

O modelo estatístico de causalidade

Ronir Raggio Luiz
Claudio José Struchiner

SciELO Books / SciELO Livros / SciELO Libros

LUIZ, RR., and STRUCHINER, CJ. *Inferência causal em epidemiologia: o modelo de respostas potenciais* [online]. Rio de Janeiro: Editora FIOCRUZ, 2002. 112 p. ISBN 85-7541-010-5. Available from SciELO Books <<http://books.scielo.org>>.



All the contents of this chapter, except where otherwise noted, is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported.

Todo o conteúdo deste capítulo, exceto quando houver ressalva, é publicado sob a licença Creative Commons Atribuição - Uso Não Comercial - Partilha nos Mesmos Termos 3.0 Não adaptada.

Todo el contenido de este capítulo, excepto donde se indique lo contrario, está bajo licencia de la licencia Creative Commons Reconocimiento-NoComercial-CompartirIgual 3.0 Unported.

O Modelo Estatístico de Causalidade

Causalidade e Estatística

A relação entre causa e estatística merece alguns comentários iniciais. Sustentados na retórica estatística de que correlação não implica causalidade e apesar do reconhecimento geral de que um experimento randomizado bem planejado pode fornecer uma poderosa ajuda na investigação de relações causais, os estatísticos abstiveram-se de pensar em causalidade de forma mais específica e audaciosa. Isso é sutilmente evidenciado quando, por exemplo, lemos o trabalho de Kempthorne (1978), um grande estatístico contemporâneo, que, muito modestamente, se preocupa em não ser arrogante quando ele, um ‘mero’ estatístico (em sua auto-avaliação), aborda a questão do que seja um conceito viável de causa.

Dempster (1990) argumenta que a primeira razão para que os estatísticos analisem causalidade é prática. O pensamento causal está profundamente presente no entendimento científico dos problemas da estatística aplicada. Por meio de alguns exemplos práticos, tal como a questão de saber se a redução da ingestão de álcool por mulheres elimina ou adia o surgimento de alguns casos de câncer de mama, o autor conclui que a coleta e a análise de dados estatísticos constituem uma porção apreciável da base empírica em que ocorrem as tentativas cientificamente confiáveis de atribuição de causalidade. Ou, como observado por Gail (1996), o pensamento estatístico,

presente na coleta e na análise de dados, tem sido crucial para o entendimento do poder e das potenciais fragilidades das evidências científicas. O resultado disso é um crescente desenvolvimento do horizonte do estatístico e dos métodos estatísticos, impulsionando o surgimento de novas metodologias e debates construtivos sobre critérios necessários para se inferirem relações causais. Esse conhecimento estatístico forma a base de grande parte da prática epidemiológica atual.

Em termos mais específicos, o processo de inferir causalidade envolve diversas tarefas estatísticas. É decisivamente importante, quando se discutem efeitos causais, começar definindo a quantidade a ser estimada. Subseqüentemente, deve-se estudar métodos para coletar dados com os quais essa quantidade possa ser estimada. E, finalmente, deve-se considerar técnicas estatísticas para inferir os efeitos com base nos dados observados (Rosenbaun & Rubin, 1984). É na etapa de definição da quantidade a ser estimada que se concentra a mudança na lógica da investigação causal devida a Rubin (1974), por meio do modelo de respostas potenciais. Entretanto, no momento da coleta de dados para estimação dessa quantidade, o esquema de amostragem necessário deve ser considerado em razão do desenho de estudo epidemiológico. Já para a etapa de inferência causal baseada nas observações, diversas técnicas podem ser utilizadas. Rubin (1990a) tem discutido e comparado essas técnicas, com especial destaque para o método bayesiano.

Assim, o problema de atribuição de causalidade, em uma leitura estatística, é interpretado como um problema de detecção de efeitos causais por meio de mensurações. Medir efeitos causais constitui, pois, a base do desenvolvimento da teoria estatística que discute causalidade, com particular interesse em epidemiologia. Nessa visão, inferir estatisticamente efeitos causais traduz-se em um processo de estimação de efeitos devido a potenciais manipulações que podem ou poderiam ser aplicadas a uma unidade em um con-

texto real com todas as suas complexidades, necessitando, portanto, de definições adequadas a essa estrutura complicada (Rubin, 1990a).

Uma primeira característica importante associada à contribuição da estatística para a discussão de causalidade, que acaba se refletindo em uma limitação, se refere à concentração do estudo nos efeitos de causas previamente determinadas, em oposição à abordagem usual, que busca as causas de efeitos observados. Essa é uma condição imposta por Holland (1986) e defendida por Rubin, o qual, de forma incisiva, expressa:

Em qualquer situação relativamente complicada do mundo real, envolvendo plantas, animais, pessoas, aviões ou reatores nucleares, eu acredito que é geralmente impossível examinar uma resposta observada e realisticamente encontrar a causa dela. A única esperança para tal atribuição é descrever a situação muito cuidadosamente para limitar o tratamento (causa) sugerido que tenha ocorrido e indicar a) quais outros fatores (outras potenciais causas) estão sendo assumidos fixados em seus valores observados; e b) quais tratamentos (causas) alternativos contrafatuais estão sendo considerados terem acontecidos preferivelmente ao tratamento observado, que é a causa postulada. (Rubin, 1990a:280)

Dempster (1990) tenta justificar essa insistência de Holland (1986) e Rubin (1990a), afirmando que os princípios estatísticos que estudam as dificuldades de se fazer induções sobre o mundo real com base em correlações empíricas não estão estreitamente ligados ao conceito de causa. Acrescenta que o estatístico deve também estar interessado na abordagem tradicional, que estuda as causas de efeitos observados, utilizando-se, entretanto, de princípios complementares àqueles descritos por Holland no Modelo de Rubin. Entretanto, essa restrição parece justificável, uma vez que tenta responder de forma mais pragmática à questão de como identificar uma causa. Não necessariamente exclui outras formas de estudar causalidade, apenas limita a participação da estatística nessa discussão. Os princípios complementares aludidos por Dempster (1990) fazem uso mais especificamente de conceitos próprios à teoria filosófica.

Outra característica importante contida nesta discussão, e que também é bastante controversa, se refere à definição do que pode ser considerado uma causa. A condição-chave dessa idéia, sob a ótica estatística, é que cada unidade deve estar potencialmente exposta à ação de qualquer uma das causas cujo efeito poderia ser medido. Assim, atributos pessoais imutáveis, tais como sexo ou raça, não podem ser vistos como causas, já que não se poderia observar seu efeito sob a condição alternativa àquela que a unidade possui. É bastante comum encontrar declarações de causalidade atribuídas a características pessoais. Mas tais declarações, na conotação de causa de Holland (1986), são sempre declarações de associação entre os valores de um atributo e uma variável resposta para as unidades de uma população. Assim, pode-se dizer, por exemplo, que certa pessoa não teve câncer de pele (melanoma) ‘porque’ era negra. No entanto, tudo o que pode ser inferido dessa declaração é que a proporção de casos de melanoma é inferior entre negros, apesar da condição causal que a declaração carrega. Tal restrição está longe de ser um consenso por parte de autores preocupados com a questão da causalidade, principalmente os filósofos. Entretanto, entre os estatísticos ela é razoavelmente bem aceita. Kempthorne (1978) diz que é sem sentido epistemológico falar de uma característica de um indivíduo ‘causando’ ou determinando uma outra característica individual. Cox (1986, 1992) parece acolher a idéia ao reconhecer que certas variáveis não podem propriamente ser vistas como causas.

Deve ser reconhecido, todavia, que se um atributo pessoal pode ou não ser uma causa depende da conotação que se dá à palavra. Mas, para o modelo estatístico a ser discutido a seguir, essa condição é teoricamente importante, pois a idéia de causa nele contida se refere à comparação de respostas potenciais sob as diferentes situações de exposição (causas ou tratamentos) consideradas.

O Modelo de Respostas Potenciais

Estrutura do modelo

Um estudo estatístico de efeitos causais é, portanto, aquele em que se comparam os resultados de dois ou mais tratamentos em uma população de unidades, onde, em princípio, cada uma destas unidades poderia ser exposta a qualquer um dos tratamentos. A estruturação desse modelo, também conhecido como Modelo de Rubin, foi pioneiramente descrita de forma eloqüente por Holland (1986) e apresentada novamente de forma sucinta em outros trabalhos (Holland & Rubin, 1988; Holland, 1989; Wainer, 1991; Halloran & Struchiner, 1995; Angrist et al., 1996; Little & Rubin, 2000). A seguir, tenta-se reproduzir quase que integralmente essa estrutura, a fim de consolidar a notação.

Os elementos essenciais que compõem o Modelo de Rubin:

- uma população de unidades, U ;
- um conjunto, K , de agentes causais bem definidos (também chamados tratamentos ou causas) para os quais cada unidade $u \in U$ possa ser exposta. Para efeito de simplificação, serão considerados apenas dois agentes causais, $K = \{t; c\}$, onde t representa tratamento ou exposição e c controle ou não exposição;
- uma resposta, Y , a variável dependente, que pode ser registrada para cada unidade após a exposição aos agentes causais em K . Y será considerada dicotômica, na medida em que em epidemiologia o principal interesse é a ocorrência ou não de uma determinada doença. A extensão para um Y qualquer é imediata.

Nesse modelo, o papel do tempo é fundamental. Quando uma unidade é exposta a uma causa, isso deve acontecer em algum tempo ou dentro de um período de tempo específico. Assim, as variáveis dividem-se em pré-

exposição, aquelas cujos valores são determinados anteriormente à exposição à causa, e pós-exposição, aquelas determinadas após. A função da variável resposta Y é medir o efeito de uma causa. Logo, se encontra na classe das variáveis pós-exposição. Isso dá surgimento a uma característica crítica do modelo, isto é, o valor de uma variável pós-exposição é potencialmente afetado pela particular causa, t ou c , para a qual a unidade é exposta. E isso é exatamente equivalente à declaração de que causas têm efeitos, que é a essência da idéia de causalidade.

Assim, no lugar de uma variável dependente simples Y , tem-se uma variável dependente Y_k para cada um dos tratamentos para os quais a unidade pode potencialmente ser exposta. Se uma unidade é exposta ao agente causal t , registra-se o valor Y_t para esta unidade. Se a mesma unidade tivesse sido exposta ao agente causal c no lugar de t , seria registrado o valor Y_c e não o valor Y_t . Formalmente, para dois tratamentos, associa-se o vetor (Y_t, Y_c) para cada unidade $u \in U$, onde $Y_k(u)$ é a resposta obtida para a unidade u quando exposta à causa $k \in K$.

O efeito da causa t sobre u quando medido por Y relativo à causa c é a diferença entre $Y_t(u)$ e $Y_c(u)$. No modelo, será representado pela diferença algébrica

$$Y_t(u) - Y_c(u).$$

Holland (1986) chama essa diferença de efeito causal de t relativo à c sobre u , quando medido por Y . Essa é a maneira com que o modelo para inferência causal expressa a mais básica de todas as declarações causais, sendo na realidade a quantidade que se gostaria de poder estimar a partir de dados observáveis e hipóteses plausíveis. Uma forma alternativa de se quantificar o efeito causal é considerar a razão entre as respostas potenciais.

Assim, dentro da situação considerada, onde tanto a variável exposição (tratamento) quanto a variável resposta (doença) são dicotômicas, assumindo-se o valor 1 para presença e 0 para ausência, o efeito causal indi-

vidual $Y_t(u) - Y_c(u)$ será igual a 1 somente se u torna-se doente quando exposto, mas não se torna doente quando não exposto. $Y_t(u) - Y_c(u)$ será 0 quando o *status* de doença de u é o mesmo, independentemente da condição de exposição. E, finalmente, será -1 se u não adoece quando exposto, mas adoece quando não exposto. Assim, para um determinado indivíduo u_0 , o valor 0 para o efeito causal $Y_t(u) - Y_c(u)$ é indicação de ausência de efeito da causa t relativa à causa c para a unidade u_0 . Já os valores 1 e -1 representam existência de efeito da causa postulada, sendo o valor -1 indicativo de uma causa preventiva – por exemplo, uma vacina.

Inferência causal está, enfim, interessada no efeito de causas sobre unidades específicas, isto é, está interessada em determinar o valor do efeito causal $Y_t(u) - Y_c(u)$. É frustrada, entretanto, por uma limitação de observação que Holland (1986) chama de Problema Fundamental da Inferência Causal. Qual seja: é impossível ‘observar’ os valores de $Y_t(u)$ e $Y_c(u)$ para a mesma unidade u e, portanto, é impossível ‘observar’ o efeito de t relativo a c sobre u . No entanto, a aplicação dessa afirmação depende da natureza dos tratamentos e das unidades sob estudo.

A aparente inviabilidade de fazer inferências causais em consequência desse problema desaparece quando se nota que a impossibilidade de observação simultânea de $Y_t(u)$ e $Y_c(u)$ não significa ausência total de informação relevante sobre estes valores. E essa informação depende da situação considerada. Holland (1986) assinala duas soluções para o problema, que chama de ‘solução científica’ e ‘solução estatística’.

A solução científica faz uso de hipóteses de homogeneidade ou estabilidade. Ao estudar o comportamento de uma peça em laboratório, um cientista pode acreditar que o valor de $Y_c(u)$ não depende do momento em que é medido – hipótese de ‘estabilidade temporal’ – e o valor de $Y_t(u)$ não é afetado pela exposição anterior da unidade u à causa c – hipótese de causa ‘transiente’. Assim, para superar o Problema Fundamental da Inferência

Causal, o cientista expõe u à causa c , mede $Y_c(u)$ e, subsequente, expõe u à causa t e mede $Y_t(u)$. A obtenção do efeito causal em nível individual $Y_t(u) - Y_c(u)$ torna-se, então, imediata. Note-se, entretanto, que o efeito causal que pode ser obtido pelo cientista é dependente de hipóteses não testáveis. Isto é, o cientista não é capaz de provar as suposições que fez de estabilidade temporal e causa transiente. Com um trabalho cuidadoso, ele pode convencer a si e aos outros de que são válidas, mas nunca poderá estar absolutamente certo disso. Essa abordagem é bastante comum em experimentos físicos e está presente em nossas pequenas avaliações de causalidade feitas diariamente.

Uma segunda maneira de aplicar a solução científica é assumir que $Y_t(u_1) = Y_t(u_2)$ e $Y_c(u_1) = Y_c(u_2)$ para duas unidades u_1 e u_2 . Ou, de forma mais geral, assumir que $Y_k(u_1) = Y_k(u_2)$ para todo par de unidades u_1 e u_2 . Essa é a hipótese que Holland (1986) chama de ‘homogeneidade de unidades’. Sob tal hipótese, a obtenção do efeito causal em nível individual também se torna imediata, uma vez que $Y_t(u) - Y_c(u) = Y_t(u_1) - Y_c(u_2)$ para todo u , u_1 e $u_2 \in U$. Apesar de também ser bastante comum em laboratórios, essa solução para o problema fundamental da inferência causal tem sido buscada por estatísticos e epidemiologistas, que procuram operacionalizá-la por meio de procedimentos de seleção de unidades. A maneira que os cientistas de laboratórios (e, de forma análoga, os epidemiologistas) utilizam para se convencer de que as unidades são homogêneas é prepará-las (ou selecioná-las) cuidadosamente, de modo que elas pareçam idênticas em todos os aspectos relevantes. Apesar de se poder fazer a hipótese de homogeneidade plausível, sua validade não é passível de ser testada principalmente devido à possibilidade de existência de outras variáveis não observadas eventualmente importantes. Na maioria dos estudos epidemiológicos, senão em todos, a hipótese de homogeneidade obtida por estratificação, isto é, pela separação das unidades em subgrupos com base em co-variáveis observadas, dificilmente seria justificada, pois raramente

existiriam situações em que todos os fatores de risco importantes são acuradamente medidos e controlados (Greenland, 1990). Entretanto, a estratificação e, num caso mais extremo, o pareamento constituem estratégias de seleção de unidades comumente utilizadas na avaliação de efeitos causais por meio de uma aproximação, tanto quanto possível, da hipótese de homogeneidade de unidades.

Como visto, estratificação é entendida como a busca de unidades homogêneas por meio da identificação de subpopulações com base em co-variáveis pré-exposição observadas. O pareamento é uma tentativa de se atingir comparabilidade entre potenciais fatores de confundimento no estágio de desenho do estudo. Isso é feito selecionando-se apropriadamente, para o estudo, unidades para formar os pares que sejam tão semelhantes quanto possível com respeito a potenciais variáveis de confundimento. Consiste, portanto, na escolha de uma ou mais unidades controle para cada unidade tratada que seja similar quanto a características mensuradas anteriormente à ação dos tratamentos.

Cabe aqui uma pequena apresentação da idéia de confundimento. Uma discussão mais detalhada será feita em tópico específico. Como exemplo, considere um estudo experimental para avaliação de efeitos causais no qual, uma vez analisados os dados apropriadamente a partir da amostra disponível, conclui-se pela existência de associação entre exposição e doença, não importando por ora a magnitude desta associação. Três situações alternativas devem ser consideradas:

- existe de fato uma associação causal entre exposição e doença e o estudo foi capaz de detectá-la;
- ‘não’ existe de fato associação causal entre exposição e doença, mas por mero acaso uma ‘infeliz’ amostra forneceu evidências de causalidade. Neste caso, a ordem de grandeza do acaso é comumente medida pelo p-valor;

- ‘não’ existe de fato associação causal, mas um viés no estudo foi responsável pela evidência amostral de causalidade quando avaliada pelo p-valor.

Em epidemiologia, os principais vieses, cujo controle é fundamental para causalidade, têm sido tradicionalmente classificados em viés de seleção, viés de informação e confundimento (ou confusão). Assim, confundimento se refere a uma explicação alternativa para uma determinada conclusão, constituindo portanto um dos principais – senão o principal – problemas para a inferência causal. Aproveitando o exemplo para antecipar um outro elemento fundamental desse modelo estatístico de causalidade, esta terceira situação, em que os dados apontam a existência de causalidade em virtude da presença de algum viés, tem sido apresentada também como um problema relacionado ao mecanismo de designação de tratamentos (Rubin, 1991). Assim, confundimento poderia ser visto também como algo devido ao emprego de um mecanismo de designação de tratamentos impróprio.

Voltando ao Problema Fundamental da Inferência Causal, de forma diferente da solução científica, a solução estatística substitui a busca do efeito causal obtido em nível individual pela busca de um efeito causal ‘típico’, em nível populacional ou subpopulacional. O efeito causal médio, T , de t relativo à c sobre U é a esperança da diferença $Y_t(u) - Y_c(u)$ para as unidades $u \in U$. Isto é,

$$T = E(Y_t - Y_c).$$

Que, pelas regras usuais de probabilidade, pode ser expresso como

$$T = E(Y_t) - E(Y_c).$$

Essa expressão revela que respostas que ‘podem ser observadas’ sobre diferentes unidades podem ser utilizadas para obter informação sobre T . Se algumas unidades são expostas à causa t , elas podem ser usadas para fornecer informação sobre $E(Y_t)$, já que este é o valor médio de Y_t sobre U , enquanto outras unidades podem ser expostas à c e se obtiver informação sobre $E(Y_c)$.

Na situação considerada, em que a resposta Y é dicotômica assumindo o valor 1 quando presente e 0 quando ausente, pode-se reconhecer no parâmetro T , definido anteriormente, a medida de efeito conhecida pelos epidemiologistas como risco atribuível. Isso porque

$$E(Y_t) = 1 \cdot \Pr(Y_t = 1) + 0 \cdot \Pr(Y_t = 0) = \Pr(Y_t = 1) = \text{Risco entre tratados} = R_t$$

De forma análoga, o risco entre controles (ou não expostos) é

$$R_c = \Pr(Y_c = 1) = E(Y_c).$$

Assim, T mede a parcela do risco entre os tratados que pode ser atribuída ao tratamento, já que remove o risco em consequência de outras causas.

O mecanismo que indica qual tratamento (exposição), t ou c , a unidade u recebe, é muito importante e envolve todas as considerações estatísticas de um experimento bem planejado, tal como a randomização ou procedimentos a ela alternativos. Assim, a solução estatística substitui o efeito causal impossível de se observar de t sobre unidades específicas, pelo possível de ser estimado efeito causal médio de t , sobre uma população de unidades. Em resumo, diante da impossibilidade de se observar $Y_t(u)$ e $Y_c(u)$ para a mesma unidade, elege-se um conjunto de unidades a serem expostas à causa t estimando-se $E(Y_t)$ e um outro conjunto a ser exposto à causa c para se estimar $E(Y_c)$.

Tal como a solução científica, a solução estatística para o Problema Fundamental da Inferência Causal pressupõe a validade de hipóteses não passíveis de serem testadas, ou apenas parcialmente testáveis. As hipóteses adjacentes à solução estatística são duas: a de independência ou a de efeito constante. Para discussão da hipótese de independência, é necessário incorporar um novo elemento ao modelo. Seja S a variável que indica a causa (tratamento) a que cada unidade em U é exposta. Isto é, $S(u) = t$ indica que a unidade é exposta à causa t e se observa $Y_t(u)$. Por sua vez, $S(u) = c$ indica exposição à c e a resposta observada é $Y_c(u)$. Assim, a resposta observada para a unidade u é $Y_{S(u)}(u)$. A variável resposta observada é, portanto,

Y_s e, mesmo que agora o modelo contenha três variáveis, S , Y_t e Y_c , o processo de observação envolve apenas duas, S e Y_s . Holland (1986) observa ainda que a distinção entre as três situações envolvidas, isto é, o processo de mensuração Y que produz a variável resposta, as duas versões da variável resposta Y_t e Y_c que correspondem a qual causa a unidade é exposta e a variável resposta observada Y_s , é muito importante e freqüentemente não está presente em discussões sobre causa. Essa distinção não surge em estudos de simples associação, mas é crucial para análise de causalidade.

Em um estudo experimental, S é construída pelo experimentador, enquanto em um estudo não controlado (observacional) é determinada por fatores além do controle do analista. Em qualquer caso, a característica crítica da noção de causa nesse modelo é que o valor de $S(u)$ para cada unidade ‘poderia ter sido diferente’. Estar sob o controle do experimentador, nesse contexto, significa que ele conhece a estrutura probabilística que indicou o tratamento $k \in K$ à unidade $u \in U$, tal como num processo de randomização em que as unidades são alocadas nos grupos de exposição com base em um mecanismo aleatório, independentemente de outras variáveis.

Em estudos observacionais, o analista procura resgatar as vantagens do processo de randomização, no qual se busca comparabilidade entre os tratamentos considerados, por meio da observação de co-variáveis relevantes (associadas ao tratamento ou à resposta) que podem ser controladas no momento do desenho do estudo ou, posteriormente, na análise. Busca-se então substituir a hipótese de independência por hipóteses menos rigorosas de independência ‘condicional’, onde o condicionamento é determinado por co-variáveis pré-exposição observadas. Assim, incorpora-se ao modelo estatístico geral para inferência causal um conjunto de q co-variáveis observáveis não afetadas pela exposição, isto é, pré-exposição.

A Tabela 1 sintetiza as informações observadas relevantes para se fazer inferência causal em um contexto geral, comumente disponíveis nos

estudos epidemiológicos. Inferir causalidade em um contexto real e complexo torna-se possível ou não dependendo da capacidade de se obter alguma informação sobre os pontos de interrogação desta tabela, isto é, sobre os valores não observados. O estabelecimento de hipóteses plausíveis, embora nem sempre passíveis de serem validadas pelos dados disponíveis, em conjunto com técnicas estatísticas apropriadas constituem importantes instrumentos para estimação ou conhecimento dos valores que teriam sido observados pelas unidades, tivessem elas recebido tratamento alternativo àquele que de fato receberam.

Tabela 1 – Exemplo de dados observáveis em um estudo de 2 tratamentos ($k=2$) em uma população de U unidades com q co-variáveis e 1 variável resposta (2 respostas potenciais)

Unidades	Co-variáveis X				S	Y	
	X_1	X_2	\dots	X_q		Y_t	Y_c
1	x_{11}	x_{12}	\dots	x_{1q}	t	$y_t(1)$?
2	x_{21}	x_{22}	\dots	x_{2q}	t	$y_t(2)$?
3	x_{31}	x_{32}	\dots	x_{3q}	c	?	$y_c(3)$
:	:	:		:	:	:	:
:	:	:		:	:	:	:
u	x_{u1}	x_{u2}	\dots	x_{uq}	c	?	$y_c(u)$
:	:	:		:	:	:	:
:	:	:		:	:	:	:
U	x_{U1}	x_{U2}	\dots	x_{Uq}	t	$y_t(U)$?

Essa estrutura permite visualizar que o processo de inferência de causalidade deve estar condicionado aos dados observados e ao padrão de observação dos dados de fato observados e dos dados perdidos ou não observados (*missing*). Os *missing* da Tabela 1, representados pelos pontos de interrogação, são inerentes à própria formulação do Modelo de Rubin, diferentemente de outros tipos de dados perdidos comumente encontrados na prática

devido a circunstâncias indesejáveis, tais como falta de resposta em um inquérito ou censura. Para fins de causalidade, a consideração de ambos os tipos de *missing* é fundamental.

Para fins de uniformização dos procedimentos estatísticos comumente utilizados para análise, pode ser conveniente a adoção de uma notação comum contida em Rosenbaum (1995). As U unidades sob investigação podem ser divididas em M estratos ou subpopulações com base em co-variáveis, isto é, com base em características mensuradas anteriormente à designação dos tratamentos. Assim, existem U_m unidades no estrato m , $m = 1, 2, \dots, M$, de modo que $U = \sum_m U_m$. Faz-se $S_m(u) = 1$ se a u -ésima unidade do estrato m recebe o tratamento t e $S_m(u) = 0$ se recebe o tratamento c . Escreve-se ainda n_m para o número de unidades tratadas no estrato m , de modo que $n_m = \sum_{u=1}^{U_m} S_m(u)$ e $0 \leq n_m \leq U_m$. E, finalmente, designe-se como \mathbf{S} o vetor coluna U -dimensional contendo os valores $S_m(u)$ para todas as unidades, isto é,

$$\mathbf{S} = \begin{bmatrix} S_1(1) \\ S_1(2) \\ \vdots \\ S_1(U_1) \\ S_2(1) \\ \vdots \\ S_M(U_m) \end{bmatrix} = \begin{bmatrix} \mathbf{S}_1 \\ \vdots \\ \mathbf{S}_M \end{bmatrix}, \text{ onde } \mathbf{S}_m = \begin{bmatrix} S_m(1) \\ \vdots \\ S_m(U_m) \end{bmatrix}$$

Esta notação compreende diversas situações. Se nenhuma co-variável é utilizada para estratificar a população, então existe um único estrato contendo todas as unidades ($M=1$). Se $U_m=2$ e $n_m=1$ para $m=1, 2, \dots, M$, então fica caracterizado um estudo pareado com base nas co-variáveis, isto é, cada um dos M pares apresenta uma unidade tratada e uma unidade controle. A situação onde $U_m \geq 2$ e $n_m=1$ para $m=1, 2, \dots, M$ é conhecida como pareamento com múltiplos controles, ou seja, existem M conjuntos de unidades, cada qual com uma unidade tratada e uma ou mais unidades controle.

Como já comentado, o efeito causal médio T é a diferença entre os dois valores esperados $E(Y_t)$ e $E(Y_c)$. Entretanto, os valores observados (S, Y_S) podem apenas fornecer informação sobre

$$E(Y_S | S = t) = E(Y_t | S = t) \text{ e } E(Y_S | S = c) = E(Y_c | S = c).$$

No entanto, é claro que $E(Y_t)$ e $E(Y_t | S = t)$ não são em geral iguais, bem como $E(Y_c)$ e $E(Y_c | S = c)$. Quando as unidades são designadas aleatoriamente às causas consideradas (t ou c), por meio de algum processo físico de randomização (tal como o lançamento de uma moeda), então a determinação de qual causa a unidade u é exposta é considerada estatisticamente independente de qualquer outra variável, incluindo Y_t e Y_c . Isso significa que se um processo de randomização é executado, então é plausível considerar S independente de Y_t e Y_c e de todas as outras variáveis sobre U . Essa é a hipótese de independência que, quando válida, valida as seguintes equações básicas:

$$E(Y_t) = E(Y_t | S = t) \text{ e } E(Y_c) = E(Y_c | S = c).$$

Logo,

$$\hat{T} = E(Y_S | S = t) - E(Y_S | S = c) = E(Y_t | S = t) - E(Y_c | S = c) = E(Y_t) - E(Y_c) = T$$

Podendo essa quantidade ser estimada a partir dos dados observados por

$$\sum_{u=1}^U Y_S(u) I_{(S(u)=t)} U_t - \sum_{u=1}^U Y_S(u) I_{(S(u)=c)} U_c,$$

onde U_t e U_c são os números de indivíduos alocados nos grupos tratado e controle, respectivamente, e I é uma variável indicadora de indivíduo tratado ou não.

Assim, o estimador $\hat{T} = E(Y_t | S = t) - E(Y_c | S = c)$, sob a hipótese de independência, é não viciado para T , o efeito causal médio, que é o parâmetro de interesse para a solução estatística. Esse estimador é de uso corrente em estudos de comparação de tratamentos e, não raro, não se encontram explicitadas as hipóteses adjacentes que viabilizam sua

avaliação. Holland (1986) chama este estimador de ‘efeito causal à primeira vista’, enquanto Smith & Sugden (1988) batizaram-no de ‘efeito causal aparente’.

Sendo T uma média, goza de todas as características e propriedades inerentes a esta estatística. Assim, se a variabilidade dos efeitos causais $Y_t(u) - Y_c(u)$ sobre todas as unidades u em U é grande, T pode não representar adequadamente o efeito causal de uma determinada unidade u_0 . Se existe interesse específico em u_0 , então T pode ser irrelevante independentemente de quão precisa tenha sido sua estimativa.

Uma outra hipótese que torna possível a inferência causal é a de ‘efeito constante’. Esta hipótese assume que o efeito de t é o mesmo para todas as unidades. Vale então a seguinte equação:

$$T = Y_t(u) - Y_c(u), \quad \forall u \in U.$$

Essa hipótese é conhecida pelo nome de ‘aditividade’ em modelos estatísticos para experimentos, uma vez que o tratamento t ‘adiciona’ uma quantidade constante T à resposta controle para cada unidade. Sob aditividade, o efeito causal é especialmente importante e fácil de estimar, pois o mesmo efeito vale para cada unidade da população. Em muitos casos em que não vale essa hipótese, já se mostrou que o efeito causal obtido no nível populacional é uma medida de pouco interesse. Por exemplo, em um estudo sobre a eficácia de uma droga nova onde se suspeita de importantes diferenças entre homens e mulheres, os efeitos causais subpopulacionais, para homens e para mulheres, é que são relevantes. Efeitos causais subpopulacionais não são potencialmente mais atraentes apenas porque estão condicionados a características observadas das unidades, mas também porque aditividade provavelmente vale mais neste nível que em um nível populacional (Rubin, 1990a). Eventualmente, havendo interesse em perseguir a validação da hipótese de efeito constante, comumente faz-se uso de transformações adequadas na variável resposta.

A hipótese de efeito constante pode ser parcialmente verificada dividindo-se U em n subpopulações U_1, U_2, \dots, U_n com base em co-variáveis. Para cada U_i , podemos estimar os efeitos causais médios subpopacionais T_1, T_2, \dots, T_n . Uma grande variabilidade entre os T_i é indicação de que a hipótese de efeito constante pode não ser válida. Caso contrário, essa hipótese pode ser plausível.

Existe uma relação entre as hipóteses de homogeneidade e efeito constante. Se vale a hipótese de homogeneidade, vale também a hipótese de efeito constante. Isso porque se, para duas unidades u_1 e u_2 , $Y_t(u_1) = Y_t(u_2)$ e $Y_c(u_1) = Y_c(u_2)$, então

$$Y_t(u_1) - Y_t(u_2) = 0 \text{ e } Y_c(u_1) - Y_c(u_2) = 0 \rightarrow Y_t(u_1) - Y_t(u_2) = Y_c(u_1) - Y_c(u_2) \therefore \\ Y_t(u_1) - Y_c(u_1) = Y_t(u_2) - Y_c(u_2) = T.$$

Se faz-se somente a hipótese de efeito constante, não se pode concluir que \hat{T} seja não viciado para o efeito causal médio T . Isso porque, sob a hipótese de efeito constante, pode-se escrever $Y_t(u) = T + Y_c(u)$ para todas as unidades. Aplicando-se o operador esperança condicionado a $S = t$ nesta última expressão, tem-se

$$E(Y_t | S = t) = T + E(Y_c | S = t).$$

Mas, $\hat{T} = E(Y_t | S = t) - E(Y_c | S = c)$. Logo,

$$\hat{T} = T + [E(Y_c | S = t) - E(Y_c | S = c)].$$

Sem a hipótese de independência, não há uma razão específica para que o termo entre colchetes acima seja igual a zero, o que seria necessário para que, somente com a hipótese de efeito constante, o estimador \hat{T} pudesse estimar sem viés o parâmetro de interesse T .

Já com a hipótese mais forte de homogeneidade de unidades, é fácil mostrar que o efeito causal aparente (\hat{T}) é não viesado para o efeito causal médio (T). De fato, sob esta hipótese, tem-se que $Y_k(u_1) = Y_k(u_2)$ para todo par de unidades u_1 e u_2 . Isto é, as respostas potenciais $Y_t(u)$ e $Y_c(u)$ são constantes para qualquer $u \in U$ (porém diferentes para a mesma uni-

dade, a menos que não haja efeito). Assim, $E(Y_t)$ e $E(Y_c)$ são também constantes para quaisquer combinações de unidades. Logo,

$$\hat{T} = E(Y_t | S = t) - E(Y_c | S = c) = E(Y_t) - E(Y_c) = T.$$

Hipótese de valor estável unidade-tratamento

Toda a estruturação do Modelo de Rubin, discutida até aqui e sintetizada na Tabela 1, só é adequada sob a condição que Rubin (1980) batizou de ‘hipótese de valor estável unidade-tratamento’ ou posteriormente, de forma mais simples, de ‘hipótese de estabilidade’ (Rubin, 1990b). Esta é simplesmente a hipótese *a priori* de que o valor de Y para a unidade u quando exposta ao tratamento t é o mesmo, independentemente de qual mecanismo é usado para designar o tratamento à unidade e de quais tratamentos as outras unidades recebem. Neste último aspecto, coincide com a hipótese de não interferência entre unidades, de Cox (1958).

Embora não seja absolutamente necessária para inferência causal, a hipótese de estabilidade é a hipótese mais simples sob a qual os efeitos causais para cada unidade, tais como $Y_t(u) - Y_c(u)$, podem ser definidos precisamente (Rubin, 1990a). Por ser uma hipótese muito forte, freqüentemente não é válida para muitas situações. Além disso, Dempster (1990) observa que a simplicidade dessa hipótese contrasta com a complexidade dos problemas de causalidade encontrados em situações reais. A hipótese de estabilidade é freqüentemente plausível em experimentos bem planejados. Em outras situações, entretanto, fazem-se necessárias considerações criteriosas antes de se legitimar causalidade com base nos efeitos observados. Quando a hipótese de estabilidade é violada, modelos especiais podem ser construídos a fim de acomodar novas condições, embora tais modelos normalmente assumam hipóteses não testáveis.

As formas mais comuns de violação da hipótese de valor estável unidade-tratamento acontecem quando não existe uniformidade na eficiência

para cada tratamento ou quando existe interferência entre as unidades. Para exemplificar o problema com a uniformidade na eficiência, suponha um estudo no qual se quer avaliar o efeito de uma droga, administrada oralmente, sobre a pressão sangüínea. O grupo tratado, então, é aquele que recebe o comprimido com a substância para a qual se supõe seja hipotensora, e o grupo controle recebe um placebo. Se, para o grupo tratado, os comprimidos variam em eficiência, isto é, se apresentam quantidades variadas da substância considerada e isso pode ter um efeito diferenciado, então o efeito causal $Y_t(u) - Y_c(u)$ não é precisamente definido, uma vez que seu valor depende especificamente de qual comprimido a unidade u recebe. A rigor, para satisfazer a hipótese de valor estável unidade-tratamento, cada comprimido deveria ser considerado um tratamento. Em respostas dicotômicas essa possível violação é menos sentida, na medida em que somente grandes variações nas ‘doses’ implicariam mudança na resposta. Um exemplo seria a avaliação do efeito do hábito de fumar sobre doença cardiovascular, onde no grupo de fumantes pode haver diferentes níveis de consumo de cigarros. Na maioria das situações práticas, procura-se garantir essa hipótese administrando-se ‘doses’ ou construindo-se grupos de tratamento tão homogêneos quanto possível. Entretanto, sua verificação normalmente não é viável com base nos dados observados.

A violação da hipótese de estabilidade devido à existência de interferência entre as unidades, por ser mais nitidamente percebida em situações reais, tem sido motivo de maior preocupação. Há interferência entre unidades quando a resposta à causa imputada (tratamento) para uma determinada unidade depende do tratamento designado a outra unidade. Logo, o efeito causal também não é definido, já que seu valor depende de unidades ‘vizinhas’. Em epidemiologia, a ocorrência de doenças contagiosas constitui bom exemplo de uma situação em que essa hipótese é claramente violada, já que estas doenças têm como característica fundamental

a presença de cadeias de transmissão de um hospedeiro para outro, de modo que um indivíduo (unidade) eventualmente torna-se infectado dependendo de contatos com outros já infectados.

Como conseqüência da violação da hipótese de estabilidade, faz-se necessária uma expansão da representação das respostas, tal como aquela da Tabela 1, adicionando-se tantas colunas de respostas potenciais quanto necessárias para se considerar a interferência entre as unidades (Rubin, 1990b). Assim, por exemplo, numa situação extremamente simples em que só exista interferência entre as unidades u_1 e u_2 , de tal forma que a resposta da unidade u_2 depende do tratamento que a unidade u_1 recebe, seriam necessárias quatro colunas para representar as respostas potenciais. Percebe-se imediatamente a dificuldade dessa abordagem para uma situação real envolvendo complexas cadeias de transmissão.

Halloran & Struchiner (1995), estudando a aplicação do Modelo de Rubin em programas de vacinação (intervenção) para doenças contagiosas, observam que a exposição à infecção (interferência) pode também ser vista como um tipo de causa ou manipulação. A exposição à infecção em doenças contagiosas dada pelos outros membros da população, diretamente ou via vetores,⁴ é fundamental não só para a transmissão como também para a avaliação dos efeitos da intervenção (vacinação). Entretanto, essa causa deve ser avaliada usando-se um diferente tipo de efeito causal, o efeito indireto. Este é, portanto, aquele efeito devido ao contato com outros indivíduos ou vetores de transmissão, diferente do efeito direto atribuível à vacinação. Assim, os autores discutem uma forma alternativa para solucionar o problema de interferência entre unidades, presente em doenças contagiosas, por meio do condicionamento da análise somente aos indivíduos expostos à infecção.

⁴ Um vetor pode ser entendido como qualquer ser vivo que veicule o agente infeccioso desde a fonte de infecção (indivíduo, macaco etc.) até o hospedeiro em potencial.

No desenvolvimento de sua análise, definem como E o indicador para exposição à infecção, onde $(E = +)$ representa exposição à infecção e $(E = -)$ representa não exposição à infecção. Os tratamentos (causas) em questão, t e c , são receber vacina e não receber vacina, e a resposta Y é também dicotômica, assumindo o valor 0 para não doente e 1 para doente. Considerando a notação já utilizada, as respostas potenciais para uma determinada unidade u quando exposta à infecção seriam, então, $(Y_t(u) | E = +)$ e $(Y_c(u) | E = +)$. O efeito causal direto condicional para a unidade u é a diferença entre essas duas respostas potenciais. Assim, analogamente a T , o efeito causal médio, os autores definem o efeito causal médio condicional à exposição à infecção como

$$T' = E(Y_t | E = +) - E(Y_c | E = +).$$

O termo $E(Y_c | E = +)$ é o quociente entre o número de indivíduos suscetíveis não vacinados que potencialmente tornam-se infectados, se expostos à infecção, e o número total de expostos. Pode, portanto, ser definido como a probabilidade de transmissão esperada em indivíduos suscetíveis não vacinados devido ao contato com indivíduos expostos já infectados. Vale a mesma interpretação para o termo $E(Y_t | E = +)$, substituindo-se não vacinado por vacinado. Se ninguém é exposto à infecção, então $(Y_c | E = -)$ e $(Y_t | E = -)$ são iguais a zero, de modo que o efeito causal médio na ausência de exposição à infecção será também zero.

Assim, o condicionamento do estudo aos indivíduos expostos à infecção tenta resgatar a hipótese de estabilidade. Isto é, condicionar o estudo aos expostos à infecção torna novamente adequada a representação das respostas potenciais em duas colunas: uma para a resposta ao tratamento ‘receber vacina’ e outra para a resposta ao controle ‘não receber vacina’. Entre as dificuldades dessa solução, Halloran & Struchiner (1995) assinalam que nem todos os indivíduos são expostos à infecção, de modo que a resposta potencial para um determinado indivíduo é dependente não só do mecanis-

mo de designação como também do potencial de exposição à infecção, caracterizando então uma abordagem contrafactual em dois níveis. Isto é, ‘se’ um indivíduo é exposto à infecção e ‘se’ é não vacinado, qual a sua resposta? Um outro problema está relacionado à impossibilidade de se distinguir, na prática, dois diferentes tipos de não exposição à infecção. Há aqueles indivíduos que, independentemente do tratamento recebido, não serão expostos à infecção, e há aqueles que não serão expostos sob particulares designações de tratamento. Por fim, dado que pessoas não expostas à infecção não são incluídas na análise, o número de indivíduos no estudo poderia ser muito reduzido, sobretudo se muitos expostos à infecção são vacinados.

Assim, inferência causal em doenças contagiosas, onde não vale a hipótese de estabilidade devido à presença de interferência entre as unidades, tem como dificuldade principal a identificação das fontes de exposição à infecção, comumente não possível em situações reais, de modo que se exigem hipóteses não testáveis específicas sobre as situações de exposição à infecção, tais como picadas de mosquitos, relações sexuais ou o número de contatos com indivíduos já infectados. Alternativamente, o estudo de intervenções em doenças contagiosas pode ser conduzido, substituindo-se as informações sobre a exposição à infecção por hipóteses parcialmente não testáveis que modelem a dinâmica da transmissão do agente infeccioso.

O Mecanismo de Designação de Tratamentos

Diferentemente de um modelo determinístico de causa que procura ‘demonstrar’ causalidade, tal como o modelo de Rothman, um modelo estatístico ou probabilístico procura ‘inferir’ causalidade. Esse modelo não necessariamente nega a visão de que a ocorrência de uma doença poderia ser perfeitamente determinada por certa quantidade de fatores, apesar de muitos

deles não serem conhecidos ou eventualmente não observados de forma apropriada. Assim, a incerteza presente nos modelos estatísticos serve sobretudo para expressar a ignorância sobre o processo causal e sobre como observá-lo na prática, e não porque se acredita ser a ocorrência da doença um fenômeno aleatório. O Modelo de Rubin, não obstante apresentar uma estrutura determinística, tem a sua discussão inserida em um contexto probabilístico em razão, principalmente, do Problema Fundamental da Inferência Causal.

É por meio do mecanismo que designa qual tratamento as unidades recebem que se identifica a natureza probabilística necessária para se tentar inferir causalidade. Mais precisamente, qualquer método estatístico para inferência causal exige especificações para os procedimentos que geram os dados observados. Isto é, faz-se necessário conhecer as estruturas probabilísticas de registro de dados por meio dos mecanismos de amostragem e de designação de tratamentos. Aqui, com o objetivo de explorar mais detalhadamente a aplicabilidade da idéia de causa contida no Modelo de Rubin, não será considerado nenhum esquema de amostragem. Assim, a aleatoriedade presente será toda devida ao mecanismo de designação de tratamentos, assumindo-se, portanto, que todas as U unidades participam do estudo. Entretanto, tem-se visto que o próprio mecanismo de designação de tratamentos pode refletir tanto o esquema de seleção das unidades de uma população finita quanto a indicação de qual tratamento cada unidade recebe (Rubin, 1978). Smith & Sugden (1988) separam estes dois mecanismos e diferenciam a análise pela ordem com que eles se sucedem: amostrando-se as unidades antes de designar o tratamento ou, alternativamente, designando-se os tratamentos às unidades e posteriormente amostrando-se unidades de cada estrato de tratamento.

Desse modo, para que se possa inferir causalidade, independentemente do procedimento estatístico a ser utilizado, é necessária a explicitação de um mecanismo de designação de tratamentos – um modelo probabilístico para o

processo pelo qual se obtêm os valores de Y que de fato são observados. Em outras palavras, é necessário especificar o processo que conduz aos valores observados de S . Comumente, para a situação dicotômica considerada, tem-se na prática imaginado as unidades recebendo um determinado tratamento, t por exemplo, com probabilidade $p(X)$ e, conseqüentemente, o tratamento alternativo com probabilidade $1 - p(X)$, onde $p(X)$ é uma função desconhecida das co-variáveis X . Assim, o mecanismo de designação de tratamentos pode ser escrito genericamente por meio da seguinte expressão probabilística, atribuindo-se os valores 0 e 1 para os tratamentos c e t , respectivamente:

$$\Pr(\mathbf{S} | \mathbf{X}) = \prod_{u=1}^U p(X_u)^{S(u)} [1 - p(X_u)]^{1-S(u)},$$

onde as variáveis em negrito indicam vetor ou matriz de dados para todas as unidades.

Essa expressão revela o mecanismo de designação como dependente de co-variáveis observáveis, a matriz $\mathbf{X}_{U \times q}$. Entretanto, um mecanismo de designação pode também ser dependente de variáveis não observáveis ou somente parcialmente observáveis, a matriz $U \times 2$ de respostas potenciais $\mathbf{Y} = (\mathbf{Y}_t, \mathbf{Y}_c)$. Seja W o conjunto de todas as outras co-variáveis pré-exposição, que diferentemente das co-variáveis X , são desconhecidas ou não passíveis de observação. Desse modo, o mecanismo de designação de tratamentos deve ser concebido de forma geral como dependente dessas três quantidades aleatórias. Simbolicamente, $\Pr(\mathbf{S} | \mathbf{X}, \mathbf{Y}, \mathbf{W})$.

Um mecanismo é dito ‘controlado’ quando está livre da dependência de \mathbf{Y} e \mathbf{W} (Smith & Sugden, 1988). Sem considerar a dependência sobre \mathbf{W} , Rubin (1990a, 1991) classifica alguns mecanismos de designação de tratamentos de acordo com suas propriedades. Assim, uma classe muito importante é aquela constituída dos mecanismos ditos ‘não confundidos’ (o neologismo é inevitável). Estes são os mecanismos que independem da variável resposta \mathbf{Y} , isto é, $\Pr(\mathbf{S} | \mathbf{X}, \mathbf{Y}) = \Pr(\mathbf{S} | \mathbf{X})$ para todos os valores possíveis

de \mathbf{S} , \mathbf{X} e \mathbf{Y} . A dependência de um mecanismo sobre \mathbf{X} pode usualmente ser tratada diretamente na análise, enquanto a dependência sobre \mathbf{Y} , em especial sobre sua parte não observada, em geral é difícil de ser considerada de maneira satisfatória. Daí a importância dessa classe de mecanismos.

Uma importante característica da especificação de um mecanismo de designação é que ele deve ser ‘probabilístico’, no sentido de que cada unidade tenha uma probabilidade de designação dos tratamentos t ou c satisfazendo a desigualdade

$$0 < \Pr(S(u) = t \mid \mathbf{X}, \mathbf{Y}) < 1, \quad \forall u \in U.$$

Se alguma unidade não tem chance de ser designada para algum tratamento, é sensato considerar o porquê e se a inclusão de tal unidade é de fato adequada ao estudo planejado (Rubin, 1991).

Um mecanismo de designação é dito ‘ignorável’ para os valores observados de X , Y e S se

$$\Pr(\mathbf{S} \mid \mathbf{X}, \mathbf{Y}) = \Pr(\mathbf{S} \mid \mathbf{X}, \mathbf{Y}_t, \mathbf{Y}_c) = \Pr(\mathbf{S} \mid \mathbf{X}, \mathbf{Y}_c).$$

Isto é, o mecanismo pode ser escrito como dependente somente dos dados observados.

Sob a condição de ser ‘não confundido’ e ‘probabilístico’, isto é, se

$$(Y_t, Y_c) \perp S \mid X \quad \text{e} \quad 0 < \Pr(S(u) = t \mid X) < 1,$$

onde o símbolo \perp significa independência, o mecanismo é conhecido como ‘fortemente ignorável’. Observa-se que se o mecanismo é ‘fortemente ignorável’, é também ‘ignorável’, mas a recíproca não é verdadeira (Rosembaum & Rubin, 1983). Essa classe de mecanismos de designação de tratamentos também está dentro da classe dos mecanismos ditos ‘controlados’.

Para inferência causal, o mecanismo de designação de tratamentos assume uma função vital. Pode-se dizer que o ceticismo que acompanha a atribuição de causa com base em estudos observacionais surge das potenciais dificuldades de se colocar um mecanismo de designação aceitável. Isto é, a principal fonte de incerteza na análise de um estudo observacional não é

geralmente o método de inferência empregado, mas sim a especificação do mecanismo de designação (Rubin, 1991). Embora na maioria das vezes esteja completamente ausente, este tipo de discussão deve ser um importante componente da análise de qualquer estudo observacional para efeitos causais.

A importância do mecanismo de designação para inferência causal pode ser ilustrada pela seguinte situação hipotética, baseada em um exemplo de Smith & Sugden (1988). Considere um estudo observacional do tipo coorte, em que indivíduos suscetíveis a determinada doença são acompanhados por certo período de tempo e alocados nos diferentes níveis de exposição (causas) sem qualquer controle sobre este mecanismo. Os dados observados sobre casos incidentes de hepatite A diante das duas causas alternativas postuladas, presença ou não de água tratada no domicílio, para as três mil crianças existentes em certa região e acompanhadas por um ano estão sumariados na Tabela 2.

Tabela 2 – Casos de hepatite A em um ano, segundo condição da água do domicílio (dados hipotéticos)

Tratamentos ou causas	Total acompanhado	Resposta (Y_s)	
		Casos (1)	Não casos (0)
Com água tratada (f)	499	165	334
Sem água tratada (c)	2.501	577	1.924
TOTAL	3.000	742	2.258

Os dados mostram um risco significativamente maior para aqueles que dispõem de água tratada no domicílio ($165/499 = 0,3307$) em relação àqueles que não possuem ($577/2501 = 0,2307$), com um efeito causal aparente ($\hat{\tau}$) de $0,3307 - 0,2307 = 0,1$. Como já dito, em epidemiologia, tal medida é conhecida como ‘risco atribuível’. O risco atribuível é usado para quantificar o risco da doença no grupo exposto (no exemplo, com água

tratada) que pode ser considerado atribuível à exposição, uma vez que remove o risco de doença presente de forma geral devido a outras causas (o risco no grupo não exposto). Assim, a interpretação do risco atribuível é dependente da suposição de que existe uma relação de causa e efeito entre exposição e doença.

Uma conclusão precipitada com base no risco atribuível obtido com base nos dados da Tabela 2 indicaria que possuir água tratada no domicílio, ao contrário de um conhecimento já consolidado, é fator causal para hepatite A. O número de casos a mais de hepatite A entre aqueles que possuem água tratada atribuível a esta condição é de 1 para cada 10 crianças. Entretanto, o quão aproximada essa estimativa está do efeito causal verdadeiro, desprezada a variação amostral, pode depender de co-variáveis observáveis ou não e também da própria resposta. Como já observado, em um modelo estatístico de causalidade, essa dependência é tratada por meio do mecanismo de designação de tratamentos. Inferir causalidade com base no efeito causal aparente sem considerar essas possíveis dependências pode levar a conclusões enganadoras, pois este efeito pode ser bem diferente do efeito causal verdadeiro. Mesmo mecanismos de designação de tratamentos controlados, que dependem apenas de co-variáveis observáveis, podem produzir grandes diferenças. Se uma co-variável X é tal que para valores grandes desta variável mais provavelmente se observa a resposta 1 para o tratamento t e a resposta 0 para o tratamento c , e se a alocação dos tratamentos é obtida inadvertidamente sem a consideração de X , então os efeitos causais aparente e verdadeiro serão bem distantes. O condicionamento do mecanismo de designação em X produziria um efeito causal aparente ajustado, de modo que o efeito causal verdadeiro poderia ser estimado sem viés.

Para o exemplo considerado, a matriz de respostas potenciais $\mathbf{Y}_{U \times 2} = (\mathbf{Y}_t, \mathbf{Y}_c)$ é uma matriz onde cada $Y_k(u)$, $k = \{t, c\}$ e $u = 1, 2, \dots, U$ assume o valor 0 ou 1. Assim, os U pares de respostas potenciais

$(Y_t(u), Y_c(u))$ podem tomar os valores (0,0), (0,1), (1,0) e (1,1). Se as frequências desses pares são hipoteticamente conhecidas e dadas pela Tabela 3, a seguir, então o efeito causal médio verdadeiro seria

$$T = E(Y_t) - E(Y_c) = \frac{300 + 100}{3000} - \frac{600 + 100}{3000} = \frac{300 - 600}{3000} = -0,1.$$

Tabela 3 – Exemplo de um mecanismo de designação não ignorável e frequência hipotética dos pares de respostas potenciais

Respostas potenciais (Y_t, Y_c)	Frequência	$\Pr(S(u)=t)$
(0,0)	2.000	0,11
(0,1)	600	0,19
(1,0)	300	0,52
(1,1)	100	0,09
TOTAL	3.000	

Suponha que o mecanismo não seja ignorável e que a probabilidade de designação do tratamento t a uma unidade dependa do par de respostas potenciais, como indicado pela mesma Tabela 3. Sob este mecanismo, as frequências de casos e não casos de hepatite A pelos níveis de exposição considerados que se espera observar podem ser calculadas. A resposta 1 (caso) é observada, ou quando t é aplicado com probabilidade 0,52 ao par de respostas potenciais (1,0), cuja frequência é de 300 crianças, ou com probabilidade 0,09 ao par (1,1), que tem frequência 100. Isto é, a frequência esperada de casos sob o tratamento t é $(300 \times 0,52) + (100 \times 0,09) = 165$. Da mesma forma, a frequência de não casos sob o tratamento c (sem água tratada), que se esperaria observar se a verdade fosse aquela representada pela Tabela 3, seria de $2000 \times (1 - 0,11) + 300 \times (1 - 0,52) = 1924$. As outras duas frequências esperadas, casos sob o tratamento c e não casos sob o tratamento t , podem ser obtidas de forma análoga, cujos resultados são 577 e 334, respectivamente.

É curioso observar que as frequências esperadas coincidem exatamente com os valores observados pelo estudo e apresentados na Tabela 2.

A finalidade desse exemplo é ilustrar quão perigosa pode ser a interpretação de resultados se baseada em estudos observacionais. Se, eventualmente, por um alguma razão o mecanismo que alocou as unidades aos tratamentos não é ignorável e este fato não é do conhecimento do analista, as estimativas resultantes sofrerão erros sistemáticos, podendo inclusive gerar uma medida de efeito de igual magnitude mas de sentido contrário, como é o caso desse exemplo, perdendo, conseqüentemente, sua interpretação causal.

O mecanismo de designação de tratamentos é o elemento, então, que assume papel dominante em uma formulação estatística de causalidade. Dependendo do tipo de estudo empregado, mecanismos particulares podem ser usados a fim de subsidiar conclusões de causalidade. Classicamente, randomização em estudos experimentais (Rubin, 1974) e, mais recentemente, o escore de propensão em estudos observacionais (Rosembaun & Rubin, 1983) têm sido advogados como mecanismos apropriados para se inferir causalidade.

Randomização

Randomização é um tema cuja literatura disponível é bastante extensa e data o início do século XX, com os trabalhos de Fisher. Apesar disso, a discussão de suas características e propriedades continua sendo extremamente atual. Formalmente, randomização proporciona um mecanismo com base no qual é possível a obtenção das propriedades probabilísticas das estimativas. As principais propriedades são:

- o efeito causal aparente $\hat{\tau}$ é uma estimativa não viesada de T , o efeito causal verdadeiro. Esta propriedade já fora observada quando se comentou sobre a hipótese de independência;

- declarações probabilísticas bem definidas podem ser feitas para indicar quão incomuns valores observados de $\hat{\tau}$ seriam sob um específico efeito causal hipotetizado. P-valores, intervalos de confiança e razões de verossimilhança para parâmetros causais são computados sob a hipótese de que as indicações alternativas de tratamento são respostas igualmente prováveis. Esta hipótese é perfeitamente justificada sob randomização.

Greenland (1990) tem dito que inferir estatisticamente causalidade com base em intervalos de confiança, p-valores ou razões de verossimilhança é de pouco significado se o mecanismo de designação à exposição é desconhecido ou sabidamente não randomizado. Um mecanismo randomizado proporciona, então, o elo entre inferência estatística e parâmetros causais. De forma análoga, amostras aleatórias fornecem a chave para o relacionamento entre inferência estatística e parâmetros populacionais.

Um importante fato enfatizado por Rubin (1978) é que randomização, quando empregada, garante independência *a priori* entre o mecanismo de designação de tratamentos e diversos fatores ou atributos das unidades, incluindo aqueles eventualmente também causais. Assim, randomização funciona como um fiador da validade da representação matemática de medidas de efeito de fatores causais observados, permitindo modelar os efeitos de fatores não observados como erros aleatórios (Dempster, 1990).

A propriedade que tem o estimador considerado – no caso de se adotar um mecanismo randomizado – de não ser tendencioso não deve trazer muito conforto para o epidemiologista. O conceito estatístico de viés e o conceito epidemiológico de confundimento, embora possam ser relacionados, são bem diferentes. Viés estatístico se refere a uma média ponderada diferente de zero das estimativas dos efeitos causais, com pesos dados pela distribuição de probabilidade de tais estimativas. Confundimento é visto como falta de comparabilidade dos grupos considerados, sendo, por-

tanto, característica de uma particular alocação. Porém, no caso artificial de se quantificar o grau de confundimento de cada possível alocação como a diferença entre a estimativa e o parâmetro causal de interesse, pode-se concluir que a não tendenciosidade de experimentos randomizados corresponde a um confundimento médio igual a zero sobre a distribuição dos resultados das possíveis alocações (Greenland, 1990). O epidemiologista, ao tentar interpretar o resultado de um estudo, está interessado no grau de confundimento de uma particular estimativa observada, cuja magnitude não é diretamente obtida por um estudo randomizado. Entretanto, sob randomização pode-se fazer a probabilidade de um confundimento grave, tal como interpretado anteriormente, tão pequena quanto se queira, aumentando-se o tamanho das coortes sob os tratamentos considerados. No entanto, esse recurso, quase sempre uma grande dificuldade em situações reais, não é uma garantia absoluta de que não há confundimento severo em um particular estudo. É ainda possível que um ‘infeliz’ resultado tenha ocorrido e, assim, o grau de confundimento seja grande. Desse modo, se as coortes são grandes e não há evidências de falta de comparabilidade entre elas, o processo de randomização proporciona um mecanismo em que se deposita alta credibilidade na hipótese de que as coortes são aproximadamente comparáveis, desde que não haja nenhuma violação grosseira do protocolo de designação dos tratamentos às unidades.

Ainda dentro dessa lógica, atualmente faz-se uma interpretação mais atraente de um ensaio randomizado. Supõe-se que randomização garanta que os grupos sob as condições de exposição consideradas são comparáveis, onde o sentido de comparabilidade é que o resultado observado para o grupo exposto teria sido o mesmo se o grupo não exposto tivesse sido exposto, e vice-versa. Isso porque, como já dito aqui, supõe-se assegurada a hipótese de que as respostas são independentes da designação dos tratamentos e de que co-variáveis não observadas são balanceadas nos dois grupos. Em ou-

tras palavras, sob randomização, não importa qual dos dois grupos de indivíduos será exposto ao fator causal considerado.

Entretanto, dependendo de características específicas de um estudo, nem mesmo a randomização garante estimativas não tendenciosas do efeito biológico de interesse. Struchiner & Halloran (1996), estudando os efeitos de vacinas para doenças contagiosas, mostraram como a exposição à infecção pode ser um fator de confundimento para o efeito causal de interesse mesmo em ensaios randomizados. Exposição à infecção, tal como picada de mosquito ou contato sexual, é vista como uma co-variável freqüentemente não registrada nos estudos que se dá exclusivamente pela ação da natureza (Halloran & Struchiner, 1995).

Sintetizando, um mecanismo de designação construído com base em algum processo randomizado, tal como o lançamento de uma moeda, onde a indicação de um tratamento se daria pela ocorrência de cara e, obviamente, a ocorrência de coroa implicaria indicação do tratamento alternativo, pode ser escrito como

$$\Pr(\mathbf{S}) = \left(\frac{1}{2}\right)^U,$$

sendo as 2^U possíveis indicações igualmente prováveis, proporcionando portanto um mecanismo independente das variáveis X e Y e, mais interessante que estas, também de W , as co-variáveis não observadas. Daí a credibilidade dada a esse mecanismo quando o interesse é atribuir causa. Para o Modelo de Rubin, a validade desta última expressão é suficiente para inferência causal.

Escore de propensão

Como visto, sob randomização, os dois grupos sob tratamento podem ser diretamente comparáveis porque provavelmente suas unidades são similares. Sem esse mecanismo, comparações diretas podem ser enganadoras,

pois as unidades expostas a um tratamento podem, em geral, diferir sistematicamente das unidades expostas ao outro tratamento. Nessas condições, escores balanceadores podem ser usados para que a comparação entre os grupos considerados seja reveladora do efeito que se quer estimar.

Um escore balanceador, $b(X)$, é uma função das co-variáveis observadas X , tal que a distribuição condicional de X dado $b(X)$ é a mesma para o grupo tratado ($S = t$) e para o grupo controle ($S = c$). Isto é, dado $b(X)$, S e X são independentes. Simbolicamente,

$$S \perp X \mid b(X).$$

O escore balanceador mais trivial é obviamente $b(X) = X$, e outros de maior interesse são funções mais complexas de X .

Rosenbaun & Rubin (1983) mostraram que se um mecanismo de designação é fortemente ignorável dado X , então a diferença entre as médias dos grupos tratamento e controle para cada valor de um escore balanceador é uma estimativa não viesada do efeito do tratamento naquele valor e, conseqüentemente, pareamento ou estratificação sobre este escore devem produzir estimativas não viesadas do efeito causal médio verdadeiro.

Assim, um outro mecanismo de designação, com indicação para utilização em estudos observacionais, é obtido com base em um particular escore balanceador conhecido como ‘escore de propensão’. Este é definido como a probabilidade condicional de designação de um particular tratamento dado um vetor de co-variáveis observadas. O interessante desse mecanismo é que o ajustamento para um escore de propensão escalar é suficiente para remover vieses, em razão de todas as co-variáveis observadas. A seleção de unidades tratadas e controles com o mesmo valor para esse escore terão a mesma distribuição de X .

Em experimentos randomizados, o escore de propensão é uma função conhecida, havendo portanto uma especificação apropriada para $\Pr(S = t \mid X)$. No entanto, em estudos observacionais, o escore de propen-

são é quase sempre uma função desconhecida, de modo que uma especificação para ele deve ser estimada com base em dados observados por meio de algum modelo adequado, como, por exemplo, o modelo logístico. Para um bayesiano, essas estimativas são probabilidades preditoras *a posteriori* de designação do tratamento t para uma unidade com vetor X de co-variáveis. Sob um modelo logístico, fazendo $t = 1$ e $c = 0$, o escore de propensão $\Pr(S = t | X)$ é dado por

$$\Pr(S = t | X) = \frac{e^{\beta_0 + \sum_{i=1}^q \beta_i x_i}}{1 + e^{\beta_0 + \sum_{i=1}^q \beta_i x_i}}$$

onde os β_i são coeficientes estimados com base nos dados observados S e X .

Já foi utilizada a notação para a probabilidade de uma determinada unidade receber um tratamento, t por exemplo, como $p(X)$. Assim, $p(X) = \Pr(S = t | X)$ e, atribuindo-se novamente os valores 1 e 0 para os tratamentos t e c , respectivamente, tem-se o seguinte mecanismo de designação, também já apresentado, construído com base no escore de propensão:

$$\Pr(\mathbf{S} | \mathbf{X}) = \prod_{u=1}^U p(X_u)^{S(u)} [1 - p(X_u)]^{1-S(u)}$$

Uma interessante utilização do escore de propensão acontece quando se criam pares de unidades tratamento-controle. Para cada unidade que recebe um tratamento, busca-se outra para ser seu controle que tenha um escore de propensão aproximadamente igual. Com isso, inferências sobre o efeito causal devido ao tratamento podem ser obtidas livres dos efeitos das variáveis usadas para construir o escore. Esse tipo de uso do escore de propensão em estudos observacionais é análogo à randomização em experimentos controlados pareados. Dentro de cada par, tratado e controle têm aproximadamente a mesma probabilidade preditora, dado X , de ser tratado. Randomização em um experimento de comparação pareado é mais interes-

sante, é claro, na medida em que sustenta a hipótese de não dependência de X bem como de W , o conjunto de co-variáveis não observadas.

Como ilustração das potencialidades desse mecanismo, considere-se a análise de causalidade desenvolvida por Rubin (1991) em que se busca estimar os possíveis efeitos em virtude da entrada no mercado de uma droga psiquiátrica com seu nome genérico em concorrência a seu nome comercial. O grupo tratado ($S=t$) foi constituído por aqueles pacientes que optaram pela troca, enquanto o grupo controle ($S=c$) consistiu daqueles que mantiveram o uso da droga com seu nome comercial. Foram observadas 53 co-variáveis pré-intervenção entre variáveis demográficas e relacionadas ao uso da droga, cujo controle pensou-se ser conveniente para que os efeitos pudessem ser medidos e atribuídos à causa em questão e não a uma possível diferença nestas co-variáveis entre os grupos. Obteve-se assim o escore de propensão $\Pr(S = t | X)$ para cada unidade, por meio de um modelo logístico para as co-variáveis X ou transformações delas. Com base nesse escore para cada unidade tratada, buscou-se uma unidade controle com o valor mais próximo. A conseqüência desse processo de pareamento foi sintetizada em uma tabela de freqüências, tal como a Tabela 4. Esta tabela conta a freqüência, por faixas de magnitude, de uma medida de viés padronizada para as 53 co-variáveis envolvidas, ou eventualmente utilizando-se transformações apropriadas, antes e depois do pareamento. Essa medida de viés foi obtida por meio da seguinte expressão:

$$\frac{\bar{x}_d}{\sqrt{\frac{(s_t^2 + s_c^2)}{2}}}$$

onde \bar{x}_d foi a diferença entre as médias dos grupos tratado e controle para cada variável (ou transformação), antes e após o pareamento, e s_t^2 e s_c^2 foram as variâncias de cada grupo antes do pareamento.

Nota-se que antes do pareamento havia duas co-variáveis com viés acima de 0,25 e 10 (quase 19%) com viés de pelo menos 0,10. Após o pareamento, houve grande concentração das variáveis com medida de viés inferior a 0,05, o que levou Rubin (1991) a considerar o procedimento bem efetivo. Isto é, os dois grupos sob investigação puderam ser considerados homogêneos, pelo menos com relação às co-variáveis observadas.

Tabela 4 – Distribuição de freqüência de uma medida de viés padronizada para 53 co-variáveis antes e após pareamento

Viés Padronizado	Antes		Após	
	N	%	N	%
Até 0,05	30	56,6	49	92,5
0,05 - 0,10	13	24,5	4	7,5
0,10 - 0,15	3	5,7	0	0,0
0,15 - 0,20	5	9,4	0	0,0
0,20 - 0,25	0	0,0	0	0,0
0,25 ou +	2	3,8	0	0,0
TOTAL	53	100,0	53	100,0

Admitindo-se esse procedimento como adequado, pode-se adotar um mecanismo de designação mais simplificado. Se, entre as U unidades presentes no estudo, encontramos $M \leq U/2$ pares, de tal forma que as unidades de cada par tratamento-controle apresentem a mesma probabilidade de serem tratadas, então o mecanismo de designação reduz-se a

$$\Pr(\mathbf{S}) = \left(\frac{1}{2}\right)^M,$$

se as unidades de cada par recebem tratamentos diferentes. Assim, cada um dos 2^M possíveis valores de \mathbf{S} que designam t a uma unidade do par e c à outra, são igualmente prováveis, e qualquer \mathbf{S} que designe as duas unidades do par o mesmo tratamento tem probabilidade 0.

Resumindo, as limitações e propriedades de um mecanismo construído com base no escore de propensão são bem conhecidas. Pareamento sobre esse escore balanceia as co-variáveis observadas X . Entretanto, diferentemente da randomização, não balanceia co-variáveis não observadas, exceto aquelas correlacionas com X . Outra questão importante é que ajustamento sobre esse escore balanceia X apenas em média. Em um particular estudo, mesmo após um ajustamento, é possível ainda encontrar por mero acaso algumas variáveis não balanceadas entre os grupos sob investigação. Daí a necessidade, quase sempre presente, de se estabelecerem hipóteses adicionais, eventualmente não testáveis, a fim de validar uma atribuição de causalidade.

A abordagem mais comumente encontrada para controle de co-variáveis eventualmente confundidoras tem se dado por meio de análises multivariadas, incorporando a noção de modelagem estatística. No entanto, a contribuição dos modelos estatísticos para inferência causal, com os quais se tenta descrever as medidas de efeito causal de interesse por meio dos coeficientes do modelo, tem sido apresentada e criticada em muitos trabalhos (Greenland, 1979; Robins & Greenland, 1986; Vandenbrouke, 1987; Greenland, 1989b). Alternativamente, apesar de só mais recentemente se encontrarem na literatura clínica e epidemiológica trabalhos que utilizem o escore de propensão como um instrumento para controle simultâneo de muitas co-variáveis observadas, Rosenbaum e Rubin têm, em dupla ou individualmente, defendido em vários trabalhos este mecanismo como apropriado para a discussão de causalidade com base em estudos observacionais. D'Agostino (1998) apresenta um excelente tutorial desse procedimento.

Inferência Estatística

Como já mencionado, o objetivo, quando se analisam os dados disponíveis, é obter alguma informação sobre os pontos de interrogação ilustrados na Tabela 1 e, assim, poder estimar efeitos causais tais como T para as unidades sob investigação. Repetindo, as interrogações se referem aos valores de Y que seriam observados se, contrariamente ao fato, as unidades tivessem recebido o tratamento alternativo àquele que receberam. Entretanto, é objetivo também estender a informação desses efeitos para unidades em que nenhum dos elementos do par de respostas potenciais (Y_t, Y_c) tenha sido observado, tais como as outras unidades da população das quais se supõe ter extraído aquelas que de fato estão sob análise, ou mesmo para unidades que futuramente possam ser expostas aos tratamentos. Em outras palavras, quer-se ‘inferir’ estatisticamente efeitos causais. Rubin (1990a, 1991) tem assinalado e comparado quatro métodos de inferência estatística, apontando a diferença entre eles como consequência da forma com que usam os dados observados para obter informação sobre os valores não observados, os quais, como se tem visto, são necessários para que se definam os efeitos causais. Mais importante que as diferentes formas de construção desses métodos é que todos compartilham a mesma estrutura conceitual de causalidade presente no Modelo de Rubin.

Os quatro métodos considerados por Rubin (1990a) como conceitualmente distintos são:

- teste de hipóteses nulas pontuais sob um mecanismo randomizado;
- inferência de parâmetros causais baseada em possíveis alocações aleatórias de tratamentos;
- inferência bayesiana para efeitos causais;
- inferência freqüentista.

Apesar do Problema Fundamental da Inferência Causal, talvez a demonstração mais simples de que é possível a estimação de efeitos causais seja possível por meio do estabelecimento de uma hipótese nula pontual tal como

$$H_0: Y_t(u) = Y_c(u), \quad \forall u \in U.$$

sob essa hipótese, \mathbf{Y} , a matriz de respostas potenciais $U \times 2$ é de fato completamente revelada, uma vez que ou $Y_t(u)$ ou $Y_c(u)$ é observado, e $Y_t(u) = Y_c(u)$. Sendo Y_s o valor observado de Y e definindo-se \mathbf{Y}_{H_0} como o valor revelado sob H_0 , tem-se $\mathbf{Y}_{H_0} = (\mathbf{Y}_S; \mathbf{Y}_S)$. Dessa forma, são especificados valores conhecidos para cada ponto de interrogação da Tabela 1. Em outras palavras, H_0 em conjunto com os dados observados implica valor específico para cada interrogação e, assim, todos os valores e todos os efeitos causais passam a ser conhecidos.

Segundo o desenvolvimento de Rubin (1990a), dois fatos importantes se seguem. Primeiro, os valores das probabilidades das possíveis designações de tratamentos, especificada pelo mecanismo de designação $\Pr(\mathbf{S} | \mathbf{X}, \mathbf{Y})$, podem ser obtidos não só para a particular designação S observada como também para qualquer outra designação possível, dados $\mathbf{X} = \mathbf{X}_{\text{obs}}$ e $\mathbf{Y} = \mathbf{Y}_{H_0}$. Em segundo lugar, os valores do vetor $\mathbf{Y}_{U \times 1} = \{Y_{S(u)}(u)\}$ podem ser encontrados para cada possível S sob H_0 . Em particular, $\mathbf{Y}_{H_0} = \mathbf{Y}_S$. Assim, o valor de qualquer estatística $Z = Z(S, X_{\text{obs}}, Y_{H_0})$ pode ser calculada não apenas para a específica designação de tratamentos observada S , mas também para todas as outras possíveis. Como resultado, escolhendo-se uma direção para valores incomuns de Z , vale o seguinte cálculo:

$$\begin{aligned} & \Pr(Z \text{ mais extremo que } Z_{\text{obs}} | \mathbf{X} = \mathbf{X}_{\text{obs}}, \mathbf{Y} = \mathbf{Y}_{H_0}) = \\ & = \sum_S \Pr(S | \mathbf{X} = \mathbf{X}_{\text{obs}}, \mathbf{Y} = \mathbf{Y}_{H_0}) \times \delta(Z \text{ mais extremo que } Z_{\text{obs}}), \end{aligned}$$

onde d é uma função indicadora, assumindo o valor 1 quando o argumento for verdadeiro e 0 quando for falso. Essa expressão fornece o nível de

significância (ou p-valor) para H_0 baseado na estatística de teste Z . Sob um mecanismo probabilístico não confundido corresponde ao teste randomizado de Fisher. Estes p-valores podem também ser calculados com mecanismo confundidos ou mesmo não probabilísticos, embora a interpretação dos resultados seja pouco interessante (Rubin, 1990a).

Já se viu que se o p-valor é pequeno, então ou um evento raro aconteceu, ou a hipótese nula não é verdadeira, ou ainda o mecanismo de designação colocado está errado. Dessa maneira, a qualquer hipótese nula pontual pode ser atribuído um p-valor, onde pontual significa que H_0 em conjunto com os valores observados Y_s , implica o conhecimento de todos os valores de Y . Esse método de inferência tem como grande vantagem sua simplicidade e imparcialidade, que são particularmente interessantes quando Z e sua direção de raridade são definidos antes que qualquer dado seja observado.

Outro aspecto interessante desse método se refere ao rico debate entre Neyman e Fisher, cuja origem pode ser atribuída à especificação da hipótese nula. Enquanto Fisher defendia uma hipótese nula tal como a apresentada acima, ou seja, de ausência de efeito em cada unidade, Neyman propunha como hipótese nula a igualdade entre os efeitos médios populacionais, isto é, $H_0: E(Y_t) = E(Y_c)$. Rubin (1990b) discute maiores detalhes sobre as seqüências da abordagem de Neyman.

Estreitamente relacionado a esse método por também funcionar sob aleatoriedade, distingue-se outro método para se inferir causalidade em que, entretanto, hipóteses nulas pontuais não assumem mais um papel central. A abordagem se concentra na inferência de parâmetros causais sob a distribuição de um estimador causal obtida com base nas possíveis e aleatórias designações dos tratamentos. Por ser mais associada à idéia de levantamentos amostrais aleatórios, Rubin (1990a) assinala que essa abordagem deve ser vista mais como uma avaliação da qualidade de procedimentos de

inferência propostos, tais como erro quadrático médio e poder, do que propriamente um método para inferência por si.

Inicialmente define-se um parâmetro causal, $T = E(Y_t - Y_c)$ por exemplo, podendo ser inclusive função de X , o que caracterizaria um parâmetro causal subpopulacional. Em seguida, busca-se um estimador $\hat{T} = \hat{T}(S, X, Y)$ tal que seja aproximadamente não tendencioso para T , isto é,

$$E(\hat{T} | X, Y) \cong T,$$

onde a aleatoriedade é devida a S com distribuição dada por $\Pr(S | X, Y)$. Em geral, a esperança de uma estatística depende de (\mathbf{X}, \mathbf{Y}) . Assim, a determinação de \hat{T} que satisfaça a esperança acima é fácil com mecanismos de designação probabilísticos não confundidos, tal como randomização, mas em mecanismos diferentes pode ser difícil ou mesmo impossível sem a definição de fortes hipóteses *a priori* (Rubin, 1990a). É necessária, ainda, a determinação de um estimador $\hat{V} = \hat{V}(S, X, Y)$ para a variância de \hat{T} que seja aproximadamente não tendencioso e que tenha variabilidade menor que a de \hat{T} . Assim, assumindo-se normalidade da distribuição de $T - \hat{T}$, inferências para T são obtidas a partir da declaração probabilística

$$(T - \hat{T}) \sim N(0; \hat{V}).$$

A escolha dos estimadores é geralmente baseada em princípios tais como variância mínima, não tendenciosidade e menor erro quadrático médio. Apesar de reconhecer que esse método proporciona resultados coincidentes com os do método anterior, Rubin (1990a) aponta a dificuldade dessa abordagem em se adequar a problemas mais complicados e reais, como por exemplo distribuições não normais. Resumindo, esse método de inferência trata \mathbf{X} e \mathbf{Y} como fixados mas \mathbf{S} como desconhecido, e obtém as características dos estimadores sob todas as possíveis designações de tratamentos dadas pelo mecanismo $\Pr(\mathbf{S} | \mathbf{X}, \mathbf{Y})$.

O método bayesiano proporciona abordagem mais direta para se inferirem efeitos causais, na medida em que se concentra na distribuição

a posteriori de T . Essa é, pois, a distribuição do parâmetro causal de interesse condicionada a valores observados e especificações de modelos probabilísticos para as variáveis envolvidas. Isto é, inferência bayesiana de efeitos causais é obtida com base na colocação de uma distribuição de probabilidade *a priori* para \mathbf{Y} dado \mathbf{X} e da especificação probabilística do mecanismo de designação de tratamentos. Enquanto no método de hipóteses nulas pontuais os pontos de interrogação da Tabela 1 eram especificados com base nos dados observados em conjunto com H_0 , no método bayesiano busca-se a distribuição deles dados todos os valores observados. Busca-se, portanto, a distribuição *a posteriori* de $\mathbf{Y}_{\bar{S}}$, a componente não observada de \mathbf{Y} . A distribuição de $\mathbf{Y}_{\bar{S}}$ pode ser escrita como

$$\Pr(\mathbf{Y}_{\bar{S}} | \mathbf{X}, \mathbf{S}, \mathbf{Y}_S) = \frac{\Pr(\mathbf{S} | \mathbf{X}, \mathbf{Y}) \Pr(\mathbf{Y} | \mathbf{X})}{\int \Pr(\mathbf{S} | \mathbf{X}, \mathbf{Y}) \Pr(\mathbf{Y} | \mathbf{X}) d\mathbf{Y}_{\bar{S}}}$$

onde $\mathbf{Y} = (\mathbf{Y}_S, \mathbf{Y}_{\bar{S}})$ ‘particiona’ \mathbf{Y} em valores observados e não observados. E, no caso de um mecanismo de designação de tratamentos não confundido, isto é, se

$$\Pr(\mathbf{S} | \mathbf{X}, \mathbf{Y}) = \Pr(\mathbf{S} | \mathbf{X}),$$

então a distribuição de $\mathbf{Y}_{\bar{S}}$ é simplificada, sendo dada por

$$\Pr(\mathbf{Y}_{\bar{S}} | \mathbf{X}, \mathbf{Y}_S) = \frac{\Pr(\mathbf{Y} | \mathbf{X})}{\int \Pr(\mathbf{Y} | \mathbf{X}) d\mathbf{Y}_{\bar{S}}}.$$

Os valores observados de $(\mathbf{X}, \mathbf{S}, \mathbf{Y}_S)$ em conjunto com a distribuição *a posteriori* de $\mathbf{Y}_{\bar{S}}$ implica distribuição para qualquer função de $(\mathbf{X}, \mathbf{S}, \mathbf{Y})$ tais como os efeitos causais médios populacional ou subpopulacional, ou mesmo efeitos causais individuais. Assim, dentro de uma estrutura bayesiana, inferências causais podem ser obtidas por exemplo a partir de $\Pr(T | \mathbf{X}, \mathbf{S}, \mathbf{Y}_S)$. Próprio desse método é a necessidade de se estabelecer um modelo para a distribuição de \mathbf{Y} dado \mathbf{X} . Usualmente apela-se para o teorema construído por de Finetti e escreve-se

$$\Pr(\mathbf{Y} | \mathbf{X}) = \int \prod_{u=1}^U f(Y_u | X_u, \theta) \Pr(\theta) d\theta$$

onde Y_u é a u -ésima linha de \mathbf{Y} e $f(Y_u | X_u, \theta)$ é o modelo comum para a distribuição condicional de \mathbf{Y} dado \mathbf{X} para cada linha de (\mathbf{Y}, \mathbf{X}) dado o parâmetro não observável θ com distribuição *a priori* $\Pr(\theta)$. Não faz muito tempo, a grande dificuldade dessa abordagem se concentrava na identificação de distribuições conjugadas, de modo a se obter um procedimento analiticamente tratável. Entretanto, essa barreira tem sido vencida com a incorporação de procedimentos de simulação.

O quarto método de inferência citado por Rubin (1990a; 1991) domina a prática. O método freqüentista é similar ao método de inferências de parâmetros causais na sua formulação, mas em vez de suas características serem obtidas sob possíveis \mathbf{S} do mecanismo de designação $\Pr(\mathbf{S} | \mathbf{X}, \mathbf{Y})$ para um fixado (\mathbf{X}, \mathbf{Y}) , são obtidas sob repetidas observações de \mathbf{Y} a partir da distribuição amostral $\Pr(\mathbf{Y} | \mathbf{X}; \theta)$, dada por

$$\Pr(\mathbf{Y} | \mathbf{X}; \theta) = \prod_{u=1}^U f(Y_u | X_u; \theta)$$

com θ constante mas desconhecido. Essa abordagem proporciona resultados estatísticos matemáticos bem conhecidos para θ , especialmente quando f segue uma distribuição normal.

Rubin (1991) critica esse método porque, apesar de formular quase todas as hipóteses que o método bayesiano formula, não permite uma especificação *a priori* para θ . Do ponto de vista prático, uma especificação para $\Pr(\theta)$ é muito menos sensível do que qualquer especificação para $f(Y_u | X_u, \theta)$, principalmente em estudos observacionais. Assim, por exemplo, enquanto na teoria bayesiana intervalos de confiança para θ têm uma interpretação probabilística imediata, dentro do método freqüentista esta interpretação depende do argumento condicional de que amostras repetidas seriam necessárias para verificar a freqüência relativa de intervalos que de fato compreendiam θ .

Uma outra forma, mais genérica, de se apresentar conceitualmente a questão da inferência é encontrada em Smith & Sugden (1988), escrevendo-se a função de verossimilhança das variáveis envolvidas. O processo de inferência causal envolve basicamente a consideração de três conjuntos de variáveis: as respostas Y , as co-variáveis conhecidas X e as desconhecidas W . Formalmente, pode-se escrever parametricamente de forma conveniente a distribuição conjunta dessas variáveis como:

$$f(\mathbf{W} | \mathbf{Y}, \mathbf{X}; \psi)g(\mathbf{Y} | \mathbf{X}; \theta)h(\mathbf{X}; \phi),$$

onde se assume que os parâmetros ψ , θ e ϕ são distintos. Tomando-se também de forma geral o mecanismo de designação como $\Pr(S | \mathbf{X}, \mathbf{Y}, \mathbf{W})$ e admitindo-se que todas as unidades estão presentes no estudo, a função de verossimilhança pode ser escrita como

$$h(\mathbf{X}; \phi) \int_W \int_S \Pr(S | \mathbf{Y}, \mathbf{X}, \mathbf{W}) f(\mathbf{W} | \mathbf{Y}, \mathbf{X}; \psi) g(\mathbf{Y} | \mathbf{X}; \theta) d\mathbf{Y} d\mathbf{W}.$$

E as condições que implicariam um mecanismo que sustentasse uma interpretação causal seriam:

$$S \perp \mathbf{Y} | \mathbf{X}, \mathbf{W} \text{ e } \mathbf{Y} \perp \mathbf{W} | \mathbf{X}.$$

Diante dessas considerações sobre os métodos de inferência estatística, percebe-se que os recursos que a estatística tradicionalmente utiliza para fazer inferências sobre associações podem ser transferidos para o caso em que se estuda causalidade com base na adoção explícita de um modelo especialmente planejado para este fim.