

AVALIAÇÃO DE POLÍTICAS

2015



FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE DE RIBEIRÃO PRETO
UNIVERSIDADE DE SÃO PAULO

“Apesar de bilhões de dólares serem gastos todos os anos com projetos sociais, muito pouco se sabe do seu impacto sobre beneficiários e sua viabilidade econômica”

BANCO MUNDIAL

“ O que conta, é o que funciona”

TONY BLAIR (1998)





1. INTRODUÇÃO

Avaliação é importante tanto para grandes Programas e Políticas Públicas Governamentais, como para Projetos Sociais desenvolvido por ONGs.



QUESTÕES IMPORTANTES A SEREM RESPONDIDAS PELA AVALIAÇÃO ECONÔMICA:

- O projeto atingiu seu público-alvo?
- O projeto teve o impacto almejado?
- Qual o retorno econômico do projeto?
- Os recursos estão sendo gastos do modo mais eficiente?



Especialmente em PAÍSES EM DESENVOLVIMENTO

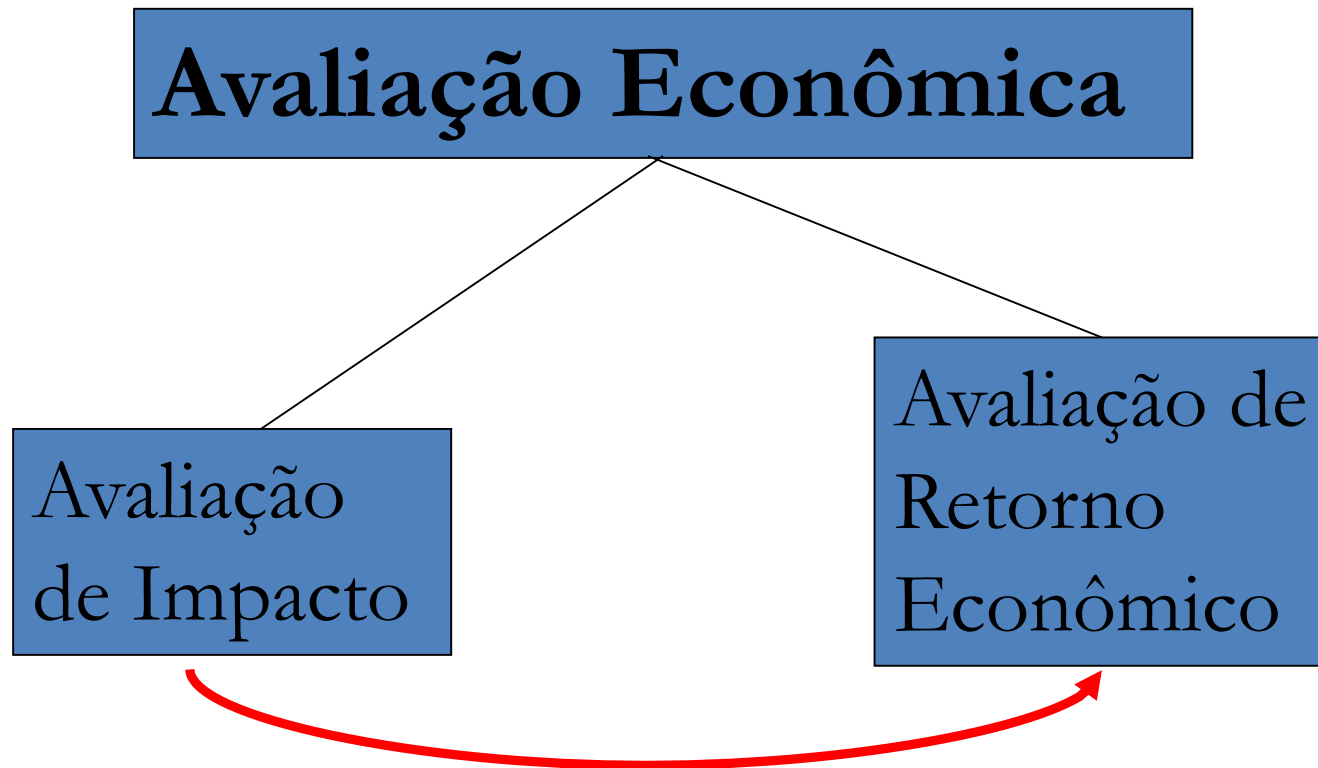
Recursos são escassos → cada R\$ gasto em programas sociais deveria almejar impacto máximo em seus beneficiários

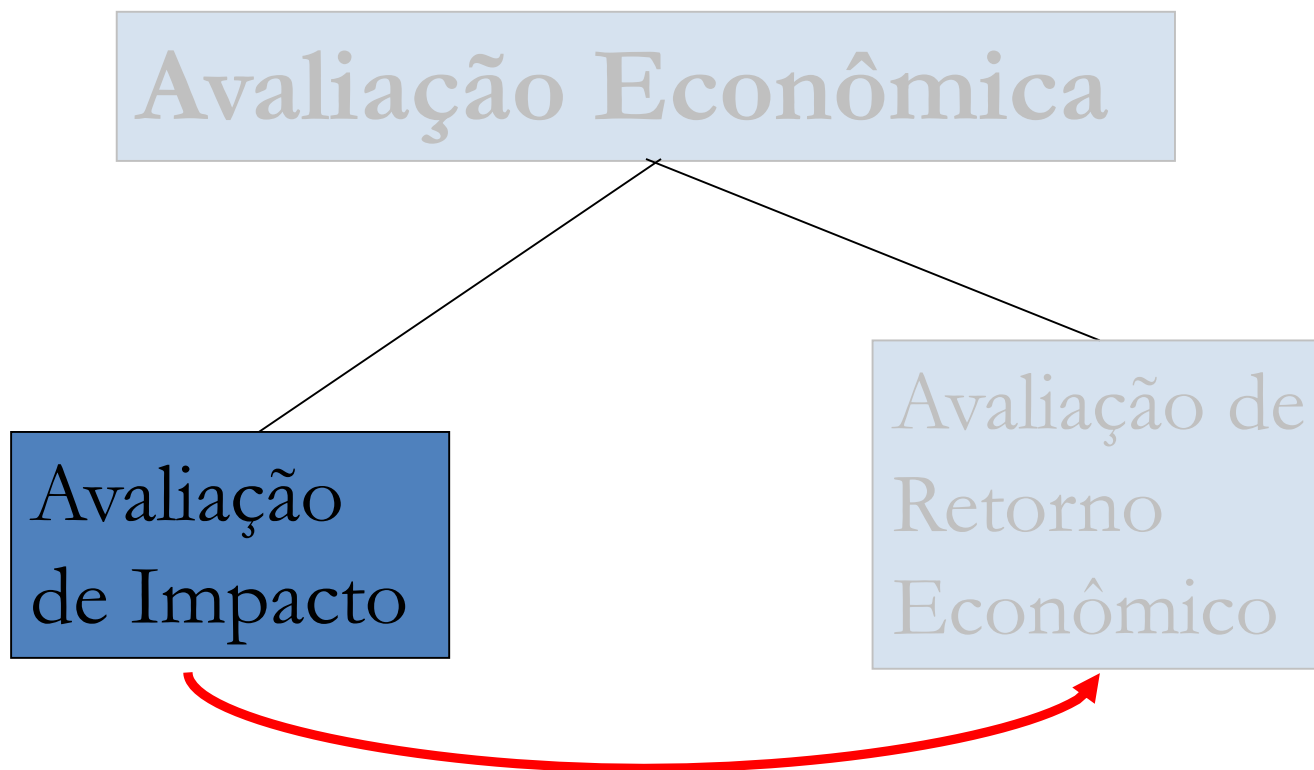


CONTRIBUIÇÃO da Avaliação Econômica para os Projetos Sociais

- Gestão: Ajudar no redesenho e melhora do próprio programa social, juntamente com o monitoramento e outros tipos de avaliação.
- “Accountability”: Ajudar na prestação de contas e obtenção de fundos para novos projetos.
- Focalização: permitir saber se o programa está chegando ao público-alvo.
- Eficiência: otimizar o uso dos recursos disponíveis







Avaliação de Impacto

Determinar:

- Se o projeto teve efeitos sobre público alvo - indivíduos, famílias ou regiões.
- Se os efeitos foram intencionais ou não-intencionais
- Se os efeitos foram REALMENTE resultado do projeto ou ocorreriam de qualquer forma



O PROBLEMA da determinação do impacto

- Como assegurar que os resultados do projeto não foram provocados por outros fatores?
- O que teria acontecido com os participantes caso o projeto nunca tivesse sido realizado?

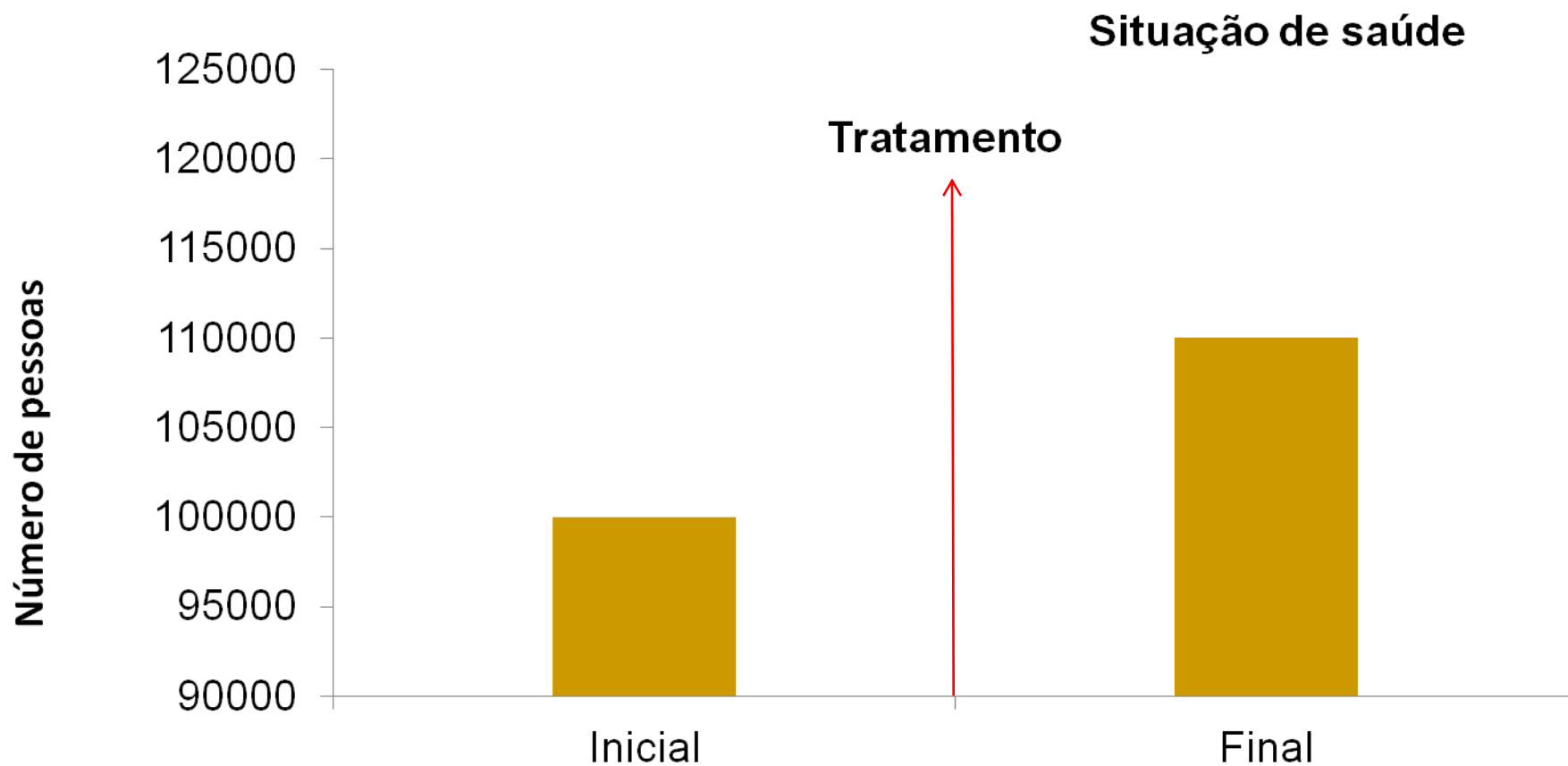
Resposta: o contrafactual



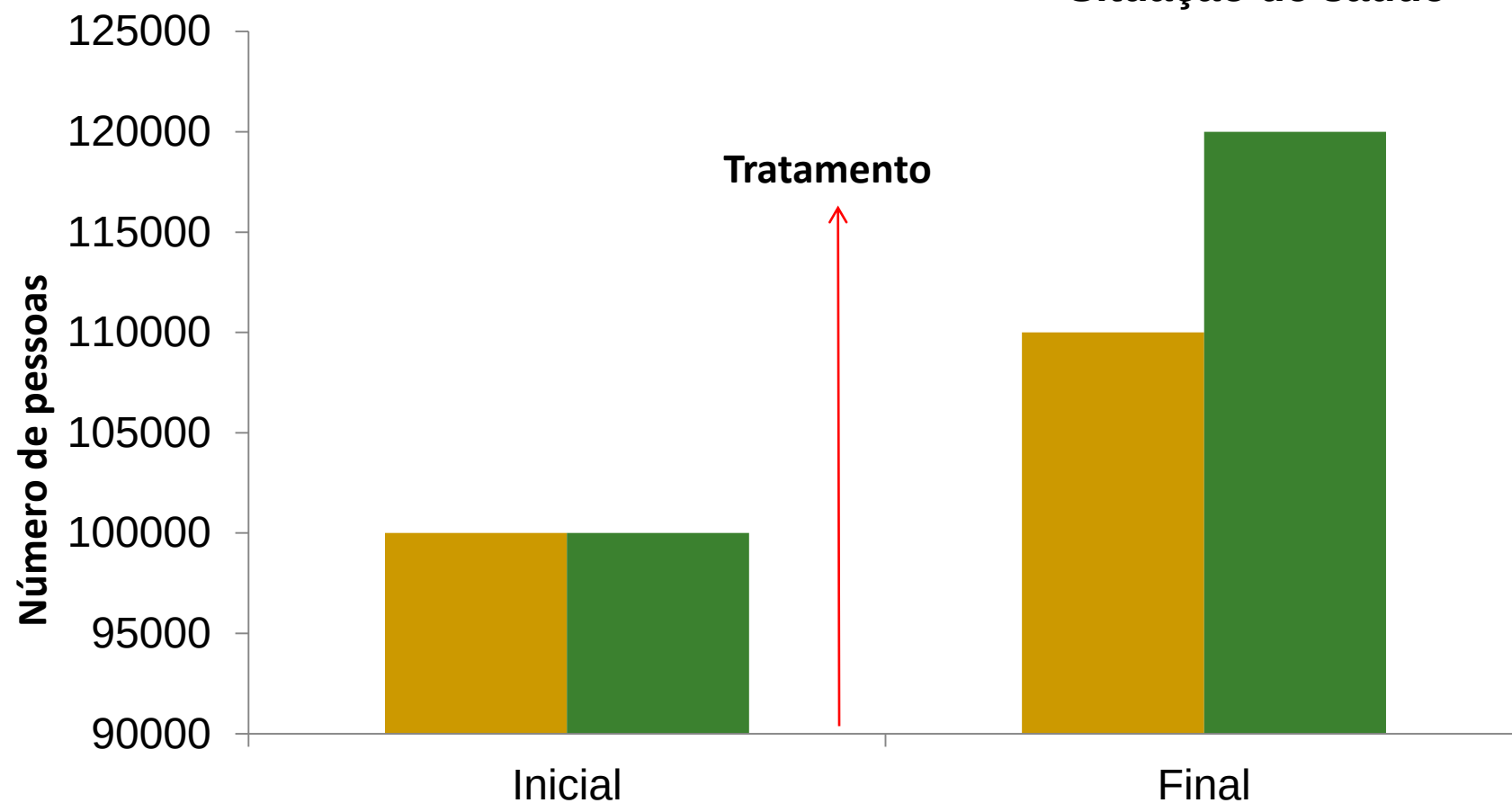
CONTRAFATUAL

- Estratégia: encontrar um contrafactual para o grupo tratado, ou seja, um grupo (de controle) que represente o que teria ocorrido com os tratados caso eles não tivessem passado pelo programa.
- **Não** se recomenda:
 - ✓ Comparar diretamente o grupo tratado com outro grupo que seja muito distinto dele;
 - ✓ Apenas comparar o grupo tratado antes e depois do programa





Situação de saúde

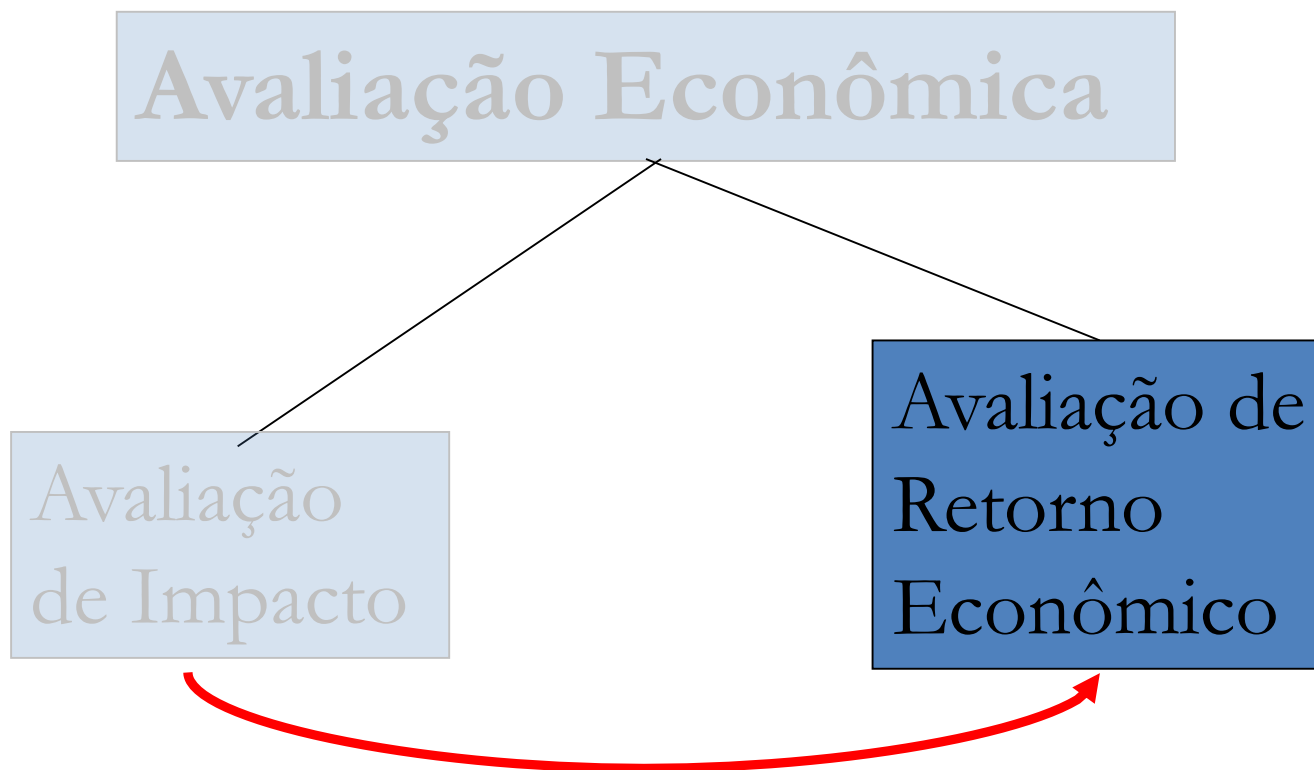


Operacionalizar o contra-factual - GARANTIR que o programa gera efeitos

Utilizar Grupos de Controle que devem ser parecido com o público-alvo, mas não afetados pelo programa

Métodos para construção do contrafactual:
Aleatorização - Sorteio - Planejamento
Técnicas Econométricas - Regressão

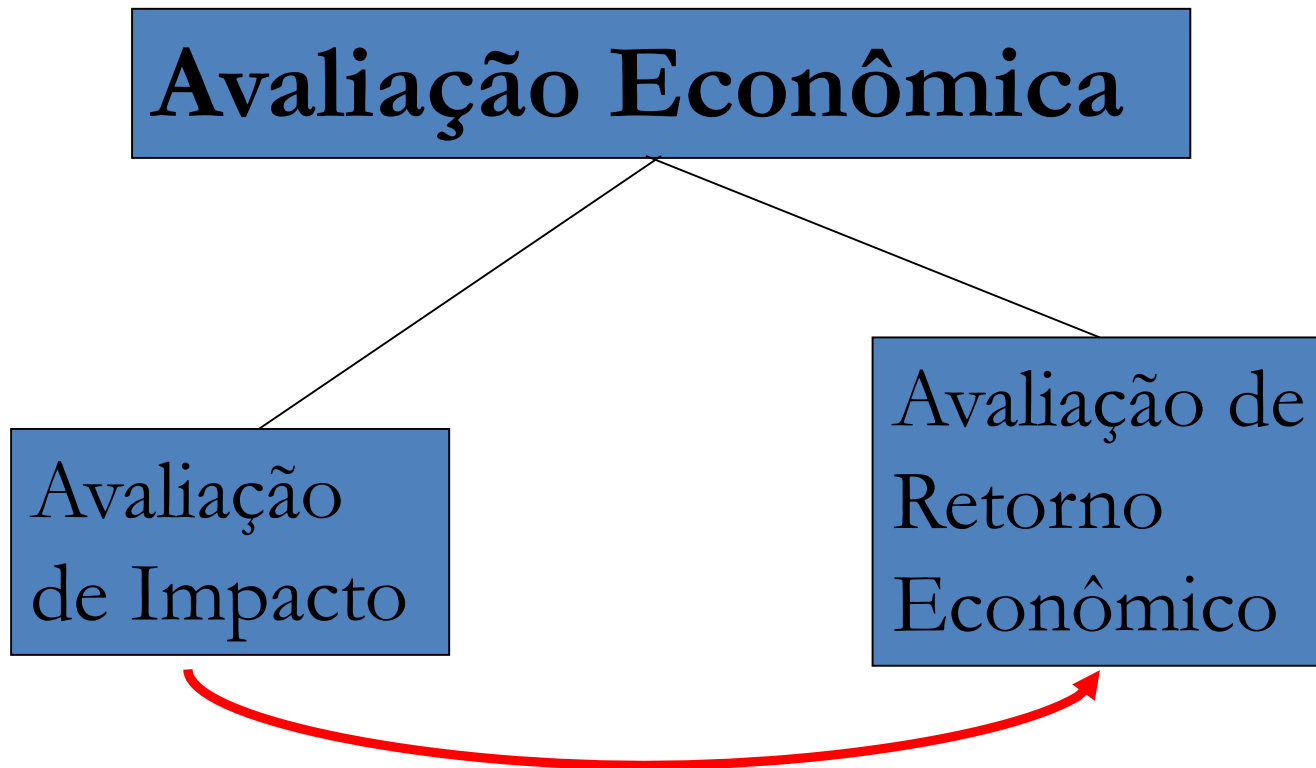




Avaliação do Retorno Econômico

- Transforma os impactos em benefícios monetários para comparar com os custos do projeto (contábeis e de oportunidade)
- Simula os fluxos de benefícios e custos ao longo do tempo
- Calcula o retorno econômico/social:
- Indicadores: Taxa Interna de Retorno (TIR) e Valor Presente Líquido (VPL)





3. Exemplo: PROGRESA

- Transferências em \$ para mães, condicional à frequência escolar e exames de saúde;
- Usando escolas e postos já existentes;
- Avaliação nos estágios iniciais do programa;
- 2.6 milhões de famílias em 72,000 localidades;
- Orçamento: 20% do orçamento nacional para diminuição da pobreza.



Condicionalidades

Educação

Criança com matrícula e frequência à escola (80% das aulas)

Saúde:

Planejamento Familiar, Pré-natal, Vacinações,
Prevenção e tratamento de diarreia, etc...



Caracterização da avaliação de impacto

Avaliação experimental: aleatorizado em nível da comunidade.
RESULTADOS CONFIÁVEIS!

Grupo de controle aderiu ao programa 2 anos depois.

Amostra: 506 comunidades.

186 controle

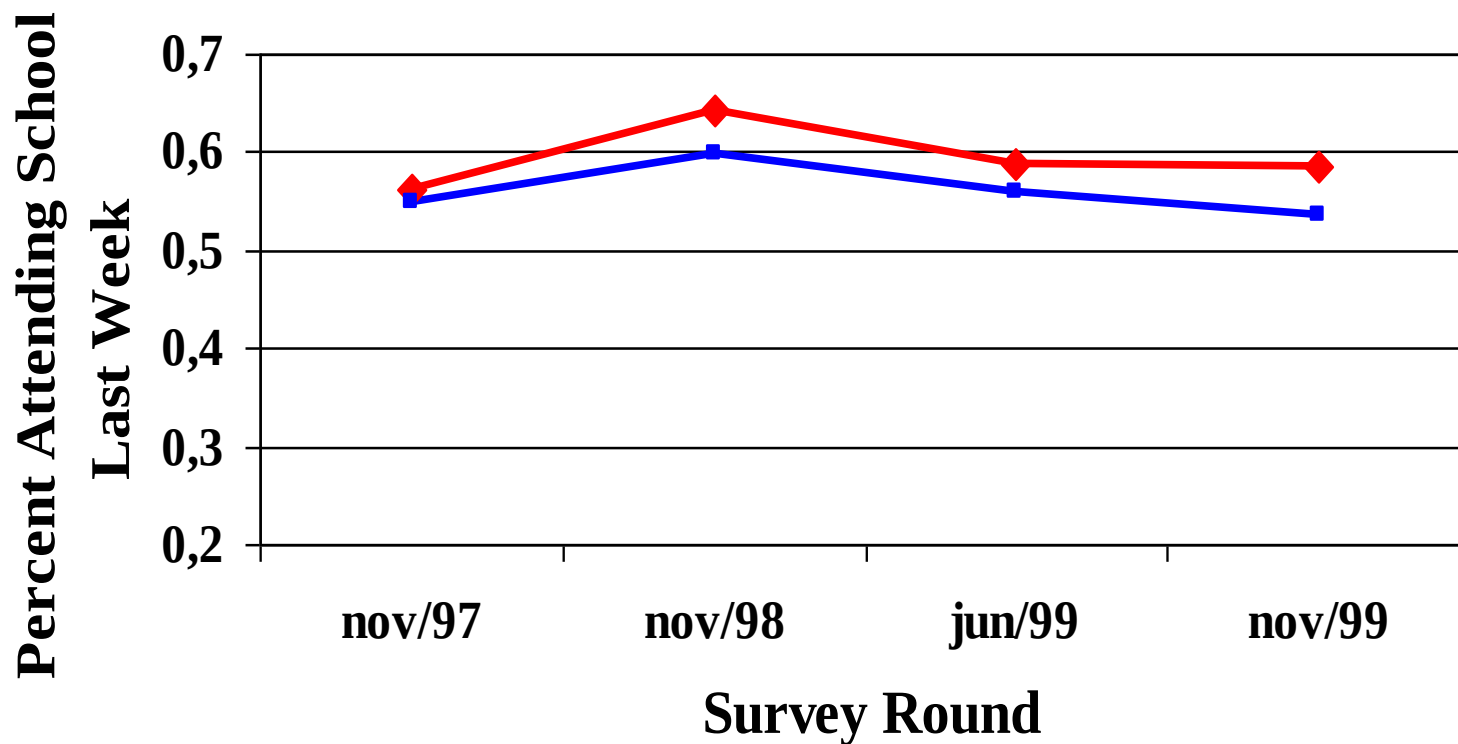
320 tratamento

24.077 Domicílios

Domicílios no grupo de tratamento e controle foram entrevistados antes e depois do começo do PROGRESA.



FREQUÊNCIA ESCOLAR - MENINOS

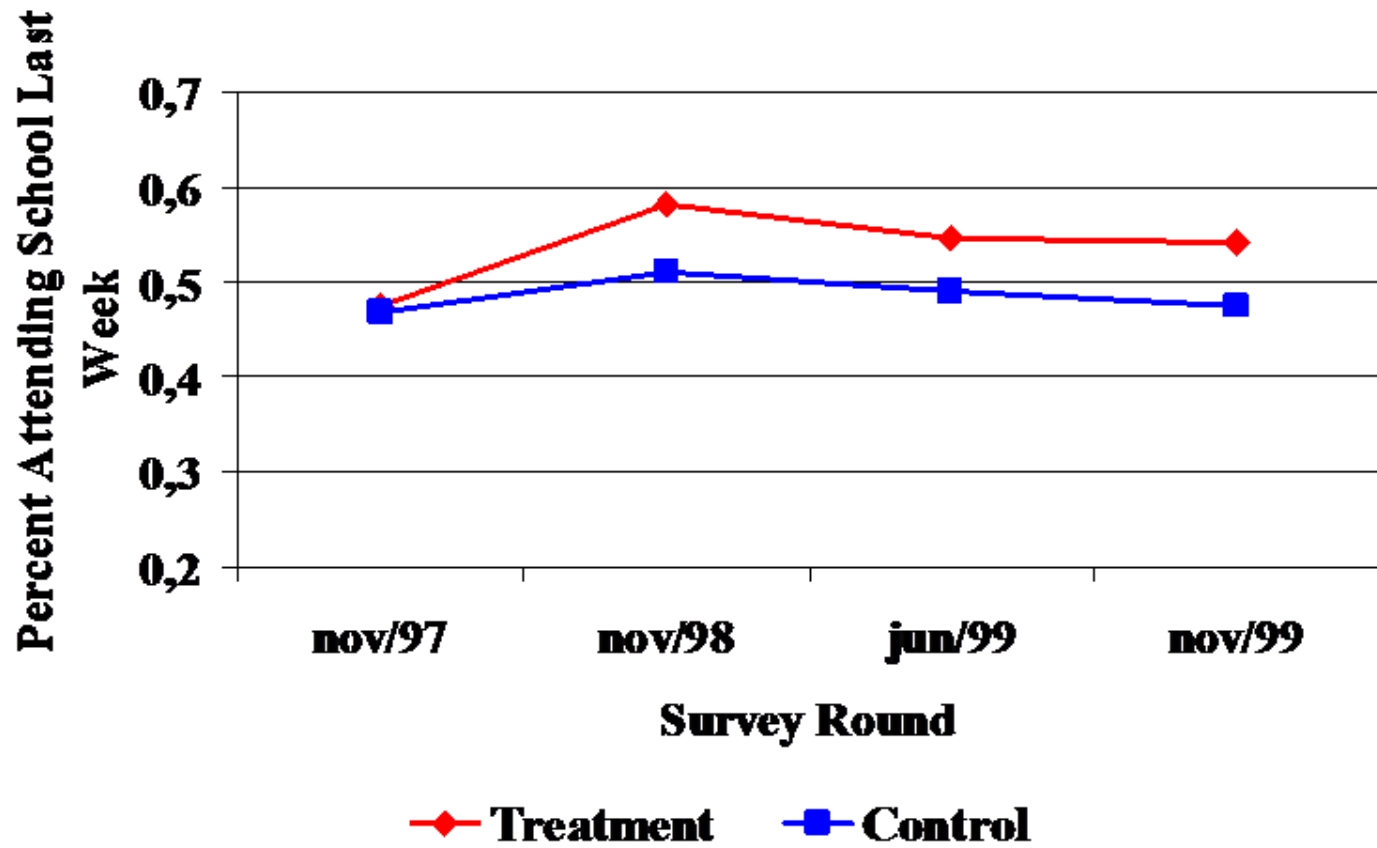


 **Treatment**
 **Control**

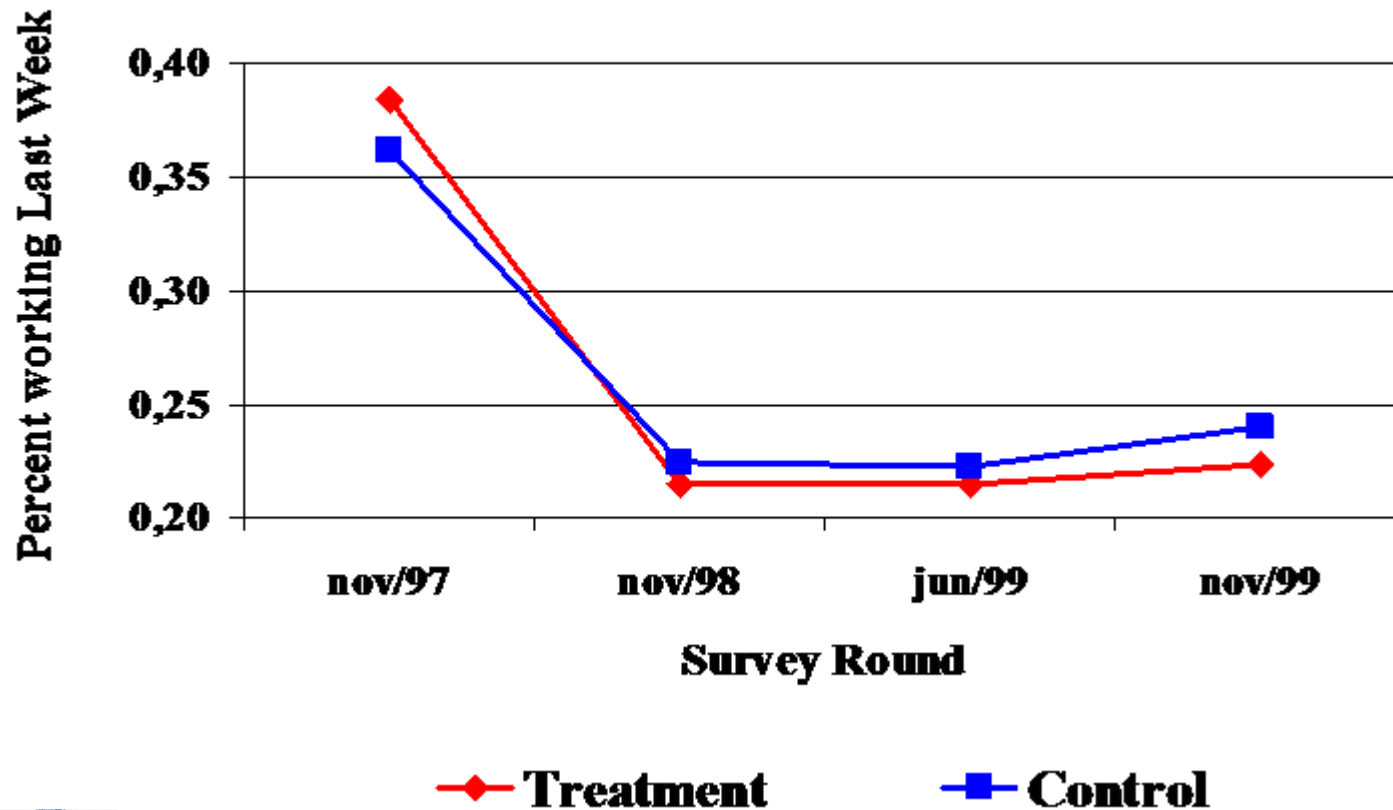
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE DE RIBEIRÃO PRETO
UNIVERSIDADE DE SÃO PAULO



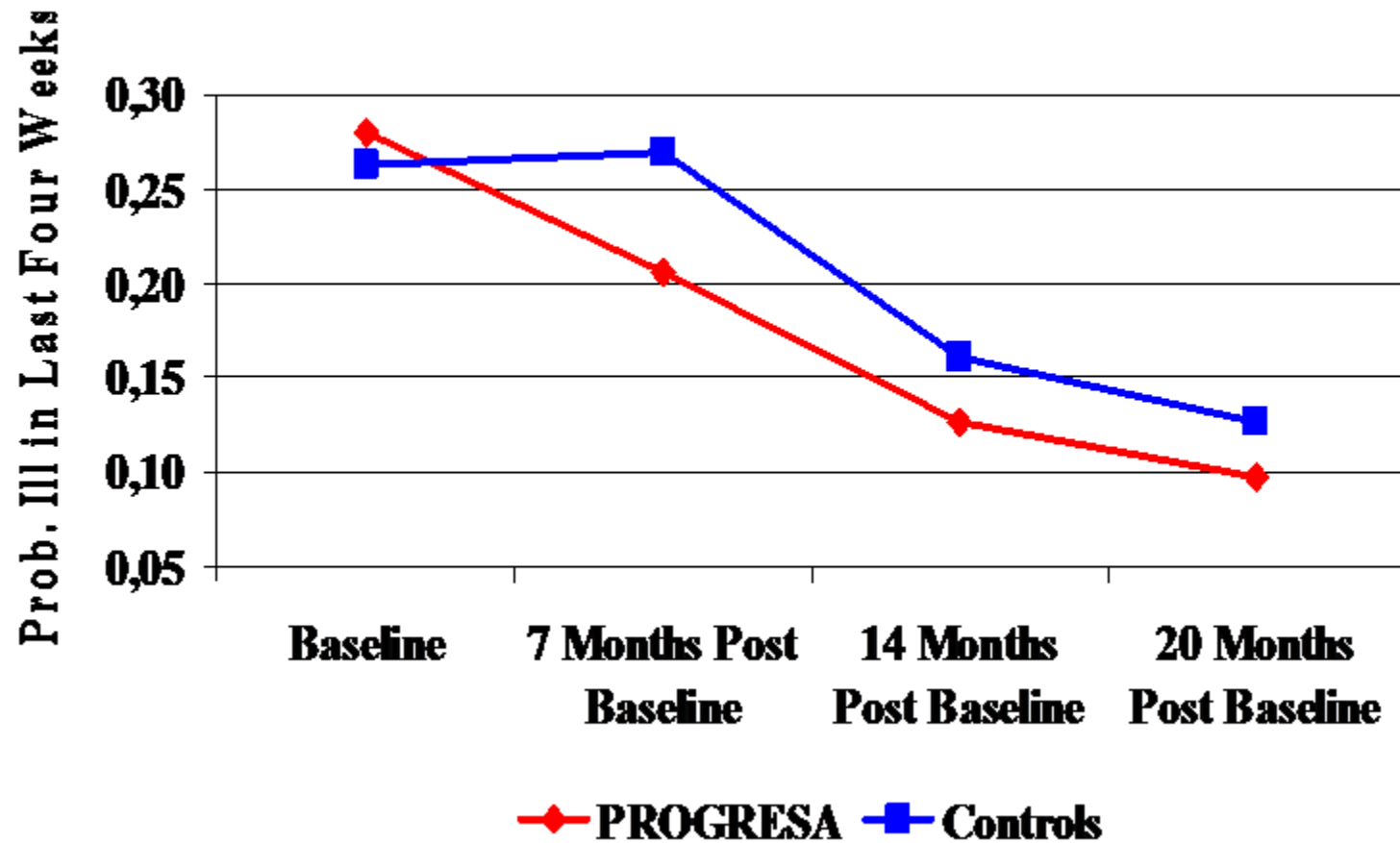
FREQUÊNCIA ESCOLAR- MENINAS



TRABALHO INFANTIL - MENINOS



INCIDÊNCIA DE DOENÇAS 3-5 ANOS



Conclusões da Avaliação de Impacto

- Efeitos significativos no crescimento das crianças (1cm);
- Consumo Domiciliar Total Aumenta;
- Domicílios do PROGRESA “comem melhor”:
Aumento nos gastos com frutas, vegetais, carnes, e produtos animais



Retorno econômico do Progressa

- Para cada 100 pesos alocados no programa, apenas 8.2 pesos são custos administrativos;
- 56% dos custos administrativos são relacionados com focalização e condicionantes;
- Taxa Interna de Retorno: 8%.



Exemplo: Projeto de Inclusão Digital

Capacitar, inserir e acompanhar jovens no mercado de trabalho, desenvolvendo a sua cidadania, a inclusão digital e a formação profissionalizante e empreendedora, por meio de uma rede de parceiros e alianças estratégicas com gestão profissional e foco em resultados.

Município de atuação: SP



Caracterização da avaliação de impacto

Avaliação não experimental: RESULTADOS devem ser olhados com cautela

GRUPO TRATAMENTO:

Indivíduos que realizaram o curso de Informática e Cidadania, inscritas no programa em 2008;

Coleta de dados de 30 indivíduos (via telefone);

GRUPO CONTROLE:

Amostra PME (Pesquisa Mensal de Emprego)

Área de análise da Região Metropolitana de São Paulo.

200 indivíduos selecionados para análise.



Resultado da avaliação de impacto

- Intencionais:

Não houve impacto na empregabilidade;

Não houve impacto no aumento da renda.

- Não Intencionais:

A chance dos indivíduos que participaram do Projeto continuarem a estudar é 25% maior em relação ao Grupo Controle;



Resultado da avaliação de retorno econômico

- Custo do programa foi de 8.720,00 reais.
- Valor presente líquido de 4.544,91 reais.
- Taxa interna de retorno de 25%.



Exemplo: Projeto de Contraturno escolar

Objetivo: Aumentar a proficiência do aluno.

Público alvo: crianças e adolescentes de 6 a 16 anos, com renda per capita inferior a meio salário mínimo e situação de vulnerabilidade social.

Ações: Oficinas lúdicas e acompanhamento escolar. Atendimento odontológico, médico, alimentar, etc.



Caracterização da avaliação de impacto

Avaliação não experimental: RESULTADOS devem ser olhados com cautela.

GRUPO TRATAMENTO:

104 alunos que participaram do projeto com início em janeiro de 2009.

GRUPO CONTROLE:

62 alunos candidatos a participarem do projeto, pertenciam à lista de espera e não foram chamados.



Resultado da avaliação de impacto

- 16% dos alunos que participaram do projeto foram aprovados por causa do projeto.
- 19% dos alunos que participaram do projeto tiveram melhoras nas notas de português.
- O programa não teve impacto nas notas de matemática.
- O programa não teve impacto na frequência escolar.



Resultado da avaliação de retorno econômico

- O custo do programa foi de 607.812,00 reais
- Valor presente líquido de -395.004,53 reais.



4. Conclusões

- Avaliação Econômica de Programas Sociais é um instrumento importante para verificar se programas estão atingindo seus objetivos.
- Importante pensar a avaliação desde o início do programa: definição da amostra, grupos de controle, coleta de informação, etc.
- Calcular impacto e retorno social: ajudar a escolher entre projetos quando recursos são escassos.

Monitoramento e Avaliação

Sistema de Monitoramento e Avaliação

- Monitoramento
- Avaliação de processos
- Avaliação de impacto
- Análise Custo-Benefício



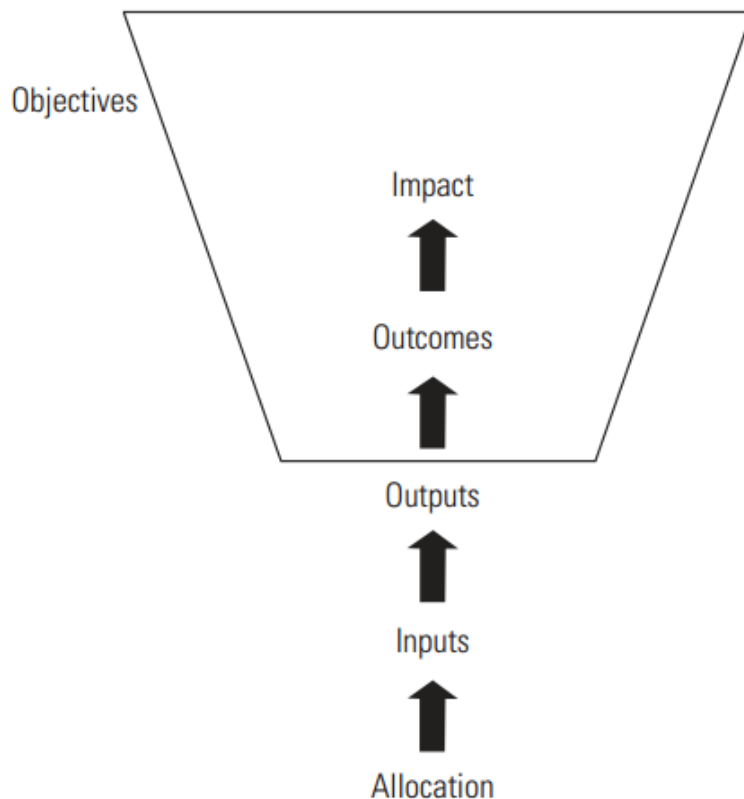
Monitoramento

- Identificar os objetivos que o programa ou estratégia foram desenhados para alcançar.
- Identify key indicators that can be used to monitor progress against these goals.
- Set targets, which quantify the level of the indicators that are to be achieved by a given date
- Establish a monitoring system to track progress toward achieving specific targets and to inform policy makers



Marco Lógico

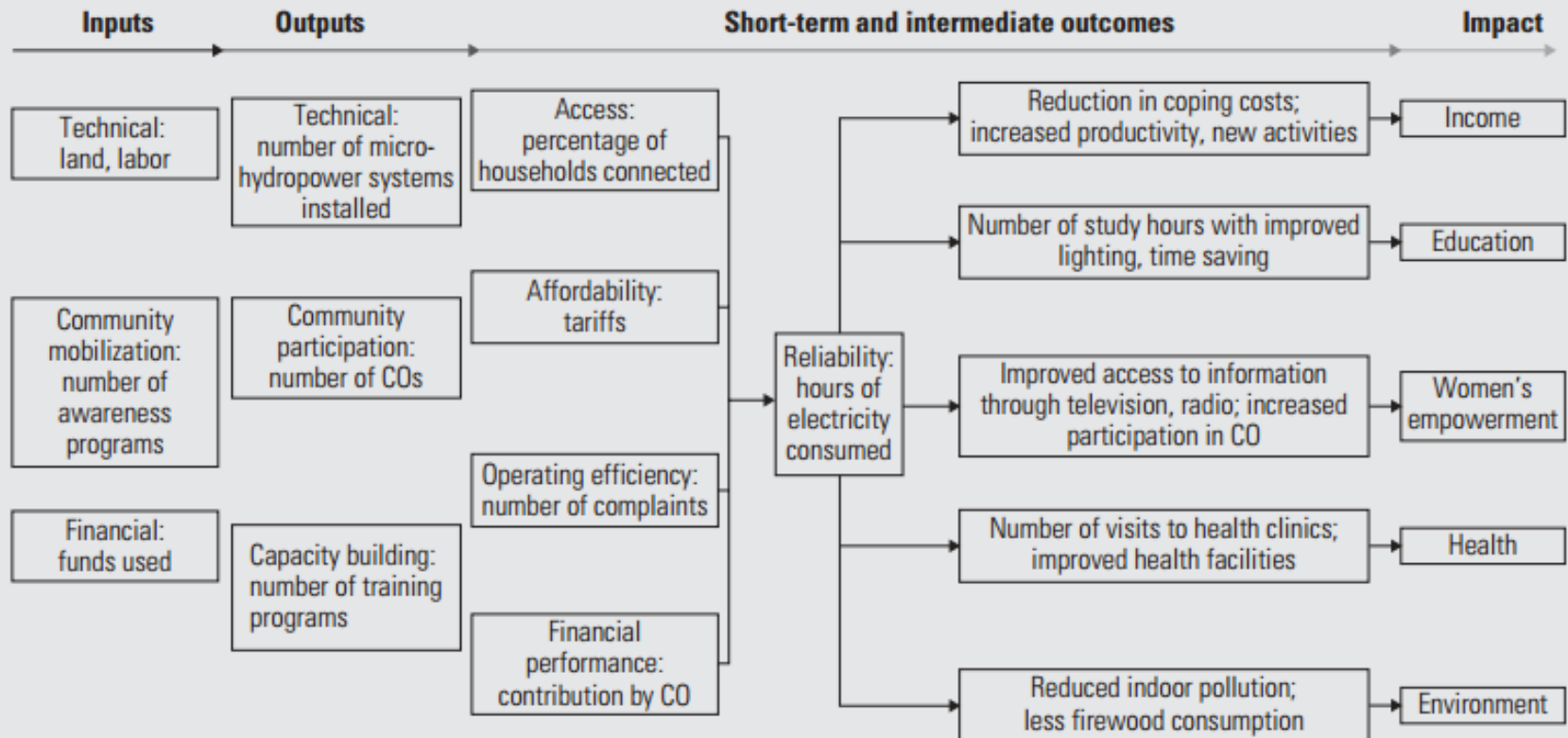
Monitoring and Evaluation Framework



- Outcome: indicador de bem estar final
- Output: indicador intermediário de mais fácil mensuração.
- Ex: Sistema de minicentraís elétricas do Nepal



BOX Figure 2.B Building up of Key Performance Indicators: Project Stage Details



Source: Banerjee, Singh, and Samad 2009.



Avaliação Operacional

- understand whether implementation of a program unfolded as planned.
- retrospective assessment based on initial project objectives, indicators, and targets from the M&E framework.
- based on interviews with program beneficiaries and with officials responsible for implementation.
- the aim is to compare what was planned with what was actually delivered



Avaliação Operacional

- includes monitoring how project money was ultimately spent or allocated across sectors (as compared to what was targeted), as well as potential spillovers of the program into non-targeted areas.



Avaliação de Impacto

- Impact evaluation is an effort to understand whether the changes in well-being are indeed due to project or program intervention.
- Impact evaluation tries to determine whether it is possible to identify the program effect and to what extent the measured effect can be attributed to the program and not to some other causes.
- Operational and impact evaluation are complementary rather than substitutes.



Avaliação Quali x Quanti

- Qualitative information (local sociocultural and institutional context, program and participant details) is essential to a sound quantitative assessment.
- Ex: qualitative information can help identify mechanisms through which programs might be having an impact;
- Ex: surveys can also identify local policy makers or individuals who would be important in determining the course of how programs are implemented,

thereby

Avaliação Quali x Quanti

- Ex: surveys can also identify local policy makers or individuals who would be important in determining the course of how programs are implemented, thereby aiding operational evaluation.
- But a qualitative assessment on its own cannot assess outcomes against relevant alternatives or *counterfactual* outcomes. That is, it cannot really indicate what might happen in the absence of the program.



Avaliação Ex-Ante e Ex-Post

- An ex-ante impact evaluation attempts to measure the intended impacts of future programs and policies, given a potentially targeted area's current situation, and may involve simulations based on assumptions about how the economy works.



Avaliação Ex-Ante e Ex-Post

- Ex post evaluations, in contrast, measure actual impacts accrued by the beneficiaries that are attributable to program intervention.
- Ex post evaluations have immediate benefits and reflect reality. These evaluations, however, sometimes miss the mechanisms underlying the program's impact on the population, which structural models aim to capture and which can be very important in understanding program effectiveness (particularly in future settings).



Avaliação Ex-Ante e Ex-Post

- Ex post evaluations can also be much more costly than ex ante evaluations because they require collecting data on actual outcomes for participant and nonparticipant groups, as well as on other accompanying social and economic factors that may have determined the course of the intervention.
- An added cost in the ex post setting is the failure of the intervention, which might have been predicted through ex ante analysis.



Problema do Contrafactual

- What would have happened to the beneficiaries if the program had not existed.
- A beneficiary's outcome in the absence of the intervention would be its *counterfactual*.
- Ex: se a desnutrição caiu depois de um programa de transferencia de renda, será que a queda pode ser atribuída a ele?



Problema do Contrafactual

- The problem of evaluation is that while the program's impact (independent of other factors) can truly be assessed only by comparing actual and counterfactual outcomes, the counterfactual is not observed.
- The challenge of an impact assessment is to create a convincing and reasonable comparison group for beneficiaries in light of this missing data.



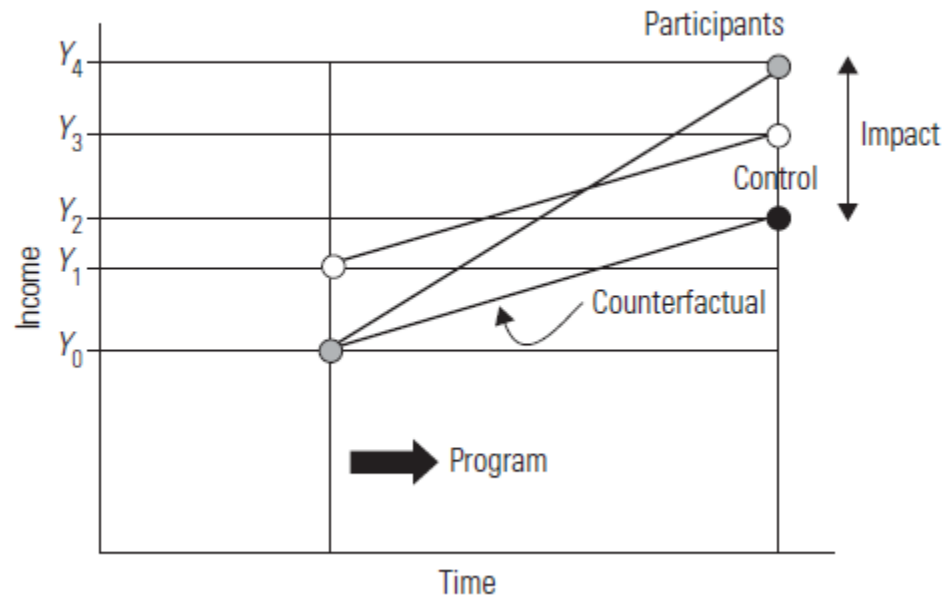
Problema do Contrafactual

- Ideally, one would like to compare how the same household or individual would have fared with and without an intervention or “treatment.”
- But one cannot do so because at a given point in time a household or an individual cannot have two simultaneous existences—a household or an individual cannot be in the treated and the control groups at the same time. Finding an appropriate counterfactual constitutes the main challenge of an impact evaluation.



Problema do Contrafactual

Figure 2.2 Evaluation Using a With-and-Without Comparison



Source: Authors' representation.



Autosseleção

$$Y_i = \alpha + \beta T_i + u_i$$

- Se T não é atribuído aleatoriamente, mas para (a) subgrupos específicos; (b) pessoas que se autosselecionam para o tratamento.
- Autosseleção pode ser baseado em características observáveis, não observáveis ou ambas.
- Em não observáveis, $cov(T_i, u_i) \neq 0$
- MQO viesado e inconsistente



Abordagem dos Resultados Potenciais

$$Y_i = \alpha + \beta T_i + u_i$$

- Abordagem proposta por Rosembaum e Rubin (1982)
- Explicita as hipóteses do modelo
- Seja Y_i a renda per capita do domicílio i . Para os que participam do tratamento, $T_i = 1$ e o valor de Y_i sob o tratamento é $Y_i(1)$
- Para não participantes ($T_i = 0$) e o valor de Y_i é $Y_i(0)$



Efeitos Tratamento

- Note que podemos usar a seguinte notação para Y_i

$$Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$$

- Podemos definir 3 efeitos tratamento:
- ATE = Efeito Tratamento Médio (populacional)

$$ATE = E[Y_i(1) - Y_i(0)]$$

- Mede o efeito de tratamento para a população de interesse



Efeitos Tratamento

- Ex: efeito de programa de voucher para estudantes carentes de 1a serie.
- Nesse caso, o efeito tratamento médio mede qual seria o efeito desse programa para o público-alvo.

$$ATE = E[Y_i(1) - Y_i(0)]$$



Efeito Tratamento sobre Tratados

- É o efeito do tratamento condicional naqueles indivíduos que foram efetivamente tratados:

$$ATT = E\{[Y_i(1)|T_i = 1] - [Y_i(0) |T_i = 1]\}$$

- Note que para cálculo do ATT, também precisamos resolver a questão do contrafactual



Efeito Tratamento sobre Não Tratados

- É o efeito do tratamento condicional naqueles indivíduos que não foram tratados:

$$ATNT = E\{[Y_i(1)|T_i = 0] - [Y_i(0) |T_i = 0]\}$$

- Note que para cálculo do ATNT, também precisamos resolver a questão do contrafactual
- Usado para avaliar o impacto de eventual expansão do programa



Abordagem ingênua e viés

- Imagine que queremos calcular o efeito médio do tratamento sobre os tratados
- Será que é uma boa ideia usar o comportamento dos não tratados como contrafactual dos tratados na ausência do tratamento?
- Talvez não, pois os grupos de tratamento e controle podem ser diferentes entre si mesmo antes do tratamento.



Abordagem ingênua e viés

- Se usarmos simplesmente a diferença entre tratados e controles nos valores de Y , estaremos incorrendo em viés:

$$D = E\{[Y_i(1)|T_i = 1] - [Y_i(0) |T_i = 0]\}$$

- Adicionando e subtraindo $E\{[Y_i(0)|T_i = 1]$ na expressão anterior:

$$D = E\{[Y_i(1)|T_i = 1] - [Y_i(0) |T_i = 1] + [Y_i(0)|T_i = 1] - [Y_i(0)|T_i = 0]\}$$



Abordagem ingênua e viés

$$D = ATT + [Y_i(0)|T_i = 1] - [Y_i(0)|T_i = 0]$$

$$D = ATT + \text{VIÉS}$$

- Ou seja, a diferença da média de tratados e não tratados não irá corresponder ao ATT, a menos que o viés seja zero.
- A hipótese de identificação que garante esse resultado é a...



Hipótese de Ignorabilidade

$$[Y_i(0), Y_i(1)] \perp T$$

- Ou seja, a atribuição dos indivíduos ao tratamento ou ao controle deve ser independente dos resultados potenciais dos mesmos.
- Violamos tal hipótese se atribuímos ao tratamento:
 - Os que chegam primeiro
 - Os mais necessitados
 - Etc.



Exemplo de violação da Ignorabilidade

	$E[Y_i(1) .]$	$E[Y_i(0) .]$
$T_i=1$ (tratamento)	10	6
$T_i=0$ (controle)	8	5



Hipótese de Ignorabilidade

- Se a hipótese de ignorabilidade for válida,
 - $[Y_i(0)|T_i = 1] = [Y_i(0)|T_i = 0]$
- Ou seja, a situação dos não tratados é um contrafactual válido para os tratados efetivamente.
- Quando isso é garantido?
- Aleatorização!!!!



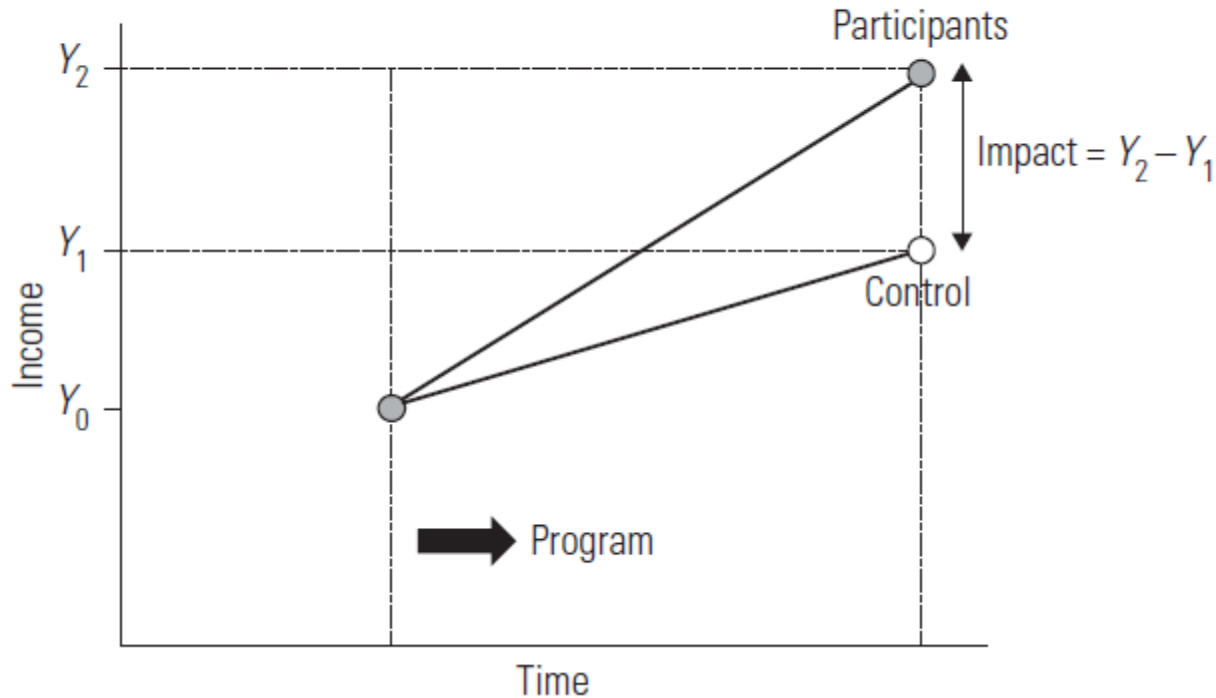
Aleatorização

- Ela envolve dividir aleatoriamente os indivíduos em 2 grupos: tratados e não tratados
- Como a divisão é aleatória, garante-se que todas as características, observáveis ou não, são igualmente distribuídas nos dois grupos antes do tratamento.
- Assim, a única diferença entre os grupos é a atribuição aleatória do tratamento.



Aleatorização

The Ideal Experiment with an Equivalent Control Group



Aleatorização

$$Y_i = \alpha + \beta T_i + u_i$$

- Com aleatorização, $cov(T_i, u_i) = 0$
- β pode ser estimado de modo não viesado por MQO, nessa regressão simples.
- Com aleatorização, $ATE = ATT = ATNT$



Aleatorização

	$E[Y_i(1) .]$	$E[Y_i(0) .]$
$T_i=1$ (tratamento)	10	5
$T_i=0$ (controle)	10	5



Métodos de Aleatorização

- Excesso de inscritos: Se há mais potenciais candidatos do que vagas, pode alocar vagas aleatoriamente entre os elegíveis. Os demais ficam como controle
- Phase-in: Adoção aleatória e gradual do programa em áreas elegíveis, de sorte que os controles são áreas que irão receber o programa futuramente.



Métodos de Aleatorização

- Aleatorização dentro do grupo: No caso anterior, se o lag de tempo de adoção do programa entre áreas for grande, pode gerar controvérsia sobre quem receberia primeiro o programa. Assim, pode-se prover o programa a alguns sub-grupos em cada área, mas tem maior risco de contaminação.



Métodos de Aleatorização

- Desenho de Anúncio Aleatório: O pesquisador aleatoriza um anúncio que incentive a participação no programa a um subconjunto de beneficiários.
- Esse anúncio pode ser usado como uma variável instrumental para a decisão de participar do programa.



Problemas da Aleatorização

- Ética: Aleatorizar, em alguns contextos, pode não ser ético.
- Validade Externa: Programas em pequena escala podem não ter seus resultados generalizáveis para outros lugares.
- Problema de Compliance: quando pessoas que deveriam receber o tratamento não recebem e vice-versa → LATE



Problemas da Aleatorização

- Efeito contaminação: Tratamento contamina o controle e vice-versa.
- Ex: pessoas mudam para cidade que passou a receber o projeto.
- Ex: pessoa que recebe visita de agente de saúde da família repassa instruções aos vizinhos.
- Hawthorne effect: pessoas mudam comp.to. ao saber que estão no experimento.



Problemas da Aleatorização

- Atrito: Pessoas saem do programa.
- Ex: imagine caso de programa de reforço escolar. Se alunos mais fracos tem mais chance de abandonar a escolar, isso pode viesar o impacto do programa positivamente.



Valor do Baseline

- Oportunidade para checar qualidade da aleatorização
- Pode afetar decisão future (Box 3.7)
- Liga Solidária (2014/2015)
- Exercícios: 1 a 6 – cap. 3 - Handbook on Impact Evaluation - Quantitative Methods and Practices - Khandker, Koolwal, Samad



Propensity Score Matching

- Quando tratamento não pode ser aleatorizado, a segunda melhor coisa a se fazer é tentar “similar” uma aleatorização.
- Isso envolve achar uma situação observacional análoga a um experimento aleatório.
- Matching procura encontrar grupo de controle que seja similar ao grupo de tratamento em **características observáveis**.



Propensity Score Matching

- Isso inclui características não afetadas pelo programa (incluindo fatores pré-programa, que não são afetados pela participação no programa).
- A ideia é parear cada participante com um não participante que seja similar em características não observáveis.
- Em seguida, calculamos a diferença no resultado médio dos dois grupos → **efeito tratamento**



Propensity Score Matching

- O problema é encontrar dois indivíduos iguais em todas as características X (mesma idade, escolaridade, região, etc) → **problema de dimensionalidade**
- Rosembaum e Rubin (1983) mostraram que o resultado de se parear os indivíduos de tratamento e controle por X é exatamente o mesmo de parear no **escore de propensão de X** .



Propensity Score Matching

- O escore de propensão é a probabilidade de participação no programa, estimada como função de características observáveis não afetadas pelo programa
- Participantes são pareados (sob diferentes métodos) com base nesse escore de propensão, uma medida unidimensional, resolvendo o problema da multidimensionalidade de X .



Propensity Score Matching

- O escore de propensão é a probabilidade de participação no programa, estimada como função de características observáveis não afetadas pelo programa
- Participantes são pareados (sob diferentes métodos) com base nesse escore de propensão, uma medida unidimensional, resolvendo o problema da multidimensionalidade de X .



Hipóteses necessárias para Propensity Score Matching capturar efeito causal

- **Hipótese de Independência Condicional (CIA)**

$$(y_0, y_1) \perp T | X$$
- Uma vez que controlamos por X , a atribuição do indivíduo ao tratamento ou controle é como se fosse aleatória.
- Chamada também de ignorabilidade ou seleção em observáveis
- Se houver seleção em *não observáveis*, viola CIA



Hipóteses necessárias para Propensity Score Matching capturar efeito causal

- CIA mais fraca utilizada para a identificação do Efeito do Tratamento sobre os Tratados (ATT):

$$Y_0 \perp D|X$$

- Cada indivíduo do grupo de tratamento tem um par no grupo de controle, cujo resultado reproduz o que ocorreria com o indivíduo na ausência de tratamento.



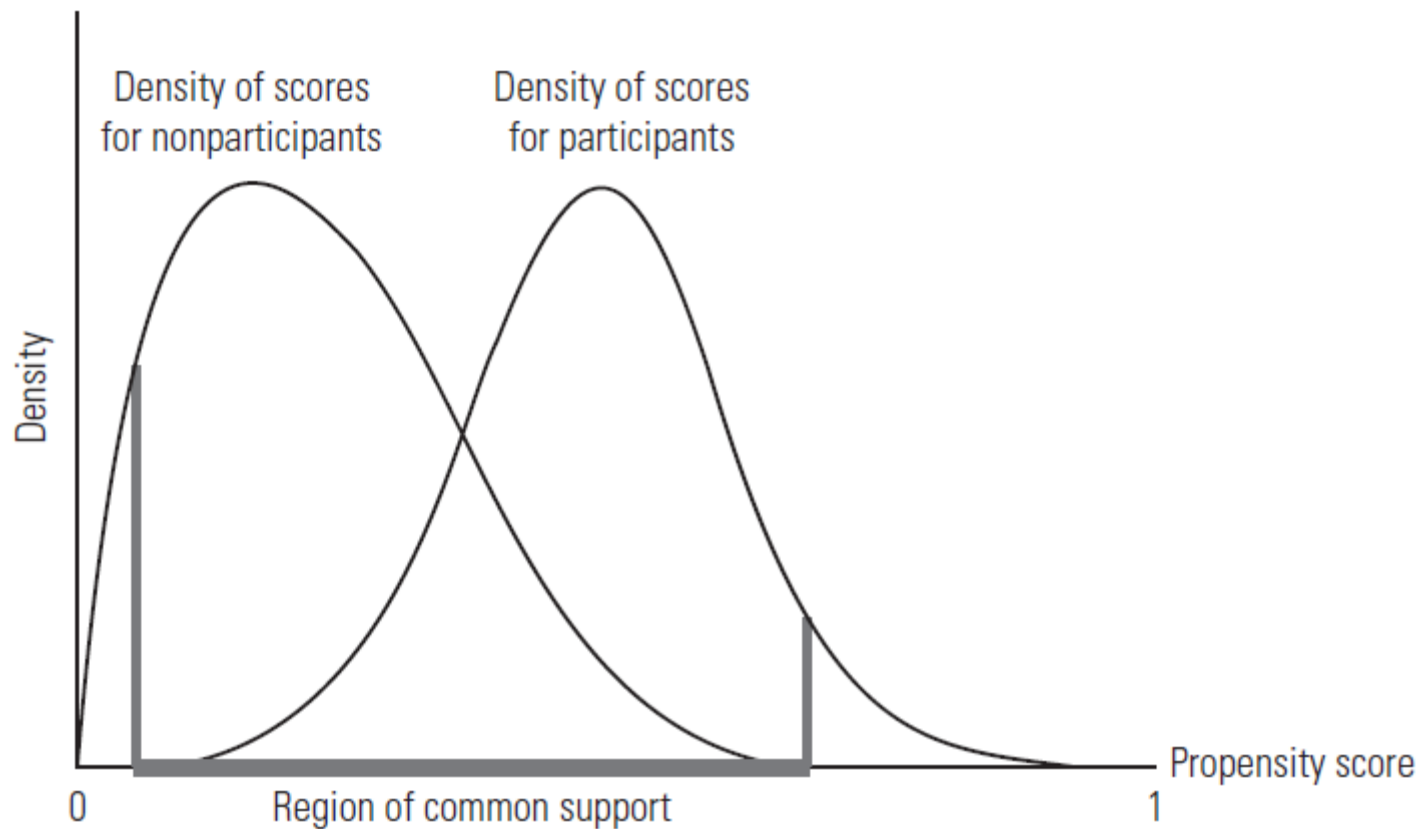
Hipóteses necessárias para Propensity Score Matching capturar efeito causal

- **Hipótese de Suporte Comum (Sobreposição)**

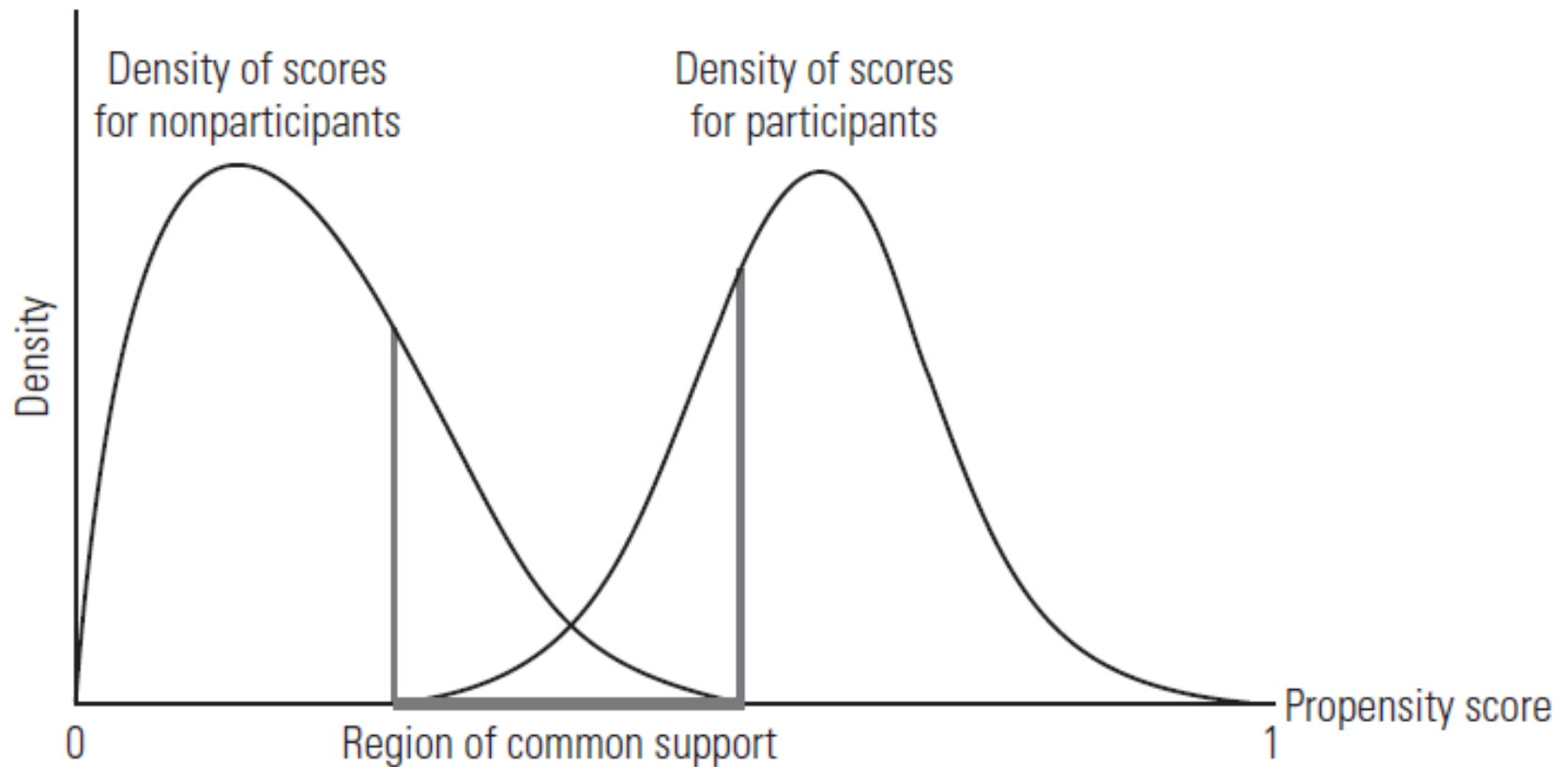
$$0 < Pr[T = 1|X] < 1$$
- Para cada valor de X existem ambos tratados e não tratados. Ou seja, não há valor de X em que todo mundo seja do tratamento.
- Não há valor de X para o qual se possa dizer com certeza a que grupo o indivíduo pertence, usando apenas informação em X.
- Ignorabilidade forte: CIA +suporte comum



Exemplo de Suporte Comum



Exemplo de Suporte Comum Fraco



Estimador de PSM

- Para ATT, hipótese de suporte comum é que:

$$Pr[D = 1|X] < 1$$
- Por que?
- Por que se $Pr[D = 1|X] = 0$, não existe para cada indivíduo do controle um par no grupo de tratamento cujo resultado potencial seja bom preditor do resultado do controle sob tratamento.



Estimador de PSM

$$ATT_{PSM} = \frac{1}{N} \left[\sum_{i \in T} y_i^T - \sum_{j \in C} \omega(i, j) y_i^C \right]$$

- Onde N é o número de indivíduos do grupo de tratamento; $\omega(i, j)$ é o peso usado para agregar os resultados de não tratados.
- Passos:
 - Estimar escore de propensão via logit ou Probit
 - Checar condição de balanceamento e suporte comum
 - Realizar o pareamento



Métodos de Pareamento

- Nearest Neighbour
- Caliper
- Radius
- Estratificação
- Kernel



Métodos de Pareamento

- Nearest Neighbour → envolve escolher com ou sem reposição



Críticas ao PSM

- Se modelo de determinação do tratamento estiver mal especificado, resultados podem ser suspeitos.
- Não requer baseline

