastante consistente, quando não há viés por desistência de participantes ou por substituição dos mesmos.

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 conseguir um emprego e aí não conseguiremos isolar o efeito específico do curso na probabilidade de conseguirem um 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 a tona questões de ordem ética. Entretanto, quando 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 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, os quais não são testáveis, apesar de muitas de suas implicações possam ser, e podem ou não ser viáveis com dados disponíveis. Métodos não experimentais podem produzir vieses de auto-seleção por causa das diferenças nas fontes e qualidade dos dados.

Infelizmente, os pesquisadores raramente têm a oportunidade de “tornar aleatórias” variáveis tais como o desempenho educacional, a imigração, ou salário mínimo por exemplo. Muitos pesquisadores precisam confiar em estudos que geralmente falham em produzir a mesma força de evidência de um experimento aleatório (ANGRIST & KRUEGER, 1998). 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.

## **2-Seleção aleatória dos grupos de tratamento (recedor da política ou programa social) e 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 controle (contrafactual) o mais parecido possível com o grupo tratamento deve ser a aleatoriedade.

A análise causal, e portanto, no caso do impacto de uma política ou programa social, está inserida em um problema que, segundo MOFFITT (2003a), é baseado na noção fundamental do contra-factual 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,  $Y_{1i}$  e  $Y_{0i}$ , em que  $Y_{1i}$  é o resultado se o indivíduo experimenta um evento particular ou toma uma determinada decisão, e  $Y_{0i}$  é o resultado caso ele não experimente o evento, mantendo o resto constante. A diferença entre os dois ( $Y_{1i} - Y_{0i}$ ) 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 contra-factual (MOFFITT, 2003a).

Autores como ANGRIST & 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 contra-factuais (*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 contra-factual.

Na prática, o pesquisador pode aprender mais sobre os resultados contra-factuais se ele considerar o experimento aleatório. Em muitas pesquisas, principalmente na área médica, acredita-se que a melhor evidência sobre o resultado contra-factual é observada a partir de um experimento aleatório. A falta deste é uma das principais razões porque a pesquisa econométrica sempre convence menos do que a pesquisa em outras ciências mais experimentais (ANGRIST & KRUEGER, 1998). Outros autores, como ROSENZWEIG & 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 sendo, conforme já mencionado acima, 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.

## **3-O procedimento da diferença na diferença**

A estratégia chamada de “diferença na 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 contra-factuais (ANGRIST & KRUEGER, 1998).

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 subtrai 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*. Este programa visava melhorar a performance de alguns estudantes com dificuldades em matemática e nas habilidades para leitura. Para tanto foram selecionados aleatoriamente, entre escolas e estudantes com dificuldades, aquelas que iriam participar do programa e aquelas que não iriam. O Programa *Balsakhi* se constituía na contratação de tutores (os *balsakhis*) que iriam fornecer, àquelas escolas e crianças selecionadas, aulas extras de matemática e leitura.

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

$$Y = \beta_0 + \beta_1 + \beta_2 + \beta_3$$

Onde:

$Y$  = Diferença nos Escores nos testes (notas após menos notas antes)

$\beta_1$  = ano posterior ao programa (=1)

$\beta_2$  = participou do programa (=1)

$\beta_3$  = interação ano \* participou.

Para identificarmos a “diferença na 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 abaixo podem ser calculados pela soma dos  $\beta$ s, notando que vários deles são iguais a zero.

	Participante	Não participante	Diferença
Antes	$\beta_0 + \beta_2$	$\beta_0$	$\beta_2$
Após	$\beta_0 + \beta_1 + \beta_2 + \beta_3$	$\beta_0 + \beta_1$	$\beta_2 + \beta_3$
Diferença	$\beta_1 + \beta_3$	$\beta_1$	$\beta_3$

Podemos observar no esquema acima que o  $\beta_3$  indica a diferença na diferença. Concretamente quando colocamos as informações do projeto exemplificado acima numa planilha estatística e efetuamos a análise de regressão obtemos os resultados a seguir:

Source	SS	df	MS	Number of obs =	24208
Model	2796447,76	3	932149.255	F( 3, 24204) =	1986.70
				Prob > F	= 0.0000

Residual		11356418,4	24204	469,195934		R-squared	=	0.1976
-----						Adj R-squared	=	0.1975
Total		14152866.2	24207	584.660063		Root MSE	=	21.661
-----								
Teste(y)		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]		
-----								
após		18,19565	,397692	45,75	0,000	17,41615	18,97515	
participante		,7492239	,3892857	1,92	0,054	-,0138002	1,512248	
partic*após		5,574539	,5571308	10,01	0,000	4,482528	6,666549	
_cons		25,40569	,277887	91,42	0,000	24,86102	25,95037	
-----								

Se substituirmos e somarmos os valores encontrados nos coeficientes de regressão calculados e fornecidos pelo software obtemos o esquema abaixo:

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

Note que os valores encontrados dizem respeito aos resultados do cálculo da estimativa de Y, por exemplo: o valor 26,16 foi obtido somando os coeficientes  $\beta_0$  e  $\beta_2$ , 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  $\beta_3$  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, fossem estatisticamente não significativas como ilustra o coeficiente  $\beta_2$ . Entretanto, a diferença apareceu como marginalmente significativa ( $P=0,054$ ), o que indica que havia diferenças entre os grupos antes do programa ter sido implantado. Cabe destacar que sempre que encontrarmos diferenças significativas entre os grupos antes do programa ser implantado, devemos pensar em acrescentar controles em nossos modelos ou nos utilizarmos de técnicas de pareamento tais como escore de propensão para pareamento, o qual será discutido na sessão 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 das 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 na diferença, a qual 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 na 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 obteremos um valor idêntico obtido no teste t e o valor do coeficiente que indica a diferença na diferença na diferença (no exemplo acima o  $\beta_3$ ). 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.

#### 4-Escore de propensão para pareamento (*Propensity score matching*)

Esta técnica visa encontrar um grupo 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 o fato da possibilidade de incomparabilidade dos grupos.

Esta técnica consiste em identificar unidades não tratadas que sejam similares as 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*). Nós 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 controle e do grupo 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 controle e tratamento. Se não forem, o processo continua, com a retirada de unidades que estão mais distantes, até que tenhamos um balanço nas observações. Uma vez que haja certo balanço, isto é, as unidades estejam comparáveis, podemos efetuar o teste de diferença de médias, da nossa variável que refere ao resultado do programa que estamos querendo estimar, entre o grupo controle e o tratamento.

Um conceito importante dentro desta técnica é a idéia de um suporte comum, o qual se constitui na região onde o balanço dos escores de propensão, entre os grupos controle e 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 tais 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).

Em função do fato de que esta técnica se baseia nas características observáveis, ele apresenta certa limitação no que diz respeito às características não observadas que podem estar na base de processos de seletividade na. Contudo, a literatura sobre esta técnica (Heckman, Ichimura and Todd, 1997, 1998 e Heckman, Ichimura, Smith and Todd, 1998) enfatiza que uma estratégia de avaliação que consegue controlar rigorosamente pelas características observadas e que consegue informações sobre o

grupo controle e o 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 70 nos EUA para oferecer experiência de trabalho a trabalhadores que estivessem enfrentando problemas econômicos e sociais. Aqueles aleatoriamente designados participar do programa participaram de 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 o fato dos candidatos terem sido aleatoriamente selecionados, o fato da avaliação ter sido feita num período de 2 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, assim sendo 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 é feito o cálculo do valor estimado ( $y$ ) na equação para todos os sujeitos (participantes e não participantes), isto é, baseado nos coeficientes da regressão logística soma-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 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 uma análise de regressão com variáveis de controle, basta fazer um teste de diferença de médias do resultado que estamos querendo observar.

Apresentamos a seguir a ilustração, 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 tratamento e controle na avaliação do impacto do Programa Nacional de Suporte ao Trabalho nos ganhos dos participantes em 1978.

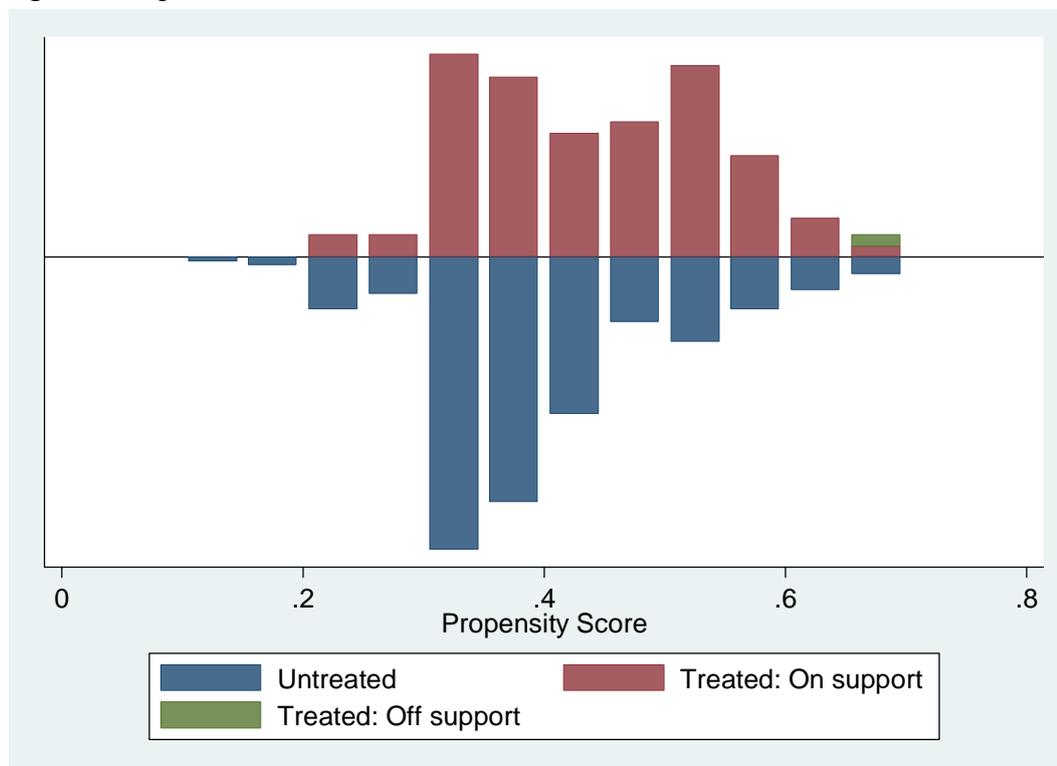
Regressão Logística	Number of obs	=	445
	LR chi2(10)	=	19.94
	Prob > chi2	=	0.0298
Log likelihood = -292.12772	Pseudo R2	=	0.0330

Participação	Progr. Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
ida	.0082756	.0145206	0.57	0.569	-.0201842 .0367355
educ	-.0828202	.072303	-1.15	0.252	-.2245314 .058891
negro	-.2215734	.3684377	-0.60	0.548	-.943698 .5005513
hisp	-.8556853	.5127832	-1.67	0.095	-1.860722 .1493513
casado	.196035	.2783724	0.70	0.481	-.3495648 .7416348
s/curso sup	-.8980952	.3145636	-2.86	<b>0.004</b>	-1.514629 -.2815618
renda 1974	-.0000447	.0000301	-1.48	0.138	-.0001037 .0000143
taxa des 1974	-.1927101	.376494	-0.51	0.609	-.9306248 .5452045
renda 1975	.0000292	.0000479	0.61	0.541	-.0000646 .0001231
taxa des.1975	-.3368532	.3213219	-1.05	0.294	-.9666325 .2929262
_cons	1.622268	1.101926	1.47	0.141	-.5374665 3.782002

Matching Method = kernel Metric = pscore

Variable	Sample	Treated	Controls	Difference
Renda após o Programa	<b>Não pareados</b>	6349.14537	4554.80229	<b>1794.34308</b>
	<b>Pareados</b>	6336.15652	4443.92743	<b>1892.2291</b>

Observamos que as diferenças nas rendas entre o grupo tratamento e o controle, após o tratamento, é bem menor quando não utilizamos o procedimento do pareamento. Especificamente esta diferença é de 1.794,34 dólares antes do pareamento e de 1.892,22 dólares após o pareamento. O gráfico abaixo ilustra a região de suporte comum entre os grupos controle e tratamento. As observações nos cantos extremos são aquelas que estão fora da região de suporte comum.



Observando a tabela a seguir podemos ver as diferenças nas variáveis observadas entre grupo tratamento e 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 controle e o tratamento.

Tabela 1 Características antes do programa, grupo de controle e de tratamento

Variable	Sample	Mean		%bias	%reduct  bias	t-test	
		Treated	Control			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	.84324	.82692	4.4		0.45	0.649
	Pareados	.85246	.86475	-3.3	24.7	-0.36	0.720
hisp	Não Pareados	.05946	.10769	-17.5		-1.78	0.076
	Pareados	.06011	.05464	2.0	88.7	0.24	0.811
Casado	Não pareados	.18919	.15385	9.4		0.98	0.327
	Pareados	.19126	.18852	0.7	92.3	0.07	0.944
s/cursosup	Não pareados	.70811	.83462	-30.4		-3.22	<b>0.001</b>
	Pareados	.71585	.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	.70811	.75	-9.4		-0.98	0.326
	Pareados	.71038	.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	.6	.68462	-17.7		-1.85	0.065
	Pareados	.60656	.63525	-6.0	66.1	-0.60	0.548

### 5-Variáveis Instrumentais

É sempre desejável procurar situações, quando trabalhamos com análise de regressão múltipla, nas quais seja razoável supor que as variáveis omitidas não sejam correlacionadas com as variáveis de interesse. Tais situações podem surgir se o pesquisador supõe aleatoriedade ou 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, 1998).

Apesar disso, evidências sobre as relações causais são possíveis de ser captadas. Alguns estudos tentam controlar as diferenças observadas a partir de técnicas de regressão, usando comparações pré-pós da mesma unidade observada para reduzir o viés das diferenças não-observadas, e usando variáveis instrumentais como uma fonte de variação quase-experimental. Estes são exemplos de tentativas em esclarecer uma relação causal.

O caminho aleatório forma uma referência conceitual na avaliação do sucesso ou fracasso de estudos que utilizam estas idéias, mesmo sabendo que é impossível, ou pelo menos impraticável, estudar algumas questões usando a idéia de aleatório (ANGRIST & KRUEGER, 1998).

Um importante pressuposto sobre o efeito causal é que ele deve ser diferente entre os indivíduos, o que implica que somente o efeito causal médio pode ser estimado.

De acordo com GUJARATI (2000), variáveis endógenas são aquelas cujos valores são determinados dentro do modelo, enquanto que as exógenas ou pré-determinadas são aquelas cujos valores são determinados fora do modelo. Este autor também salienta que é responsabilidade do pesquisador especificar quais variáveis são endógenas e quais são exógenas.

A variável-resposta,  $Y_i$ , é afetada por características observáveis e por não-observáveis (imensuráveis). O problema de endogeneidade surge porque o evento (ou ação) é também afetado por características imensuráveis. 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 não-observáveis que afetam a variável-resposta, uma correlação espúria será estabelecida entre o evento (ou a ação) e a variável-resposta<sup>‡</sup>  $Y_i$  (MOFFITT, 2003a).

Um ponto básico na literatura sobre o efeito causal, é que este não pode ser estimado sem algum tipo de hipótese ou restrição, a princípio, devido a inerente não-observação do resultado contrafactual. A maioria dos problemas causais importantes na pesquisa econômica e na populacional surge quando a variável que afeta  $Y_i$  é endógena, o que equivale a dizer que os valores da variável podem se diferenciar de maneiras não-observadas. Este problema produz uma estimativa incorreta do efeito causal em  $Y_i$ , correspondendo assim, a um coeficiente de regressão não-ajustado. Isto significa que hipóteses ou estratégias de identificação devem ser feitas (MOFFITT, 2003a).

### 5.1 Estratégias de identificação para relações causais

Segundo GUJARATI (2000), “o problema de identificação surge porque diferentes conjuntos de coeficientes estruturais podem ser compatíveis com o mesmo conjunto de dados” (p. 663).

A estratégia de identificação é a combinação entre a variação identificada (rotulada) em uma variável causal, e o uso de uma técnica econométrica particular que explora esta informação. Algumas estratégias de identificação podem ser mencionadas como, por exemplo, o controle por variáveis de confundimento, por variáveis instrumentais, pelo experimento natural, entre outras.

Uma estratégia simples é o controle por variáveis de confundimento. 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 como o *background*, pode diminuir a magnitude do coeficiente de

---

<sup>‡</sup> Por exemplo, considere a relação entre a mãe fumante e o peso do filho ao nascer. Se além de fumante, uma mãe também tem desvantagens em seu *background*, que não podem ser medidas, e que levam a um baixo peso ao nascer de seu filho, poderá ser observado uma correlação negativa entre mães fumantes e o peso do filho ao nascer, não (somente) devido ao ato de fumar por si só. Mães fumantes poderiam ter filhos com baixo peso, mesmo se não fumassem.

educação na regressão, implicando que parte da associação entre educação e renda pode ser atribuída a outras variáveis, e não só a educação (ANGRIST & KRUEGER, 1998).

Um problema pode surgir quando se controla através da regressão. É necessário saber se as estimativas são realmente sensíveis à inclusão de variáveis-controle adicionais. Por exemplo, o impacto negativo da habilidade do indivíduo no coeficiente de educação pode ser eliminado ou mesmo revertido, se são considerados dois fatores: erro de mensuração no regressor de interesse e o uso de uma variável endógena para servir de *proxy* para a habilidade. Tais problemas levam a busca de outros métodos de estimação.

Algumas estratégias de identificação são baseadas em variáveis instrumentais. Estas podem ser tratadas como uma variação exógena para a aproximação de um teste aleatório. O método utilizando variáveis instrumentais pode eliminar o viés de endogeneidade, ao encontrar uma variável que seja correlacionada com o regressor e não com a variável-resposta (ANGRIST & KRUEGER, 1998).

Da mesma forma, GUJARATI (2000) confirma que variável instrumental é uma “*proxy*” para variável explicativa estocástica, que embora deva ser altamente correlacionada com esta variável, não pode ser relacionada com a perturbação estocástica (termo do erro). Ao encontrar uma *proxy* para o regressor endógeno, pode-se aplicar diretamente o método dos mínimos quadrados ordinários (MQO).

Uma maneira de se empregar uma variável instrumental é a partir do método de mínimos quadrados em dois estágios (MQ2S), que envolve duas sucessivas aplicações do MQO.

Na economia, as variáveis instrumentais são relativamente comuns. O modelo de regressão linear pode ilustrar esta estratégia:

$$Y_i = \alpha + \beta_i T_i + \lambda X_i + e_i \quad (1)$$

$$T_i = \delta + \theta X_i + \phi Z_i + u_i \quad (2)$$

onde  $Y_i$  é a variável-resposta do indivíduo  $i$ ;  $T_i$  é a variável *dummy* que é igual a 1 se  $i$  experimenta o evento e igual a 0 se não experimenta;  $X_i$  é uma variável-controle exógena;  $Z_i$  é uma variável-controle exógena que afeta a probabilidade de experimentar o evento, mas não afeta diretamente  $Y_i$ ; e  $e_i$  e  $u_i$  são os termos do erro.  $\beta_i$  é o efeito do evento na variável-resposta. A MQO (1) gera uma estimativa enviesada e inconsistente deste efeito se  $T_i$  é endógeno ou auto-selecionado. A MQO (2) assume que  $T_i$  é afetado por  $Z_i$ , que não está presente na equação (1) e assim, não é esperado que afete diretamente  $Y_i$ .

Neste exemplo,  $Z_i$  é a variável instrumental. Para esta ser válida, duas exigências são impostas:  $Z_i$  deve afetar  $T_i$  ao mesmo tempo em que não deve ser correlacionada com  $Y_i$ .

Existem muitos mecanismos pelos quais  $Z_i$  pode afetar  $Y_i$ . Quando uma hipótese de identificação é feita, a escolha de  $Z_i$  depende necessariamente de hipóteses que não são formalmente testáveis, o que implica na necessidade de uma explicação teórica que guie a escolha de  $Z_i$ . Tal explicação deve considerar o argumento de exogeneidade (MOFFITT, 2003a).

A maior dificuldade a respeito das variáveis instrumentais é encontrar uma que seja realmente válida. Apesar das vantagens relacionadas à utilização destas variáveis, GUJARATI (2000) enfatiza que “(...) achar boas *proxies* nem sempre é fácil, fazendo com que o método de variável instrumental tenha pouco uso prático (...)” (p. 611). O mesmo argumento (ou explicação) foi utilizado por MARTELETO (2001), ao se abster de tratar do viés de simultaneidade entre fecundidade e investimento em capital humano

dos filhos. De acordo com esta autora, não havia variáveis instrumentais convincentes disponíveis para o seu estudo.

Algumas vezes é suficiente saber somente qual é o efeito total de um instrumento na variável-resposta, não sendo necessário saber como o efeito ocorre através do evento de interesse. Este tipo de análise é conhecido como forma reduzida. Assim, o que caracteriza este método é a falta de estimação do canal através do qual o instrumento afeta o resultado. De acordo com GUJARATI (2000), estas equações expressam a variável endógena somente como uma função das exógenas e do termo do erro.

Como exemplo concreto do uso de variáveis instrumentais podemos citar a pesquisa sobre o impacto da Reforma da Previdência Rural Brasileira, 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 da 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 seus pais idosos dados que estes últimos, após a reforma, tiveram um acréscimo em suas rendas. 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. 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, ao invés da renda domiciliar diretamente.

No exemplo acima 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 sendo estimamos esta variável instrumental e a utilizamos no nosso modelo principal ao invés de utilizarmos diretamente a variável renda.

### **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 experimento ou não) e possuir uma vasta gama de dados (variáveis quantitativas e qualitativas, numa planilha de dados, passíveis de serem analisadas 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 desnecessárias 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 do programa. Estes aspectos são extremamente relevantes naquelas ocasiões 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 é 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 acima parte-se para a identificação de possíveis fontes de viéses. Estas podem se originar de duas fontes. Uma do fato dos 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 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 poder integrar este programa as famílias devem ter uma renda domiciliar máxima X. Então, quando queremos verificar se a participação no programa gerou algum impacto na frequência das crianças à escola, montamos uma equação na qual variáveis que explicam a frequência na escola podem ser as mesmas que explicam a probabilidade de ser elegível para participar no 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 asqueles viéses. 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 soa mais ético (priorizar aqueles realmente mais carentes).

Dentre os métodos para lidar com o viés da falta de comparabilidade dos grupos apresentamos aqui 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 na diferença. Neste procedimento o avaliador compara o resultado a ser avaliado no grupo tratamento (que recebeu o programa) e 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 a 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 (enquête). 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 obtido dentro do mesmo contexto que o tratamento foi, e também é necessário que tenhamos as mesmas informações (variáveis) coletadas para ambo 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 estamos querendo avaliar). Se tais variáveis existirem então elas identificam a fonte de variação exógena nos resultados atribuídos ao programa.

Como conclusão final salientamos que neste artigo apresentamos alguns dos principais conceitos, princípios e técnicas necessários no processo de avaliação de um programa ou política social. Destacamos que precisamos ter um conjunto de informações imprescindíveis, tais como 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 sua posterior avaliação após sua implementação, se assim não for os procedimentos técnicos neste artigo apresentados, podem se tornar impossíveis de serem aplicados.

Contudo, temos consciência 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 de ser feita. Temos ciência, também, 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 vir a trazer grandes obstáculos à realização de uma avaliação realmente séria e baseada em critérios isentos e científicos.

## Referências

HARTZ Z.M.A. & POUVOURVILLE, G. (1998). Avaliação da eficiência em saúde: a eficiência em questão. *Ciência & Saúde Coletiva*, III(1): 68-82.

MOKATE, K. M. (2002). “Convirtiendo el ‘monstruo’ en aliado: a avaliação como herramienta de la gerencia social”. *Revista do Serviço Público*, 8(1), 91-136.

Glewwe, Paul, Michael Kremer, Sylvie Moulin and Eric Zitzewitz. “Retrospective vs prospective analyses of school inputs: the case of flip charts in Kenya” *Journal of Development Economics*, 74(1). 2004.

Glewwe, Paul, Nauman Ilias and Michael Kremer (2003), “Teacher Incentives”, Mimeo, Harvard. Kremer, Michael and Daniel Chen (2001), “An Interim Report on a Teacher Attendance Incentive Program in Kenya”, Mimeo, Harvard University.

Kremer, Michael, Edward Miguel and Rebecca Thornton (2004), “Incentives to Learn”, NBER Working Paper #10971.

Kremer, Michael, and Christel Vermeesh (2005) “School committee empowerment: preliminary notes”, Mimeo, Harvard.

“Remedying Education: Evidence from Two Randomized Experiments in India” is available for download at [www.povertyactionlab.org/papers](http://www.povertyactionlab.org/papers)

ANGRIST, J., EVANS, W. N. Children and their parents’ labor supply: evidence from exogenous variation in family size. **The American Economic Review**, v.88, n.3, p.450-477, June 1998.

ANGRIST, J., KRUEGER, A. Empirical strategies in labor economics. In.: ASHENFELTER, O., CARD, D. (Eds.) **The handbook of labor economics**. Amsterdam: North-Holland, 1999.v.3A, Chapter 23. (Handbooks in economics, 5)

GUJARATI, D.N. **Econometria básica**. 3.ed. São Paulo: Makron Books, 2000. 846p.

MOFFITT, R. **Casual analysis in population research: an economist’s perspective**. Johns Hopkins University, 2003b. Disponível em: <http://www.econ.jhu.edu/People/Moffitt/causal.pdf>> Acesso em: 19 set. 2006.

MOFFITT, R. **Remarks on the analysis of casual relationships in population research**. Johns Hopkins University, 2003a. Disponível em: <http://www.econ.jhu.edu/People/Moffitt/causal.pdf>> Acesso em: 16 dez. 2006.

MUNIZ, J. O. **Uma aproximação empírica para o trabalho infantil no Brasil**. 2000. 60 f. Monografia - Faculdade de Ciências Econômicas, Universidade Federal de Minas Gerais, 2000.

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.

RUA, M.G. (2000). “Avaliação de Políticas, Programas e Projetos: Notas Introdutórias”. *Mimeo*.

MOKATE, K. M. (2002). “Convirtiendo el ‘monstruo’ en aliado: a avaliação como herramienta de la gerencia social”. *Revista do Serviço Público*, 8(1), 91-136.

ROSENZWEIG, M. R, WOLPIN, K, I. Natural “natural experiments” in economics. **Journal of Economics Literature**. v.38, n.4, p.827-874, Dez. 2000.