



Inferência Causal e Escore de Propensão

Rosemeire Leovigildo Fiaccone

Prof. Associada do Instituto de Matemática e Estatística da UFBA

Pesquisadora colaboradora CIDACS-Fiocruz Bahia



INSTITUTO DE
**MATEMÁTICA
E ESTATÍSTICA**
UFBA
1968-2020

Moises Augusto

Graduando do Bacharelado em Estatística do
Instituto de Matemática e Estatística da UFBA

1

1

Referências

Livros:

- Gertler P., Martinez S., Premand P., Rawlings L.B., Vermeersch C.M.J. (2011). Impact Evaluation in Practice. Washington, DC: The World Bank.
- Khandker S.R., Koolwal G. B., Samad H.A. (2010). Handbook on Impact Evaluation: Quantitative Methods and Practices. Washington, DC: The World Bank
- Pearl, J., Glymour, M. and Jewell, N.P., 2016. Causal inference in statistics: A primer. John Wiley & Sons.
- Hernán, M.A. and Robins, J.M., 2020. Causal inference: what if. Boca Raton: Chapman & Hill/CRC, 2020

Artigos:

- Holland, P. (1986) "Statistics and Causal Inference", with discussion and rejoinder. *Journal of the American Statistical Association*, 81, 945-970.
- Imbens, G.W. and J.M. Wooldridge (2009) "Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature*, vol. 47(1), 5-86.
- Pearl J. (2009) "Causal inference in statistics: An overview". *Statistics Surveys* Vol. 3 (2009) 96–146.
- Rosenbaum, Paul R. (2005), "Observational Studies", *Encyclopedia of Statistics in Behavioral Science*, Vol. 3, p. 1451-1462

2

2

Sumário



Breve introdução a inferência causal

- Causalidade
- Correlação e causalidade
- Inferência causal: Modelos de resposta potenciais
- O problema fundamental da inferência causal
- Uso no contexto de avaliação de impacto

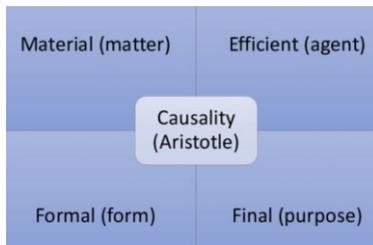
Escore de propensão

- Definição e pressupostos
- Formas de utilização do escore de propensão
- Avaliação do balanceamento e suporte comum
- Implementação no R

3

3

Inferência Causal

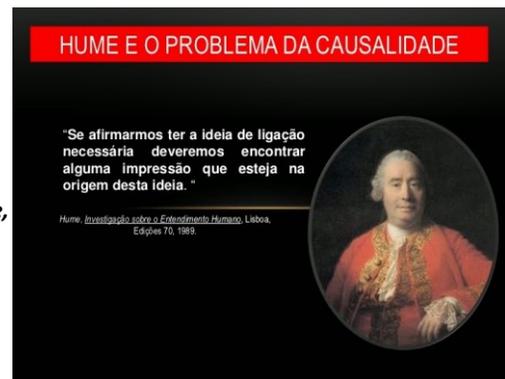


We do not have knowledge of a thing until we have grasped its why, that is to say, its cause.

— Aristotle, *Physics*

A noção de causa e efeito não é uma propriedade da natureza, mas uma criação humana para ordenar o que, em essência, é desordenado

- David Hume (1711-1776)

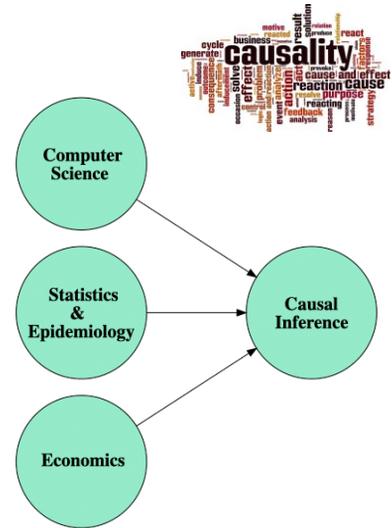


4

Inferência Causal

Conceito ligado as questões

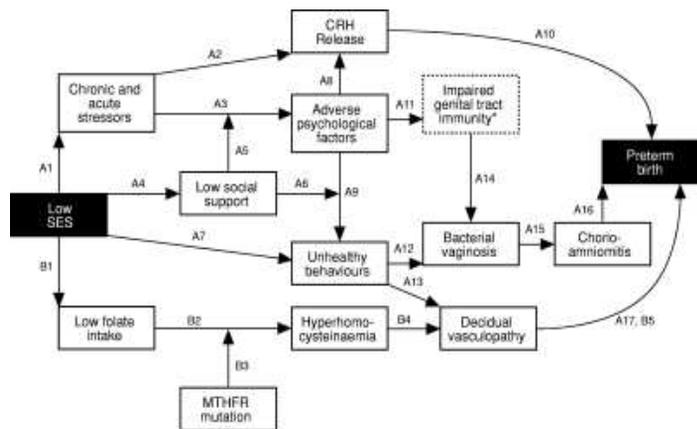
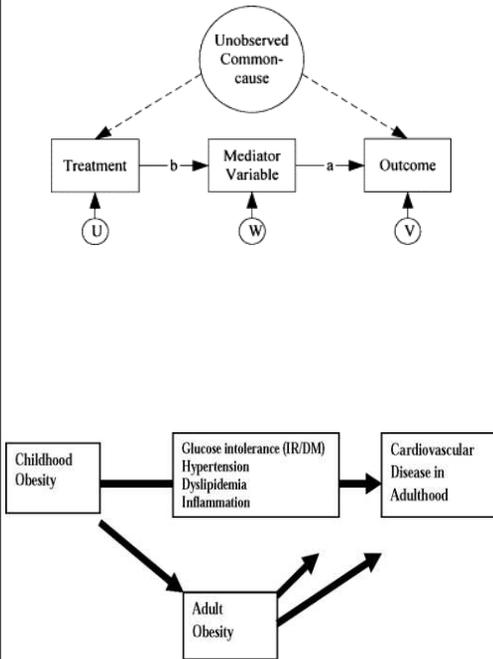
- O que significa causalidade?
- Quais são as causas de um determinado efeito?
- Qual é o mecanismo causal?



5

5

Socio-economic disparities in preterm birth: causal pathways and mechanisms



Paediatric Perinatal Epid, Volume: 15, Issue: s2, Pages: 104-123, First published: 23 December 2001, DOI: (10.1046/j.1365-3016.2001.00012.x)

6

6

Inferência Causal



- Para entender a causalidade relativa a um conjunto de dados precisamos de
 - Uma definição de “causalidade”,
 - Um método que formalize suposições causais, ou seja, definir modelos causais,
 - Um método que permita relacionar o modelo causal com as características do conjunto de dados,
 - Um método que permita tirar conclusões a partir da combinação de suposições causais consideradas em um modelo aplicado a um conjunto de dados.

7

7

Inferência Causal: O que tem de especial?

- Análise estatística tradicionalmente exclui o vocabulário causal da linguagem matemática
 - análise padrão é inferir associações entre variáveis,
 - estimar a probabilidade (ou equivalentemente verossimilhança) de eventos passados e futuros,
 - atualizar a probabilidade de eventos à luz de novas evidências ou novas medidas, usando amostras extraídas de uma população.
- A análise causal incorpora pressupostos sobre o processo de geração de dados, tipicamente derivados de teorias substantivas
 - objetivo é inferir sobre as prováveis consequências da mudança de um estado do mundo (condição de tratamento/decisão).

Medir os efeitos causais

8

8

Correlação e Causalidade



- Correlação **NÃO IMPLICA** em causalidade.

Barnard, G. A. (1982), "Causation," in *Encyclopedia of Statistical Sciences* (Vol. 1), eds. S. Kotz, N. Johnson, and C. Read, New York: John Wiley, pp. 387–389.

- Associação significa que duas variáveis se movem juntas

Lembrete: duas variáveis aleatórias T e Y são marginalmente independentes se

$$P(Y=y/T=t) = P(Y=y / T=t^*) \quad \forall y, t \text{ e } t^* \quad \Rightarrow \quad Y \perp T$$



- **Causação** significa que uma das variáveis está causando (pelo menos parte) os movimentos na outra

9

9

Causação



- **Condições necessárias para causação**

X tem um efeito causal em Y se:

1. X e Y estão associados (condição de relacionamento)
2. X precede temporalmente/logicamente Y (condição antecedente)
3. A associação observada entre X e Y não pode ser explicada por fatores não controlados no estudo (condição de falta de explicações alternativas)

Nota: também pode acontecer que não encontremos nenhuma associação entre duas variáveis por causa de terceiros fatores! Então, sempre precisamos para verificar a condição 3 (e 2)

10

10

Guia segundo Bradford Hill : Causalidade em Pesquisa na Área de Saúde

- Força da associação - magnitude do efeito
- Consistência - generabilidade da associação em várias populações
- Especificidade - causa é específica para o efeito
- Temporalidade - causa precede efeito
- Relação dose-resposta - aumento da exposição aumenta efeito
- Plausibilidade – Coerência e analogia
- Evidência experimental - usado para mostrar o efeito

Contudo, nenhum dos critérios com exceção da “temporalidade” (causa precede o efeito) são necessários ou suficiente

11

11

Procedimentos de Inferência Causal



Estruturas (ou modelos) para inferência causal :

- *Modelos de Equações Estruturais* (Wright, 1928)
- Os diagramas causais ou os diagramas chamados DAGs, sigla em inglês para *Directed Acyclic Graphs* (Pearl, 1995 e 2009, Dawid, 2002, VanderWeele, Robins, 2007)
- *Modelos de respostas potencias* ou estruturas contrafactuais (Neyman, 1923; Rubin, 1974; Imbens and Rubin, 2015; Hernan and Robins, 2020) ←

12

12

Modelos de Respostas Pontenciais

- Referido como modelo Neyman-Rubin ou modelo Rubin.
 - Fundamentado no trabalho estatístico sobre experimentos aleatórios de Fisher (1918, 1925)
 - Proposto por Neyman (1923) no contexto de estudos agrícolas
 - Estendido por Rubin (1974, 1978) para aplicar a estudos não randomizados
- Denominado de Modelo Causal de Rubin (Holland, 1986) por visualizar a inferência causal como um problema de dados omissos com a modelagem explícita do mecanismo de atribuição como um processo para revelar o dados observados
- Estabelece uma relação causal entre uma causa potencial (tratamento) e seu efeito (resultado).
- Obs: Variáveis como raça, gênero e tempo não são “causas” bem definidas na estrutura de resposta potencial

13

13

Terminologia / Notação



- Unidade (i) : A pessoa, lugar ou coisa sobre a qual um tratamento irá operar, em um determinado momento
- Tratamento (T) : Uma intervenção, cujos efeitos (em algum medida das unidades) o investigador deseja avaliar em relação a nenhuma intervenção (ou seja, o controle)

Exemplo: Estudo sobre formas de reduzir o nível de colesterol das pessoas, utilizar uma dieta vegetariana é o tratamento ($T=1$) e ausência de dieta ($T=0$) seria o controle
- Resultados Potenciais [$Y(1)$ e $Y(0)$]: Os valores da medida de interesse de uma unidade após (a) aplicação do tratamento e (b) não aplicação do tratamento (ou seja, sob controle)
- **Efeito Causal:** Para cada unidade, a comparação do resultado potencial sob tratamento e o resultado potencial sob controle, ou seja, $Y_i(1) - Y_i(0)$
- Covariáveis observáveis ou confundidores (X)
- Covariáveis ou confundidores não observáveis (U)

14

14

Modelo de Respostas Potenciais

- O tratamento T tem um **efeito causal** sobre uma resposta individual Y se

$$Y(1) \neq Y(0) \text{ para o indivíduo } i.$$

- Efeito causal individual **não** pode ser identificado! impossível observar simultaneamente o valor de Y quando submetidos ao tratamento e ao não-tratamento na mesma unidade
- Somente uma resposta potencial é observada em um dado ponto no tempo para cada unidade

Problema fundamental da inferência causal

Grupo	$Y(1)$	$Y(0)$
Tratamento	Observável (Y)	Contrafactual
Controle	Contrafactual	Observável (Y)

15

15

Modelo de Respostas Potenciais



- Essa estrutura de resposta potencial** é um problema de dados omissos, ou seja, qual seria a resposta potencial contrafactual de $Y_i(T)$?

Por exemplo, o que teria acontecido com $Y_i(T)$ caso essa unidade i tivesse sido submetida ao controle?

- Então a resposta potencial e o mecanismo de atribuição determinam conjuntamente o valor da resposta observada e da resposta omissa

$$Y_i^{OBS} = Y_i(T_i) = T_i * Y_i(1) + (1 - T_i) * Y_i(0)$$

$$Y_i^{OMIS} = Y_i(1 - T_i) = (1 - T_i) Y_i(1) + T_i * Y_i(0)$$

- As principais suposições da identificação na inferência causal estão no mecanismo de atribuição

16

16

Modelo de Respostas Potenciais

- Como solução, no caso em que as unidades são indivíduos, faz-se uso da estatística substituindo a impossibilidade de observar o efeito causal em um indivíduo específico pela possibilidade de estimar o "**efeito causal médio**" em uma população de indivíduos.
- Rubin (citado por Holland13) aponta que uma condição necessária para identificação de um contrafactual é a Suposição de Valor Estável da Unidade de Tratamento (SUTVA, Stable-Unit-Treatment-Value Assumption)

i) Sem "interferência" ou "efeito de transbordamento" entre as unidades

$Y_i(1)$ e $Y_i(0)$ NÃO são afetados por qual ação outra unidade j recebeu => indivíduos são independentes entre si

- Esta premissa garantiria que o efeito médio estimado a partir da amostra seja igual ao efeito médio calculado a partir da população

ii) Nenhuma interação entre as unidades

Não há versões ocultas dos tratamentos e cada unidade é afetada pelo próprio status de tratamento

- Não há características relevantes do tratamento que variem entre as unidade de um mesmo grupo

Suposição também conhecida como suposição do equilíbrio parcial

17

17

Modelo de Respostas Potenciais

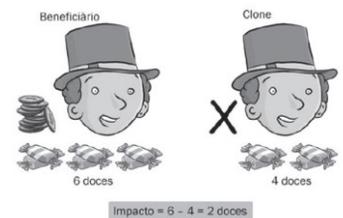


- Encontrar um grupo de controle que possa **replicar** o grupo tratado caso ele não tivesse passado pela intervenção – **contrafactual**

- Grupo de comparação deve ser igual ao grupo sob tratamento nos seguintes aspectos:

- São idênticos na ausência do programa: características compatíveis;
- Têm reação similar se expostos ao programa;
- Nenhum dos dois pode estar exposto diferencialmente a outras intervenções durante o período de avaliação.

Figura 3.1 O Clone Perfeito



- A escolha desse grupo de comparação envolverá o uso de **procedimentos** e **hipóteses** cuja finalidade básica é **minimizar** o que se costuma denominar de **viés de seleção**

18

18

Modelo de Respostas Potenciais

- A ideia básica é a comparação dos resultados do grupo de tratamento aos de um grupo controle representando o contrafactual (respostas potenciais semelhantes)
- *No caso de intervenções aleatórias:*
- Temos $Y_i(1)$ (observado)
 - precisamos prever a resposta se fosse controle
 - usaremos a resposta média no grupo controle para prever

T	Y_i^{OBS}	$Y_i(0)$	$Y_i(1)$
0	$Y_i(0)$	Ok	?
1	$Y_i(1)$?	ok

T	Y_i^{OBS}	$Y_i(0)$	$Y_i(1)$
0	$Y_i(0)$	ok	?
1	$Y_i(1)$?	ok

- **Problema:** Nos estudos observacionais essa condição em geral não é verdadeira devido aos **confundidores**
 - Uma característica (sócio-demográfica) que afeta a probabilidade de receber o tratamento além de poder estar relacionada a resposta de interesse
 - Pode haver desequilíbrio dessas características entre os grupos pelo fato do tratamento não ser randomizado
- **Ajuste:** Compare-se os resultados de unidades tratadas/controle que são semelhantes em termos de todos os fatores de confundimento relevantes
- Temos $Y_i(1)$ (observado)
 - precisamos prever a resposta se fosse controle
 - usaremos a resposta média no grupo controle para prever **utilizando as unidades com características similares**

19

19

Por que estudar Inferência Causal?



- Necessidade de entender os mecanismos que explicam fenômenos estudados a partir de dados coletados
 - Exemplo: Suponha que desejamos avaliar o impacto de um programa de assistência técnica sobre a renda de um conjunto de produtores agrícolas
- Avaliação de impacto tenta fazer um nexo de causalidade entre um programa ou intervenção e um conjunto de resultados
 - Avaliação de Impacto Ambiental (AIA) busca identificar, mitigar e avaliar os potenciais impactos socioambientais de obras, atividades ou projetos de significativo impacto ambiental.
 - Na área de saúde é a combinação de procedimentos, métodos e ferramentas com os quais políticas, programas e projetos podem ser julgados pelos seus efeitos potenciais na saúde da população, assim como pela distribuição de tais efeitos dentro da população

20

20

Exemplos

Box 1.1: Evaluations and Political Sustainability

The Progresa/Oportunidades Conditional Cash Transfer Program in Mexico

In the 1990s, the government of Mexico launched an innovative conditional cash transfer (CCT) program called "Progresa." Its objectives were to provide poor households with short-term income support and to cre-

BOX 2.1 Case Study: PROGRESA (Oportunidades) in Mexico

Monitoring was a key component of the randomized program PROGRESA (now called Oportunidades) in Mexico, to ensure that the cash transfers were directed accurately. Program officials foresaw several potential risks in implementing the program. These risks included the ability to

- ❖ Suporte de renda para famílias de baixa renda incentivando investimento nas crianças;
- ❖ Participação condicional às crianças irem regularmente à escola e visitarem centros de saúde

Box 1.2: Evaluating to Improve Resource Allocations

Family Planning and Fertility in Indonesia

In the 1970s, Indonesia's innovative family planning efforts gained international recognition for their success in decreasing the

average age had finished primary education. During the same period as the family planning program, however, the government

Recursos realocados para programas de escolaridade de mulheres

21

21

Exemplos de Perguntas de Avaliação



- Qual é o efeito do programa de Subsídio ao Seguro Saúde sobre os gastos das famílias com saúde?
- Que combinação de campanhas pelo correio e orientação familiar funciona melhor para estimular o aleitamento materno como única fonte de alimentação?
- Os programas de transferência de renda geram resultados melhores na saúde das crianças?
- Os programas de capacitação profissional aumentam a renda dos indivíduos capacitados?

22

22

- Esta pergunta pode ser aplicada a muitos contextos em saúde, por exemplo:



Qual o impacto da Estratégia de Saúde da Família no perfil de morbidade hospitalar no município de Minas Gerais?

Qual o impacto do Programa Bolsa Família no status nutricional das crianças beneficiárias?

Impacto da Estratégia Saúde da Família sobre o perfil de morbidade hospitalar em Minas Gerais*
The Family Health Strategy impact on the hospital morbidity profile in Minas Gerais

Antônio Gonçalves Maciel^a, Antônio Pires Cabral^a, Francisco José Lopes de Sousa Diniz^a

RESUMO Este estudo verifica a potencialidade da Estratégia Saúde da Família em reduzir a morbidade hospitalar por condições sensíveis à atenção primária em Minas Gerais. Utilizou-se a análise longitudinal retrospectiva em três níveis de agregação: a nível agregado (município de referência), a partir de indicadores socioeconômicos, observou-se o percentual de cobertura capaz de explicar a redução das taxas de morbidade hospitalar. Foi demonstrada a redução de 68,67 pontos nas taxas de hospitalizações por causas não sensíveis, contra o aumento de 1,87 pontos nas taxas de hospitalizações por causas não sensíveis. Os resultados refletem que a estratégia tem implicações para a gestão da saúde no estado de Minas Gerais.

PALAVRAS-CHAVE Avaliação em saúde, Indicadores de morbimortalidade, Nível de saúde, Morbidade, Gestão de serviços de saúde.

ABSTRACT This study verifies the Family Health Strategy potential in reducing the hospital morbidity in the state of Minas Gerais by the indicator of hospitalization for sensitive conditions in primary care. The retrospective longitudinal analysis was used in a three-level system. The techniques used the statistical regression model and, according to the final outcomes, observed the percentage of coverage able to impact in reducing rates of hospital morbidity. The rate of hospitalization for sensitive conditions in primary care was reduced by 68,67 points, against a 1,87 points increase on the reduction for non-sensitive conditions. The results indicate the implications this study has for the health management in Minas Gerais.

KEYWORDS Health evaluation, Indicators of morbidity and mortality, Health status, Morbidity, Health services administration.

Effects of a conditional cash transfer programme on child nutrition in Brazil

Rômulo Paes-Sousa,^a Leonor Maria Pacheco Santos^a & Édina Shisue Mizak^b

Objective To examine the association between Brazil's Bolsa Família programme (BFP), which is the world's largest conditional cash transfer programme, and the anthropometric indicators of nutritional status in children.

Methods Using the opportunity provided by vaccination campaigns, the Brazilian government promotes Health and Nutrition Days to estimate the prevalence of anthropometric deficits in children. Data collected in 2005–2006 for 22 375 impoverished children under 5 years of age were employed to estimate nutritional outcomes among recipients of Bolsa Família. All variables under study, namely child birth weight, lack of birth certificate, educational level and gender of family head, access to piped water and electricity, height for age, weight for age and weight for height, were converted into binary variables for regression analysis.

Findings Children from families exposed to the BFP were 20% more likely to have normal height for age than those from non-exposed families; the difference also applied to weight for age. No statistically significant deficit in weight for height was found. Stratification by age group revealed 19% and 41% higher odds of having normal height for age at 12–35 and 36–59 months of age, respectively, in children receiving Bolsa Família, and no difference at 0–11 months of age.

Conclusion The BFP can lead to better nutritional outcomes in children 12 to 59 months of age. Longitudinal studies are needed to confirm these findings.

Importante



- A questão básica da avaliação de impacto constitui, essencialmente, um problema de inferência causal.

Qual é o impacto (ou efeito causal) de um programa sobre um resultado de interesse?

Observação: Não é uma questão simples determinar se a relação é causal.

Parâmetros Causais

- Dois parâmetros são mais frequentemente estimados na literatura: ATE e ATT.
- ATE - Efeito causal médio do tratamento em toda a população:

$$ATE = E[Y(1)-Y(0)] = E[Y(1)] - E[Y(0)]$$
- Qual é o efeito esperado do tratamento em uma unidade selecionada aleatoriamente da população?"
- Heckman et al (1997) observam que esta estimativa pode não ser relevante para os líderes governamentais porque inclui o efeito sobre as pessoas para as quais o programa nunca foi planejado

Exemplo: Se um programa é direcionado especificamente para indivíduos com baixa renda familiar, há pouco interesse no efeito de tal programa para um milionário.

25

25

Parâmetros Causais



- ATT: O efeito causal médio do tratamento na subpopulação de tratados (ATT):

$$ATT = E(Y(1) - Y(0) | T = 1) = E(Y(1) | T = 1) - E(Y(0) | T = 1)$$

“Qual é o efeito esperado do tratamento em uma unidade selecionada aleatoriamente da subpopulação de tratados?”

- Esse parâmetro incide diretamente sobre os participantes reais do tratamento, expressando o ganho bruto realizado do programa e pode ser comparado com seus custos, ajudando a decidir se o programa é bem-sucedido ou não .
- Da mesma forma, podemos definir (raramente usado) o efeito causal médio do tratamento na subpopulação de não tratados (ATU):

$$ATU = E [Y(1) - Y(0) | T = 0]$$

26

26

Parâmetros Causais

- Às vezes é útil calcular ATE ou ATT para subpopulações de unidades porque o tratamento pode ser mais eficaz para indivíduos com certas características ou pode ser eficaz para alguns grupos e não para outros.

A campanha de informação sobre o HIV é mais eficaz se a informação for dada a homens ou mulheres?

- Seja X = gênero; poderíamos calcular o ATE e ATT para homens e mulheres separadamente, então condicional a X

$$ATE(x) = E(Y(1) - Y(0) | X = x) = E(Y(1) | X = x) - E(Y(0) | X = x)$$

$$ATT(x) = E(Y1 - Y0 | T = 1, X = x) = E(Y1 | T = 1, X = x) - E(Y0 | T = 1, X = x)$$

- Se o tratamento for igualmente eficaz (ou ineficaz) para todos, os parâmetros condicionais serão todos iguais ao parâmetro geral: $ATE(x) = ATE$ e $ATT(x) = ATT$

27

27

ATT – Possibilidade de Viés



- Lembrete $ATT = E(Y(1) | T = 1) - E(Y(0) | T = 1)$
- O primeiro componente, $E(Y(1) | T = 1)$ pode ser estimado usando dados sobre resultados para unidades tratadas, ou seja, $E(Y1 | T = 1) = E(Y^{OBS} | T = 1)$
- O segundo componente, $E(Y(0) | T = 1)$ não tem como ser estimado diretamente (o resultado sob controle ($Y(0)$) nunca é observado para tratado!) e precisamos de um grupo de controle adequado para isso
- Em geral não podemos simplesmente tomar os resultados de todas as unidades de controle como substitutos. Ou seja, $E(Y(0) | T = 1) \neq E(Y^{OBS} | T = 0) = E(Y(0) | T = 0)$
- A diferença é : $VIÉS = E(Y(0) | T = 1) - E(Y^{OBS} | T = 0)$
- Devido: diferenças pré-tratamento entre os grupos de tratados e controles

Formas de minimizar esse viés e consequentemente estimar a quantidade faltante dependerá do tipo de abordagem metodológica utilizada para estimar o efeito de uma avaliação de impacto

28

28

Fonte do Viés

- Segundo Heckman et al (1997) o viés pode ser decomposto em três componentes :
- Presença de unidades em um grupo que não podem encontrar comparação adequada no outro grupo – **Ausência de sobreposição** (suporte comum) de X (características) no grupo tratado e controle
- Presença de desequilíbrio nos confundidores observados entre os grupos de unidades tratadas e de controle - **Seleção em observáveis**
- Desequilíbrio em fatores de confusão não observados entre os grupos de tratados e unidades de controle – **Omissão de confundidores**

29

29

Type of outcome	Scale	Potential outcome notation	Effect measure	Example interpretation
Continuous	Additive	$E(Y^{a=1} - Y^{a=0})$	Causal mean difference	'An average increase in systolic blood pressure by 10 mm Hg'
	Multiplicative	$\frac{E(Y^{a=1})}{E(Y^{a=0})}$	Causal mean ratio	'An average increase in systolic blood pressure by a factor of 1.1' or 'by 10%'
Binary	Additive	$p(Y^{a=1} = 1) - p(Y^{a=0} = 1)$	Causal risk difference	'An average increase in the risk of stroke by 5 percentage points'
	Multiplicative	$\frac{p(Y^{a=1} = 1)}{p(Y^{a=0} = 1)}$	Causal risk ratio	'An average increase in the risk of stroke by a factor of 1.5'

Igelström E, Craig P, Lewsey J, et al. J (2022). Epidemiol Community Health2022;76: 960–966.

A escala da medida de efeito pode ser aditiva e multiplicativa

30

30

Igelström E, Craig P, Lewsey J, et al. J (2022). Epidemiol Community Health 2022;76:960–966.

Vários efeitos causais do tratamento podem ser distinguidos, dependendo de como a exposição é definida e qual população alvo é considerada

Necessário especificar qual o efeito de tratamento a estimar pois pode diferir em termos de tamanho, tipo de viés e interpretação

Table 2 Definitions of different types of treatment effect

Effect	Potential outcome notation	Description
Average treatment effect (ATE)	$E(\gamma^{a=1} - \gamma^{a=0})$	The difference between the average outcome when everyone is exposed, and the average outcome when nobody is.
Average treatment effect in the treated (ATT)	$E(\gamma^{a=1} - \gamma^{a=0} A = 1)$	The ATE in the subpopulation of individuals who were actually exposed.
Average treatment effect in the untreated (ATU/ATUT)	$E(\gamma^{a=1} - \gamma^{a=0} A = 0)$	The ATE in the subpopulation of individuals who were actually unexposed.
Intention-to-treat effect (ITT)	$E(\gamma^{z=1} - \gamma^{z=0})$	Average effect of being assigned to (but not necessarily receiving) the exposure.
Complier average causal effect (CACE) or local average treatment effect (LATE)	$E(\gamma^{a=1} - \gamma^{a=0} A^{z=0} = 0, A^{z=1} = 1)$	The ATE among the 'compliers', that is, the subpopulation whose exposure status was affected by the assignment mechanism.
A denotes actual exposure status (a=1 for exposed, a=0 for unexposed). Z denotes assignment to the exposure, which may or may not have been adhered to (z=1 for assignment to the exposure, z=0 for assignment away from the exposure).		

31

Métodos de Avaliação de Impacto



- Procedimentos baseados: estudos experimentais e não experimentais (ou experimentos naturais-Natural Experiments)
- Importante :
 - ✓ Estabelecer grupo de controle
 - ✓ Necessidade de compreensão do processo de atribuição de características do grupo controle
 - ✓ Considerar tipo de efeito causal a ser estimado e se o mesmo se aplica a toda a população
- Métodos dependem da adequabilidade aos diferentes pressupostos
 - ✓ Entendimento do processo que determina a exposição à intervenção é fundamental para concepção e condução dos estudos NE

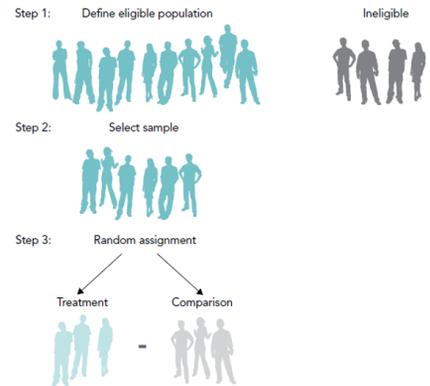
32

32

Métodos de Avaliação de Impacto

- **Métodos de seleção aleatória** (RCT, em inglês)

- Padrão-ouro na avaliação de impacto
- Evita vies de seleção
- A avaliação do impacto precisaria ser feita no mesmo nível da aleatorização
- Exemplo: Progres, Programa de Saneamento no setor rural da Bolívia , etc



33

33

Métodos de Avaliação de Impacto



- **Métodos de seleção aleatória** - Os grupos tratamento e controle são considerados similares tanto para características observadas quanto não observadas e por isso oferecem uma boa estimativa do contrafactual

- $Y(1), Y(0) \perp\!\!\!\perp T \Rightarrow E[Y(0) | T=1] = E[Y(0) | T=0]$

- **O vies** = $E[Y(0) | T=1] - E[Y^{OBS} | T=0] = 0$

- A estimativa do ATT pode ser estimado

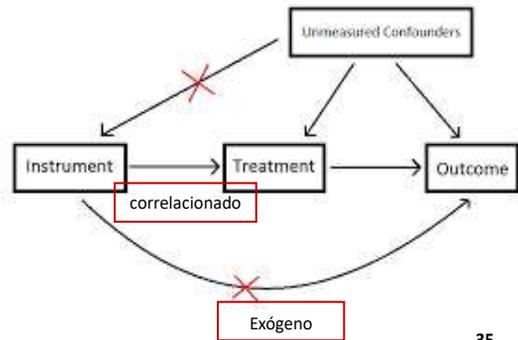
$$ATT = \frac{\sum_{i:T_i=1} Y_i^{obs}}{n_T} - \frac{\sum_{i:T_i=0} Y_i^{obs}}{n_C}$$

34

34

Métodos de Avaliação de Impacto

- Programas que apresentam cumprimento parcial, inscrições voluntárias ou cobertura universal
- **Variável instrumental** : baseia-se em alguma fonte externa de variação para determinar o status do tratamento
- Suposições:
 - Instrumento válido -> correlacionado com o tratamento
 - Exógeno -> não afeta o desfecho exceto pela influência da probabilidade de receber o tratamento



35

35

Métodos de Avaliação de Impacto

Diferença em diferenças : compara as *mudanças* nos resultados ao longo do tempo entre os inscritos em um programa (o grupo de tratamento) e os não inscritos (o grupo de comparação).

Permite corrigir quaisquer diferenças entre os grupos de tratamento e de comparação que sejam constantes ao longo do tempo.

Suposições:

- ✓ Observações antes e depois da intervenção(panel data)
- ✓ Há o controle entre as diferenças observadas e não observadas nas características fixas dos grupos e é, portanto, menos propensa a viés de variáveis omitidas causadas por fatores de confusão não medidos ou erro de medição

Cálculo do método diferença em diferenças

	Depois	Antes	Diferença
Tratamento/inscritos	B	A	$B - A$
Comparação/não inscritos	D	C	$D - C$
Diferença	$B - D$	$A - C$	$DD = (B - A) - (D - C)$

36

36

Griffin et al.
BMC Medical Research Methodology (2021) 21:279
https://doi.org/10.1186/s12874-021-01471-y

BMC Medical Research
Methodology

RESEARCH Open Access

Moving beyond the classic difference-in-differences model: a simulation study comparing statistical methods for estimating effectiveness of state-level policies

Beth Ann Griffin^{1*}, Megan S. Schuler¹, Elizabeth A. Stuart², Stephen Patrick³, Elizabeth McNeer¹, Rosanna Smart⁴, David Powell¹, Bradley D. Stein⁵, Terry L. Schell⁶ and Rosalie Liccardo Pacula⁶

JAMA Internal Medicine | Original Investigation | HEALTH CARE POLICY AND LAW

Association Between State Laws Facilitating Pharmacy Distribution of Naloxone and Risk of Fatal Overdose

Rahi Abouk, PhD; Rosalie Liccardo Pacula, PhD; David Powell, PhD

IMPORTANCE Given high rates of opioid-related fatal overdoses, improving naloxone access has become a priority. States have implemented different types of naloxone access laws (NALs) and there is controversy over which of these policies, if any, can curb overdose deaths. We hypothesize that NALs granting direct authority to pharmacists to provide naloxone will have the greatest potential for reducing fatal overdoses.

OBJECTIVES To identify which types of NALs, if any, are associated with reductions in fatal overdoses involving opioids and examine possible implications for nonfatal overdoses.

DESIGN, SETTING, AND PARTICIPANTS State-level changes in both fatal and nonfatal overdoses from 2005 to 2016 were examined across the 50 states and the District of Columbia after adoption of NALs using a difference-in-differences approach while estimating the magnitude of the association for each year relative to time of adoption. Policy environments across full state populations were represented in the primary data set. The association for 3 types of NALs was associated: NALs providing direct authority to pharmacists to prescribe, NALs providing indirect authority to prescribe, and other NALs. The study was conducted from January 2017 to January 2019.

EXPOSURES Fatal and nonfatal overdoses in states that adopted NAL laws were compared with those in states that did not adopt NAL laws. Further consideration was given to the type of NAL passed in terms of its association with these outcomes. We hypothesize that NALs granting direct authority to pharmacists to provide naloxone will have the greatest potential for reducing fatal overdoses.

Supplemental content
CME Quiz at jamanetwork.com/learning and CME Questions page 856

37

37

Métodos de Avaliação de Impacto

Uso de índices para implementação de programas, ou seja, quando atribuição ao tratamento depende de variáveis observadas mensuradas

Desenho de regressão descontínua (RDD) - Explora os critérios de elegibilidade do programa para comparar participantes e controles.

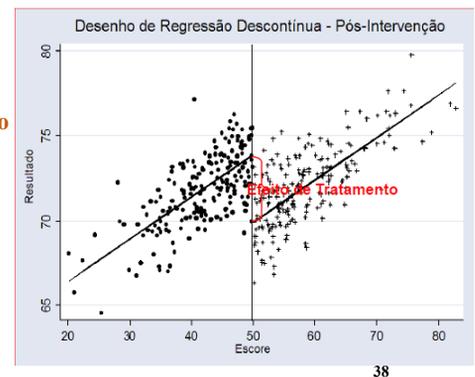
Índices são comuns na focalização de programas sociais:

Programas de combate à pobreza => domicílios abaixo de um determinado índice de pobreza

Programa de pensão => população acima de certa idade

Avaliação : comparar domicílios logo acima e logo abaixo do ponto de corte

(-) Efeito médio local (as vezes não generalizável)



38

Regressão Descontínua

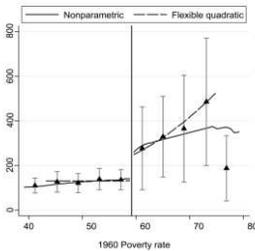
Exemplos na Avaliação de Políticas

DOES HEAD START IMPROVE CHILDREN'S LIFE CHANCES? EVIDENCE FROM A REGRESSION DISCONTINUITY DESIGN*

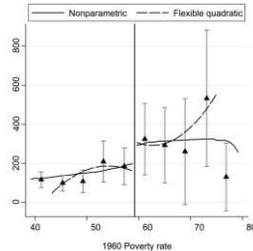
JENS LUDWIG AND DOUGLAS L. MILLER

This paper exploits a new source of variation in Head Start funding to identify the program's effects on health and schooling. In 1965 the Office of Economic Opportunity (OEO) provided technical assistance to the 300 poorest counties to develop Head Start proposals. The result was a large and lasting discontinuity in Head Start funding rates at the OEO cutoff for grant-writing assistance. We find evidence of a large drop at the OEO cutoff in mortality rates for children from causes that could be affected by Head Start, as well as suggestive evidence for a positive effect on educational attainment.

Panel A: 1968 Head Start funding per 4 year old



Panel B: 1972 Head Start funding per 4 year old



Evidência de descontinuidade do tratamento no ponto de corte

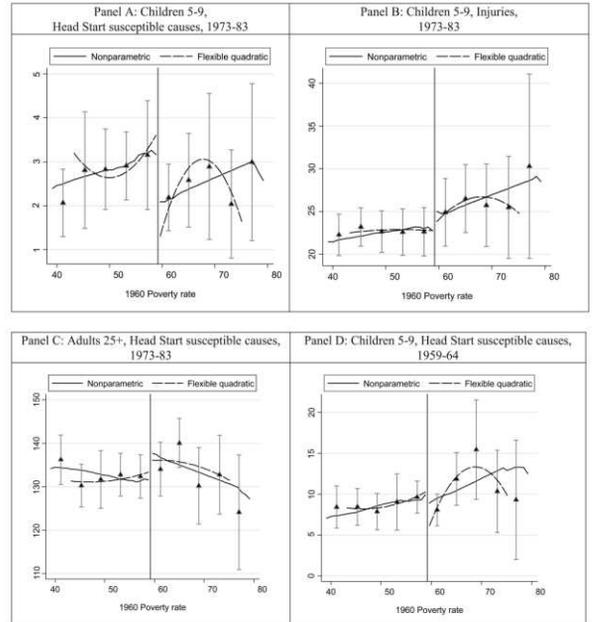
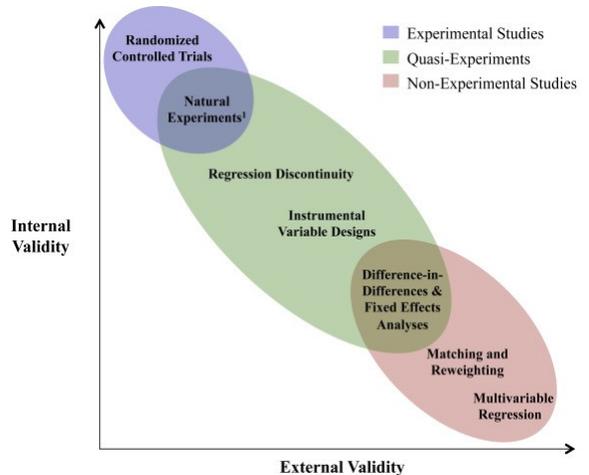


FIGURE IV
Estimated Discontinuities at OEO Cutoff in Mortality Rates per 100,000 for Children and Adults, from Causes Affected by Head Start and from Injuries

39

- Diferentes métodos, diferentes perguntas de avaliação e conclusões
- Necessário escolhas entre métodos de forma coerente com as questões que se quer responder
- Esforço no uso métodos quasi-experimentais de avaliação de impacto em saúde
- Combinações: diff-diff e pareamento

Modelo teóricos causais robustos!!



Geldsetzer P, Fawzi W. Quasi-experimental study designs series—paper 2: complementary approaches to advancing global health knowledge. *Journal of clinical epidemiology*. 2017 Sep 1;89:12-6.

40

40

Pareamento



- Utilizados desde a primeira metade do século 20 (Greenwood, 1945; Chapin, 1947).
- Base teórica desenvolvida a partir de 1970.
- De forma geral, a idéia é igualar ou equilibrar a distribuição de covariáveis nos grupos tratado e controle.
- Objetivo: encontrar a melhor forma de escolher os grupos tratados e controle para comparação

41

41

Pareamento

- Estágios-chave em qualquer estudo para estimar efeito de uma intervenção:
 - **Estágio 1: Planejamento/Desenho**
 - Uso de informações do *baseline* para planejar um estudo não experimental como se fosse randomizado, sem acesso às informações do desfecho.
 - Etapa onde as ferramentas de pareamento são utilizadas.
 - **Estágio 2: Análise do desfecho**
 - Comparação do desfecho entre grupos de intervenção → redução viés.
 - Só é realizada após conclusão da etapa anterior.
- Apesar do pareamento ser geralmente usado para estimar efeitos causais, algumas vezes é usado para responder questões não causais.

42

42

Pareamento

- Indicado em situações onde atribuição ao tratamento depende das características observadas
- Pressuposto: Independência condicional**
 - Respostas potenciais Y são independentes da alocação ao tratamento T, ou seja:

$$(Y_i^T, Y_i^C) \perp T_i | X_i$$

- Nos subgrupos específicos da amostra da avaliação, a seleção ao tratamento ocorre de forma “aleatória”, ou seja, a seleção ao tratamento ocorre de forma tão boa quanto “aleatória” dentro de cada subgrupo formado pelas características observadas
- Ex: Idade jovem jovem tem a mesma probabilidade de receber o tratamento, embora possa diferir de um grupo para outro
 estado civil solteiro casado
 gênero feminino feminino

Importante: Fatores não observados não afetam a participação.

Pareamento



Ideal

Unidade de análise	Desfecho (tratamento)	Desfecho (comparação)	Efeito do tratamento
1	Y_1^T	Y_1^C	$Y_1^T - Y_1^C$
2	Y_2^T	Y_2^C	$Y_2^T - Y_2^C$
⋮	⋮	⋮	⋮
i	Y_i^T	Y_i^C	$Y_i^T - Y_i^C$
⋮	⋮	⋮	⋮
n	Y_n^T	Y_n^C	$Y_n^T - Y_n^C$

Não temos como ter essa diferença

$$\hat{E}(Y^T - Y^C) = \sum(Y_i^T - Y_i^C)/n$$

Intervenção: Participação num programa de treinamento

Resposta de interesse : renda

Cada unidade tem duas resposta potencial Y^T e Y^C

Pareamento

Unidade de análise	Característica	Desfecho (tratamento)	Desfecho (comparação)
1	X_1	Y_1^T	?
2	X_2	?	Y_2^C
⋮	⋮	⋮	⋮
i	⋮	?	Y_i^C
⋮	⋮	⋮	⋮
n	X_n	Y_n^T	?

Problema fundamental da inferência causal

Vamos supor que unidades 1, i e n tem a mesma característica X

As unidades 1 e n receberam tratamento

Unidade de análise	Característica	Desfecho (tratamento)	Desfecho (comparação)
1	X_1	Y_1^T	?
2	X_2	?	Y_2^C
⋮	⋮	⋮	⋮
i	X_i	?	Y_i^C
⋮	⋮	⋮	⋮
n	X_n	Y_n^T	?

45

45

Pareamento



Unidade de análise	Característica	Desfecho (tratamento)	Desfecho (comparação)	Efeito do tratamento
1	X_1	Y_1^T	?	
2	X_2	?	Y_2^C	
⋮	⋮	⋮	⋮	⋮
i	X_i	$\left[\frac{Y_1^T + Y_n^T}{2} \right]$	Y_i^C	$\left[\frac{Y_1^T + Y_n^T}{2} \right] - Y_i^C$
⋮	⋮	⋮	⋮	⋮
n	X_n	Y_n^T	?	

← O contrafacto para a unidade i
O efeito causal para unidade i

Lembrete: A estimativa do efeito é válida desde que fatores não observados não afetam a participação no programa de intervenção.

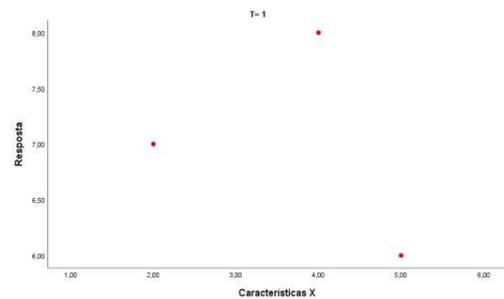
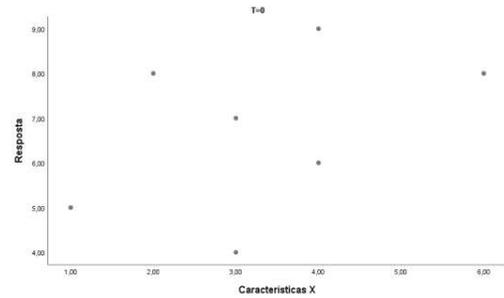
$$(Y_i^T, Y_i^C) \perp T_i | X_i$$

46

46

Pareamento

i	X_i	T_i	Y_i
1	2	1	7
2	4	1	8
3	5	1	6
4	3	0	4
5	2	0	8
6	3	0	7
7	1	0	5
8	4	0	9
9	4	0	6
10	6	0	8



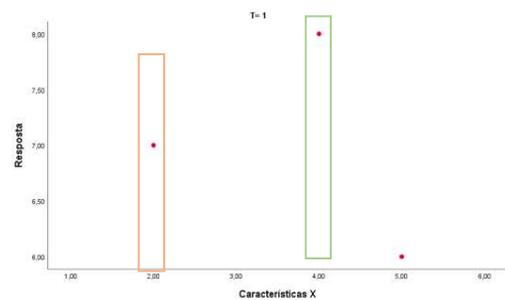
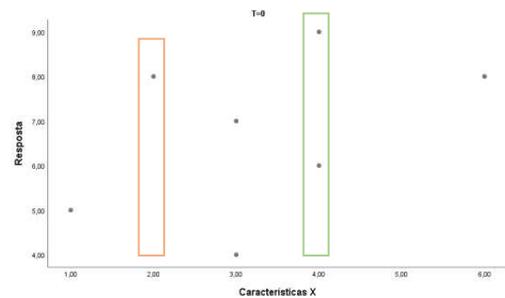
47

Pareamento

i	X_i	T_i	Y_i	\hat{y}_i^c
1	2	1	7	8
2	4	1	8	7.5
3	5	1	6	
4	3	0	4	
5	2	0	8	
6	3	0	7	
7	1	0	5	
8	4	0	9	
9	4	0	6	
10	6	0	8	

Pareamento exato

Um contrafacto aproximado – média das observações



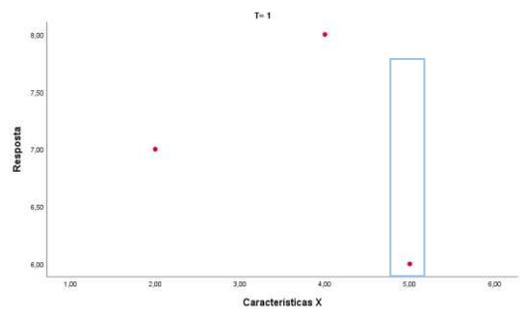
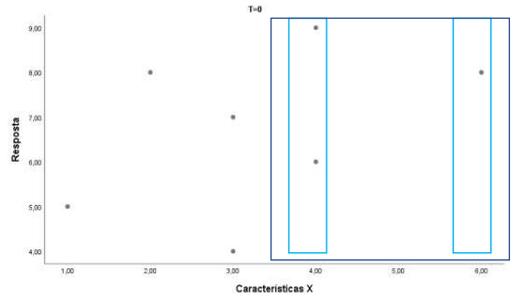
48

Pareamento

i	X _i	T _i	Y _i	\widehat{Y}_i^c
1	2	1	7	8
2	4	1	8	7.5
3	5	1	6	7.7
4	3	0	4	
5	2	0	8	
6	3	0	7	
7	1	0	5	
8	4	0	9	
9	4	0	6	
10	6	0	8	

Não Exato Área de suporte comum

Pareamento pelo vizinho mais próximo



Simple and Bias-Corrected Matching Estimators for Average Treatment Effects
 Alberto Abadie and Guido Imbens
 NBER Technical Working Paper No. 283
 October 2002
 JEL No. C100, C130, C140, J240, J310

ABSTRACT

Matching estimators for average treatment effects are widely used in evaluation research despite the fact that their large sample properties have not been established in many cases. In this article, we develop a new framework to analyze the properties of matching estimators and establish a number of new results. First, we show that matching estimators include a conditional bias term which may not vanish at a rate faster than root-N when more than one continuous variable is used for matching. As a result, matching estimators may not be root-N-consistent. Second, we show that even after removing the conditional bias, matching estimators with a fixed number of matches do not reach the semiparametric efficiency bound for average treatment effects, although the efficiency loss may be small. Third, we propose a bias-correction that removes the conditional bias asymptotically, making matching estimators root-N-consistent. Fourth, we provide a new estimator for the conditional variance that does not require consistent nonparametric estimation of unknown functions. We apply the bias-corrected matching estimators to the study of the effects of a labor market program previously analyzed by Lalonde (1986). We also carry out a small simulation study based on Lalonde's example where a simple implementation of the biascorrected matching estimator performs well compared to both simple matching estimators and to regression estimators in terms of bias and root-mean-squared-error. Software for implementing the proposed estimators in STATA and Matlab is available from the authors on the web.

Alberto Abadie
 John F. Kennedy School of Government
 Harvard University
 79 John F. Kennedy Street
 Cambridge, MA 02138
 and NBER
 alberto_abadie@harvard.edu

Guido Imbens
 Department of Economics, and Department
 of Agricultural and Resource Economics
 University of California at Berkeley
 661 Evans Hall #3880
 Berkeley, CA 94720-3880
 and NBER
 imbens@econ.berkeley.edu



Pareamento com reposição

Pareamento

i	X _i	T _i	Y _i	\widehat{Y}_i^C	\widehat{Y}_i^T
1	2	1	7	8	7
2	4	1	8	7.5	8
3	5	1	6	7.7	6
4	3	0	4	4	
5	2	0	8	8	
6	3	0	7	7	
7	1	0	5	5	
8	4	0	9	9	
9	4	0	6	6	
10	6	0	8	8	

observado

i	X _i	T _i	Y _i	\widehat{Y}_i^C	\widehat{Y}_i^T
1	2	1	7	8	7
2	4	1	8	7.5	8
3	5	1	6	7.7	6
4	3	0	4	4	7.5
5	2	0	8	8	7
6	3	0	7	7	7.5
7	1	0	5	5	7
8	4	0	9	9	8
9	4	0	6	6	8
10	6	0	8	8	6

Contrafactual

51

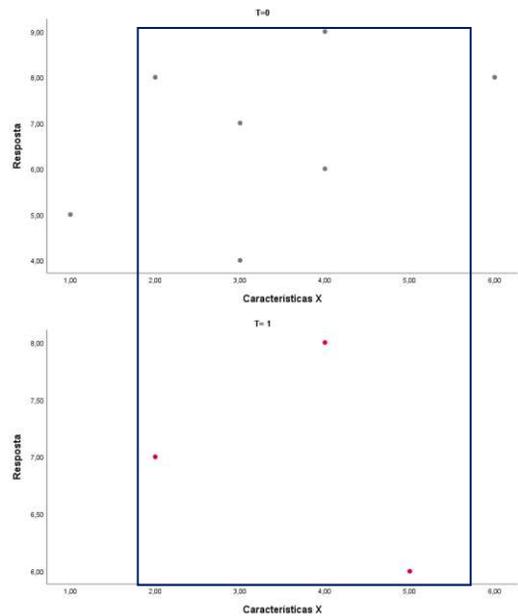
51

Pareamento

i	X _i	T _i	Y _i	\widehat{Y}_i^C	\widehat{Y}_i^T
1	2	1	7	8	7
2	4	1	8	7.5	8
3	5	1	6	7.7	6
4	3	0	4	4	7.5
5	2	0	8	8	7
6	3	0	7	7	7.5
7	1	0	5	5	7
8	4	0	9	9	8
9	4	0	6	6	8
10	6	0	8	8	6

As unidades 7 e 10 estão fora da área de suporte comum em termos de X,

só tem observações no grupo de comparação, talvez o valor do contrafacto seja de qualidade inferior quando comparado as outras observações que estão na área de suporte comum



52

52

Pareamento

i	X _i	T _i	Y _i	\widehat{Y}_i^c	\widehat{Y}_i^t	Eff _i
1	2	1	7	8	7	-1
2	4	1	8	7.5	8	0.5
3	5	1	6	7.7	6	-1.7
4	3	0	4	4	7.5	3.5
5	2	0	8	8	7	-1
6	3	0	7	7	7.5	0.5
7	1	0	5	5	7	2
8	4	0	9	9	8	-1
9	4	0	6	6	8	2
10	6	0	8	8	6	-2

$$ATE_{NN} = 0.183$$

ATE representa o efeito médio em toda amostra

$$ATT_{NN} = -0.733$$

ATT representa o efeito médio da intervenção entre aqueles que de fato receberam a intervenção.

• Foco no efeito sobre aqueles que o programa almeja

Pesquisadores devem decidir se ATE ou ATT é a medida de maior utilidade e interesse em seus estudos observacionais

53

53

Pareamento



• “Maldição da dimensionalidade”:

- ✓ Uso de pequeno número de variáveis facilita o pareamento perfeito, mas corre-se o risco de não considerar variáveis importantes;
- ✓ Uso de grande número de variáveis dificulta encontrar o par perfeito.
- ✓ Uma forma de resolver é através do método do **pareamento com escores de propensão** (*propensity score matching - PSM*) (Rosenbaum e Rubin, 1983)
 - A idéia é escolher um grupo de comparação com propensão similar ao grupo tratado
 - Como a alocação não é verdadeiramente aleatória, PSM pertence à categoria de métodos **quasi-experimentais**.

54

54

Análise via escores de propensão

- A análise via escores de propensão utiliza as covariáveis desbalanceadas entre os grupos de tratamento e controle em um modelo logístico para prever a probabilidade de receber tratamento

$$P(T=1|X) = \pi(x)$$

- Probabilidade condicional de tratamento em função de variáveis preditoras observadas (características pré-tratamento)
- Segundo Rosenbaum e Rubin (1983) se o conjunto de informações justifica o pareamento em X (ou seja, todos os fatores de confusão relevantes estão incluídos em X), então também justifica o pareamento no escore de propensão.

55

55

Escore de Propensão- Estratégias de Análise

Objetivo: remover efeitos de confundimento ao estimar efeito da intervenção

- Pareamento por escore de propensão (PSM)
- Estratificação (ou subclassificação) baseado em PS
- Full matching
- Ponderação usando PS:
 - IPTW (*inverse probability of treatment weighting*)
 - Kernel matching
 - Local linear weighting
 - Overlapping weighting (área de suporte comum)
- Ajuste de modelo de regressão utilizado o escore como covariável

56

56

Pareamento pelo Escore de Propensão

- **A ideia é encontrar um par para cada indivíduo do grupo tratado:**

- Com todos os indivíduos (trt e controle), calcula-se a **probabilidade de participar** da intervenção baseada em características observadas : **escore de propensão**

$$P(T = 1 | X)$$

- **Escore varia entre 0 e 1**: sumariza as características que influenciam a participação no programa (medida de equilíbrio);
- Após cálculo do escore, tratados e controles com escores de propensão mais próximos são pareados: tentativa de **proxy para contrafactual**.
- Sob certas condições, mostrou-se que **PSM faz um bom pareamento** (Rosenbaum e Rubin, 1983)

57

57

Pareamento pelo Escore de Propensão

É útil quando se acredita que apenas características observadas podem afetar a participação na intervenção:

- Tenta capturar os efeitos de diferentes covariáveis X na participação usando um escore ou índice de propensão.
- Indivíduos que não forem pareados são excluídos da análise.
- PSM será viesado se covariáveis que determinam participação não forem incluídas, mas

inclusão de número muito grande de X deve ser evitada porque pode aumentar erros-padrão e resultar em predição perfeita $\{P(T=1|X)=1\}$

58

58

Pareamento pelo Escore de Propensão

Escore de Propensão (PS) (Rosenbaum e Rubim, 1983) $e_i = P(T_i = 1 | X_i)$

- Condicional ao PS, a distribuição das covariáveis mensuradas no *baseline* é similar entre tratados e não tratados.
- Se a alocação do tratamento é “ignorável” condicional às covariáveis:
 - ✓ Então a alocação do tratamento também é “ignorável” condicional ao PS.
 - ✓ Este fato justifica o pareamento baseado no PS no lugar de usar o conjunto completo de covariáveis.

Suposições atendidas => estimação não viesada do efeito pela comparação de médias, por exemplo, usando grupos com mesmo PS.

59

59

Pareamento pelo Escore de Propensão

PS (propensity score) geralmente estimado via regressão logística:

- Obtém-se os valores preditos para T: probabilidade estimada de participação → escore de propensão
- Outros métodos também podem ser usados na estimação de PS: métodos de particionamento recursivos ou métodos *tree-based*, redes neurais, CART (*Classification and Regression Tree*) etc...
- Para variável tratamento politômica: probito multinomial (Imbens, 2000; Lechner, 2001):

É computacionalmente intensivo mas baseado em suposições mais fracas do que a logística multinomial.

Lembrete: Não é de interesse as estimativas do modelo, colinearidade, bondade de ajuste: maior ênfase no **balanceamento** das covariáveis.

60

60

Pareamento pelo Escore de Propensão

Validade depende das seguintes condições:

- **Independência condicional** $(Y(1), Y(0)) \perp T \mid X$ (Não pode ser testada diretamente)

Respostas potenciais Y são independentes da alocação ao tratamento T ,
Fatores não observados não afetam a participação
(forte suposição - não existem confundidores não mensurados)

$$(Y(1), Y(0)) \perp T \mid e(X)$$

Rosenbaum & Rubin (1983) que se a seleção do tratamento é ignorável dado X , então é válido para $e(x)$

- **Suporte comum** (ou sobreposição) $0 < P(T = 1 \mid X) < 1$

Todos os indivíduos tem uma probabilidade positiva de serem alocados no grupo intervenção

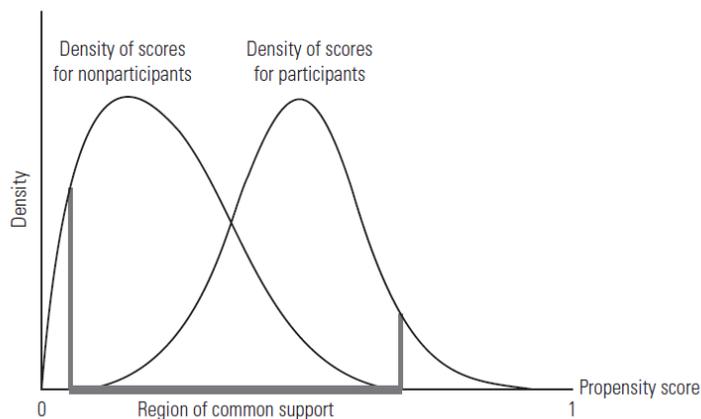
Assegura que observações no grupo tratado têm observações no grupo controle “próximas” em relação à distribuição de PS.

Efetividade do PSM depende também em ter um número grande e aproximadamente igual de participantes e não participantes para que esta região possa ser encontrada

61

61

Figure 4.1 Example of Common Support

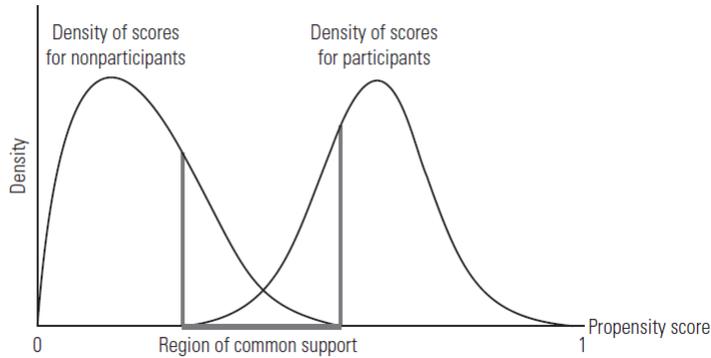


Source: Authors' representation.

62

62

Figure 4.2 Example of Poor Balancing and Weak Common Support



Source: Authors' representation.

A exclusão de indivíduos fora da região de suporte comum pode levar a viés amostral:

- Exame das características das observações excluídas

63

63

Estimando ATT usando PSM

- Assumindo que independência condicional e suporte comum são válidos:
- Estimação pode ser feita usando diferença média em Y no suporte comum, ponderando as unidades de comparação através do PS:

$$\text{ATT}_{\text{PSM}} = E_{P(X) | T=1} \{E[Y^T | T=1, P(X)] - E[Y^C | T=0, P(X)]\}$$

- Este estimador pode ainda ser reescrito como:

Validade interna

$$\Rightarrow \text{ATT}_{\text{PSM}} = \frac{1}{N_T} \left[\sum_{i \in T} Y_i^T - \sum_{j \in C} \omega(i, j) Y_j^C \right]$$

onde $N_T = \#$ participantes (grupo trt), $w(i, j) =$ peso usado para agregar participantes i com não participantes j (definido de acordo com o método de pareamento).

64

64

Estimando ATT ou ATE ?

- É preciso definir claramente o efeito causal de interesse (comparação) fornecendo teoria ou evidência preliminar sobre sua plausibilidade e viabilidade
- Importante ter bom entendimento do mecanismo de atribuição de tratamento: quais fatores entraram na escolha ou decisão de qual tratamento os indivíduos receberam e, em seguida, idealmente, ter todos os fatores importantes observados
- Pergunta: É viável todos os elegíveis receberem o tratamento/intervenção de interesse?
 - ✓ Afirmativo => ATE (população alvo para inferência)
 - ✓ Negativo -> ATT (somente uma população específica de elegíveis podem receber a intervenção/tratamento)

65

65

Lembrete : Uso do Escore de Propensão

- Pareamento por escore de propensão (PSM)
- Estratificação (ou subclassificação) baseado em PS
- Full matching
- Ponderação usando PS:
 - IPTW (*inverse probability of treatment weighting*)
 - Kernel matching (mais peso as unidades mais próximas que serão pareadas)
 - Local linear weighting
 - Overlapping weighting
- Ajuste de modelo de regressão utilizado o escore como covariável

66

66

Comparação de diferentes métodos usando PS

Método	Vantagens	Limitações
Estratificação	<ul style="list-style-type: none"> Retém dados de todos os participantes Oportunidade de explorar interações entre o tratamento e o EP no risco associado ao desfecho Fornecer estimativas de efeito para cada estrato 	<ul style="list-style-type: none"> Apresenta pior desempenho em conjuntos de dados com desfechos raros, especialmente quando o número de estratos é grande Pode não ser eficiente em cenários de confundimento muito elevado
Pareamento	<ul style="list-style-type: none"> Fornecer equilíbrio nas covariáveis na maioria das circunstâncias Simples de analisar, apresentar e interpretar 	<ul style="list-style-type: none"> Observações "incomparáveis" (extremas), tendem a ser excluídas da análise – reduz a precisão das estimativas
Ponderação pelo inverso da probabilidade de tratamento (IPTW)	<ul style="list-style-type: none"> Retém dados de todos os participantes do estudo Fácil de implementar Cria uma pseudo-população perfeitamente balanceada 	<ul style="list-style-type: none"> Pode ser instável quando há pesos extremos
Uso do EP como covariada de ajuste	<ul style="list-style-type: none"> Fácil implementação Performance adequada em cenários de confundimento baixo ou moderado 	<ul style="list-style-type: none"> Produz resultados muito semelhantes (e não necessariamente superiores) ao tradicional ajuste de covariável

Adaptado de: Elze MC, et al. Comparison of propensity score methods and covariate adjustment: evaluation in 4 cardiovascular studies. Journal of the American College of Cardiology. 2017 Jan 24;69(3):345-57.

67

67

Comparação de diferentes métodos usando PS

- Muitos estudos têm demonstrado que:
 - PSM elimina uma maior proporção das diferenças sistemáticas devido a características do *baseline* do que os métodos de estratificação com PS ou ajuste de modelo com PS;
 - ✓ Em alguns cenários, PSM e IPTW removem diferenças sistemáticas de modo similar;
 - ✓ Mas, em outros cenários, PSM foi superior a IPTW.
 - Estimação de efeito médio da intervenção usando estratificação com PS resulta em maior viés do que através do uso de estimadores ponderados.
 - Ajuste com PS como covariável e IPTW são métodos mais sensíveis à precisão da estimação do PS.
- Métodos apropriados de diagnóstico existem para cada um dos métodos usando PS:
 - PSM, Estratificação com PS e IPTW:
 - ✓ adequação da especificação do modelo para PS → estimação direta do efeito do tratamento usando amostra pareada, estratificada ou ponderada.
 - Ajuste de covariável no modelo de regressão:
 - requer especificação da adequação do modelo para o desfecho.

68

68

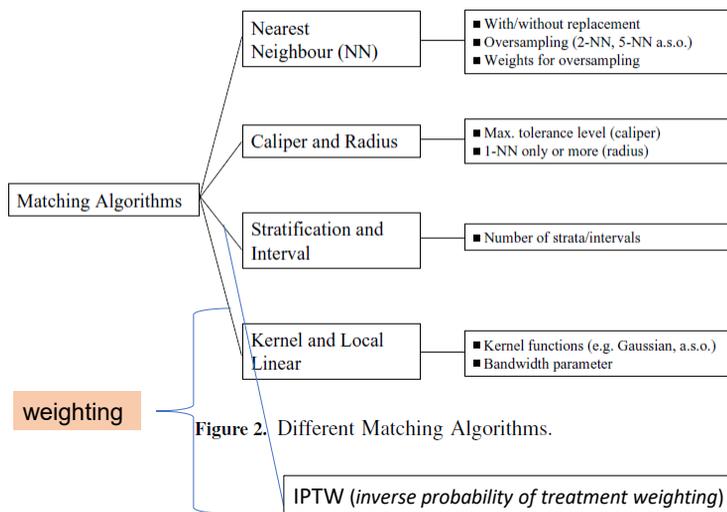
Métodos para Pareamento

- **Diferentes metodologias no pareamento usando Escore de Propensão:**
 - **Pareamento por vizinho mais próximo (*nearest-neighbor (NN)*)**
 - Escolha realizada entre indivíduos com PS mais próximos (mínima diferença pode ser algo grande – pareamento ruim)
 - **Pareamento por Raio (*radius*)**
 - Parear baseando-se em um “*threshold*” (amplitude para o valor da similaridade) ou tolerância na distância máxima do PS
 - **Pareamento por intervalo ou por estratificação**
 - Em cada intervalo, o efeito do programa é definido pela diferença das médias das respostas entre observações nos grupos tratamento e controle e uma média ponderada destas estimativas produz uma medida global de impacto
 - **Pareamento por kernel**
 - usam uma média ponderada de **TODOS** os não participantes para construir o contrafactual pareado para cada participante
 - **Pareamento local linear (*local linear matching (LLM)*)**
 - Estima uma regressão não paramétrica com ponderação local (*lowess*) da resposta do grupo controle na vizinhança de cada observação do grupo tratado

69

69

Escores de Propensão – Estratégias de Análise



Terminologia pode ser confusa e depende da área!

70

70

Table 1. Trade-offs in Terms of Bias and Efficiency.

Decision	Bias	Variance
Nearest neighbour matching: multiple neighbours/single neighbour with caliper/without caliper	(+)/(–) (–)/(+)	(–)/(+) (+)/(–)
Use of control individuals: with replacement/without replacement	(–)/(+)	(+)/(–)
Choosing method: NN matching/Radius matching KM or LLM/NN methods	(–)/(+) (+)/(–)	(+)/(–) (–)/(+)
Bandwidth choice with KM: small/large	(–)/(+)	(+)/(–)
Polynomial order with LPM: small/large	(+)/(–)	(–)/(+)

KM, kernel matching, LLM; local linear matching; LPM, local polynomial matching NN, nearest neighbour; increase; (+); decrease (–).

Assintoticamente todos os estimadores usando PSM devem produzir mesmos resultados. (Caliendo e Kopeinig, 2008)

Ponderação pelo Escore de Propensão

- A ideia da ponderação pelo inverso da probabilidade de receber o tratamento (IPTW) é usar o escore de propensão para formar um peso.
- A ponderação é uma estratégia que tem sido largamente utilizada em pesquisa por amostragem (Horvitz & Thompson, 1952 – técnica estatística para calcular estatísticas para uma população diferente daquela em que os dados foram coletados).
- IPTWs são usados para criar uma pseudo-população em que as covariáveis e a atribuição de tratamento são independentes de uns aos outros (uma propriedade que esperaríamos na randomização).
- Espera-se que nessa pseudo-população, a distribuição das covariáveis entre os grupos seja balanceada
- O termo pseudo-população reflete o fato de que as observações/grupos ponderados não são idênticos à população de interesse. Porém, este grupo ponderado poderia ter sido selecionado de uma população em que não existia confundimento.

Ponderação pelo Escore de Propensão

- IPTW: *inverse probability of treatment weighting*.

- PS usado como peso amostral para estimar ATE:

$$w_i = \frac{T_i}{e_i} + \frac{(1-T_i)}{1-e_i}$$

- Peso = inverso da probabilidade de receber o tratamento que o indivíduo de fato recebeu.
- Uso dos dados da amostra completa.
- IPTW foi proposto originalmente como um método direto de padronização em modelos (Rosenbaum, 1987)

73

73

Ponderação pelo Escore de Propensão

- Vários estimadores foram propostos (Lunceford e Davidian, 2004) baseados em IPTW (descrição de propriedades teóricas e performance dos estimadores).

- Um estimador para ATE é: $ATE = \frac{1}{n} \sum_{i=1}^n \frac{T_i Y_i}{e_i} - \frac{1}{n} \sum_{i=1}^n \frac{(1-T_i) Y_i}{1-e_i}$

- Modelos de regressão com ponderação usando IPTW também têm sido descritos (Joffe et al, 2004).

- Limitação do uso de pesos:

- Pesos podem ser imprecisos ou instáveis se a probabilidade de receber o tratamento for muito baixa.
- Para resolver isto, pesos estabilizadores têm sido propostos (Robins et al, 2000)

74

74

Ponderação pelo Escore de Propensão

- Peso para estimação de ATT, com ponderação nas *odds* (Morgan e Todd, 2008):

$$w_{i,ATT} = T_i + \frac{(1-T_i)e_i}{1-e_i}$$

- Neste caso, indivíduos tratados recebem peso = 1.
- Assim, ambos grupos são ponderados para representar o grupo tratado.

75

75



American Journal of Epidemiology
© The Author(s) 2020. Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Vol. 188, No. 12
DOI: 10.1093/aje/kwaa527
Advance Access publication:
July 8, 2020

Original Contribution

Conditional Cash Transfer Program and Leprosy Incidence: Analysis of 12.9 Million Families From the 100 Million Brazilian Cohort

Julia M. Pescarini*, Elizabeth Williamson, Maria Y. Ichihara, Rosemeire L. Fiaccone, Laura Forastiere, Anna Ramond, Joilda Silva Nery, Maria Lucia F. Penna, Agostino Strina, Sandra Reis, Liam Smeeth, Laura C. Rodrigues, Elizabeth B. Brickley, Gerson O. Penna, and Mauricio L. Barreto

* Correspondence to Dr. Julia M. Pescarini, Centro de Integração de Dados e Conhecimentos para Saúde (Cidacs), Fundação Oswaldo Cruz, R. Mundo, 121 – Trobogy, CEP 41301-110, Salvador, Brazil (e-mail: juliapescarini@gmail.com).

Initially submitted September 18, 2019; accepted for publication June 21, 2020.

Pareamento pelo escore de propensão utilizando abordagem do vizinho mais próximo com a distância máxima de 0.01

We performed 1:1 nearest-neighbor matching with a caliper of 0.05, allowing a same nonbeneficiary family to match with more than 1 beneficiary family (i.e., matching with replacement) (20). We compared the difference in the distribution of PS covariates between beneficiary and nonbeneficiary families using the standardized mean difference to assess balance of potential confounders before and after matching (standardized mean difference of >0.1 was taken to indicate potential confounding by that characteristic) (21).

76

76

STUDY PROTOCOL

Evaluating the effect of Bolsa Familia, Brazil's conditional cash transfer programme, on maternal and child health: A study protocol

PLOS ONE | <https://doi.org/10.1371/journal.pone.0268500> May 23, 2022

Illa Rocha Falcao^{1,2*}, Rita de Cássia Ribeiro-Silva^{1,2}, Flávia José Oliveira Alves^{2,3}, Naiá Ortelan², Natanuel J. Silva², Rosemeire L. Fiaccone^{2,4}, Marcia Furquim de Almeida⁵, Júlia M. Pescarini^{2,6}, Cinthia Soares Lisboa^{2,7}, Elzo Pereira Pinto Júnior², Enny S. Paixão^{2,6}, Andrea J. F. Ferreira^{2,3}, Camila Silveira Silva Teixeira^{2,3}, Aline dos Santos Rocha^{1,2}, Srinivasa Vittal Katikireddi⁸, M. Sanni Ali⁸, Ruth Dundas⁸, Alastair Leyland⁸, Laura C. Rodrigues^{2,6}, Maria Yury Ichihara^{2,3}, Mauricio L. Barreto^{2,3}

“ The models will be weighted by the Inverse Probability Treatment Weighting (IPTW) and by the Kernel weights. Balancing will be performed before and after weighting to ensure that the procedure used controlled for the available confounders ”

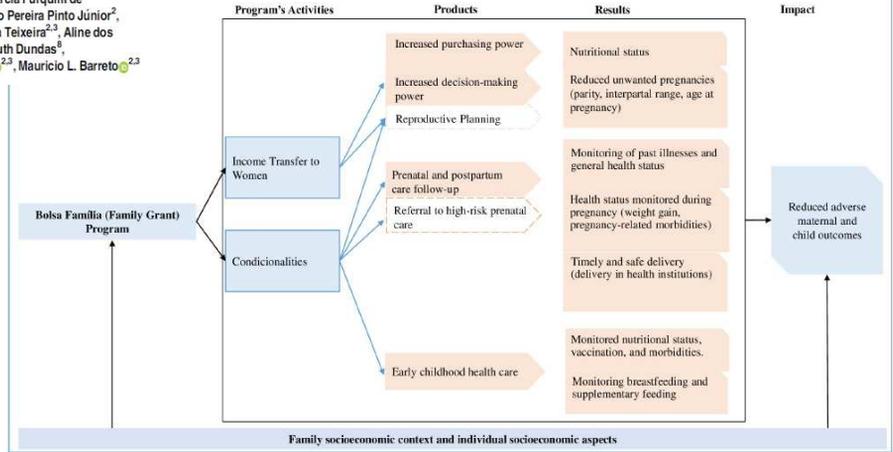


Fig 1. Logical model of the impact of the Bolsa Familia Program (BFP) in reducing adverse maternal and child outcomes.

PLOS MEDICINE

RESEARCH ARTICLE

Conditional cash transfer program and child mortality: A cross-sectional analysis nested within the 100 Million Brazilian Cohort

Dandara Ramos^{1,2*}, Nívea B. da Silva^{1,2*}, Maria Yury Ichihara^{1,2}, Rosemeire L. Fiaccone^{1,2}, Daniela Almeida^{1,4}, Samila Sena¹, Poliana Rebouças^{1,2}, Elzo Pereira Pinto Júnior¹, Enny S. Paixão^{1,5}, Sanni Ali^{1,5}, Laura C. Rodrigues^{1,5}, Mauricio L. Barreto^{1,2}

1 Center for Data and Knowledge Integration for Health (CIDACS), Fundação Oswaldo Cruz, Salvador, Bahia, Brazil, 2 Institute of Collective Health, Federal University of Bahia, Salvador, Bahia, Brazil, 3 Statistics Department, Institute of Mathematics and Statistics, Federal University of Bahia, Salvador, Bahia, Brazil, 4 Computer Science Department, Institute of Mathematics and Statistics, Federal University of Bahia, Salvador, Bahia, Brazil, 5 Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, United Kingdom

* These authors contributed equally to this work.
* dandara.ramos@ufba.br



World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:7, No:6, 2013

Kernel Matching versus Inverse Probability Weighting: A Comparative Study

Andy Handouyaha, Tony Haddad, and Frank Eaton

International Journal of Mathematical and Computational Sciences (2013)

After we estimated the PSs, we performed a kernel matching procedure, as proposed by Heckman et al. [31]. Kernel-based matching estimation is a non-parametric approach that uses weighted averages of all individuals in the control group to construct the counterfactual outcome, where more weight is given to units close to the one that needs to be matched

Kernel Matching pode ser mais vantajoso quando comparado aos métodos tradicionais de PSM (nnm, caliper, multivariate distance matching) - evidência de redução significativa de viés nos erros padrão

Considerações sobre métodos com ponderação

- Maior desvantagem: variância pode ser muito grande se existem pesos nas extremidades (PS próximos a 0 ou 1)
 - Sugestões de controle:
 - *Weight trimming*: considerar valores máximos para o peso; valores maiores que o especificado são fixados no máximo.
- Devido à sensibilidade aos pesos nestes métodos, deve-se ter mais cuidado na avaliação da acurácia das estimativas do PS.
- Estratégias alternativas incluem **métodos “doubly-robust”** (Band e Robins, 2005):
 - Combina o modelo de regressão para o desfecho com um modelo para a intervenção (PS) para estimar o efeito causal de uma intervenção no desfecho.

79

79

Uso do Escore de propensão – Breve Comparação

Pareamento

- Unidades são pareadas com base na sua probabilidade de seleção ao tratamento (escore de propensão) – *Propensity Score Matching* (PSM)
- Pareamento pode ser com ou sem reposição do tipo:
 - Exato
 - Caliper (estabelecimento de uma distância máxima, ex: .05)
 - Vizinho mais próximo

Ponderação

- Análise ponderada pelo inverso da probabilidade de seleção ao tratamento – *Inverse probability treatment weighting* (IPTW ou IPW)
- Unidades de tratamento e controle são rebalanceadas com base nas suas características observáveis

80

80

Diagnóstico do Balanceamento

- **PS verdadeiro: escore de balanceamento**
 - A distribuição das covariáveis no *baseline* é independente do grupo de tratamento, condicional ao PS verdadeiro.
- **Avaliar se o modelo para o PS está corretamente especificado**
 - Exame da distribuição das covariáveis mensuradas no *baseline* entre grupos de tratamento para indivíduos com mesmo PS.
 - Se diferenças ainda existirem, modelo com problemas de especificação.
 - Em PSM, avaliação é feita comparando indivíduos dos grupos de tratamento que foram pareados por PS.
 - Usando estratificação, avaliação é feita dentro de cada estrato.
 - Em IPTW, avaliação é feita comparando indivíduos dos grupos de tratamento na amostra ponderada pelo inverso da PS.

81

81

Diagnóstico do Balanceamento

- **Comparação de médias e medianas de covariáveis contínuas e da distribuição das covariáveis categóricas entre grupos de tratamento.**
 - Uso de diferença padronizada (ou “viés padronizado”)

$$d = \frac{\bar{X}_{treat} - \bar{X}_{control}}{\sqrt{\frac{S^2_{treat} + S^2_{control}}{2}}} \quad d = \frac{\hat{p}_{treat} - \hat{p}_{control}}{\sqrt{\frac{\hat{p}_{treat}(1 - \hat{p}_{treat}) + \hat{p}_{control}(1 - \hat{p}_{control})}{2}}}$$

- Medida não é influenciada pelo tamanho amostral.
- Não existe consenso sobre *threshold* para indicar desbalanceamento:
 - Diferença padrão < 0.1 indica diferença negligenciável (Normand et al., 2001)
 - Valor absoluto das diferenças deve ser < 0.25 (Rubin, 2001)

82

82

3086

P. C. AUSTIN

Table I. Comparison of baseline characteristics between treated and untreated subjects in the original unmatched sample.

Variable	Statin: no (N = 6055)	Statin: yes (N = 3049)	P-value	Standardized differences
<i>Demographic characteristics</i>				
Age	68.1 ± 13.8	63.4 ± 12.4	<0.001	0.355
Female	2241 (37.0 Per cent)	887 (29.1 Per cent)	<0.001	0.167
<i>Presenting signs and symptoms</i>				
Acute CHF/pulmonary edema	316 (5.2 Per cent)	122 (4.0 Per cent)	0.010	0.057
Cardiogenic shock	46 (0.8 Per cent)	12 (0.4 Per cent)	0.038	0.046
<i>Cardiac risk factors</i>				
Diabetes	1561 (25.8 Per cent)	774 (25.4 Per cent)	0.684	0.009
Current smoker	2004 (33.1 Per cent)	1070 (35.1 Per cent)	0.057	0.042
Hyperlipidemia	1138 (18.8 Per cent)	1761 (57.8 Per cent)	<0.001	0.910
Hypertension	2681 (44.3 Per cent)	1453 (47.7 Per cent)	0.002	0.068
Family history of CAD	1762 (29.1 Per cent)	1177 (38.6 Per cent)	<0.001	0.204

83

83

3088

P. C. AUSTIN

Table II. Comparison of baseline characteristics between treated and untreated subjects in the propensity-score matched sample.

Variable	Statin: no (N = 2430)	Statin: yes (N = 2430)	Standardized difference
<i>Demographic characteristics</i>			
Age	63.4 ± 13.2	63.5 ± 12.6	0.011
Female	717 (29.50 Per cent)	736 (30.3 Per cent)	0.017
<i>Presenting signs and symptoms</i>			
Acute CHF/pulmonary edema	94 (3.9 Per cent)	93 (3.8 Per cent)	0.002
Cardiogenic shock	15 (0.6 Per cent)	12 (0.5 Per cent)	0.017
<i>Cardiac risk factors</i>			
Diabetes	609 (25.1 Per cent)	606 (24.9 Per cent)	0.003
Current smoker	883 (36.3 Per cent)	872 (35.9 Per cent)	0.009
Hyperlipidemia	1135 (46.7 Per cent)	1,142 (47.0 Per cent)	0.006
Hypertension	1109 (45.6 Per cent)	1,118 (46.0 Per cent)	0.007
Family history of CAD	898 (37.0 Per cent)	912 (37.5 Per cent)	0.012

84

84

Diagnóstico do Balanceamento

- Outra medida usada na avaliação inclui:
 - Razão das variâncias dos PS entre grupos controle e tratado
 - devem estar entre 0.5 e 2 (Rubin, 2001)
- Testes de hipóteses e valores de p não devem ser usados:
 - Balanceamento é uma propriedade “*in-sample*”
 - Procedimentos podem enganar porque mudanças podem estar associadas com poder estatístico.
- Métodos gráficos:
 - Boxplots, funções de distribuição acumuladas, gráficos não paramétricos de densidades para comparar distribuição das variáveis contínuas entre os grupos na amostra pareada com PS (Austin, 2009).

85

85

Diagnóstico do Balanceamento - Métodos gráficos

- Exame da distribuição dos PS nas amostras original e nos grupos pareados:

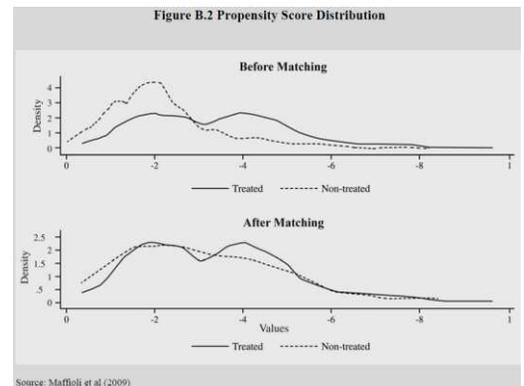
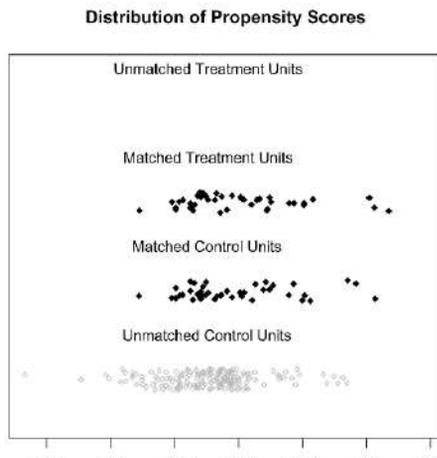


Figure 1. Matches chosen using 1:1 nearest neighbor matching on propensity score. Black dots indicate matched individuals; grey unmatched individuals. Data from Stuart and Green (2008).

86

86

Diagnóstico do Balanceamento – Maiores detalhes

STATISTICS IN MEDICINE
Statist. Med. 2009; **28**:3083–3107
 Published online 15 September 2009 in Wiley InterScience
 (www.interscience.wiley.com) DOI: 10.1002/sim.3697

Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples

Peter C. Austin^{1,2,3,*†}

87

87

Seleção de variáveis para o modelo do PS

- Na prática, não é fácil classificar as variáveis do *baseline* em potenciais ou verdadeiros confundidores.
 - Definição precisa ser feita com base na literatura disponível sobre o tópico.
 - Devido à falta de informação, em muitos casos é mais provável a inclusão de todas as covariáveis mensuradas no *baseline*.
 - Salienta-se que apenas covariáveis mensuradas no *baseline* podem ser incluídas no modelo para PS e **nunca** covariáveis tempo-dependentes ou mensuradas posteriormente que podem ser influenciadas ou modificadas pelo tratamento.
- **Conceito chave na determinação das covariáveis:**
 - “*strong ignorability*”: assume que não há diferenças não observadas entre os grupos de intervenção, condicional nas variáveis observadas.

Variáveis que são afetadas pela intervenção NÃO DEVEM ser incluídas no processo de pareamento:

- Estas variáveis devem ser incluídas no modelo para o desfecho

88

88

Seleção de variáveis para o modelo do PS

- Não existe consenso sobre quais variáveis devem ser incluídas no modelo.
Propostas:
 - 1) Todas as covariáveis mensuradas no baseline.
 - 2) Todas as covariáveis do baseline associadas com a alocação do tratamento (argumentos teóricos).
 - 3) Todas as covariáveis do baseline associadas com o desfecho: potenciais confundidores.
 - 4) Todas as covariáveis do baseline associadas com a alocação do tratamento e com o desfecho: verdadeiros confundidores.
- Alguns estudos têm avaliado a inclusão destes conjuntos de covariáveis no modelo.
 - Uso de quaisquer dos conjuntos de covariáveis resultou em balanceamento entre os grupos.
 - Uso das estratégias (3) e (4) resultou em estimativas do efeito do tratamento mais precisas.
 - Alguns autores indicam preferencialmente uso da estratégia (4): redução de viés e aumento de precisão.

89

89

Passos para os métodos com PS

1. Definir grupo tratamento (e grupo controle)
2. **Selecionar covariáveis** para compor o escore de propensão
3. Verificar o balanceamento antes da implementação do escore de propensão
4. Estimar os escores de propensão
5. Usar o escore de propensão: pareamento, ponderação, subclassificação
6. Verificação do balanceamento depois da implementação do escore de propensão
7. Cálculo do efeito causal em modelos ponderados ou com pareamento pelo escore de propensão
8. Interpretar os resultados Segundo o tipo da escala da variável resposta

90

90

Overlap Weighting – Ponderação baseado no suporte comum

- A ponderação pelo inverso da probabilidade de tratamento utiliza todos os elementos da amostra que recebem tratamento ou não tratamento.
 - ✓ Pode acontecer um peso excessivamente grande nos indivíduos que recebem uma das opções de tratamento e consequentemente não representariam bem um contrafacto para o outro grupo
 - ✓ Uma probabilidade muito pequena de ser tratado (escore de propensão) terá um peso grande podendo levar a super-representação desse(s) indivíduos. Peso pequeno se a probabilidade de ser tratado é grande (sub-representação)
- Abordagem alternativa de ponderação em relação ao procedimento IPTW no que diz respeito a presença de valores extremos nos escores de propensão (próximos a 0 ou 1).
 - ✓ Comum usar procedimentos de estabilização desses pesos extremos ou ainda eliminar as observações com pesos extremos utilizando um limiar
 - ✓ Ambas estratégias podem gerar um aumento na variância da medida de efeito => problema na Inferência

91

91

Overlap Weighting – Ponderação baseado no suporte comum

- Escore de propensão $e(X) = Pr(Z = 1|X)$
- Abordagem IPTW
$$\begin{cases} w(X_i) = \frac{1}{e(X_i)} & \text{for } Z_i = 1, \\ w(X_i) = \frac{1}{1-e(X_i)} & \text{for } Z_i = 0 \end{cases}$$
- Overlap weighting
$$\begin{cases} w(X_i) \propto 1 - e(X_i) & \text{for } Z_i = 1, \\ w(X_i) \propto e(X_i) & \text{for } Z_i = 0 \end{cases}$$
- Esses pesos irão variar de 0 a 1 reduzindo o problema do aumento da variância do efeito em situações envolvendo pesos extremos no procedimento IPTW

92

92

Escores de propensão - *Limitações*

- Embora os procedimentos de pareamento possam ser aplicados em muitos contextos, independentemente das regras de alocação do programa, eles possuem várias limitações:
 - Exigem um conjunto extenso de dados e grandes amostras que, mesmo quando disponíveis, não impedem a falta de suporte comum entre o grupo de tratamento - ou de inscritos - e o conjunto de não participantes
 - O pareamento somente pode ser executado com base em características observadas; por definição, não podemos incorporar características não observadas no cálculo do escore de propensão
 - Sendo assim, para que o procedimento de pareamento identifique um grupo de comparação válido, precisamos ter certeza de que não existe nenhuma diferença sistemática em características não observadas entre as unidades de tratamento e de comparação correspondentes que possa influenciar o resultado (Y)

93

93

Considerações adicionais

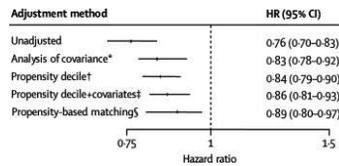
- **Maior parte dos métodos discutidos refere-se aos seguintes cenários:**
 - Tratamento definido em um ponto particular do tempo
 - Covariáveis mensuradas algum tempo antes do tratamento
 - Desfechos mensurados após o tratamento
- **Outras situações mais complexas envolvem**
 - Amostras complexas , por exemplo, amostragem por conglomerado
 - Dados com estrutura multinível (Arpino e Mealli (2011) e Arpino e Cannas (2016)) ou estudos longitudinais com intervenção tempo dependente
 - Estudos onde os indivíduos podem entrar e sair da intervenção, ou onde a data de alocação do tratamento não está definida para o grupo controle.
 - ✓ Métodos úteis nestes casos:
 - Modelos estruturais marginais (Robins et al., 2000)
 - *Balanced risk set matching* (Li et al., 2001)

94

94

Não esquecer

- Validade de PSM depende de suposições que devem ser verificadas, quando possível.
- Diferentes métodos podem ser usados para parear participantes (P) e não participantes (NP) de acordo com PS:
 - Modelo para obter PS (variáveis, modelo);
 - Algoritmo de pareamento
- A escolha de um método de pareamento particular pode afetar a estimação do efeito de intervenção
 - Comparação de resultados entre diferentes métodos pode revelar se a estimação é robusta.



95