EPI66 - Tópicos de Pesquisa I

Uso de DAGs para a identificação de confundidores na pesquisa em saúde

Ricardo de Souza Kuchenbecker
Rodrigo Citton P. dos Reis - citton.padilha@ufrgs.br

Universidade Federal do Rio Grande do Sul
Programa de Pós-Graduação em Epidemiologia

Porto Alegre, 2023

PPGEPI
Programa de Pós-Graduação em Epidemiologia

Introdução

Introdução

Epidemiologia[1]

"é o estudo da distribuição e de determinantes de estados ou eventos relacionados com a saúde em populações especificadas e com a aplicação desse estudo para controlar problemas de saúde".

- O que deve ser visto com atenção nessa definição é que inclui tanto as descrições do conteúdo da disciplina quanto proposta ou aplicação para as quais as investigações epidemiológicas são realizadas.

[1] Porta M: A Dictionary of Epidemiology, 5th ed. New York, Oxford University Press, 2008.

Quais são os objetivos específicos da epidemiologia?

1. Identificar a **etiologia** ou a **causa de uma doença** e os fatores de risco relevantes.
2. Determinar a extensão da doença encontrada em uma comunidade.
3. Estudar a história natural e o prognóstico da doença.
4. Avaliar medidas preventivas e terapêuticas e modelos novos ou existentes de assistência à saúde.
5. Fornecer fundamentos para o desenvolvimento de políticas públicas relacionando problemas ambientais, questões genéticas e outras no que diz respeito à prevenção de doenças e promoção da saúde.

Causalidade na pesquisa em saúde

A história da epidemiologia nos apresenta uma série de modelos e esquemas analíticos para avaliar questões causais.

- Postulados de Henle-Koch;
- "Critérios" de Hill;
- Modelo de causa suficiente e causas componentes de Rothman;
- Modelo causal de Rubin;
- Diagramas causais, entre outros.

O que é inferência causal?

O que é inferência causal?

inferência causal

é a ciência de **inferir a presença e a magnitude das relações de causa e efeito a partir dos dados**.

- Como epidemiologistas, estatísticos, sociólogos, etc., e de fato como seres humanos, é algo sobre o qual **sabemos bastante**.

O que é inferência causal?

Exemplo

Suponha que um estudo encontre uma **associação** entre a propriedade paterna de gravata de seda e a mortalidade infantil. Com base nisso, o governo implementa um programa no qual **cinco gravatas de seda são distribuídas a todos os homens com idade entre 18 e 45 anos**, com o objetivo de reduzir a mortalidade infantil.

- Nós todos concordamos que isso é uma bobagem!
- Isso porque entendemos a diferença entre associação e causalidade.

O que é inferência causal?

A área de inferência causal consiste em (pelo menos) três partes:

1. Uma **linguagem formal** para definir inequivocamente conceitos causais.
2. **Diagramas causais:** uma ferramenta para exibir claramente nossas suposições causais.
3. **Métodos de análise** (isto é, métodos estatísticos) que podem nos ajudar a tirar conclusões causais mais confiáveis a partir dos dados disponíveis.

Um pouco de dor de cabeça!

Um exemplo

- 12 senhoras estão sofrendo de **dor de cabeça**.
- Algumas tomam **aspirina**; outras não.
- Uma hora depois, perguntamos para cada uma delas se a dor de cabeça **sumiu (passou)**.

Os dados observados

	Z (tomou aspirina?)	R (dor de cabeça sumiu?)
Mary	0	0
Anna	1	0
Emma	1	1
Elizabeth	0	0
Minnie	0	1
Margaret	1	0
Ida	1	0
Alice	0	0
Bertha	0	1
Sarah	0	0
Annie	0	1
Clara	1	1

Os dados observados

	Z (tomou aspirina?)	R (dor de cabeça sumiu?)
Mary	0	0
Anna	1	0
Emma	1	1
Elizabeth	0	0
Minnie	0	1
Margaret	1	0
Ida	1	0
Alice	0	0
Bertha	0	1
Sarah	0	0
Annie	0	1
Clara	1	1

Os dados observados

- Emma tomou aspirina ($Z = 1$) e a sua dor de cabeça passou ($R = 1$).
- **A aspirina causou o desaparecimento da sua dor de cabeça?**

Questões Causais

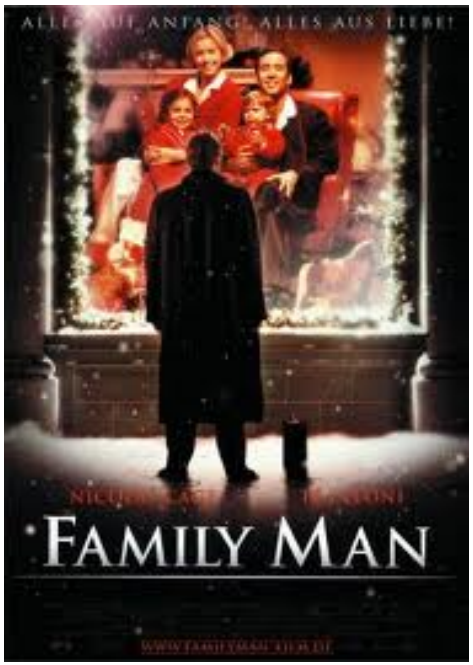
- Esta é uma questão causal, uma questão sobre os **efeitos causados pelos tratamentos**.
- A questão começa com dois possíveis tratamentos para a dor de cabeça[2].
- Para um indivíduo específico é perguntado:
  - O que aconteceria com esse indivíduo sob o primeiro tratamento?
  - O que aconteceria com o indivíduo sob o segundo tratamento?
  - O indivíduo se sairia melhor sob o primeiro, em vez do segundo tratamento?
  - O desfecho seria o mesmo sob os dois tratamentos?

efeitos causais

são comparações de desfechos (resultados, ou respostas) potenciais sob tratamentos alternativos.

[2] Consideramos apenas dois níveis de tratamento por uma questão de simplicidade. Esta ideia pode ser generalizada para múltiplos níveis de tratamento e para outros regimes de tratamento mais gerais.

Desfechos potenciais no senso comum



(The Family Man, 2000) Jack Campbell é um investidor de Wall Street jovem e solteiro vivendo uma vida de rico em Nova Iorque. Ele se surpreende quando sua ex-namorada, Kate, tentou ligar para ele após anos sem se verem. Após uma conversa com o seu mentor na empresa, Jack resolve pensar se responderia a esta chamada no dia seguinte. Naquela noite de Natal, porém, ele resolve ir a pé até sua casa, passando por uma loja de conveniências no caminho e convencendo para que um vencedor da loteria, irritado, chamado Cash, não atirasse no vendedor. Ele oferece ajuda à Cash antes de ir dormir em sua cobertura.

Tudo muda num passe de mágica quando na manhã seguinte ele acorda em um quarto no subúrbio de Nova Jersey com Kate, a sua atual esposa, com quem anteriormente ele havia deixado de se casar e ainda com duas crianças que ele sequer conhecia, Jack percebe então que esta é justamente **a vida que ele teria** se não tivesse se transformado em um investidor financeiro quando jovem. Ao invés disso, ele tem uma vida modesta, onde ele é um vendedor de pneus e Kate é uma advogada não-remunerada.

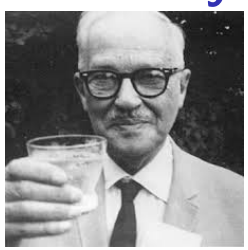
Desfechos potenciais no senso comum

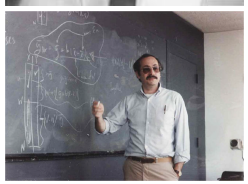
“Num bosque amarelo dois caminhos se separavam,
E lamentando não poder seguir os dois [...]”

Desfechos potenciais no senso comum

Desfechos potenciais: breve histórico

- Na estatística, a ideia de definir efeitos causais como comparações de desfechos potenciais sob tratamentos alternativos é creditada a **Jerzy Neyman**[3] e **Donald Rubin**[4].





- Neyman introduziu a ideia no contexto de experimentos aleatorizados nos anos 1920.
- Rubin desenvolveu a ideia em outras áreas de inferência causal.

---

[3] Splawa-Neyman, J., Dabrowska, D.M., Speed, T.P. On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9. *Statistical Science* 5:465-472, 1990.

[4] Rubin, D.B. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66:688–701, 1974.

A estrutura da inferência causal

- Z é o tratamento atribuído: tomou aspirina?
- R é o desfecho/resposta: dor de cabeça sumiu?
- r_C e r_T representam as **desfechos potenciais**.
  - r_C é a desfecho/resposta que teria sido observada caso a aspirina NÃO tivesse sido tomada.
  - r_T é a desfecho/resposta que teria sido observada caso a aspirina tivesse sido tomada.
- **Uma** destas respostas **é observada**: se $Z = 0$, r_C é observada; se $Z = 1$, r_T é observada (ou seja, $R = Z \times r_T + (1 - Z) \times r_C$)[5].
- A outra é **contrafactual**[6].

[5] Muitas vezes, referida como a **suposição de consistência**.
[6] Não-observada, ausente.

Os dados ideais

	r_C	r_T
Mary	0	0
Anna	1	0
Emma	0	1
Elizabeth	0	0
Minnie	1	1
Margaret	0	0
Ida	0	0
Alice	0	0
Bertha	1	0
Sarah	0	0
Annie	1	1
Clara	0	1

Os dados ideais

Com o par de repostas potenciais, podemos responder as seguintes perguntas:

- A aspirina causou o desaparecimento da dor de cabeça de Emma?
  - E de Margaret?
  - E de Clara?
  - E de Alice?

Os dados ideais

i	r_{C_i}	r_{T_i}	$\delta_i = r_{T_i} - r_{C_i} \neq 0$? (efeito causal?)
Mary	0	0	Não
Anna	1	0	Sim, prejudicial
Emma	0	1	Sim, benéfico
Elizabeth	0	0	Não
Minnie	1	1	Não
Margaret	0	0	Não
Ida	0	0	Não
Alice	0	0	Não
Bertha	1	0	Sim, prejudicial
Sarah	0	0	Não
Annie	1	1	Não
Clara	0	1	Sim, benéfico

O problema fundamental da inferência causal

i	r_{C_i}	r_{T_i}	Z_i	R_i
Mary	0	?	0	0
Anna	?	0	1	0
Emma	?	1	1	1
Elizabeth	0	?	0	0
Minnie	1	?	0	1
Margaret	?	0	1	0
Ida	?	0	1	0
Alice	0	?	0	0
Bertha	1	?	0	1
Sarah	0	?	0	0
Annie	1	?	0	1
Clara	?	1	1	1

Efeitos causais populacionais

- $\delta_i = r_{T_i} - r_{C_i} =$?, para todo indivíduo i, pois um dos desfechos potenciais nunca é observada.
- Um objetivo menos ambicioso é focar no **efeito causal médio** (ou **efeito causal em nível populacional**)[7]:

$$\overline{\delta} = \overline{r}_T - \overline{r}_C.$$

- No caso em que a **resposta** é **dicotômica**, temos

$$\overline{\delta} = \Pr(r_T = 1) - \Pr(r_C = 1).$$

[7] Utilizando o operador $\mathrm{E}[\cdot]$, temos que $\overline{\delta} = \mathrm{E}[\delta_i] = \mathrm{E}[r_{T_i}] - \mathrm{E}[r_{C_i}]$.

Efeitos causais populacionais

i	r_{C_i}	r_{T_i}	$\delta_i = r_{T_i} - r_{C_i} \neq 0$? (efeito causal?)
Mary	0	0	Não
Anna	1	0	Sim, prejudicial
Emma	0	1	Sim, benéfico
Elizabeth	0	0	Não
Minnie	1	1	Não
Margaret	0	0	Não
Ida	0	0	Não
Alice	0	0	Não
Bertha	1	0	Sim, prejudicial
Sarah	0	0	Não
Annie	1	1	Não
Clara	0	1	Sim, benéfico

- $\overline{r}_T = \Pr(r_T = 1) = 4/12$ e $\overline{r}_C = \Pr(r_C = 1) = 4/12$, e portanto,

$$\overline{\delta} = \overline{r}_T - \overline{r}_C = \frac{4}{12} - \frac{4}{12} = 0.$$

- Ou seja, concluímos que **não existe** efeito causal em nível populacional.

Efeitos causais populacionais

- Em verdade, **não sabemos** r_T para cada indivíduo, então **não podemos simplesmente estimar** $\Pr(r_T = 1)$ como a proporção de todos os indivíduos com $r_T = 1$.
- Da mesma forma, **não podemos simplesmente estimar** $\Pr(r_C = 1)$ como a proporção de todos os indivíduos com $r_C = 1$.
- Assim, **não podemos estimar** facilmente $\overline{\delta} = \Pr(r_T = 1) - \Pr(r_C = 1)$ pelo mesmo motivo que não podemos estimar $\delta_i = r_{T_i} - r_{C_i}$.
- A inferência causal **é toda sobre a escolha de quantidades dos dados observados** (isto é, envolvendo Z, R e outras variáveis observadas) **que representam substitutos razoáveis** para quantidades hipotéticas tais como $\overline{\delta}$, que envolvem contrafactuais não observáveis.

Quando associção é igual a causação?

- O que pode ser um bom substituto para $\Pr(r_T = 1)$?
  - Que tal $\hat{r}_T = \Pr(R = 1|Z = 1)$?
  - Esta é a proporção de "dor de cabeça desapareceu" entre aquelas senhoras que realmente tomaram a aspirina.
  - Isso é o mesmo que $\Pr(r_T = 1)$?
    - Somente se aquelas que tomaram a aspirina forem **intercambiáveis**[8] com aquelas que não o fizeram.
- Este seria o caso se a escolha de tomar a aspirina fosse feita de forma **aleatória**.
- É por isso que **experimentos aleatorizados** são o **padrão-ouro** para inferir efeitos causais.

[8] Diz-se que o par de desfechos potenciais é **independente** da alocação ao tratamento. Ou seja, $\{r_{T_i}, r_{C_i}\} \perp\!\!\!\perp Z_i$.

Quando associção é igual a causação?

i	r_{C_i}	r_{T_i}	Z_i	R_i
Mary	0	?	0	0
Anna	?	0	1	0
Emma	?	1	1	1
Elizabeth	0	?	0	0
Minnie	1	?	0	1
Margaret	?	0	1	0
Ida	?	0	1	0
Alice	0	?	0	0
Bertha	1	?	0	1
Sarah	0	?	0	0
Annie	1	?	0	1
Clara	?	1	1	1

- $\hat{r}_T = \Pr(R = 1|Z = 1) = 2/5$ e $\hat{r}_C = \Pr(R = 1|Z = 0) = 3/7$, e portanto,

$$\hat{\delta} = \hat{r}_T - \hat{r}_C = \frac{2}{5} - \frac{3}{7} = -\frac{1}{35}.$$

- Se assumirmos "associação = causação", concluiremos que a aspirina foi, em média, prejudicial.

Mas, e se . . .

. . . as senhoras com uma dor de cabeça **mais forte** (grave) fossem **mais propensas** a tomarem a aspirina?

- Neste caso, "associação $\neq$ causalidade"!

Levando em conta a gravidade

- Suponha que perguntamos a cada uma das 12 senhoras no início do estudo: **"sua dor de cabeça é forte?"**.
  - Então, poderíamos propor que, depois de levar em conta a gravidade, a decisão de tomar ou não a aspirina fosse efetivamente tomada de forma aleatória.
- Suponha que X denota a gravidade. Então, sob essa suposição, **dentro dos estratos** de X, os indivíduos expostos e não expostos podem ser **intercambiáveis**.
  - Isso é chamado de **intercambiabilidade (permutabilidade) condicional** (dado X)[9].
- Sob intercambiabilidade condicional dada X, "associação = causação" **dentro dos estratos** de X.

[9] Ou seja, $\{r_{T_i}, r_{C_i}\} \perp\!\!\!\perp Z_i | X_i$.

Levando em conta a gravidade

i	r_{C_i}	r_{T_i}	Z_i	R_i	X_i
Mary	0	0	0	0	1
Anna	1	0	1	0	0
Emma	0	1	1	1	0
Elizabeth	0	0	0	0	1
Minnie	1	1	0	1	0
Margaret	0	0	1	0	1
Ida	0	0	1	0	1
Alice	0	0	0	0	0
Bertha	1	0	0	1	1
Sarah	0	0	0	0	0
Annie	1	1	0	1	0
Clara	0	1	1	1	1

Estratificando por gravidade

No estrato $X = 0$

- $\hat{r}_T = \Pr(R = 1|Z = 1) = 1/2$ e $\hat{r}_C = \Pr(R = 1|Z = 0) = 2/4$, e portanto,

$$\hat{\delta} = \hat{r}_T - \hat{r}_C = \frac{1}{2} - \frac{2}{4} = 0.$$

No estrato $X = 0$:

- $\hat{r}_T = \Pr(R = 1|Z = 1) = 1/3$ e $\hat{r}_C = \Pr(R = 1|Z = 0) = 1/3$, e portanto,

$$\hat{\delta} = \hat{r}_T - \hat{r}_C = \frac{1}{3} - \frac{1}{3} = 0.$$

Estratificando por gravidade

- Ou seja, dentro dos estratos **não existe** efeito causal.
- Com alguma técnica para combinar os resultados dos estratos **(Cochran-Mantel-Haenszel)**, chegaremos a mesma conclusão que no caso em que "conhecíamos" o par de desfechos potenciais para cada senhora no estudo.

Exemplo da dor de cabeça: breves conclusões

- De maneira mais geral, se **existe** um efeito causal de Z em R, mas também uma associação não-causal **(efeito de confusão)** devido a X, então o **efeito causal** será estimado **com viés**, a menos que **estratifiquemos/condicionemos** em X.
- A **intercambiabilidade condicional** é o principal critério que nos permite fazer declarações causais usando **dados observacionais**.
- Assim, precisamos identificar, se possível, um conjunto de **(co)variáveis/confundidores** $X_1, X_2, \ldots$, de tal forma que a intercambiabilidade condicional é válida, dado este conjunto de variáveis.

Exemplo da dor de cabeça: breves conclusões

- Na vida real, pode haver muitas variáveis candidatas X.
- Estes podem ser causalmente inter-relacionados de uma maneira muito complexa.
- Decidir se os indivíduos expostos e os não expostos são condicionalmente intercambiáveis, dado $X_1, X_2, \ldots$, **requer conhecimento detalhado do assunto**.

Os

diagramas causais

podem nos ajudar a usar esse conhecimento para determinar se a intercambiabilidade condicional é válida ou não.

Bons estudos!