

Questões epidemiológicas relacionadas

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.

Questões Epidemiológicas Relacionadas

Até aqui se tem visto como a estatística pode contribuir para a discussão sobre inferência causal, procurando explorar as propriedades e limitações de um modelo estatístico de causalidade devido a Rubin. É interessante, também, discutir as interfaces dessa contribuição com a teoria e a prática epidemiológica, abordando algumas questões metodológicas relacionadas.

A importância do instrumental estatístico no desenvolvimento da teoria epidemiológica tem cada vez mais ficado evidente. Embora construída sobre o tripé clínica, medicina social e estatística, a epidemiologia tem experimentado nos últimos anos um *boom* de reconhecimento graças, em parte, à utilização e ao aperfeiçoamento de técnicas quantitativas avançadas especialmente desenvolvidas para responder à complexidade dos problemas por ela enfrentados. Assim, a incorporação de um modelo estatístico de causalidade parece ser relevante por ser esta, a causalidade, o grande desafio da investigação epidemiológica.

Pode-se entender inferência epidemiológica como o processo de obtenção de inferências, tais como a predição de ocorrências de doenças ou a identificação de suas causas, com base em dados epidemiológicos, isto é, dados relacionados à ocorrência de doenças em populações. Essas inferências devem poder ser feitas sem os benefícios diretos de evidências experimentais e também sem a necessidade de uma teoria estabelecida sobre a etiologia

da doença, ou seja, sem o estabelecimento de um mecanismo causal. Cabe aqui um parêntese, já que o termo ‘população’ em epidemiologia merece atenção especial. Apesar de os epidemiologistas observarem indivíduos, as interpretações dos resultados são baseadas na combinação de dados de muitas unidades. Razões para o uso de populações em epidemiologia estão relacionadas a dois grandes objetivos: avaliar programas de melhora do *status* de saúde de grupos específicos e, mais interessante que isso, permitir que uma investigação faça inferências causais, usando métodos estatísticos, sobre as relações entre certas exposições e o *status* de saúde.

Como exemplo de que inferências epidemiológicas devem ser possíveis mesmo sem os benefícios de um estudo experimental, considere como problema a verificação da hipótese de que cafeína é fator causal para doença cardiovascular. Essa hipótese pode ser testada por meio da estimação do risco atribuível a esse fator, que, como já visto, é obtido pela diferença entre o risco de doença cardiovascular entre indivíduos usuários dessa substância e o risco entre indivíduos não usuários. Entretanto, diferentemente de uma exposição experimental, o uso de cafeína é de opção pessoal e tem sido apontada sua correlação com diversas outras preferências, principalmente o fumo. Portanto, mesmo que cafeína não tenha nenhum efeito sobre o risco de doença cardiovascular, não se esperaria encontrar uma equivalência de riscos entre usuários e não usuários de cafeína. Assim, a análise deve ser conduzida necessariamente através de um estudo observacional que seja capaz de corrigir eventuais diferenças entre os grupos considerados. É essa tentativa de correção que tem estimulado muito não só o aparecimento de novas técnicas estatísticas como também o estabelecimento de critérios de diagnóstico e controle dessas técnicas, de modo a validar ou não as hipóteses a elas adjacentes.

Um desafio dos estudos epidemiológicos é, portanto, a estimação isolada do risco atribuível (ou de qualquer outra medida de efeito pertinente)

a uma determinada exposição – no exemplo, uso de cafeína –, livre dos efeitos devidos a outros fatores, de modo que as eventuais conclusões de causalidade sejam válidas. No entanto, nesse momento surge uma outra característica fundamental da pesquisa epidemiológica. Eventualmente, a estimação independente de uma medida de efeito de determinado fator pode ser de pouca valia se existe um outro fator que age sinergicamente com ele. Existe sinergismo quando a presença de um fator modifica o efeito biológico do outro e um caso particular acontece quando os indivíduos têm a doença somente se expostos aos dois fatores, mas não a um deles isoladamente. Sinergismo – conceito da biologia – é, na prática, comumente avaliado por meio do conceito estatístico de interação. Diz-se que existe interação estatística entre dois fatores, A e B, se é necessário um parâmetro adicional para descrever adequadamente o risco de doença em virtude da exposição conjunta a esses fatores. Em oposição, dois fatores podem ser considerados estatisticamente independentes se o risco de doença, dada uma exposição conjunta a esses fatores, pode ser adequadamente modelado como uma função dos efeitos separados, onde esta função pode ser aditiva, multiplicativa ou, ainda, assumir outra forma. Assim, a determinação se existe ou não interação estatística depende do modelo estatístico adotado (Siemiatycki & Thomas, 1981).

Desse modo, ao lado do confundimento, a interação de fatores, entendida, portanto, como a interdependência entre dois fatores que alteram a magnitude de um dado efeito, é uma outra importante preocupação metodológica. Entretanto, o interesse principal em uma discussão de causalidade presente nos estudos epidemiológicos é o julgamento da validade da associação observada entre exposição e doença.

O Conceito de Validade

Validade pode ser entendida como a capacidade de um particular estudo epidemiológico encontrar critérios lógicos que apontem para ausência de vieses. Assim, um estudo de associação entre uma exposição e uma doença é considerado válido, e com conseqüente interpretação causal, se está livre de vieses que possam, alternativamente à exposição proposta, explicar seus resultados.

Os vieses possíveis em um particular estudo podem ser devidos a várias fontes, e a maior parte deles é comum a qualquer desenho epidemiológico. Tradicionalmente, como já dito aqui, os vieses têm sido classificados como viés de informação, viés de seleção e confundimento. Greenland (1991) propõe uma classificação mais abrangente, que, embora não tenha ecoado na literatura, parece representar uma estrutura mais conveniente para discutir a validade dos estudos e o processo de inferência epidemiológica. Em sua abordagem, qualquer estudo epidemiológico cujo objetivo seja a avaliação de causalidade deve considerar as seguintes condições: validade de comparação, validade de seguimento, validade de especificação e validade de mensuração. Em estudos de caso-controle, especificamente, além de serem mais sujeitos a fontes adicionais de erros de mensuração, deve-se considerar também a validade de seleção de casos e controles. Percebe-se imediatamente a complexidade que envolve a validade de determinada medida de efeito, visto que é necessário o cumprimento de uma série de condições suficientes. Essas quatro condições são vistas como condições de validade ‘internas’, pois se referem apenas à estimação de efeitos dentro dos grupos de exposição considerados, sem preocupação quanto à generalização dos resultados para outras populações. São condições suficientes para validade, mas não necessárias, no sentido de que uma eventual violação pode não produzir viés.

Pode-se acreditar que haja validade de comparação se as unidades que de fato foram expostas, caso não o tivessem sido, teriam apresentado aproximadamente a mesma distribuição para a variável resposta (proporção de doença) que as unidades não expostas. Vale o mesmo raciocínio para as unidades não expostas: tivessem elas sido expostas, se esperaria observar o mesmo resultado que o das unidades de fato expostas. Validade de comparação simplesmente significa que a distribuição da variável resposta para as unidades de um determinado grupo de tratamento (exposição) prediz o que teria acontecido com o grupo de tratamento alternativo, tivessem as unidades deste último grupo experimentado a condição do outro. Ou ainda, uma outra forma de dizer isso é afirmar que os grupos tratado (exposto) e controle (não exposto) são ‘comparáveis’ ou ‘permutáveis’ com respeito à variável resposta, independentemente da situação de exposição que eventualmente cada unidade experimente. Falta de validade de comparação deve resultar em uma estimativa viesada da medida de efeito causal considerada e diz-se que há ‘confundimento’ na estimativa. Como se trata da principal preocupação da investigação de causa, essa questão é retomada adiante em um item específico, no qual é discutida mais detalhadamente, procurando-se interfaces entre a forma com que a epidemiologia trata a questão e o Modelo de Rubin.

Validade de seguimento acontece quando, dentro de cada grupo de tratamento considerado (tratado e controle), o risco de censura⁵ não está associado ao risco de doença. Desse modo, dada a validade de seguimento, espera-se que o risco de doença dentro de cada grupo em um determinado momento seja o mesmo tanto para unidades perdidas (ou retiradas do estudo por razões concorrentes) até este momento quanto para unidades acompanhadas além deste momento. No exemplo do uso de cafeína como fator

⁵ Censura diz respeito à finalização do período de seguimento por uma razão diferente daquela que se esperaria observar, no caso, ocorrência da doença em questão.

causal para doença cardiovascular, poder-se-ia imaginar um viés de seguimento, já que tabagismo, que está associado à exposição em questão, relaciona-se também a uma taxa maior de mortalidade por doença cardiovascular e também por outras doenças. Ou seja, indivíduos expostos à cafeína estariam potencialmente mais sujeitos à censura – por causas de morte relacionadas ao tabagismo – do que os indivíduos não expostos.

Quando se discutiram os mecanismos de designação de tratamentos, particularmente o escore de propensão, verificou-se a necessidade de se estabelecer um modelo estatístico. Mais genericamente, as técnicas estatísticas assumem algum tipo de modelo para o processo de geração dos dados a serem analisados. E, na ausência de randomização ou amostras aleatórias, raramente será possível identificar um modelo absolutamente correto. Logo, algum tipo de erro de especificação deve ser esperado. No caso de respostas dicotômicas, o modelo amostral tradicionalmente empregado é o binomial. Um modelo amostral deve ser entendido como uma expressão matemática que descreve a probabilidade de observação dos possíveis dados como função de parâmetros, quase sempre desconhecidos. Em um estudo de coorte, por exemplo, a validade de um modelo binomial depende da hipótese de que os riscos dentro de cada grupo de exposição permanecem constantes por todo o período de acompanhamento. Mais especificamente, o risco de doença cardiovascular não é constante ao longo do tempo para as coortes acompanhadas uma vez que, à medida que os indivíduos vão envelhecendo, tal risco vai aumentando. Logo, a utilização do modelo binomial, nesse exemplo, não gozaria de validade de especificação.

Além da fonte de erro devido ao modelo amostral, falta de validade de especificação pode ser devida também a um modelo estrutural, isto é, a uma expressão matemática que descreve os parâmetros do modelo amostral como função de co-variáveis, tal como um modelo logístico. Associado a qualquer modelo estrutural encontra-se sempre um conjunto de hipóteses que para

serem validadas são comumente confrontadas com os dados observados. Na literatura, é comum encontrar o modelo estrutural incorporado ao modelo amostral, sendo essa combinação designada genericamente como ‘modelo estatístico’. Assim, de uma estimativa de uma medida de efeito pode-se dizer que tem validade de especificação se é obtida com base em um modelo estatístico correto, ou pelo menos aproximadamente correto. E se o modelo amostral ou o modelo estrutural usado na análise for incorreto, as estimativas resultantes podem estar viesadas. Como comentário adicional, mesmo quando um eventual erro de especificação não conduza a um viés, testes estatísticos ou intervalos de confiança para as estimativas ficam comprometidos (Greenland, 1991).

Já foi ressaltado que a contribuição da estatística para o processo de inferir causalidade passa necessariamente pela mensuração de variáveis que permitam a quantificação dos efeitos causais. Desse modo, medir adequadamente as variáveis envolvidas é condição *sine qua non* para uma possível interpretação de causa. Entretanto, do ponto de vista prático ou operacional, qualquer estudo está sujeito a erros de mensuração por várias fontes, o que obviamente pode contribuir para um possível viés na estimativa. Assim, pode-se dizer que uma estimativa de uma determinada medida de efeito goza de validade de mensuração se ela não sofre de vieses em consequência de erros de mensuração nas variáveis envolvidas no estudo. Para ajudar na tarefa de identificar possíveis erros de mensuração, pode ser útil classificá-los de acordo com suas fontes. E erros de uma determinada fonte podem ser ainda classificados de acordo com características que são preditoras da direção do viés que eles produzem.

Greenland (1991) propõe um esquema de classificação para os erros de mensuração bastante abrangente, dividindo-os quanto à fonte, quanto ao tipo e quanto à dependência entre eles. Quanto à fonte, os erros podem ser classificados como erros de procedimento – tais como instrumentos mal

calibrados ou falta de lembrança em estudos retrospectivos, erros de uma variável *proxy*, ou seja, erros por se utilizar uma variável substituta à variável de interesse – e erros de construção, que surgem de ambigüidades na definição das variáveis. Independentemente da fonte, os erros podem ser classificados como diferenciais ou não diferenciais, se a direção ou a magnitude dos erros tende a variar ou não com os valores verdadeiros de outras variáveis, respectivamente. Finalmente, diz-se que erros na mensuração de duas variáveis são dependentes se a magnitude ou a direção do erro cometido na mensuração de uma variável está associada à magnitude ou à direção do erro cometido em uma outra variável. Se não existe associação entre os erros, diz-se que são independentes. Questão importante suscitada por essa classificação é uma maior ‘valorização’ da validade de mensuração. Se suspeita-se que os ‘instrumentos’ utilizados para medir as variáveis envolvidas não medem adequadamente aquilo que se pensa que estejam medindo, então pouca credibilidade deveria ser dada a qualquer resultado eventualmente obtido. Por exemplo, conceitos epidemiológicos tais como ‘sensibilidade’ e ‘especificidade’ dizem respeito ao poder, ou às probabilidades, que tem um teste diagnóstico de classificar como doente ou não doente indivíduos que de fato estejam doentes ou não doentes, respectivamente. Logo, conhecer a magnitude dessas probabilidades seria necessário para, pelo menos, identificar o sentido do erro.

Diante dessas considerações, em que se registra uma série de condicionantes acerca da validade de um estudo, parece justificável a insegurança sempre presente quando o assunto é estabelecimento de causa. O conhecimento dos possíveis vieses envolvidos em um particular estudo é, portanto, fundamental para que, sempre que possível, possam ser verificados a partir dos dados disponíveis. Fazendo uma associação com a teoria contida no Modelo de Rubin, a discussão de alguns possíveis vieses fica embutida nas hipóteses relacionadas ao modelo, que precisariam, portanto, ser testadas. De novo, nem sempre as hipóteses adjacentes podem ser postas à prova.

Confundimento e Permutabilidade

Confundimento tem sido reconhecido como um dos principais problemas da pesquisa epidemiológica e já foi aqui citado diversas vezes. Em razão disso, muito se tem escrito e já há bastante tempo sobre esse problema metodológico, registrando-se um rico debate não só sobre abordagens alternativas para seu controle, mas também sobre o seu próprio conceito. Do ponto de vista epidemiológico, vê-se confundimento como um problema surgido a partir de uma diferença intrínseca aos riscos de doença entre as populações exposta e não exposta (Greenland & Robins, 1986). Isso significa que a diferença existiria mesmo se a exposição estivesse ausente em ambas as populações. Retomando o exemplo em que a exposição sob consideração é o uso de cafeína, esperar-se-ia encontrar riscos para doença cardiovascular diferentes entre usuários e não usuários de cafeína, mesmo se nenhum daqueles indivíduos usuários tivesse de fato sido exposto à cafeína. Essa diferença poderia ser explicada pela falta de comparabilidade, ou confundimento, devida à associação entre cafeína e tabagismo. O conceito de confundimento está baseado, portanto, na idéia de comparabilidade.

Entretanto, parece que o próprio conceito do problema ainda é motivo de discussão, sendo o termo usado comumente para se referir a três problemas distintos. A forma mais antiga se relaciona a um tipo de viés na estimação de um efeito causal, utilizado na literatura epidemiológica e sociológica. Um conceito mais recente associa confundimento à idéia de ‘não colapsibilidade’, isto é, uma certa medida de efeito bruta é diferente daquela obtida pela combinação desta medida com estratos de uma terceira variável. Um terceiro conceito está relacionado à impossibilidade de separação do efeito principal e o efeito de interação, presente comumente em uma análise de variância. É difícil distinguir adequadamente esses conceitos. Em particular, os conceitos de confundimento como um viés na estimação do efeito e como ‘não

colapsibilidade' são freqüentemente tratados como idênticos. Greenland et al. (1999) abordam essa questão e explicitam essas diferenças. Independentemente do conceito, o importante é que, para se fazer inferência causal, é necessária a garantia de comparabilidade entre os grupos considerados.

Greenland (1989a), ao comentar o trabalho de Wickramaratne & Holford (1987) sobre confundimento em estudos epidemiológicos, identifica três ambientes apropriados para essa discussão. O primeiro corresponde exatamente ao Modelo de Rubin, o qual, segundo ele, é a melhor formalização atualmente disponível para se tratar do problema de confundimento. Como foi possível observar aqui, essa abordagem explora a noção de 'não ignorabilidade' da designação dos tratamentos. Um segundo ambiente, imaginado por Greenland & Robins (1986), trata confundimento sob o conceito de permutabilidade. Essa questão é retomada adiante por apresentar forte ligação com o Modelo de Rubin. A terceira abordagem, devida a Gail (1986), trata confundimento sob a noção de ajustamento para co-variáveis balanceadas.

Greenland & Robins (1986) discutem o conceito de confundimento epidemiológico baseados em um modelo desenvolvido para se estudar efeitos individuais. Sob esse modelo, percebem confundimento como um problema de não identificabilidade de parâmetros. Isto é, valores distintos de um parâmetro de interesse desconhecido podem gerar a mesma distribuição de dados observados. Associam também a esse conceito a idéia de permutabilidade, cujo significado é aquele próprio à teoria bayesiana, qual seja: são esperados os mesmos resultados caso as condições de exposição dos grupos sejam permutadas (Migon & Gamerman, 1999). Esse aspecto corresponde, na realidade, à hipótese de homogeneidade desenvolvida no Modelo de Rubin. Aliás, como poderá ser observado, este modelo de efeitos individuais assemelha-se bastante ao Modelo de Rubin, visto, porém, sob uma ótica exclusivamente epidemiológica.

Fazendo-se uma reprodução simplificada de Greenland & Robins (1986), considere-se a situação na qual se deseja estudar o efeito de um fator de exposição dicotômico sobre o risco de uma doença em um período de tempo a risco especificado. Existem quatro tipos de indivíduos de acordo com suas possíveis respostas aos tratamentos considerados. Assim, um indivíduo classificado como tipo 1 é aquele que se torna doente independentemente da exposição, e o tipo 4 é aquele imune à doença. Os indivíduos tipo 2 ou 3 são os suscetíveis. A Tabela 5 ilustra esse fato, onde o valor 1 é indicativo de presença de doença e o valor 0, de ausência. Após observar um único indivíduo do grupo tratado sobre o período de tempo a risco e verificar que ele contraiu a doença, não é possível dizer, sem uma informação adicional, se ele é do tipo 1 – aquele indivíduo condenado a adoecer sob qualquer condição – ou do tipo 2, aquele indivíduo suscetível à doença pela exposição. Da mesma forma, se ele não tivesse contraído a doença também não seria possível dizer a qual tipo pertenceria: ao 3 ou ao 4. Ou seja, independentemente do que é observado, não é possível dizer se a exposição tem efeito. Esse é um problema de não identificabilidade.

Tabela 5 – Tipos de indivíduos quanto a suas possíveis respostas (1 ou 0) a um tratamento dicotômico

Tipo e descrição dos indivíduos	Tratamento		Tipo de efeito
	Tratado	Controle	
Tipo 1: Indivíduo condenado a adoecer	1	1	Sem efeito
Tipo 2: Indivíduo suscetível	1	0	Causal
Tipo 3: Indivíduo suscetível	0	1	Preventivo
Tipo 4: Indivíduo imune à doença	0	0	Sem efeito

Nota: um efeito preventivo é também um efeito causal.

Se, adicionalmente ao indivíduo exposto, observa-se um indivíduo não exposto, quatro possíveis resultados se apresentam:

- a) ambos os indivíduos adoecem;
- b) somente o indivíduo exposto adoecer;
- c) somente o não exposto adoecer;
- d) nenhum dos dois adoecer.

Pela mesma razão exposta, não é possível saber se a exposição tem efeito. Entretanto, se é razoável supor que os dois indivíduos são do mesmo tipo, então torna-se possível deduzir se a exposição tem ou não algum efeito. Isso porque, na hipótese de que os indivíduos são do mesmo tipo, as respostas (a) e (d) significam que a exposição não tem efeito. Do mesmo modo, as respostas (b) e (c), quando combinadas com a hipótese, implicam causalidade. Assim, a combinação da resposta observada com a hipótese de equivalência dos indivíduos implica identificabilidade do efeito. Essa equivalência pode ser colocada em termos de permutabilidade dos indivíduos. Se as condições de exposição dos dois indivíduos tivessem sido trocadas (permutadas), o mesmo resultado seria obtido.

Para completar a apresentação da noção de confundimento por meio do modelo de efeitos individuais de Greenland & Robins (1986), é necessária a consideração de uma população de indivíduos. Considere uma coorte de U_t indivíduos expostos a um tratamento t , inicialmente livres de doença, acompanhados ao longo de um período a risco. Seja p_j , $j = 1, 2, 3$ e 4 , a proporção de indivíduos dessa coorte que são do tipo j , de acordo com a Tabela 5. Considere também uma coorte de U_c indivíduos controles (ou não expostos) e equivalentes proporções de indivíduos do tipo j dadas por q_j . Ao final do tempo considerado, os resultados poderiam ser resumidos em uma tabela 2x2 tal como a Tabela 6, a seguir.

Tabela 6 – Frequências observadas de doentes, como função dos tipos de indivíduos, segundo coortes a risco

Grupos de exposição	Total acompanhado	Frequência	
		Doentes	Não Doentes
Tratado (t)	U_t	$D_t = (p_1 + p_2) \times U_t$	$ND_t = (p_3 + p_4) \times U_t$
Controle (c)	U_c	$D_c = (q_1 + q_3) \times U_c$	$ND_c = (q_2 + q_4) \times U_c$

Para atribuição de causa, seria necessário conhecer no grupo tratado as proporções indicadoras de existência de efeito, ou seja, os valores p_2 e p_3 , embutidos nas frequências observadas D_t e ND_t . O risco atribuível à exposição (\hat{T}), obtido pela diferença entre os riscos nas populações expostas (\hat{R}_t) e não expostas (\hat{R}_c) e que tem representado a medida de efeito causal de interesse, é dado por

$$\hat{T} = \hat{R}_t - \hat{R}_c = \frac{D_t}{U_t} - \frac{D_c}{U_c}$$

Só com o valor calculado de \hat{T} não é possível dizer se efetivamente há algum efeito, pois para qualquer valor de \hat{T} , positivo, negativo ou zero, não se pode identificar as proporções p_2 ou p_3 . Fazendo-se, entretanto, a hipótese de comparabilidade entre as coortes, isto é, assumindo-se que a proporção de indivíduos que adoeceriam se a exposição estivesse ausente é a mesma para ambas as coortes, seria obtida apenas uma identificabilidade parcial. Isso porque, sob essa condição, vale a equação $q_1 + q_3 = p_1 + p_3$. Logo,

$$\hat{T} = \frac{D_t}{U_t} - \frac{D_c}{U_c} = (p_1 + p_2) - (q_1 + q_3) = (p_1 + p_2) - (p_1 + p_3) = p_2 - p_3$$

Então, se $\hat{T} > 0$ implica que p_2 também maior que zero, o que significa que no estudo houve alguns indivíduos que tiveram a doença ‘por causa’ da exposição. Da mesma forma, $\hat{T} < 0$ significa que alguns tiveram a doença ‘prevenida’ pela exposição, isto é, $p_3 > 0$. Porém, se $\hat{T} = 0$, só se pode deduzir que $p_2 = p_3$. Para se concluir que não houve efeito, e assim atingir completa

identificabilidade, é necessária a hipótese adicional de que p_3 (ou p_2 , alternativamente) seja igual a zero, isto é, assumir que a exposição nunca previne. Assim, sob essa hipótese adicional, $\hat{\tau}=0$ implica $p_2=0$, indicando portanto não ter havido efeito. A hipótese de que $q_1 + q_3 = p_1 + p_3$ pode ser vista também como uma hipótese de permutabilidade parcial, e dir-se-ia que se as condições de exposição fossem trocadas, o valor observado do risco na ausência da exposição teria sido o mesmo. Para uma completa permutabilidade seria necessária a hipótese adicional de que $q_1 + q_2 = p_1 + p_2$, ou seja, existiria a mesma relação entre exposição e risco se as condições de exposição fossem permutadas.

A finalidade desta seção foi apresentar a discussão de confundimento tal como ela tem estado presente em um contexto epidemiológico. Diante das características desse modelo determinístico de efeitos individuais, usado por Greenland & Robins (1986) para discutir confundimento, uma analogia com o Modelo de Rubin parece inevitável. Nesse nível individual, a falta de identificabilidade dos efeitos, remediada pela hipótese de permutabilidade, corresponde ao Problema Fundamental da Inferência Causal, cuja solução correspondente seria a hipótese de homogeneidade. Também pode ser percebida, embutida nos quatro tipos de indivíduos citados, a noção de respostas potenciais, própria do Modelo de Rubin. Cada combinação possível de exposição e resposta potencial no Modelo de Rubin corresponderia a um dos quatro tipos de indivíduos no modelo de efeitos individuais.

Do ponto de vista da formulação do Modelo de Rubin, a discussão sobre confundimento está concentrada no mecanismo de designação de tratamentos. E, como já visto, existe uma classe de mecanismos ditos ‘não confundidos’, que são aqueles que não dependem da variável resposta \mathbf{Y} , isto é, $\Pr(\mathbf{S} \mid \mathbf{X}, \mathbf{Y}) = \Pr(\mathbf{S} \mid \mathbf{X})$ para todos os valores possíveis de \mathbf{S} , \mathbf{X} e \mathbf{Y} , sendo a dependência sobre \mathbf{X} passível de ser controlada na medida em que são covariáveis observadas. Se adiciona-se a condição $0 < \Pr(S(u) = t \mid X) < 1$,

viu-se que o mecanismo é conhecido como sendo ‘fortemente ignorável’. Assim, o Modelo de Rubin trata o problema de confundimento por meio da idéia de ‘ignorabilidade’.

Os Estudos Epidemiológicos

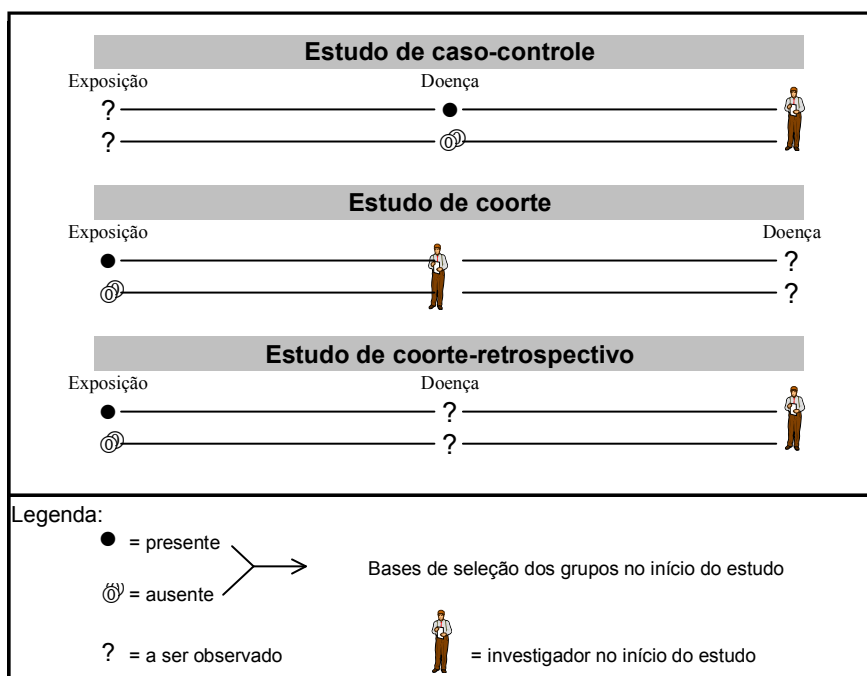
De forma geral, os estudos epidemiológicos podem ser divididos entre aqueles experimentais e aqueles observacionais, dependendo do controle de que o investigador dispõe sobre o mecanismo que designa a condição de exposição das unidades. Entre os observacionais, destacam-se os estudos de coorte, os estudos de caso-controle e os estudos de coorte-retrospectivo, de acordo com a forma com que são coletadas as unidades e o *timing* de observação.⁶ A Figura 2, adiante, ilustra esse conceito. Embora a epidemiologia disponha de outros tipos de estudos, tais como os estudos ecológicos, ou ainda particulares desenhos dentro dos citados, essa classificação é suficiente para se explorar de forma geral a noção de causa contida nos estudos epidemiológicos.

Em um estudo experimental, que no ambiente epidemiológico é conhecido como um ensaio clínico ou um estudo de intervenção, a designação dos tratamentos às unidades é controlada pelo experimentador. Isto é, os tratamentos são designados para as unidades com base em um mecanismo casual, comumente um gerador de números aleatórios, que é de conhecimento do experimentador, de modo que só o acaso determina quem recebe qual tratamento. Sob essa condição, espera-se, portanto, que as unidades recebendo os diferentes tratamentos sejam comparáveis, tal como discutido no capítulo sobre randomização. Entretanto, o acaso poderia em princípio fazer os grupos tratado e controle diferirem de maneira significativa, não

⁶ *Timing* refere-se à relação cronológica entre as observações dos *status* de doença e exposição e suas ocorrências de fato.

sendo difícil quantificar o seu impacto potencial e distingui-lo do efeito do tratamento. Testes estatísticos usuais e intervalos de confiança fazem precisamente isso. Sob esse desenho, portanto, toda a teoria desenvolvida para o Modelo de Rubin se aplica diretamente. A causa a ser avaliada bem como a causa alternativa são estabelecidas *a priori* e, uma vez alocadas aleatoriamente nas *U* unidades investigadas, qualquer medida de efeito pode ser obtida diretamente, com a crença geral de que irá satisfazer aos critérios de validade discutidos – salvo pelo acaso e por particulares situações –, tal como comentado na seção sobre randomização.

Figura 2 – *Timing* dos principais estudos epidemiológicos observacionais em relação à exposição e à resposta



Fonte: Hennekens & Buring (1987).

Um estudo observacional, entretanto, pode ser entendido genericamente como uma investigação empírica de tratamentos (ou exposições) e dos efeitos que eles causam, na qual, entretanto, diferentemente de um estudo experimental, o investigador não tem controle sobre a designação dos tratamentos considerados para as unidades. Isto é, as unidades não são designadas para os tratamentos por um dispositivo aleatório criado por um experimentador. Como já visto aqui, esse fato faz dos estudos observacionais mais sujeitos a vieses e conseqüentemente menos poderosos para avaliação de efeitos causais que os estudos experimentais. Assim, estudos observacionais são tipicamente empregados quando por alguma razão um estudo experimental não é possível, ou por questões éticas ou pela própria factibilidade do estudo. Assim, nos estudos observacionais, para compensar a insegurança quanto à comparabilidade dos grupos devida à falta de controle sobre o mecanismo de designação de tratamentos, o investigador deve se armar com uma teoria estatística pertinente. Essa teoria deve compreender, portanto, uma estrutura e um conjunto de ferramentas que forneçam medidas que considerem as evidências do estudo.

Procedimentos de ajustamento comumente utilizados tais como pareamento e estratificação, que procuram viabilizar a hipótese de homogeneidade de unidades e que podem, por exemplo, ser operacionalizados pelo escore de propensão, não garantem que os grupos tratado e controle sejam comparáveis sob todos os aspectos relevantes. Por exemplo, o pareamento garante que os dois grupos são comparáveis apenas quanto a algumas características previamente definidas. Se os grupos não são comparáveis antes da exposição aos tratamentos em razão de algum aspecto não considerado no pareamento, então uma eventual diferença na resposta observada pode ser apenas um reflexo dessa diferença inicial. O problema é de fato grave quando os grupos não são comparáveis e os dados disponíveis não são suficientes ou fracassam na tentativa de revelar esse fato. Técnicas

específicas, tais como a Análise de Sensibilidade, têm sido desenvolvidas para discutir essa questão. Se a estimativa de um efeito causal é insensível a variações plausíveis em supostos valores de co-variáveis não observadas, então uma interpretação causal se torna mais defensável. Assim, enquanto os procedimentos de ajuste comumente utilizados podem tentar corrigir vieses já conhecidos, os vieses ocultos, isto é, aqueles que necessitam de informações adicionais não observadas para serem visualizados, são tratados sob a técnica de Análise de Sensibilidade (Rosembaun, 1995).

Entre os diversos desenhos observacionais, um estudo de coorte ocupa uma posição privilegiada, sobretudo porque qualquer medida de risco pode ser obtida diretamente. Entretanto, a falta de controle do mecanismo de designação pode inviabilizar uma conclusão causal se não houver uma preocupação específica com os diversos vieses possíveis. No exemplo do uso de cafeína como fator causal para doença cardiovascular, viu-se que estimativas não ajustadas para uma medida de efeito pertinente deve ser confundida por muitas variáveis, tal como tabagismo, e deve também apresentar falta de validade de seguimento. Desse modo, o número de variáveis que deveriam ser controladas é muito grande para que um controle adequado fosse feito por estratificação. Assim, dado que a verdadeira dependência funcional dos riscos de doença cardiovascular para uso de cafeína e eventuais variáveis confundidoras é desconhecida, as estimativas obtidas por modelos multivariados provavelmente também estariam viesadas. E, mesmo que esse viés não fosse importante, as estimativas ainda permaneceriam confundidas em virtude da incapacidade de se medir acuradamente todas as variáveis envolvidas. Dado esse exemplo, no qual se constata que em um estudo de coorte devem existir diversas fontes de vieses de magnitudes desconhecidas e direções diferentes, qualquer conclusão de causalidade deve ser vista com muita cautela. No entanto, os elementos essenciais que constituem o Modelo de Rubin são facilmente identificados em um estudo de coorte. O desafio

seria, portanto, a identificação de um mecanismo ignorável, de um mecanismo que fosse dependente apenas dos dados observados. Essa não é obviamente uma tarefa simples, pois pressupõe necessariamente o estabelecimento de hipóteses adjacentes. E na possibilidade de tais hipóteses não serem passíveis de testes à luz dos dados observados, qualquer conclusão de causalidade deve vir acompanhada explicitamente da crença de que são válidas.

A situação é particularmente mais desfavorável quando o investigador está posicionado cronologicamente após a ocorrência da doença (Figura 2), quando a investigação causal é desenvolvida com base em um estudo de caso-controle ou de um estudo de coorte-retrospectivo. Assim, apesar da boa aderência dos estudos de coortes às condições do Modelo de Rubin, principalmente devido à forma de observação, isto é, buscando-se o efeito de causas postuladas e não as causas do efeito observado, as dificuldades práticas desses estudos têm estimulado um desenvolvimento teórico de desenhos de estudo retrospectivos, particularmente os estudos de caso-controle.

A característica básica que distingue um estudo de caso-controle é que a seleção das unidades é intencionalmente baseada na resposta dos indivíduos. Um grande complicador desse desenho é o fato de que a obtenção de medidas de efeito que estejam associadas explicitamente à noção de risco não é imediata, e depende da possibilidade de inclusão ou não dos casos no grupo dos controles selecionados. Aliás, a constituição de um grupo controle adequado é, sem dúvida, o grande desafio dos estudos e tem ocupado grande parte do tempo dos autores preocupados com tal questão. Segundo Greenland (1991), se os indivíduos que se tornam casos no período de risco considerado são ineleáveis para inclusão no grupo controle, como nos tradicionais desenhos de caso-controle, a hipótese de raridade da doença em questão será necessária para se estimarem riscos relativos com base em dados. Se, entretanto, tal como nos novos desenhos de caso-controle, esses indivíduos, mesmo sendo casos, são elegíveis para o grupo controle, então a hipótese de raridade pode ser dispensada.

As limitações do método caso-controle são bem conhecidas. Falta de validade devido a viés de seleção, mais freqüentemente causado por níveis altos de recusa à participação no estudo, gerando controles não representativos da população sob risco; erros de mensuração, particularmente devido a diferentes vieses de recordação entre casos e controles; e confundimento, a sempre presente possibilidade de que a associação encontrada é resultado de variáveis escondidas que falseiam uma associação causal, constituem as principais fragilidades desse estudo (Breslow, 1996). Entre suas vantagens, destacam-se as reduções tanto no tempo quanto no número de indivíduos necessários para atingir o mesmo poder estatístico que se atingiria em um estudo de coorte. Desse modo, adicionalmente às quatro condições de validade já apresentadas, deve-se acrescentar a condição de validade de seleção, tanto de casos quanto de controles. Haverá validade de seleção de casos quando o número de casos estudados – provavelmente menor que a população de casos, por alguma razão tal como a incapacidade de registrar todos –, que tenham ocorrido durante certo período, forneça estimativas não viesadas das prevalências, para a população de casos, das diferentes causas (níveis de exposição) consideradas. Validade de seleção de controles acontece sob a mesma lógica, substituindo casos por controles.

Com relação a procedimentos de ajuste, uma importante característica distingue os estudos de caso-controle e os de coorte. Nos estudos de coorte, pareamento refere-se à seleção das subcoortes de exposição, de modo a se forçar distribuições semelhantes entre as subcoortes para os fatores utilizados no pareamento. Esse procedimento deve prevenir confundimento para as variáveis envolvidas no pareamento, a menos que haja outra questão tal como uma falta de validade de seguimento relacionada a alguma variável do pareamento. Nos estudos de caso-controle, pareamento refere-se à seleção de indivíduos de maneira a forçar que as distribuições dos fatores considerados entre casos e controles sejam similares. Entretanto, sob esse

desenho, pareamento não previne confundimento. De fato, é sabido que um estudo de caso-controle pareado produz viés de seleção que pode ser, todavia, controlado pelo fator de pareamento. Um exemplo simples, presente em Greenland (1991), pode ser útil para esclarecer. Considere uma população sob risco, metade homem, em que os homens tendem a beber menos café que as mulheres e cerca de 75% dos casos de uma certa doença é de homens. Uma seleção de controles não viesada deveria ser constituída de 50% de homens. Se, entretanto, faz-se um pareamento dos controles aos casos pela variável sexo, cerca de 75% dos controles seriam constituídos de homens. E, visto que homens bebem menos café e seriam majoritários no grupo controle, a freqüência de usuários de cafeína estaria diminuída no grupo controle, subestimando a verdadeira proporção de usuários de cafeína na população sob risco. Como resultado, uma medida de efeito não ajustada seria superestimada. No entanto, as prevalências de usuários de cafeína no grupo controle específicas por sexo não são alteradas pelo pareamento. Logo, o viés de seleção poderia ser removido pelo ajustamento para a própria variável utilizada no pareamento. A conclusão é que pareamento pode exigir controle na análise para a variável utilizada no pareamento. Existe vasta literatura que discute os méritos desse procedimento; entre eles, a eventual redução na variância dos estimadores ajustados é considerado o principal. Assim, a fim de evitar um aumento no número de variáveis a serem controladas, o pareamento, quando indicado, deveria ser limitado apenas às variáveis que de fato fossem necessárias. Entretanto, a utilização já discutida do score de propensão como instrumento para a construção de pares poderia ser uma excelente alternativa.

Do ponto de vista do Modelo de Rubin, a variável resposta para um estudo de caso-controle é a variável dicotômica que indica se a unidade é um caso ou um controle, isto é,

$$Y_s = \begin{cases} 1, & \text{caso a unidade seja um caso} \\ 0, & \text{caso a unidade seja um controle.} \end{cases}$$

E agora, diferentemente da situação ideal, deve-se buscar a causa desse efeito já observado. Nesse tipo de desenho, os parâmetros associados são as proporções populacionais (ou probabilidades), entre casos e controles, de indivíduos com e sem a exposição de interesse, tal como representadas na Tabela 7.

Como visto, os parâmetros de interesse em um estudo de causalidade são $E(Y_t)$ e $E(Y_c)$, que, quando se trata de variáveis dicotômicas, podem ser escritos como $\Pr(Y_t = 1)$ e $\Pr(Y_c = 1)$. Logo, as probabilidades retrospectivas da Tabela 7, genericamente dadas por $\Pr(S = k | Y_s = y)$, não apresentam nenhuma interpretação causal, pois sequer se referem ao evento correto. Entretanto, fazendo-se uso do teorema de Bayes, pode-se escrever

$$\Pr(Y_s = y | S = k) = \Pr(S = k | Y_s = y) \frac{\Pr(Y_s = y)}{\Pr(S = k)},$$

mas, $\Pr(Y_s = y | S = k) = \Pr(Y_k = y | S = k)$ e

$$\Pr(S = k) = \sum_y \Pr(S = k | Y_s = y) \Pr(Y_s = y). \text{ Logo,}$$

$$\Pr(Y_k = y | S = k) = \frac{\Pr(S = k | Y_s = y) \Pr(Y_s = y)}{\sum_y \Pr(S = k | Y_s = y) \Pr(Y_s = y)}.$$

Tabela 7 – Probabilidades populacionais retrospectivas

Tratamentos ou causas	Resposta (Y_s)	
	Casos ($Y_s=1$)	Controles ($Y_s=0$)
$S = t$	$\Pr(S=t / Y_s=1)$	$\Pr(S=t / Y_s=0)$
$S = c$	$\Pr(S=c / Y_s=1)$	$\Pr(S=c / Y_s=0)$
TOTAL	1	1

As probabilidades $\Pr(Y_k = y | S = k)$ são ditas ‘prospectivas’ porque os eventos condicionantes acontecem em algum tempo antes do evento de interesse. E são aquelas passíveis de ser obtidas diretamente por meio de estudos prospectivos, experimentais ou de coorte. A Tabela 8 ilustra essas probabilidades. A expressão anterior revela que as probabilidades prospectivas

podem ser determinadas com base nas probabilidades retrospectivas e nas proporções de casos e controles da população.

Com base nos parâmetros $E(Y_t) = \Pr(Y_t = 1)$ e $E(Y_c) = \Pr(Y_c = 1)$, viu-se que foi construída a medida de efeito causal conhecida como ‘risco atribuível’, definida como a diferença entre esses dois parâmetros. Entretanto, outras medidas de efeito de interesse podem ser definidas. O risco relativo, definido como

$$RR = \frac{\Pr(Y_t = 1)}{\Pr(Y_c = 1)}$$

e o *odds ratio*, cuja expressão correspondente é

$$OR = \frac{\Pr(Y_t = 1)}{\Pr(Y_t = 0)} \bigg/ \frac{\Pr(Y_c = 1)}{\Pr(Y_c = 0)},$$

são largamente utilizados na literatura epidemiológica.

Tabela 8 – Probabilidades populacionais prospectivas

Tratamentos ou causas	Resposta (Y)		TOTAL
	Casos (Y=1)	Controles (Y=0)	
$S = t$	$\Pr(Y_t=1 / S=t)$	$\Pr(Y_t=0 / S=t)$	1
$S = c$	$\Pr(Y_c=1 / S=c)$	$\Pr(Y_c=0 / S=c)$	1

Dentro de um estudo caso-controle tradicional em que a hipótese de raridade da doença não pareça razoável, a idéia de risco fica comprometida. Nesse caso, o parâmetro causal de interesse se limita ao *odds ratio*, que, com base na Tabela 7, pode ser estimado por

$$\hat{OR}_r = \frac{\Pr(S = t \mid Y_s = 1)}{\Pr(S = c \mid Y_s = 1)} \bigg/ \frac{\Pr(S = t \mid Y_s = 0)}{\Pr(S = c \mid Y_s = 0)}$$

Para verificar se essa medida pode ter uma interpretação causal, basta observar que ela se iguala à que seria obtida prospectivamente. Aplicando-se o teorema de Bayes às probabilidades que constituem \hat{OR}_r e observando-se, como antes, que

$$\Pr(Y_S = y | S = k) = \Pr(Y_k = y | S = k),$$

chega-se à mesma expressão para o *odds ratio*, que pode ser obtida com base na Tabela 8, dada por

$$\hat{OR}_p = \frac{\Pr(Y_t = 1 | S = t)}{\Pr(Y_t = 0 | S = t)} \bigg/ \frac{\Pr(Y_c = 1 | S = c)}{\Pr(Y_c = 0 | S = c)} = \hat{OR}_r$$

Apesar disso, ainda não é possível dizer se um particular estudo de caso-controle gozaria de uma interpretação causal. A expressão acima está para *OR* assim como

$$\hat{T} = E(Y_t | S = t) - E(Y_c | S = c) = \Pr(Y_t = 1 | S = t) - \Pr(Y_c = 1 | S = c)$$

está para

$$T = E(Y_t) - E(Y_c) = \Pr(Y_t = 1) - \Pr(Y_c = 1)$$

Logo, sem uma informação adicional não se pode falar sobre causalidade. No caso de \hat{T} , viu-se que sob randomização o problema estaria resolvido. Entretanto, um estudo de caso-controle nunca é randomizado. Assim, a alternativa consiste na observação de um conjunto de co-variáveis X e na busca de um modelo para o mecanismo de designação que seja ignorável, de modo que, condicionalmente a X , S e Y , sejam independentes.

Nossa finalidade, aqui, foi tentar inserir os desenhos tradicionalmente utilizados na pesquisa epidemiológica para atribuição de causa à lógica de causalidade presente no Modelo de Rubin. Se, em um nível mais geral, os estudos experimentais se contrapõem aos estudos observacionais pelo controle que o investigador possui sobre o mecanismo de designação de tratamentos, especificamente nos estudos observacionais, há também uma contraposição, porém muito menos significativa, dada pelos estudos de caso-controle e de coorte que se refere à noção apresentada de *timing* de observação. Assim, para o estabelecimento das condições de validade devido a Greenland (1991), essas questões foram consideradas, e, sob este aspecto, os estudos de coorte-retrospectivo se alinham aos estudos de caso-controle.