

# Epidemiologia Moderna

3ª Edição

Kenneth J. Rothman  
Sander Greenland  
Timothy L. Lash



### **Revisão técnica desta edição**

Álvaro Vigo

Bacharel em Estatística pela UFRGS. Mestre em Estatística pela Universidade Estadual de Campinas (UNICAMP). Doutor em Epidemiologia pela UFRGS. Professor Adjunto do Departamento de Estatística da UFRGS.

Helena Barreto dos Santos

Graduada em Medicina pela UFRGS. Mestre em Epidemiologia pela Universidade Federal de Pelotas (UFPel). Doutora em Epidemiologia pela UFRGS. Assessora de Planejamento e Avaliação do Hospital de Clínicas de Proto Alegre.

Jandyrá Maria Guimarães Fachel

Graduada em Matemática pela UFRGS. Mestre em Estatística pela USP. Doutora em Estatística pela University of London. Professora Titular do Departamento de Estatística da UFRGS e do Programa de Pós-Graduação em Epidemiologia da UFRGS.

Lavínia Schüller-Faccini

Graduada em Medicina pela UFRGS. Doutora em Genética e Biologia Molecular pela UFRGS. Pós-Doutora em Toxicologia Reprodutiva pela Universidade de Toronto, Canadá. Professora Associada do Departamento de Genética da UFRGS.

Luciana Neves Nunes

Graduada em Estatística pela UFRGS. Mestre em Epidemiologia pela UFPel. Doutora em Epidemiologia pela UFRGS. Professora Adjunta do Departamento de Estatística da UFRGS.

Patrícia Klarmann Ziegelmann

Bacharel em Estatística pela UFRGS. Mestre em Estatística pela USP. Doutora em Estatística pela University of Kent, Canterbury. Professora Adjunta do Departamento de Estatística da UFRGS e do Programa de Pós-Graduação em Ciências Médicas da UFRGS.

Sotero Serrate Mengue

Doutor em Ciências Farmacêuticas pela UFRGS. Professor do Programa de Pós-Graduação em Epidemiologia da Faculdade de Medicina da UFRGS.

Vivian Cristine Luft

Graduada em Nutrição pela UFRGS. Mestre e Doutora em Epidemiologia pela UFRGS. Membro de Grupo de Pesquisa do Programa de Pós-Graduação em Epidemiologia da UFRGS. Bolsista de Desenvolvimento Tecnológico e Industrial Nível 1 pelo CNPQ.

**Kenneth J. Rothman**  
Vice President, Epidemiology  
Research  
RTI Health Solutions  
Professor of Epidemiology and  
Medicine  
Boston University  
Boston, Massachusetts

**Sander Greenland**  
Professor of Epidemiology and  
Statistics  
University of California  
Los Angeles, California

**Timothy L. Lash**  
Associate Professor of  
Epidemiology and Medicine  
Boston University  
Boston, Massachusetts

# Epidemiologia Moderna

3ª Edição

**Tradução**

Geraldo Serra

**Consultoria e supervisão desta edição**

Maria Inês Schmidt

Graduada em Medicina pela Faculdade Católica de Medicina de Porto Alegre. Mestre em Epidemiologia pela University of North Carolina. PhD em Epidemiologia pela University of North Carolina. Professora Associada do Departamento de Medicina Social da Universidade Federal do Rio Grande do Sul (UFRGS) e do Programa de Pós-Graduação em Epidemiologia da UFRGS.

Suzi Alves Camey

Bacharel em Estatística pela UFRGS. Mestre em Matemática pela UFRGS. Doutora em Estatística pela Universidade de São Paulo (USP). Professora Adjunta do Departamento de Estatística da UFRGS e do Programa de Pós-Graduação em Epidemiologia da UFRGS.

Versão impressa  
desta obra 2011



2011

Obra originalmente publicada sob o título *Modern Epidemiology*, 3rd Edition  
ISBN 9780781755641

© 2008 by Lippincot Williams & Wilkins

Published by arrangement with Lippincott Williams & Wilkins/Wolters Kluwer Health Inc. USA

Indicações, reações colaterais e programação de dosagens estão precisas nesta obra, mas poderão sofrer mudanças com o tempo. Recomenda-se ao leitor sempre consultar a bula do medicamento antes de sua administração. Os autores e editores não podem ser responsabilizados por erros ou omissões ou quaisquer consequências advindas da aplicação incorreta de informação contida nesta obra.

Capa: *Mário Röhne*

Preparação de original: *Marcia Rolin Serafini*

Leitura final: *Luiza Signorelli Germano*

Editora Sênior – Biociências: *Letícia Bispo*

Editora responsável por esta obra: *Laura Ávila de Souza*

Editoração eletrônica: *Techbooks*



R845e Rothman, Kenneth J.  
Epidemiologia moderna [recurso eletrônico] / Kenneth J.  
Rothman, Sander Greenland, Timothy L. Lash ; tradução:  
Geraldo Serra. – 3. ed. – Dados eletrônicos. – Porto Alegre :  
Artmed, 2011.

Editado também como livro impresso em 2011.  
ISBN 978-85-363-2588-0

1. Epidemiologia. I. Greenland, Sander. II. Lash, Timothy L.  
III. Título.

CDU 616-036.22

Catálogo na publicação: Ana Paula M. Magnus – CRB 10/2052

Reservados todos os direitos de publicação, em língua portuguesa, à  
ARTMED® EDITORA S.A.

Av. Jerônimo de Ornelas, 670 - Santana

90040-340 Porto Alegre RS

Fone (51) 3027-7000 Fax (51) 3027-7070

É proibida a duplicação ou reprodução deste volume, no todo ou em parte, sob quaisquer formas ou por quaisquer meios (eletrônico, mecânico, gravação, fotocópia, distribuição na Web e outros), sem permissão expressa da Editora.

SÃO PAULO

Av. Embaixador Macedo Soares, 10.735 - Pavilhão 5 - Cond. Espace Center

Vila Anastácio 05095-035 São Paulo SP

Fone (11) 3665-1100 Fax (11) 3667-1333

SAC 0800 703-3444

IMPRESSO NO BRASIL

PRINTED IN BRAZIL

# Prefácio e agradecimentos

Esta 3ª edição de *Epidemiologia Moderna* chega mais de 20 anos depois da 1ª, que consistia em um volume muito menor, com um só autor, e esboçava os conceitos e os métodos de uma disciplina em rápido crescimento. A 2ª edição, publicada 12 anos mais tarde, foi uma transição importante, pois o livro cresceu juntamente com a área. Foi incluído um novo autor, e houve expansão de tópicos em várias subdisciplinas, com a contribuição de especialistas convidados. Agora, com a ajuda de um terceiro autor, a nova edição revisa o conteúdo do modo abrangente e introduz novos tópicos essenciais aos epidemiologistas do século XXI.

Esta edição mantém a organização básica da 2ª, sendo dividida em quatro seções. A Seção I “Conceitos básicos”, agora compreende cinco capítulos, em vez de quatro, com a realocação do Capítulo 5, “Conceitos de interação”, que corresponde ao Capítulo 18 da 2ª edição. O tópico de interação pertence, de forma adequada, aos conceitos básicos, embora um leitor que queira ampliar a compreensão prática dos princípios epidemiológicos possa adiar a leitura desse tópico para depois da Seção II, “Delineamento e condução do estudo”. Foi acrescentado um novo capítulo a respeito de diagramas causais, o qual consideramos colocar na Seção I, pois envolve assuntos básicos na conceituação de relações entre variáveis de estudos. No entanto, esse material invoca conceitos que parecem mais ligados à análise de dados e pressupõe o conhecimento de delineamento, motivo pelo qual optamos por colocá-lo no começo da Seção III, “Análise de dados”. Aqueles com formação básica em epidemiologia podem ler o Capítulo 12 em conjunto com os Capítulos 2 e 4, para adquirir uma base

meticulosa dos conceitos envolvendo relações causais e não causais entre variáveis. Outra adição importante é um capítulo na Seção III, intitulado “Introdução à estatística bayesiana”, o qual, esperamos, estimulará os epidemiologistas a considerar e aplicar métodos bayesianos a cenários epidemiológicos. O capítulo anterior sobre análise de sensibilidade, agora intitulado “Análise de viés”, foi revisado e expandido substancialmente para incluir métodos probabilísticos, os quais penetraram na epidemiologia a partir dos campos de risco e de análise de políticas. A aplicação rígida de interpretações de estatística frequentista sobre os dados tem dominado a pesquisa biomédica (assim como muitas outras ciências). Assim, esperamos que os novos capítulos da Seção III ajudem a liberar os epidemiologistas das algemas da estatística frequentista e a deixá-los mais abertos a abordagens mais flexíveis, realistas e profundas de análise e inferência.

Como nas edições anteriores, a Seção IV compreende tópicos adicionais que são mais especializados do que aqueles considerados nas três primeiras partes do livro. Embora os métodos de campo ainda tenham aplicação ampla na pesquisa epidemiológica, houve um crescimento de trabalhos de epidemiologia baseados em fontes de dados existentes, tais como os de registros e ações médicas. Dessa forma, o capítulo sobre métodos de campo foi movido da Seção II para a Seção IV, e um capítulo intitulado “Usando dados secundários” foi acrescentado. Outro acréscimo é o capítulo sobre epidemiologia social, e a cobertura da epidemiologia molecular foi acrescida ao capítulo sobre epidemiologia genética. Muitos desses capítulos podem ser de interesse principalmente

daqueles que estão focados em uma área em particular, tal como epidemiologia reprodutiva ou epidemiologia das doenças infecciosas, que têm preocupações metodológicas distintas, embora os tópicos levantados mereçam ser considerados por todo epidemiologista que deseje tornar-se proficiente no ramo. Tópicos como estudos ecológicos e metanálise têm um interesse amplo, que perpassa os objetos de estudo de subdisciplinas. O rastreamento teve seu capítulo próprio na 2ª edição; seu conteúdo foi incorporado ao capítulo revisado sobre epidemiologia clínica.

O escopo da epidemiologia tornou-se grande demais para um só livro: assim, esperamos familiarizar aqueles que desejam compreender os conceitos e os métodos da epidemiologia, abordando os assuntos que são centrais à disciplina e apontando o caminho das referências-chave para estudos posteriores. Embora suas edições prévias tenham sido adotadas como livro-texto em muitos programas de ensino de epidemiologia, este livro não foi escrito como um texto de referência para um curso específico, nem contém exercícios ou perguntas de revisão, como em muitos textos de cursos. Alguns leitores podem considerá-lo mais valioso como um livro de referência ou de leitura complementar, para usar junto com livros menores, tais como Kelsey et al. (1996), Szklo e Nieto (2000), Savitz (2001), Koepsell e Weiss (2003), ou Checkoway et al. (2004). No entanto, há subgrupos de capítulos que poderiam formar a base de um livro-texto para cursos de métodos epidemiológicos. Por exemplo, um curso sobre teoria e métodos epidemiológicos poderia se basear nos Capítulos 1 a 12; um curso mais abreviado poderia se basear nos Capítulos 1 a 4 e 6 a 11; um curso curto sobre os fundamentos da teoria epidemiológica poderia se basear nos Capítulos 1 a 5, mais o Capítulo 12. Presumindo-se uma formação em epidemiologia básica, uma introdução à análise de dados epidemiológicos poderia se basear nos Capítulos 9, 10 e 12 a 19, ao passo que um curso

mais avançado, detalhando análise causal e de regressão, poderia se basear nos Capítulos 2 a 5, 9, 10 e 12 a 21. Muitos dos outros capítulos também se encaixariam em outras coleções sugeridas, dependendo do programa e do currículo.

Muitos tópicos são discutidos em várias seções do livro, porque pertencem a mais de um aspecto da ciência. Para facilitar o acesso a todas as seções relevantes que se relacionam a determinado assunto, indexamos o texto minuciosamente. Assim, recomendamos que o índice seja consultado por aqueles que desejam ler nossa discussão completa de tópicos específicos.

Esperamos que esta edição represente um recurso para professores, estudantes e praticantes da epidemiologia; tentamos ser tão acurados quanto possível, mas reconhecemos que qualquer trabalho dessa abrangência conterá equívocos e omissões. Agradecemos aos leitores de edições anteriores que trouxeram tais itens à nossa atenção. Pretendemos continuar nossa prática de colocar tais correções em uma página da internet e incorporá-las a impressões subsequentes. Por favor, consulte <http://www.lww.com/ModernEpidemiology> para encontrar informações mais recentes sobre erratas.\*

Também somos gratos a muitos colegas que revisaram seções deste texto e forneceram *feedback* útil. Embora não possamos mencionar todos que auxiliaram nesse aspecto, devemos agradecimentos especiais a Onyebuchi Arah, Mathew Fox, Jamie Gradus, Jennifer Hill, Katherine Hoggatt, Marshal Joffe, Ari Lipsky, James Robins, Federico Soldani, Henrik Toft Sørensen, Soe Thwin e Tyler VanderWeele. Uma versão inicial do Capítulo 18 apareceu no *International Journal of Epidemiology* (2006;35:765-778), reproduzida com permissão da Oxford University Press. Finalmente, agradecemos a Mary Anne Armstrong, Alan Dyer, Gary Friedman, Ulrik Gerdorf, Paul Sorlie e Katsuhiko Yano, por fornecer informações não publicadas, usadas nos exemplos do Capítulo 33.

**Kenneth J. Rothman  
Sander Greenland  
Timothy L. Lash**

---

\* As erratas divulgadas até a data de 30/11/10 já foram incorporadas ao texto.

# Colaboradores

**Allen J. Wilcox**

Senior Investigator  
Epidemiology Branch  
National Institute of Environmental Health  
Sciences/NIH  
Durham, North Carolina

**Barbara E. Mahon**

Assistant Professor  
Department of Epidemiology and Pediatrics  
Boston University  
Novartis Vaccines and Diagnostics  
Boston, Massachusetts

**C. Robert Horsburgh, Jr.**

Professor of Epidemiology, Biostatistics and  
Medicine  
Department Epidemiology  
Boston University School of Public Health  
Boston, Massachusetts

**Charles Poole**

Associate Professor  
Department of Epidemiology  
University of North Carolina at Chapel Hill,  
School of Public Health  
Chapel Hill, North Carolina

**Clarice R. Weinberg**

National Institute of Environmental Health  
Sciences  
Biostatistics Branch  
Research Triangle Park, North Carolina

**Hal Morgenstern**

Professor and Chair  
Department of Epidemiology  
University of Michigan School of Public Health  
Ann Arbor, Michigan

**Irva Hertz-Picciotto**

Professor  
Department of Public Health  
University of California, Davis  
Davis, California

**Keith O'Rourke**

Visiting Assistant Professor  
Department of Statistical Science  
Duke University  
Durham, North Carolina  
Adjunct Professor  
Department of Epidemiology and Community  
Medicine  
University of Ottawa  
Ottawa, Ontario  
Canada

**Jack Cahill**

Vice President  
Department of Health Studies Sector  
Westat, Inc.  
Rockville, Maryland

**James W. Buehler**

Research Professor  
Department of Epidemiology  
Rollins School of Public Health  
Emory University  
Atlanta, Georgia

**Jay S. Kaufman**

Associate Professor  
Department of Epidemiology  
University of North Carolina at Chapel Hill,  
School of Public Health  
Chapel Hill, North Carolina

**Jørn Olsen**

Professor and Chair  
Department of Epidemiology  
UCLA School of Public Health  
Los Angeles, California

**M. Maria Glymour**

Robert Wood Johnson Foundation Health and  
Society Scholar  
Department of Epidemiology  
Mailman School of Public Health  
Columbia University  
Nova Iorque, Nova Iorque  
Department of Society, Human Development  
and Health  
Harvard School of Public Health  
Boston, Massachusetts

**Marta Gwinn**

Associate Director  
Department of Epidemiology  
National Office of Public Health Genomics  
Centers for Disease Control and Prevention  
Atlanta, Georgia

**Muin J. Khoury**

Director  
National Office of Public Health Genomics  
Centers for Disease Control and Prevention  
Atlanta, Georgia

**Noel S. Weiss**

Professor  
Department of Epidemiology  
University of Washington  
Seattle, Washington

**Patricia Hartge**

Deputy Director  
Department of Epidemiology and Biostatistics  
Program  
Division of Cancer Epidemiology and  
Genetics  
National Cancer Institute, National Institutes of  
Health  
Rockville, Maryland

**Robert C. Millikan**

Professor  
Department of Epidemiology  
University of North Carolina at Chapel Hill,  
School of Public Health  
Chapel Hill, North Carolina

**Walter C. Willett**

Professor and Chair  
Department of Nutrition  
Harvard School of Public Health  
Boston, Massachusetts



# Sumário

<b>1</b>	<b>Introdução</b>	<b>11</b>	<b>8</b>	<b>Estudos de caso-controle</b>	<b>136</b>
	<i>Kenneth J. Rothman, Sander Greenland e Timothy L. Lash</i>			<i>Kenneth J. Rothman, Sander Greenland e Timothy L. Lash</i>	
<b>Seção I</b>			<b>9</b>		
<b>Conceitos básicos</b>			<b>Validade em estudos epidemiológicos</b>		
					<b>156</b>
				<i>Kenneth J. Rothman, Sander Greenland e Timothy L. Lash</i>	
<b>2</b>	<b>Causalidade e inferência causal</b>	<b>15</b>	<b>10</b>	<b>Precisão e estatística em estudos epidemiológicos</b>	<b>179</b>
	<i>Kenneth J. Rothman, Sander Greenland, Charles Poole e Timothy L. Lash</i>			<i>Kenneth J. Rothman, Sander Greenland e Timothy L. Lash</i>	
<b>3</b>	<b>Medidas de ocorrência</b>	<b>46</b>	<b>11</b>	<b>Estratégias de delineamento para melhorar a acurácia de estudos</b>	<b>202</b>
	<i>Sander Greenland e Kenneth J. Rothman</i>			<i>Kenneth J. Rothman, Sander Greenland e Timothy L. Lash</i>	
<b>4</b>	<b>Medidas de efeito e medidas de associação</b>	<b>67</b>	<b>12</b>	<b>Diagramas causais</b>	<b>219</b>
	<i>Sander Greenland, Kenneth J. Rothman e Timothy L. Lash</i>			<i>M. Maria Glymour e Sander Greenland</i>	
<b>5</b>	<b>Conceitos de interação</b>	<b>89</b>	<b>Seção III</b>		
	<i>Sander Greenland, Timothy L. Lashman e Kenneth J. Rothman</i>		<b>Análise de dados</b>		
<b>Seção II</b>			<b>13</b>	<b>Fundamentos da análise de dados epidemiológicos</b>	<b>253</b>
<b>Delineamento e condução do estudo</b>				<i>Sander Greenland e Kenneth J. Rothman</i>	
<b>6</b>	<b>Tipos de estudos epidemiológicos</b>	<b>107</b>	<b>14</b>	<b>Introdução à estatística categórica</b>	<b>283</b>
	<i>Kenneth J. Rothman, Sander Greenland e Timothy L. Lash</i>			<i>Sander Greenland e Kenneth J. Rothman</i>	
<b>7</b>	<b>Estudos de coorte</b>	<b>123</b>	<b>15</b>	<b>Introdução à análise estratificada</b>	<b>307</b>
	<i>Kenneth J. Rothman e Sander Greenland</i>			<i>Sander Greenland e Kenneth J. Rothman</i>	

<b>16</b>	<b>Aplicações dos métodos de análise estratificada</b>	<b>335</b>	<b>25</b>	<b>Estudos ecológicos</b>	<b>599</b>
	<i>Sander Greenland</i>			<i>Hal Morgenstern</i>	
<b>17</b>	<b>Análise de fatores de exposições e desfechos politômicos</b>		<b>26</b>	<b>Epidemiologia social</b>	<b>622</b>
	<i>Sander Greenland</i>			<i>Jay S. Kaufman</i>	
<b>18</b>	<b>Introdução à estatística bayesiana</b>	<b>386</b>	<b>27</b>	<b>Epidemiologia das doenças infecciosas</b>	<b>642</b>
	<i>Sander Greenland</i>			<i>C. Robert Horsburgh Jr. e Barbara E. Mahon</i>	
<b>19</b>	<b>Análise de Viés</b>	<b>405</b>	<b>28</b>	<b>Epidemiologia genética e molecular</b>	<b>660</b>
	<i>Sander Greenland e Timothy L. Lash</i>			<i>Muin J. Khoury, Robert Millikan e Marta Gwinn</i>	
<b>20</b>	<b>Introdução a modelos de regressão</b>	<b>447</b>	<b>29</b>	<b>Epidemiologia nutricional</b>	<b>679</b>
	<i>Sander Greenland</i>			<i>Walter C. Willett</i>	
<b>21</b>	<b>Introdução à modelagem por regressão</b>	<b>490</b>	<b>30</b>	<b>Epidemiologia ambiental</b>	<b>700</b>
	<i>Sander Greenland</i>			<i>Irva Hertz-Picciotto</i>	
			<b>31</b>	<b>Tópicos metodológicos em epidemiologia reprodutiva</b>	<b>726</b>
				<i>Clarice R. Weinberg e Allen J. Wilcox</i>	
			<b>32</b>	<b>Epidemiologia clínica</b>	<b>751</b>
				<i>Noel S. Weiss</i>	
			<b>33</b>	<b>Metanálise</b>	<b>763</b>
				<i>Sander Greenland e Keith O'Rourke</i>	
<b>Seção IV</b>			<b>Referências</b>	<b>799</b>	
<b>Tópicos especiais</b>			<b>Índice</b>	<b>857</b>	
<b>22</b>	<b>Vigilância</b>	<b>537</b>			
	<i>James W. Buehler</i>				
<b>23</b>	<b>Uso de dados secundários</b>	<b>563</b>			
	<i>Jorn Olsen</i>				
<b>24</b>	<b>Métodos de campo em epidemiologia</b>	<b>577</b>			
	<i>Patricia Hartge e Jack Hill</i>				

# Introdução

Kenneth J. Rothman, Sander Greenland e  
Timothy L. Lash

**E**mbora algumas pesquisas epidemiológicas de excelência tenham sido conduzidas antes do século XX, um corpo sistematizado de princípios que possibilitasse o delineamento e a avaliação de estudos epidemiológicos só começou a se formar na segunda metade daquele mesmo século. Tais princípios evoluíram em conjunto com uma explosão de pesquisas epidemiológicas, e sua evolução continua até hoje.

Vários estudos epidemiológicos em grande escala exerceram grande influência sobre a saúde. Por exemplo, os ensaios de intervenção comunitária de suplementação de flúor na água, que foram iniciados durante a década de 1940, levaram à disseminação da prevenção primária de cáries dentárias (Ast, 1965). O Framingham Heart Study (Estudo Framingham), iniciado em 1949, destaca-se entre várias pesquisas de seguimento de longo prazo sobre doenças cardiovasculares, que contribuíram de modo importante para a compreensão das causas desse enorme problema de saúde pública (Dawber et al., 1957; Kannel et al., 1961, 1970; McKee et al., 1971). Esse estudo notável continua a produzir achados valiosos, mesmo depois de mais de 60 anos de seu início (Kannel e Abbott, 1984; Sytkowski et al., 1990; Fox et al., 2004; Elias et al., 2004; [www.nhlbi.nih.gov/about/framingham](http://www.nhlbi.nih.gov/about/framingham)). O conhecimento obtido a partir deste e de outros estudos epidemiológicos similares ajudou a refrear a epidemia moderna de mortalidade cardiovascular nos Estados Unidos, que atingiu seu ápice em meados de 1960 (Stallones, 1980). O maior experimento humano formal já realizado foi com a vacina Salk, em 1954, do qual várias centenas de milhares de crianças em idade escolar participaram (Francis et al., 1957). Esse estudo forneceu a primeira base prática para a prevenção da poliomielite parálitica.

Na mesma época, houve a publicação de muitos estudos epidemiológicos sobre os efeitos do uso do fumo. Esses trabalhos levaram, por fim, a um relatório que foi um marco, *Smoking and Health*, divulgado pelo *Surgeon General*\* (United States Department of Health, Education and Welfare, 1964), o primeiro de muitos relatos sobre os efeitos adversos do hábito de fumar na saúde, publicados pelo *Surgeon General* ([www.cdc.gov/Tobacco/sgr/index.htm](http://www.cdc.gov/Tobacco/sgr/index.htm)).

Desde aquele primeiro relatório, a pesquisa epidemiológica tem atraído constantemente a atenção pública. A mídia, impelida por uma “maré alta” de preocupação da sociedade com assuntos de saúde e ambiente, levou muitos estudos epidemiológicos à proeminência. Alguns desses estudos eram controversos, e algumas das pesquisas que mais receberam atenção estavam relacionadas com os seguintes temas:

- *Influenza* aviária
- Síndrome respiratória aguda grave (SARS)
- Terapia de reposição hormonal e doença cardíaca
- Ingestão de carboidratos e saúde
- Vacinação e autismo

---

\* N. de T.: Cargo equivalente ao de Ministro da Saúde, no Brasil.

- Tampões e síndrome do choque tóxico
- *Bendectin* e defeitos congênitos
- Fumantes passivos e saúde
- Síndrome da imunodeficiência adquirida (Aids)
- O efeito do dietilestilbestrol nos descendentes

Em alguns casos, o desacordo sobre pontos conceituais e metodológicos básicos levou a diferenças profundas na interpretação de dados. Em 1978, surgiu uma controvérsia sobre o possível efeito carcinogênico dos estrogênios exógenos sobre o endométrio: vários estudos de caso-controle tinham relatado uma associação extremamente forte, com até 15 vezes de aumento do risco (Smith et al, 1975; Ziel e Finkle, 1975; Mack et al., 1976). Um grupo argumentava que um viés de seleção era responsável pela maior parte da associação observada (Horwitz e Feinstein, 1978), ao passo que outros alegavam que o delineamento alternativo proposto por Horwitz e Feinstein introduzia um viés de seleção para baixo bem mais forte do que qualquer tendenciosidade para cima que ele removesse (Hutchinson e Rothman, 1978; Jick et al., 1979; Greenland e Neutra, 1981). Tais discussões sobre conceitos fundamentais sugerem que as fundações metodológicas da ciência ainda não haviam sido estabelecidas e que a epidemiologia permanecia jovem em termos conceituais.

O último terço do século XX contemplou um crescimento rápido na compreensão e na síntese de conceitos epidemiológicos. O principal estímulo para esse crescimento conceitual parece ter sido prática e controvérsia. A explosão da atividade epidemiológica acentuou a necessidade de melhorar a compreensão dos fundamentos teóricos. Por exemplo, os estudos iniciais sobre fumo e câncer de pulmão (p. ex., Wynder e Graham, 1950; Doll e Hill, 1952) foram notáveis cientificamente, não somente por seus achados substanciais, mas também porque demonstraram a eficácia e a grande eficiência do estudo de caso-controle. Controvérsias sobre o delineamento apropriado de caso-controle levaram ao reconhecimento da importância de relacionar tais estudos a uma fonte populacional subjacente (Sheehe, 1962; Miettinen, 1976a; Cole, 1979; ver Cap. 8). De modo semelhante, a análise de dados do Estudo Framingham estimulou o desenvolvimento do modelo metodológico mais popular na epidemiologia atual, a regressão logística múltipla (Cornfield, 1962; Truett et al., 1967; ver Cap. 20).

Apesar do ímpeto crescente da atividade epidemiológica no final do século XX, as evidências indicam que a epidemiologia permanece em um estágio inicial de desenvolvimento (Pearce e Merletti, 2006). Nos últimos anos, os conceitos epidemiológicos continuaram a evoluir rapidamente, talvez porque o escopo, a atividade e a influência da epidemiologia continuam a aumentar. Esse aumento tem sido acompanhado pelas dores do crescimento, refletindo amplamente a preocupação sobre a validade dos métodos usados na pesquisa epidemiológica e a confiabilidade dos resultados. A disparidade entre os resultados de estudos randomizados (Writing Group for the Woman's Health Initiative Investigators, 2002) e não randomizados (Stampfer e Colditz, 1991) sobre a associação entre terapia de reposição hormonal e doença cardiovascular representa um dos exemplos mais recentes, e de alto perfil, de hipóteses supostamente estabelecidas pela epidemiologia observacional e, subseqüentemente, contraditórias (Davey Smith, 2004; Prentice et al., 2005).

A epidemiologia frequentemente está na mira do público, o que a torna um alvo para críticas. Tais críticas têm sido ampliadas, ocasionalmente, para uma desconfiança quanto aos métodos da epidemiologia em si, indo além do ceticismo em relação a achados específicos e tornando-se um criticismo generalizado em relação à investigação epidemiológica (Taubes, 1995, 2007). Entretanto, essas críticas, embora duras de aceitar, devem ser bem-vindas pelos cientistas. Todos nós aprendemos melhor a partir de nossos erros, e há muito que os epidemiologistas podem fazer para aumentar a confiabilidade e a utilidade de seus achados. Fornecer aos leitores a base para alcançar tal meta é o objetivo deste livro.

## Seção I

# Conceitos básicos



# Causalidade e inferência causal

Kenneth J. Rothman, Sander Greenland,  
Charles Poole e Timothy L. Lash

## Causalidade 15

- Um modelo de causa suficiente e causas componentes 16
- A necessidade de uma categoria de referência específica 18
- Aplicação do modelo de causa suficiente à epidemiologia 18
- Probabilidade, risco e causas 20
- Força dos efeitos 21
- Interação entre causas 24
- Proporção de doença devido a causas específicas 25
- Período de indução 27

- Escopo do modelo 29
- Outros modelos de causalidade 30

## Filosofia da inferência científica 30

- Indutivismo 31
- Refutacionismo 32
- Consenso e naturalismo 33
- Bayesianismo 35
- Impossibilidade de prova científica 37

## Inferência causal em epidemiologia 38

- Testes de teorias epidemiológicas concorrentes 38
- Crítérios causais 39

## CAUSALIDADE

Uma compreensão rudimentar de causa e efeito parece ser adquirida pela maioria das pessoas por si próprias, muito mais cedo do que lhes poderia ter sido ensinado por alguém. Mesmo antes que possam falar, muitas crianças compreendem a relação entre chorar e o aparecimento do pai ou de outro adulto e a relação entre esse aparecimento e o ato de ser carregada ou alimentada. Um pouco mais tarde, as crianças desenvolvem teorias sobre o que acontece quando um copo com leite é derrubado e quando um interruptor na parede é movimentado de sua posição de repouso para outra. Enquanto tais teorias estão sendo formuladas, uma teoria causal mais geral também está sendo formada. A teoria mais geral postula que alguns eventos ou estados da natureza são causas de efeitos específicos. Sem uma teoria geral de causalidade, não haveria uma estrutura sobre a qual apoiar a essência de muitas teorias causais específicas, das quais se necessita para sobreviver.

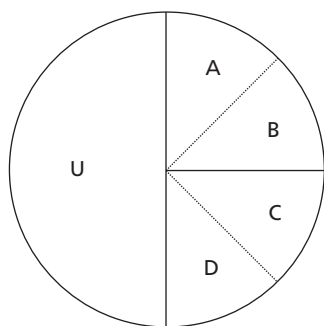
Entretanto, os conceitos de causalidade que são estabelecidos no início da vida são demasiadamente primitivos para servir de base para as teorias científicas. Essa insuficiência pode ser especialmente verdadeira nas ciências da saúde e sociais, nas quais as causas típicas não são necessárias nem suficientes para produzir efeitos de interesse. Portanto, como tem sido reconhecido há muito tempo em epidemiologia, há necessidade de se desenvolver um modelo conceitual mais refinado, que possa servir como ponto de partida para discussões sobre causalidade. Em particular, tal modelo deve abordar problemas de causalidade multifatorial, confundimento, interdependência de efeitos, efeitos diretos e indiretos, categorias de causalidade e sistemas ou redes de causalidade (MacMahon e Pugh, 1967; Susser, 1973). Este capítulo descreve um ponto de partida, o modelo de causa suficiente componente (ou modelo de causa suficiente), que se tem comprovado útil na elucidação de certos conceitos em

mecanismos individuais de causalidade. O Capítulo 4 introduz o modelo de causalidade largamente utilizado, potencial-desfecho, ou contrafactual, que é útil para relacionar a causalidade em nível individual com a causalidade em nível populacional, ao passo que o Capítulo 12 introduz modelos causais gráficos (diagramas causais), que são especialmente úteis para modelagem de sistemas causais.

Exceto quando especificado de modo diferente (em particular, no Cap. 17, sobre doenças infecciosas), em todo o livro, partiremos do pressuposto de que doença refere-se a um evento não recorrente, tal como óbito ou primeira ocorrência de uma enfermidade, e que o desfecho de cada indivíduo ou unidade de estudo (p. ex., um grupo de pessoas) não é afetado pelas exposições e pelos desfechos de outros indivíduos ou unidades. Embora esse pressuposto não simplifique muito nossa discussão e seja razoável em muitas situações, não se aplica a fenômenos contagiosos, tais como comportamentos e doenças transmissíveis. No entanto, todas as definições e a maioria dos pontos que nós levantamos (especialmente no que tange à validade) se aplicam de modo mais geral. Também é essencial compreender situações mais simples antes de lidar com as complexidades criadas pela interdependência causal de indivíduos ou unidades.

## Um modelo de causa suficiente e causas componentes

Inicialmente, é preciso definir *causa*. Uma definição da causa da ocorrência de uma doença específica é um evento, uma condição ou uma característica antecedente necessários para o surgimento de tal doença no momento em que ela ocorreu, contanto que outras condições sejam fixas. Em outras palavras, a causa da ocorrência de uma doença é um evento, uma condição ou uma característica que precedeu o início da doença, e que, se tal evento, condição ou característica tivesse sido diferente de modo específico, tal doença não teria ocorrido de modo algum, ou não teria ocorrido até algum tempo mais tarde. Sob essa definição, se alguém está caminhando por uma pista congelada, cai e fratura o quadril, pode haver uma longa lista de causas. Essas causas poderiam incluir o tempo no dia do incidente, o fato de que a pista não foi limpa para pedestres, a escolha do tipo de calçado pela vítima, a falta de um corrimão, e assim por diante. As causas requeridas para essa pessoa quebrar seu quadril, nessa ocasião em particular, podem ser retratadas com o diagrama de causa suficiente, na Figura 2.1. Por *causa suficiente*, nós queremos dizer um mecanismo causal completo, um conjunto mínimo de condições e de eventos que sejam suficientes para o desfecho ocorrer. O círculo na figura compreende cinco segmentos, cada um dos quais representa um componente causal que deve estar presente, ou deve ter acontecido, para que a pessoa frature seu quadril naquele instante. O primeiro componente, chamado A, representa o mau tempo. O segundo, chamado B, representa uma pista para pedestres que não foi limpa. O terceiro, C, é uma má escolha do tipo de calçado. O quarto componente, D, é a falta de um corrimão. O componente final, chamado U, representa todos os outros eventos, condições ou características não especificados que devem estar presentes, ou ter ocorrido, no exemplo da queda que levou a um quadril fraturado. Para efeitos etiológicos, tais como a causalidade de doença, muitos, possivelmente todos, os componentes de uma causa suficiente podem ser desconhecidos (Rothman,



**FIGURA 2.1** • Ilustração da constelação de causas componentes que constituem uma causa suficiente para fratura do quadril, para uma pessoa em particular em um tempo particular. **A** representa mau tempo, **B** representa uma pista para pedestres que não foi limpa, **C** representa uma má escolha de tipo de calçado, **D** representa a falta de um corrimão e **U** representa todos os outros eventos, condições ou características não especificados que devem estar presentes, ou devem ter ocorrido, no exemplo da queda que levou a um quadril fraturado.



1976a). Geralmente, incluímos uma causa componente, rotulada como U, para representar o conjunto de fatores desconhecidos.

Todas as causas componentes da causa suficiente são necessárias e devem estar presentes, ou ter ocorrido, no exemplo da queda, para a pessoa fraturar o quadril. Nenhuma é supérflua, o que significa que bloquear a contribuição de qualquer causa componente previne a atuação da causa suficiente. Para muitas pessoas, o pensamento causal primitivo persiste nas tentativas de encontrar causas isoladas como explicações para fenômenos observados. Entretanto, a experiência e o raciocínio mostram que o mecanismo causal para qualquer efeito deve consistir em um conjunto de componentes que atuam em conjunto (Mill, 1862; Mackie, 1965). Na etiologia de doenças, uma causa suficiente é um conjunto de condições suficientes para garantir que o desfecho ocorrerá. Portanto, completar uma causa suficiente é equivalente ao início da doença. Início aqui pode se referir à instalação do estágio mais precoce do processo mórbido, ou a qualquer transição de uma fase bem definida e prontamente caracterizada para a próxima, tal como o princípio de sinais ou sintomas.

Considere novamente a função do corrimão em causar a fratura do quadril. A ausência de tal corrimão pode desempenhar um papel causal em algumas causas suficientes, mas não em outras, a depender de circunstâncias como o tempo, o nível de embriaguez do pedestre e incontáveis outros fatores. Nossa definição liga a falta de um corrimão a esse quadril quebrado, e não implica que a sua falta por si só fosse suficiente para que essa fratura ocorresse. Com essa definição de causa, nenhum evento específico, condição ou característica é suficiente para produzir a doença. A definição não descreve um mecanismo causal completo, mas apenas um componente dele. Dizer que a ausência de um corrimão é uma causa componente de um quadril fraturado não implica, entretanto, que toda pessoa caminhando pela pista terá uma fratura de quadril. Não implica, tampouco, que, se for instalado um corrimão com propriedades suficientes para prevenir aquele quadril fraturado, ninguém quebrará o quadril naquela mesma pista. Pode haver outras causas suficientes pelas quais uma pessoa possa sofrer uma fratura de quadril. Cada uma de tais causas suficientes seria ilustrada por seu próprio diagrama, semelhante à Figura 2.1. A primeira dessas causas suficientes, a ser completada pelo acúmulo simultâneo de todas as suas causas componentes, seria aquela que ilustra o mecanismo pelo qual a fratura de quadril ocorre em uma pessoa em particular. Se nenhuma causa suficiente for completada enquanto uma pessoa passa ao longo da pista, então nenhuma fratura de quadril ocorrerá durante o curso daquela caminhada.

Como já foi assinalado, uma característica do conceito ingênuo de causalidade é a presunção de uma correspondência um-para-um entre a causa observada e o efeito. Sob esse ponto de vista, cada causa é tida como “necessária” e “suficiente” para produzir o efeito, especialmente quando a causa é uma ação ou um evento observável que é cronologicamente próximo ao efeito. Assim, o acionar de um interruptor parece ser a causa singular a fazer com que a luz elétrica se acenda. Há causas menos evidentes, contudo, que também operam para produzir o efeito: uma lâmpada em condições funcionais em um ponto de iluminação, fiação intacta do interruptor até a lâmpada e voltagem para produzir uma corrente quando o circuito é fechado. Para se conseguir o efeito de acender a luz, cada um desses componentes é tão importante quanto a movimentação do interruptor, porque a mudança em qualquer desses componentes da constelação causal prevenirá o efeito. O termo *causa necessária* é reservado, portanto, a um tipo particular de causa componente sob o modelo da causa suficiente. Se qualquer das causas componentes aparecer em toda causa suficiente, então aquela causa componente é denominada uma causa componente “necessária”. Para a doença ocorrer, toda e qualquer causa componente necessária deve estar presente, ou deve ter ocorrido. Por exemplo, pode-se rotular uma causa componente com o requisito para que se tenha um quadril para sofrer uma fratura de quadril. Toda causa suficiente que leve a uma fratura de quadril deve ter aquela causa componente presente, porque, para se fraturar um quadril, é preciso que se tenha um quadril.

O conceito de causas componentes complementares será útil em aplicações à epidemiologia que vêm em seguida. Para cada causa componente em uma causa suficiente, o conjunto das outras causas componentes naquela causa suficiente compreende as causas componentes complementares.

Por exemplo, na Figura 2.1, a causa componente A (mau tempo) tem como suas causas componentes complementares as rotuladas como B, C, D e U. A causa componente B (falta de limpeza da pista para pedestres) tem como causas componentes complementares as rotuladas como A, C, D e U.

## A necessidade de uma categoria de referência específica

As causas componentes devem ser definidas em relação a uma alternativa claramente especificada, ou a uma categoria de referência (chamada, simplesmente, de referência). Considere novamente a falta de um corrimão ao longo da pista. Para dizer que essa condição é uma causa componente do quadril fraturado, temos que especificar uma condição alternativa com a qual contrastar a causa. A mera presença de um corrimão não seria suficiente. No final das contas, a fratura do quadril ainda poderia ter ocorrido na presença de um corrimão, se fosse baixo demais, ou estivesse envelhecido e fosse feito de madeira que estivesse podre. Precisaríamos especificar a presença de um corrimão suficientemente alto e firme para evitar a queda, ou seja, para que o corrimão fosse uma causa componente do quadril quebrado.

Para enxergar a necessidade de especificar um evento, uma condição ou uma característica alternativa, assim como a causal, considere um exemplo de um homem que tomou altas doses de ibuprofeno por vários anos e desenvolveu uma úlcera gástrica. O uso de ibuprofeno pelo homem causou sua úlcera? Seria possível pressupor, a princípio, que o contraste natural seria o que teria acontecido se o homem nada tivesse tomado. Dada uma razão forte para tomar o ibuprofeno, entretanto, tal alternativa pode não fazer sentido. Se a alternativa especificada fosse tomar acetaminofeno, uma droga diferente que poderia ter sido indicada para seu problema, e ele não tivesse desenvolvido a úlcera com o uso do acetaminofeno, então poderíamos dizer que usar o ibuprofeno teria causado a úlcera. Porém, o ibuprofeno não teria causado a úlcera se a alternativa especificada fosse tomar ácido acetilsalicílico, e, tendo ele tomado ácido acetilsalicílico, ainda tivesse desenvolvido a úlcera. A necessidade de especificar a alternativa em uma medida preventiva é ilustrada por uma manchete de jornal que dizia: “Carne malpassada baixa risco de câncer do cólon”. Seria essa a história de um estudo epidemiológico, comparando a taxa de câncer de cólon de um grupo de pessoas que comia carne vermelha malpassada com a taxa de câncer de cólon de um grupo de vegetarianos? Não, o estudo comparou pessoas que comiam carne vermelha malpassada com pessoas que comiam carne vermelha bem cozida. A mesma exposição, consumo regular de carne vermelha malpassada, poderia ter um efeito preventivo, quando contrastada com carne vermelha bem cozida, e um efeito causal, ou nenhum efeito, em comparação com a dieta vegetariana. Um evento, uma condição, ou uma característica não constituem uma causa sozinhos, como uma propriedade intrínseca que se possui isoladamente, mas como parte de um contraste causal com um evento, uma condição ou uma característica alternativa (Lewis, 1973; Rubin, 1974; Greenland et al., 1999a; Maldonado e Greenland, 2002; ver Cap. 4).

## Aplicação do modelo de causa suficiente à epidemiologia

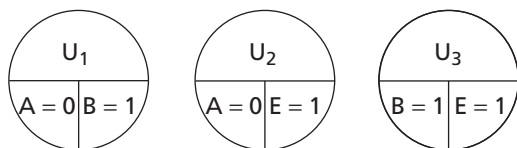
A introdução precedente aos conceitos de causas suficientes e causas componentes fornece o léxico para a aplicação do modelo à epidemiologia. Por exemplo, fumar tabaco é uma causa de câncer de pulmão, mas, por si só, não é uma causa suficiente, como pode ser demonstrado pelo fato de que a maioria dos fumantes não desenvolve esse tipo de câncer. Em primeiro lugar, o termo *fumar* é impreciso demais para ser útil além de uma descrição casual. Deve-se especificar o tipo de fumo (p. ex., cigarro, charuto, cachimbo ou ambiental), se é filtrado ou não, a maneira e a frequência da inalação, a idade de iniciação ao fumo e a duração do hábito de fumar. E, como o ato de fumar for definido, sua alternativa também precisa ser determinada. É fumar nada, fumar menos, fumar outra coisa? Igualmente importante, mesmo que fumar e sua alternativa sejam ambos definidos explicitamente, o fumo não causará câncer em todas as pessoas. Então, quem é suscetível a esse efeito do fumo? Ou, para colocar em outros termos, quais são os outros componentes da constelação causal que atuarão com o fumo para produzir câncer de pulmão nessa comparação?

A Figura 2.2 exibe um diagrama esquemático de três causas suficientes que poderia ser completado durante o seguimento de um indivíduo. As três condições ou eventos – A, B e E – foram definidas como variáveis binárias, de modo que só podem assumir os valores 0 ou 1. Com o código A usado na figura, sua categoria de referência,  $A = 0$ , algumas vezes é causador, mas sua categoria de índice,  $A = 1$ , nunca o é. Essa situação ocorre porque duas causas suficientes contêm uma causa componente rotulada “ $A = 0$ ”, mas nenhuma causa suficiente contém uma causa componente rotulada “ $A = 1$ ”. Um exemplo de uma condição ou evento desse tipo poderia ser  $A = 1$ , tomar um suplemento multivitamínico diariamente, e  $A = 0$ , não tomar suplemento de vitamina. Com os códigos de B e E usados no exemplo ilustrado pela Figura 2.2, as categorias de índice,  $B = 1$  e  $E = 1$ , algumas vezes são causadores, mas suas categorias de referência,  $B = 0$  e  $E = 0$ , nunca o são. Para cada variável, as categorias de índice e de referência podem representar apenas dois estados ou eventos alternativos dentre muitas possibilidades. Assim, o código B pode ser  $B = 1$ , fumar 20 cigarros por dia durante 40 anos, e  $B = 0$ , fumar 20 cigarros por dia durante 20 anos, seguidos por 20 anos sem fumar. E poderia ser codificado  $E = 1$ , viver em uma área urbana com renda média baixa e com alta desigualdade de renda, e  $E = 0$ , morar em uma área urbana com renda média alta e com baixa desigualdade de renda.

$A = 0$ ,  $B = 1$  e  $E = 1$  são causas componentes individuais das causas suficientes na Figura 2.2.  $U_1$ ,  $U_2$  e  $U_3$  representam conjuntos de causas componentes.  $U_1$ , por exemplo, é o conjunto de todos os outros componentes que não sejam  $A = 0$  e  $B = 1$  necessários para completar a primeira causa suficiente na Figura 2.2. Se decidíssemos não especificar  $B = 1$ , então  $B = 1$  se tornaria parte do conjunto de componentes que são causalmente complementares a  $A = 0$ ; em outras palavras,  $B = 1$  seria então absorvida por  $U_1$ .

Cada uma das três causas suficientes representadas na Figura 2.2 é minimamente suficiente para produzir a doença no indivíduo. Isto é, somente um desses mecanismos precisa ser completado para a doença ocorrer (suficiência), e não há causa componente supérflua em nenhum mecanismo (minimidade) – cada componente é uma parte necessária daquele mecanismo causal específico. Uma causa componente específica pode desempenhar um papel em um, em vários ou em todos os mecanismos causais. Como observado, uma causa componente que aparece em todas as causas suficientes é designada causa *necessária* do desfecho. Por exemplo, a infecção por HIV é um componente de todas as causas suficientes da síndrome de imunodeficiência adquirida (Aids), e, portanto, é uma causa necessária da Aids. Tem sido sugerido que tais causas sejam denominadas “universalmente necessárias”, em reconhecimento de que cada componente de uma causa suficiente é necessário para que a causa suficiente (mecanismo) ocorra (Poole, 2001a).

A Figura 2.2 não ilustra aspectos do processo causal, tais como sequência ou tempo de ação das causas componentes, dose ou outras complexidades. Estes podem ser especificados na descrição do contraste de categorias de índice e de referência que definem cada causa componente. Assim, se o desfecho é câncer de pulmão e o fator B representa fumar cigarros, ele pode ser definido mais explicitamente como fumar pelo menos 20 cigarros sem filtro por dia durante pelo menos 40 anos, começando na idade de 20 anos ou antes ( $B = 1$ ), ou fumar 20 cigarros sem filtro por dia durante pelo menos 20 anos\*, começando na idade de 20 anos ou antes, e depois não fumar cigarros pelos próximos 20 anos ( $B = 0$ ).



**FIGURA 2.2** • Três classes de causas suficientes de uma doença (causas suficientes I, II e III, da esquerda para a direita).

\* N. de T.: Durante pelo menos 20 anos não no original, mas a lógica indica que o correto é assim.

Na especificação de uma causa componente, os dois lados do contraste causal do qual ela é composta devem ser definidos de acordo com escolhas ou opções realistas. Se prescrever um placebo não for uma opção terapêutica realista, um contraste causal entre um tratamento novo e um placebo, em um ensaio clínico, pode ser questionado por sua relevância duvidosa em relação à prática médica. De maneira semelhante, antes de se dizer que os contraceptivos orais aumentam o risco de morte ao longo de 10 anos (p. ex., por infarto do miocárdio ou por acidente vascular cerebral), devemos considerar qual é a alternativa. Se a alternativa for engravidar, então o risco de morte inerente ao parto pode ser maior do que o risco de usar contraceptivos orais, fazendo desses medicamentos uma prevenção, em vez de uma causa. Se a alternativa for um método contraceptivo igualmente eficaz, sem efeitos colaterais sérios, então os contraceptivos orais podem ser descritos como causa de morte.

Para compreender a prevenção em um esquema de causa suficiente componente, postulamos que a condição alternativa (na qual uma causa componente esteja ausente) previne o desfecho relativo à presença da causa componente. Assim, um efeito preventivo de um fator é representado ao se especificar sua causa alternativa como uma causa componente. Um exemplo é a presença de  $A = 0$  como uma causa componente nas duas primeiras causas mostradas na Figura 2.2. Um outro exemplo seria definir uma variável  $F$  (não representada na Fig. 2.2), como “com vacinação ( $F = 1$ ) ou sem vacinação ( $F = 0$ )”. A prevenção da doença por meio de vacinação ( $F = 1$ ) seria expressa no modelo de causa suficiente componente como causação da doença por não se vacinar ( $F = 0$ ). Essa ilustração não é problemática porque, uma vez que ambos os lados de um contraste causal tenham sido especificados, causação e prevenção são meramente dois lados da mesma moeda.

Sheps (1958) uma vez perguntou: “Nós devemos contar os vivos ou os mortos?”. A morte é um evento, mas a sobrevivência não é. Assim, para usar o modelo de causa suficiente componente, devemos contar os mortos. Essa restrição ao modelo pode ter implicações substanciais. Por exemplo, algumas medidas e fórmulas aproximam-se de outras somente quando o desfecho é raro. Quando a sobrevida é rara, a morte é comum. Nesse caso, o uso do modelo causa suficiente componente para prover de informação a análise nos impedirá de tirar vantagem das aproximações de desfecho raro.

Similarmente, as etiologias de desfechos adversos à saúde, que são condições ou estados, mas não eventos, devem ser ilustradas no modelo causa suficiente pela reversão do código do desfecho. Considere a espinha bífida, que é uma falha de fechamento total do tubo neural durante a gestação. Não há um ponto no tempo no qual se possa dizer que a espinha bífida tenha ocorrido. Seria complicado definir o “tempo de incidência” da espinha bífida como a idade gestacional, na qual o fechamento completo do tubo neural geralmente ocorre. O modelo de causa suficiente componente seria mais adequado para esse caso ao definir o evento de fechamento completo (sem espinha bífida) como o desfecho, e enxergar as condições, os eventos ou as características que impeçam esse evento benéfico como causas da condição adversa, espinha bífida.

## Probabilidade, risco e causas

Na linguagem do cotidiano, “risco” frequentemente é usado como um sinônimo para probabilidade. Também é comumente usado como sinônimo de “perigo”, como em “viver perto de uma usina de energia nuclear é um risco que deve ser evitado”. Infelizmente, no linguajar epidemiológico, e mesmo na literatura erudita, “risco” é usado frequentemente para muitos conceitos distintos: taxa, razão de taxas, risco relativo, chances\* de incidência, prevalência, etc. A definição mais específica e, portanto, a mais útil, de *risco* é “probabilidade de um evento ocorrer durante um período de tempo especificado”.

O termo *probabilidade* tem múltiplos significados. Um deles é frequência relativa de um evento; outro é que probabilidade é a tendência, ou a propensão, de uma entidade produzir um evento. Um terceiro significado é que a probabilidade mede o grau de certeza de que um evento ocorrerá. Quando

\* N. de T.: Em inglês, *odds*, aqui traduzido como “chances”, é a razão da probabilidade de ocorrência de um evento pela probabilidade de não ocorrência. Por exemplo, se as chances são iguais a 2, significa que a probabilidade do evento ocorrer é 2 vezes a probabilidade dele não ocorrer.

se diz que “a probabilidade de morte em acidentes automobilísticos quando viajando a  $>120$  km/h é alta”, o que se quer dizer é que a proporção de acidentes que geram óbitos é mais elevada quando envolve veículos se deslocando a  $>120$  km/h, do que quando envolve veículos se deslocando em velocidades mais baixas (uso da frequência), que os acidentes com veículos em alta velocidade têm uma tendência maior de resultar em mortes (uso da propensão) do que os acidentes com veículos em velocidade mais baixa, ou que aquele que fala está mais seguro de que um óbito acontecerá em um acidente em alta velocidade do que em um acidente em velocidade mais baixa (uso da certeza).

O uso da frequência para “probabilidade” e para “risco”, ao contrário dos usos da propensão e da certeza, não admite significado para a noção de “risco” de um indivíduo, além da frequência relativa de 100% de um evento ocorrer e de 0% de ele não ocorrer. Essa restrição de riscos individuais de 0 ou 1 só pode ser relaxada para permitir valores intermediários, pela reinterpretação de tais declarações como a frequência com que o desfecho seria observado em amostra aleatória de uma população muito grande de pessoas, que se julga ser “semelhante” ao indivíduo de alguma forma (p. ex., idade, sexo e hábito de fumar). Se essa interpretação for aceita, quer alguma amostragem real tenha sido conduzida ou não, a noção de risco individual é substituída pela noção de frequência do evento em questão na população grande da qual o indivíduo foi amostrado. Com essa visão, o risco mudará conforme o modo pelo qual agruparmos os indivíduos para avaliar frequências. Um julgamento subjetivo entrará no quadro, inevitavelmente, na decisão sobre quais características usar para a formação de grupos. Por exemplo, deve-se levar em consideração o consumo de tomates para definir a classe de homens que são “semelhantes” a um determinado sujeito, com o propósito de determinar o risco de um diagnóstico de câncer de próstata entre seu 60º e 70º aniversários? Em caso afirmativo, qual estudo ou metanálise devem ser considerados nesse fragmento de informação?

A menos que encontremos um conjunto de condições e eventos no qual a doença não ocorra absolutamente, é sempre uma hipótese de trabalho razoável que alguns componentes causais permaneçam desconhecidos, não importa o quanto se saiba sobre a etiologia de uma doença. Podemos estar inclinados a atribuir um risco igual para todos os indivíduos cuja situação para alguns componentes seja conhecida e idêntica. Podemos dizer, por exemplo, que os homens que são grandes fumantes de cigarros têm um risco ao longo da vida de aproximadamente 10% de desenvolver câncer de pulmão. Alguns interpretam essa declaração como significando que todos os homens estariam sujeitos a uma probabilidade de 10% de câncer de pulmão caso se tornassem grandes fumantes, como se a ocorrência de câncer de pulmão, além do tabagismo, fosse puramente uma questão do acaso. Esse ponto de vista é indefensável. Uma probabilidade pode ser 10% condicional a um tipo de informação, e mais alta ou mais baixa do que 10%, se também condicionarmos outras informações relevantes. Por exemplo, seria possível dizer que homens que são grandes fumantes e que trabalharam por muitos anos em ocupações com níveis historicamente altos de exposição a fibras de asbestos no ar teriam um risco de câncer de pulmão ao longo da vida relativamente mais alto do que 10%.

Independentemente de interpretarmos probabilidade como frequência relativa ou como grau de certeza, a atribuição de riscos iguais reflete meramente o agrupamento em particular. Em nossa ignorância, o melhor que podemos fazer na avaliação de risco é classificar as pessoas de acordo com indicadores de risco mensurados, e então atribuir o risco médio observado em uma classe às pessoas pertencentes à mesma classe. Conforme o conhecimento ou a especificação de indicadores de risco adicional se expandem, as estimativas de risco atribuídas a pessoas se afastarão da média, de acordo com a presença ou a ausência de outros fatores que predizem o desfecho.

## Força dos efeitos

O modelo causal exemplificado pela Figura 2.2 pode facilitar a compreensão de alguns conceitos-chave, tais como *força do efeito* e *interação*. Como uma ilustração de força do efeito, a Tabela 2.1 exibe a frequência de oito padrões possíveis de exposição a A, B e E em duas populações hipotéticas. Agora, os gráficos de setores na Figura 2.2 ilustram classes de mecanismos. O primeiro, por exemplo, representa todas as causas suficientes que, não importa quais outras causas componentes elas possam

TABELA 2.1

Frequências de exposição e riscos individuais em duas populações hipotéticas de acordo com as combinações possíveis das três causas componentes especificadas na Figura 2.2

Exposições			Causa suficiente completada	Risco	Frequência do padrão de exposição	
A	B	E			População 1	População 2
1	1	1	III	1	900	100
1	1	0	Nenhuma	0	900	100
1	0	1	Nenhuma	0	100	900
1	0	0	Nenhuma	0	100	900
0	1	1	I, II, ou III	1	100	900
0	1	0	I	1	100	900
0	0	1	II	1	900	100
0	0	0	Nenhuma	0	900	100

conter, têm em comum o fato de que incluem  $A = 0$  e  $B = 1$ . Os constituintes de  $U_1$  podem, e geralmente o fazem, diferir de indivíduo para indivíduo. Para simplificar, podemos supor, de modo um tanto irrealista, que  $U_1$ ,  $U_2$  e  $U_3$  sempre estão presentes, ou têm ocorrido para todos, e que a Figura 2.2 representa todas as causas suficientes.

Sob essas suposições, a resposta de cada indivíduo ao padrão de exposição em uma dada linha pode ser encontrada na coluna de resposta. A resposta aqui é o risco de desenvolver uma doença durante um período de tempo especificado, que é o mesmo para todos os indivíduos. Para simplificar, é empregado um modelo de risco determinístico, no qual os riscos individuais só podem ser iguais aos valores 0 ou 1 e a nenhum valor intermediário. Um modelo estocástico de risco individual relaxaria essa restrição e permitiria que os riscos individuais ficassem entre 0 e 1.

A proporção que contrai a doença, ou a proporção de incidência, em qualquer subpopulação na Tabela 2.1, pode ser encontrada por meio da soma do número de pessoas em cada padrão de exposição com um risco individual de 1 dividida pelo tamanho da subpopulação. Por exemplo, se a exposição A não for considerada (p. ex., se ela não foi mensurada), o padrão de proporções de incidência na população 1 seria aquele na Tabela 2.2.

Em um exemplo de como as proporções na Tabela 2.2 foram calculadas, revisemos como a proporção de incidência entre pessoas na população 1, com  $B = 1$  e  $E = 0$ , foi calculada: havia 900 pessoas com  $A = 1$ ,  $B = 1$  e  $E = 0$ ; nenhuma se tornou caso porque não há causas suficientes que possam culminar na ocorrência da doença durante o período do estudo em pessoas com essa combinação de

TABELA 2.2

Proporções de incidência (PI) para combinações de causas componentes B e E na população hipotética 1, pressupondo-se que a causa componente A não foi mensurada

	B = 1, E = 1	B = 1, E = 0	B = 0, E = 1	B = 0, E = 0
Casos	1.000	100	900	0
Total	1.000	1.000	1.000	1.000
PI	1,00	0,10	0,90	0,00



categorias de exposição. (Há duas causas suficientes que contêm  $B = 1$  como uma causa componente, mas uma delas contém a causa componente  $A = 0$  e a outra contém a causa componente  $E = 1$ . A presença de  $A = 1$  ou  $E = 0$  bloqueia esses mecanismos etiológicos.) Houve 100 pessoas com  $A = 0$ ,  $B = 1$  e  $E = 0$ ; todas se tornaram casos porque tinham  $U_1$ , o conjunto de complementos causais para a classe de causas suficientes, contendo  $A = 0$  e  $B = 1$ . Assim, entre todas as 1.000 pessoas com  $B = 1$  e  $E = 0$ , houve 100 casos resultando em uma proporção de incidência de 0,10.

Se fôssemos mensurar a força do efeito pela diferença das proporções de incidência, a Tabela 2.2 sugere que, para a população 1,  $E = 1$  tem um efeito muito mais forte do que  $B = 1$ , porque  $E = 1$  parece aumentar a proporção de incidência em 0,9 (em ambas as categorias de  $B$ ), ao passo que  $B = 1$  eleva a proporção de incidência em apenas 0,1 (em ambas as categorias de  $E$ ). A Tabela 2.3 mostra os resultados análogos para a população 2. Embora os membros dessa população tenham exatamente os mesmos mecanismos causais da população 1 operando entre eles, as forças relativas dos fatores causais  $E = 1$  e  $B = 1$  parecem invertidas, usando novamente a diferença de proporções de incidência como a medida de força.  $B = 1$  parecem ter agora um efeito muito mais forte sobre as proporções de incidência do que  $E = 1$ , apesar do fato de  $B$  e  $E$  não estarem associados em qualquer população, como é evidente a partir da prevalência 50% de  $E = 1$  entre aqueles com  $B = 1$  e entre aqueles com  $B = 0$ . O efeito de  $E$ , mas não de  $B$ , é estimável a partir das Tabelas 2.2 e 2.3. O efeito de  $B$  não é estimável porque ele é confundido por  $A$ , mas nós ignoraremos isto aqui. O tópico de confundimento será introduzido no Capítulo 4.

A diferença geral de proporções de incidência contrastando  $E = 1$  e  $E = 0$  é  $(1.900/2.000) - (100/2.000) = 0,9$ , na população 1, e  $(1.100/2.000) - (900/2.000) = 0,1$ , na população 2. A diferença-chave entre as populações 1 e 2 é a diferença na prevalência das categorias sob as quais  $E = 1$  age para aumentar o risco; isto é, a presença de  $A = 0$  ou  $B = 1$ , mas não de ambas. (Quando  $A = 0$  e  $B = 1$ , a primeira das três causas suficientes na Figura 2-2 já é completa. Quando  $E = 1$ , a segunda e a terceira também são completas, mas como o risco já é 1,  $E = 1$  não aumenta o risco de ninguém). A prevalência da condição, “ $A = 0$  ou  $B = 1$ , mas não ambas” é  $1.800/2.000 = 90\%$  em ambas as categorias de  $E$ , na população 1. Na população 2, essa prevalência é de apenas  $200/2.000 = 10\%$  em ambas as categorias de  $E$ . Essa diferença na prevalência das categorias, suficiente para  $E = 1$  aumentar o risco, explica a diferença na força do efeito de  $E = 1$ , conforme mensurada pela diferença nas proporções de incidência.

Conforme observado, o conjunto de todas as outras causas componentes em todas as causas suficientes, no qual um fator causal participa, é chamado de *complemento causal* do fator. Assim, no exemplo supracitado,  $A = 0$ ,  $B = 1$ ,  $U_2$  e  $U_3$  compõem o complemento causal de  $E = 1$ . Esse exemplo mostra que a força do efeito de um fator sobre a ocorrência de uma doença em uma população, mensurada como a diferença absoluta nas proporções de incidência, depende da prevalência de seu complemento causal. Um fator terá um efeito forte quando mensurado pela diferença de proporções dos que adquirem a doença se seu complemento causal for comum. Em caso contrário, um fator com um complemento causal raro terá um efeito fraco.

**TABELA 2.3**

**Proporções de incidência (PI) para combinações de causas componentes B e E na população hipotética 2, pressupondo-se que a causa componente A não foi mensurada**

	B = 1, E = 1	B = 1, E = 0	B = 0, E = 1	B = 0, E = 0
Casos	1.000	900	100	0
Total	1.000	1.000	1.000	1.000
PI	1,00	0,90	0,10	0,00

Se a força do efeito for mensurada pela razão de proporções de adoecer, em oposição à diferença, então a força dependerá de mais do que o complemento causal de um fator. Particularmente, a força depende também do quão comuns ou raros são os componentes de causas suficientes nas quais o fator causal especificado *não* desempenha um papel. Nesse exemplo, ainda, o efeito de  $E = 1$ , mensurado em termos de razão, depende da prevalência do complemento causal de  $E = 1$  e da prevalência da conjunção de  $A = 0$  e  $B = 1$ . Se muitas pessoas têm tanto  $A = 0$  quanto  $B = 1$ , a proporção de incidência basal (i.e., a proporção de pessoas não E, ou “não expostas”, adquirindo a doença) será alta, e a proporção adquirindo a doença devido a E será comparativamente baixa. Se poucas pessoas tiverem tanto  $A = 0$  quanto  $B = 1$ , a proporção de incidência de basal será baixa, e a proporção adquirindo a doença devido a  $E = 1$  será comparativamente alta. Assim, a força do efeito, mensurada pela razão das proporções de incidência, depende de mais condições do que apenas da força do efeito, medida pela diferença da proporção de incidência.

Independentemente de como a força do efeito de um fator causal for mensurada, o significado em saúde pública daquele efeito não implica num grau correspondente de significância etiológica. Cada causa componente em uma dada causa suficiente tem a mesma significância etiológica. Dado um mecanismo causal específico, qualquer das causas componentes pode ter efeitos fortes ou fracos, usando-se a medida de diferença ou a de razão. As identidades reais dos componentes de uma causa suficiente fazem parte da mecânica da causalidade, ao passo que a força do efeito de um fator depende da distribuição temporal específica de seu complemento causal (se a força for mensurada em termos absolutos) mais a distribuição dos componentes de todas as causas suficientes nas quais o fator não desempenha um papel (se a força for mensurada em termos relativos). Durante um período de tempo, a força do efeito de um dado fator sobre a ocorrência de doença pode mudar, porque a prevalência de seu complemento causal em vários mecanismos também pode variar, mesmo que os mecanismos causais nos quais o fator e seus cofatores atuam permaneçam inalterados.

## Interação entre causas

Duas causas componentes atuando na mesma causa suficiente podem ser definidas como *interagindo causalmente* para produzir doença. Essa definição deixa em aberto muitos mecanismos possíveis para a interação, inclusive aqueles nos quais dois componentes interagem de um modo físico direto (p. ex., duas drogas que reagem para formar um subproduto tóxico), e aqueles nos quais um componente (o *iniciador* do par) altera um substrato de modo que o outro componente (o *promotor* do par) possa agir. Entretanto, essa definição exclui qualquer situação na qual um componente E seja meramente uma causa de um outro componente F, sem efeito algum de E sobre a doença, exceto por meio do componente F que ele causa.

Atuar na mesma causa suficiente não é o mesmo que uma causa componente agindo para produzir uma segunda causa componente, e esta, então, continuar para produzir a doença (Robins e Greenland 1992, Kaufman et al., 2004). Em um exemplo da distinção, se fumar cigarros (*versus* nunca fumar) é uma causa componente da aterosclerose, e aterosclerose (*versus* sem aterosclerose) causa infarto do miocárdio, tanto o fumo quanto a aterosclerose seriam causas componentes (cofatores) em certas causas suficientes de infarto do miocárdio. As causas componentes não necessariamente apareceriam na mesma causa suficiente. Em vez disso, para uma causa suficiente, envolvendo aterosclerose como uma causa componente, haveria uma outra causa suficiente na qual a causa componente aterosclerose seria substituída por todas as causas componentes que levam à aterosclerose, inclusive o fumo. Assim, uma relação causal sequencial entre tabagismo e aterosclerose não seria o bastante para que estes interajam sinergicamente na etiologia do infarto do miocárdio, como causa suficiente. Em vez disso, a sequência causal significa que o fumo pode agir indiretamente, por meio da aterosclerose, para provocar infarto do miocárdio.

Agora suponhamos que, talvez em adição ao mecanismo citado, o fumo reduza o tempo de coagulação e cause, assim, trombos que bloqueiam as artérias coronárias, se elas estiverem estreitadas



pela aterosclerose. Esse mecanismo seria representado por uma causa suficiente contendo tanto o fumo quanto a aterosclerose como componentes e constituiria, dessa forma, uma interação sinérgica entre tabagismo e aterosclerose para causar infarto do miocárdio. A presença de uma causa suficiente, contudo, não nos diria se o fumo também contribuiu para o infarto do miocárdio por causar aterosclerose. Assim, o modelo básico de causa suficiente não nos alerta para efeitos indiretos (efeitos de algumas causas componentes mediados por outras causas componentes no modelo). Os Capítulos 4 e 12 introduzem modelos potencial-desfecho e gráficos mais adequados para exibir os efeitos indiretos e os mecanismos sequenciais mais gerais, ao passo que o Capítulo 5 discute em detalhe a interação como definida no arcabouço potencial-desfecho, e sua relação com a interação definida no modelo causa suficiente.

## Proporção de doença devido a causas específicas

Na Figura 2.2, presumindo-se que as três causas suficientes no diagrama sejam as únicas operando, qual fração da doença é causada por  $E = 1$ ?  $E = 1$  é uma causa componente da doença em dois dos mecanismos de causa suficiente, II e III, de modo que toda doença que surgir de qualquer desses mecanismos seja atribuível a  $E = 1$ . Note que, em pessoas com o padrão de exposição  $A = 0$ ,  $B = 1$ ,  $E = 1$ , todas as três causas suficientes seriam completadas. O primeiro dos três mecanismos a ser completado seria aquele que realmente produz um certo caso. Se o primeiro completado fosse o mecanismo II ou o III, o caso seria atribuível causalmente a  $E = 1$ . Entretanto, se o mecanismo I fosse o primeiro completado,  $E = 1$  não seria parte da causa suficiente produzindo aquele caso. Sem saber os tempos em que os três mecanismos foram completados entre pessoas com o padrão de exposição  $A = 0$ ,  $B = 1$ ,  $E = 1$ , não podemos dizer quantos dos 100 casos na população 1, ou dos 900 na população 2, são atribuíveis etiologicamente a  $E = 1$ .

Cada um dos casos, cuja etiologia é atribuível a  $E = 1$ , também pode ser atribuído às outras causas componentes nos mecanismos causais em que  $E = 1$  atua. Cada causa componente interage com seus fatores complementares para produzir doença, de modo que cada caso da enfermidade pode ser atribuído a todas as causas componentes na causa suficiente completada. Observe, porém, que as frações atribuíveis adicionadas por meio de causas componentes da mesma doença não somam 1, embora haja uma tendência errônea a pensar que sim. Para ilustrar o erro nessa tendência, note que uma causa componente necessária aparece em cada causa suficiente completada da doença, e assim, por si própria, tem uma fração atribuível de 1, sem contar as frações atribuíveis para outras causas componentes. Posto que todo caso da doença pode ser atribuído a toda causa componente em seu mecanismo causal, as frações atribuíveis para diferentes causas componentes geralmente somarão mais do que 1, e não há limite superior para essa soma.

Um debate recente com relação à proporção de fatores de risco para cardiopatia coronária, atribuível a causas componentes em particular, ilustra o tipo de erros em inferência que podem surgir quando se pensa que a soma é restrita a 1. O debate gira em torno de ser a proporção de doença cardíaca coronariana atribuível ao colesterol sanguíneo elevado, à hipertensão arterial e ao hábito de fumar cigarros igual a 75 ou “somente a 50%” (Magnus e Beaglehole, 2001). Se for verdadeira a primeira hipótese, então, de acordo com alguns argumentos, a pesquisa por causas adicionais seria de utilidade limitada (Beaglehole e Magnus, 2002), porque apenas 25% dos casos “permanecem sem explicação”. Ao pressupor que a proporção explicada por causas componentes ainda desconhecidas não pode exceder a 25%, aqueles que apoiam essa opinião deixam de reconhecer que casos provocados por uma causa suficiente, que contenha qualquer subconjunto das três supostas causas, também poderiam conter causas componentes desconhecidas. Casos provenientes de causas suficientes com esse conjunto de causas componentes superpostas poderiam ser prevenidos por intervenções direcionadas às três causas citadas, ou por intervenções relacionadas às causas ainda desconhecidas, quando forem identificadas. As últimas intervenções poderiam reduzir o ônus da doença em muito mais do que 25%.

Em outro exemplo, em uma coorte de fumantes de cigarro expostos a arsênico por trabalhar em uma fundição, uma taxa de câncer de pulmão estimada em 75% foi atribuível ao ambiente ocupacio-

nal e de 65% a seu hábito de fumar (Pinto et al., 1978; Hertz-Picciotto et al., 1992). Não há problemas com tais cifras, que simplesmente refletem a etiologia multifatorial da doença. Assim como com a cardiopatia coronariana: se 75% são atribuíveis ao colesterol sanguíneo alto, à hipertensão e ao fumar cigarros, 100% ainda podem ser atribuíveis a outras causas, conhecidas, suspeitas e ainda a serem descobertas. Algumas dessas causas participarão nos mesmos mecanismos causais como colesterol elevado, pressão arterial alta e fumo de cigarros. Beaglehole e Magnus estavam corretos em pensar que se as três causas componentes especificadas se combinassem para explicar 75% das doenças cardiovasculares (DCV), e se nós, de algum modo, as eliminássemos, haveria somente 25% de casos de DCV remanescentes. Porém, até que os 75% fossem eliminados, qualquer componente novo descoberto poderia causar até 100% das DCV que temos atualmente.

A noção de que intervenções direcionadas a colesterol sanguíneo alto, hipertensão e tabagismo pudessem eliminar 75% da cardiopatia coronariana é irreal, dadas as estratégias de intervenção atualmente disponíveis. Embora seja possível algum progresso para reduzir o efeito desses fatores de risco, é improvável que qualquer deles possa ser completamente erradicado de alguma população grande em curto prazo. Estimativas do efeito em saúde pública de eliminar as próprias doenças como causas de morte (Murray et al., 2002) são ainda mais remotas, porque deixam de levar em conta todos os efeitos das intervenções necessárias para se conseguir a eliminação da doença, inclusive efeitos colaterais imprevistos (Greenland, 2002a, 2005a).

O debate sobre atribuir cardiopatia coronariana a causas componentes traz reminiscências de uma discussão anterior com relação a causas de câncer. Em seu trabalho largamente citado, *The Causes of Cancer*, Doll e Peto (1981, Tabela 20) criaram uma tabela dando suas estimativas da fração de todos os cânceres causados por vários agentes. As frações somaram aproximadamente 100%. Embora os autores reconhecessem que qualquer caso poderia ser provocado por mais de um agente (o que significa que, com agentes suficientes, as frações atribuíveis somariam bem mais do que 100%), referira-se a essa situação como uma “dificuldade” e uma “anomalia”, que optaram por ignorar. Subsequentemente, um dos autores admitiu que a fração atribuível podia somar mais do que 100% (Peto, 1985). Não é uma dificuldade nem uma anomalia, nem algo que se possa ignorar com segurança, mas simplesmente uma consequência do fato de que nenhum evento tem um só agente como causa. A fração de doença que pode ser atribuída a causas conhecidas crescerá sem limites à medida que mais causas forem descobertas. Somente a fração de doença atribuível a uma causa componente isolada não pode ultrapassar 100%.

De modo semelhante, provocou repercussão um pronunciamento, em 1960, de que cerca de 90% dos casos de câncer têm causas ambientais (Higginson, 1960). Nesse caso, pensou-se em ambiente representando todas as causas componentes não genéticas, e assim incluiu-se não apenas o ambiente físico, mas também o ambiente social e todo o comportamento humano que não seja determinado geneticamente. Dessa forma, causas componentes ambientais devem estar presentes em alguma extensão em cada causa suficiente de uma doença. Portanto, a estimativa de Higginson de 90% foi subestimada.

Pode-se mostrar também que 100% de qualquer doença são herdados, mesmo quando fatores ambientais são causas componentes. MacMahon (1968) citou o exemplo dado por Hogben (1933) sobre canelas amarelas, um traço ocorrendo em certas linhagens de aves alimentadas com milho amarelo. Tanto um conjunto específico de genes como uma dieta de milho amarelo são necessários para produzir canelas amareladas. Um fazendeiro com várias linhagens de aves que alimentasse todas apenas com milho amarelo consideraria as canelas amarelas como uma condição genética, porque somente uma linhagem ficou com canelas daquela cor, apesar de todas as linhagens receberem a mesma dieta. Um outro fazendeiro que possuísse apenas a linhagem capaz de ficar com as canelas amarelas, mas que alimentasse algumas das aves com milho amarelo e outras com milho branco, consideraria as canelas amarelas uma condição determinada ambientalmente, porque depende da dieta. Em seres humanos, o retardo mental causado pela fenilcetonúria é considerado por muitos puramente genético. Entretanto, esse retardo pode ser prevenido com sucesso pela intervenção dietética, o que demonstra a presença de uma causa ambiental. Na verdade, canelas amarelas, fenilcetonúria e outras condições

são determinadas por uma interação de genes e ambiente. Não faz sentido alocar uma porção da causalidade a genes ou ao ambiente separadamente quando ambos podem atuar juntos em causas suficientes.

No entanto, muitos pesquisadores têm comparado a ocorrência de doenças em gêmeos idênticos e não idênticos para estimar a fração de doença que é hereditária. Esses estudos e outros índices de hereditariedade avaliam apenas o papel relativo de causas ambientais e genéticas de doenças em um cenário particular. Por exemplo, algumas causas genéticas podem ser causas componentes necessárias de todo mecanismo causal. Se todos em uma população possuem um conjunto idêntico de genes que causam doença, entretanto, seu efeito não é incluído nos índices de hereditariedade, apesar do fato de que os genes são causas da doença. Os dois fazendeiros do exemplo ofereceriam valores muito diferentes quanto à hereditariedade das canelas amarelas, apesar de a condição ser sempre 100% dependente da presença de certos genes.

Cada caso de cada doença tem algumas causas componentes ambientais e algumas genéticas, e, portanto, cada caso pode ser atribuído tanto aos genes quanto ao ambiente. Nenhum paradoxo existe, contanto que seja compreendido que as frações de doença atribuíveis a genes e ao ambiente se superpõem. Assim, se debates sobre que proporção de todas as ocorrências de uma doença é genética e que proporção é ambiental presumirem que as cotas devem somar 100%, serão falaciosos e provocarão distração quanto a pesquisas mais úteis.

Em um nível mais geral, a questão sobre uma dada doença ter ou não uma “etiologia multifatorial” pode ser respondida afirmativamente de uma vez por todas. Todas as doenças têm etiologias multifatoriais. Portanto, é completamente indigno de nota, para uma dada doença, ter tal etiologia, e não se deve gastar tempo ou dinheiro em pesquisas que tentem responder se uma doença em particular tem, ou não, uma etiologia multifatorial; todas têm. O propósito da pesquisa etiológica é identificar componentes de tais etiologias.

## Período de indução

Gráficos de setores de causas suficientes e suas componentes, tais como aqueles na Figura 2.2, não são adequados para fornecer um modelo conceitual para o *período de indução*, que pode ser definido como o período de tempo da ação causal até o início da doença. Não há como dizer, a partir de um gráfico de setores de uma causa suficiente, quais componentes afetam um ao outro, quais componentes devem vir antes ou depois de outros, para quais componentes a ordem temporal é irrelevante, etc. A informação crucial sobre o ordenamento temporal deve ser fornecida em uma descrição separada das inter-relações entre os componentes de uma causa suficiente.

Se, na causa suficiente I, a sequência de ação das causas componentes especificadas deve ser  $A = 0$ ,  $B = 1$ , e estivermos estudando o efeito de  $A = 0$ , o qual (suponhamos) atua em um ponto no tempo estreitamente definido, não observamos o aparecimento da doença imediatamente após  $A = 0$  ocorrer. A doença só aparece depois que a sequência é completada, de modo que haverá um retardo enquanto  $B = 1$  ocorre (juntamente com os componentes do conjunto  $U_1$  que não estão presentes, ou que não ocorreram quando  $A = 0$  aconteceu). Quando  $B = 1$  atuar e se for a última de todas as causas componentes (inclusive aquelas no conjunto de condições e eventos não especificados representados por  $U_1$ ), a doença ocorrerá. O intervalo entre a ação de  $B = 1$  e a ocorrência da doença é o tempo de indução para o efeito de  $B = 1$  na causa suficiente I.

No exemplo dado anteriormente de um problema de equilíbrio levando a uma queda e a uma lesão do quadril, o tempo de indução entre o começo do problema e a lesão do quadril subsequente pode ser longo, se o problema de equilíbrio for causado por um traumatismo craniano antigo, ou curto, se for causado por embriaguez. No último caso, o tempo de indução pode até ser instantâneo, se o definirmos como álcool sanguíneo além de certo nível. Em geral, condições e eventos podem ser causas componentes, e duas ou mais causas componentes podem ter o mesmo tempo de indução, incluindo o zero.

Definir um período de indução de interesse equivale a especificar as características das causas componentes de interesse. Um exemplo claro de um tempo de indução prolongado é a relação causa-efeito entre a exposição de um feto feminino ao dietilestilbestrol (DES) e o desenvolvimento subsequente de adenocarcinoma vaginal. O câncer é diagnosticado, geralmente, entre as idades de 15 e 30 anos. Visto que a exposição causal ao DES ocorre no início da gravidez, há um tempo de indução de cerca de 15 a 30 anos para a ação carcinogênica do DES. Durante esse tempo, presumivelmente, outras causas estão operando; algumas evidências sugerem que ação hormonal durante a adolescência pode ser parte do mecanismo (Rothman, 1981).

É incorreto caracterizar a doença em si como tendo um período de indução longo ou curto. O tempo de indução pode ser conceituado apenas em relação a uma causa componente específica, operando em uma causa suficiente específica. Assim, dizemos que o tempo de indução relacionando o DES ao carcinoma de células claras da vagina é de 15 a 30 anos, mas não devemos dizer que 15 a 30 anos seja o tempo de indução para o carcinoma de células claras em geral. Como cada causa componente, em qualquer mecanismo causal, pode atuar em um tempo diferente das outras causas componentes, cada uma pode ter seu tempo de indução próprio. Para a causa componente que atua por último, o tempo de indução é igual a zero. Se outra causa componente do carcinoma de células claras da vagina que atuasse na adolescência fosse identificada, teria um tempo de indução para sua ação carcinogênica muito mais curto do que o do DES. Assim, o tempo de indução caracteriza um par causa-efeito específico, e não apenas o efeito.

Na carcinogênese, os termos *iniciador* e *promotor* têm sido usados para referir algumas das causas componentes de câncer que agem cedo e tarde, respectivamente, no mecanismo causal. O próprio câncer tem sido caracterizado, frequentemente, como um processo mórbido com longo tempo de indução. Essa caracterização, contudo, é um efeito equivocado, porque qualquer componente de ação tardia no processo causal, tal como um promotor, terá um tempo de indução curto. Na verdade, por definição, o tempo de indução será sempre zero para, pelo menos, uma causa componente, a última a atuar. A visão errônea de que as doenças, ao contrário das relações causa-doença, têm períodos de indução longos ou curtos pode ter implicações importantes para a pesquisa. Por exemplo, a visão dos cânceres de adultos como “doenças de latência longa” pode induzir alguns pesquisadores a ignorar evidências de efeitos etiológicos ocorrendo relativamente tarde no processo que culmina em cânceres diagnosticados clinicamente. No outro extremo, o desinteresse rotineiro por exposições ocorrendo na primeira ou na segunda década de vida em estudos de carcinogênese ocupacional, como um dos exemplos principais, pode ter inibido a descoberta de causas ocupacionais com períodos de indução muito longos.

A doença, uma vez iniciada, não é necessariamente aparente. O intervalo de tempo entre a ocorrência irreversível de doença e a detecção tem sido denominado *período latente* (Rothman, 1981), embora outros tenham usado esse termo como sinônimo de período de indução. Ainda outros usam *período latente* para designar o tempo total entre ação causal e detecção da doença. Usamos *período de indução* para descrever o tempo da ação causal para a ocorrência irreversível de doença, e *período latente* para significar o tempo da ocorrência da doença até sua detecção. Algumas vezes, o período latente pode ser reduzido por melhora dos métodos de detecção da doença. O período de indução, entretanto, não pode ser reduzido pela detecção precoce da doença, porque a ocorrência desta marca o fim do período de indução. A detecção mais precoce da doença, contudo, pode reduzir o período de indução aparente (o tempo entre ação causal e detecção da doença), porque o tempo, quando a doença é detectada, do ponto de vista prático, geralmente é utilizado para marcar o tempo de ocorrência da doença. Assim, doenças como cânceres de crescimento lento podem aparentar ter períodos de indução longos em relação a muitas causas, porque têm períodos latentes longos. O período latente, diferentemente do período de indução, é uma característica da doença e do esforço de detecção aplicado à pessoa com a doença.

Embora não seja possível reduzir o período de indução propriamente dito pela detecção mais precoce da doença, pode ser possível observar-se estágios intermediários de um mecanismo causal. O interesse aumentado em marcadores biológicos, tais como os adutos de DNA, é um exemplo da tentativa de focar causas mais próximas da ocorrência da doença, ou efeitos mais próximos à ocorrência

da causa. Tais marcadores biológicos podem, contudo, refletir os efeitos de agentes com atuação mais precoce sobre a pessoa.

Alguns agentes podem ter uma ação causal por encurtar o tempo de indução de outros agentes. Suponha que a exposição ao fator  $X = 1$  leve à epilepsia após um intervalo de 10 anos, em média. Pode ser que a exposição a uma droga,  $Z = 1$ , encurtasse esse intervalo para 2 anos.  $Z = 1$  está atuando como um catalisador, ou como uma causa, de epilepsia? A resposta é: ambos; um catalisador é uma causa. Sem  $Z = 1$ , a epilepsia ocorre 8 anos mais tarde do que com  $Z = 1$ , assim, podemos dizer que  $Z = 1$  causa o início de epilepsia precoce. Não é suficiente argumentar que a epilepsia teria ocorrido de qualquer maneira. Em primeiro lugar, ela não teria acontecido naquele momento, e o momento de ocorrência é parte da nossa definição de um evento. Em segundo, a epilepsia só ocorrerá mais tarde se o indivíduo sobreviver os 8 anos adicionais, o que não é certeza. O agente  $Z = 1$  não só determina quando a epilepsia ocorre, como também determina se ela ocorre. Assim, podemos chamar qualquer agente que atua como catalisador de um mecanismo causal, acelerando um período de indução para outros agentes, de causa em seu próprio direito. Da mesma forma qualquer agente que adie o início de um evento, extraíndo o período de indução para um outro agente, é um preventivo. Não seria surpreendente demais igualar adiamento a prevenção: nós, rotineiramente, usamos tal equação quando empregamos o eufemismo de que “prevenimos” a morte, a qual, na verdade, pode somente ser adiada. O que prevenimos é morte em um dado momento, em favor de morte mais tardiamente.

## Escopo do modelo

A principal utilidade desse modelo de causas suficientes e suas componentes está em sua capacidade de fornecer uma estrutura conceitual geral, mas prática, para problemas causais. A tentativa de fazer a proporção de doença atribuível a várias causas componentes somar 100% é um exemplo de uma falácia que é exposta pelo modelo (embora MacMahon e outros fossem capazes de invocar as canelas amarelas e a fenilcetonúria para expor tal falácia muito antes que o modelo de causas suficiente componente fosse descrito formalmente [MacMahon e Pugh, 1967, 1970]). O modelo torna claro que, por causa de interações, não há limite para a soma dessas proporções. Conforme veremos no Capítulo 5, a avaliação epidemiológica das próprias interações pode ser esclarecida, em alguma extensão, com o auxílio do modelo.

Embora o modelo pareça lidar qualitativamente com a ação das causas componentes, pode ser ampliado para incluir a dependência da dose, postulando um conjunto de causas suficientes, cada uma delas contendo como componente uma dose diferente do agente em questão. Doses pequenas podem requerer um grupo de causas complementares maior ou mais raro para completar uma causa suficiente em comparação a um requerido por doses grandes (Rothman, 1976a), caso em que é particularmente importante especificar ambos os lados do contraste causal. Dessa maneira, o modelo pode explicar o fenômeno de um período de indução mais curto acompanhando doses de exposição maiores, porque um conjunto menor de componentes complementares seria necessário para completar a causa suficiente.

Aqueles que acreditam que o acaso deve desempenhar um papel em qualquer mecanismo complexo poderiam se opor à confusão desse modelo aparentemente determinístico. Um modelo probabilístico (estocástico) poderia ser invocado para descrever uma relação dose-resposta, por exemplo, sem a necessidade de uma multidão de mecanismos causais diferentes. O modelo simplesmente relacionaria a dose de exposição à probabilidade do efeito ocorrer. Para aqueles que creem que praticamente todos os eventos contêm algum elemento do acaso, os modelos causais determinísticos podem parecer representar mal o indeterminismo do mundo real. Contudo, o modelo determinístico aqui apresentado pode incorporar o “acaso”; uma maneira seria encarar o acaso, ou pelo menos alguma parte da variabilidade que chamamos de “acaso”, como o resultado de eventos determinísticos que estão além dos limites atuais de conhecimento ou da possibilidade de observação.

Por exemplo, o desfecho de uma moeda que cai após ser arremessada, geralmente é considerado um evento do acaso. Na mecânica clássica, entretanto, o desfecho do lançamento pode ser

determinado completamente, em teoria, pela aplicação das leis da física e de uma descrição suficiente das condições iniciais. Para colocar isso em termos mais familiares para os epidemiologistas, considere a explicação pela qual um indivíduo adquire câncer de pulmão. Cem anos atrás, quando pouco se sabia sobre a etiologia do câncer de pulmão, um cientista poderia ter dito que era uma questão de acaso. Atualmente, poderíamos dizer que o risco depende de quanto o indivíduo fuma, o quanto ele foi exposto a asbestos e radônio, e assim por diante. No entanto, reconhecer essa dependência mexe com o limite da ignorância; não o elimina. Ainda se pode perguntar o que determina se um indivíduo que tenha fumado uma quantidade específica, e que tenha uma quantidade específica de exposição a todos os outros fatores de risco conhecidos, desenvolverá câncer de pulmão. Alguns o desenvolverão e outros não, e se todos os fatores de risco conhecidos já foram levados em consideração, o que sobrou ainda podemos descrever como acaso. É verdade que podemos explicar muito mais sobre a variabilidade na ocorrência de câncer de pulmão atualmente do que no passado, ao levar em consideração fatores causais conhecidos, porém, nos limites de nosso conhecimento, ainda atribuímos a variabilidade remanescente ao que chamamos de acaso. Dessa forma, o acaso é visto como um termo que cobre toda a nossa ignorância sobre explicações causais.

Até aqui, desprezamos considerações mais sutis sobre fontes de imprevisibilidade nos eventos, tais como o comportamento caótico (no qual até mesmo a menor dúvida sobre as condições iniciais leva a uma vasta incerteza sobre os desfechos) e a incerteza da mecânica quântica. Em cada uma dessas situações, um componente de modelo aleatório (estocástico) pode ser essencial para qualquer esforço útil de modelagem. Tais componentes também podem ser introduzidos no modelo conceitual anterior, tratando as causas componentes não mensuradas do modelo como eventos aleatórios, de modo que o modelo causal baseado em componentes de causas suficientes possa ter elementos aleatórios. Um exemplo é a alocação do tratamento em ensaios clínicos randomizados (Poole 2001a).

## Outros modelos de causalidade

O modelo de causa suficiente componente é apenas um de vários modelos de causação que podem ser úteis para se obter percepção sobre conceitos epidemiológicos (Greenland e Brumback, 2002; Greenland, 2004a). Ele retrata mecanismos causais qualitativos entre membros de uma população, de maneira que sua unidade fundamental de análise seja o mecanismo causal em vez de uma pessoa. Muitos conjuntos diferentes de mecanismos podem levar ao mesmo padrão de doença em uma população, de modo que o modelo de causa suficiente componente envolve especificação de detalhes que estão além do escopo de dados epidemiológicos. Além disso, o modelo não incorpora elementos que refletem distribuições de fatores ou sequências causais na população, que são cruciais para a compreensão de confundimento e de outros vieses.

Outros modelos de causação, tais como modelos potencial-desfecho (contrafactuais) e modelos gráficos, fornecem representações diretas de conceitos epidemiológicos, como confundimento e outros vieses, e podem ser aplicados a níveis de análise mecanicistas, individuais ou populacionais. Modelos potencial-desfecho (Caps. 4 e 5) especificam em detalhe o que aconteceria a indivíduos ou populações sob padrões alternativos possíveis de intervenções ou de exposições e também trazem à tona problemas na definição operacional de causas (Greenland, 2002a, 2005a; Hernán, 2005). Os modelos gráficos (Cap. 12) exibem amplas suposições qualitativas sobre direções causais e independências. Ambos os tipos de modelo têm relações íntimas com os modelos de equações estruturais que são populares nas ciências sociais (Pearl, 2000; Greenland e Brumback, 2002) e podem ser incluídos em uma teoria geral de causalidade longitudinal (Robins, 1997).

## FILOSOFIA DA INFERÊNCIA CIENTÍFICA

A inferência causal pode ser vista como um caso especial dentro do processo mais geral do raciocínio científico. A literatura sobre esse tópico é muito vasta para que possamos revisá-la minuciosamente,



mas forneceremos uma visão geral de certos pontos relevantes à epidemiologia, correndo o risco de algum excesso de simplificação.

## Indutivismo

A ciência moderna surgiu por volta dos séculos XVI e XVII, quando as demandas de conhecimento de tecnologias emergentes (tais como artilharia e navegação transoceânica) estimularam investigações sobre as origens do saber. Uma codificação inicial do método científico foi o *Novum Organum* de Francis Bacon, que, em 1620, apresentou uma visão indutivista da ciência. Essa filosofia diz que o raciocínio científico depende de generalizações, ou induções, a partir de observações das leis gerais da natureza; é dito que as observações induzem à formulação de uma lei natural na mente do cientista. Assim, um indutivista teria dito que a observação de Jenner sobre a ausência de varíola entre mulheres que ordenhavam vacas induziu na sua mente a teoria de que a varíola bovina (comum entre as ordenhadeiras) conferia imunidade à varíola. A filosofia indutivista atingiu uma espécie de pináculo nos cânones de John Stuart Mill (1862), que evoluíram para os critérios inferenciais ainda em voga hoje em dia.

A filosofia indutivista foi um grande passo adiante do escolasticismo medieval, pois demandava que um cientista fizesse observações cuidadosas sobre as pessoas e a natureza, em vez de apelar para a fé, aos textos arcaicos ou às autoridades. No entanto, o filósofo escocês David Hume, no século XVIII, descreveu uma deficiência perturbadora no indutivismo. Um argumento indutivo não carregava força lógica; em vez disso, tal argumento representava nada mais que uma *presunção* de que certos eventos seguiriam, no futuro, o mesmo padrão que haviam seguido no passado. Assim, argumentar que a varíola bovina causava imunidade contra a varíola, porque ninguém contraía varíola depois de ter varíola bovina, correspondia a uma presunção injustificável de que o padrão observado até então (nenhuma varíola após varíola bovina) continuaria durante o futuro. Hume enfatizou que, mesmo para as suposições que soam mais razoáveis, não há necessidade lógica por trás da argumentação indutiva. A preocupação central para Hume (1739) foi a inferência causal e a falha da indução em lhe fornecer uma base:

Assim, não apenas nossa razão nos falha quanto à descoberta da conexão final de causas e efeitos, mas mesmo depois que a experiência nos tenha informado de sua conjunção constante, é impossível para nós nos satisfazer conosco próprios pelo nosso raciocínio, porque nós deveríamos estender aquela experiência além daqueles exemplos particulares, que caíam sob nossa observação. Nós supomos, mas não somos capazes de provar, que deva haver uma semelhança entre aqueles objetos, dos quais nós tivemos experiência, e aqueles que ficam além do alcance de nossa descoberta.

Em outras palavras, nenhum número de repetições de uma sequência específica de eventos, como o aparecimento de luz após o acionar do interruptor, pode comprovar uma conexão causal entre o acionamento do interruptor e o acendimento da luz. Não importa quantas vezes a luz apareça depois de o interruptor ter sido pressionado, a possibilidade de ocorrência coincidente não pode ser excluída. Hume assinalou que observadores não podem perceber conexões causais, mas somente uma série de eventos. Bertrand Russell (1945) ilustrou esse ponto com o exemplo de dois relógios precisos, que soam perpetuamente no horário, com um deles mantendo o tempo levemente adiantado em relação ao outro. Embora um bata a hora invariavelmente antes do outro, não há uma conexão causal direta de um para o outro. Portanto, atribuir uma interpretação causal ao padrão de eventos não pode ser uma extensão lógica de nossas observações isoladamente, porque os eventos podem estar ocorrendo juntos devido apenas a uma causa anterior compartilhada ou algum erro sistemático nas observações.

A inferência causal baseada somente na mera associação de eventos constitui uma falácia lógica conhecida como *post hoc ergo propter hoc* (em latim, depois disso, portanto por causa disso). Essa

falácia é exemplificada pela inferência de que o canto de um galo é necessário para o sol se erguer, porque o nascer do sol sempre é precedido pelo canto do galo.

A falácia *post hoc* é um caso especial de uma falácia lógica mais geral conhecida como a *falácia de afirmar o consequente*, que toma o seguinte formato geral: sabemos que se H é verdadeiro, B deve ser verdadeiro; e sabemos que B é verdadeiro; portanto, H deve ser verdadeiro. Essa falácia é usada rotineiramente por cientistas na interpretação de dados. É usada, por exemplo, quando se argumenta o seguinte: se o serviço de esgoto causa doenças do coração, então as taxas de doença cardíaca devem ser mais elevadas onde ele está disponível; as taxas de cardiopatia são realmente mais altas onde existe serviço de esgoto; portanto, o serviço de esgoto causa doenças do coração. Aqui, H é a hipótese “serviço de esgoto causa doenças do coração” e B é a observação “as taxas de doença do coração são mais altas onde o serviço de esgoto está disponível”. O argumento logicamente é infundado, pelo fato de que podemos imaginar muitas maneiras pelas quais as suposições podem ser verdadeiras, mas as conclusões, falsas; por exemplo, o desenvolvimento econômico pode levar a frequências mais elevadas, tanto de serviços de esgoto quanto de cardiopatia, sem qualquer efeito do serviço de esgoto sobre as doenças do coração. Nesse caso, entretanto, também sabemos que uma das suposições não é verdadeira – especificamente, a suposição “se H é verdadeiro, B deve ser verdadeiro”. Essa forma particular de falácia exemplifica o problema do *confundimento*, que discutiremos detalhadamente em capítulos posteriores.

Bertrand Russell (1945) satirizou a falácia deste modo:

“Se p, então q; q é verdadeiro; portanto, p é verdadeiro”. Por exemplo, “se porcos têm asas, então alguns animais alados são bons para comer; agora, alguns animais alados são bons para comer; portanto, porcos têm asas”. Essa forma de inferência é chamada de “método científico”.

## Refutacionismo

Russell não estava sozinho em seu lamento sobre a falta de lógica do raciocínio científico praticado comumente. Muitos filósofos e cientistas a partir da época de Hume tentaram estabelecer uma base lógica segura para o raciocínio científico. Em 1920, os mais notáveis foram os da escola dos positivistas lógicos, que buscavam uma lógica para a ciência, que pudesse levar inevitavelmente a conclusões corretas em matemática. Outros filósofos e cientistas, contudo, tinham começado a suspeitar que hipóteses científicas nunca poderiam ser comprovadas ou estabelecidas como verdadeiras, em qualquer sentido lógico. Por exemplo, vários filósofos notaram que as declarações científicas só podem ser consideradas consistentes com a observação, mas não podem ser comprovadas ou desmentidas em qualquer lógica ou senso matemático “hermético” (Duhem, 1906, trad. 1954; Popper 1934, trad. 1959; Quine, 1951). Esse fato algumas vezes é chamado de problema da *não identificação* ou *indeterminação* de teorias por observações (Curd e Cover, 1998). Particularmente, as observações disponíveis são sempre consistentes com várias hipóteses que são, elas próprias, mutuamente inconsistentes, o que explica porque (como Hume notou) as teorias científicas não podem ser logicamente comprovadas. Especificamente, a consistência entre uma hipótese e as observações não é prova da hipótese, porque sempre podemos inventar hipóteses alternativas que são exatamente tão consistentes quanto as observações.

Em contraste, uma observação válida que é inconsistente com uma hipótese implica que a hipótese como declarada é falsa e, assim, refuta a hipótese. Se você torcer o pescoço do galo antes que ele cante e mesmo assim o sol nascer, você refutou que o canto do galo seja uma causa necessária para o nascer do sol. Alternativamente, considere um programa de pesquisa hipotético para aprender o ponto de ebulição da água (Magee, 1985). Um cientista que ferver água em um recipiente aberto e medir, repetidamente, o ponto de ebulição a 100°C, nunca provará, não importa quantas repetições confirmadoras estejam envolvidas, que 100°C sempre será o ponto de fervura. Em contrapartida, uma só tentativa de ferver água em uma vasilha fechada ou em grande altitude refutará a proposição de que a água sempre ferve a 100°C.



De acordo com Popper, a ciência avança por um processo de eliminação que chamou de “conjectura e refutação”. Os cientistas formam hipóteses baseadas em intuição, em conjectura e em experiência prévia. Bons cientistas usam lógica dedutiva para inferir previsões a partir da hipótese e, então, comparam observações com previsões. As hipóteses cujas previsões concordam com observações são confirmadas (Popper usou o termo “corroboradas”) apenas no sentido de que elas podem continuar a ser usadas como explicações de fenômenos naturais. Em qualquer momento, entretanto, elas podem ser refutadas por observações adicionais e ser substituídas por outras hipóteses que sejam mais consistentes com as observações. Essa visão da inferência científica às vezes é chamada de *refutacionismo* ou *falsificacionismo*. Os refutacionistas consideram a indução como sendo uma muleta psicológica: observações repetidas, de fato, não induzem à formulação de uma lei natural, mas apenas à crença de que tal lei tenha sido encontrada. Para um refutacionista, somente o conforto psicológico fornecido pela indução explica por que ela ainda tem defensores.

Uma maneira de resgatar o conceito de indução do estigma de pura ilusão é ressuscitá-lo como um fenômeno psicológico, como Hume e Popper acharam que a indução fosse, mas que desempenhe um papel legítimo na formulação de hipóteses. A filosofia de conjectura e refutação não impõe restrições à origem de conjecturas. Mesmo delírios são permitidos como hipóteses, portanto hipóteses indutivamente inspiradas e, embora psicológicas, são pontos de partida válidos para avaliação científica. Essa concessão não admite um papel lógico para a indução na confirmação de hipóteses científicas, mas permite que o processo de indução desempenhe um papel, junto com a imaginação, no ciclo científico de conjectura e refutação.

A filosofia de conjectura e refutação tem implicações profundas para a metodologia científica. O conceito popular de um cientista reunindo evidências de forma obstinada, para dar suporte a uma tese favorita, é passível de objeção do ponto de vista da filosofia refutacionista, porque encoraja os cientistas a considerarem suas próprias teorias favoritas como propriedade intelectual, a serem confirmadas, comprovadas e, quando todas as evidências estiverem presentes, moldadas em pedra e defendidas como lei natural. Tais atitudes dificultam a avaliação crítica, o intercâmbio e o progresso. A abordagem de conjectura e aprovação, em contraste, estimula os cientistas a considerarem múltiplas hipóteses e a buscar testes cruciais que decidam entre hipóteses antagônicas, tornando falsa uma delas. Como o objetivo é tornar falsa uma ou mais teorias, há incentivo para despersonalizá-las. Críticas dirigidas a uma teoria não precisam ser vistas como crítica à pessoa que a propôs. Tem sido sugerido que a razão pela qual certos campos da ciência avançam rapidamente enquanto outros fenecem, é que aqueles em progresso rápido são impulsionados por cientistas que estão ocupados construindo e testando hipóteses concorrentes; enquanto os outros “estão doentes, porque esqueceram a necessidade de hipóteses alternativas e de refutação” (Platt, 1964).

O modelo refutacionista da ciência tem várias lições valiosas para a condução da pesquisa, especialmente a necessidade de buscar explicações alternativas para as observações, em vez de focar a quimera de procurar uma “prova” científica para alguma teoria favorita. No entanto, ele é vulnerável a críticas de que as observações (ou, alguns diriam, suas interpretações) sejam, elas próprias, carregadas de teoria (algumas vezes chamada a *tese de Duhem-Quine*; Curd e Cover, 1998). Assim, observações nunca podem fornecer o tipo de refutações definitivas que constituem a marca das explicações populares de refutacionismo. Por exemplo, pode haver vieses não controlados, e até mesmo não imaginados, que tenham tornado inválidas nossas observações refutatórias; aclamar a refutação é assumir como verdadeira a teoria incomprovada de que nenhum viés como tal exista. Em outras palavras, não só as teorias são indeterminadas por observações, como também o são as refutações, as quais estão cheias de teorias. O resultado é que a certeza lógica sobre a verdade, ou sobre a falsidade, de uma teoria intrinsecamente consistente, é impossível (Quine, 1951).

## Consenso e naturalismo

Alguns filósofos de ciência do século XX, mais notadamente Thomas Kuhn (1962), enfatizaram o papel da comunidade científica no julgamento da validade de teorias científicas. Esses críticos do

modelo conjectura-e-refutação sugeriram que a refutação de uma teoria envolve a escolha de uma opção. Toda observação é dependente de teorias. Por exemplo, observar as luas de Júpiter através de um telescópio nos parece uma observação direta, mas somente porque a teoria da óptica sobre a qual se baseia o telescópio é tão bem aceita. Quando confrontado com uma observação refutatória, um cientista encara a escolha de rejeitar ou a validade da teoria sendo testada a validade da observação refutatória, a qual deve estar baseada em teorias científicas que não estejam corretas (Haack, 2003). Observações que estão falsificando exemplos de teorias podem, às vezes, ser tratadas como “anomalias”, toleradas sem falsificar a teoria, na esperança de que as anomalias possam, eventualmente, vir a ser explicadas. Um exemplo epidemiológico é a observação de que fumantes com inalação superficial tinham taxas de câncer de pulmão mais altas do que os que inalam profundamente. Essa anomalia foi explicada, finalmente, quando se observou que os tecidos de partes mais altas do pulmão são mais suscetíveis a tumores associados ao fumo, e que o alcatrão do fumo inalado de forma superficial tende a se depositar mais alto nos pulmões (Wald, 1985).

Em outros exemplos, as anomalias podem levar, eventualmente, à derrubada da doutrina científica corrente, assim como a mecânica de Newton foi deslocada (permanecendo apenas como aproximação de primeira ordem) pela teoria da relatividade. Kuhn afirmou que em cada ramo da ciência o ponto de vista científico prevalente, que ele denominou “ciência normal”, ocasionalmente, sofre desvios importantes que correspondem a revoluções científicas. Essas revoluções assinalam uma decisão da comunidade científica de descartar a infraestrutura científica, antes de falsificar uma nova hipótese que não pode ser enxertada facilmente nela. Kuhn e outros têm argumentado que o consenso da comunidade científica determina o que é considerado aceito e o que é refutado.

Os críticos de Kuhn caracterizaram essa descrição como processo irracional, “uma matéria para psicologia de massa” (Lakatos, 1970). Aqueles que acreditam em uma estrutura racional para a ciência consideram a visão de Kuhn uma descrição lamentavelmente real de muito daquilo que passa por atividade científica, mas não uma prescrição para qualquer boa ciência. Embora muitos filósofos modernos rejeitem demarcações rígidas e formulações para a ciência, como o refutacionismo, sustentam, contudo, que a ciência é fundamentada na razão, ainda que, possivelmente, no senso comum informal (Haack, 2003). Outros vão além de Kuhn e enfatizam que as tentativas de impor uma estrutura ou metodologia singular e racional à ciência tornam a imaginação deficiente e representam uma receita para a mesma espécie de repressão autoritária das ideias que os cientistas tiveram que enfrentar ao longo da história (Feyerabend, 1975 e 1993).

O debate filosófico sobre a descrição de Kuhn da ciência depende de ele ter pretendido descrever apenas o que tem acontecido historicamente na ciência ou o que deveria acontecer; assunto sobre o qual o próprio Kuhn (1970) não foi completamente claro:

Os [meus] comentários de Kuhn sobre o desenvolvimento científico [...] devem ser lidos como descrições ou prescrições? A resposta, é claro, é que eles devem ser lidos de ambas as maneiras ao mesmo tempo. Se eu tenho uma teoria sobre como e por que a ciência funciona, ela deve ter, necessariamente, implicações para o modo como os cientistas devem se comportar, para que seu empreendimento floresça.

A ideia de que a ciência é um processo sociológico, quer considerado descritivo, quer normativo, é uma tese interessante, assim como a ideia de que pela observação de como os cientistas trabalham, possamos aprender sobre como eles deveriam trabalhar. A última ideia tem levado ao desenvolvimento da filosofia *naturalista* da ciência, ou “estudos da ciência”, que examina os desenvolvimentos científicos em busca de pistas sobre que tipos de métodos os cientistas precisam, e desenvolvem, para as descobertas e invenções bem-sucedidas (Callebaut, 1993; Giere, 1999).

Independentemente de desenvolvimentos filosóficos, suspeitamos que a maioria dos epidemiologistas (e a maioria dos cientistas) continuará trabalhando como se o seguinte ponto de vista clássico fosse correto: a meta final da inferência científica é capturar algumas verdades objetivas sobre o mundo material no qual vivemos, e qualquer teoria de inferência, idealmente, deve ser avaliada pelo

quão bem ela nos leva a essas verdades. Contudo, esse ideal é impossível de se operacionalizar, pois se algum dia encontrarmos verdades finais, não teremos uma maneira de saber isso com certeza. Assim, aqueles que mantêm o ponto de vista de que a verdade científica não é arbitrária, concedem, entretanto, que nosso conhecimento dessas verdades será sempre uma tentativa. Para os refutacionistas, esse caráter tentativo tem uma qualidade assimétrica, mas tal assimetria é menos acentuada para outros. Podemos crer que sabemos que uma teoria é falsa porque ela falha constantemente nos testes aos quais a submetemos, mas nossos testes podem ser falhos, pois envolvem raciocínio e percepção sensorial imperfeitos. Tampouco podemos saber que uma teoria é verdadeira, mesmo que ela passe por cada teste que arquitetemos, pois ela pode falhar num teste ainda não imaginado.

Poucos, se existirem, discordarão que uma teoria de inferência deva ser avaliada, pelo menos em parte, por quão bem ela nos leva a detectar erros em nossas hipóteses e observações. Há, porém, muitas outras atividades inferenciais além da avaliação de hipóteses, como a predição ou a previsão de eventos, e as tentativas subsequentes de controlá-los (o que, é claro, requer informações causais). Os estatísticos, e não os filósofos, têm confrontado mais frequentemente esses problemas na prática, de modo que não deve ser surpreendente que as principais filosofias relacionadas a isso emergiram da estatística, e não da filosofia.

## Bayesianismo

Há uma outra filosofia de inferência que, como a maioria, mantém uma visão objetiva da verdade científica e uma visão do conhecimento como tentativo, ou incerto, mas que foca a avaliação do conhecimento em vez da verdade. Como o refutacionismo, a forma moderna dessa filosofia evoluiu dos escritos de pensadores do século XVIII. Os argumentos focais apareceram primeiramente em um ensaio axial pelo reverendo Thomas Bayes (1764), e a partir daí a filosofia é referida, geralmente, como bayesianismo (Howson e Urbach, 1993). E foi o renomado matemático e cientista francês, Pierre Simon de Laplace, quem primeiro deu a essa filosofia um formato estatístico aplicado. No entanto, ela não alcançou uma expressão completa até depois da 1ª Guerra Mundial, mais notadamente nos escritos de Ramsey (1931) e DeFinetti (1937); e, como o refutacionismo, não começou a aparecer na epidemiologia até a década de 1970 (p. ex., Cornfield, 1976).

O problema central abordado pelo bayesianismo é o seguinte: na lógica clássica, um argumento dedutivo não pode fornecer informação alguma sobre a verdade ou a falsidade de uma hipótese científica, a menos que se possa ter 100% de certeza sobre a verdade das suposições do argumento. Considere o argumento lógico chamado *modus tollens*: “se H implica B, e B é falso, então H deve ser falso”. Esse argumento é logicamente válido, mas a conclusão segue apenas as presunções de que as suposições “H implica B” e “B é falso” são declarações verdadeiras. Se essas suposições forem afirmações sobre o mundo físico, não há possibilidade de sabermos se elas são corretas com 100% de certeza, porque todas as observações estão sujeitas a erro. Além disso, a alegação de que “H implica B” dependerá, frequentemente, de sua própria cadeia de deduções, cada qual com suas próprias suposições, das quais não podemos ter certeza.

Por exemplo, se H for “assistir televisão causa homicídios” e B for “as taxas de homicídio são mais altas onde as televisões são mais comuns”, a primeira suposição usada em *modus tollens* para testar a hipótese de que assistir televisão causa homicídios será: “se assistir televisão causa homicídios, as taxas de homicídio são mais altas onde as televisões são mais comuns”. A validade dessa suposição é duvidosa – afinal de contas, mesmo que a televisão realmente cause homicídios, as taxas de homicídio podem ser baixas onde as televisões são comuns por causa de vantagens socioeconômicas nessas áreas.

Continuando a raciocinar desse modo, poderíamos chegar a uma situação mais pessimista do que até mesmo a imaginada por Hume. Não somente a indução carece de fundamentação lógica, como a *dedução* tem utilidade científica limitada, porque não podemos garantir a verdade de todas as suposições, mesmo que um argumento lógico seja válido. A resposta bayesiana a esse problema é

parcial, na qual é feita uma demanda severa sobre o cientista e colocada uma limitação severa sobre os resultados. Ela diz, grosseiramente, isto: se você pode atribuir um grau de certeza, ou probabilidade pessoal, às suposições de seu argumento válido, você pode usar toda e qualquer regra da teoria de probabilidades para derivar uma certeza para a conclusão, e essa certeza será uma consequência logicamente válida de suas certezas originais. Um fato inescapável é que sua certeza conclusiva, ou *probabilidade a posteriori*, pode depender fortemente do que você usou como certezas iniciais ou como *probabilidades a priori*. Se tais certezas iniciais não forem as mesmas que as de um colega, este colega pode muito bem atribuir uma certeza a uma conclusão diferente da que você derivou. Com o acúmulo de evidências consistentes, entretanto, os dados podem, geralmente, forçar até mesmo probabilidades *a priori* extremamente díspares a convergir para probabilidades *a posteriori* similares.

Tendo em vista que as probabilidades *a posteriori* resultantes de uma inferência bayesiana dependem da pessoa que supriu as certezas iniciais, e assim podem variar entre indivíduos, é dito que as inferências são subjetivas. Essa subjetividade da inferência bayesiana frequentemente é confundida com um tratamento subjetivo da verdade. Não somente tal visão do bayesianismo é incorreta, como é diametralmente oposta à filosofia bayesiana. A abordagem bayesiana representa uma tentativa construtiva de lidar com o dilema de que leis e fatos científicos não devem ser tratados como sabidos com certeza, ao passo que a lógica dedutiva clássica só gera conclusões quando alguma lei, fato ou conexão é asseverada com 100% de certeza.

Uma crítica comum à filosofia bayesiana é que ela desvia a atenção dos objetivos clássicos da ciência, como a descoberta de como o mundo funciona, para estados psicológicos chamados de “certezas”, “probabilidades subjetivas” ou “graus de crença” (Popper, 1959). Essa crítica, todavia, deixa de reconhecer a importância do estado mental de um cientista na determinação de quais teorias testar, e de que testes aplicar, além da consequente influência daquele estado sobre o estoque de dados disponíveis para inferência e da influência dos dados sobre os estados mentais.

Outra réplica a essa crítica é que os cientistas já usam dados para influenciar seus graus de crença e não temem expressá-los. O problema é que o processo convencional é informal, intuitivo e inefável, portanto não está sujeito a críticas; na pior das circunstâncias, ele frequentemente equivale a nada mais do que os especialistas anunciando que viram as evidências, e isto é o quão seguros estão. Como os cientistas alcançaram essa certeza é deixado obscuro ou, colocado de outra maneira, não é “transparente”. O problema é que ninguém, nem mesmo um especialista, é muito bom em formular certezas, informal e intuitivamente, que predigam satisfatoriamente fatos e eventos futuros (Kahne-man et al., 1982; Gilovich, 1993; Piattelli-Palmarini, 1994; Gilovich et al., 2002). Uma razão para esse problema é que vieses e preconceitos anteriores podem, facilmente, influenciar os julgamentos dos especialistas. Os métodos bayesianos forçam os especialistas a “pôr suas cartas na mesa” e a especificar explicitamente a força de suas crenças prévias e por que têm tais crenças, a defender tais especificações contra argumentos e evidências e a atualizar seus graus de certeza com novas evidências, de maneira que não violem a probabilidade lógica.

Em qualquer contexto de pesquisa, haverá um número ilimitado de hipóteses que poderiam explicar um fenômeno observado. Alguns argumentam que o progresso é melhor ajudado por testes severos (empiricamente desafiadores) daquelas explicações que pareçam mais prováveis à luz de pesquisas anteriores, de modo que os defeitos das teorias atualmente “recebidas” possam ser descobertos mais rapidamente. De fato, muitas pesquisas em certos campos adotam esse formato, como quando previsões teóricas sobre massa de partículas são submetidas a testes cada vez mais precisos, em experiências de física. Esse processo não envolve a mera repetição aperfeiçoada de estudos anteriores. Em vez disso, envolve testes de previsões da teoria, previamente não testadas, porém importantes. Além disso, há um imperativo de tornar a base para crenças prévias criticável e defensável. Dizer que as probabilidades *a priori* podem diferir entre as pessoas não significa dizer que todas as crenças sejam baseadas na mesma informação, nem que todas sejam igualmente sustentáveis.

As probabilidades de hipóteses auxiliares também são importantes no delineamento e na interpretação do estudo. A falha de uma teoria passar em um teste pode levar à sua rejeição mais rapidamente.

te, quando as hipóteses auxiliares das quais o teste depende possuam alta probabilidade. Essa observação fornece a lógica pela preferência por estudos caso-controle em “nichos” (nos quais os controles são selecionados de uma lista da população-fonte para os casos), em vez dos estudos caso-controle “de base hospitalar” (nos quais os controles são “selecionados” pela ocorrência ou pelo diagnóstico de uma ou mais doenças que não aquela que define os casos), porque os primeiros têm menos mecanismos para seleção tendenciosa de sujeitos e, portanto, têm uma probabilidade maior de seleção sem vieses.

Mesmo que se discutam os argumentos citados, a maioria dos epidemiologistas deseja alguma maneira de expressar os graus variáveis de certeza sobre valores possíveis de uma medida de efeito, à luz de dados disponíveis. Tais expressões, inevitavelmente, devem ser derivadas diante de incerteza considerável sobre detalhes metodológicos e vários eventos que levam aos dados disponíveis e podem ser extremamente sensíveis ao raciocínio usado em sua derivação. Por exemplo, como discutiremos mais extensamente no Capítulo 19, os intervalos de confiança convencionais quantificam apenas o erro aleatório sob suposições frequentemente questionáveis e, assim, não devem ser interpretados como medidas de incerteza total, especialmente em estudos não experimentais. Como observado anteriormente, a maioria das pessoas, inclusive cientistas, raciocina pobremente diante da incerteza. No mínimo, a filosofia bayesiana subjetiva fornece uma metodologia para o raciocínio sólido sob incerteza e, em particular, alerta contra estar abertamente certo sobre suas próprias conclusões (Greenland, 1998a, 1988b, 2006a; ver também Caps. 18 e 19).

Tais alertas são ecoados pela filosofia refutacionista. Como Peter Medawar (1979) colocou: “Eu não posso dar a qualquer cientista de qualquer idade um conselho melhor que este: a intensidade da convicção de que uma hipótese é verdadeira não tem influência sobre ela ser verdadeira ou não”. Nós acrescentaríamos dois pontos. Primeiro, a intensidade da convicção de que uma hipótese é falsa não tem influência sobre ela ser falsa realmente ou não. Segundo, os métodos bayesianos não confundem crenças com evidências. Eles usam evidências para modificar crenças, o que os cientistas fazem em qualquer evento, mas, frequentemente, de modos implícitos, intuitivos e incoerentes.

## Impossibilidade de prova científica

O debate vigoroso é uma característica da filosofia científica moderna, não menos na epidemiologia do que em outras áreas (Rothman, 1998). As filosofias divergentes da ciência podem ser reconciliadas? Haack (2003) sugeriu que o empreendimento científico está estreitamente a resolver um vasto enigma coletivo de palavras cruzadas. Em áreas nas quais a evidência é apertadamente entrelaçada, é mais racional colocar a confiança nas respostas, mas em áreas com escassez de informações, as teorias podem ser um pouco melhores do que palpites baseados em informações. Em relação ao método científico, Haack (2003) disse que “há menos no ‘método científico’ do que aquilo que o olho enxerga”. A investigação científica é categoricamente diferente de outros tipos? Não. A pesquisa científica é uma continuidade da pesquisa empírica do cotidiano – apenas melhor.”

Talvez o fio comum mais importante que surge das filosofias debatidas é que a prova é impossível na ciência empírica. Esse simples fato é especialmente importante para os epidemiologistas de observação, que frequentemente enfrentam a crítica de que a prova é impossível na epidemiologia, com a implicação de que ela é possível em outras disciplinas científicas. Tal crítica pode derivar de uma visão em que os experimentos constituem a fonte definitiva do conhecimento científico. Essa visão é equivocada em pelo menos duas questões. Em primeiro lugar, a natureza não experimental de uma ciência não elimina descobertas científicas impressionantes; a miríade de exemplos inclui a tectônica de placas, a evolução das espécies, os planetas em órbita de outras estrelas e os efeitos de fumar cigarros sobre a saúde humana. Mesmo quando são possíveis, os experimentos (inclusive ensaios randomizados) não fornecem nada que se aproxime de prova e, de fato, podem ser contravertidos, contraditórios ou não reproduzíveis. Se os ensaios clínicos randomizados fornecessem provas, nunca precisaríamos fazer mais de um deles sobre uma dada hipótese. Nem a ciência física, nem a experimental são imunes a tais problemas, como demonstrado por episódios como a “descoberta” experimental (mais tarde refutada) da fusão a frio (Taubes, 1993).

Alguns cientistas experimentais sustentam que as relações epidemiológicas são apenas sugestivas e acreditam que estudos de laboratório detalhados de mecanismos em indivíduos isolados podem revelar relações causa-efeito com certeza. Essa visão omite o fato de que *todas* as relações são sugestivas, exatamente como é discutido por Hume. Mesmo a dissecação mecanicista mais cuidadosa e detalhada de eventos individuais não pode fornecer mais do que associações, embora em um nível mais refinado. Os estudos de laboratório frequentemente envolvem um grau de controle do observador que não pode ser aproximado na epidemiologia; é somente esse controle e não o nível de observação, que pode fortalecer as inferências dos estudos de laboratório. E novamente, tal controle não garante contra o erro. Além disso, nem cientistas, nem tomadores de decisão com frequência são persuadidos facilmente quando apenas a evidência mecanicista do laboratório está disponível.

Todos os frutos do trabalho científico, na epidemiologia ou em outras disciplinas, são, no máximo, tentativas de formulações de uma descrição da natureza, mesmo quando o trabalho em si é efetuado sem erros. O caráter tentativo de nosso conhecimento não impede as aplicações práticas, mas deve nos manter céticos e críticos não só em relação ao trabalho de todos os outros, como também do nosso próprio. Algumas vezes, hipóteses etiológicas desfrutam de um grau de certeza extremamente elevado, universalmente, ou quase universalmente, compartilhado. A hipótese de que fumar cigarros causa câncer é um dos exemplos mais bem conhecidos. Essas hipóteses se elevam acima da aceitação “tentativa” e representam o mais próximo da “prova” a que podemos chegar. Porém, mesmo essas hipóteses não são “provas” com o grau de certeza absoluta que acompanha a prova de um teorema matemático.

## INFERÊNCIA CAUSAL EM EPIDEMIOLOGIA

O conhecimento etiológico sobre hipóteses epidemiológicas frequentemente é escasso, fazendo das próprias hipóteses, às vezes, pouco mais do que declarações vagas de associação causal entre exposição e doença, tal como “o fumo causa doença cardiovascular”. Essas hipóteses vagas têm apenas consequências vagas, que podem ser difíceis de testar. Para lidar com essa falta de clareza, os epidemiologistas geralmente procuram testar a negação da hipótese causal, isto é, a hipótese nula de que a exposição *não* tem uma relação causal com a doença. Então, qualquer associação observada pode refutar potencialmente a hipótese, estando-se sujeito à suposição (hipótese auxiliar) de que vieses e flutuações do acaso não são os únicos responsáveis pela observação.

## Testes de teorias epidemiológicas concorrentes

Se o mecanismo causal é declarado com bastante especificidade, as observações epidemiológicas podem fornecer testes cruciais de hipóteses causais não nulas, concorrentes. Por exemplo, quando a síndrome do choque tóxico foi primeiramente estudada, havia duas hipóteses concorrentes sobre o agente causal. Em uma hipótese, seria um produto químico no absorvente interno, de modo que as mulheres usando esse tipo de absorvente estariam expostas ao agente diretamente. Na outra hipótese, o absorvente atuaria como um meio de cultura para estafilococos, os quais produziram uma toxina. Ambas as hipóteses explicavam a relação da ocorrência de choque tóxico com o uso do absorvente interno. Entretanto, levaram a predições opostas sobre a relação entre a frequência de troca dos absorventes e a taxa de choque tóxico. Na hipótese de um agente químico, a troca mais frequente levaria à maior exposição ao agente, e à possível absorção de uma dose global maior. Essa hipótese predizia que as mulheres que trocassem de absorventes mais frequentemente teriam uma taxa mais elevada do que aquelas que os mudassem com frequência menor. A hipótese do meio de cultura predisse que as mulheres que mudassem de absorventes frequentemente teriam uma taxa mais baixa do que aquelas que o faziam com menor frequência, porque uma duração de uso curta para cada absorvente preveniria a multiplicação de estafilococos o bastante para produzir uma dose lesiva de toxina. Assim, a pesquisa epidemiológica, ao mostrar que a troca pouco frequente de absorventes estava associada a uma taxa mais alta de choque tóxico, refutou a teoria química na forma apresentada. Houve, contudo, uma



terceira hipótese, de que um produto químico em alguns absorventes internos (p. ex., conteúdo de oxigênio) melhoraria o desempenho destes como meio de cultura. Essa hipótese do promotor químico fez a mesma predição sobre a associação com a frequência de troca de absorventes que a hipótese da toxina microbiana (Lanes e Rothman, 1990).

Outro exemplo de uma teoria que pode ser facilmente testada por dados epidemiológicos relaciona-se com a observação de que as mulheres que faziam terapia de reposição hormonal com estrógenos tinham uma taxa consideravelmente elevada de câncer de endométrio. Horwitz e Feinstein (1978) conjecturaram uma teoria concorrente para explicar a associação: propuseram que as mulheres que tomavam estrogênio experimentavam sintomas como sangramento, que as induziriam a consultar o médico. A avaliação diagnóstica resultante levaria à detecção do câncer de endométrio em um estágio mais precoce nessas mulheres, em comparação com aquelas que não estavam usando estrogênio. Horwitz e Feinstein argumentaram que a associação surgia desse viés de detecção, alegando que, sem a avaliação diagnóstica induzida pelo sangramento, muitos desses cânceres não seriam detectados de modo algum. Muitas observações epidemiológicas foram utilizadas para avaliar essas hipóteses concorrentes. A teoria de detecção de viés predisse que as mulheres que tinham usado estrogênio apenas por um período curto teriam a taxa mais elevada, pois os sintomas relacionados com o uso de estrogênio, que levariam à consulta médica, tendiam a aparecer logo depois do início do uso. Visto que a associação entre uso recente de estrogênio e câncer do endométrio foi a mesma, tanto para as usuárias de curta duração quanto para as de uso prolongado, a teoria do viés de detecção foi refutada como uma explicação para todos, salvo para uma pequena fração: os casos de câncer do endométrio ocorrendo após uso de estrogênio. A refutação da teoria do viés de detecção também dependeu de muitas outras observações. De importância especial foi a implicação da teoria de que deve haver um enorme número de cânceres de endométrio não detectados na população típica de mulheres, para explicar a taxa muito maior observada nas usuárias de estrogênio, implicação que não foi apoiada por observações ulteriores (Hutchison e Rothman, 1978).

O exemplo do câncer do endométrio ilustra um ponto crítico na compreensão do processo de inferência causal em estudos epidemiológicos: muitas das hipóteses sendo avaliadas na interpretação de estudos epidemiológicos são hipóteses auxiliares, porque são independentes da presença, da ausência ou da direção de qualquer conexão causal entre a exposição em estudo e a doença. Por exemplo, explicações de como tipos específicos de viés podiam ter distorcido uma associação entre exposição e doença são as alternativas comuns à hipótese primária do estudo. Muito da interpretação de estudos epidemiológicos resume-se ao teste de tais explicações auxiliares para associações observadas.

## Critérios causais

Na prática, como os epidemiologistas separam explicações causais das não causais? Apesar das críticas filosóficas da inferência indutiva, considerações orientadas indutivamente são usadas, com frequência, como critérios para fazer tais inferências (Weed e Gorelic, 1996). Se um conjunto de critérios causais necessários e suficientes pudesse ser usado para distinguir entre relações causais e não causais em estudos epidemiológicos, o trabalho do cientista seria consideravelmente facilitado. Com tais critérios, todas as preocupações sobre a lógica, ou a sua falta, na inferência causal poderiam ser incluídas: seria necessário apenas consultar uma lista de critérios para ver se uma relação seria causal. Sabemos, pela filosofia anteriormente revisada, que não existe um conjunto de critérios suficientes. No entanto, listas de critérios causais têm se tornado populares, possivelmente porque parecem proporcionar um mapa rodoviário para um território complicado e talvez por sugerirem hipóteses a serem avaliadas em um dado problema.

Um conjunto de critérios comumente usados foi baseado em uma lista de considerações ou de “pontos de vista”, proposta por Sir Austin Bradford Hill (1965). A lista de Hill foi uma expansão de uma lista oferecida previamente no relatório histórico *Smoking and Health* (1964) do *Surgeon General* dos Estados Unidos, que, por sua vez, foi precedida pelos cânones indutivos de John Stuart Mill

(1862) e pelas normas dadas por Hume (1739). Subsequentemente, outros, especialmente Susser, desenvolveram mais considerações causais (Kaufman e Poole, 2000).

Hill sugeriu que as seguintes considerações na tentativa de distinguir entre associações causais e não causais já eram “bem definidas e além do que gostaríamos de atribuir à obra do acaso”: (1) força, (2) consistência, (3) especificidade, (4) temporalidade, (5) gradiente biológico, (6) plausibilidade, (7) coerência, (8) evidências experimentais e (9) analogia. Hill salientou que inferências causais não podem ser baseadas em um grupo de regras, condenou a ênfase sobre os testes de significância estatística e reconheceu a importância de muitos outros fatores na tomada de decisões (Phillips e Goodman, 2004). Apesar disso, o ponto de vista equivocado, porém popular, de que suas considerações deveriam ser usadas como critérios para inferência causal, torna necessário examiná-las em detalhe.

### **Força**

Hill argumentou que associações fortes são particularmente convincentes, porque, para associações mais fracas, é “mais fácil” imaginar o que chamaríamos hoje de um confundidor não mensurado, que poderia ser responsável pela associação. Vários anos antes, Cornfield e colaboradores (1959) tiraram conclusões semelhantes. Eles se concentraram em um confundidor hipotético isolado, o qual, por si próprio, explicaria inteiramente uma associação observada. Expressaram uma forte preferência por medidas de razão de força, em oposição a mensurações de diferenças, e focaram em como a estimativa observada de uma razão de risco fornece um mínimo para a associação que um confundidor completamente explicativo deve ter com a exposição (em vez de um mínimo para a associação confundidor/doença). De importância especial, Cornfield e colaboradores reconheceram que apenas ter uma associação fraca não elimina uma conexão causal (Rothman e Poole, 1988). Hoje, algumas associações, como aquelas entre fumo e doença cardiovascular ou entre fumaça de tabaco ambiental e câncer de pulmão, são aceitas pela maioria como causais, embora as associações sejam consideradas fracas.

Exemplos opostos de associações fortes, mas não causais, também não são difíceis de encontrar; qualquer estudo com um confundidor forte ilustra o fenômeno. Por exemplo, considere a relação forte entre síndrome de Down e ordem de nascimento, que é confundida pela relação entre essa síndrome e a idade materna. É claro que uma vez identificado o fator de confundimento, a associação é diminuída pelo controle desse fator.

Esses fatores nos lembram que uma associação forte não é necessária nem suficiente para a causalidade, e que a fraqueza não é necessária nem suficiente para a ausência de causalidade. Uma associação forte tem peso apenas sobre hipóteses de que a associação se deve inteira, ou parcialmente, a confundidores não mensurados ou a outra fonte modesta de viés.

### **Consistência**

Para a maioria dos observadores, consistência refere-se à observação repetida de uma associação, em populações diferentes, sob diferentes condições. A falta de consistência, entretanto, não exclui uma associação causal, porque alguns efeitos são produzidos por suas causas somente sob circunstâncias incomuns. Mais precisamente, o efeito de um agente causal não pode ocorrer a menos que as causas componentes complementares atuem, ou já tenham atuado, para completar uma causa suficiente. Essas condições nem sempre serão satisfeitas. Assim, transfusões podem causar infecção com o vírus da imunodeficiência humana, mas nem sempre o fazem: o vírus também precisa estar presente. O uso de absorvente interno pode causar a síndrome do choque tóxico, mas só raramente, quando outras condições, talvez desconhecidas, forem encontradas. A consistência só é aparente depois que todos os detalhes relevantes de um mecanismo causal são compreendidos, o que equivale a dizer “muito raramente”. Além do mais, pode ser esperado que mesmo estudos exatamente dos mesmos fenômenos gerem resultados diferentes, simplesmente porque diferem em seus métodos e erros aleatórios. A consistência serve apenas para eliminar hipóteses de que a associação é atribuível a algum fator que varie ao longo dos estudos.



Um erro na implementação de critérios de consistência é tão comum que merece menção especial. Alega-se, algumas vezes, que uma literatura ou um grupo de resultados é inconsistente, simplesmente porque alguns resultados são “estatisticamente significativos” e outros não são. Essa espécie de avaliação é completamente falaciosa, mesmo que seja aceito o uso de métodos de testes de significância. Os resultados (estimativas de efeito) de um grupo de estudos poderiam ser todos idênticos, mesmo que muitos fossem significativos e muitos não fossem, a diferença em significância que surge somente por causa de diferenças nos desvios-padrão ou no tamanho dos estudos. Inversamente, os resultados podem estar significativamente em conflito, mesmo que todos fossem não significativos individualmente, simplesmente porque, em conjunto, um efeito pode ser aparente em alguns subgrupos, mas não em outros (ver Cap. 33). A falácia de julgar a consistência pela comparação de valores *P*, ou significância estatística, não é eliminada pela “padronização” de estimativas (i.e., dividi-las pelo desvio-padrão do desfecho, multiplicá-las pelo desvio-padrão da exposição, ou ambos); de fato isso é pior, pois tal padronização pode criar diferenças onde não existem, ou mascarar diferenças verdadeiras (Greenland et al., 1986, 1991; ver Caps. 21 e 33).

### **Especificidade**

O critério da especificidade tem duas variantes. A primeira é que uma causa leva a um efeito isolado, não a múltiplos efeitos. A outra é que um efeito tem uma causa, não causas múltiplas. Hill mencionou ambas. O primeiro critério, especificidade dos efeitos, foi usado como um argumento a favor de uma interpretação causal da associação entre fumo e câncer de pulmão, em um ato de raciocínio circular, favorecendo comparações de razões, e não de diferenças, como as medidas de força apropriadas. Quando medidas de razões foram examinadas, a associação do fumo a doenças pareceu “quantitativamente específica” para o câncer de pulmão. Quando medidas de diferença foram examinadas, a associação pareceu ser inespecífica, com várias doenças (outros cânceres, doença cardíaca coronariana, etc.) sendo pelo menos tão fortemente associadas com o fumo, quanto o câncer de pulmão o era. Hoje, sabemos que o fumo afeta o risco de muitas doenças, e que as comparações de diferenças estavam retratando acuradamente essa falta de especificidade. Infelizmente, contudo, o episódio histórico do debate sobre fumo e saúde é citado frequentemente hoje em dia como justificativa para o critério da especificidade e para usar comparações de razões para medir a força da associação. As lições adequadas que se deve aprender com aquele episódio seriam justamente o oposto.

Weiss (2002) argumentou que a especificidade pode ser usada para distinguir algumas hipóteses causais de hipóteses não causais, quando a hipótese causal prediz uma relação com um desfecho, mas nenhuma relação com outro. Seu argumento é persuasivo quando, somada à hipótese causal, se tem uma hipótese alternativa não causal, que prediz uma associação inespecífica. Weiss ofereceu o exemplo da sigmoidoscopia de rastreamento, que foi associada em estudos de caso-controle com uma redução de 50 a 70% na mortalidade de tumores distais do reto e tumores do cólon distal, dentro do alcance do sigmoidoscópio, mas a nenhuma diminuição de mortalidade por tumores em outras partes do cólon. Se o efeito da sigmoidoscopia de rastreamento não fosse específico para os tumores do cólon distal, isso daria suporte não a todas as teorias não causais para explicar a associação, como Weiss sugeriu, mas somente àquelas teorias não causais que teriam predito uma associação inespecífica. Assim, a especificidade só pode entrar em jogo quando pode ser deduzida logicamente da hipótese causal em questão e quando a inespecificidade de uma ou mais hipóteses pode ser deduzida logicamente a partir de uma ou mais hipóteses não causais.

### **Temporalidade**

Temporalidade refere-se à necessidade de que a causa preceda o efeito no tempo. Esse critério é indiscutível, desde que qualquer observação alegada de causação envolva a suposta causa *C* precedendo o suposto efeito *D*. Entretanto, *não* quer dizer que uma ordem reversa no tempo seja evidência contra a hipótese de que *C* causa *D*. Em vez disso, as observações em que *C* se seguiu

à D mostram meramente que C não poderia ter causado D em tais exemplos; elas não fornecem evidências a favor, ou contra, à hipótese de que C possa causar D naqueles exemplos em que ele precede D. Somente se for encontrado que C não pode preceder D, podemos eliminar a hipótese causal de que C *poderia* causar D.

### **Gradiente biológico**

Gradiente biológico refere-se à presença de uma curva de dose-resposta ou exposição-resposta, com uma forma esperada. Embora Hill se refira a um gradiente “linear”, sem especificar a escala, um gradiente linear em uma escala, tal como risco, pode ser distintamente não linear em uma outra escala, tal como o logaritmo de risco, as chances ou o logaritmo de chances. Poderíamos relaxar a expectativa de linear para estritamente monótona (aumentando ou diminuindo constantemente), ou ainda, para meramente monótona (um gradiente que nunca muda de direção). Por exemplo, fumar mais significa maior exposição a carcinógenos e mais lesão tecidual, daí maior oportunidade para carcinogênese. Algumas associações causais, todavia, mostram um aumento rápido de resposta (um efeito limiar aproximado), em vez de uma tendência estritamente monótona. Um exemplo é a associação entre DES e adenocarcinoma vaginal. Uma explicação possível é que todas as doses de DES que foram administradas eram suficientemente grandes para produzir o efeito máximo. Sob essa hipótese, para todas as mulheres expostas ao DES, o desenvolvimento da doença dependeria inteiramente de outras causas componentes.

O tópico um tanto controverso sobre consumo de álcool e mortalidade é outro exemplo. As taxas de mortalidade são mais altas entre os abstêmios do que entre os que bebem moderadamente, mas se elevam mais entre os que bebem muito. Há debate considerável sobre quais partes da curva dose-resposta em forma de J têm relação causal com o consumo de álcool e quais partes são artefatos não causais, surgindo de confundimento ou de outros vieses. Alguns estudos parecem encontrar somente uma relação crescente entre consumo de álcool e mortalidade, possivelmente porque as categorias de consumo de álcool sejam demasiado amplas para se distinguir taxas diferentes entre consumidores moderados e abstêmios ou, possivelmente, porque têm menos confundimento na extremidade mais baixa da escala de consumo.

Associações que realmente mostram uma tendência monótona na frequência da doença, com níveis crescentes de exposição, não são, necessariamente, causais. Confundimento pode resultar em uma relação monótona entre um fator de risco não causal e uma doença, se o próprio fator de confundimento demonstrar um gradiente biológico em sua relação com a doença. A relação entre ordem de nascimento e síndrome de Down, mencionada anteriormente, mostra um forte gradiente biológico, que reflete meramente a relação progressiva entre idade materna e ocorrência de síndrome de Down.

Esses tópicos implicam que a existência de uma associação monótona não é necessária nem suficiente para uma relação causal. Uma relação não monótona só refuta aquelas hipóteses causais específicas o bastante para predizer uma curva dose-resposta monótona.

### **Plausibilidade**

Plausibilidade refere-se ao fato de uma associação ser cientificamente plausível. Mais do que qualquer outro critério, esse mostra o quão estreitamente os sistemas de critérios causais estão focados na epidemiologia. O ponto de partida é uma associação epidemiológica. Ao nos perguntarmos se ela é causal ou não, uma das considerações que levamos em conta é sua plausibilidade. De uma perspectiva menos provinciana, toda a tarefa da inferência causal seria vista como o ato de determinar o quão plausível é uma *hipótese* causal. Algumas das considerações que levaríamos em conta seriam associações epidemiológicas, caso estivessem disponíveis. Frequentemente não estão, porém, a inferência causal deve ser feita com influxos da toxicologia, da farmacologia, da biologia básica e de outras ciências.

Da mesma forma que a epidemiologia não é essencial para a inferência causal, a plausibilidade pode mudar com o tempo. Sartwell (1960) enfatizou esse ponto, citando observações de Cheever em

1861, que fez comentários sobre a etiologia do tifo antes que seu modo de transmissão (por meio de piolhos do corpo) fosse conhecido:

Não poderia ser mais ridículo para o estranho que passou a noite na terceira classe de um navio emigrante atribuir o tifo, que ele contraiu lá, aos parasitas com os quais os corpos dos enfermos podiam estar infestados. Uma causa adequada, razoável por si própria, precisa corrigir as coincidências da simples experiência.

O que era para Cheever uma explicação implausível, acabou sendo a explicação correta, porque eram realmente os parasitas que causavam a infecção pelo tifo. Esse é o problema da plausibilidade: com muita frequência, não se baseia na lógica ou em dados, mas somente em crenças prévias. Isso não significa que o conhecimento biológico deva ser descartado quando uma nova hipótese está sendo avaliada, mas apenas para assinalar a dificuldade na aplicação daquele conhecimento.

A abordagem bayesiana da inferência tenta lidar com esse problema requerendo que se quantifique, em uma escala de probabilidades (0 a 1), a certeza que se tem sobre crenças anteriores, assim como nas novas hipóteses. Essa quantificação expõe publicamente o dogmatismo, ou a mente aberta, do analista, com valores de certeza próximos de 1 ou de 0, revelando o seu forte comprometimento a favor ou contra uma hipótese. Ela também fornece meios de se testar aquelas crenças quantificadas contra novas evidências (Howson e Urbach, 1993). Apesar disso, nenhuma abordagem pode transformar plausibilidade em um critério causal objetivo.

### Coerência

Extraído de *Smoking and Health* (1964) do *Surgeon General* dos Estados Unidos, o termo *coerência* implica que uma interpretação de causa e efeito para uma associação não entre em conflito com o que se sabe da história natural e da biologia da doença. Os exemplos para coerência dados por Hill, tais como o efeito histopatológico do fumo sobre o epitélio brônquico (em referência à associação entre fumo e câncer de pulmão) ou a diferença na incidência de câncer de pulmão por sexo, poderiam ser considerados, de certa forma, exemplos de plausibilidade, bem como de coerência; a distinção parece ser tênue. Hill enfatizou que a ausência de informação coerente, diferenciada, aparentemente, da presença desta, não deve ser tomada como evidência contra uma associação ser considerada causal. Em contrapartida, a presença de informação conflitante pode realmente refutar uma hipótese, mas é preciso lembrar sempre que a informação conflitante pode estar errada, ou mal interpretada. Um exemplo mencionado anteriormente é a “anomalia de inalação” do fumo e câncer de pulmão; o fato de que o excesso de cânceres de pulmão vistos entre fumantes parecia estar concentrado em bloqueios nas vias aéreas pulmonares superiores. Vários observadores interpretaram essa anomalia como evidência de que os cigarros não eram responsáveis pelo excesso. Outras observações, entretanto, sugeriram que os carcinógenos veiculados pelo cigarro depositavam-se, preferencialmente, onde o excesso era observado, e assim a anomalia era de fato compatível com um papel causal para os cigarros (Wald, 1985).

### Evidência experimental

Para observadores diferentes, a evidência experimental pode referir-se a ensaios clínicos, a experimentos de laboratório com roedores ou com outros organismos não humanos, ou a ambos. Evidências de experimentos humanos, contudo, raramente estão disponíveis para questões de pesquisa epidemiológica, e a evidência animal relaciona-se a diferentes espécies e, geralmente, a categorias de exposição muito diferentes daqueles que os seres humanos experimentam. A incerteza nas extrapolações de animais para seres humanos domina, frequentemente, a incerteza das avaliações quantitativas de risco (Freedman e Zeisel, 1988; Crouch et al., 1997).

Para Hill, entretanto, evidência experimental significava outra coisa: é a “evidência experimental” ou “semiexperimental” obtida pela redução ou pela eliminação de uma exposição supos-

tamente nociva e verificação subsequente do declínio da frequência da doença. Hill chamou isso de evidência mais forte possível de causalidade que pode ser obtida. Essa evidência pode ser falha, contudo, pois a abordagem “semiexperimental” nada mais é do que uma análise de tendência no tempo “antes e depois”, que pode ser confundida ou sofrer outros tipos de viés, por muitas alterações seculares concomitantes. Além do mais, mesmo que a remoção da exposição realmente reduza causalmente a frequência da doença, isso pode não ser consequência da razão etiológica hipotética. A drenagem de um pântano próximo de uma cidade, por exemplo, reduziria de forma previsível e causal a taxa de febre amarela, ou de malária, nessa cidade, no verão seguinte. Porém, seria um erro chamar essa observação de evidência mais forte possível de um papel causal dos miasmas (Poole, 1999).

### Analogia

Qualquer que seja o discernimento derivado de analogia, ele é prejudicado pela imaginação inventiva dos cientistas, que podem encontrar analogias em toda parte. No máximo, a analogia fornece uma fonte de hipóteses mais elaboradas sobre as associações em estudo; a ausência de tais analogias reflete apenas falta de imaginação ou de experiência, não torna a hipótese falsa.

Poderíamos achar ingênuos os exemplos de Hill, nos quais o raciocínio por analogia a partir das tragédias da talidomida e da rubéola tornou mais provável que outros medicamentos e infecções pudessem causar outros defeitos congênitos. Entretanto, tal raciocínio é comum; suspeitamos que a maioria das pessoas acha mais plausível que o fumo possa causar, por exemplo, câncer de estômago, por causa de suas associações, algumas amplamente aceitas como causais, com cânceres de outros órgãos internos e gastrintestinais. Aqui vemos como o critério da analogia pode estar em desacordo com qualquer dos dois critérios de especificidade. Quanto mais apta a analogia, menos específicos os efeitos de uma causa, ou menos específicas as causas de um efeito.

### Resumo

Os padrões de evidência epidemiológica oferecidos por Hill estão carregados de reservas e de exceções. O próprio Hill foi ambivalente sobre sua utilidade. Ele não usava a palavra “critérios”. Ele os chamava de “pontos de vista” ou de “perspectivas”. No entanto, perguntava: “Em que circunstâncias podemos passar dessa *associação* observada para um veredicto de *causação*?”\*. Ainda assim, apesar de falar de veredictos sobre causação, Hill discordava de que existissem quaisquer “regras de evidência firmes e diretas” pelas quais julgar causação: “Nenhum dos meus nove pontos de vista pode trazer evidência indisputável a favor, ou contra, à hipótese de causa e efeito, e nenhum deles pode ser exigido como um *sine qua non*” (Hill, 1965).

Na verdade, como observado, o quarto ponto de vista, temporalidade, é um *sine qua non* para explicações causais de associações observadas. No entanto, esse ponto de vista não resiste à hipótese de que uma exposição seja capaz de causar uma doença em situações ainda não observadas (quer no passado, quer no futuro). Pois, suponhamos que cada caso exposto de doença já relatado tivesse sofrido a exposição após desenvolver a doença. Essa relação temporal reversa implicaria que a exposição não tivesse causado a doença entre esses casos relatados, e assim refutaria a hipótese de que teria causado a doença. No entanto, essa mesma relação temporal invertida *não* refutaria a hipótese de que a exposição fosse *capaz* de causar a doença, ou de que tivesse causado a doença em casos não observados. Significaria apenas que não teríamos evidências epidemiológicas válidas relevantes para aquela hipótese, pois ainda não teríamos visto o que aconteceu com aqueles casos expostos antes da ocorrência da doença, em relação àqueles não expostos. Além disso, o que parece ser uma sequência causal poderia representar causalidade reversa se os sintomas pré-clínicos da doença levassem

\* N. de T.: A ênfase é do original.

à exposição, e então a doença seguisse abertamente, como quando pacientes sentindo dores tomam analgésicos, que pode ser o resultado de doença que é mais tarde diagnosticada, e não a causa.

Afora a temporalidade, não há critério necessário ou suficiente para determinar se uma associação observada é causal. Somente quando uma hipótese causal é elaborada em uma extensão na qual se pode prever uma forma particular de consistência, de especificidade, de gradiente biológico, e assim por diante, os “critérios causais” podem entrar em jogo na avaliação de hipóteses causais, e mesmo assim eles não entram na avaliação da hipótese geral em si, mas apenas na avaliação de algumas hipóteses específicas, deixando outras não testadas.

Essa conclusão concorda com a visão de Hume e de muitos outros, de que as inferências causais não podem atingir a certeza de deduções lógicas. Embora alguns cientistas continuem a desenvolver considerações causais como auxílio à inferência (Susser, 1991), outros argumentam que é prejudicial enevoar o processo de inferência ao considerar critérios de *check list* (Lanes e Poole, 1984). Uma abordagem refutacionista, intermediária, busca transformar os critérios propostos em testes de hipóteses causais (Maclure, 1985; Weed, 1986). Tal abordagem ajuda a evitar a tentação de usar critérios causais simplesmente para reforçar teorias prediletas que estejam à mão e, em vez disso, permite que os epidemiologistas se concentrem na avaliação de teorias causais concorrentes, utilizando observações cruciais. Embora essa abordagem refutacionista à inferência causal possa parecer estar em choque com a implantação comum dos pontos de vista de Hill, ela procura, em verdade, responder à questão fundamental colocada por Hill e ao propósito final dos pontos de vista que ele promulgou:

O que [os nove pontos de vista] podem fazer, com força maior ou menor, é nos ajudar a decidir sobre a questão fundamental – há qualquer outro modo de explicar o conjunto de fatos diante de nós, há alguma outra resposta igualmente, ou mais, provável do que causa e efeito? (Hill, 1965)

A frase crucial “igualmente ou mais provável do que causa e efeito” nos sugere uma avaliação subjetiva da certeza, ou da probabilidade, da hipótese causal em questão, em relação a uma outra hipótese. Embora Hill escrevesse em um tempo em que expressar incerteza não era comum em estatística, pela sua declaração parece que, para ele, a inferência causal é um assunto subjetivo do grau de crença, de certeza, ou de convicção pessoal. Em qualquer caso, essa visão é precisamente aquela das estatísticas bayesianas subjetivas (Cap. 18).

Não é surpreendente que estudos de casos (p. ex., Weed e Gorelick, 1996) e inquéritos de epidemiologistas (Holman et al., 2001) mostrem, ao contrário da retórica que frequentemente acompanha as invocações de critérios causais, que os epidemiologistas *não* concordaram sobre um conjunto de critérios causais ou sobre como aplicá-los. Em um estudo, no qual se pediu a epidemiologistas para empregar critérios causais a resumos fictícios de literaturas epidemiológicas, a concordância foi apenas levemente maior do que teria sido esperado pelo acaso (Holman et al., 2001). O uso típico de critérios causais deve construir uma defesa de uma posição a favor ou contra causalidade a que se tenha chegado por outros meios não declarados. Os autores remexem e escolhem entre os critérios que adotam, e os definem e ponderam em maneiras *ad hoc* que dependem somente das exigências da discussão em pauta. Dessa forma, os critérios causais parecem funcionar menos como padrões ou princípios e mais como valores (Poole, 2001b) e variam entre cientistas, e até mesmo dentro do trabalho de um só cientista, dependendo do contexto e do tempo. Assim, se é que existem critérios causais objetivos, eles ainda precisam ser identificados.

# Medidas de ocorrência

Sander Greenland e Kenneth J. Rothman

## Tempos de incidência 47

### Taxas de incidência 48

- Pessoa-tempo e tempo da população 48
- Taxas populacionais e taxas individuais 48
- Interpretação apropriada das taxas de incidência 49
- Taxas de eventos recorrentes 50

### Tipos de populações 51

- Populações fechadas 51
- Populações abertas 52
- Populações *versus* coortes 53
- Estado de equilíbrio 54
- Relação de taxas de incidência com tempos de incidência em populações especiais 54

### Outros tipos de taxas 55

### Proporções de incidência e proporções de sobrevida 55

## Relações entre medidas de incidência 57

- Fórmula produto-limite 58
- Fórmula exponencial 58
- Aplicações com riscos competitivos 60
- Relação entre proporção de sobrevida e tempo médio de incidência 61
- Resumo 61

## Limitações e generalizações das medidas básicas de ocorrência 62

### Prevalência 62

- Uso da prevalência na pesquisa etiológica 63
- Prevalência, incidência e duração média 63

### Idade média no evento 65

### Padronização 65

Neste capítulo, começamos a abordar os elementos básicos, os conceitos e as ferramentas da epidemiologia. Um bom ponto de partida é definir epidemiologia. Infelizmente, parece que existem mais definições de epidemiologia do que epidemiologistas. Alguns a têm definido de acordo com seus métodos. Embora os métodos da epidemiologia possam ser característicos, é mais típico definir um ramo da ciência em termos de objeto de estudo do que de metodologia. MacMahon e Pugh (1970) ofereceram uma definição amplamente aceita, a qual atualizamos levemente: epidemiologia é o estudo da distribuição e dos determinantes da frequência de doenças em populações humanas. Uma definição baseada em seu objeto de estudo semelhante tem sido atribuída a Gaylord Anderson (Cole, 1979), que definiu epidemiologia simplesmente como o estudo da *ocorrência* de enfermidades. Embora distinções razoáveis possam ser feitas entre os termos *doença* e *enfermidade*, aqui os trataremos como sinônimos.

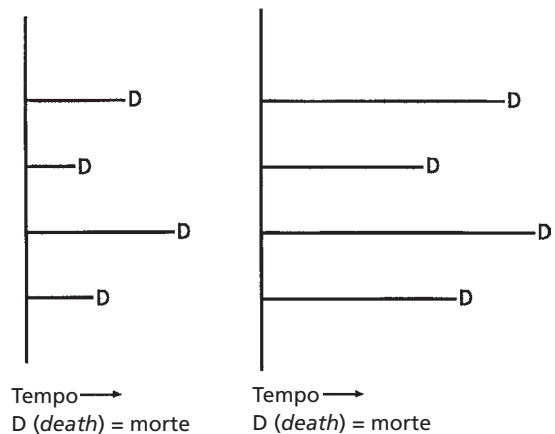
Reconhecendo o amplo espectro atual da epidemiologia, podemos defini-la como o estudo da distribuição de estados de saúde e eventos em populações. Com essa definição, pretendemos abranger não apenas doença e enfermidade, mas também estados fisiológicos, tais como pressão arterial, mensurações psicológicas, como o escore de depressão e desfechos positivos, tais como imunidade a doenças. Outras ciências, como a clínica médica, também estão voltadas para o estudo da saúde e da doença, porém, na epidemiologia, o foco são as distribuições na população.

O objetivo de grande parte da pesquisa epidemiológica é obter uma estimativa válida e precisa do efeito de uma causa potencial sobre a ocorrência de doença, a qual, frequentemente, é um desfecho binário (ou/ou), tal como “morto/vivo”. Para tanto, um epidemiologista deve ser capaz de mensurar a frequência de ocorrência de doenças, quer em termos absolutos, quer em termos relativos. Concentraremos nosso foco sobre quatro medidas básicas da frequência de doenças. *Tempos de incidência* são simplesmente os tempos, após um evento referencial comum, em que novos casos da doença ocorrem entre membros da população. *Taxa de incidência* mede a ocorrência de novos casos da doença por unidade de pessoa-tempo. *Proporção de incidência* mensura a proporção de pessoas que desenvolvem nova doença durante um período de tempo especificado. *Prevalência*, uma medida de *status* em vez de ocorrência de casos novos de doença, mede a proporção de pessoas que têm a doença em um tempo específico. Discutiremos, também, como essas medidas são generalizadas para desfechos mensurados em uma escala mais complexa do que uma dicotomia, tais como função pulmonar, contagem de linfócitos ou título de anticorpos. E, por fim, descreveremos como as medidas podem ser *padronizadas*, ou agregadas como médias das distribuições de fatores relacionados à saúde na população, para a obtenção de medidas de ocorrência sumarizadas.

## TEMPOS DE INCIDÊNCIA

Na tentativa de medir a frequência da ocorrência de doença em uma população, é insuficiente o mero registro do número de pessoas ou da proporção da população que é afetada. Também é necessário levar em consideração o tempo passado antes que a doença ocorra, assim como o período de tempo durante o qual os eventos são contabilizados. Considere a frequência de óbito. Posto que todas as pessoas cedo ou tarde são afetadas, o tempo transcorrido entre o nascimento e o óbito torna-se o fator determinante da taxa de ocorrência de morte. Se, em média, a morte chega mais cedo para pessoas de uma população do que para as de uma outra, é natural dizer que a primeira população tem uma taxa de mortalidade mais alta do que a segunda. O tempo é o fator que diferencia as duas situações mostradas na Figura 3.1.

Em um estudo epidemiológico, podemos mensurar o tempo de eventos na vida de uma pessoa em relação a qualquer um dentre vários eventos de referência. Ao usar a idade, por exemplo, o evento de referência é o nascimento, mas, em vez disso, poderíamos usar o início de um tratamento, ou o começo de uma exposição, como evento de referência. O evento de referência pode ocorrer em um tempo que é único para cada pessoa, como é o caso do nascimento, mas ele também poderia ser estabelecido como um valor em comum, tal como um dia escolhido do calendário. O tempo do evento de referência determina o tempo original, ou *tempo zero*, para a mensuração da sequência temporal dos eventos.



**FIGURA 3.1** • Dois padrões diferentes de mortalidade.



Dado um desfecho ou “incidente” de interesse, o *tempo de incidência* de uma pessoa para esse desfecho é definido como o período do tempo zero ao tempo em que o desfecho ocorrer, se ele ocorrer. Os sinônimos para tempo de incidência incluem *tempo de evento*, *tempo de falha* e *tempo de ocorrência*. Um homem que teve seu primeiro infarto do miocárdio no ano 2000, com a idade de 50 anos, tem um tempo de incidência de 2000, em tempo de calendário (ocidental), e um tempo de incidência de 50, em tempo de idade. O tempo de incidência de uma pessoa é indefinido se aquela pessoa nunca experimentar o desfecho. Sob essa convenção, uma mulher que passou por uma histerectomia com 45 anos, sem nunca ter tido câncer do endométrio, é classificada como tendo um tempo de incidência de câncer do endométrio não especificado, porém maior do que 45 anos. Diz-se então que a histerectomia *cersura* a incidência de câncer de endométrio da mulher na idade de 45 anos.

Há muitas maneiras de sumarizar a distribuição dos tempos de incidência em populações se não houver censura. Por exemplo, pode-se observar o tempo médio, o tempo mediano e outras medidas de resumo. Tais abordagens são usadas, comumente, com tempo para a morte, para o qual a média, ou *expectativa de vida*, é uma medida popular para comparação do estado de saúde de populações. Se há censura, entretanto, a tarefa de sumarização torna-se mais complicada, e os epidemiologistas têm se voltado, tradicionalmente, para os conceitos envolvendo *pessoa-tempo em risco* para lidar com essa situação.

O termo *idade média no óbito* merece atenção especial, pois é usado, algumas vezes, para denotar expectativa de vida, mas é utilizado, frequentemente, para indicar uma quantidade completamente diferente, a saber, a média de idade daqueles falecendo em um determinado ponto no tempo. A última quantidade é denominada mais precisamente de média *transversal* de idade no óbito. As duas quantidades podem estar muito afastadas. As comparações de média transversal em um evento (tal como morte) podem ser bastante ilusórias quando se tenta inferir as causas do evento. Discutiremos esses problemas mais adiante neste capítulo.

## TAXAS DE INCIDÊNCIA

### Pessoa-tempo e tempo da população

Os epidemiologistas frequentemente estudam desfechos que não são inevitáveis ou que podem não ocorrer durante o período de observação. Em tais situações, o cenário dos tempos de incidência para um evento específico em uma população não será definido, ou observado, precisamente. Um modo de lidar com essa complicação é desenvolver medidas que cubram a extensão do tempo em que cada indivíduo estava na *população exposta ao risco* para o evento, isto é, o período de tempo durante o qual o evento era uma possibilidade e teria sido contabilizado como um evento na população, caso ele ocorresse. Essa extensão, ou intervalo de tempo, é designada como contribuição *pessoa-tempo* do indivíduo.

A soma desses pessoas-tempo em todos os membros da população é chamada de *pessoa-tempo em risco total*, ou *tempo em risco da população*. Esse pessoa-tempo total deve ser diferenciado do tempo de relógio, pois é um somatório do tempo que ocorre simultaneamente para muitas pessoas, ao passo que o tempo do relógio não o é. O pessoa-tempo em risco total representa meramente o total de tempo durante o qual os inícios de doença poderiam ocorrer, e seriam considerados eventos ocorrendo na população de interesse.

### Taxas populacionais e taxas individuais

Definimos a *taxa de incidência* da população como o número de casos novos da doença (número incidente) dividido por pessoa-tempo durante o período:

$$\text{Taxa de incidência} = \frac{\text{número de inícios da doença}}{\sum_{\text{pessoas}} \text{tempo passado na população}}$$



Essa taxa também tem sido chamada *taxa pessoa-tempo*, *densidade de incidência*, *força de morbidade* (ou *força de mortalidade*, em relação a óbitos), *taxa de azar* e *intensidade de doença*, embora os três últimos termos sejam mais comumente usados para se referir ao limite teórico aproximado de uma taxa de incidência quando a unidade de medida de tempo se aproxima de zero.

Quando o período de risco é de duração fixa  $\Delta t$ , a proporção do período que uma pessoa passa na população em risco é sua quantidade de pessoa-tempo dividida por  $\Delta t$ . Portanto, o tamanho médio da população durante o período é:

$$\bar{N} = \sum_{\text{pessoas}} \frac{\text{tempo passado na população}}{\Delta t}$$

Então, o pessoa-tempo em risco total durante o período é igual ao produto do tamanho médio da população durante o período,  $\bar{N}$ , pela duração fixa do período de risco,  $\Delta t$ . Se designarmos o número incidente por  $A$ , então a taxa de incidência é igual a  $A/(\bar{N} \times \Delta t)$ . Essa formulação mostra que a taxa de incidência tem unidades de tempo inversas (por ano, por mês, por dia, etc.). Assim, as unidades ligadas a uma taxa de incidência podem ser escritas como  $\text{ano}^{-1}$ ,  $\text{mês}^{-1}$  ou  $\text{dia}^{-1}$ . Os únicos desfechos elegíveis para serem contabilizados no numerador de uma taxa de incidência são aqueles que ocorrem com pessoas que estejam contribuindo com tempo no denominador da mesma taxa, no momento do início da doença. Da mesma forma, somente o tempo contribuído por pessoas elegíveis para contabilização no numerador, caso sofram tal evento, deve ser contabilizado no denominador.

Outra maneira de expressar uma taxa de incidência populacional é como uma média ponderada pelo tempo das taxas individuais. Uma taxa individual é  $0/(\text{tempo passado na população}) = 0$ , se o indivíduo não experimentar o evento, ou então é  $1/(\text{tempo passado na população})$ , se o indivíduo vier a experimentar o evento. Temos, então, que o número de inícios de doença  $A$  é:

$$A = \sum_{\text{pessoas}} [(\text{tempo passado na população}) (\text{taxa individual})]$$

e assim,

$$\text{Taxa de incidência} = \frac{\sum_{\text{pessoas}} (\text{tempo passado na população}) (\text{taxa individual})}{\sum_{\text{pessoas}} \text{tempo passado na população}}$$

Essa fórmula mostra que a taxa de incidência ignora a distinção entre indivíduos que não contribuem para o número incidente  $A$ , porque só estiveram brevemente na população, e aqueles que não contribuem, porque estiveram na população por um tempo longo, mas nunca contraíram a doença (p. ex., indivíduos imunes). Dessa forma, a taxa de incidência lida com o problema de censura, ignorando distinções potencialmente importantes entre aqueles que *não* tiveram a doença. Embora a noção de uma taxa de incidência seja capital em epidemiologia, a formulação precedente mostra que ela não pode englobar todos os aspectos da ocorrência de doenças. Essa limitação também é mostrada pela observação de que uma taxa de 1 caso/(100 anos) =  $0,01 \text{ ano}^{-1}$  poderia ser obtida pelo acompanhamento de 100 pessoas por, em média, 1 ano e pela observação de um caso, mas também poderia ser obtida pelo seguimento de duas pessoas por 50 anos e pela ocorrência de um caso; um cenário muito diferente. Para distinguir essas situações, medidas de ocorrência mais detalhadas também são necessárias, tais como tempo de incidência.

## Interpretação apropriada das taxas de incidência

Além da falta de sensibilidade quanto a distinções importantes, as taxas de incidência têm dificuldades de interpretação, por serem confundidas, frequentemente, com riscos (probabilidades). Essa confusão surge quando se deixa de levar em conta a dependência da porção numérica de uma taxa das unidades usadas para sua expressão.

A porção numérica de uma taxa de incidência tem um limite inferior igual a zero e nenhum limite superior, que é o intervalo da razão de uma quantidade não negativa para uma positiva. As duas quantidades são o número de eventos no numerador e a pessoa-tempo no denominador. Pode ser surpreendente que uma taxa de incidência possa exceder o valor de 1, o que pareceria indicar que mais de 100% de uma população fosse afetada. No máximo 100% das pessoas em uma população podem contrair uma doença, mas a taxa de incidência não mede a proporção de uma população que contrai a doença e, na verdade, não é mesmo uma proporção. Lembre-se de que a taxa de incidência é mensurada em unidades do inverso do tempo. Entre 100 pessoas, não mais do que 100 mortes podem ocorrer, mas esses 100 óbitos podem ocorrer em 10.000 pessoas-ano, em 1.000 pessoas-ano, em 100 pessoas-ano ou em 1 pessoa-ano (se as 100 mortes acontecerem depois de uma média de 3,65 dias cada, como num combate militar). Uma taxa de incidência de 100 casos (ou óbitos) por 1 pessoa-ano poderia ser expressa como:

$$100 \frac{\text{casos}}{\text{pessoa-ano}}$$

Ela também poderia ser expressa como:

$$10.000 \frac{\text{casos}}{\text{pessoa-século}}$$

$$8,33 \frac{\text{casos}}{\text{pessoa-mês}}$$

$$1,92 \frac{\text{casos}}{\text{pessoa-semana}}$$

$$0,27 \frac{\text{casos}}{\text{pessoa-dia}}$$

O valor numérico de uma taxa de incidência não é interpretável, porque depende da seleção da unidade de tempo. Assim, na apresentação de taxas de incidência, é essencial fornecer a unidade de tempo utilizada para calcular a porção numérica. Tal unidade geralmente é escolhida para assegurar que a taxa mínima tenha pelo menos um dígito à esquerda da casa decimal. Por exemplo, uma tabela de taxas de incidência de 0,15, 0,04 e 0,009 casos por pessoa-ano pode ser multiplicada por 1.000 para ser exibida como 150, 40 e 9 casos por 1.000 pessoas-ano. Pode-se usar uma unidade tão grande quanto 1.000 pessoas-ano, não importando se as observações foram coletadas durante 1 ano de tempo, durante 1 semana de tempo ou durante uma década, da mesma maneira como se pode medir a velocidade de um veículo em termos de quilômetros por hora, mesmo que a mensuração seja feita por apenas poucos segundos.

## **Taxas de eventos recorrentes**

As taxas de incidência, frequentemente, incluem somente a primeira ocorrência de início de doença como um evento elegível para o numerador da taxa. Para muitas condições irreversíveis, tais como a esclerose múltipla, a cirrose ou a morte, uma pessoa pode experimentar, no máximo, somente um início. Para algumas doenças que podem recorrer, como a rinite, podemos querer medir simplesmente a incidência da “primeira” ocorrência, ou a primeira ocorrência após um período pré-especificado livre da doença, embora a doença possa ocorrer repetidamente. Para outras doenças, tais como câncer ou cardiopatia, a primeira ocorrência, frequentemente, é de interesse maior para o estudo etiológico do que as ocorrências subsequentes na mesma pessoa, porque o primeiro evento, ou seus tratamentos médicos, afetam a taxa de ocorrências subsequentes. Portanto, é típico que os eventos no numerador de uma taxa de incidência correspondam à primeira ocorrência de uma doença em particular, mesmo naqueles exemplos em que é possível para uma pessoa ter mais de uma ocorrên-

cia. Neste livro, presumiremos que estamos lidando com primeiras ocorrências, exceto quando declarado diferentemente. Conforme explicado mais adiante, as abordagens para primeiras ocorrências estendem-se, naturalmente, a ocorrências subsequentes, pela restrição da população em risco com base em ocorrência prévia.

Quando os eventos contabilizados no numerador de uma taxa de incidência são primeiras ocorrências de uma doença, então o tempo contribuído por cada pessoa na qual a doença se desenvolve deve terminar com o início da doença. A razão é que a pessoa não é mais elegível para apresentar o evento (a primeira ocorrência só pode acontecer uma vez por pessoa), portanto, não há mais informação sobre primeira ocorrência a ser obtida pela observação continuada dessa pessoa. Assim, cada indivíduo que experimente o desfecho deve contribuir com tempo para o denominador até a ocorrência de tal evento, mas não depois. Além disso, para o estudo de primeiras ocorrências, o número de inícios da doença no numerador da taxa de incidência é também uma contagem de pessoas experimentando o evento, porque só pode ocorrer um evento por pessoa.

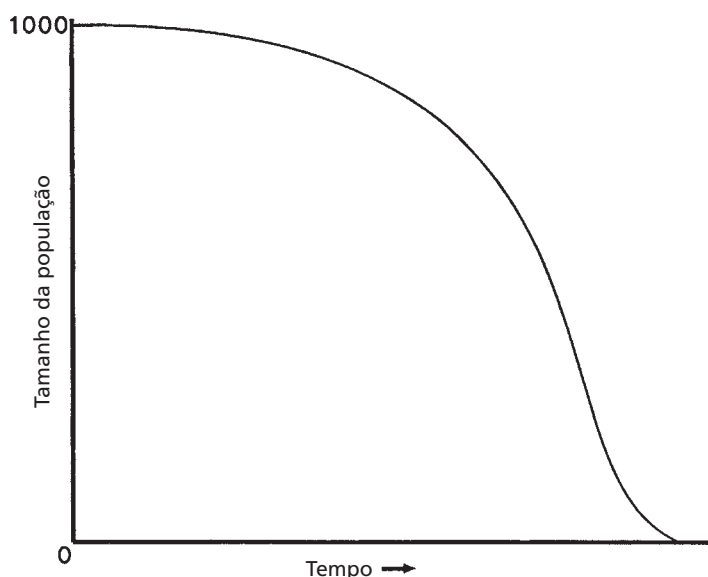
Um epidemiologista que deseje estudar tanto as primeiras ocorrências de doença como as subsequentes pode decidir não distinguir entre ocorrências iniciais e mais tardias e simplesmente contar todos os eventos que ocorreram na população sob observação. Nesse caso, o tempo acumulado no denominador da taxa não cessaria a ocorrência do desfecho, porque um evento adicional poderia acometer a mesma pessoa. Geralmente, entretanto, há bastante distinção biológica entre a primeira ocorrência e as subsequentes para justificar que sejam mensuradas separadamente. Uma abordagem é definir a “população exposta ao risco” de maneira diferente para cada ocorrência do evento: a população em risco para o primeiro evento consistiria em pessoas que não tivessem experimentado a doença antes; a população em risco para o segundo evento (que é a primeira recorrência) seria limitada àqueles que tivessem experimentado o evento uma vez, e apenas uma vez, e assim por diante. Dessa forma, estudos de cânceres secundários seriam restritos à população daqueles que sobreviveram ao seu primeiro câncer. Uma dada pessoa deve contribuir com tempo para o denominador da taxa de incidência para primeiros eventos somente até o tempo em que a doença ocorra pela primeira vez. Nesse ponto, a pessoa deve cessar a contribuição de tempo para o denominador daquela taxa e começar a contribuir com tempo para o denominador da taxa de medida da segunda ocorrência. Se e quando houver um segundo evento, a pessoa deve parar a contribuição de tempo para a taxa de mensuração da segunda ocorrência e começar a contribuir com tempo para o denominador da taxa de medida da terceira, e assim por diante.

## TIPOS DE POPULAÇÕES

### Populações fechadas

Dada uma escala de tempo particular para exibição da incidência, podemos distinguir as populações que estão *fechadas* ou *abertas* naquela escala. Uma população fechada não adiciona membros novos com o passar do tempo e só perde membros por morte, ao passo que uma população aberta pode adquirir membros com o tempo, por meio de imigração ou de nascimento, ou perder membros que ainda estão vivos, por meio de emigração, ou ambos. (Alguns demógrafos e ecologistas usam uma definição mais ampla para população fechada, na qual nascimentos são permitidos, mas não imigração ou emigração. Membros dessa população só podem deixá-la por falecimento.)

Suponha que coloquemos em um gráfico a experiência de sobrevida de uma população fechada que começa com 1.000 pessoas. Visto que a morte chegará para todos, depois de um período de tempo suficiente, o número original de 1.000 terá reduzido para zero. Um gráfico do tamanho da população com o tempo poderia se aproximar daquele na Figura 3.2. A curva faz um declive, porque à medida que as 1.000 pessoas da população vão morrendo, a população em risco de morte vai sendo reduzida. A população é fechada porque só consideramos o destino das 1.000 pessoas presentes no tempo zero. A experiência de pessoa-tempo dessas 1.000 pessoas é representada pela área sob a curva, no diagrama. Quando cada pessoa morre, a curva marca um ponto para baixo; aquela pessoa não mais contribui para o denominador pessoa-tempo da taxa de óbitos (mortalidade). A contribuição de cada pessoa é



**FIGURA 3.2** • Tamanho de uma população fechada de 1.000 pessoas, pelo tempo.

exatamente igual ao período de tempo no qual é acompanhada do princípio ao fim. Nesse exemplo, como toda a população é acompanhada até o óbito, o final é a morte da pessoa. Em outros exemplos, a contribuição da experiência de pessoa-tempo continuaria até o início da doença ou até algum tempo arbitrário de corte da observação, o que ocorresse mais cedo.

Suponha que somássemos a experiência de pessoa-tempo total dessa população fechada de 1.000 e obtivéssemos um total de 75.000 pessoas-ano. A taxa de mortalidade seria  $(1.000/75.000) \times \text{ano}^{-1}$ , porque as 75.000 pessoas-ano representariam a experiência de todas as 1.000 pessoas até suas mortes. Além do mais, se o tempo é mensurado desde o início do seguimento, o tempo médio de morte nessa população fechada seria  $75.000 \text{ pessoas-ano} / 1.000 \text{ pessoas} = 75 \text{ anos}$ , que é o inverso da taxa de mortalidade.

Uma população fechada com uma taxa de mortalidade constante no decorrer do tempo declinaria em tamanho de forma exponencial (que é o que se quer dizer pelo termo *declínio exponencial*). Na prática, entretanto, as taxas de mortalidade em uma população fechada mudam com o tempo, porque a população envelhece à medida que o tempo progride. Consequentemente, a curva de declínio de uma população humana fechada nunca é exponencial. A *metodologia de tábua de vida* é um procedimento pelo qual a taxa de mortalidade (ou a taxa de doença) de uma população fechada é avaliada dentro de pequenos intervalos sucessivos de idade ou de tempo, de modo que a dependência da mortalidade em razão da idade ou do tempo possa ser elucidada. Com qualquer método, entretanto, é importante distinguir os efeitos relacionados com a idade daqueles relativos a outros eixos de contagem do tempo, porque a idade de cada pessoa aumenta diretamente com um aumento ao longo de qualquer linha de tempo. Por exemplo, a idade de uma pessoa aumenta com o crescimento da duração do emprego, com a progressão do tempo do calendário e com o tempo crescente a partir do início do seguimento.

## Populações abertas

Uma população aberta difere de uma fechada, pois a população exposta ao risco está aberta a novos membros que não se qualificavam inicialmente. Um exemplo de população aberta é a população de um país. As pessoas podem ingressar na população aberta por meio de vários mecanismos. Algumas

podem nascer nela, outras podem migrar para ela. Para uma população aberta de pessoas que tenham atingido uma idade específica, as pessoas podem tornar-se elegíveis para entrar na população ao envelhecer até tal idade. Similarmente, as pessoas podem sair por falecimento, por envelhecer até uma idade fora da faixa etária definida, por emigração ou por ficarem doentes (o último método de saída aplica-se apenas se os primeiros surtos de uma doença estão sendo estudados). As pessoas também podem sair de uma população aberta e depois reingressar, por exemplo, emigrando da área geográfica na qual a população está localizada, e mais tarde retornando para a mesma área.

A distinção entre populações fechadas e abertas depende, em parte, do eixo de tempo usado para descrever a população, assim como de qual maneira se define a afiliação. Todas as pessoas que alguma vez usaram uma droga em particular constituíam uma população fechada, se o tempo fosse medido a partir do início de uso da droga. Essas pessoas, entretanto, comporiam uma população aberta no tempo do calendário, porque novos usuários poderiam se acumular durante um período de tempo. Se, como nesse exemplo, a condição de membro da população sempre começa com um evento, tal como o início de um tratamento, e nunca termina daí em diante, a população é fechada ao longo do eixo de tempo que marca esse evento como o tempo zero para cada membro, porque todos os novos membros somente ingressam quando experimentam o evento. A mesma população, no entanto, estará aberta ao longo da maioria dos outros eixos de tempo. Se a condição de membro pode ser interrompida por eventos mais tardios, que não a morte, trata-se de uma população aberta ao longo de qualquer eixo de tempo.

Pelas definições supracitadas, qualquer população de estudo com perdas no seguimento é aberta. Por exemplo, a condição de membro de uma população em estudo pode ser definida, em parte, por estar sob vigilância ativa para doença; em tal caso, os membros que se perdem do seguimento, por definição, saíram da população, mesmo que ainda estejam vivos e, afora isso, estão elegíveis para o estudo. É prática comum analisar tais populações usando-se o tempo de começo da observação, um eixo ao longo do qual nenhuma imigração pode ocorrer (por definição, o tempo zero é quando a pessoa entra no estudo). Tais populações podem ser designadas como “fechadas à esquerda” e frequentemente são denominadas “coortes fixas”, embora o termo *coorte* seja mais usado para se referir a um conceito diferente, que discutiremos em seguida.

## Populações versus coortes

O termo *população*, como o usamos aqui, tem um elemento intrinsecamente temporal e potencialmente dinâmico: uma pessoa pode ser um membro em um certo tempo, não ser membro mais tarde, voltar a ser membro, e assim por diante. Esse uso é o senso mais comum de população, como com a população de uma cidade ou de um país. O termo *coorte* algumas vezes é usado para descrever qualquer população de estudo, mas o reservamos para um conceito mais estreito, que é o de um grupo de pessoas para as quais a condição de membro é definida de modo permanente, ou o de uma população na qual a condição de membro é determinada totalmente por um só evento definidor e, assim, torna-se permanente. Um exemplo de uma coorte seria o dos membros da classe de formandos de uma escola em um dado ano. A lista dos membros da coorte é fixada por ocasião da graduação, e não aumentará. Outros exemplos incluem a coorte de todas as pessoas que já usaram um medicamento, e a coorte de pessoas recrutadas para um estudo de seguimento. No último caso, a população de estudo pode começar com todos os membros da coorte, mas pode reduzir, gradualmente, quando aqueles recrutados inicialmente são perdidos no seguimento. Os que são perdidos do seguimento permanecem membros da coorte inicialmente recrutada, embora não estejam mais na população do estudo. Com essa definição, os membros de qualquer coorte constituem uma população fechada ao longo do eixo de tempo no qual o evento definidor (p. ex., nascimento com síndrome de Down ou recrutamento para estudo) é tomado como tempo zero. Uma *coorte de nascimento* é a coorte definida, em parte, por haver nascimento em um tempo específico (p. ex., todas as pessoas nascidas na Etiópia em 1990 constituem a coorte de nascimento etíope para 1990).

## Estado de equilíbrio

Se o número de pessoas ingressando na população é contrabalançado pelo número dos que saem, em qualquer período de tempo, dentro de categorias de idade, de sexo e de outros determinantes de risco, diz-se que a população é *estacionária*, ou está em um *estado de equilíbrio*. O estado de equilíbrio é uma propriedade que só pode ocorrer em populações abertas, e não em populações fechadas. É possível, contudo, existir uma população em estado de equilíbrio na qual não esteja ocorrendo imigração, nem emigração; essa situação requereria que os nascimentos se contrabalançassem perfeitamente com os óbitos na população. O gráfico do tamanho de uma população aberta em estado de equilíbrio é simplesmente uma linha horizontal, conforme mostrado na Figura 3.3. As pessoas estão entrando e saindo da população continuamente.

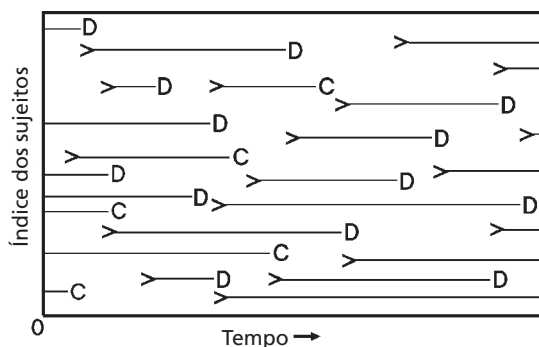
No diagrama, o símbolo > representa uma pessoa ingressando na população, um segmento de reta representa a experiência de pessoa-tempo dessa pessoa, e o término de um segmento de reta representa o fim de sua experiência. Um terminal D indica que a experiência terminou por causa do início de doença, e um terminal C sinaliza que ela acabou por outras razões. Teoricamente, qualquer intervalo de tempo fornecerá uma boa estimativa da taxa de incidência em uma população estacionária.

## Relação de taxas de incidência com tempos de incidência em populações especiais

O inverso do tempo é um conceito complicado que não oferece um alcance intuitivo de uma taxa de incidência. A medida, entretanto, realmente tem uma conexão íntima com mensurações de ocorrência mais interpretáveis em populações fechadas. Em referência à Figura 3.2, pode-se ver que a área sob a curva é igual a  $N \times T$ , onde  $N$  é o número de pessoas começando em uma população fechada, e  $T$  é o tempo médio até a morte. A taxa de mortalidade por tempo médio, então, é  $N/(N \times T) = 1/T$ ; isto é, a taxa de mortalidade iguala-se ao inverso do tempo médio até a morte.

Em uma população estacionária sem migração, a taxa de incidência bruta de um desfecho inevitável, tal como a morte, será igual ao inverso do tempo médio passado na população até que o desfecho ocorra (Morrison, 1979). Assim, em uma população estacionária sem migração, uma taxa de mortalidade de  $0,04 \text{ ano}^{-1}$  seria traduzida em um tempo médio da entrada até o óbito de 25 anos. De modo semelhante, em uma população estacionária sem migração, a idade média de morte se iguala à expectativa de vida. O tempo passado na população até que o desfecho ocorra é designado, às vezes, como o *tempo de espera* até que o evento aconteça e corresponde ao tempo de incidência, quando o tempo é medido a partir da entrada na população.

Se o desfecho de interesse não é morte, mas o início de doença ou morte por uma causa específica, a interpretação do tempo médio deve ser modificada para levar em consideração os *riscos competitivos*, que são eventos que “competem” com o desfecho de interesse, para remover pessoas



**FIGURA 3.3** • Composição de uma população aberta em estado de equilíbrio aproximado, por tempo; > indica entrada na população, D indica início da doença e C indica saída da população sem doença.

da população exposta ao risco. Mesmo que não haja risco competitivo, a interpretação das taxas de incidência, como o inverso do tempo médio de espera, não será válida se houver migração (tal como perda do seguimento), e a idade média no óbito não mais será igual à expectativa de vida. Por exemplo, a taxa de mortalidade para os Estados Unidos em 1977 foi de  $0,0088 \text{ ano}^{-1}$ . Em um estado de equilíbrio, essa taxa corresponderia a uma duração de vida média, ou expectativa de vida, de 114 anos. Outras análises, entretanto, indicam que a expectativa de vida real em 1977 era de 73 anos (Alho, 1992). A discrepância é um resultado da imigração e da falta de um estado de equilíbrio. Observe que a suposição de não migração é inviável dentro de faixas etárias específicas, pois as pessoas estão sempre “migrando” para dentro e para fora de grupos etários, à medida que envelhecem.

## OUTROS TIPOS DE TAXAS

Em adição a números de casos por unidade de pessoa-tempo, algumas vezes é útil examinar números de eventos por outra unidade. Em serviços de saúde e epidemiologia de doenças infecciosas, as curvas epidêmicas são ilustradas, frequentemente, em termos do número de casos por unidade de tempo, também chamada de *taxa absoluta*,

$$\frac{\text{nº de inícios de doença}}{\text{período de tempo de observação}}$$

ou  $A/\Delta t$ . Visto que a taxa pessoa-tempo é simplesmente essa taxa absoluta dividida pelo tamanho médio da população durante o período de tempo, ou  $A/(\bar{N} \times \Delta t)$ , a taxa pessoa-tempo tem sido chamada de *taxa relativa* (Elandt-Johnson, 1975); a taxa absoluta é relativa ao, ou “ajustada para”, tamanho médio da população.

Algumas vezes, é útil expressar taxas de eventos em unidades que não envolvam diretamente o tempo. Um exemplo comum é a expressão de fatalidades por modalidade de viagem na relação passageiros/milhas, comparando-se a segurança do trem comercial e a da viagem aérea. Nesse caso, pessoas-milha substitui pessoas-tempo no denominador da taxa. Como as taxas com tempo no denominador, a porção numérica de tais taxas é completamente dependente da escolha das unidades de medida; uma taxa de 1,6 mortes por  $10^6$  passageiros/milhas equivale a uma taxa de 1 morte por  $10^6$  passageiros/quilômetros.

O conceito central do uso preciso do termo *taxa de incidência* é a expressão da mudança de número incidente em relação à alteração de uma outra quantidade, de modo que a taxa de incidência sempre tenha uma dimensão. Assim, uma taxa pessoa-tempo expressa o aumento do número incidente que esperamos pelo aumento de uma unidade de pessoa-tempo. Uma taxa absoluta expressa o aumento de número incidente que esperamos pelo aumento de uma unidade no tempo de relógio, e uma taxa pessoas-milhas expressa o aumento de número incidente que esperamos pelo aumento de uma unidade em milhas dos passageiros.

## PROPORÇÕES DE INCIDÊNCIA E PROPORÇÕES DE SOBREVIVÊNCIA

Em um dado intervalo de tempo, também podemos expressar o número incidente de casos em relação ao tamanho da população exposta ao risco. Se mensuramos o tamanho da população no começo de um intervalo de tempo e ninguém ingressar na população (imigrar), ou deixá-la (emigrar), depois do início do intervalo, tal taxa se torna a proporção de pessoas que passam a ser casos entre aqueles na população no início do intervalo. Chamamos essa quantidade de *proporção de incidência*, que também pode ser definida como a proporção de uma população fechada exposta ao risco, que se torna doente dentro de um dado período de tempo. Essa quantidade é chamada, às vezes, de *incidência cumulativa*, mas esse termo também é usado para outra quantidade, que discutiremos mais tarde. Um termo mais tradicional para proporção de incidência é *taxa de ataque*, mas reservamos o termo *taxa* para taxas de incidência pessoa-tempo.



Se *risco* é definido como a probabilidade de que uma doença se desenvolva em uma pessoa dentro de um intervalo de tempo especificado, então proporção de incidência é uma medida, ou estimativa, de risco médio. Embora esse conceito de risco aplique-se a indivíduos, ao passo que a proporção de incidência se aplica a populações, a proporção de incidência às vezes é chamada de risco. Esse uso é consistente com a visão de que os riscos individuais referem-se meramente à frequência relativa de doença em um grupo de indivíduos, como o que está em discussão. *Risco médio* é um sinônimo mais preciso, o qual usaremos algumas vezes.

Outra maneira de expressar a proporção de incidência é por uma simples média das proporções individuais. A última é 0, para aqueles que não tiveram o evento, ou é 1, para aqueles que o experimentaram. O número de inícios de doença  $A$  é, então, a soma das proporções individuais,

$$A = \sum_{\text{pessoas}} \text{proporções individuais}$$

e assim,

$$\text{Proporção de incidência} = \frac{\sum_{\text{pessoas}} \text{proporções individuais}}{\text{tamanho inicial da população}} = A/N$$

Se as proporções individuais são chamadas de “riscos individuais”, esta fórmula mostra outro entendimento, no qual a proporção de incidência também é um “risco médio”. Isso também torna claro que a proporção de incidência ignora a quantidade de pessoa-tempo contribuída pelos indivíduos e, assim, despreza ainda mais informações do que a taxa de incidência, embora tenha uma interpretação mais intuitiva.

Como qualquer proporção, o valor de uma proporção de incidência varia de 0 a 1 e não tem dimensão. A proporção de incidência não é interpretável, contudo, sem a especificação do período de tempo ao qual se aplica. Uma proporção de incidência de morte de 3% significa algo muito diferente quando se refere a um período de 40 anos do que quando se refere a um período de 40 dias.

Uma medida útil, complementar à proporção de incidência, é a *proporção de sobrevida*, que pode ser definida como a proporção de uma população fechada exposta ao risco que *não* fica doente dentro de um dado período de tempo. Se  $R$  e  $S$  denotam as proporções de incidência e sobrevida, então  $S = 1 - R$ . Outra medida comumente usada é chamada de *chance de incidência*, definida como  $R/S = R/(1 - R)$ , a razão da proporção contraindo a doença sobre a proporção não a contraindo. Se  $R$  é pequena,  $S \approx 1$  e  $R/S \approx R$ ; isto é, a chance de incidência se aproximará da proporção de incidência quando ambas as quantidades forem pequenas. Caso contrário, posto que  $S < 1$ , as chances de incidência serão maiores do que a proporção de incidência e, diferentemente da última, poderão exceder a 1.

Para intervalos de tempo suficientemente curtos, há uma relação muito simples entre a proporção de incidência e a taxa de incidência de um evento não recorrente. Considere uma população fechada durante um intervalo  $t_0$  a  $t_1$ , e que  $\Delta t = t_1 - t_0$  seja o comprimento do intervalo. Se  $N$  é o tamanho da população em  $t_0$ , e  $A$  é o número de inícios de doença durante o intervalo, então as proporções de incidência e de sobrevida durante o intervalo são  $R = A/N$  e  $S = (N - A)/N$ . Agora suponha que o intervalo de tempo seja curto o bastante para que o tamanho da população decline apenas levemente durante o intervalo. Então,  $N - A \approx N$ ,  $S \approx 1$ , e assim  $R/S \approx R$ . Além do mais, o tamanho médio da população exposta ao risco será aproximadamente  $N$ , de sorte que o pessoa-tempo em risco total durante o intervalo será aproximadamente  $N\Delta t$ . Assim, a taxa de incidência ( $I$ ) durante o intervalo será aproximadamente  $A/N\Delta t$ , e obteremos:

$$R = A/N = (A/N\Delta t)\Delta t \approx I\Delta t \text{ e } R \approx R/S$$

Posto em palavras, a proporção de incidência, as chances de incidência e a quantidade  $I\Delta t$  serão todas próximas umas das outras se a população exposta ao risco declinar apenas levemente durante o intervalo. Podemos fazer com que essa aproximação se mantenha dentro de uma acurácia de  $1/N$ ,



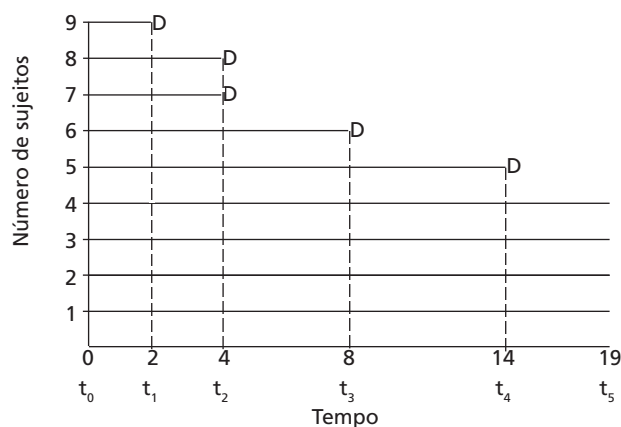
tornando  $\Delta t$  tão curto que não mais do que uma pessoa saia da população exposta ao risco durante o intervalo. Assim, dado um intervalo de tempo suficientemente curto, pode-se simplesmente multiplicar a taxa de incidência pelo período de tempo para se aproximar da proporção de incidência. Essa aproximação oferece outra interpretação para a taxa de incidência: pode ser vista como o valor limite da razão entre o risco médio e a duração do tempo em risco, quando esta última se aproxima de zero.

Um tipo específico de proporção de incidência é a *taxa de letalidade*, que é a proporção de incidência de morte entre aqueles nos quais uma doença se desenvolve (portanto, ela não é uma taxa em nosso entendimento, e sim uma proporção). O período de tempo para mensuração da taxa de letalidade frequentemente não é declarado, mas sempre é melhor especificá-lo.

## RELAÇÕES ENTRE MEDIDAS DE INCIDÊNCIA

A ocorrência de doença em uma população reflete dois aspectos de experiências individuais: a quantidade de tempo em que o indivíduo está exposto ao risco na população e se o indivíduo realmente tem o evento focal (p. ex., contrai a doença) durante aquele tempo. Medidas de incidência diferentes sumarizam aspectos distintos da distribuição dessas experiências. O tempo médio de incidência é o tempo médio até um evento, e a proporção de incidência é o “risco” médio de um evento (onde “risco” é 1 ou 0, conforme o evento tenha ocorrido ou não no período de risco). Cada uma dessas medidas é fácil de captar intuitivamente, mas, frequentemente, não são fáceis de estimar, ou mesmo de definir. Em contraste, a taxa de incidência pode ser aplicada à situação comum na qual o tempo de risco e a ocorrência do evento podem ser determinados sem ambiguidade para todos. Infelizmente, pode ser difícil compreender de maneira correta o que a taxa indica sobre as dimensões diferentes das distribuições do evento, e assim é útil compreender sua relação com os tempos de incidência e com as proporções de incidência. Essas relações constituem um componente central dos tópicos de análise de sobrevivência e de análise do tempo de falha em estatística (Kalbfleisch e Prentice, 2002; Cox e Oakes, 1984).

Há relações relativamente simples entre a proporção de incidência de um evento inevitável, não recorrente (tal como morte), e a taxa de incidência em uma população fechada. Para ilustrá-las, consideraremos a pequena população fechada mostrada na Figura 3.4. O tempo em risco (história de risco) de cada membro está grafado do mais curto, no alto, ao mais longo, embaixo. Cada história termina com um D, indicando a ocorrência do evento de interesse, ou no fim do seguimento, em  $t_5 = 19$ . O tempo inicial é denotado como  $t_0$  e aqui é igual a 0. Cada vez que um ou mais eventos ocorrem, isso é marcado por uma linha vertical tracejada, os tempos de eventos isolados são assinalados como  $t_i$  (o que ocorreu mais cedo) a  $t_4$ , e o fim do seguimento é denotado por  $t_5$ . Marcamos o número de eventos no tempo  $t_k$  como  $A_k$ , o número total de pessoas expostas ao risco no tempo  $t_k$  (inclusive as pessoas  $A_k$  que experimentam o evento) como  $N_k$ , e o número de pessoas vivas ao fim do seguimento como  $N_5$ .



**FIGURA 3.4** • Exemplo de uma população fechada pequena com fim do seguimento aos 19 anos.

## Fórmula produto-limite

A Tabela 3.1 mostra a história da população durante o período de seguimento de 20 anos na Figura 3.4, em termos de  $t_k$ ,  $A_k$  e  $N_k$ . Observe que, pelo fato de a população ser fechada e o evento ser inevitável, o número remanescente exposto ao risco após  $t_k$ ,  $N_{k+1}$ , é igual a  $N_k - A_k$ , que é o número em risco até  $t_k$  menos o número experimentando o evento em  $t_k$ . A proporção da população permanecendo em risco até  $t_k$ , a qual também continua exposta ao risco após  $t_k$ , é, portanto:

$$s_k = \frac{N_k - A_k}{N_k} = \frac{N_{k+1}}{N_k}$$

Podemos ver agora que a proporção da população original que permanece em risco ao fim do seguimento é

$$S = N_5/N_1 = (N_5/N_4)(N_4/N_3)(N_3/N_2)(N_2/N_1) = s_4 s_3 s_2 s_1$$

que pela Tabela 3.1 é

$$S = (4/5)(5/6)(6/8)(8/9) = 4/9$$

Esta fórmula de multiplicação diz que a proporção de sobrevida durante todo o intervalo de tempo na Figura 3.4 é só o produto das proporções de sobrevida para cada subintervalo de  $t_{k-1}$  a  $t_k$ . Em sua forma mais geral,

$$S = \prod_{k=1}^5 \frac{N_k - A_k}{N_k} \quad [3.1]$$

Essa fórmula é chamada de fórmula de *Kaplan-Meier*, ou fórmula de *produto-limite* (Kalbfleisch e Prentice, 2002; Cox e Oakes, 1984).

## Fórmula exponencial

Agora, considere que  $T_k$  seja o pessoa-tempo em risco na população total, durante o subintervalo de  $t_{k-1}$  a  $t_k$ , e que  $\Delta t_k = t_k - t_{k-1}$  seja o comprimento do subintervalo. Visto que a população é de tamanho constante  $N_k$  durante esse subintervalo, e todos ainda presentes contribuem  $\Delta t_k$  unidades pessoa-tempo em risco, o pessoa-tempo em risco total no intervalo é  $N_k \Delta t_k$ , de modo que a taxa de incidência no tempo seguinte a  $t_{k-1}$  até (mas não além de)  $t_k$  é

$$I_k = \frac{A_k}{N_k \Delta t_k}$$

Porém, a proporção de incidência durante o mesmo subintervalo é igual a  $I_k \Delta t_k$ , de modo que a proporção de sobrevida no intervalo é

$$s_k = 1 - I_k \Delta t_k$$

Assim, podemos substituir  $1 - I_k \Delta t_k$  no lugar de  $s_k$  na equação anterior de  $S$ , a proporção de sobrevida geral, para obter

$$\begin{aligned} S &= (1 - I_5 \Delta t_5)(1 - I_4 \Delta t_4)(1 - I_3 \Delta t_3)(1 - I_2 \Delta t_2)(1 - I_1 \Delta t_1) \\ &= [1 - (0/5)][1 - (1/30)6][1 - (1/24)4][1 - (2/16)2][1 - (1/18)2] \\ &= 4/9 \end{aligned}$$

como antes.

**TABELA 3.1****Tempos de eventos e intervalos para a população fechada na Figura 3.4**

	Começo	Tempos dos desfechos ( $t_k$ )				Fim
	0	2	4	8	14	19
Índice ( $k$ )	0	1	2	3	4	5
Nº de desfechos ( $A_k$ )	0	1	2	1	1	0
Nº em risco ( $N_k$ )	9	9	8	6	5	4
Proporção sobrevivendo ( $S_k$ )		8/9	6/8	5/6	4/5	4/4
Duração do intervalo ( $\Delta t_k$ )		2	2	4	6	5
Pessoa-tempo ( $N_k \Delta t_k$ )		18	16	24	30	20
Taxa de incidência ( $I_k$ )		1/18	2/16	1/24	1/30	0/20

Se cada uma das proporções de incidência do subintervalo de  $I_k \Delta t_k$  é pequena ( $< 0,10$  ou próximo disso), podemos simplificar a última fórmula pelo fato de que, para  $x$  pequeno,

$$1 - x \approx \exp(-x)$$

Tomando  $x = I_k \Delta t_k$  nesta fórmula de aproximação, obtemos  $1 - I_k \Delta t_k \approx \exp(-I_k \Delta t_k)$ , e assim

$$\begin{aligned} S &\approx \exp(-I_5 \Delta t_5) \exp(-I_4 \Delta t_4) \cdots \exp(-I_1 \Delta t_1) \\ &= \exp(-I_5 \Delta t_5 - I_4 \Delta t_4 - \cdots - I_1 \Delta t_1) \\ &= \exp\left(-\sum_{k=1}^5 I_k \Delta t_k\right) \end{aligned}$$

que pela Tabela 3.1, obtém-se

$$\exp[-0(5) - (1/30)6 - (1/24)4 - (2/16)2 - (1/18)2] = 0,483$$

não muito longe do valor anterior de  $4/9 = 0,444$ . Finalmente, usamos o fato de que a proporção de incidência para o período total é  $1 - S$  para obter

$$R = 1 - S \approx 1 - \exp\left(-\sum_{k=1}^5 I_k \Delta t_k\right) \quad [3-2]$$

A última fórmula é citada em muitos livros e chamada algumas vezes de *fórmula exponencial* para relacionar taxas a proporções de incidência. A soma no expoente,  $\sum_k I_k \Delta t_k$ , é chamada algumas vezes de *incidência cumulativa* (Breslow e Day, 1980) ou *azar cumulativo* (Kalbfleisch e Prentice, 2002; Cox e Oakes, 1984). De modo confuso, o termo *incidência cumulativa* é usado mais frequentemente para denotar a proporção de incidência. O azar cumulativo, embora não tenha unidade, não é uma proporção e excederá 1,0 quando a proporção de incidência ultrapassar  $1 - e^{-1} = 0,632$ .

Queremos enfatizar as suposições que usamos para derivar a fórmula exponencial na equação 3.2:

1. A população é fechada.
2. O evento em estudo é inevitável (não há risco competitivo).

3. O número de eventos  $A_k$  em cada tempo  $t_k$  de evento é uma pequena proporção do número em risco  $N_k$  naquele tempo (i.e.,  $A_k/N_k$  sempre é pequeno).

Se a população não é muito pequena, podemos sempre forçar que a suposição 3 se mantenha, medindo o tempo de modo que cada evento ocorra em seu próprio tempo (de modo que somente um evento ocorra em cada  $t_k$ ). Na Tabela 3.1, a discrepância entre o  $R$  verdadeiro de  $5/9 = 0,555$  e o valor da fórmula exponencial de  $1 - 0,483 = 0,517$  é um tanto pequena, considerando que  $A_k/N_k$  fica tão grande quanto  $2/8 = 0,25$  (em  $t_2$ ).

As suposições 1 e 2 também foram usadas para derivar a fórmula produto-limite na equação 3.1. Elas raramente são satisfeitas, entretanto são frequentemente negligenciadas em apresentações e aplicações das fórmulas. Alguma forma da suposição de população fechada ( $n^\circ 1$ ) é essencial, porque a proporção de incidência só é definida com referência a populações fechadas. Um uso importante das fórmulas produto-limite e exponencial é, contudo, na tradução de estimativas de taxa de incidência de populações abertas para estimativas de proporção de incidência em uma população fechada de interesse. Ao assumir que as taxas de incidência nas duas populações são as mesmas em cada tempo, pode-se justificar a substituição das proporções de sobrevida  $(N_k - A_k)/N_k$ , ou das taxas de incidência observadas na população aberta na fórmula produto-limite, ou fórmula exponencial. Esta suposição frequentemente é plausível quando a população aberta que se observa é um subgrupo da população fechada de interesse, como em um estudo de coorte com perdas do seguimento que não estão relacionadas com o risco.

## Aplicações com riscos competitivos

Quando riscos competitivos removem pessoas da população em risco, a aplicação das fórmulas do produto-limite e exponencial requer novos conceitos e suposição. Considere as taxas de incidência específicas por subintervalo para nossa população fechada de interesse. Quando riscos competitivos estão presentes, a fórmula produto-limite na equação 3.1 para a proporção de sobrevida  $S$  não é mais válida, porque os riscos competitivos podem remover pessoas adicionais entre tempos de início de doença, e neste caso  $N_{k+1}$  será menor do que  $N_k - A_k$ . Além disso, quando riscos competitivos ocorrem entre  $t_{k-1}$  e  $t_k$ , o tamanho da população não será constante durante o subintervalo. Consequentemente, o pessoa-tempo no intervalo  $k$  não será igual a  $N_k \Delta t_k$ , e  $I_k \Delta t_k$  não se igualará a  $A_k/N_k$ . Assim, a fórmula exponencial da equação 3.2 não será válida se ocorrerem riscos competitivos. A quantidade de erro dependerá da frequência dos eventos que são os riscos competitivos.

Podemos, contudo, levantar a seguinte questão: qual seria a proporção de incidência durante o intervalo total se não tivesse ocorrido risco competitivo? Essa quantidade às vezes é chamada de *risco condicional* do desfecho (condicionado à remoção de riscos competitivos). Um menos a fórmula produto-limite da equação 3.1 dá uma estimativa dessa quantidade, sob a suposição de que as taxas de incidência específicas por subintervalo não mudariam se nenhum risco competitivo ocorresse (*riscos competitivos independentes*). Um menos a fórmula exponencial da equação 3.2 permanece uma aproximação dela, se  $A_k/N_k$  sempre for pequena. A suposição de que as taxas não mudariam se os riscos competitivos fossem removidos é feita por quase todos os métodos de análise de sobrevida para estimar riscos condicionais, mas isso requer escrutínio cuidadoso. Sob condições que eliminassem os riscos competitivos, as taxas de incidência do desfecho em estudo provavelmente poderiam mudar. Suponha, por exemplo, que o desfecho de interesse seja câncer do cólon. Os riscos competitivos incluiriam mortes por quaisquer outras causas que não esse tipo de câncer. A remoção de tantos riscos seria praticamente impossível, mas uma tentativa de minimizá-los poderia envolver intervenções dietéticas para prevenir óbitos por outros cânceres e por cardiopatia. Se tais intervenções fossem efetivas, também poderiam reduzir as taxas de câncer do cólon, violando, assim, a suposição de que as taxas específicas não se modificariam se nenhum risco competitivo ocorresse.

Devido à falta de praticidade em remover riscos competitivos sem alterar as taxas da doença em estudo, muitos autores alertam contra a interpretação de estimativas de sobrevida com base na remoção estatística de riscos competitivos como sendo a incidência que realmente ocorreria se tais riscos

forem removidos (Cox e Oakes, 1984; Prentice e Kalbfleisch, 1988; Pepe e Mori, 1993; Kalbfleisch e Prentice, 2002; Greenland, 2002a, 2005a; Alberti et al., 2003). Uma abordagem discutivelmente mais prática é focar o risco (probabilidade) do desfecho sem remover os riscos competitivos, o que se chama algumas vezes de “risco não condicional”. Essa quantidade pode ser estimada sem suposição alguma sobre remoção de riscos competitivos (Benichou e Gail, 1990a; Pepe e Mori, 1993; Gooley et al., 1999; Kalbfleisch e Prentice, 2002).

Um uso mais defensável da fórmula produto-limite é estimar proporções de sobrevida na presença de censura (p. ex., perda do seguimento); ver Capítulo 16 para detalhes. Para esse uso, faz-se a suposição de que a censura é aleatória (p. ex., a experiência dos sujeitos perdidos difere apenas de modo aleatório daquela experiência dos que não foram perdidos). Essa suposição frequentemente é questionável, mas sua abordagem é viável no delineamento e na análise de um estudo por meio de mensuração e de ajuste para fatores que afetem a censura e o desfecho (Kalbfleisch e Prentice, 2002).

## Relação entre proporção de sobrevida e tempo médio de incidência

Retornando à situação mais simples de um desfecho inevitável não recorrente em uma população fechada, derivaremos uma equação, relacionando proporções de sobrevida ao tempo médio de incidência. Primeiramente, podemos escrever o pessoa-tempo em risco total sobre o intervalo total na Figura 3.4 como:

$$N_1 \Delta t_1 + \dots + N_5 \Delta t_5 = \sum_{k=1}^5 N_k \Delta t_k = 18 + 16 + 24 + 30 + 20 = 108 \text{ pessoas-ano}$$

Assim, o tempo médio em risco contribuído pelos membros da população durante o intervalo, é:

$$\frac{1}{N_0} \sum_{k=1}^5 N_k \Delta t_k = \sum_{k=1}^5 \left( \frac{N_k}{N_0} \right) \Delta t_k = \frac{1}{9} 108 = 12 \text{ anos}$$

Observe que  $N_k/N_0$  é apenas a proporção que permanece exposta ao risco até  $t_k$ , isto é, a proporção de sobrevida de  $t_0$  a  $t_k$ . Se representarmos essa proporção  $N_k/N_0$  por  $S_{0,k}$  (para diferenciá-la das proporções específicas por subintervalo de  $s_k$ ), o tempo médio em risco pode ser escrito

$$\sum_{k=1}^5 S_{0,k} \Delta t_k = \left( \frac{9}{9} \right) 2 + \left( \frac{8}{9} \right) 2 + \left( \frac{6}{9} \right) 4 + \left( \frac{5}{9} \right) 6 + \left( \frac{4}{9} \right) 5 = 12 \text{ anos}$$

como antes. Agora suponha que o intervalo é estendido para adiante no tempo, até que a população inteira tenha experimentado o desfecho de interesse, como na Figura 3.2. O tempo médio em risco, então, será igual ao tempo médio de incidência, de modo que o tempo médio de incidência será computável a partir das proporções de sobrevida usando a última fórmula. As proporções de sobrevida podem, por sua vez, ser calculadas a partir das taxas de incidência específicas por subintervalo, como descrito anteriormente.

## Resumo

Os três tipos amplos de medidas de frequência de doenças – tempo de incidência, taxa de incidência e proporção de incidência (e sobrevida) – estão ligados por fórmulas matemáticas simples, quando se considera um evento inevitável não recorrente em uma população fechada, com seguimento até que todos tenham experimentado o evento. As relações matemáticas tornam-se mais complexas quando se consideram eventos com riscos competitivos, populações abertas ou períodos de risco truncados. As interpretações tornam-se particularmente problemáticas quando riscos competitivos estão presentes.

## LIMITAÇÕES E GENERALIZAÇÕES DAS MEDIDAS BÁSICAS DE OCORRÊNCIA

Todas as medidas mencionadas podem ser vistas como tipos de médias aritméticas, passíveis de serem ponderadas. Consequentemente, é objetivo estendê-las a desfechos que sejam mais detalhados do que tudo ou nada. Considere como alvo, por exemplo, o título de anticorpos após vacinação de pessoas sem anticorpos. Poderíamos examinar o número médio de dias até que o título atinja qualquer nível dado  $L$  ou a proporção que alcance  $L$  dentro de um determinado número de dias. Poderíamos também examinar a taxa na qual as pessoas atinjam  $L$ . Para dar cobertura ao fato de que um título tem muitos níveis possíveis, poderíamos examinar essas medidas para diferentes valores de  $L$ .

Médias não podem capturar todos os aspectos relevantes de uma distribuição e podem até ser enganosas, exceto por algumas distribuições especiais. Por exemplo, já ilustramos como uma taxa de incidência – por si própria – deixa de distinguir entre muitas pessoas acompanhadas por um período curto e poucas pessoas seguidas por um longo período. Similarmente, uma proporção de incidência de, por exemplo, 0,30 durante 10 anos não nos diz se os 30% de pessoas que tiveram o evento o experimentaram todas dentro de 2 anos, ou somente depois de 8 anos, ou tiveram os eventos espalhados por todo o período. Um tempo médio de incidência de 2 meses entre 101 pessoas não nos diz se essa média foi derivada de todas as 101 pessoas tendo o evento em 2 meses ou de 100 pessoas experimentando o evento em 1 mês, e 1 pessoa o fazendo em 102 meses, ou algo entre esses extremos.

Um modo de lidar com esses assuntos, sem fazer suposições sobre a distribuição da população, é focar em como a distribuição do desfecho muda com o tempo; introduzimos essa ideia anteriormente pela construção de proporções de sobrevida e de taxas de incidência dentro de intervalos de tempo (estratos). Pelo uso de intervalos de tempo bastante curtos, podemos ver claramente como os eventos se distribuem no tempo. Por exemplo, podemos observar se nossa taxa geral de eventos reflete muitas pessoas tendo o evento cedo e poucas o exibindo muito tardiamente, ou qualquer coisa intermediária; também é possível descrever a distribuição de eventos recorrentes, tais como ataques de asma.

Podemos generalizar nossa ideia de evento para representar qualquer transição de um estado para outro, não apenas de “sem doença” para “doente”, mas (digamos) de uma pressão arterial diastólica de 90 mmHg em um intervalo de tempo para 100 mmHg no próximo. Podemos examinar tais transições em uma escala tão fina quanto nossas mensurações permitam, por exemplo, ao estudar a evolução das contagens de linfócitos CD4 ao longo do tempo, podemos imaginar a taxa de transições da contagem  $x$  para a contagem  $y$  por unidade de tempo, para cada combinação sensível de  $x$  e  $y$ . Obviamente, esse ponto de vista generalizado implica um quadro muito mais complexo da população. Estudar essa complexidade é o assunto da *análise longitudinal de dados*. Fora alguns comentários breves no Capítulo 21, não discutiremos mais esse campo importante e vasto. Há muitos livros devotados a diferentes aspectos do tópico; por exemplo, Manton e Stallard (1988) fornecem uma abordagem de dinâmica populacional (demográfica), ao passo que Diggle e colaboradores (2002) cobrem abordagens baseadas em análise estatística de estudos de coortes, e van der Laan e Robins (2003) focam o modelo causal longitudinal.

## PREVALÊNCIA

Ao contrário das medidas de incidência, cujo foco é sobre novos eventos ou *alterações* de estados de saúde, a *prevalência* foca estados existentes. A prevalência de uma condição em um ponto no tempo pode ser definida como a proporção de uma população naquele estado, naquele tempo; assim, a prevalência de uma doença é a proporção da população com a doença no tempo especificado. Os termos *prevalência ponto*, *proporção de prevalência* e *taxa de prevalência* são usados, algumas vezes, como sinônimos. A prevalência generaliza para estados de saúde com múltiplos níveis. Por exemplo, ao considerar a saúde cardiovascular, poderíamos examinar a prevalência de níveis diferentes da pressão arterial diastólica e da sistólica em repouso; isto é, poderíamos estudar a distribuição completa da pressão arterial, e não apenas a prevalência de estar acima de certos pontos de corte clínicos.

O *grupo de prevalência* é o subgrupo da população com a dada condição. Uma pessoa que morre com a condição, ou por ela (p. ex., pela doença sob consideração), é removida do grupo de prevalência; consequentemente, a morte diminui a prevalência. As pessoas podem sair do grupo de prevalência ao se recuperarem da condição ou ao emigrarem da população. Doenças com taxas de incidência elevadas podem ter baixa prevalência se forem rapidamente fatais ou se curarem com brevidade. No entanto, doenças com taxas de incidência muito baixas podem ter uma prevalência substancial se não forem fatais, porém incuráveis.

## Uso da prevalência na pesquisa etiológica

Raramente a prevalência é de interesse direto em aplicações etiológicas da pesquisa epidemiológica. Posto que a prevalência reflete tanto a taxa de incidência quanto a duração da doença, os estudos de prevalência, ou estudos baseados em casos prevalentes, fornecem associações que refletem tanto os determinantes de sobrevida com a doença quanto as causas desta. O estudo da prevalência pode ser ilusório na situação paradoxal na qual a maior sobrevida de uma doença e, portanto, uma prevalência maior, resulta da ação de agentes preventivos que mitigam a doença uma vez que ela ocorra. Em tal situação, o agente preventivo pode estar associado positivamente à prevalência da doença e ser mal interpretado como um agente causal.

Entretanto, para pelo menos uma classe de doenças, a saber, as anomalias congênitas, a prevalência é a medida geralmente empregada. A proporção de bebês nascidos com alguma malformação é uma proporção de prevalência, não uma taxa de incidência. A incidência de malformações refere-se à sua ocorrência entre as populações suscetíveis de embriões. Muitas malformações levam à morte embrionária ou fetal precoce, a qual é classificada, se reconhecida, como um aborto, e não um nascimento, com ou sem malformação. Assim, os bebês malformados no nascimento representam apenas aqueles fetos que sobreviveram o bastante com suas malformações para serem registrados como nascimentos. A frequência de tais bebês entre todos os nascimentos é na verdade uma medida de prevalência, o ponto de referência no tempo sendo o momento do nascimento. A medida classifica a população de recém-nascidos quanto a seu *status* de doença, malformados ou não, no momento do nascimento. Esse exemplo ilustra que o tempo de referência para a prevalência não precisa ser um ponto no tempo do calendário, pode ser um ponto em outra escala de tempo, tal como a idade ou o tempo desde o tratamento.

Para estudar causas, seria mais útil medir a incidência do que a prevalência de malformações congênitas. Infelizmente, raramente é possível mensurar a taxa de incidência de malformações, porque a população em risco – embriões jovens – é difícil de apurar, e aprender sobre a ocorrência e o tempo das malformações entre os embriões é igualmente problemático. Consequentemente, nessa área de pesquisa, os casos incidentes geralmente não são estudados, e a maioria dos pesquisadores se acomoda no estudo da prevalência ao nascer, que teoricamente é menos desejável, mas muito mais prático.

A prevalência algumas vezes é usada para mensurar a ocorrência de doenças degenerativas sem um momento de início claro. Também é utilizada em estudos de soroprevalência da incidência de infecções, especialmente quando a infecção tem uma longa fase assintomática (silenciosa), que só pode ser detectada por testes sorológicos. A infecção pelo vírus da imunodeficiência humana (HIV) é um exemplo excelente (Brookmeyer e Gail, 1994). Aqui, a incidência de infecção pode ser calculada retroativamente pela incidência do início dos sintomas (síndrome de imunodeficiência adquirida, ou Aids), e sua prevalência pelo uso de suposições e de dados sobre a duração da fase assintomática. É claro que nas aplicações epidemiológicas fora da pesquisa etiológica, tais como o planejamento de recursos e as instalações para a saúde, a prevalência pode ser uma medida mais relevante do que a incidência.

## Prevalência, incidência e duração média

Frequentemente, o estudo da prevalência em lugar da incidência é racionalizado com base na relação simples entre as duas medidas, que é pertinente sob certas condições muito especiais. Examinaremos



essas condições cuidadosamente, com o objetivo de explicar por que elas raramente fornecem uma base segura para estudar a prevalência como um substituto da incidência.

Lembre-se de que uma população estacionária tem um número igual de pessoas entrando e saindo durante uma unidade de tempo. Suponha que tanto a população em risco quanto o grupo de prevalência sejam estacionários, e que todos estão expostos ao risco ou têm a doença. Então, o número de pessoas entrando no grupo de prevalência em qualquer período de tempo será contrabalançado com o grupo que sai:

$$\text{Influxo (para o grupo de prevalência)} = \text{efluxo (do grupo de prevalência)}$$

As pessoas podem entrar no grupo de prevalência a partir da população sadia e por imigração de outra população, enquanto estiverem doentes. Suponha que não haja imigração nem emigração, de modo que ninguém ingressa ou sai do grupo de prevalência, salvo por início da doença, morte ou cura. Se o tamanho da população é  $N$  e o do grupo de prevalência é  $P$ , então o tamanho da população exposta ao risco que alimenta o grupo de prevalência será  $N - P$ . Além disso, durante qualquer intervalo de tempo de duração  $\Delta t$ , o número de pessoas que ingressam no grupo de prevalência será

$$I(N - P)\Delta t$$

onde  $I$  é a taxa de incidência, e o efluxo do grupo de prevalência será

$$I'P\Delta t$$

onde  $I'$  representa a taxa de incidência de saída do grupo de prevalência, isto é, o número dos que saem dividido pela experiência de pessoa-tempo daqueles no grupo de prevalência. Portanto, na ausência de migração, o inverso de  $I'$  igualará à duração média da doença,  $\bar{D}$ , que é o tempo médio até a morte ou a cura. Então

$$\text{Influxo} = I(N - P)\Delta t = \text{efluxo} = (1/\bar{D})P\Delta t$$

o que gera

$$\frac{P}{N - P} = I \cdot \bar{D}$$

$P/(N - P)$  é a razão de pessoas doentes para pessoas sem doença na população ou, equivalentemente, a razão da proporção de prevalência para a proporção de sadios. (Chamamos aqueles que estão sem a doença de sadios, pelo fato de querermos dizer que eles não têm uma doença específica, o que não implica uma ausência de todas as doenças.)

A razão  $P/(N - P)$  é chamada de *chances de prevalência*; é a razão entre a proporção de uma população que tem a doença a proporção dos que não têm a doença. Conforme exibido, a chance de prevalência é igual à taxa de incidência vezes a duração média da doença. Se a prevalência é pequena, digamos  $<0,1$ , então

$$\text{Proporção de prevalência} \approx I \cdot \bar{D}$$

porque a proporção de prevalência se aproximará da chance de prevalência, para valores pequenos da prevalência. Sob a suposição de grupo de prevalência estacionário, sem migração para dentro ou para fora (Freeman e Hutchinson, 1980),

$$\text{Proporção de prevalência} = \frac{I \cdot \bar{D}}{1 + I \cdot \bar{D}}$$

que pode ser obtida da expressão supracitada para a chance de prevalência,  $P/(N - P)$ .

Como a proporção de incidência, a proporção de prevalência é não dimensional, com um intervalo de 0 a 1. As equações citadas estão de acordo com esses requisitos, porque, em cada



uma delas, a taxa de incidência, com uma dimensionalidade do inverso do tempo, é multiplicada pela duração média da doença, que tem a dimensionalidade de tempo, gerando um produto sem dimensão. Além disso, o produto  $I \times \overline{D}$  está no intervalo de 0 a infinito, que corresponde ao intervalo da chance de prevalência, ao passo que a expressão

$$\frac{I \cdot \overline{D}}{1 + I \cdot \overline{D}}$$

está sempre no intervalo de 0 a 1, correspondente ao intervalo de uma proporção.

Infelizmente, as fórmulas citadas têm utilidade prática limitada, por causa da suposição de não migração e porque não se aplicam à prevalência específica por idade (Miettinen, 1976a). Se considerarmos o grupo de prevalência de, por exemplo, diabéticos que têm 60 a 64 anos de idade, poderemos ver que esse grupo experimenta imigração considerável de diabéticos mais jovens envelhecendo para dentro do grupo e emigração considerável de membros do grupo ficando mais velhos e saindo. De modo mais geral, devido à relação muito forte da idade com a maioria das doenças, quase sempre precisamos considerar subpopulações específicas por idade ao estudar padrões de ocorrência. Sob tais condições, a análise apropriada requer fórmulas mais elaboradas, que forneçam a prevalência como uma função de incidência específica por idade, por duração e por outros parâmetros populacionais (Preston, 1987; Manton e Stallard, 1988; Keiding, 1991; Alho, 1992).

## IDADE MÉDIA NO EVENTO

Entende-se, geralmente, que a expectativa de vida refere-se à idade média no óbito em uma coorte ou em uma população fechada definida por uma coorte, tal como todas as pessoas nascidas em um ano em particular. Como tal, a expectativa de vida só pode se desdobrar durante um intervalo de tempo desde a instalação da coorte até a morte de seu último membro sobrevivente (o que pode representar mais de um século). Em contraste, a idade média no óbito refere-se, em geral, à idade média de pessoas falecendo em um intervalo de tempo específico, estreito. Por exemplo, a idade média no óbito entre as pessoas morando no Vietnã em 2010 representa experiências de pessoas nascidas em, grosseiramente, de 1900 a todos os anos seguintes até 2010. Ela é altamente influenciada pelo tamanho em anos da população que contribui para o cálculo e por mudanças na expectativa de vida, ao longo das coortes de nascimento contribuindo para ela. Assim como a prevalência, ela é um atributo transversal (ponto isolado no tempo) de uma população, e a população pode ser aberta.

Se a população estivesse em estado de equilíbrio durante um século e continuasse assim por outro século ou mais, a expectativa de vida daqueles nascidos hoje poderia se assemelhar à idade média no óbito atualmente. A realidade em muitos locais, contudo, é que o *número* de nascimentos por ano aumentou substancialmente durante o século XX, e, assim, a proporção em faixas etárias mais jovens aumentou. Essa alteração por si só puxou a idade média no óbito para baixo, pois ela foi mais influenciada pelo número aumentado de mortes em faixas etárias mais jovens. Mudanças na expectativa real de vida exerceriam outros efeitos, possivelmente na direção oposta. A consequência é que a idade média no óbito pode diferir consideravelmente da expectativa de vida em qualquer ano.

Essas mesmas forças e outras forças semelhantes também afetam a idade média de ocorrência de outros eventos, tais como a ocorrência de uma doença em particular. Comparações entre tais médias misturam as forças que afetam a taxa de ocorrência de doença com alterações demográficas que influenciam a estrutura etária da população. Tais comparações, portanto, são inerentemente tendenciosas como estimativas de associações causais.

## PADRONIZAÇÃO

A noção de padronização é essencial para muitas técnicas analíticas em epidemiologia, inclusive para técnicas para o controle de confundimento (Caps. 4 e 15) e para sumarização de ocorrência e efeitos

(Caps. 20 e 21). Assim, taxas e proporções padronizadas surgirão em vários pontos deste livro, nos quais serão descritas com mais detalhes. A padronização de medidas de ocorrência nada mais é do que a média ponderada dessas medidas. O uso apropriado da ideia, contudo, requer atenção cuidadosa para tópicos de ordenação causal, aos quais retornaremos em discussões posteriores. Também observamos que o termo *padronização* tem uma definição inteiramente diferente em outros ramos da estatística, nos quais ele significa a reexpressão de quantidades em unidades de desvio-padrão, uma prática que pode levar a distorções graves de estimativas de efeito (Greenland et al., 1986, 1991; ver Cap. 33).

Para ilustrar a ideia básica de padronização, suponha que nos seja dada uma distribuição de pessoas-tempo específica para uma série de variáveis, por exemplo, as pessoas-ano em risco observadas dentro das categorias etárias 50 a 59 anos, 60 a 69 anos e 70 a 74 anos, para homens e mulheres em Quebec, no ano 2000. Deixemos que  $T_1, T_2, \dots, T_6$  seja a distribuição de pessoas-ano nas seis categorias de idade/sexo desse exemplo. Suponha, também, que nos sejam fornecidas as seis taxas de incidência específicas por idade e sexo  $I_1, I_2, \dots, I_6$  correspondentes aos estratos específicos idade/sexo. A partir dessa distribuição e desse conjunto de taxas, podemos computar uma média ponderada das taxas com pesos da distribuição,

$$I_s = \frac{I_1 T_1 + \dots + I_6 T_6}{T_1 + \dots + T_6} = \frac{\sum_{k=1}^6 I_k T_k}{\sum_{k=1}^6 T_k}$$

O numerador de  $I_s$  pode ser reconhecido como o número de casos que seriam vistos em uma população que tivesse a distribuição pessoas-tempo  $T_1, T_2, \dots, T_6$  e essas taxas específicas por estrato. O denominador de  $I_s$  é o pessoas-tempo total nessa população. Portanto,  $I_s$  é a taxa que seria vista em uma população com a distribuição  $T_1, T_2, \dots, T_6$  e as taxas específicas  $I_1, I_2, \dots, I_6$ .

$I_s$  é tradicionalmente chamada de uma *taxa padronizada*, e  $T_1, T_2, \dots, T_6$ , *distribuição padrão* na qual se baseia  $I_s$ .  $I_s$  representa a taxa geral que seria observada em uma população cujas pessoas-tempo seguissem a distribuição padrão, e cujas taxas específicas fossem  $I_1, I_2, \dots, I_6$ .

O processo de padronização também pode ser conduzido com proporções de incidência ou de prevalência. Suponha, por exemplo, que tenhamos uma distribuição de pessoas  $N_1, N_2, \dots, N_6$ , em vez de pessoas-tempo em risco, e um conjunto correspondente de proporções de incidência específicas por estrato  $R_1, R_2, \dots, R_6$ . Dessa distribuição e desse conjunto de proporções, podemos calcular o risco médio ponderado

$$R_s = \frac{R_1 N_1 + \dots + R_6 N_6}{N_1 + \dots + N_6} = \frac{\sum_{k=1}^6 R_k N_k}{\sum_{k=1}^6 N_k}$$

que é o *risco padronizado* baseado na distribuição  $N_1, N_2, \dots, N_6$ . A padronização também pode ser aplicada usando-se outras medidas (tais como tempos médios de incidência ou pressões sanguíneas médias), em vez das taxas ou das proporções.

Visto que as taxas que se aplicam a uma população podem afetar a distribuição pessoas-tempo, a taxa padronizada não é necessariamente a taxa que descreve o que aconteceria a uma população com a distribuição padrão  $T_1, \dots, T_6$  se as taxas específicas  $I_1, I_2, \dots, I_6$  fossem aplicadas a essa população. Uma discrepância análoga pode surgir para chances padronizadas. Esse problema pode distorcer inferências baseadas na comparação de taxas e chances padronizadas e será mais discutido no próximo capítulo. O problema não aparece ao se considerar riscos padronizados, porque a distribuição inicial  $N_1, \dots, N_6$  não pode ser afetada pelos riscos subsequentes  $R_1, \dots, R_6$ .

# Medidas de efeito e medidas de associação

Sander Greenland, Kenneth J. Rothman e Timothy L. Lash

## Medidas de efeito 67

- Medidas de diferença 68
- Medidas de razão 69
- Medidas relativas de excesso 70
- Dependência do estado de efeito nulo sobre a medida de efeito 71
- A natureza teórica das medidas de efeito 71
- Definindo a exposição em medidas de efeito 71
- Efeitos mediados por riscos competitivos 72

## Associação e confundimento 73

- Medidas de associação 73
- Confundimento 74
- Confundidores 75
- Um modelo simples que distingue causalidade de associação 76

## Relações entre medidas de efeito 78

- Relações entre riscos relativos 78
- Modificação da medida de efeito (heterogeneidade) 79

Relação entre medidas estrato-específicas e medidas gerais 79

## Frações atribuíveis 80

- Frações de excesso 80
- Frações etiológicas 81
- Probabilidade de causalidade e suscetibilidade à exposição 82
- Uma nota sobre terminologia 83

## Generalizando definições de efeito 84

## Frações atribuíveis à população e frações de impacto 85

## Medidas padronizadas de associação e efeito 85

- Razões de morbidade padronizadas (RMP) 87

## Razões de prevalências 88

## Outras medidas 88

**S**uponha que desejamos estimar o efeito de uma exposição sobre a ocorrência de uma doença. Por motivos explicados adiante, não podemos observar, ou mesmo estimar, esse efeito diretamente. Em vez disso, observamos uma associação entre a exposição e a doença entre sujeitos do estudo, que estima uma associação populacional. A associação observada será um substituto pobre para o efeito desejado se ela for uma estimativa pobre da associação populacional, ou se a própria associação populacional não estiver próxima do efeito de interesse. Os Capítulos 9 até 12 abordam problemas específicos que surgem na conexão entre associações e efeitos observados. O presente capítulo define efeitos e associações em populações e os conceitos básicos necessários para conectá-los.

## MEDIDAS DE EFEITO

Os epidemiologistas usam o termo *efeito* em dois sentidos. Em um sentido, qualquer caso de uma dada doença pode ser o efeito de uma dada causa. *Efeito* é usado dessa maneira para significar o ponto terminal de um mecanismo causal, identificando o tipo de desfecho que uma causa produz. Por exemplo, podemos dizer que a infecção pelo vírus da imunodeficiência humana (HIV) é um efeito do compartilhamento de agulhas para uso de drogas. Esse uso do termo *efeito* meramente identifica

a infecção por HIV como uma consequência da atividade de compartilhar agulhas. Outros efeitos da exposição, tal como infecção por hepatite B, também são possíveis.

Em um sentido mais epidemiológico, um *efeito* de um fator é uma mudança de característica de uma população, que é causada em consequência de o fator estar em um nível *versus* outro. As características populacionais de foco tradicional na epidemiologia são as medidas de frequência de doenças, como descrito no Capítulo 3. Se a frequência da doença é medida em termos de taxa ou proporção de incidência, então o efeito é a mudança de tal taxa ou proporção provocada por um fator específico. Podemos dizer que para os usuários de drogas, o efeito de compartilhar agulhas, em comparação com não o fazer, é aumentar o risco médio de infecção por HIV de 0,001 para 0,01 em 1 ano. Embora seja costumeiro usar o artigo definido ao se referir a esse segundo tipo de efeito (o efeito de compartilhar agulhas), isso não significa que este seja o único efeito de compartilhar agulhas. Um aumento do risco de hepatite ou de outras doenças permanece possível, e o aumento do risco de infecção por HIV pode diferir entre populações e no tempo.

Em epidemiologia, é costume referir-se a características causais potencialmente como *exposições*. Assim, *exposição* pode designar um comportamento (p. ex., compartilhamento de agulhas), um tratamento ou outra intervenção (p. ex., um programa educativo sobre os perigos de compartilhar agulhas), um traço (p. ex., um genótipo), uma exposição no sentido comum (p. ex., uma transfusão de sangue contaminado) ou mesmo uma doença (p. ex., diabete como uma causa de morte).

Os efeitos populacionais são expressos mais comumente como efeitos por meio de taxas ou de proporções de incidência, mas outras medidas baseadas em tempos de incidência, ou prevalências, também podem ser usadas. Análises epidemiológicas que focam o tempo de sobrevivência até a morte ou a recorrência de doença são exemplos de análises que medem efeitos sobre tempos de incidência. *Medidas de efeito absoluto* são diferenças nas medidas de ocorrência, e *medidas de efeito relativo* são razões de medidas de ocorrência. Outras medidas dividem efeitos absolutos pela medida de ocorrência.

Para simplificar, nossas descrições básicas serão de efeitos em coortes, que são grupos de indivíduos. Como mencionado no Capítulo 3, cada coorte define uma população fechada que começa no momento em que o grupo é definido. Entre as medidas que consideraremos, somente aquelas envolvendo taxas de incidência podem ser generalizadas diretamente para populações abertas.

## Medidas de diferença

Considere uma coorte acompanhada durante um tempo específico ou por intervalo de idade – por exemplo, de 2000 a 2005, ou das idades de 50 a 69 anos. Se podemos imaginar a experiência dessa coorte durante o mesmo intervalo de tempo sob duas condições diferentes – digamos, exposta e não exposta –, então podemos indagar qual a taxa de incidência de qualquer desfecho sob as duas condições. Assim, podemos considerar uma coorte de fumantes e uma exposição que consistisse em enviar pelo correio, a cada membro da coorte, um folheto sobre programas atuais de cessação do fumo no município de residência do membro da coorte. Poderíamos, então, perguntar qual seria a taxa de incidência de câncer do pulmão nessa coorte se fizéssemos essa intervenção, e qual seria essa taxa de incidência se não a realizássemos. Essas histórias de intervenção representam alternativas mutuamente exclusivas para a coorte. Assim, as duas taxas de incidência representam *desfechos potenciais* alternativos para a coorte.

Podemos chamar a diferença entre as duas taxas de efeito absoluto da exposição utilizada (envio de brochura por correio) de taxa de incidência, ou *diferença de taxas causal*. Para ser breve, poderíamos nos referir à diferença de taxas causal como o excesso da taxa devido ao programa (o qual seria negativo se o programa prevenisse alguns casos de câncer de pulmão).

De modo paralelo, poderíamos perguntar qual seria a proporção de incidência se efetuássemos essa intervenção, e qual seria essa proporção de incidência se não o fizéssemos. Chamamos a diferença entre as duas proporções de *efeito absoluto* de nossa intervenção sobre a proporção de incidência, ou *diferença de riscos causal*, ou, para abreviar, *excesso de risco*. Também de maneira paralela, a diferença da média de anos de vida livres de câncer de pulmão vividos durante o intervalo, sob as condições de intervenção e sem intervenção, é um outro efeito absoluto do tratamento.

Para ilustrar de forma simbólica as medidas citadas, suponha que temos uma coorte de tamanho  $N$  definida no começo de um intervalo de tempo fixo, e que qualquer pessoa viva sem a doença está exposta ao risco desta. Além disso, suponha que se todo membro da coorte for exposto durante o intervalo,  $A_1$  casos ocorrerão e o tempo total em risco será  $T_1$ , mas se nenhum membro da mesma coorte for exposto durante o intervalo,  $A_0$  casos ocorrerão e o tempo total em risco será  $T_0$ . Então, a diferença de taxas causal será

$$\frac{A_1}{T_1} - \frac{A_0}{T_0}$$

a diferença de riscos causal será

$$\frac{A_1}{N} - \frac{A_0}{N}$$

e a diferença causal no tempo médio livre de doença será

$$\frac{T_1}{N} - \frac{T_0}{N} = \frac{T_1 - T_0}{N}$$

Quando o desfecho é a morte, o negativo da diferença de tempos médios,  $T_0/N - T_1/N$ , frequentemente é chamado de *anos de vida perdidos* em consequência da exposição. Cada uma dessas medidas compara a ocorrência de doença tomando diferenças, por isso são chamadas de medidas de *diferença*, ou medidas absolutas. Elas são expressas em unidades de suas medidas componentes: casos por unidade pessoa-tempo, para a diferença de taxas, casos por pessoas que iniciaram o seguimento, para a diferença de riscos, e unidades de tempo, para a diferença de tempos médios.

## Medidas de razão

Mais comumente, as medidas de efeito são calculadas tomando-se razões. Exemplos de tais medidas de razão (ou relativas) são a *razão de taxas causais*,

$$\frac{A_1/T_1}{A_0/T_0} = \frac{I_1}{I_0}$$

onde  $I_j = A_j/T_j$  é a taxa de incidência sob a condição  $j$  ( $1 =$  expostos,  $0 =$  não expostos); a *razão de riscos causal*,

$$\frac{A_1/N}{A_0/N} = \frac{A_1}{A_0} = \frac{R_1}{R_0}$$

onde  $R_j = A_j/N$  é a proporção de incidência (risco médio) sob a condição  $j$ ; e a *razão causal de tempo livre de doença*,

$$\frac{T_1/N}{T_0/N} = \frac{T_1}{T_0}$$

Em contraste com as medidas de diferença, as medidas de razão não têm dimensão, porque as unidades se cancelam na divisão.

A razão de taxas e a razão de riscos são denominadas, frequentemente, *riscos relativos*. Algumas vezes, esse termo também é aplicado a razões de chances, embora desencorajemos tal uso. “Risco relativo” pode ser o termo mais comum em epidemiologia. Seu uso é tão amplo que se deve, frequentemente, examinar os detalhes do delineamento e da análise, para discernir qual medida de razão está sendo estimada, ou discutida.

As três medidas de razão estão relacionadas pela fórmula

$$\frac{R_1}{R_0} = \frac{R_1 N}{R_0 N} = \frac{A_1}{A_0} = \frac{I_1 T_1}{I_0 T_0}$$

que deriva do fato de que o número de casos é igual à taxa da doença vezes o tempo em risco. Uma quarta medida relativa pode ser construída a partir da chance de incidência. Se escrevermos  $S_1 = 1 - R_1$  e  $S_0 = 1 - R_0$ , a *razão de chances causal* é então

$$\frac{R_1/S_1}{R_0/S_0} = \frac{A_1/(N - A_1)}{A_0/(N - A_0)}$$

## Medidas relativas de excesso

Quando um risco relativo é maior do que 1, refletindo um efeito médio que é causal, é expresso algumas vezes como *excesso de risco relativo*, que pode se referir à razão de excesso de taxa causal,

$$RT - 1 = \frac{I_1}{I_0} - 1 = \frac{I_1 - I_0}{I_0}$$

onde  $RT = I_1/I_0$  é a razão de taxas causal. Similarmente, a razão de excesso de risco causal é

$$RR - 1 = \frac{R_1}{R_0} - 1 = \frac{R_1 - R_0}{R_0}$$

onde  $RR = R_1/R_0$  é a razão de riscos causal. Essas fórmulas mostram como excessos de riscos relativos são iguais à taxa, ou à diferença de riscos, dividida pela (relativa à) taxa de risco sem exposição ( $I_0$  ou  $R_0$ ), e, assim, são chamadas algumas vezes de diferença relativa, ou medidas relativas de excesso.

Mais frequentemente, o excesso da taxa é expresso em relação à  $I_1$ , como

$$\frac{I_1 - I_0}{I_1} = \frac{I_1/I_0 - 1}{I_1/I_0} = \frac{RT - 1}{RT} = 1 - \frac{1}{RT}$$

onde  $RT = I_1/I_0$  é a razão de taxas causal. De modo semelhante, o excesso de risco frequentemente é expresso em relação a  $R_1$ , como

$$\frac{R_1 - R_0}{R_1} = \frac{R_1/R_0 - 1}{R_1/R_0} = \frac{RR - 1}{RR} = 1 - \frac{1}{RR}$$

onde  $RR = R_1/R_0$  é a razão de riscos causal. Nessas duas medidas o excesso de risco, ou risco atribuível à exposição, é expresso como uma fração da taxa total, ou risco sob exposição; daí,  $(RT - 1)/RT$  pode ser chamada a *fração de taxa*, e  $(RR - 1)/RR$ , a *fração de risco*. Frequentemente, são designadas como *frações atribuíveis*.

Numerosas outras medidas também são designadas como frações atribuíveis. Especialmente, as frações de taxa e de risco que acabamos de definir são confundidas, frequentemente, com uma quantidade distinta chamada de *fração etiológica*, que não pode ser expressa como uma função simples das taxas ou riscos. Discutiremos esses problemas em detalhe mais tarde, quando uma outra medida relativa de excesso aparecerá.

Pretendia-se que as medidas relativas de excesso fossem usadas com exposições que tivessem um efeito causal. Elas se tornam negativas e, portanto, difíceis de interpretar com um efeito preventivo. Uma modificação oportuna para lidar com exposições preventivas é intercambiar as quantidades de expostos e não expostos nas medidas. As medidas que surgem do intercâmbio de  $I_1$  com  $I_0$  e de  $R_1$  com  $R_0$  nas frações atribuíveis têm sido chamadas de *frações preveníveis*, e são facilmente interpretadas. Por exemplo,  $(R_0 - R_1)/R_0 = 1 - R_1/R_0 = 1 - RR$  é a fração do risco que não está sob exposição,

que poderia ser prevenido pela exposição. Em estudos sobre vacinas, essa medida também é conhecida como a *eficácia da vacina*.

## Dependência do estado de efeito nulo sobre a medida de efeito

Se as medidas de ocorrência comparadas não variarem com a exposição, a medida de efeito será igual a 0, se ela for uma medida de diferença, ou de diferença relativa, e igual a 1 se for uma medida de razão. Nesse caso, dizemos que o efeito é *nulo* e que a exposição não tem efeito sobre a medida de ocorrência. Esse estado de efeito nulo não depende do modo pelo qual a medida de ocorrência é comparada (diferença, razão, etc.), mas pode depender da medida de ocorrência. Por exemplo, o risco médio de morte aos 150 anos é sempre 100%, e, assim, a diferença de riscos causal aos 150 anos é sempre 0, para qualquer exposição conhecida; nada ainda foi descoberto que previna a morte à idade de 150. No entanto, muitas exposições (tais como o uso de fumo) mudarão o risco de morte à idade de 60 anos relativamente à não exposição e, assim, terão uma diferença de riscos causal não zero aos 60 anos; tais exposições também mudarão a taxa de mortalidade e o tempo médio de morte.

## A natureza teórica das medidas de efeito

As definições de medidas de efeito supracitadas são chamadas, às vezes, de *contrafactuais*, ou definições de desfecho potencial. Tais definições podem ser rastreadas nos escritos de Hume, no século XVIII. Embora elas tenham sido pouco explicadas até o último terço do século XX, estavam sendo usadas por cientistas (inclusive epidemiologistas e estatísticos) muito tempo antes; ver Lewis (1973), Rubin (1990a), Greenland (2000a, 2004a) e Greenland e Morgenstern (2001), para referências anteriores.

Essas medidas de efeito são chamadas de medidas contrafactuais, porque pelo menos uma das duas categorias nas definições das medidas de efeito deve ser contrária ao fato. A coorte pode ser exposta ou tratada (p. ex., foi enviado um folheto a todos os membros), ou não tratada (não foi enviado folheto a ninguém). Se a coorte for tratada, então a categoria não tratada será contrafactual, e se não for tratada, a categoria tratada será contrafactual. Ambas as categorias podem ser contrafactuais, como ocorreria se o folheto fosse enviado por correio a somente uma parte da coorte. Os desfechos das condições (p. ex.,  $I_1$  e  $I_0$ , ou  $R_1$  e  $R_0$ , ou  $T_1/N$  e  $T_0/N$ ) permanecem como potenciais até que um tratamento seja aplicado à coorte (Rubin, 1974, 1990a; Greenland, 2000a, 2004a).

Um aspecto importante das medidas de efeito contrafactuais é que envolvem duas categorias distintas: uma categoria de índice, que geralmente envolve alguma exposição ou tratamento, e uma categoria de referência – tal como nenhum tratamento – contra a qual essa exposição ou tratamento será avaliado. Perguntar qual o efeito da exposição não tem sentido nem referência a alguma outra categoria. No exemplo precedente sobre fumo, o efeito de um folheto enviado pelo correio só é definido em relação a nenhum folheto. Poderíamos ter perguntado, em vez disso, sobre o efeito de um folheto em relação a quatro folhetos; essa comparação é muito diferente de uma *versus* nenhum folheto.

Outro aspecto importante das medidas de efeito é que elas nunca são observadas separadamente de alguma medida de ocorrência componente. Se, no mesmo exemplo, enviamos uma postagem para toda a coorte, a diferença de taxas comparando o desfecho com nenhuma postagem,  $I_1 - I_0$ , não é observada diretamente; observamos apenas  $I_1$ , que é a soma daquela medida de efeito e da taxa (contrafactual) de nenhuma postagem,  $I_0$ :  $I_1 = (I_1 - I_0) + I_0$ . Portanto, o pesquisador enfrenta o problema de separar a medida de efeito  $I_1 - I_0$  da taxa de não exposição  $I_0$ , tendo observado apenas sua soma,  $I_1$ .

## Definindo a exposição em medidas de efeito

Posto que definimos efeitos com referência a uma única coorte sob duas condições distintas, devemos ser capazes de descrever significativamente cada condição para uma coorte (Rubin, 1990a; Greenland, 2002a, 2005a; Hernán, 2005; Maldonado e Greenland, 2002). Considere, por exemplo, o efeito do sexo



(masculino *versus* feminino) sobre doença do coração. Para essas palavras terem conteúdo, devemos ser capazes de imaginar uma coorte de homens, sua incidência de cardiopatia e qual teria sido tal incidência se esses indivíduos, em vez de homens, fossem mulheres. O absurdo dessa proposta revela o significado vago do efeito do sexo. Para atingir um nível razoável de precisão científica, o efeito do sexo poderia ser substituído por conceitos mecanicistas mais precisos, tais como efeitos hormonais, efeitos da discriminação e efeitos de outros fatores associados ao sexo, que expliquem a associação de sexo com incidência. Com tais conceitos, podemos imaginar o que significaria para os homens ter sua exposição trocada: tratamentos hormonais, operações de troca de sexo, e assim por diante.

As considerações precedentes enfatizam a necessidade de se definir a categoria-índice e a categoria de referência em detalhes substanciais, para facilitar a interpretação dos resultados. Por exemplo, em um estudo sobre os efeitos do fumo, uma definição da categoria-índice para um fumante atual poderia dar conta da frequência do fumo (cigarros por dia), da sua duração (em anos) e da idade em que o tabagismo começou. Similarmente, a definição da ausência de exposição para a categoria de referência – com relação à dose, à duração e ao período de indução – deve receber tanta atenção quanto a definição da presença de exposição. Embora seja comum caracterizar todas as pessoas que não satisfazem a definição de fumante atual como não expostas, tal definição poderia diluir o efeito, ao incluir ex-fumantes e fumantes ocasionais do grupo de referência.

Se as definições da categoria-índice e da categoria de referência forem suficientemente precisas, depende, em parte, do desfecho sob estudo. Por exemplo, um estudo do efeito do fumo atual sobre a ocorrência de câncer do pulmão deve estabelecer condições mínimas para frequência, duração e período de indução do hábito de fumar cigarros para definir o grupo exposto e estabelecer condições máximas (talvez zero) para essas mesmas características, para definir o grupo de não expostos. Ex-fumantes não preencheriam os critérios para a categoria-índice, nem para a categoria de referência. Em contraste, um estudo do efeito do fumo atual sobre a ocorrência de lesões em um incêndio residencial, poderia definir qualquer hábito de fumar atual para definir o grupo exposto e qualquer categoria de não fumante atual para definição do grupo não exposto, mesmo que este último incluía ex-fumantes (presumindo-se que os ex-fumantes e os que nunca fumaram tenham o mesmo risco de incêndio domiciliar).

## Efeitos mediados por riscos competitivos

Conforme discutido no Capítulo 3, a presença de riscos competitivos leva a várias complicações quando se está interpretando medidas de incidência. Tais complexidades são transpostas para a interpretação de medidas de efeito. Em particular, a interpretação de comparações simples de proporções de incidência deve ser ponderada pelo fato de que refletem efeitos de exposição sobre riscos competitivos, assim como ocorrências individuais da doença em estudo. Uma consequência desses efeitos é que a exposição pode afetar o tempo em risco para a doença em estudo. Para tomar um exemplo extremo, suponha que a exposição fosse um tratamento antifumo, e a “doença” fosse ser atropelado por um motorista bêbado. Se o tratamento antifumo fosse mesmo moderadamente eficaz na redução do uso de tabaco, provavelmente levaria a uma redução das mortes, assim prolongando o tempo de vida, o que aumentaria a oportunidade de ser atropelado por um motorista bêbado. O resultado seria mais atropelamentos por motoristas embriagados para aqueles expostos, e, portanto, um risco maior sob exposição.

Esse risco maior de ser atropelado por um motorista bêbado é um efeito genuíno do tratamento antifumo, embora indireto e não intencional. O mesmo tipo de efeito surge de qualquer exposição que altere o tempo em risco de outros desfechos. Assim, o hábito de fumar pode reduzir o risco médio de morte acidental simplesmente pela redução do tempo de exposição ao risco de acidente. De modo semelhante e bastante distante de algum efeito biológico direto, o tabagismo pode reduzir o risco médio de doença de Alzheimer, de doença de Parkinson e de outras moléstias dos idosos, simplesmente por diminuir a possibilidade de uma vida longa o bastante para adquirir essas doenças. Esse efeito indireto ocorre mesmo se olharmos para um intervalo de tempo estreito, tal como um risco em 2 anos, em vez de um risco para a vida toda: mesmo dentro de um intervalo de 2 anos, o fumo poderia causar alguns óbitos e assim reduzir o tempo populacional em risco, levando a menos casos daquelas doenças.

Embora os efeitos que acabamos de descrever sejam efeitos de exposição real, os pesquisadores, tipicamente, querem remover esses efeitos ou fazer ajustes para eles e focar efeitos mais diretos da exposição sobre a doença. Medidas de taxas levam em conta mudanças no tempo de risco, produzidas pela exposição de maneira simples, por mensurar o número de eventos da doença relativo ao tempo em risco. Na verdade, se não houver uma tendência na taxa da doença no decorrer do tempo, e o único efeito da exposição sobre a ocorrência da doença for alterar o tempo em risco, a razão e a diferença de taxa serão nulas (1 e 0, respectivamente). Se, contudo, houver tendências da doença no tempo, então, mesmo as medidas de taxas incorporarão alguns efeitos da exposição sobre o tempo em risco; quando isso acontece, comparações de taxa estratificadas por tempo (análise de sobrevida; ver Cap. 16) são necessárias para justificar tais efeitos.

As estimativas típicas de risco tentam “ajustar” riscos competitivos, usando métodos que estimam o risco contrafactual da doença em estudo se os riscos competitivos forem removidos. Como mencionado no Capítulo 3, uma objeção a esses métodos é que a condição contrafactual não é clara: como são removidos os riscos competitivos? A incidência da doença em estudo dependeria fortemente dessa resposta. Tais problemas são paralelos aos problemas de definição de exposição nas medidas de efeito: como trocar a exposição para não exposição, ou vice-versa? A maioria dos métodos faz as suposições implausíveis de que a exposição poderia ser completamente removida sem afetar a taxa de riscos competitivos e de que os riscos competitivos poderiam ser removidos sem afetar a taxa da doença em estudo. Essas suposições raramente, ou nunca, são justificáveis. Uma abordagem mais geral trata a doença em estudo e os riscos competitivos como partes de um desfecho *multivariado* ou *multidimensional*. Essa abordagem pode reduzir a dependência de suposições implausíveis e também responde ao argumento de que uma exposição não deve ser considerada isoladamente, em especial quando os efeitos da exposição e dos riscos competitivos implicam custos e benefícios muito diferentes (Greenland, 2002a, 2005a).

Devido às complexidades que derivam de se adotar uma abordagem mais geral, não faremos maiores elucubrações sobre riscos competitivos. No entanto, os leitores devem estar atentos para os problemas que podem surgir, quando a exposição pode ter efeitos fortes sobre outras doenças e não sobre a que se está estudando.

## ASSOCIAÇÃO E CONFUNDIMENTO

Visto que a população isolada em uma definição de efeito só pode ser observada sob uma das duas condições na definição (e, algumas vezes, sob nenhuma das duas), encaramos um problema especial na estimação do efeito: devemos predizer acuradamente a magnitude da ocorrência sob condições que não ocorreram ou que, de fato, não ocorrerão. Em outras palavras, devemos predizer certos desfechos sob o que são, ou se tornarão, condições contrafactuais. Por exemplo, podemos observar  $I_1 = 50$  mortes por 100.000 pessoas-ano em uma coorte-alvo de fumantes durante um seguimento de 10 anos e perguntar que redução de taxa teria sido conseguida se esses fumantes tivessem deixado de fumar no início do seguimento. Aqui, observamos  $I_1$  e precisamos de  $I_0$  (a taxa que teria ocorrido sob completa cessação do fumo) para completar  $I_1 - I_0$ .

Uma vez que  $I_0$  não é observada, precisamos predizer o que ela teria sido. Para fazer isso, precisaríamos recorrer a dados de pessoas não expostas, tais como os dados de uma coorte que não foi exposta. A partir de tais dados, construiríamos uma previsão de  $I_0$ . Nem esses dados, nem a previsão derivada deles, são parte da medida de efeito; são apenas ingredientes de nosso processo de estimação. Nós os utilizamos para construir uma medida de *associação*, a qual esperamos que equivalha à medida de efeito de interesse.

### Medidas de associação

Considere uma situação na qual contrastamos uma medida de ocorrência em *duas* populações diferentes. Por exemplo, poderíamos tomar a razão das taxas de incidência de câncer entre o sexo masculino e o feminino no Canadá. Essa razão das taxas de câncer comparando subpopulações de homens e de mulheres *não* é uma medida de efeito, porque suas duas taxas componentes referem-se a grupos de

pessoas diferentes. Nessa situação, dizemos que a razão de taxas é uma *medida de associação*; nesse exemplo, é uma medida da associação entre sexo e incidência de câncer.

Como um outro exemplo, poderíamos contrastar a taxa de incidência de cáries dentárias em crianças de uma comunidade no ano anterior ao início e no terceiro ano após a introdução de fluoretação do suprimento de água. Se tomarmos a diferença das taxas nesses períodos (anterior e posterior), essa diferença *não* é uma medida de efeito, porque suas duas taxas componentes referem-se a duas subpopulações diferentes, uma antes da fluoretação e outra depois. Pode haver uma superposição considerável, ou mesmo total, das crianças presentes nos dois períodos. No entanto, as experiências comparadas envolvem períodos de tempo diferentes, então dizemos que a diferença de taxas é uma medida de associação. Nesse exemplo, trata-se de uma medida de associação entre fluoretação e incidência de cáries dentárias.

Podemos resumir a distinção entre medidas de efeito e medidas de associação do seguinte modo: uma medida de efeito compara o que aconteceria a *uma* população sob dois cursos, ou condições, de vida possíveis, mas distintos, dos quais no máximo um pode ocorrer (p. ex., um banimento de toda a propaganda do fumo *versus* um banimento apenas da propaganda de televisão). É um conceito teórico (alguns diriam “metafísico”), pois é logicamente impossível observar a população sob ambas as condições e, portanto, verificar diretamente o tamanho do efeito (Maldonado e Greenland, 2002). Em contraste, uma medida de associação compara o que acontece em duas populações distintas, embora elas possam corresponder a uma só população em períodos de tempo diferentes. Sujeitos a limitações físicas e sociais, podemos observar ambas as populações e, assim, verificar diretamente uma associação.

## Confundimento

Dada a natureza observável das medidas de associação, é tentador substituí-las por medidas de efeito (talvez depois de fazer alguns ajustes). É até mesmo mais natural dar explicações causais para associações observadas em termos de diferenças óbvias entre as populações comparadas. No exemplo precedente de cáries dentárias, é tentador atribuir um declínio da incidência após a fluoretação ao próprio ato da adição de flúor. Analisemos em detalhe como tal inferência se traduz em medidas de efeito e de associação.

O efeito que queremos mensurar é o que a fluoretação tem sobre a taxa. Para mensurar esse efeito, devemos comparar a taxa real sob fluoretação com a taxa que teria ocorrido *no mesmo período de tempo*, se a fluoretação *não* houvesse sido introduzida. Não podemos observar a última taxa, pois a fluoretação foi introduzida, e, portanto, a taxa de não fluoretação naquele período de tempo é contrafactual. Assim, substituímos pela taxa no período de tempo antes da fluoretação. Ao fazê-lo, substituímos uma medida de associação (a diferença de taxas antes e depois da fluoretação) por aquilo que realmente nos interessa: a diferença entre a taxa com fluoretação e o que aquela taxa teria sido sem fluoretação no período de tempo pós-fluoretação.

Essa substituição seria ilusória, posto que a taxa antes da fluoretação não é igual – nem deveria ser trocada por ela – à taxa contrafactual (i.e., a taxa que teria ocorrido no período pós-fluoretação, se a fluoretação não tivesse sido introduzida). Se as duas não são iguais, então a medida de associação não será equivalente à medida de efeito pela qual ela foi substituída. Em tal circunstância, dizemos que a medida de associação é *confundida* (para nossa medida de efeito desejada). Outras maneiras de expressar a mesma ideia são que a diferença de taxas antes e depois é confundida pela diferença de taxas causal, ou que há *confundimento* na diferença antes e depois (Greenland e Robins, 1986, Greenland et al., 1999b, Greenland e Morgenstern, 2001). Contudo, se a taxa antes da fluoretação realmente for igual à taxa contrafactual pós-fluoretação, então a medida de associação é igual a nossa medida de efeito desejada, e dizemos que a diferença antes/depois não é confundida, ou que não há confundimento nessa diferença.

As definições precedentes aplicam-se a razões, assim como a diferenças. Visto que razões e diferenças contrastam com as mesmas quantidades subjacentes, o confundimento de uma medida de razão implica confundimento da medida de diferença correspondente, e vice-versa. Se o valor subs-

tituído na taxa, ou no risco, contrafactual, não igualar àquela taxa, ou ao risco, tanto a razão quanto a diferença serão confundidas.

As definições supracitadas também se estendem imediatamente a situações nas quais as quantidades contrastadas são riscos médios, tempos de incidência, chances ou prevalências. Por exemplo, pode-se querer estimar o efeito da fluoretação sobre a prevalência de cáries 3 anos depois que a fluoretação começou. Aqui a taxa contrafactual necessária, mas não observada, é qual teria sido a prevalência de cáries 3 anos depois do início da fluoretação, se esta, de fato, não tivesse iniciado. Poderíamos substituir aquela taxa contrafactual pela prevalência de cáries no tempo em que a fluoretação começou. É possível (embora, talvez, raro na prática) que uma medida de efeito seja confundida, mas não uma outra, se as duas medidas de efeito derivam de medidas subjacentes de frequência de doença diferentes (Greenland et al., 1999b). Por exemplo, teoricamente poderia haver confundimento na taxa de frequência, mas não na taxa de risco, ou poderia haver confundimento na razão de riscos aos 5 anos, mas não na razão de riscos aos 10 anos.

Um ponto de confusão na literatura é a falha em reconhecer que as chances de incidência são medidas baseadas em risco, e, consequentemente, as razões de chances de incidência serão confundidas sob exatamente as mesmas circunstâncias que as razões de riscos (Miettinen e Cook, 1981; Greenland e Robins, 1986; Greenland, 1987a, Greenland et al., 1999b). A confusão surge por causa da peculiaridade de que a razão de chances causal para uma coorte inteira pode ser mais próxima da nulidade do que qualquer razão de chances causal específica por estrato. Tal impossibilidade de colapsibilidade da razão de chances causal algumas vezes é confundida com confundimento, muito embora isso nada tenha a ver com o último fenômeno; isso será discutido mais adiante, em uma seção posterior.

## Confundidores

Considere novamente o exemplo sobre fluoretação. Suponha que dentro do ano seguinte ao início da fluoretação, programas educativos de higiene dental fossem implantados em algumas das escolas na comunidade. Se tais programas fossem efetivos, então (outras coisas sendo iguais), em consequência, alguma redução da incidência de cáries teria ocorrido. Assim, mesmo que a fluoretação não tivesse começado, a incidência de cáries teria declinado no período pós-fluoretação. Em outras palavras, os programas isoladamente teriam feito com que a taxa contrafactual em nossa medida de efeito fosse mais baixa do que a taxa pré-fluoretação que a substitui. Como resultado, a medida de associação (que é a diferença de taxas antes/depois) deve ser maior do que a taxa de efeito desejada (a diferença de taxas causal). Nessa situação, dizemos que os programas *confundiram* a medida de associação, ou que os efeitos dos programas são confundidos com o efeito da fluoretação na medida de associação. Também dizemos que os programas são *confundidores* da associação e que a associação é confundida pelos programas.

Confundidores são fatores (exposições, intervenções, tratamentos, etc.) que explicam ou produzem toda, ou parte, da diferença entre a medida de associação e a medida de efeito que seria obtida com uma taxa contrafactual ideal. No presente exemplo, os programas explicam por que a associação antes/depois exagera o efeito da fluoretação: a diferença, ou a razão antes/depois, inclui os efeitos de programas assim como os da fluoretação. Para um fator explicar essa discrepância e, assim, confundir, o fator deve afetar, ou pelo menos predizer, o risco ou a taxa no grupo não exposto (de referência), e não ser afetado pela exposição, ou pela doença. No exemplo precedente, presumimos que a presença dos programas de higiene dental nos anos depois da fluoretação seriam inteiramente responsáveis pela discrepância entre a taxa pré-fluoretação e a taxa (contrafactual) que teria ocorrido 3 anos depois da fluoretação, se esta não tivesse sido introduzida.

Uma grande parte dos métodos epidemiológicos dedica-se a evitar, ou ajustar (controlar), o confundimento. Tais métodos baseiam-se, inevitavelmente, na coleta e no uso apropriado de mensurações de confundidores. Retornaremos repetidamente a esse assunto. Por enquanto, simplesmente observaremos que a maioria dos métodos de ajuste fundamentais baseia-se na noção de *estratificação* de confundidores. Se fizermos nossas comparações dentro de níveis apropriados de um confundidor, este não poderá confundir as comparações. Por exemplo, poderíamos limitar nossas comparações

antes/depois da fluoreação a escolas em estados nos quais nenhum programa de higiene dental tivesse sido introduzido. Em tais escolas, introduções de programas não poderiam ter tido um efeito (porque nenhum programa estava presente), de modo que efeitos desses programas não poderiam explicar qualquer declínio seguinte à fluoreação.

Um modelo simples que distingue causalidade de associação

Podemos clarificar a diferença entre medidas de efeito e medidas de associação, assim como o papel de confundimento e confundidores, examinando medidas de risco sob um modelo simples potencial-desfecho para uma coorte de indivíduos (Greenland e Robins, 1986).

A Tabela 4.1 apresenta a composição de duas coortes, coorte 1 e coorte 0. Suponha que a coorte 1 é uniformemente exposta a algum agente de interesse, tal como uma postagem com material sobre cessação do fumo, e que a coorte 0 não é exposta, isto é, não recebe tal material. Os indivíduos nas coortes são classificados por seus desfechos quando expostos e quando não expostos:

- 1. Tipo1, ou pessoas “condenadas”, para as quais a exposição é irrelevante, porque a doença ocorre com ou sem exposição.
- 2. Tipo 2, ou pessoas “causais”, nas quais a doença ocorre se, e somente se, elas forem expostas.
- 3. Tipo 3, ou pessoas “preventivas”, nas quais a doença ocorre se, e apenas se, elas não forem expostas.
- 4. Tipo 4, ou pessoas “imunes”, para as quais a exposição novamente é irrelevante, porque a doença não ocorre, com ou sem exposição.

Entre os expostos, somente as pessoas tipo 1 e tipo 2 adquirem a doença, de modo que a proporção de incidência na coorte 1 é  $p_1 + p_2$ . Se, contudo, a exposição tivesse estado ausente nessa coorte, apenas as pessoas tipo 1 e tipo 3 teriam contraído a doença, de sorte que a proporção de incidência teria sido  $p_1 + p_3$ . Portanto, a mudança absoluta na proporção de incidência da coorte 1 causada pela exposição, ou a diferença de riscos causal é  $(p_1 + p_2) - (p_1 + p_3) = p_2 - p_3$ , ao passo que a mudança relativa, ou razão de riscos causal é  $(p_1 + p_2)/(p_1 + p_3)$ . De modo similar, a chance de incidência é  $(p_1 + p_2)/[1 - (p_1 + p_2)] = (p_1 + p_2)/(p_3 + p_4)$ , porém teria sido  $(p_1 + p_3)/[1 - (p_1 + p_3)] = (p_1 + p_3)/(p_2 + p_4)$  se a exposição tivesse estado ausente; assim, a mudança relativa na chance de incidência (a razão de chances causal) é

$$\frac{(p_1 + p_2)/(p_3 + p_4)}{(p_1 + p_3)/(p_2 + p_4)}$$

TABELA 4.1  
Um modelo elementar de tipos causais e sua distribuição em duas coortes distintas

Tipo	Resposta <sup>a</sup> sob		Descrição	Proporção de tipos em	
	Exposição	Não exposição		Coorte 1 (Expostos)	Coorte 0 (Não expostos)
1	1	1	Condenados	$p_1$	$q_1$
2	1	0	Exposição é causal	$p_2$	$q_2$
3	0	1	Exposição é preventiva	$p_3$	$q_3$
4	0	0	Imune	$p_4$	$q_4$

<sup>a</sup>1, contrai a doença; 0, não contrai a doença.  
Fonte: Reproduzida de Greenland S, Robins JM. Identifiability, exchangeability and epidemiological confounding. *Int J Epidemiol*. 1986; 15:413-419.

Números iguais de tipos causais (tipo 2) e de tipos preventivos (tipo 3) na coorte 1 correspondem a  $p_2 = p_3$ . A equivalência de  $p_2$  e  $p_3$  implica que a diferença de riscos causal  $p_2 - p_3$  será 0, e o risco causal e razões de chances serão 1. Assim, esses valores das medidas de efeito causal *não* correspondem a ausência efeito, mas, em vez disso, a um equilíbrio entre efeitos causais e preventivos.

A hipótese de nenhum efeito é chamada, algumas vezes, de hipótese nula exata, e aqui corresponde a  $p_2 = p_3 = 0$ . A hipótese nula exata é um caso especial da hipótese nula usual, em que a diferença de riscos é zero ou a razão de riscos é 1, o que corresponde aos efeitos causal e preventivo equilibrando-se um no outro, para produzir  $p_2 = p_3$ . Somente se estivermos seguros de que uma direção do efeito não acontece (ou  $p_2 = 0$ , ou  $p_3 = 0$ ), podemos dizer que uma diferença de riscos igual a 0, ou uma razão de riscos igual a 1, corresponde a nenhum efeito; caso contrário, só podemos dizer que aqueles valores correspondem a nenhum efeito *líquido*. De maneira mais geral, as medidas de efeito populacionais correspondem somente a efeitos líquidos: uma diferença de riscos representa apenas a alteração líquida no risco médio produzido pela exposição.

Entre os não expostos, somente as pessoas tipo 1 e tipo 3 adquirem a doença, de modo que a proporção de incidência na coorte 0 é  $q_1 + q_3$  e a chance de incidência é  $(q_1 + q_3)/(q_2 + q_4)$ . Portanto, a diferença e a razão das proporções de incidência nas coortes são  $(p_1 + p_2) - (q_1 + q_3)$  e  $(p_1 + p_2)/(q_1 + q_3)$ , ao passo que a razão de chances de incidência é

$$\frac{(p_1 + p_2)/(p_3 + p_4)}{(q_1 + q_3)/(q_2 + q_4)}$$

Essas medidas comparam duas coortes diferentes, a de expostos e a de não expostos, e são medidas de associação e não causais. Elas somente se igualam a suas correspondentes causais se  $q_1 + q_3 = p_1 + p_3$ , isto é, somente se a proporção de incidência da coorte 0 equivaler ao que a coorte 1 teria experimentado, se a exposição fosse ausente. Se  $q_1 + q_3 \neq p_1 + p_3$ , então a quantidade  $q_1 + q_3$  não é uma substituição válida para  $p_1 + p_3$ . Nesse caso, a diferença de riscos de associação, a razão de riscos e a razão de chances são confundidas pela discrepância entre  $q_1 + q_3$  e  $p_1 + p_3$ , assim dizemos que confundimento está presente nas comparações de risco.

O confundimento corresponde à diferença entre a quantidade contrafactual desejada  $p_1 + p_3$ , e a substituta observada  $q_1 + q_3$ . Essa diferença nasce das diversidades entre as coortes de expostos e não expostos, com relação a outros fatores que afetam o risco de doença, os confundidores. O controle do confundimento seria conseguido se pudéssemos estratificar as coortes sobre um conjunto suficiente desses confundidores, ou sobre fatores associados a eles, para produzir estratos dentro dos quais o contrafactual e seu substituto fossem iguais, isto é, dentro dos quais não ocorresse confundimento.

O confundimento depende da coorte para a qual estamos estimando os efeitos. Suponha que estamos interessados no efeito relativo que a exposição terá sobre o risco na coorte 0. Esse efeito seria mensurado pela razão causal para a coorte 0:  $(q_1 + q_2)/(q_1 + q_3)$ . Visto que a coorte 0 não é exposta, não observamos  $q_1 + q_2$ , o risco médio que ela teria se exposta; isto é,  $q_1 + q_2$  é contrafactual. Se substituirmos o risco médio real da coorte 1,  $p_1 + p_2$ , por esta média de risco contrafactual na coorte 0, obteríamos a mesma razão de riscos de associação usada antes:  $(p_1 + p_2)/(q_1 + q_3)$ . Mesmo que essa razão de associação igualasse à razão de riscos causal para a coorte 1 (o que só ocorre se  $p_1 + p_3 = q_1 + q_3$ ), ela não igualaria à razão de riscos causal para a coorte 0, a menos que  $p_1 + p_2 = q_1 + q_2$ . Para ver isto, suponha que  $p_1 = p_2 = p_3 = q_1 = q_3 = 0,1$  e  $q_2 = 0,3$ . Então  $p_1 + p_3 = q_1 + q_3 = 0,2$ , mas  $p_1 + p_2 = 0,2 \neq q_1 + q_2 = 0,4$ . Assim, não há confundimento no uso da razão de associação  $(p_1 + p_2)/(q_1 + q_3) = 0,2/0,2 = 1$  para a razão causal na coorte 1,  $(p_1 + p_2)/(p_1 + p_3) = 0,2/0,2 = 1$ , no entanto há confundimento no uso da razão de associação pela razão causal na coorte 0, pois esta última é  $(q_1 + q_2)/(q_1 + q_3) = 0,4/0,2 = 2$ . Esse exemplo mostra que a presença de confundimento pode depender da população escolhida como o alvo de inferência (a população-alvo), assim como da população eleita para fornecer uma substituta da quantidade contrafactual no alvo (a população de referência). Pode depender, também, do período de tempo em questão.



Diagramas causais (modelos gráficos) fornecem modelos visuais para distinguir causalidade de associação e, assim, para definir e detectar o confundimento (Pearl, 1995, 2000; Greenland et al., 1999a; Cap. 12). Modelos de desfecho potencial e modelos gráficos podem ser ligados por meio de uma terceira classe de modelos causais, chamados de *equações estruturais*, e levar aos mesmos critérios operacionais para detecção e controle de confundimento (Greenland et al., 1999a; Pearl, 2000; Greenland e Brumback, 2002).

## RELAÇÕES ENTRE MEDIDAS DE EFEITO

### Relações entre riscos relativos

Recorde-se do Capítulo 3, em que uma população fechada durante um intervalo de duração  $\Delta t$ , a proporção de incidência  $R$ , a taxa  $I$  e a chance  $R/S$  (onde  $S = 1 - R$ ) estarão relacionadas por  $R \approx I\Delta t \approx R/S$ , se o tamanho da população em risco declinar apenas levemente durante o intervalo (o que implica que  $R$  deve ser pequena e  $S = 1 - R \approx 1$ ). Suponha agora que contrastamos a experiência da população durante o intervalo sob duas condições, exposição e não exposição, e que o tamanho da população em risco declinaria apenas levemente sob cada condição. Então, a aproximação precedente implica que

$$\frac{R_1}{R_0} \approx \frac{I_1 \Delta t}{I_0 \Delta t} = \frac{I_1}{I_0} \approx \frac{R_1/S_1}{R_0/S_0}$$

onde  $S_1 = 1 - R_1$  e  $S_0 = 1 - R_0$ . Em outras palavras, as razões dos riscos, as taxas e as chances serão aproximadamente iguais sob condições adequadas. A condição de que tanto  $R_1$  quanto  $R_0$  sejam pequenas é suficiente para garantir que tanto  $S_1$  quanto  $S_0$  sejam próximos de 1, caso em que a razão de chances se aproximará da razão de riscos (Cornfield, 1951). Para a razão de taxas se aproximar da razão de riscos, devemos ter  $R_1/R_0 \approx I_1 T_1 / I_0 T_0 \approx I_1 I_0$ , o que requer que a exposição afete apenas de forma irrelevante o pessoa-tempo em risco (i.e., que  $T_1 \approx T_0$ ). Ambas as condições seriam satisfeitas se o tamanho da população em risco declinasse por não mais do que um pequeno percentual durante o intervalo, independentemente do *status* da exposição.

A ordem das três razões (risco, taxa e chance) em relação à nulidade é previsível. Quando  $R_1 > R_0$ , temos  $S_1 = 1 - R_1 < 1 - R_0 = S_0$ , de modo que  $S_0/S_1 > 1$  e

$$1 < \frac{R_1}{R_0} < \frac{R_1}{R_0} \frac{S_0}{S_1} = \frac{R_1/S_1}{R_0/S_0}$$

No entanto, quando  $R_1 < R_0$ , temos  $S_1 > S_0$ , de modo que  $S_0/S_1 < 1$  e

$$1 > \frac{R_1}{R_0} > \frac{R_1}{R_0} \frac{S_0}{S_1} = \frac{R_1/S_1}{R_0/S_0}$$

Assim, quando a exposição afeta o risco médio, a razão de riscos será mais próxima da nulidade (1) do que a razão de chances.

Agora suponha que, como esperaríamos normalmente, o efeito da exposição sobre pessoa-tempo em risco está na direção oposta de seu efeito sobre o risco, de modo que  $T_1 < T_0$  se  $R_1 > R_0$ , e  $T_1 > T_0$  se  $R_1 < R_0$ . Então, se  $R_1 > R_0$ , temos  $T_1/T_0 < 1$  e assim

$$1 < \frac{R_1}{R_0} = \frac{I_1 T_1}{I_0 T_0} < \frac{I_1}{I_0}$$

e se  $R_1 < R_0$ , temos  $T_1/T_0 > 1$  e assim

$$1 > \frac{R_1}{R_0} = \frac{I_1 T_1}{I_0 T_0} > \frac{I_1}{I_0}$$



Assim, quando a exposição afeta o risco médio, esperaríamos, normalmente, que a razão de riscos estivesse mais próxima da nulidade que a razão de taxas. Ainda, a razão de taxa estará mais próxima da nulidade do que a razão de chances (Greenland e Thomas, 1982). Assim, geralmente esperaríamos que a razão de riscos fosse mais próxima da nulidade, a razão de chances mais distante da nulidade, e que a razão de taxas caísse entre a razão de riscos e a razão de chances.

## Modificação da medida de efeito (heterogeneidade)

Suponha que dividamos nossa população em duas ou mais categorias, ou estratos. Em cada estrato, podemos calcular uma medida de efeito de nossa escolha. Essas medidas de efeito específicas por estrato podem, ou não, ser iguais uma à outra. Raramente teríamos qualquer razão para supor que elas seriam iguais. Se elas não são iguais, dizemos que a medida de efeito é *heterogênea* ou *modificada*, ou que *varia* ao longo dos estratos. Se elas são iguais, dizemos que a medida é *homogênea*, *constante* ou *uniforme*, ao longo dos estratos.

Um ponto importante sobre modificação da medida de efeito é que, se os efeitos estão presentes, geralmente o caso será que não mais que uma das medidas de efeito discutidas será uniforme ao longo dos estratos. De fato, se a exposição tem qualquer efeito sobre uma medida de ocorrência, no máximo uma das medidas de efeito de razão ou diferença pode ser uniforme ao longo dos estratos. Como um exemplo, suponha que entre os homens o risco médio fosse 0,50, se a exposição estivesse presente, mas 0,20, se a exposição estivesse ausente, ao passo que entre as mulheres o risco médio fosse 0,10, se a exposição estivesse presente, mas 0,04, se estivesse ausente. Então, a diferença de riscos causal para os homens seria  $0,50 - 0,20 = 0,30$ , cinco vezes a diferença para mulheres, de  $0,10 - 0,04 = 0,06$ . Em contraste, tanto para homens como para mulheres, a razão de riscos causal é  $0,50/0,20 = 0,10/0,04 = 2,5$ . Agora suponha que mudemos esse exemplo para tornar as diferenças de riscos uniformes, por exemplo, colocando o risco dos homens expostos em 0,26, em vez de 0,50. Então, ambas as diferenças de riscos seriam 0,06, mas a razão de riscos masculina seria  $0,26/0,20 = 1,3$ , muito menor do que a razão de riscos feminina de 2,5. Finalmente, se modificarmos o exemplo, tornando o risco masculino exposto 0,32 em vez de 0,50, a diferença de riscos masculina seria 0,12, o dobro da diferença de riscos feminina de 0,06, mas a razão masculina seria 1,6, menos de dois terços da razão feminina de 2,5. Assim, a presença, a direção e o tamanho da modificação da medida de efeito pode ser dependente da escolha da medida (Berkson, 1958; Brumback e Berg, 2008).

## Relação entre medidas estrato-específicas e medidas gerais

A relação de medidas de efeito específicas por estrato com a medida de efeito para uma coorte inteira pode ser sutil. Para diferenças e razões de riscos causais, a medida para a coorte inteira deve cair em algum lugar no meio das medidas estrato-específicas. Para as razões de chance, entretanto, a razão de chances causal para o total da coorte pode ser mais próxima da nulidade do que qualquer das razões de chances causais para os estratos (Miettinen e Cook, 1981; Greenland, 1987a; Greenland et al., 1999b). Esse fenômeno bizarro algumas vezes é referido como *não colapsidade* da razão de chances causal. O fenômeno tem levado alguns autores a criticar a razão de chances como uma medida de efeito, exceto como uma aproximação a razões de riscos e de taxas. (Miettinen e Cook, 1981; Greenland, 1987a; Greenland et al., 1999b; Greenland e Morgenstern, 2001).

Como exemplo, suponha que temos uma coorte com 50% de homens, e que entre os homens o risco médio fosse 0,50, se a exposição estivesse presente, mas 0,20, se a exposição estivesse ausente, ao passo que entre as mulheres o risco médio seria 0,08, se a exposição estivesse presente, mas 0,02 se estivesse ausente. Então, a razão de chances causal seria

$$\frac{0,50/(1 - 0,50)}{0,20/(1 - 0,20)} = 4,0 \text{ para homens} \quad \text{e} \quad \frac{0,08/(1 - 0,08)}{0,02/(1 - 0,02)} = 4,3 \text{ para mulheres}$$

Para o total da coorte, o risco médio, se a exposição estivesse presente, seria exatamente a média dos riscos médios masculino e feminino,  $0,5(0,50) + 0,5(0,08) = 0,29$ ; da mesma forma, o risco médio com a exposição ausente seria  $0,5(0,20) + 0,5(0,02) = 0,11$ . Assim, a razão de chances causal para o total da coorte seria

$$\frac{0,29/(1 - 0,29)}{0,11/(1 - 0,11)} = 3,3$$

que é menor do que a razão de chances masculina e do que a razão de chances a feminina. Essa não colapsidade pode ocorrer porque, ao contrário da diferença e da razão de riscos, a razão de chance causal para a coorte total não é uma média ponderada das razões de chances causais específicas por estrato (Greenland, 1987a). Isso não deve ser confundido com o fenômeno do confundimento (Greenland et al., 1999b), que foi discutido antes. As taxas de razão causais e diferenças de taxas também podem exibir não colapsidade sem confundimento (Greenland, 1996a). Em particular, a razão de taxas causal para uma coorte total pode ser mais próxima da nulidade do que todas as razões de taxas causais estrato-específicas. Para demonstrar, estendemos o exemplo precedente como a seguir: suponha que o período de risco fosse o ano de 1º de janeiro a 31 de dezembro de 2000, que todas as pessoas que adoecessem ficassem doentes em 1º de janeiro, e que ninguém mais fosse removido do risco durante o ano. Então, as taxas seriam proporcionais às chances, porque nenhum dos casos contribuiria com uma quantidade significativa de pessoa-tempo. Em consequência, as razões de taxas causais para homens e mulheres seriam 4,0 e 4,3, ao passo que a razão de taxas causal para toda a coorte seria somente 3,3.

Como discutido antes, risco, taxa e razões de chance se aproximarão uns dos outros se a população diminuir de tamanho apenas levemente durante o período de risco, independente da exposição. Se essa condição se mantiver em todos os estratos, a razão de taxas e a razão de chances se aproximarão da razão de riscos nos estratos, e por isso ambas as medidas serão aproximadamente colapsáveis quando a razão de riscos for colapsável.

## FRAÇÕES ATRIBUÍVEIS

Freqüentemente são vistas medidas que tentam avaliar o impacto de uma exposição em saúde pública, medindo sua contribuição à incidência total sob exposição. Por conveniência, nos referiremos a toda a família de tais medidas fracionadas como *frações atribuíveis*. Os termos *risco atribuível por cento*, ou apenas *risco atribuível*, são freqüentemente usados como sinônimos, embora “risco atribuível” seja usado também para denotar a diferença de riscos (MacMahon e Pugh, 1970; Szklo e Nieto, 2000; Koepsell e Weiss, 2003). Tais frações podem ser divididas em duas classes amplas, que denominaremos *frações de excesso* e *frações etiológicas*.

Uma dificuldade fundamental é que as duas classes, costumam ser confundidas; entretanto, as frações de excesso podem ser muito menores do que as frações etiológicas, mesmo que a doença seja raras, ou que outras condições razoáveis sejam satisfeitas. Outra dificuldade é que as frações etiológicas não são estimáveis somente a partir de estudos epidemiológicos, mesmo que tais estudos sejam perfeitamente válidos: suposições sobre o mecanismo biológico subjacente devem ser introduzidas para estimar as frações etiológicas, e as estimativas serão muito sensíveis a tais suposições.

## Frações de excesso

Uma família de frações atribuíveis baseia-se em recalcular uma diferença de incidências como uma proporção, ou fração da incidência total sob exposição. Tal medida é  $(A_1 - A_0)/A_1$ , a quantidade de casos em excesso devido à exposição, que tem sido chamada de *fração de excesso* (Greenland e Robins, 1988). Em uma coorte, a fração da proporção de incidência exposta  $R_1 = A_1/N$  que é atribuível à exposição é exatamente igual à fração de excesso:

$$\frac{R_1 - R_0}{R_1} = \frac{A_1/N - A_0/N}{A_1/N} = \frac{A_1 - A_0}{A_1}$$

onde  $R_0 = A_0/N$  é o que a proporção de incidência seria sem exposição. Comparando essa fórmula com a fórmula anterior para a fração de risco ( $R_1 - R_0/R_1 = (RR - 1)/RR$ ), veremos que em uma coorte a quantidade de casos em excesso e a fração de risco são iguais.

A fração de taxa  $(I_1 - I_0)/I_1 = (RT - 1)/RT$  é frequentemente igualada, por equívoco, com a fração de excesso  $(A_1 - A_0)/A_1$ . Para verificar que as duas não são iguais, deixemos que  $T_1$  e  $T_0$  representem o tempo total em risco, que seria experimentado pela coorte sob exposição e sob não exposição durante o intervalo de interesse. A fração de taxa, então, é igual a

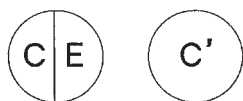
$$\frac{A_1/T_1 - A_0/T_0}{A_1/T_1}$$

Se a exposição tem algum efeito, e a doença remove as pessoas de risco ulterior (como quando a doença é irreversível), então  $T_1$  será menor do que  $T_0$ . Assim, a última expressão não pode igualar à fração de excesso  $(A_1 - A_0)/A_1$  porque  $T_1 \neq T_0$ , embora, se o efeito de exposição sobre o tempo total em risco for pequeno,  $T_1$  será próximo de  $T_0$ , e, assim, a fração de taxa se aproximará da fração de excesso.

## Frações etiológicas

Suponha que todas as causas suficientes de uma doença em particular fossem divididas em dois grupos, aqueles que contêm a exposição e aqueles que não contêm a exposição, e que a exposição nunca seja preventiva. Essa situação é resumida na Figura 4.1.  $C$  e  $C'$  podem representar muitas combinações diferentes de componentes causais. Cada um dos dois grupos de causas suficientes representa uma variedade de mecanismos causais para doença teoricamente grandes, talvez tantos quanto um mecanismo distinto para cada caso que ocorra. A doença pode ocorrer com ou sem  $E$ , a exposição de interesse. Os mecanismos causais são agrupados no diagrama conforme eles contenham, ou não, a exposição. Dizemos que a exposição pode causar doença, se ela o fizer sob pelo menos algum grupo de condições  $C$ . Dizemos que a exposição  $E$  causou a doença, se uma causa suficiente que contenha  $E$  for a primeira causa suficiente a ser completada.

A princípio, parece simples perguntar qual fração dos casos foi causada pela exposição. Chamaremos essa fração de *fração etiológica*. Visto que podemos estimar o número total de casos, poderíamos estimar a fração etiológica, se pudéssemos estimar o número de casos causados por  $E$ . Infelizmente, esse número não é estimável a partir de dados de incidência comuns, porque a observação de um caso exposto não revela o mecanismo que o causou. Particularmente, pessoas que têm a exposição podem desenvolver a doença por um mecanismo que não inclua a exposição. Por exemplo, um fumante pode desenvolver câncer de pulmão por meio de algum mecanismo que não envolva o fumo (p. ex., envolvendo asbestos, ou exposição à radiação, sem contribuição alguma do tabagismo). Para tais casos de câncer do pulmão, o fumo foi incidental; não contribuiu para a causação do câncer. Não há uma maneira geral de dizer que fatores são responsáveis por um dado caso. Portanto, os casos expostos incluem alguns casos de doença causados pela exposição, se a exposição for realmente uma causa, e alguns casos de doença que ocorrem por meio de mecanismos que não envolvem a exposição.



**FIGURA 4.1** • Dois tipos de causas suficientes de uma doença.

A taxa ou proporção de incidência observada entre os expostos reflete a incidência de casos em ambos os grupos de causas suficientes, representados na Figura 4.1. A incidência de causas suficientes contendo  $E$  poderia ser encontrada subtraindo-se a incidência de causas suficientes que não contêm  $E$ . A última incidência não pode ser estimada se não podemos distinguir casos nos quais a exposição desempenhou um papel etiológico, daqueles para os quais a exposição foi irrelevante (Greenland e Robins, 1988; Greenland, 1999a). Assim, se  $I_1$  é a taxa de incidência de doença em uma população quando a exposição está presente, e  $I_0$  é a taxa naquela população quando a exposição está ausente, a diferença de taxas  $I_1 - I_0$  não é igual, necessariamente, à taxa de doença surgindo de causas suficientes com a exposição como uma causa componente e não precisa sequer estar próxima daquela taxa.

Para entender a fonte dessa dificuldade, imagine uma coorte na qual, para cada membro, o complemento causal da exposição,  $C$ , será completado antes que a causa suficiente,  $C'$ , esteja completa. Se a coorte não é exposta, cada caso de doença deve ser atribuível à causa  $C'$ . Porém, se a coorte estiver exposta desde o início do seguimento, todos os casos de doença ocorrerão quando  $C$  estiver completa ( $E$  já estando presente), e, assim, cada caso da doença deve ser atribuível à causa suficiente contendo  $C$  e  $E$ . Então, a taxa de incidência de casos causados pela exposição é  $I_1$ , quando a exposição está presente, não  $I_1 - I_0$ , e assim a fração de casos causados pela exposição é 1, ou 100%, embora a fração da taxa  $(I_1 - I_0)/I_1$  possa ser muito pequena.

Frações de excesso e frações de taxa frequentemente são interpretadas de modo incorreto como frações etiológicas. O exemplo precedente mostra que essas frações podem ser bem menores do que a fração etiológica: no exemplo, a fração de taxa será próxima de 0 se a diferença de taxas for pequena em relação a  $I_1$ , mas a fração etiológica permanecerá 1, independente de  $A_0$  ou  $I_0$ . Robins e Greenland (1989a, 1989b) e Beyea e Greenland (1999) apresentam condições sob as quais a fração de taxa e a fração etiológica são iguais, mas tais condições não são testáveis com dados epidemiológicos, e raramente têm alguma evidência de suporte, ou plausibilidade genuína (Robins e Greenland, 1989a, 1989b). Uma condição citada algumas vezes é que a exposição age de modo independente das causas de segundo plano, que serão mais examinadas em uma seção posterior. Sem tais suposições, contudo, o máximo que podemos dizer é que a fração de excesso fornece um limite mais baixo sobre a fração etiológica.

Uma condição que é irrelevante, porém fornecida algumas vezes, é que a doença é rara. Para ver que essa condição é irrelevante, observe que o exemplo anterior não fez uso da frequência absoluta da doença; as frações excessiva e de taxa ainda podem estar próximas de 0, mesmo que a fração etiológica esteja próxima de 1. A raridade da doença apenas aproxima as frações de casos e taxas uma da outra, da mesma forma que traz as razões de riscos e de taxas para perto uma da outra (presumindo que a exposição não tenha um grande efeito sobre o pessoa-tempo); ela não aproxima a fração de taxa da fração etiológica.

## Probabilidade de causalidade e suscetibilidade à exposição

Para ilustrar a diferença entre frações de excesso e fração etiológica, suponha que em um dado tempo em uma coorte, a fração  $F$  de realizações de  $C'$  foi precedida por realizações de  $C$ . Novamente, nenhum caso pode ser atribuível à exposição se a coorte não for exposta. Porém, se a coorte for exposta, uma fração  $F$  dos casos  $A_0$ , que teriam ocorrido sem exposição, agora será causada pela exposição. Além disso, pode haver casos causados pela exposição nos quais a doença nunca tenha ocorrido. Deixemos que  $A_0$  e  $A_1$  sejam os números de casos que ocorreriam durante um dado intervalo, quando a exposição estivesse ausente e presente, respectivamente. Uma fração  $1 - F$  de  $A_0$  casos não seria afetada pela exposição; para tais casos, as realizações de  $C'$  precedem as realizações de  $C$ . O produto  $A_0(1 - F)$  é o número de casos não afetados pela exposição. A subtração desse produto de  $A_1$  dá  $A_1 - A_0(1 - F)$  para o número de casos nos quais a exposição desenvolveu um papel etiológico. A fração de casos  $A_1$  atribuível à  $C$  (uma causa suficiente com exposição) é, assim,

$$\frac{A_1 - A_0(1 - F)}{A_1} = 1 - (1 - F)/RR \quad [4.1]$$

Se escolhermos um caso aleatoriamente, essa fórmula de fração etiológica iguala à probabilidade de que a exposição causou aquele caso, ou a *probabilidade de causalidade* para o caso. Embora seja de grande interesse biológico e legal, essa probabilidade não pode ser estimada epidemiologicamente, se nada é sabido sobre a fração  $F$  (Greenland e Robins, 1988, 2000; Greenland, 1999a; Beyea e Greenland, 1999; Robins e Greenland, 1989a, 1989b). Esse problema é mais discutido no Capítulo 16, sob o tópico de estimação da fração atribuível.

Para exposições preventivas, deixemos agora que  $F$  seja a fração de casos expostos  $A_1$  para os quais a doença teria sido causada por um mecanismo requerendo ausência de exposição (i.e., não exposição, ou não  $E$ ), se a exposição estivesse ausente. Então, o produto  $A_1(1 - F)$  é o número de casos não afetados pela exposição; a subtração desse produto de  $A_0$  dá  $A_0 - A_1(1 - F)$  para o número de casos nos quais a exposição desempenharia um papel preventivo. A fração de casos não expostos  $A_0$  que foram causados por não exposição (i.e., atribuíveis a uma causa suficiente com não exposição) é, assim,

$$[A_0 - A_1(1 - F)]/A_0 = 1 - RR(1 - F)$$

Como com a fração etiológica, essa fração não pode ser estimada, se nada se sabe sobre  $F$ .

Retornando a uma exposição causal, é presumido comumente, frequentemente sem uma declaração ou evidência apoiadora, que o completamento de  $C$  e  $C'$  ocorre independentemente na coorte, de modo que a probabilidade de “susceptibilidade” à exposição,  $\Pr(C)$ , pode ser derivada pelas leis básicas de probabilidade para eventos independentes. Agora,  $\Pr(C') = A_0/N = R_0$ ; assim, sob independência,

$$\begin{aligned}\Pr(C \text{ ou } C') &= A_1/N = R_1 \\ &= \Pr(C) + \Pr(C') - \Pr(C)\Pr(C') \\ &= \Pr(C) + R_0 - \Pr(C)R_0\end{aligned}\quad [4.2]$$

O rearranjo produz

$$\Pr(C) = \frac{A_1/N - A_0/N}{1 - A_0/N} = \frac{R_1 - R_0}{1 - R_0}\quad [4.3]$$

A expressão da direita é a diferença de riscos causal dividida pela proporção sobrevivente sob não exposição. Daí, a equação pode ser reescrita

$$\Pr(C) = (R_1 - R_0)/S_0 = (S_0 - S_1)/S_0 = 1 - S_1/S_0$$

Essa medida foi derivada primeiro por Sheps (1958), que se referiu a ela como a *diferença relativa*; foi proposta mais tarde como um índice de susceptibilidade a efeitos de exposição por Khoury e colaboradores (1989a) com base na presunção de independência. Porém, como com a condição de independência, não se pode verificar a equação 4.3 a partir somente de dados epidemiológicos, e é raramente plausível, se é que ocorre, verificá-la sobre fundamentos biológicos.

## Uma nota sobre terminologia

Mais do que com outros conceitos, há uma terminologia profundamente inconsistente e confusa na literatura sobre frações atribuíveis. Levin (1953) usou o termo *proporção atribuível* para sua medida original de impacto populacional de doença, a qual, em nossos termos, é uma fração de excesso ou fração de risco. Muitos textos epidemiológicos, daí em diante, usaram o termo *risco atribuível* para se referir à diferença de riscos  $R_1 - R_0$ , e chamaram a medida de Levin de um *percentual de risco atribuível* (p. ex., MacMahon e Pugh, 1970; Koepsell e Weiss, 2003). Por volta de 1970, contudo, porções da literatura de bioestatística começaram a chamar a medida de Levin de *risco atribuível* (p. ex., Walter, 1976; Breslow e Day, 1980), e, infelizmente, parte da literatura epidemiológica os

acompanhou. Alguns epidemiologistas se esforçaram para manter a distinção, pela introdução do termo *fração atribuível* para o conceito de Levin (Ouellet et al., 1979; Deubner et al., 1980); outros adotaram o termo *fração etiológica* para o mesmo conceito, e assim o confundiram com a fração de casos causados pela exposição. O termo *risco atribuível* continua a ser usado para conceitos completamente diferentes, tais como diferença de riscos, fração de risco, fração de taxa e a fração etiológica. Por causa dessa confusão, recomendamos que o termo *risco atribuível* seja totalmente evitado, e que o termo *fração etiológica* não seja usado para medidas relativas de excesso.

## GENERALIZANDO DEFINIÇÕES DE EFEITO

Por conveniência, apresentamos as definições anteriores para a situação na qual podemos imaginar a coorte de interesse sujeita a uma de duas condições, tratamentos (intervenções) ou categorias de exposição distintos, durante o (ou no início do) intervalo de tempo de interesse. Pensamos nessas exposições, normalmente, aplicadas separadamente para cada membro da coorte. Porém, para estudar intervenções em saúde pública, devemos generalizar nosso conceito de exposição a populações gerais e permitir a variação em efeitos de exposição entre indivíduos e subgrupos. Consideraremos, daqui em diante, a “exposição” de uma população como se referindo ao *padrão* de exposição (ou tratamento) entre indivíduos na população. Isto é, consideraremos os subscritos 1 e 0 para denotar *distribuições* diferentes de exposição ao longo da população. Nessa visão, as medidas de efeito referem-se a comparações de distribuições de desfechos sob pares diferentes de padrões de distribuição ao longo da população de interesse (Greenland, 2002a; Maldonado e Greenland, 2002). Para ilustrar esse conceito epidemiológico geral de efeito, suponha que nossa população compreende apenas três membros no começo de um intervalo de 5 anos, cada um dos quais fuma um maço de cigarros por dia no início do intervalo. Vamos dar a essas pessoas números de identificação 1, 2 e 3, respectivamente. Suponha que estamos preocupados com o efeito das distribuições (padrões) diferentes da postagem com material antitumo, sobre a experiência com mortalidade dessa população durante o intervalo. Um padrão de exposição possível é

Pessoa 1: Postagem do material no começo do intervalo e a cada quadrimestre.

Pessoa 2: Postagem do material no começo do intervalo e anualmente.

Pessoa 3: Nenhum material postado.

Chamemos esse de padrão de exposição 0 ou padrão de referência. Outro padrão possível é:

Pessoa 1: Nenhum material postado.

Pessoa 2: Postagem do material no início do intervalo e anualmente.

Pessoa 3: Postagem do material no começo do intervalo e a cada quadrimestre.

Chamemos esse de padrão de exposição 1 ou o índice padrão; ele difere do padrão 0 apenas porque os tratamentos das pessoas 1 e 3 foram intercambiados.

Sob ambos os padrões, um terço da população recebe materiais anualmente, um terço os recebe a cada quadrimestre e um terço não recebe. Entretanto, é perfeitamente razoável que o padrão 0 possa produzir um desfecho diferente do padrão 1. Por exemplo, suponha que a pessoa 1 simplesmente descarte os materiais sem abri-los e, sob qualquer dos padrões, continue fumando e morra no quarto ano, por um câncer relacionado com o fumo. A pessoa 2 recebe o mesmo tratamento sob ambos os padrões; suponha que, sob qualquer padrão, ela faleça no primeiro ano, por um infarto do miocárdio. Porém, suponha que a pessoa 3 continuasse fumando sob o padrão 0, até que no terceiro ano ela falecesse por um acidente vascular cerebral relacionado com o fumo, ao passo que, no padrão 1, ela lesse os materiais recebidos, tivesse sucesso em deixar de fumar no segundo ano, e, em consequência, não sofresse acidente vascular cerebral, ou outra causa de morte, antes do fim do seguimento.

O total de óbitos e o tempo vivido sob o padrão de exposição 0 seria  $A_0 = 3$  (todos morrem) e  $T_0 = 4 + 1 + 3 = 8$  anos, ao passo que o total de mortes e o tempo de vida no padrão de exposição 1



seria  $A_1 = 2$  e  $T_1 = 4 + 1 + 5 = 10$  anos. Os efeitos do padrão 1 *versus* padrão 0, nessa população, seriam, assim, diminuir a taxa de incidência de  $3/8 = 0,38$  por ano para  $2/10 = 0,20$  por ano, uma diferença de taxas causal de  $0,20 - 0,38 = -0,18$  por ano, e uma razão de taxas causal de  $0,20/0,38 = 0,53$ ; para diminuir a proporção de incidência de  $3/3 = 1,00$  para  $2/3 = 0,67$ , uma diferença de riscos causal de  $0,67 - 1,00 = -0,33$  e uma razão de riscos causal de  $0,67/1,00 = 0,67$ ; e para aumentar o total de anos de vida de 8 para 10 anos. A fração de mortes que é passível de prevenção pelo padrão 1 é  $(3 - 2)/3 = 0,33$ , que iguala à fração de óbitos sob o padrão 0, para a qual a mudança para o padrão 1 teria relevância etiológica. Em contraste, a fração da taxa “prevenida” (removida) pelo padrão 1 em relação ao padrão 0 é  $(0,38 - 0,20)/0,38 = 1 - 0,53 = 0,47$  e representa somente a redução de taxa sob o padrão 1; ela não iguala a uma fração etiológica.

Esse exemplo ilustra dois pontos-chave que os epidemiologistas devem ter em mente quando estiverem interpretando medidas de efeito:

1. Os efeitos sobre taxas de incidência não são os mesmos que os efeitos sobre proporções de incidência (riscos médios). A terminologia comum, tal como “risco relativo”, atrai confusão entre as medidas de efeito. A menos que o desfecho seja incomum para todos os padrões de exposição sob estudo durante o intervalo de interesse, o tipo de risco relativo deve ser mantido distinto. No exemplo precedente, a razão de taxas foi 0,53, ao passo que a razão de riscos foi 0,67. Da mesma forma, o tipo de fração atribuível deve ser mantido distinto. A fração de mortes passíveis de prevenção foi 0,33, ao passo que a fração da taxa passível de prevenção foi 0,47.
2. Nem todos os indivíduos respondem de modo semelhante a exposições ou a tratamentos. Portanto, nem sempre é suficiente distinguir padrões de exposição por resumos simples, tais como “80% expostos” *versus* “20% expostos”. No exemplo precedente, ambos os padrões de exposição tiveram um terço da população recebendo materiais pelo correio a cada quadrimestre e um terço os recebendo anualmente, de modo que os padrões foram indistinguíveis com base na prevalência da exposição. Os efeitos foram produzidos inteiramente pelas diferenças de resposta das pessoas tratadas.

## FRAÇÕES ATRIBUÍVEIS À POPULAÇÃO E FRAÇÕES DE IMPACTO

Vê-se frequentemente *percentual de risco atribuível à população*, ou *fração atribuível à população*, definido como a redução de incidência que seria conseguida se a população tivesse sido inteiramente não exposta, em comparação com seu padrão atual (real) de exposição. Esse conceito, devido a Levin (1953, que o chamou de uma *proporção atribuível*), é um caso especial de definição de fração atribuível com base em padrão de exposição. Em particular, é uma comparação da incidência (ou taxa, ou número de casos, que devem ser mantidos distintos) sob o padrão observado de exposição, com a incidência sob um padrão contrafactual, no qual a exposição, ou o tratamento, está inteiramente ausente da população.

A remoção completa de uma exposição é, frequentemente, muito irrealista, tal como com o tabagismo e com a poluição do ar; mesmo com restrições legais e programas de cessação ou de despoluição, muitas pessoas continuarão a se expor, ou a ser expostas. Uma medida que dá margem a essas realidades é a *fração de impacto* (Morgenstern e Bursic, 1982), que é uma comparação da incidência sob o padrão de exposição observado com a incidência sob um padrão contrafactual, no qual a exposição só é parcialmente removida da população. Novamente, esse é um caso especial de nossa definição atribuível com base no padrão de exposição.

## MEDIDAS PADRONIZADAS DE ASSOCIAÇÃO E EFEITO

Considere novamente o conceito de padronização, conforme introduzido no final do Capítulo 3. Dada uma distribuição padrão  $T_1, \dots, T_K$  de pessoas-tempo ao longo de categorias  $K$ , ou estratos definidos



por uma ou mais variáveis e um esquema  $I_1, \dots, I_K$  de taxas de incidência naquelas categorias, temos a taxa padronizada

$$I_s = \frac{\sum_{k=1}^K T_k I_k}{\sum_{k=1}^K T_k}$$

que é a média das  $I_k$  ponderadas pelas  $T_k$ . Se  $I_1^*, \dots, I_K^*$  representa outro esquema de taxas para as mesmas categorias, e

$$I_s^* = \frac{\sum_{k=1}^K T_k I_k^*}{\sum_{k=1}^K T_k}$$

é a taxa padronizada para esse esquema, então

$$IR_s = \frac{I_s}{I_s^*} = \sum T_k I_k / \sum T_k I_k^*$$

é chamada uma *razão de taxas padronizadas*. O aspecto definidor dessa razão é que a mesma distribuição-padrão é usada para ponderar a taxa do numerador e do denominador. Similarmente,

$$ID_s = \sum T_k I_k / \sum T_k - \sum T_k I_k^* / \sum T_k = \sum (T_k I_k - T_k I_k^*) / \sum T_k$$

é chamada a *diferença de taxas padronizadas*; note que ela não é somente uma diferença de taxas padronizadas, mas é também uma média ponderada das diferenças de taxas específicas por estrato  $I_k - I_k^*$ , usando os mesmos pesos que foram utilizados para a padronização (a  $T_k$ ).

Suponha que  $I_1, \dots, I_K$  representam as taxas observadas ou previstas para estratos de uma dada população-alvo, se ela é exposta a alguma causa ou preventivo de doença,  $T_1, \dots, T_K$  são as pessoas-tempo observadas em estratos daquela população, e  $I_1^*, \dots, I_K^*$  representam as taxas previstas ou observadas para estratos da população, se ela não é exposta. A suposição então é que  $RT_s = I_s/I_s^*$  e  $ID_s$  são os efeitos de exposição nessa população, comparando as taxas gerais (brutas) que ocorreriam sob condições de exposição distintas. Essa interpretação presume, entretanto, que a distribuição de pessoas-tempo não seria afetada pela exposição.

Se  $I_1^*, \dots, I_K^*$  representam taxas contrafactuais, em vez de reais, digamos, porque a população realmente foi exposta, então  $I_s^*$  não precisa representar a taxa geral que ocorreria na população se a exposição fosse removida. Por exemplo, a modificação em taxas da  $I_k$  para a  $I_k^*$  poderia deslocar a distribuição pessoa-tempo  $T_1, \dots, T_K$  para  $T_1^*, \dots, T_K^*$ . Além disso, como discutido anteriormente, a exposição poderia afetar riscos competitivos, e esse efeito também poderia desviar a distribuição pessoa-tempo. Se esse desvio for grande, a razão e a diferença de taxas padronizadas não refletirão adequadamente o efeito real da exposição sobre a taxa de doença (Greenland, 1996).

Há umas poucas condições especiais sob as quais o efeito da exposição sobre pessoa-tempo não afetará a razão de taxas padronizadas. Se as razões estrato-específicas  $I_k/I_k^*$  são constantes ao longo das categorias, a razão de taxas padronizadas será igual a essa razão constante estrato-específica. Se a exposição tem apenas um efeito pequeno sobre pessoa-tempo, então, não importa a distribuição pessoa-tempo usada como o padrão, a diferença entre uma razão padronizada e o efeito real também será pequena. Em geral, contudo, é preciso estar atento para o fato de que uma suposição essencial é necessária para permitir que se interprete uma razão de taxas padronizadas como uma medida de efeito, mesmo que não haja problema metodológico com as observações. Analogamente, a diferença

de taxas padronizadas não será uma medida de efeito, salvo quando a exposição não afetar a distribuição pessoa-tempo, ou quando existirem outras condições especiais, tais como diferenças de taxas constante  $I_k - I_k^*$  ao longo das categorias.

As proporções de incidência têm denominadores  $N_1, \dots, N_K$ , que não são afetados por alterações de taxas, ou por riscos competitivos. Assim, se esses denominadores são usados para criar razões e diferenças de riscos padronizados, as medidas resultantes podem ser interpretadas como medidas de efeito, sem a necessidade das suposições especiais requeridas para interpretação de razões e diferenças de taxas padronizadas.

## Razões de morbidade padronizadas (RMP)

Quando a distribuição de pessoas-tempo expostas fornece o padrão, a razão de taxas padronizadas adota uma forma simplificada. Suponha que  $T_1, \dots, T_K$  representam as pessoas-tempo expostas,  $A_1, \dots, A_K$  sejam o número de casos entre os expostos,  $I_1, \dots, I_K$  sejam as taxas nos casos expostos e  $I_1^*, \dots, I_K^*$ , as taxas que teriam ocorrido nos casos expostos, se não tivessem sido expostos. Então, em cada estrato temos  $T_k I_k = A_k$ , e assim a razão de taxas padronizadas torna-se

$$\sum T_k I_k / \sum T_k I_k^* = \sum A_k / \sum T_k I_k^*$$

O numerador dessa razão é só o número total de casos expostos ocorrendo na população. O denominador é o número de casos que seria esperado ocorrer na ausência de exposição, se a exposição não afetasse a distribuição de pessoa-tempo. Essa razão de casos observados para esperados é chamada de *razão de morbidade padronizada* (RMP), *razão de incidência padronizada* (RIP), ou, quando o desfecho é óbito, a *razão de mortalidade padronizada*. Quando proporções de incidência são usadas em lugar das taxas de incidência, o mesmo tipo de simplificação ocorre ao se tomar a distribuição de pessoas expostas como o padrão: a razão de riscos padronizada reduz-se a uma razão de casos observados para esperados.

Muitos estudos ocupacionais e ambientais, que examinam populações de trabalhadores expostos, tentam estimar as RMP pelo uso de categorias de idade/sexo/raça como estratos, e usando, então, taxas específicas de idade/sexo/raça da população geral, em lugar das taxas contrafactuais desejadas  $I_1^*, \dots, I_K^*$ . Um problema importante com essa prática é aquele do *confundimento de resíduo*. Normalmente haverá muitas outras diferenças entre a população geral além de suas distribuições por idade, sexo e raça (diferenças em tabagismo, em assistência à saúde, etc.), e algumas dessas diferenças confundirão a razão padronizada resultante. Esse é um exemplo do problema mais comum de confundimento de resíduo na epidemiologia observacional, ao qual retornaremos em capítulos posteriores.

As RMP estimadas por meio de categorias de exposição, ou de populações diferentes, algumas vezes são comparadas diretamente entre si para avaliar uma tendência de dose-resposta, por exemplo. Geralmente tais comparações não são totalmente padronizadas, porque a RMP de cada categoria de exposição é ponderada pela distribuição de pessoa-tempo, ou pessoas, daquela categoria, e essas ponderações não são necessariamente comparáveis ao longo das categorias de exposição. O resultado é o confundimento de resíduo pelas variáveis usadas para criar os estratos, assim como por variáveis não medidas (Yule, 1934; Breslow e Day, 1987; Greenland, 1987e). Há, contudo, várias circunstâncias sob as quais essa diferença em ponderações não levará a confundimento importante (além do problema do confundimento de resíduo discutido).

Uma circunstância é quando as populações comparadas diferem pouco em sua distribuição de pessoa-tempo ao longo dos estratos (p. ex., quando elas têm distribuições idade/sexo/raça similares). Outra circunstância é quando os fatores de estratificação têm pouco efeito sobre o desfecho em estudo (o que não é comum; idade e sexo são fortemente relacionados à maioria dos desfechos). Outra circunstância, ainda, é quando as razões estrato-específicas são quase constantes ao longo dos estratos (nenhu-

ma modificação da razão pelas variáveis de padronização) (Breslow e Day, 1987). Embora nenhuma dessas circunstâncias possa se manter exatamente, a primeira e a última, juntas, frequentemente são grosseiramente aproximadas; quando isso ocorre, a falta de padronização mútua entre as RMP comparadas levará a pouca distorção. Então, a atenção pode voltar-se para muitos outros problemas de validade, que são o pesadelo dos estudos de RMP, tais como confundimento de resíduo, dados perdidos e erros de mensuração (ver Caps. 9 e 19). Se, entretanto, não se pode estar confiante de que o viés devido à comparação direta de RMP seja pequeno, as estimativas devem ser baseadas em um só padrão comum aplicado ao risco em todos os grupos, ou em um modelo de regressão que dê conta das diferenças entre as populações comparadas e os efeitos da exposição sobre pessoa-tempo (Cap. 20).

## RAZÕES DE PREVALÊNCIAS

No Capítulo 3, mostramos que as chances de prevalência bruta, ( $PO = prevalence\ odds$ ), é igual à taxa de incidência bruta,  $I$ , vezes a duração média da doença,  $\bar{D}$ , quando a população em risco e o grupo de prevalência são estacionários, e não há migração para dentro ou para fora do grupo de prevalência. Declarando novamente essa relação, separadamente para uma só população, sob exposição e sob não exposição, ou para uma população exposta e uma não exposta, temos

$$PO_1 = I_1 \bar{D}_1 \quad e \quad PO_0 = I_0 \bar{D}_0 \quad [4.5]$$

onde os subscritos 1 e 0 referem-se a expostos e a não expostos, respectivamente. Se a duração média da doença for a mesma, independente da exposição, isto é, se  $\bar{D}_1 = \bar{D}_0$ , a razão de chances de incidência, POR (*prevalence odds ratio*), bruta será igual à razão da taxa de incidência bruta RT:

$$POR = \frac{PO_1}{PO_0} = \frac{I_1}{I_0} = IR \quad [4.6]$$

Infelizmente, se a exposição afetar a mortalidade, também alterará a distribuição etária da população. Assim, como as pessoas mais velhas tendem a morrer mais cedo, a exposição afetará indiretamente a duração média, de modo que  $\bar{D}_1$  não será igual a  $\bar{D}_0$ . Em tal caso, a equação 4.6 não se manterá exatamente, embora ainda possa se manter aproximadamente (Newman, 1988).

## OUTRAS MEDIDAS

As medidas que discutimos não são, de modo algum, exaustivas de todas aquelas que foram propostas. Nem todas as medidas de efeito propostas se enquadram em nossa definição de medida de efeito – isto é, um contraste do desfecho de uma população *única* sob duas condições *diferentes*. Exemplos de medidas que *não* são medidas de efeito por nossa definição incluem coeficientes de correlação e medidas de redução de variância correlatas (Greenland, et al., 1986, 1991). Exemplos de medidas que são medidas de efeito por nossa definição, mas que não são discutidas em detalhes aqui, incluem expectativa de perda de anos de vida (Murray et al., 2002), assim como períodos de avanço de risco e de taxa (Brenner et al., 1993).

A perda de anos de vida,  $T_0/N - T_1/N$ , e a medida de razão correspondente,  $T_0/T_1$ , têm algumas vantagens dignas de nota sobre a taxa convencional e sobre as medidas de efeito baseadas em risco. Elas não estão sujeitas aos problemas de falta de estabilidade que surgem para as frações etiológicas (Robins e Greenland, 1991), nem a preocupações com os efeitos da exposição sobre o tempo em risco. De fato, elas representam o efeito da exposição sobre o tempo em risco. Elas são, contudo, mais difíceis de estimar estatisticamente a partir de dados epidemiológicos típicos, especialmente quando apenas dados de caso-controle (Cap. 8) estão disponíveis, o que pode explicar, em parte, sua popularidade limitada até então (Boshuizen e Greenland, 1997).

# Conceitos de interação

Sander Greenland, Timothy L. Lashman  
e Kenneth J. Rothman

## Interação estatística e modificação de medida de efeito 90

Dependência de escala da modificação de medida de efeito 90

## Interações biológicas 93

Desfechos potenciais para duas variáveis 93

Relação de distribuições de tipo de resposta a riscos médios 95

Relação de distribuições de tipo de resposta à aditividade 97

A impossibilidade da identificação de tipos de resposta de interação 98

Interações sob o modelo de causa suficiente 99

Relação entre o desfecho-potencial e os de interação de causa suficiente 100

Interação biológica *versus* estatística 102

## Interações em saúde pública 102

O conceito de interação é centrado na ideia de que o efeito de uma exposição, comparado com uma categoria de referência sem exposição, pode depender da presença de uma ou mais outras exposições. Um exemplo bem conhecido é concernente ao efeito da exposição ocupacional à poeira de asbestos sobre o risco de câncer de pulmão, o qual depende do estado de fumante (Berry e Liddell, 2004). Como uma ilustração hipotética, suponha que examinemos o risco médio em 10 anos de câncer de pulmão em um cenário ocupacional e encontremos que, entre trabalhadores com asbestos do sexo masculino, não fumantes, esse risco é de 3/1.000, e que o risco correspondente em homens comparáveis não fumantes que não trabalharam com asbestos é de 1/1.000. Suponha, também, que o risco é de 20/1.000 entre trabalhadores com asbestos do sexo masculino fumantes e é de 10/1.000 em homens comparáveis que fumavam e não trabalhavam com asbestos. Então, a razão de riscos associando o trabalho com asbestos com o risco de câncer de pulmão é de  $3/1 = 3$ , em não fumantes, maior do que a razão de riscos de  $20/10 = 2$ , nos fumantes. Em contraste, a diferença de riscos é  $3 - 1 = 2/1.000$ , em não fumantes, menor do que a diferença de riscos de  $20 - 10 = 10/1.000$ , entre os fumantes. Assim, quando se usa a medida de razão, parece que a associação entre exposição a asbestos e risco de câncer de pulmão é maior em não fumantes do que em fumantes. Quando usamos a medida de diferença, entretanto, parece que a associação é consideravelmente *menor* para não fumantes do que para fumantes.

A possível dependência de escala de uma avaliação de interação ilustra o tipo de assunto que complica a compreensão do conceito. Em verdade, o conceito de interação gerou muito debate quando, pela primeira vez, se tornou um foco para epidemiologistas, como visto em Rothman (1974, 1976a, 1976b), Koopman, (1977), Kupper e Hogan (1978), Walter e Holford (1978) e Siemiatycki e Thomas (1981). A literatura decorrente identificou numerosas distinções e conceitos, cujo delineamento tem ajudado a iluminar discordâncias iniciais e tem apontado o caminho para a elaboração de conceitos adicionais de interação; para exemplos, ver Blot e Day (1979), Rothman e colaboradores

(1980), Saracci (1980), Koopman (1981), Walker (1981), Miettinen (1982b), Weinberg (1986), Greenland e Poole (1988), Weed e colaboradores (1988), Thompson (1991), Greenland (1993b), Darroch e Borkent (1994), Darroch (1997) e VanderWeele e Robins (2007a, 2008a).

Em adição à dependência de escala, outro problema é a ambiguidade do termo *interação*, que tem sido usado para vários conceitos distintos, estatísticos, biológicos e de saúde pública. A falha na distinção entre esses conceitos foi responsável por muito da controvérsia inicial (Blot e Day, 1979; Saracci, 1980; Rothman e colaboradores, 1980). Uma vez feitas essas distinções, permanece a questão do que pode ser aprendido sobre interação a partir de dados epidemiológicos (Thompson, 1991).

O capítulo presente fornece definições e faz distinções entre conceitos de interação. O Capítulo 16 descreve como a análise estratificada pode ser usada para estudar interações e apresenta as limitações de tais métodos. Começamos pela discussão da interação estatística, um conceito que se refere a associações, causais ou não. A interação estatística é dependente de escala. Quando nenhum viés está presente, de modo que as associações observadas estimam, validamente, efeitos causais de interesse, a interação estatística corresponde à modificação de medida de efeito. Depois de discutir a relação da interação estatística com a modificação de efeito, discutimos modelos para interação biológica. Mostramos que, quando efeitos são mensurados por diferenças de riscos causal, e a interação biológica é definida como modificação de tipos de resposta potencial, a interação biológica está implicada em afastamentos da aditividade de efeitos. Também mostramos que a interação biológica pode estar presente mesmo quando há aditividade de efeitos, quando há tipos opostos de interação que se cancelam um ao outro, deixando o efeito líquido aditivo. Então, comparamos esse modelo de desfecho-potencial de interação biológica àquele baseado no modelo de causa suficiente-componente introduzido no Capítulo 2. Concluimos com uma discussão da interação em saúde pública.

## INTERAÇÃO ESTATÍSTICA E MODIFICAÇÃO DE MEDIDA DE EFEITO

Quando nenhum viés está presente, a definição de interação que é usada, frequentemente, em livros e programas computacionais de estatística (particularmente para análise de variância) é logicamente equivalente à definição de modificação de medida de efeito, ou heterogeneidade de efeito. Interação é frequentemente descrita como “afastamento da aditividade de efeitos na escala de desfecho escolhida”. Assim, os métodos para análise de interações estatísticas podem ser vistos como métodos para analisar a modificação de medida de efeito, sob a suposição de que todo viés tenha sido controlado adequadamente (ver Cap. 15).

Conforme visto no exemplo anterior dos efeitos de asbestos e fumo, a presença ou a ausência de interação estatística entre dois fatores  $X$  e  $Z$  depende da escala com a qual se escolhe mensurar sua associação. Suponha que tanto  $X$  quanto  $Z$  tenham efeitos, e a diferença de riscos para um permanece constante ao longo de níveis do outro, de modo que não há modificação das diferenças de riscos (i.e., há homogeneidade das diferenças de riscos). Se não há viés (de modo que as associações igualam os efeitos), esse estado de coisas corresponde a nenhuma interação estatística *na escala da diferença de riscos* para o efeito, porque o efeito combinado de  $X$  e  $Z$  sobre o risco pode ser calculado simplesmente pela soma das diferenças de riscos separadas para  $X$  e  $Z$ . No exemplo, há modificação de medida de efeito, ou interação estatística, na escala de diferença, porque o risco adicionado pela exposição a asbestos foi maior entre fumantes do que entre não fumantes. Também houve modificação de medida de efeito, ou interação estatística, entre asbestos e fumo na escala de razão de riscos para o efeito, porque a quantidade que os asbestos multiplicaram o risco foi menor entre fumantes do que entre não fumantes.

### Dependência de escala da modificação de medida de efeito

Como explicado no Capítulo 4, se tanto  $X$  quanto  $Z$  têm efeitos e não há modificação (heterogeneidade) das diferenças de riscos de um fator pelo outro fator, deve haver modificação das razões de riscos. Inversamente, se  $X$  e  $Z$  têm efeitos, e não há modificação das razões de riscos, deve haver modificação das diferenças de riscos. Comumente, tanto as diferenças de riscos quanto as razões de riscos para um

fator são heterogêneas através das categorias do outro. Naquele caso, as diferenças de riscos e as razões de riscos podem ser modificadas em direções opostas, como visto no exemplo para asbestos e fumo.

Para explicar por que a homogeneidade da medida de efeito em uma escala requer heterogeneidade da medida de efeito na outra escala, quando ambos os fatores têm efeitos, examinaremos primeiro o caso em que as diferenças de riscos são homogêneas e as razões de riscos são heterogêneas. Nós então examinaremos o caso oposto.

Para começar, escreva  $R_{ij}$  para o risco médio (proporção de incidência) quando  $X = i$  e  $Z = j$ , como na Tabela 5.1. Suponha que a diferença de riscos para  $X = 1$  versus  $X = 0$ , quando  $Z = 0$  (que é  $R_{10} - R_{00}$ ), seja igual à diferença de riscos para  $X = 1$  versus  $X = 0$ , quando  $Z = 1$  (que é  $R_{11} - R_{01}$ ):

$$R_{11} - R_{01} = R_{10} - R_{00} \quad [5.1]$$

Subtraindo-se  $R_{00}$  de cada lado e rearrumando, pode-se reescrever esta equação como

$$R_{11} - R_{00} = (R_{10} - R_{00}) + (R_{01} - R_{00}) \quad [5.2]$$

Esta equação mostra que a diferença de riscos para mudar o *status* da exposição de  $X = Z = 0$  para  $X = Z = 1$  pode ser encontrada simplesmente *somando* a diferença de riscos para  $X = 1$  versus  $X = 0$ , quando  $Z = 0$ , à diferença de riscos para  $Z = 1$  versus  $Z = 0$ , quando  $X = 0$ . Se dividirmos ambos os lados da equação 5.1 por  $R_{00}$  (o risco quando  $X = 0, Z = 0$ ), obteremos

$$\frac{R_{11}}{R_{00}} - \frac{R_{01}}{R_{00}} = \frac{R_{10}}{R_{00}} - 1 \quad [5.3]$$

Subtraindo 1 de cada lado da mesma equação e rearrumando, podemos reescrever esta equação em termos das razões de riscos excessivo:

$$\frac{R_{11}}{R_{00}} - 1 = \left( \frac{R_{10}}{R_{00}} - 1 \right) + \left( \frac{R_{01}}{R_{00}} - 1 \right) \quad [5.4]$$

Se tanto  $X$  como  $Z$  têm efeitos, a aditividade do excesso de razão de riscos na equação 5.4 implica que  $R_{11}/R_{01} \neq R_{10}/R_{00}$ ; isto é, a razão de riscos para  $X = 1$  versus  $X = 0$ , quando  $Z = 1$  ( $R_{11}/R_{01}$ ), não pode ser igual à razão de riscos para  $X = 1$  versus  $X = 0$ , quando  $Z = 0$  ( $R_{10}/R_{00}$ ). Nós alcançamos essa conclusão porque a igualdade

$$R_{11}/R_{01} = R_{10}/R_{00} \quad [5.5]$$

implica a possibilidade de *multiplicação* das razões de riscos:

$$R_{11}/R_{00} = (R_{11}/R_{01})(R_{01}/R_{00}) = (R_{10}/R_{00})(R_{01}/R_{00}) \quad [5.6]$$

**TABELA 5.1**

**Notação para riscos de duas variáveis (1,0) binárias de exposição**

	<b>Z = 1</b>	<b>Z = 0</b>	<b>Diferença de riscos</b>	<b>Razão de riscos</b>
<b>X = 1</b>	$R_{11}$	$R_{10}$	$R_{11} - R_{10}$	$R_{11}/R_{10}$
<b>X = 0</b>	$R_{01}$	$R_{00}$	$R_{01} - R_{00}$	$R_{01}/R_{00}$
<b>Diferença de riscos</b>	$R_{11} - R_{01}$	$R_{10} - R_{00}$		
<b>Razão de riscos</b>	$R_{11}/R_{01}$	$R_{10}/R_{00}$		

o que contradiz a equação 5.4, a menos que  $R_{10}/R_{00} = 1$  ou  $R_{01}/R_{00} = 1$ , porém nenhuma dessas razões de riscos pode ser igual a 1 quando tanto  $X$  como  $Z$  afetam o risco.

Para mostrar que a homogeneidade da razão de riscos requer heterogeneidade da diferença de riscos, comece por presumir nenhuma modificação da razão de riscos, de modo que a equação 5.5 se mantenha. Então, a equação 5.6 também deve manter-se, e podemos tirar o logaritmo de ambos os lados para obter a equação

$$\ln(R_{11}/R_{00}) = \ln(R_{10}/R_{00}) + \ln(R_{01}/R_{00}) \quad [5.7]$$

ou

$$\ln(R_{11}) - \ln(R_{00}) = \ln(R_{10}) - \ln(R_{00}) + \ln(R_{01}) - \ln(R_{00}) \quad [5.8]$$

A equação 5.7 mostra que o logaritmo da razão de riscos, para mudar o *status* da exposição de  $X = Z = 0$  para  $X = Z = 1$ , pode ser encontrada pela simples adição do logaritmo da razão de riscos para  $X = 1$  versus  $X = 0$ , quando  $Z = 0$ , ao logaritmo da razão de riscos para  $Z = 1$  versus  $Z = 0$ , quando  $X = 0$ . Assim, a homogeneidade (nenhuma modificação) da razão de riscos corresponde à aditividade (nenhuma interação estatística) na escala de risco logarítmico para o desfecho (equação 5.8). Os efeitos combinados são, simplesmente, a soma de efeitos na escala de risco logarítmico. Além disso, se tanto  $X$  quanto  $Z$  têm efeitos diferentes de zero, e esses efeitos são aditivos na escala de risco logarítmico, os efeitos não podem ser aditivos na escala de risco. Isto é, a ausência de interação estatística na escala de risco logarítmica (equação 5.7) implica a presença de interação estatística na escala de diferença de riscos, se ambos os fatores têm efeitos e não há viés.

Visto que a equação 5.7 de adição do logaritmo de riscos é equivalente à equação 5.6 de razão de riscos multiplicativa, aditividade do logaritmo da razão de riscos corresponde à capacidade de multiplicação da razão de riscos. Assim, “nenhuma interação multiplicativa” é descrito, frequentemente, como “nenhuma interação estatística na escala de logaritmo da razão de riscos”. Infelizmente, como a maioria das estatísticas epidemiológicas baseia-se em modelos multiplicativos, desenvolveu-se o mau hábito de abandonar a palavra *multiplicativa* e alegar que “não há interação”, toda vez que se acredita que os dados são consistentes com a equação 5.5 ou 5.6. Tal uso impreciso é convidativo à confusão com outros conceitos de interação. Para evitar tal confusão, aconselhamos que deva se referir à escala, ou à medida, que se está examinando com frases mais precisas, tais como “nenhuma heterogeneidade de razão de riscos foi evidente”, “nenhuma heterogeneidade de diferença de riscos foi evidente”, “nenhum afastamento da capacidade multiplicativa da razão de riscos foi evidente” ou “nenhum afastamento da capacidade aditiva da diferença de riscos foi evidente”, conforme for apropriado. O termo *modificação de efeito* também é ambíguo, e aconselhamos, novamente, o uso de termos mais precisos, tais como “modificação de diferença de riscos” ou “modificação de razão de riscos”, conforme for apropriado.

Outra fonte de ambiguidade é o fato de que as equações 5.1 a 5.8 inclusive podem ser reescritas usando-se um tipo diferente de medida de desfecho, tal como taxas, chances, prevalências, médias ou outras medidas, em lugar de riscos  $R_{ij}$ . Cada medida de desfecho leva a uma escala diferente para interação estatística e a um conceito correspondente de modificação de medida de efeito e heterogeneidade de efeito. Assim, quando ambos os fatores têm efeitos, a ausência de interação estatística em qualquer escala particular implica, necessariamente, a presença de interação estatística em muitas outras escalas.

Considere agora as medidas relativas de risco: razões de risco, razões de taxas e razões de chances. Se o risco de doença é baixo em todos os níveis das variáveis do estudo (i.e., menos que cerca de 0,1), a ausência de interação estatística para uma dessas medidas de razão implica a ausência de interação estatística para as outras duas medidas. Para riscos maiores, contudo, a ausência de interação estatística para uma medida de razão implica que deva haver alguma modificação das outras duas medidas de razão, quando ambos os fatores têm efeitos. Por exemplo, a ausência de modificação da razão de chances,



$$\frac{R_{11}/(1 - R_{11})}{R_{01}/(1 - R_{01})} = \frac{R_{10}/(1 - R_{10})}{R_{00}/(1 - R_{00})} \quad [5.9]$$

é equivalente a nenhuma interação multiplicativa na escala de chances. Porém, se  $X$  e  $Z$  têm efeitos, então a equação 5.9 implica que deve haver modificação da razão de riscos, de modo que as equações 5.6 a 5.8 inclusive não podem se manter, a menos que todos os riscos sejam baixos. De modo similar, a equação 5.9 também implica modificação da razão de taxas. Resultados paralelos são aplicáveis às medidas de diferença: se o risco da doença é sempre baixo, a ausência de interação estatística para uma das três (diferença de riscos, diferença de taxas ou diferença de chances) implica a ausência de interação estatística para as outras duas. Inversamente, se o risco de doença é alto, a ausência de interação estatística para uma medida de diferença implica que deve haver alguma modificação das outras duas, quando ambos os fatores têm efeitos.

Os exemplos precedentes e a álgebra demonstram que a interação estatística é um fenômeno cuja presença ou ausência, assim como magnitude, é determinada pela escala escolhida para medir afastamentos da capacidade aditiva de efeitos. Para evitar ambiguidade, deve-se especificar precisamente a escala na qual se está mensurando tais interações. Ao fazê-lo, é indesejável usar um termo tão vago como *interação*, porque frases mais precisas podem sempre substituir o uso do conceito equivalente de modificação de medida de efeito ou de heterogeneidade da medida de efeito.

## INTERAÇÕES BIOLÓGICAS

Há duas abordagens principais para o tópico de interação biológica (causal). Uma delas baseia-se em delinear mecanismos específicos de interação. O conceito de *interação mecanicista* raramente recebe uma definição precisa, mas sua intenção é abranger a noção de reações diretas, físicas ou químicas, entre exposições, seus metabólitos, ou seus produtos de reação, em indivíduos ou dentro de vetores de transmissão de exposição. Exemplos incluem a inibição da nitrosação gástrica das aminas e amidas da dieta pelo ácido ascórbico, e a supressão dos radicais livres nos tecidos por uma miscelânea de antioxidantes.

A descrição de um mecanismo pelo qual tais interações têm lugar, não leva imediatamente a previsões precisas sobre observações epidemiológicas. Uma razão é que raramente, se é que ocorre, um mecanismo proposto, que pode ser responsável por todos os casos observados de doença, ou por todos os efeitos de todos os fatores de risco, é mensurado e não mensurado. Ruído de fundo, sob a forma de efeitos não explicados e de interações biológicas com outros fatores, pode apagar facilmente qualquer padrão buscado pelo pesquisador. No entanto, têm havido esforços para testar hipóteses sobre mecanismos e interações biológicas, usando-se modelos abstratos simplificados. Tais esforços têm sido concentrados largamente na epidemiologia do câncer; para exemplo, ver Moolgavkar (1986, 2004).

Uma limitação-chave desse e de outros esforços com modelos biológicos é que qualquer padrão de dados especificado pode ser previsto a partir de numerosos mecanismos, ou de modelos para desenvolvimento de doenças, dissimilares (Siemiatycki e Thomas, 1981; Thompson, 1991), mesmo que nenhum viés esteja presente. Em resposta a essa limitação, vários autores definem interações biológicas dentro do contexto de um modelo causal geral, de modo que esse modelo não dependa de qualquer modelo mecanicista específico para o processo patológico. Descrevemos tais definições. A primeira definição, baseada no modelo causal *potencial-desfecho* ou *contrafactual* descrito no Capítulo 4, tem uma longa história na farmacologia (pelo menos desde 1920), e é chamada, algumas vezes, definição de interação dependente de ação. A segunda definição, baseada no modelo de *causa suficiente* descrito no Capítulo 2, tem sido mais comum em epidemiologia. Após fornecer essas definições, descreveremos como elas estão relacionadas logicamente uma à outra.

## Desfechos potenciais para duas variáveis

Considere o seguinte exemplo: suponha que desejamos estudar os efeitos de duas variáveis fixas  $X$  e  $Z$  sobre a mortalidade  $D$  em 10 anos, em uma coorte fechada. Se  $X$  e  $Z$  são indicadores binários, há

quatro combinações de exposição possíveis que cada pessoa na coorte poderia ter:  $X = Z = 0$ ,  $X = 1$  e  $Z = 0$ ,  $X = 0$  e  $Z = 1$ , ou  $X = Z = 1$ . Além disso, cada pessoa tem um dos dois desfechos possíveis para cada uma das quatro combinações: elas sobrevivem aos 10 anos ( $D = 0$ ) ou não sobrevivem ( $D = 1$ ). Isso significa que há  $2 \cdot 2 \cdot 2 \cdot 2 = 2^4 = 16$  tipos possíveis de pessoas na coorte, de acordo como a pessoa responderia a cada uma das quatro combinações de exposição.

Esses 16 tipos de pessoa são mostrados na Tabela 5.2. As colunas 2 a 5 da tabela mostram o desfecho ( $Y = 1$ , se doença se desenvolve,  $Y = 0$ , se doença não se desenvolve) para o tipo de pessoa na linha, sob a combinação de exposições mostrada no cabeçalho da coluna. Para cada tipo, podemos definir o risco sob cada combinação de  $X$  e  $Z$ , como o desfecho  $Y$  sob aquela combinação. Assim, para um dado tipo de resposta,  $R_{11}$  é 1 ou 0, conforme se  $Y$  é 1 ou 0, quando  $X = 1$  e  $Z = 1$ , e assim por diante para as outras combinações de  $X$  e  $Z$ . Podemos definir, então, várias diferenças de riscos para cada tipo. Por exemplo,  $R_{11} - R_{01}$  e  $R_{10} - R_{00}$  dão os efeitos de mudar de  $X = 0$  para  $X = 1$ , e  $R_{11} - R_{10}$  e  $R_{01} - R_{00}$  dão os efeitos de mudar de  $Z = 0$  para  $Z = 1$ . Essas diferenças podem ser 1, 0, ou  $-1$ , que correspondem a efeito causal, a nenhum efeito e a efeito preventivo da mudança.

Também podemos definir a diferença entre essas diferenças de riscos. Um fato útil é que a diferença das diferenças de riscos para mudar  $X$  é igual à diferença das diferenças de riscos para mudar  $Z$ :

$$(R_{11} - R_{01}) - (R_{10} - R_{00}) = (R_{11} - R_{10}) - (R_{01} - R_{00}) \tag{5.10}$$

**TABELA 5.2**

**Tipos de resposta possíveis (desfechos potenciais) para duas variáveis binárias de exposição  $X$  e  $Z$  e uma variável binária de desfecho  $Y$**

Tipo	Desfecho (risco) $Y$ quando a combinação de exposição é				Contraste de interação (diferença em diferenças de riscos) $IC = R_{11} - R_{10} - R_{01} + R_{00}$ e descrição do tipo causal
	$X = 1$ $Z = 1$	$X = 0$ $Z = 1$	$X = 1$ $Z = 0$	$X = 0$ $Z = 0$	
1	1	1	1	1	0 nenhum efeito (condenada)
2*	1	1	1	0	$-1$ causação isolada mais conjunta por $X = 1$ e $Z = 1$
3*	1	1	0	1	1 $Z = 1$ bloqueia efeito de $X = 1$ (antagonismo preventivo)
4	1	1	0	0	0 $X = 1$ ineficaz, $Z = 1$ causal
5*	1	0	1	1	1 $X = 1$ bloqueia efeito de $Z = 1$ (antagonismo preventivo)
6	1	0	1	0	0 $X = 1$ causal, $Z = 1$ ineficaz
7*	1	0	0	1	2 bloqueio mútuo (antagonismo preventivo)
8*	1	0	0	0	1 $X = 1$ mais $Z = 1$ causal (sinergismo causal)
9*	0	1	1	1	$-1$ $X = 1$ mais $Z = 1$ preventivo (sinergismo preventivo)
10*	0	1	1	0	$-2$ bloqueio mútuo (antagonismo causal)
11	0	1	0	1	0 $X = 1$ preventivo, $Z = 1$ ineficaz
12*	0	1	0	0	$-1$ $X = 1$ bloqueia efeito de $Z = 1$ (antagonismo causal)
13	0	0	1	1	0 $X = 1$ ineficaz, $Z = 1$ preventivo
14*	0	0	1	0	$-1$ $Z = 1$ bloqueia efeito de $X = 1$ (antagonismo causal)
15*	0	0	0	1	1 prevenção isolada mais conjunta por $X = 1$ e $Z = 1$
16	0	0	0	0	0 nenhum efeito (imune)

\* Definido como tipo de resposta de interação na presente discussão (tipos com um contraste de interação que não zero).

Essa equação nos diz que a mudança no efeito de  $X$ , quando nos movemos ao longo das categorias de  $Z$ , é o mesmo que a mudança no efeito de  $Z$ , quando nos movemos ao longo das categorias de  $X$ . A equação se mantém para cada tipo de resposta. A partir daqui, chamaremos a diferença das diferenças de riscos na equação 5.10 de *contraste de interação*, ou IC (*interaction contrast*).

Observe primeiramente que a equação 5.10 e, portanto, o contraste de interação, é igual a  $R_{11} - R_{10} - R_{01} + R_{00}$ . A coluna final da Tabela 5.2 fornece esse contraste de interação para cada tipo de resposta, junto com frases descrevendo o processo causal, levando ao desfecho (doença ou sem doença) em cada tipo de pessoa. Para seis tipos – 1, 4, 6, 11, 13, 16 – pelo menos um fator nunca teve um efeito, e assim não pode haver interação, porque ambos os fatores devem ter um efeito para que haja uma interação. O contraste de interação é igual a 0 para esses seis tipos. Os outros 10 tipos (marcados com um asterisco) podem ser vistos como exibindo algum tipo de interação (ou interdependência) dos efeitos dos dois fatores ( $X$  e  $Z$ ); para esses 10 tipos, o contraste de interação não é 0.

O aspecto definidor desses 10 tipos de interação é que não podemos dizer qual será o efeito de  $X$  (causar, prevenir ou não ter efeito sobre a doença), a menos que saibamos qual é o valor daquela pessoa para  $Z$  (e, inversamente, não podemos saber o efeito de  $Z$  sem saber o valor de  $X$  daquela pessoa). Em outras palavras, para um tipo de interação, o efeito de um fator depende do estado da pessoa para o outro fator. Uma descrição igualmente apta é dizer que cada fator modifica o efeito do outro. Infelizmente, o termo *modificação de efeito* tem sido usado, frequentemente, como uma contração de *modificação de medida de efeito*, que mostramos ser equivalente à interação estatística e que é dependente de escala, em contraste com os 10 tipos de interação apresentados na Tabela 5.2.

Alguns dos tipos de resposta na Tabela 5.2 são facilmente reconhecidos como interações. Para o tipo 8, cada fator causa a doença se, e somente se, o outro fator estiver presente; assim, ambos os fatores precisam estar presentes para que a doença ocorra. Por isso, diz-se que esse tipo representa efeitos sinérgicos. Para o tipo 10, cada fator causa a doença se o outro fator é ausente, mas previne a doença se o outro fator está presente. Por isso, diz-se que esse tipo representa efeitos mutuamente antagônicos.

Outros tipos de interação nem sempre são reconhecidos como exibindo efeitos interdependentes. Por exemplo, o tipo 2 tem sido descrito simplesmente como um tipo para o qual ambos os fatores podem ter um efeito (Miettinen, 1982b). Observe, contudo, que a presença de ambos os fatores pode levar a uma interação competitiva: para uma pessoa tipo 2, cada fator causará doença quando o outro estiver ausente, mas nenhum dos fatores pode ter um efeito sobre o desfecho em estudo ( $Y = 0$  ou  $1$ ), uma vez que o outro esteja presente. Assim, cada fator afeta o desfecho em estudo somente na ausência do outro, e, assim, pode ser dito que os dois fatores interagem antagonicamente para esse desfecho (Greenland e Poole, 1988).

## Relação de distribuições de tipo de resposta a riscos médios

Uma coorte com mais do que poucas pessoas é, inevitavelmente, uma mistura de tipos de resposta diferentes. Para examinar coortes, voltaremos a usar  $R_{11}$ ,  $R_{10}$ ,  $R_{01}$ ,  $R_{00}$ , para denotar os riscos médios (proporções e incidência) em uma coorte; esses riscos representam médias dos desfechos (riscos) dos tipos de resposta na população em discussão. Pode-se pensar sobre os riscos mostrados na Tabela 5.2 como casos especiais, em que a coorte só tem um membro.

Para calcular os riscos médios, deixemos  $p_k$  ser a proporção de pessoas tipo  $k$  na coorte ( $k = 1, \dots, 16$ ). Um aspecto útil da Tabela 5.2 é que podemos computar o risco médio da coorte sob qualquer uma das quatro combinações listadas de exposição à  $X$  e à  $Z$ , pela adição da  $p_k$  para a qual há um “1” na coluna de interesse. Assim, obtemos as seguintes fórmulas gerais:

$$\begin{aligned}
R_{11} &= \text{risco médio se } X \text{ e } Z \text{ são } 1 \\
&= p_1 + p_2 + p_3 + p_4 + p_5 + p_6 + p_7 + p_8 \\
R_{01} &= \text{risco médio se } X \text{ é } 0 \text{ e } Z \text{ é } 1 \\
&= p_1 + p_2 + p_3 + p_4 + p_9 + p_{10} + p_{11} + p_{12} \\
R_{10} &= \text{risco médio se } X \text{ é } 1 \text{ e } Z \text{ é } 0 \\
&= p_1 + p_2 + p_5 + p_6 + p_9 + p_{10} + p_{13} + p_{14} \\
R_{00} &= \text{risco médio se } X \text{ e } Z \text{ são } 0 \\
&= p_1 + p_3 + p_5 + p_7 + p_9 + p_{11} + p_{13} + p_{15}
\end{aligned}$$

Para uma coorte na qual nenhum dos 10 tipos de interação está presente, a relação de risco aditivo (equação 5.2) emerge entre os riscos médios (proporções de incidência), que seriam observados sob diferentes padrões de exposição (Greenland e Poole, 1988). Sem interação, somente  $p_1$ ,  $p_4$ ,  $p_6$ ,  $p_{11}$ ,  $p_{13}$ , e  $p_{16}$  são diferentes de zero. Nessa situação, as proporções de incidência sob os quatro padrões de exposição serão como a seguir:

$$\begin{aligned}
R_{11} &= \text{risco médio se } X \text{ e } Z \text{ são } 1 = p_1 + p_4 + p_6 \\
R_{01} &= \text{risco médio se } X \text{ é } 0 \text{ e } Z \text{ é } 1 = p_1 + p_4 + p_{11} \\
R_{10} &= \text{risco médio se } X \text{ é } 1 \text{ e } Z \text{ é } 0 = p_1 + p_6 + p_{13} \\
R_{00} &= \text{risco médio se } X \text{ e } Z \text{ são } 0 = p_1 + p_{11} + p_{13}
\end{aligned}$$

Então, as diferenças de riscos separadas para os efeitos de  $X = 1$  isoladamente e  $Z = 1$  isoladamente (em relação a  $X = Z = 0$ ) somam-se à diferença de riscos para o efeito de  $X = 1$  e  $Z = 1$  juntos:

$$R_{11} - R_{00} = p_4 + p_6 - (p_{11} + p_{13})$$

Rearrmando o lado direito da equação, temos

$$R_{11} - R_{00} = (p_6 - p_{13}) + (p_4 - p_{11})$$

Somando-se  $p_{13}$  ao termo entre parênteses da esquerda e subtraindo-se a mesma do termo da direita, e subtraindo-se  $p_{11}$  do termo entre parênteses da esquerda e somando-se a mesma ao da direita, obtemos

$$R_{11} - R_{00} = (p_6 + p_{13} - p_{11} - p_{13}) + (p_4 + p_{11} - p_{11} - p_{13}) \quad [5.11]$$

Substituindo das definições de proporções de incidência sem interação, temos

$$R_{11} - R_{00} = (R_{10} - R_{00}) + (R_{01} - R_{00})$$

Esta equação é idêntica à equação 5.2, e assim é equivalente à equação 5.1, que corresponde a nenhuma modificação das diferenças de riscos. Há uma diferença crucial na interpretação, todavia: a equação 5.2 é *descritiva* das diferenças em risco entre coortes de estudo *diferentes*; em contraste, a equação 5.10 é uma relação *causal* entre riscos, porque se refere a riscos que seriam observados na *mesma* coorte em estudo, sob categorias de exposição diferentes. A mesma coorte não pode ser observada sob categorias de exposição diferentes, assim devemos usar a equação descritiva 5.2 como um substituto para a equação causal 5.11. Esse uso requer ausência de confundimento ou padronização dos riscos para ajustar para o confundimento. O restante da presente discussão será concernente apenas à equação causal 5.11, e, portanto, não envolve preocupação com relação a confundimento ou a outro viés. A discussão também se aplica a situações envolvendo a equação 5.2, nas quais o viés esteja ausente, ou tenha sido completamente controlado (p. ex., todo confundimento tenha sido removido por meio de padronização).

Quatro pontos importantes merecem destaque. Primeiro, a álgebra precedente mostra que afastamentos da aditividade causal (equação 5.11) só podem ocorrer se tipos de interação causais estiverem presentes na coorte. Assim, a *observação* de não aditividade de diferenças de riscos (afastamentos da equação 5.2) implicará a presença de tipos de interação em uma coorte, contanto que as relações descritivas observadas sem viés representem as relações causais na coorte. Segundo, tipos de interação podem estar presentes e, ainda assim, as relações aditivas (equações 5.11 e 5.2) podem se manter. Essa circunstância pode ocorrer porque tipos de interação diferentes podem contrabalançar os efeitos, um do outro, sobre o risco médio. Por exemplo, suponha que, além dos tipos sem interação, houvesse pessoas tipo 2 e tipo 8 em proporções exatamente iguais ( $p_2 = p_8 > 0$ ). Então

$$R_{11} - R_{00} = p_2 + p_4 + p_6 + p_8 - (p_{11} + p_{13})$$

Rearrmando, a adição de  $p_{13}$  ao termo contido no parênteses da esquerda e sua subtração do contido no parênteses da direita, e a adição de  $p_{11}$  ao termo contido no parênteses da direita e sua subtração do termo contido no parênteses da esquerda, temos

$$\begin{aligned} R_{11} - R_{00} &= (p_8 + p_6 + p_{13} - p_{11} - p_{13}) + (p_2 + p_4 + p_{11} - p_{11} - p_{13}) \\ &= (R_{10} - R_{00}) + (R_{01} - R_{00}) \end{aligned}$$

Podemos resumir esses dois pontos como a seguir: afastamentos da aditividade implicam a presença de tipos de interação, mas a atividade não implica a ausência de tipos de interação.

O terceiro ponto é que o afastamento da aditividade de risco implica a presença de tipos de interação, quer estejamos estudando fatores causais, quer preventivos (Greenland e Poole, 1988). Para constatar isso, observe que os argumentos precedentes não fazem suposições sobre a ausência de tipos causais (tipos 4 e 6, na ausência de interação) ou de tipos preventivos (tipos 11 e 13, na ausência de interação). Levanta-se esse ponto em contraste a tratamentos anteriores, nos quais as interações preventivas tinham que ser estudadas usando-se modelos multiplicativos (Rothman, 1974; Walter e Holford, 1978).

O quarto ponto é que as definições de tipos de resposta (e daí as interações) dadas anteriormente são específicas para o desfecho particular em estudo. Se, em nosso exemplo, mudarmos para mortalidade em 5 anos, é possível que muitas pessoas que faleceriam dentro de 10 anos sob alguma combinação de exposições (e, assim, estariam entre os tipos 1 até 15 na Tabela 5.2) não morressem dentro de 5 anos. Por exemplo, uma pessoa que fosse tipo 8, considerando-se uma mortalidade em 10 anos, poderia ser tipo 16, ao se considerar uma mortalidade de 5 anos. De modo semelhante, é possível que uma pessoa que morreria dentro de 10 anos, se, e somente se, ela fosse exposta a um dos fatores, falecesse dentro de 5 anos se, e apenas se, fosse exposta a ambos os fatores. Tal pessoa seria um tipo 2 (ação competitiva) para a mortalidade em 10 anos, mas um tipo 8 (ação sinérgica) para a mortalidade em 5 anos. Para evitar a dependência do tipo de resposta sobre o tempo de seguimento, pode-se basear as definições de tipo de resposta no tempo de incidência, e não no risco (Greenland, 1993b).

## Relação de distribuições de tipo de resposta à aditividade

O contraste de interação  $IC = R_{11} - R_{10} - R_{01} + R_{00}$  corresponde ao afastamento da diferença de riscos, contrastando  $X = 1$  e  $Z = 1$  com  $X = 0$  e  $Z = 0$ , do que seria esperado se nenhum tipo de interação estivesse presente (i. e., se a diferença de riscos para  $X = Z = 1$  *versus*  $X = Z = 0$  fosse apenas a soma da diferença de riscos para  $X = 1$  *versus*  $X = 0$  e da diferença de riscos para  $Z = 1$  *versus*  $Z = 0$ ). Em termos algébricos, temos

$$IC = (R_{11} - R_{00}) - (R_{10} - R_{00}) - (R_{01} - R_{00}) \quad [5.12]$$

Substituindo-se os riscos, nesta fórmula, pelas proporções de tipos de resposta, e simplificando-se, obtemos

$$IC = (p_3 + p_5 + 2p_7 + p_8 + p_{15}) - (p_2 + p_9 + 2p_{10} + p_{12} + p_{14}) \quad [5.13]$$

Assim, o IC é composto de proporções de todos os 10 tipos de interação e será zero se nenhum tipo de interação estiver presente. As proporções de tipo 7 e 10 têm um peso duas vezes maior do que as proporções dos outros tipos de interação, porque correspondem aos tipos para os quais os efeitos de  $X$  reverterem ao longo dos estratos de  $Z$ . A equação 5.13 ilustra os dois primeiros pontos supracitados: o afastamento da aditividade ( $IC \neq 0$ ) implica a presença de tipos de interação, porque  $IC \neq 0$  requer que alguns tipos de interação estejam presentes; mas a aditividade ( $IC = 0$ ) não implica a ausência de tipos de interação, porque o  $IC$  pode ser zero, mesmo quando algumas proporções dentro dele não são zero. Esse fenômeno ocorre quando contribuições negativas de alguns tipos de interação ao  $IC$  se equilibram com as contribuições positivas de outros tipos de interação.

Os afastamentos da aditividade podem ser separados em duas classes. A *superaditividade* (também chamada de *transaditividade*) é definida como um afastamento “positivo”, o que, para riscos, corresponde a  $IC > 0$ , ou

$$R_{11} - R_{00} > (R_{10} - R_{00}) + (R_{01} - R_{00})$$

*Subaditividade* é um afastamento “negativo”, o que, para riscos, corresponde a  $IC < 0$ , ou

$$R_{11} - R_{00} < (R_{10} - R_{00}) + (R_{01} - R_{00})$$

Os afastamentos da aditividade de risco têm implicações especiais quando podemos presumir que nenhum dos dois fatores jamais é preventivo (nenhum dos fatores será preventivo na presença *ou* na ausência do outro, o que exclui os tipos 3, 5, 7, e de 9 a 15, inclusive). Sob essa suposição, o contraste de interação simplifica-se para

$$IC = p_8 - p_2$$

A superaditividade ( $IC > 0$ ) mais nenhuma prevenção implica, então, que  $p_8 > p_2$ . Visto que  $p_2 \geq 0$ , a superaditividade mais nenhuma prevenção implica que respostas sinérgicas (pessoas tipo 8) devem estar presentes ( $p_8 > 0$ ). O inverso, contudo, é falso; a presença de respostas sinérgicas não implica superaditividade, porque podemos ter  $p_2 > p_8 > 0$ , caso em que a subaditividade se manteria. Subaditividade mais nenhuma prevenção implica que  $p_8 < p_2$ . Como  $p_8 \geq 0$ , a subaditividade ( $IC < 0$ ) mais nenhuma prevenção implica que respostas competitivas (pessoas tipo 2) devem estar presentes ( $p_2 > 0$ ). No entanto, o inverso novamente é falso: a presença de respostas competitivas não implica subaditividade, porque podemos ter  $p_8 > p_2 > 0$ , caso em que a superaditividade seria mantida.

## A impossibilidade da identificação de tipos de resposta de interação

Os dados epidemiológicos sobre riscos ou taxas, mesmo que perfeitamente válidos, não podem, isoladamente, determinar os tipos de resposta em particular que estão presentes ou ausentes. Em especial, nunca se pode inferir que um tipo particular de interação na Tabela 5.2 esteja ausente, e a inferência de presença deve tornar não testáveis as suposições sobre ausência de outros tipos de resposta. Como um resultado, inferências sobre a presença de tipos de resposta específicos devem depender de suposições muito restritivas sobre ausência de outros tipos de resposta.

Não se pode inferir a presença de um tipo de resposta em particular, mesmo quando interações estatísticas *qualitativas* estão presentes entre as medidas de efeito reais, isto é, quando o efeito real de um fator reverte inteiramente a direção com a mudança dos níveis do outro fator. Tais reversões podem se originar a partir de combinações inteiramente distintas de tipos de interação. A interação qualitativa demonstra apenas que tipos de interação devem estar presentes.

**TABELA 5.3**

**Exemplo de duas coortes com proporções diferentes de tipos de resposta que produzem o mesmo contraste de interação**

Coorte nº 1						Coorte nº 2					
Tipo de resposta	Proporção	$R_{11}$	$R_{10}$	$R_{01}$	$R_{00}$	Tipo de resposta	Proporção	$R_{11}$	$R_{10}$	$R_{01}$	$R_{00}$
1	0,1	0,1	0,1	0,1	0,1	1	0,1	0,1	0,1	0,1	0,1
7	0	—	—	—	—	7	0,2	0,2	—	—	0,2
8	0,2	0,2	—	—	—	8	0	—	—	—	—
15	0,2	—	—	—	0,2	15	0	—	—	—	—
16	0,5	—	—	—	—	16	0,7	—	—	—	—
Total	1,0	0,3	0,1	0,1	0,3	Total	1,0	0,3	0,1	0,1	0,3

Considere o exemplo das duas coortes mostrado na Tabela 5.3, para o qual as proporções dos tipos de resposta são diferentes. Em ambas as coortes, os riscos em várias combinações de  $X$  e  $Z$  são idênticos, e assim também o são as medidas de efeito. Por exemplo, a diferença de riscos para  $X$  quando  $Z = 1$  ( $R_{11} - R_{01}$ ) é igual a 0,2 e quando  $Z = 0$  ( $R_{10} - R_{00}$ ) é igual a  $-0,2$ , uma interação estatística qualitativa. Assim, essas duas coortes completamente diferentes produzem contrastes de interação idênticos ( $IC = 0,4$ ). Na primeira coorte, os dois tipos de interação são aqueles para os quais  $X$  só tem um efeito na presença de  $Z$ , e esse efeito é causal (tipo 8), e aqueles para os quais  $X$  só tem um efeito na ausência de  $Z$ , e esse efeito é preventivo (tipo 15). Na segunda coorte, o único tipo de interação presente é aquele no qual o efeito de  $X$  é causal quando  $Z$  está presente, e preventivo quando  $Z$  está ausente (tipo 7). Em outras palavras, mesmo que víssemos os efeitos reais, livres de qualquer viés ou erro, não poderíamos distinguir se a interação estatística qualitativa surgiu porque pessoas diferentes são afetadas por  $X$  em estratos de  $Z$  diferentes ( $p_8 = p_{15} = 0,2$ ,  $p_{16} = 0,5$ ) ou porque as mesmas pessoas são afetadas, mas nestes indivíduos os efeitos de  $X$  se invertem ao longo dos estratos de  $Z$  ( $p_7 = 0,2$ ,  $p_{16} = 0,7$ ).

### Interações sob o modelo de causa suficiente

No Capítulo 2, definimos que a interação biológica entre duas ou mais causas componentes significa que as causas participam na mesma causa suficiente. Aqui, uma causa componente para um indivíduo é idêntica a um fator de risco causal, ou categoria da variável, cuja ocorrência contribui para completar uma causa suficiente. Mecanismos causais diferentes correspondem a causas suficientes de doença diferentes. Se duas causas componentes atuam para produzir doença em uma causa suficiente comum, alguns casos de doença podem surgir, nos quais as duas causas componentes compartilham a responsabilidade causal. Na ausência de qualquer das componentes, esses casos não ocorreriam. Sob o modelo de causa suficiente, essa coparticipação em uma causa suficiente é definida como interação sinérgica entre as componentes, ação conjunta causal ou sinergia.

Também pode haver mecanismos que requeiram ausência de um fator e presença do outro, para produzir doença. Esses mecanismos correspondem a uma causa suficiente na qual a ausência de um fator e a presença de outro fator são as causas componentes. A falta de ocorrência de doença, porque ambos os fatores estavam presentes, pode ser definida como uma interação antagonônica entre as componentes, ou antagonismo.



Se dois fatores nunca participam em conjunto na mesma causa suficiente por sinergia ou por antagonismo, então nenhum caso da doença pode ser atribuído a sua ação conjunta. A ausência de interação biológica, ou a independência de efeitos de dois fatores, significa, então, que nenhum caso da doença foi causado, ou prevenido, pela presença conjunta dos fatores.

Enfatizamos que duas causas componentes podem participar no mesmo mecanismo causal, sem atuar ao mesmo tempo. Expandindo um exemplo do Capítulo 2, contrair uma infecção viral pode causar um distúrbio permanente do equilíbrio em uma pessoa. Anos mais tarde, durante um tempo gélido, a pessoa pode escorregar e fraturar o quadril enquanto estava caminhando ao longo de uma trilha, porque o distúrbio tornou o equilíbrio mais difícil. A infecção viral de anos antes interagiu com o tempo gélido (e com a escolha do tipo de sapato, com a falta de um corrimão, etc.) para causar a fratura de quadril. Tanto a infecção viral quanto o tempo gélido são causas componentes no mesmo mecanismo causal, apesar de suas ações estarem separadas por muitos anos.










Dissemos que dois fatores podem “interagir” competindo para causar doença, mesmo que nem eles nem sua ausência compartilhem uma causa suficiente, porque somente uma causa suficiente completa é necessária para a doença ocorrer, e, assim, todas as causas suficientes competem para causar a doença. Considere causas de morte: dirigir sem cintos de segurança pode ser uma causa componente de uma lesão fatal (a primeira causa suficiente completa), que previne a morte por todas as outras causas suficientes (tal como câncer de pulmão fatal) e suas componentes (tal como fumo). Assim, dirigir sem cintos de segurança previne mortes por tabagismo, porque mata algumas pessoas que, caso contrário, continuariam vivas para morrer de doença relacionada com o fumo.

## Relação entre o desfecho-potencial e os modelos de interação de causa suficiente

Há uma conexão lógica e direta entre as duas definições de interação biológica discutidas até aqui, a qual pode ser explorada para fornecer uma ligação entre o modelo de causa suficiente (Cap. 2) e medidas de incidência (Greenland e Poole, 1988). Para construir essa conexão, a Figura 5.1 exhibe as nove causas suficientes possíveis quando conseguimos distinguir somente duas variáveis binárias,  $X$  e  $Z$ . O  $U_k$  em cada círculo representa todas as causas componentes (outras que não  $X = 1$  ou  $X = 0$ , e  $Z = 1$  ou  $Z = 0$ ) que são necessárias para completar a causa suficiente. Dizemos que uma pessoa está em risco de, ou suscetível à, causa suficiente  $k$  ( $k = A, B, C, D, E, F, G, H, I$ ) se  $U_k$  está presente para aquela pessoa, isto é, se a causa suficiente  $k$  está completa, exceto por alguma contribuição necessária de  $X$  ou  $Z$ . Note que uma pessoa pode estar em risco de nenhuma, uma ou várias causas suficientes. Dos nove tipos de causas suficientes na Figura 5.1, quatro ( $F, G, H, I$ ) são exemplos de ação conjunta causal (interação biológica no sentido de causa suficiente).

Podemos deduzir o tipo de resposta causal de qualquer indivíduo, dada a sua situação, quanto a causas suficientes. Em outras palavras, podemos deduzir à qual linha da Tabela 5.2 um indivíduo pertence, se soubermos as causas suficientes para as quais está em risco. Por exemplo, qualquer pessoa em risco para a causa suficiente  $A$  está condenada à doença, não importando a presença de  $X$  ou  $Z$ , de modo que a pessoa é de resposta tipo 1 na Tabela 5.2. Também, uma pessoa em risco para as causas suficientes  $B$  e  $C$ , mas para nenhuma outra, adquirirá a doença, a menos que  $X = Z = 0$ , portanto ela é de resposta tipo 2. De modo similar, uma pessoa em risco para as causas suficientes  $F$ ,  $G$  e  $H$ , mas para nenhuma outra, também contrairá a doença, a menos que  $X = Z = 0$ , então deve ser de resposta tipo 2.

Várias outras combinações de causas suficientes gerarão uma pessoa de resposta tipo 2. Em geral, combinações completamente diferentes de suscetibilidades a causas suficientes podem produzir o mesmo tipo de resposta, de modo que o modelo de causa suficiente é mais “refinado” ou mais detalhado do que o modelo desfecho-potencial (tipo de resposta) dos efeitos das mesmas variáveis (Greenland e Poole, 1988; Greenland e Brumback, 2002; VanderWeele e Hernán, 2006;

Tipo de causa suficiente		Descrição
A		X e Z irrelevantes
B		X = 1 necessário, Z irrelevante
C		Z = 1 necessário, X irrelevante
D		X = 0 necessário, Z irrelevante
E		Z = 0 necessário, X irrelevante
F		X = 1 e Z = 1 necessários
G		X = 1 e Z = 0 necessários
H		X = 0 e Z = 1 necessários
I		X = 0 e Z = 0 necessários

U = todos os outros componentes da causa suficiente

**FIGURA 5.1** • Enumeração dos nove tipos de causas suficientes para duas variáveis de exposição dicotômicas

VanderWeele e Robins, 2007a). Em outras palavras, para cada tipo de resposta em um modelo de desfecho-potencial, podemos construir pelo menos um e, frequentemente, vários modelos de causa suficiente que produzem o mesmo tipo de resposta. No entanto, há uns poucos tipos de resposta que correspondem a uma única causa suficiente. Um exemplo é o tipo de resposta sinérgica (tipo 8 na Tab. 5.2), para o qual a doença resulta se, e somente se,  $X = 1$  e  $Z = 1$ . O padrão de suscetibilidade que resulta em tal resposta sinérgica é um no qual a pessoa está em risco apenas para a causa suficiente  $F$ . A causa suficiente  $F$  corresponde exatamente à causação sinérgica, ou ação conjunta causal, de  $X = 1$  e  $Z = 1$  no modelo desfecho-potencial. Assim, a presença de respondedores sinérgicos (tipo 8 na Tab. 5.2) corresponde à presença de ação sinérgica (causa  $F$  na Fig. 5.1).

VanderWeele e Robins (2007a) mostraram que a presença de resposta de interação tipo 7, 8, 10, 12, 14 ou 15 implica a presença de ação conjunta causal, isto é, a presença de uma causa suficiente da forma  $F$ ,  $G$ ,  $H$  ou  $I$  (que eles tomam como sua definição de interação biológica). Em contraste, os outros quatro tipos de resposta acima definidos como interações (2, 3, 5, e 9) não implicam ação conjunta causal, isto é, a resposta de tipos 2, 3, 5 e 9 pode ocorrer, mesmo que nenhuma ação conjunta causal esteja presente. Por essa razão, VanderWeele e Robins (2007a) definem apenas os tipos 7, 8, 10, 12, 14 e 15 como refletindo ação interdependente, a fim de induzir uma correspondência com ação conjunta no modelo de causa suficiente. Os quatro tipos que eles excluem (tipos 2, 3, 5 e 9) são aqueles para os quais a doença ocorre sob 3 das 4 combinações de valores possíveis de  $X$  e  $Z$ .

Conforme mostrado anteriormente, podemos inferir que tipos de resposta sinérgica estão presentes a partir de superaditividade das diferenças de riscos causal, se presumirmos que nenhum dos fatores jamais é preventivo. Visto que nenhuma ação preventiva significa que nem  $X = 0$ , nem  $Z = 0$  agem em uma causa suficiente, podemos inferir a presença de ação sinérgica (causa suficiente  $F$ ), a partir de superaditividade, se presumirmos que as causas suficientes  $D$ ,  $E$ ,  $G$ ,  $H$  e  $I$  estão ausentes (há as causas suficientes que contêm  $X = 0$  ou  $Z = 0$ ). Sem presumir nenhuma ação preventiva, VanderWeele e Robins (2007a) mostraram que se  $R_{11} - R_{01} - R_{10} > 0$  (uma condição mais forte do que a superaditividade), então a causa suficiente  $F$  deve estar presente, isto é, deve haver sinergia entre  $X = 1$  e  $Z = 1$ . Os autores também dão condições análogas para se inferir a presença das causas suficientes  $G$ ,  $H$  e  $I$ .

A análise de interação é descrita adiante, no Capítulo 16.

## Interação biológica versus estatística

Alguns autores têm argumentado que os fatores que atuam em estágios diferentes de um modelo com estágios múltiplos são exemplos de ações interdependentes com efeito multiplicativo (Siemiatycki e Thomas, 1981). Pelas definições que usamos, entretanto, ações em estágios diferentes de um modelo com múltiplos estágios estão interagindo uma com a outra, justamente como a infecção viral e a pista escorregadia interagiram no exemplo, para produzir um quadril fraturado. Assim, não chamaríamos tais ações de independentes. Além do mais, não consideramos que a aditividade de diferença de riscos seja uma relação natural entre efeitos que ocorrem. Embora a ausência completa de interações implique em aditividade de risco, raramente esperaríamos observar aditividade de diferença de riscos, porque raramente esperaríamos que fatores agissem independentemente em todas as pessoas.

De modo mais geral, reiteramos que a interação estatística – modificação de medida de efeito – não deve ser confundida com interação biológica. Mais importante, quando dois fatores têm efeitos, a homogeneidade da razão de riscos, embora mal interpretada, frequentemente, como indicativa da ausência de interação biológica, implica justamente o oposto, isto é, a *presença* de interações biológicas. Essa conclusão segue-se porque, conforme mostrado anteriormente, a homogeneidade de uma medida de razão implica heterogeneidade (e, por conseguinte, não aditividade) da medida de diferença correspondente. Essa não aditividade, por sua vez, implica a presença de algum tipo de interação biológica.

## INTERAÇÕES EM SAÚDE PÚBLICA

Presumindo-se que custos ou benefícios de exposições ou de intervenções são mensurados pelo excesso ou pela redução da carga de casos que eles produzem, vários autores têm proposto que os afastamentos de aditividade da carga de casos (números incidentes), ou incidências, correspondam à interação da saúde pública (Blot e Day, 1979; Rothman e colaboradores, 1980; Saracci, 1980). A lógica é que, se as cargas de casos excessivas produzidas por cada fator não são aditivas, deve-se sa-

ber as categorias de todos os fatores, a fim de prever o impacto em saúde pública da remoção, ou da introdução, de cada uma delas (Hoffman et al., 2006).

Como um exemplo, podemos retornar à interação entre fumo e exposição a asbestos, examinada no começo do capítulo. Lembre que no exemplo hipotético o risco médio de mortalidade em 10 anos, em uma coorte de fumantes expostos a asbestos foi 0,020, mas teria sido 0,003, se todos os membros da coorte deixassem de fumar no início do seguimento, teria sido 0,010, se apenas a exposição a asbestos tivesse sido prevenida e teria declinado para 0,001 se todos deixassem de fumar e a exposição a asbestos tivesse sido prevenida. Esses efeitos não são aditivos, porque

$$\begin{aligned} R_{11} - R_{00} &= 0,020 - 0,001 = 0,019 > (R_{10} - R_{00}) + (R_{01} - R_{00}) \\ &= (0,003 - 0,001) + (0,010 - 0,001) = 0,011 \end{aligned}$$

Se houvesse 10.000 trabalhadores expostos, a prevenção da exposição a asbestos teria reduzido a carga de casos de  $(0,020)10.000 = 200$  para  $(0,010)10.000 = 100$ , se os hábitos de fumar não mudassem, mas teria reduzido a carga de casos de  $0,003(10.000) = 30$  para  $0,001(10.000) = 10$ , se todos também deixassem de fumar no início do seguimento. Assim, o benefício de prevenir a exposição a asbestos (em termos de redução da mortalidade) teria sido cinco vezes maior se ninguém deixasse de fumar do que se todos o fizessem. Somente se as diferenças de riscos fossem aditivas, a redução da mortalidade seria a mesma, independentemente do fumo. Afora isso, os hábitos de fumar da coorte não podem ser ignorados quando se estima o benefício de prevenir a exposição a asbestos. Conforme discutido no Capítulo 2, a remoção completa da exposição geralmente não é factível, mas o mesmo ponto se aplica à remoção parcial da exposição. O benefício da remoção parcial de um fator pode ser muito sensível à distribuição de outros fatores entre aqueles nos quais o fator é removido, bem como ser sensível aos meios de remoção.

Se os benefícios de saúde pública não são mensurados usando-se a redução da carga de casos, mas, em vez disso, são avaliados pelo uso de outra medida de benefício (p. ex., pela expectativa de ganho de anos de vida ou pela redução do custo da assistência à saúde), então a interação de saúde pública corresponderia à não aditividade por aquela medida, e não à carga de casos ou a diferenças de riscos. O conceito geral é que as interações de saúde pública correspondem a uma situação na qual os custos ou benefícios de se alterar um fator devem levar em conta a prevalência de outros fatores. Posto que a presença e a extensão das interações em saúde pública podem variar com a medida do benefício, o conceito é paralelo, do ponto de vista algébrico, a certos tipos de interação estatística, ou modificação de medida de efeito, e, assim, métodos estatísticos para estudar o último fenômeno também podem ser utilizados para estudar interação em saúde pública. O estudo da interação em saúde pública difere, contudo, porque a escolha da medida é ditada pelo contexto de saúde pública, e não por conveniência estatística ou por suposições biológicas.



## Seção II

# **Delineamento e condução do estudo**





# Tipos de estudos epidemiológicos

Kenneth J. Rothman, Sander Greenland e Timothy L. Lash

## Estudos experimentais 108

- Randomização 108
- Validade versus considerações éticas em experimentos em sujeitos humanos 109
- Ensaios clínicos 110
- Ensaios de campo 112
- Intervenção comunitária e ensaios randomizados por *cluster* 113

## Estudos não experimentais 114

- Estudos de coorte 116
- Estudos de caso-controle 117
- Estudos prospectivos *versus* estudos retrospectivos 117
- Estudos transversais 119
- Estudos de mortalidade proporcional 119
- Estudos ecológicos 121
- Geração de hipótese *versus* rastreamento de hipótese 121

Os delineamentos epidemiológicos compreendem ensaios experimentais e não experimentais. O experimento é um exemplo emblemático da atividade científica. Mas o que constitui um experimento? Na linguagem comum, um experimento refere-se a qualquer ensaio ou teste. Por exemplo, um professor pode introduzir novos métodos de ensino como um experimento. Para muitos cientistas, porém, o termo tem um significado mais específico: um experimento é um conjunto de observações, conduzidas sob circunstâncias controladas, no qual o cientista manipula as condições para averiguar que efeito, se é que há algum, tal manipulação tem sobre as observações. Alguns poderiam ampliar essa definição, para incluir observações controladas sem manipulação das condições. Assim, as observações astronômicas durante o eclipse solar de 1919, as quais corroboraram a teoria geral da relatividade de Einstein, têm sido referidas, frequentemente, como um experimento. Para os epidemiologistas, entretanto, a palavra *experimento* geralmente implica que o pesquisador manipula a exposição atribuída aos participantes do estudo. Portanto, a epidemiologia experimental é limitada, por definição, a tópicos para os quais a categoria de exposição pode ser manipulada. Visto que os sujeitos dessas manipulações são humanos, a epidemiologia experimental é mais limitada ainda, eticamente, a estudos nos quais se espera que as designações para exposição não causem dano.

Quando os experimentos epidemiológicos satisfazem condições mínimas de viabilidade e de ética, seu delineamento objetiva reduzir a variação do desfecho do estudo que poderia ser atribuída a outros fatores (fatores externos à associação) e levar em consideração, de forma acurada, a variação remanescente. Geralmente, há duas ou mais formas de intervenção. As designações para intervenção são determinadas pelo pesquisador, comumente, pela aplicação de um esquema randomizado de alocação. O propósito da alocação aleatória é criar grupos que diferenciem-se apenas aleatoriamente no tempo da alocação com relação à ocorrência subsequente do desfecho do estudo. Os experimentos epidemiológicos incluem ensaios clínicos (com pacientes como sujeitos), ensaios de campo (com intervenções designadas para membros individuais da comunidade) e ensaios de intervenção comunitária (com intervenções designadas para comunidades inteiras).

Quando os experimentos são impraticáveis, ou antiéticos, os epidemiologistas delineiam estudos não experimentais (também conhecidos como observacionais), em uma tentativa de simular o que poderia ter sido aprendido se um experimento fosse conduzido. Em estudos não experimentais, o pesquisador é um observador, e não um agente que designa intervenções. Os quatro tipos principais de estudos epidemiológicos não experimentais são: estudos de coortes, nos quais todos os sujeitos em uma população-fonte são classificados de acordo com o seu *status* de exposição e são acompanhados durante um tempo para verificação da incidência de doença; *status* de caso-controle, nos quais a amostra e os casos provenientes de uma população-fonte são classificados conforme seu estado de exposição; estudos transversais, inclusive estudos de prevalência, nos quais se averigua o *status* quanto à exposição e à doença em um tempo particular; e estudos ecológicos, nos quais as unidades de observação são grupos de pessoas.

## ESTUDOS EXPERIMENTAIS

Um experimento típico em sujeitos humanos cria grupos experimentais que são expostos a diferentes tratamentos ou agentes. Em um experimento simples de dois grupos, um grupo recebe um tratamento e o outro não. Idealmente, os grupos experimentais são idênticos com relação a fatores externos à associação que afetam o desfecho de interesse, de modo que, se o tratamento não tivesse efeito, desfechos idênticos seriam observados entre os grupos. Esse objetivo poderia ser atingido se fosse possível controlar todas as condições relevantes que pudessem afetar o desfecho sob estudo. Nas ciências biológicas, entretanto, as condições que afetam a maior parte dos desfechos são tão complexas e extensas e, em geral, desconhecidas e assim não podem ser tornadas uniformes. Daí, haverá variação no desfecho, mesmo na ausência de um efeito do tratamento. Essa “variação biológica” reflete variação no conjunto de condições que produz o efeito.

Assim, na experimentação biológica, não é possível criar grupos por meio dos quais somente o tratamento em estudo varia. Em vez disso, o pesquisador pode decidir criar grupos nos quais se espera que o efeito dos fatores externos à associação seja pequeno. Por exemplo, pode ser impossível fazer com que todos os animais em um experimento comam exatamente a mesma quantidade de alimento; contudo, se essa variação puder ser mantida pequena, pouco contribuirá para a variação no desfecho entre os grupos.

Geralmente, o pesquisador ficaria satisfeito se o efeito de fatores externos entre os grupos fosse substancialmente menor do que o efeito esperado do tratamento em estudo. Frequentemente, porém, nem mesmo isso pode ser alcançado. Em tal caso, o experimento deve ser delineado de modo que a variação do desfecho, devido a fatores externos, possa ser mensurada acuradamente, e assim levada em conta nas comparações entre os grupos de tratamento.

## Randomização

No começo do século XX, R. A. Fisher e outros desenvolveram uma base prática para delineamentos experimentais que leva em conta, com precisão, a variabilidade externa entre as unidades experimentais (sejam elas objetos, animais, pessoas ou comunidades). Essa base é chamada de *randomização* (alocação aleatória) de tratamentos, ou exposições, entre as unidades: cada unidade é designada para um tratamento usando-se um mecanismo aleatório de alocação, tal como jogar uma moeda para o alto. Tal mecanismo não tem relação com os fatores estranhos que afetam o desfecho, de modo que qualquer associação entre a alocação de tratamento que ele produza e aqueles fatores estranhos será aleatória. A variação no desfecho ao longo dos grupos de tratamento, que não seja resultante de efeitos do tratamento, pode, assim, ser atribuída a essas associações aleatórias, e por isso ser chamada, justificavelmente, de variação aleatória.

Uma hipótese sobre o tamanho do efeito do tratamento, tal como a hipótese nula, corresponde a uma distribuição de probabilidade específica para os potenciais desfechos sob aquela hipótese. Essa distribuição de probabilidade pode ser comparada com a associação observada entre tratamento e re-

sultados. A comparação liga estatística e inferência, o que explica por que muitos métodos estatísticos, tais como análise de variância, estimam a variação aleatória de desfecho dentro e entre os grupos. Um estudo com distribuição aleatória do tratamento permite que se calcule a probabilidade da associação observada sob várias hipóteses de como a alocação do tratamento afeta o desfecho. Em particular, se a alocação é aleatória, e não tem efeito sobre o desfecho, exceto através de tratamento, qualquer variação sistemática (não aleatória) do desfecho com a alocação deve ser atribuída a um efeito do tratamento.

Os cientistas conduziram experimentos durante séculos, antes que a ideia de alocação desfecho se cristalizasse, e os experimentos que têm pouca variação de desfecho por efeito externo (como frequentemente ocorre nas ciências físicas) não necessitam do método. Entretanto, alguns cientistas sociais e epidemiologistas identificam o termo *experimento* apenas com um ensaio randomizado. Algumas vezes, o termo quase-experimento é usado para se referir a estudos controlados, nos quais a exposição foi designada pelo pesquisador sem utilizar randomização (Cook e Campbell, 1979).

### **Validade versus considerações éticas em experimentos em sujeitos humanos**

Em um experimento, aqueles que são designados a um tratamento experimental só estão expostos porque o pesquisador alocou o sujeito à exposição. Em um experimento puramente científico, a razão para designar a exposição específica para um sujeito em particular é apenas para maximizar a validade do estudo. Os passos considerados necessários para se atingir essa meta ocorrem, geralmente, de acordo com um protocolo de estudo. O único motivo para a alocação é manter a conformidade com o protocolo e não satisfazer as necessidades do sujeito.

Por exemplo, suponha que uma médica tratando cefaleia tenha prescrito um medicamento patenteado para seus pacientes abastados, e um equivalente genérico para seus pacientes indigentes, porque a confiabilidade presumivelmente maior da versão patenteada, no julgamento dela, não compensaria o custo maior para aqueles com recursos modestos. Se a médica quisesse comparar os efeitos dos dois medicamentos entre seus pacientes, ela não estaria conduzindo um experimento válido, apesar de ela própria ter alocado as exposições. Posto que a alocação foi baseada, em parte, em fatores que poderiam influenciar o desfecho, tal como a riqueza, seria esperado haver diferenças entre os grupos de tratamento, mesmo que os medicamentos tivessem o mesmo efeito sobre o desfecho, isto é, seria esperado haver confundimento (ver Cap. 4). Para conduzir um experimento válido, a médica teria que alocar as drogas conforme um protocolo que não levasse a um desequilíbrio sistemático de causas externas de cefaleia entre os grupos de tratamento. A alocação da exposição em experimentos é delineada para ajudar o estudo, e não o sujeito individual. Se ela é feita para ajudar o sujeito, poderia ser feito um estudo não experimental, mas ele não seria considerado um experimento, por causa do confundimento que o critério de designação do tratamento poderia induzir.

Visto que as metas da pesquisa, e não as necessidades do sujeito, determinam a alocação da exposição, restrições éticas limitam seriamente as circunstâncias nas quais experimentos válidos em sujeitos humanos são factíveis. Os experimentos em sujeitos humanos somente são eticamente permissíveis quando a adesão ao protocolo científico não conflitar com os melhores interesses do sujeito. Especificamente, deve haver garantia razoável de que não existe uma maneira conhecida e viável pela qual um sujeito participante possa ser tratado melhor do que as possibilidades de tratamento que o protocolo fornece. Desse requisito, vem a restrição de que quaisquer exposições ou tratamentos dados a sujeitos devem ser limitados a preventivos potenciais da doença. Essa limitação, sozinha, classifica a maioria da pesquisa etiológica na variedade não experimental.

Entre as implicações éticas mais específicas está a de que os sujeitos admitidos no estudo não devem, por causa disso, ser privados de alguma forma preferível de tratamento, ou de prevenção, que não esteja incluída no estudo. Esse requisito implica que a melhor terapia disponível deve ser incluída para fornecer uma referência (comparação) para qualquer tratamento novo. Outro requisito ético,

conhecido como *equipotência*, declara que as possibilidades de tratamento incluídas no ensaio devem ser igualmente aceitáveis, dado o conhecimento atual. A equipotência restringe severamente o uso de placebos: a Declaração de Helsinque diz que é antiético incluir uma terapia com placebo como uma das pontas de um ensaio clínico, se já existe um remédio ou um preventivo aceito para o desfecho (Associação Médica Mundial, [www.wma.net/e/policy/b3.htm](http://www.wma.net/e/policy/b3.htm); Rothman e Michels, 2002).

Mesmo com essas limitações, muitos experimentos epidemiológicos são conduzidos (alguns deles, infelizmente, ignoram princípios éticos, tais como a equipotência). A maioria são ensaios clínicos, que são estudos epidemiológicos avaliando tratamentos para pacientes que já adquiriram a doença (*ensaio* é usado como um sinônimo de *experimento*). Experimentos epidemiológicos que visam a avaliar preventivos primários (agentes com os quais se pretende primariamente prevenir a instalação da doença) são menos comuns do que os ensaios clínicos; esses estudos são ensaios de campo ou ensaios de intervenção comunitária.

## Ensaios clínicos

Um ensaio clínico é um experimento com pacientes humanos como sujeitos. A meta da maioria dos ensaios clínicos é avaliar a cura potencial de uma doença ou encontrar uma prevenção de complicações de doença, tais como morte, incapacidade ou declínio da qualidade de vida. As exposições em tais ensaios não são preventivos primários, porque não previnem a ocorrência da doença ou a condição inicial, mas são preventivos dela. Por exemplo, uma modificação de dieta depois que um indivíduo sofre um infarto do miocárdio pode prevenir um segundo infarto e a morte subsequente, agentes quimioterápicos dados a pacientes com câncer podem prevenir a recorrência do câncer, e drogas imunossupressoras dadas a pacientes com transplantes podem prevenir a rejeição.

Sujeitos em ensaios clínicos de prevenção de complicações devem ser diagnosticados como tendo a doença em questão e ser admitidos no estudo cedo o bastante após o diagnóstico, para permitir que a alocação ocorra de modo apropriado. Sujeitos cuja enfermidade seja leve ou grave demais para permitir a forma de tratamento, ou a terapia alternativa, sendo estudada, devem ser excluídos. A alocação ao tratamento deve ser designada de tal forma a minimizar diferenças entre os grupos de tratamento com relação a fatores externos que possam afetar a comparação. Por exemplo, se alguns médicos participantes do estudo fossem favoráveis à nova terapia, é concebível que eles pudessem influenciar a alocação ao novo tratamento de, por exemplo, seus próprios pacientes ou, talvez, dos pacientes mais seriamente afetados. Se os pacientes mais seriamente acometidos tendessem a receber o novo tratamento, haveria confundimento (ver Cap. 4), e a validade da avaliação do novo tratamento ficaria comprometida.

Para evitar isso e problemas correlatos, é desejável alocar os tratamentos em ensaios clínicos, de modo a permitir que se leve em conta possíveis diferenças entre os grupos de tratamento com relação a características “basais” não mensuradas. Como parte dessa meta, o mecanismo de alocação deve impedir manipulação de alocação não definida pelo protocolo. É quase universalmente aceito que a randomização é a melhor maneira de lidar com preocupações sobre confundimento por características basais não mensuradas e por manipulação pela equipe na alocação do tratamento (Byar et al., 1976; Peto et al., 1976; Gelman et al., 2003). A validade do ensaio depende fortemente da extensão em que o protocolo de alocação aleatória é o único determinante dos tratamentos recebidos. Quando essa condição é satisfeita, o confundimento devido a fatores não mensurados pode ser considerado aleatório e ser tratado com procedimentos estatísticos padronizados, e sua magnitude diminui à medida que aumenta o número de randomizados (Greenland e Robins, 1986; Greenland, 1990). Quando a condição não é satisfeita, contudo, confundidores não mensurados podem atrapalhar a estatística, exatamente como nos estudos observacionais. Mesmo quando a condição é satisfeita, a possibilidade de generalização dos resultados do ensaio pode ser afetada por arrolamento seletivo. Os participantes do ensaio, frequentemente, não refletem a distribuição por sexo, por idade, por raça e por etnia da população-alvo (Murthy et al, 2004; Heiat et al., 2002). Por motivos explicados no Capítulo 8, as populações representativas de um estudo raramente são cientificamente ótimas. Entretanto, quando a eficácia do tratamento é modificada por sexo, idade, raça,

etnia ou por outros fatores, e a população do estudo difere da população que estaria recebendo o tratamento com relação a essas variáveis, o efeito médio do estudo será diferente do efeito médio entre aqueles que receberiam o tratamento. Nessas circunstâncias, a extrapolação dos resultados do estudo é tênue ou injustificada, e pode ser necessário restringir as inferências a subgrupos específicos se o tamanho de tais subgrupos o permitir.

Dado que o tratamento depende da alocação aleatória, e não da tomada de decisão de paciente e de médico quanto ao tratamento, o arrolamento de pacientes em um ensaio requer que seu consentimento seja informado. No mínimo, o consentimento informado exige que os pacientes compreendam (a) que eles estão participando de um trabalho de pesquisa com uma duração declarada, (b) o propósito da pesquisa, os procedimentos que serão seguidos e quais procedimentos são experimentais, (c) que sua participação é voluntária e que podem se retirar em qualquer momento e (d) os riscos e benefícios potenciais associados a sua participação.

Embora os métodos de randomização, frequentemente, aloquem os sujeitos a tratamentos em proporções aproximadamente iguais, essa equivalência nem sempre é ótima. A verdadeira equipotência provê uma lógica para proporções de alocação iguais, porém, frequentemente, há uma hipótese de que um tratamento é mais efetivo, com base em uma lógica biológica, em estudos anteriores ou até em dados preliminares do mesmo estudo. Em tais circunstâncias, probabilidades iguais de alocação podem ser uma barreira ao arrolamento. A randomização adaptativa (Armitage, 1985), ou alocação desequilibrada (Avins, 1998), permite que mais sujeitos no ensaio recebam o tratamento que se espera ser mais efetivo, com pouca redução na potência.

Sempre que possível, os ensaios clínicos devem tentar empregar o método “cego” na alocação ao tratamento. Idealmente, o indivíduo que faz a alocação, o paciente e o avaliador do desfecho devem ignorar qual foi o tratamento designado. O método “cego” previne certos vieses que poderiam afetar a alocação, a avaliação ou a adesão. O mais importante é manter o avaliador “cego”, especialmente se a avaliação do desfecho é subjetiva, como acontece com um diagnóstico clínico. (Alguns desfechos, tais como a morte, serão relativamente insuscetíveis a viés na avaliação.) O conhecimento do paciente sobre a designação do tratamento pode afetar a adesão ao regime terapêutico e pode enviesar percepções de sintomas que podem afetar a avaliação do desfecho. Os estudos, nos quais tanto o avaliador quanto o paciente são “cegos” na alocação do tratamento, são conhecidos como *estudos duplo-cegos*. Um estudo no qual o indivíduo que faz a alocação desconhece qual tratamento é qual (tal como pode acontecer se os tratamentos são pílulas codificadas, e quem aloca não sabe o código) pode ser descrito como *triplo-cego*, embora esse termo seja usado frequentemente para implicar que o analista dos dados (em adição ao paciente e ao avaliador) não sabe qual grupo de pacientes sob análise recebeu qual tratamento.

Dependendo da natureza da intervenção, pode não ser possível, ou prático, evitar o conhecimento da alocação de todos esses participantes. Por exemplo, um tratamento pode ter efeitos colaterais bem conhecidos que possibilitem aos pacientes identificá-lo. O pesquisador precisa estar ciente disso e relatar essas possibilidades, de modo que os leitores possam avaliar se toda, ou parte, de qualquer associação relatada pode ser atribuível à falta do método “cego”.

Se não há tratamento aceito para a condição sendo estudada, pode ser útil empregar um placebo como o tratamento de comparação, quando restrições éticas não impedirem. *Placebos* são tratamentos inertes dos quais não se pretende outro efeito que não o benefício psicológico de receber um tratamento, o que, por si, pode ter um efeito poderoso. Esse benefício psicológico é denominado de *efeito placebo*, mesmo que ocorra em pacientes recebendo tratamento ativo. Pelo emprego de um placebo, um pesquisador pode ser capaz de realizar o controle do componente psicológico de receber um tratamento e de estudar os benefícios não psicológicos de uma nova intervenção. Além disso, empregar um placebo facilita o cegamento, quando não há outro tratamento em comparação. Tais benefícios podem ser incompletos, entretanto, se efeitos colaterais visíveis do tratamento ativo ampliarem a resposta placebo (o componente psicológico do tratamento) entre aqueles recebendo o tratamento ativo. Placebos não são necessários quando o objetivo do ensaio é apenas comparar tratamentos diferentes, um com o outro. No entanto,

mesmo sem placebos, deve-se estar alerta para a possibilidade de um efeito placebo, ou de diferenças de adesão, devido a diferenças em efeitos colaterais visíveis entre os tratamentos ativos que são alocados.

A não adesão ou não concordância ao tratamento alocado resulta em uma discrepância entre o tratamento designado e o tratamento realmente recebido pelos participantes do ensaio. A prática padrão baseia todas as comparações na alocação do tratamento e não no tratamento recebido. Essa prática é chamada de princípio da intenção-de-tratar, porque a análise se baseia no tratamento pretendido, não no tratamento recebido. Embora esse princípio ajude a preservar a validade dos testes para efeitos de tratamento, tende a produzir viés nas estimativas de tais efeitos; por isso, alternativas têm sido desenvolvidas (Goetghebuer et al., 1998). A adesão pode ser mensurada, algumas vezes, pela indagação direta de sujeitos sobre sua concordância, pela obtenção de dados relevantes (p. ex., pedir que os comprimidos não usados sejam devolvidos) ou por medidas bioquímicas. Essas medidas de adesão podem ser usadas para ajustar estimativas de efeitos de tratamento pelo uso de métodos especiais, nos quais a randomização desempenha o papel de uma *variável instrumental* (Sommer e Zeger, 1991; Angrist et al., 1996; Greenland, 2000b; Cap. 12).

A maioria dos ensaios é monitorada enquanto está sendo conduzida por um Comitê ou Comissão de Monitoração de Segurança e Dados (Data and Safety Monitoring Committee or Board-DSMB). O objetivo primário desses comitês é garantir a segurança dos participantes do ensaio (Wilhelmsen, 2002). O comitê revê resultados de estudos, inclusive estimativas dos efeitos principais do tratamento, bem como ocorrência de eventos adversos, para determinar se o ensaio deve ser interrompido antes do seu término programado. A lógica para a suspensão precoce pode ser (a) o aparecimento de um efeito favorecendo um tratamento tão fortemente, que deixaria de ser ético randomizar novos pacientes para o tratamento alternativo ou negar acesso dos pacientes inscritos ao tratamento favorecido, (b) a ocorrência de efeitos adversos em taxas consideradas inaceitáveis, dado o benefício esperado do tratamento ou os resultados do ensaio ou (c) a determinação de que os resultados razoavelmente esperados não sejam mais de valor suficiente para continuação do ensaio. As deliberações do DSMB envolvem a consideração de tópicos de medicina, de ética, de lei, de estatística e de custos, para a decisão sobre continuar um ensaio. Dada a complexidade dos assuntos, a composição do DSMB deve compreender um âmbito diversificado de treinamentos e experiências, e assim, frequentemente, inclui clínicos, estatísticos, estudiosos da ética, nenhum dos quais tendo interesse material no resultado do ensaio.

As regras estatísticas frequencistas comumente usadas por DSMB, para determinar se um ensaio deve ser suspenso, foram desenvolvidas para garantir que o acaso de um erro Tipo I (rejeição incorreta da hipótese nula principal de nenhum efeito do tratamento; ver Cap. 10) não excedesse a um nível pré-especificado (o nível alfa) durante as análises intermediárias planejadas (Armitage et al., 1969). Apesar dessas metas, os membros do DSMB podem interpretar mal resultados intermediários (George et al., 2004), e a adesão estrita a essas regras de suspensão pode originar resultados falsos (Wheatley e Clayton, 2003). Interromper um ensaio precocemente, por causa do aparecimento de um efeito favorável a um tratamento, resultará, frequentemente, em uma estimativa excessiva do benefício verdadeiro do tratamento (Pocock e Hughes, 1989). Além disso, ensaios suspensos precocemente podem não possibilitar seguimento suficiente para se observar eventos adversos associados ao tratamento favorecido (Cannistra, 2004), particularmente se tais eventos são complicações crônicas. Alternativas bayesianas têm sido sugeridas para melhorar muitas dessas deficiências (Berry, 1993; Carlin e Sargent, 1996).

## Ensaio de campo

Nos ensaios de campo, ao contrário dos ensaios clínicos, os sujeitos não são definidos pela presença de doença, ou pela apresentação para assistência clínica; em vez disso, o foco é sobre a ocorrência inicial de doença. Pacientes em um ensaio clínico podem enfrentar as complicações de sua doença, com alta probabilidade, durante um tempo relativamente curto. Em contraste, o risco de doença incidente entre sujeitos vivendo livremente é, em geral, muito mais baixo. Em consequência, os ensaios de campo, geralmente, requerem um número muito maior de sujeitos do que os ensaios clínicos e são muito mais caros. Além do mais, como os sujeitos não estão sob assistência ativa à saúde e, portanto,



não se deslocam para tratar-se, um ensaio de campo requer, frequentemente, visitação aos sujeitos no trabalho, na residência ou na escola, ou o estabelecimento de centros, a partir dos quais a pesquisa pode ser conduzida e aos quais se insiste que os sujeitos compareçam. Esses aspectos do delineamento dos ensaios de campo aumentam o seu custo.

O gasto com os ensaios de campo limita seu uso ao estudo da prevenção de doenças extremamente comuns ou extremamente graves. Vários ensaios de campo foram conduzidos para determinar a eficácia de doses altas de vitamina C na prevenção do resfriado comum (Karlowski et al., 1975; Dykes e Meier, 1975). A poliomielite paralítica, uma doença rara, mas séria, foi um problema de saúde pública suficiente para justificar o que pode ter sido o maior experimento com seres humanos já tentado, o ensaio da vacina Salk, no qual a vacina, ou um placebo, foi administrada em centenas de milhares de crianças em idade escolar (Francis et al., 1955). Quando o desfecho da doença ocorre raramente, é mais eficiente estudar sujeitos que se pensa estarem em maior risco. Assim, o ensaio da vacina contra hepatite B foi realizado em uma população de homossexuais do sexo masculino da cidade de Nova Iorque, entre os quais a infecção por essa doença ocorre com frequência muito maior do que a usual entre os demais moradores dessa metrópole (Szmunes, 1980). Similarmente, o efeito da cessação do uso de duchas vaginais sobre o risco de doença inflamatória foi estudado em mulheres com uma história recente de doença sexualmente transmitida, um fator de risco forte para doença inflamatória pélvica (Rothman et al., 2003).

Um raciocínio análogo frequentemente é aplicado ao delineamento de ensaios clínicos que podem se concentrar em pacientes com alto risco de desfechos adversos. Visto que os pacientes que já tiveram um infarto do miocárdio estão em alto risco de um segundo infarto, vários ensaios clínicos do efeito de baixar os níveis séricos de colesterol sobre o risco de infarto do miocárdio foram empreendidos em tais pacientes (Leren, 1966; Detre e Shaw, 1974). É muito mais oneroso conduzir um ensaio delineado para estudar o efeito da redução do colesterol sérico sobre a primeira ocorrência de um infarto do miocárdio, porque muito mais sujeitos precisam ser incluídos para fornecer um número razoável de eventos do desfecho em estudo. O Ensaio de Intervenções de Fatores de Risco Múltiplos (*Multiple Risk Factor Intervention Trial* – MRFIT) foi um ensaio de campo de vários preventivos primários do infarto do miocárdio, inclusive a dieta. Embora esse ensaio admitisse apenas indivíduos de alto risco, e se esforçasse para reduzir o risco por meio de várias intervenções simultâneas, envolveu 12.866 sujeitos e um custo de US\$115 milhões (mais de meio bilhão de dólares em 2006) (Kolata, 1982).

Como em ensaios clínicos, as exposições em ensaios de campo devem ser alocadas de acordo com um protocolo que minimize a variação externa entre os grupos, por exemplo, pela remoção de qualquer decisão sobre alocação por parte da equipe do estudo. Um esquema aleatório de alocação é, novamente, uma escolha ideal, mas as dificuldades de implantar tal esquema em um ensaio de campo em larga escala podem superar as vantagens. Por exemplo, pode ser conveniente distribuir vacinações a grupos em lotes que são manuseados de forma idêntica, especialmente se for difícil armazenar e transportar a vacina. Tais considerações práticas podem ditar o uso de protocolos modificados de randomização, tais como randomização por *cluster* (que será explicada adiante). Como tais modificações podem afetar seriamente a qualidade das informações e a interpretação dos achados experimentais, as vantagens e desvantagens precisam ser analisadas cuidadosamente.

## Intervenção comunitária e ensaios randomizados por *cluster*

O ensaio de intervenção comunitária é uma extensão do ensaio de campo, que envolve intervenção em uma base comunitária. Conceitualmente, a distinção depende de a intervenção ser implantada separadamente para cada indivíduo ou não. Por exemplo, enquanto uma vacina normalmente é administrada individualmente às pessoas, a fluoretação da água para prevenção de cáries dentárias é aplicada, geralmente, em suprimentos de água individuais. Consequentemente, a fluoretação da água foi avaliada por ensaios de intervenção comunitária, nos quais comunidades inteiras foram selecionadas, e a exposição (tratamento da água) foi alocada em base comunitária. Outros exemplos de intervenções que poderiam ser implantadas em uma base comunitária incluem os programas de ressuscitação de emergência com resposta rápida e programas educacionais



conduzidos por meio de mídia de massa, tais como o Projeto de Prevenção de Queimaduras em Massachusetts (MacKay e Rothman, 1982).

Algumas intervenções são implantadas mais convenientemente em grupos de sujeitos menores que em comunidades inteiras. Intervenções dietéticas podem ser feitas com mais praticidade na família ou no domicílio. Intervenções ambientais podem afetar um escritório inteiro, uma fábrica ou um prédio residencial. Equipamentos esportivos protetores podem ter que ser alocados a um time inteiro ou a uma liga. Grupos de intervenção podem ser unidades do exército, classes escolares, ocupantes de veículos ou qualquer outro grupo, cujos membros estejam expostos à intervenção simultaneamente. A base científica dos experimentos usando tais intervenções é idêntica àquela dos ensaios de intervenção comunitária. O que coloca todos esses estudos à parte dos ensaios de campo é o fato de que as intervenções são alocadas a grupos e não a indivíduos.

Diz-se dos ensaios de campo, nos quais o tratamento é alocado a grupos de participantes de forma aleatória, que são randomizados em *cluster*. Quanto maior o tamanho do grupo a ser randomizado, em relação ao tamanho total do estudo, menos é conseguido pela alocação aleatória. Se apenas duas comunidades estão envolvidas em um estudo, uma delas receberá a intervenção e a outra não, tal como no ensaio de fluoração da água de Newburgh-Kingston (Ast et al., 1956), não importa se a comunidade que recebe o flúor é designada aleatoriamente ou não. Diferenças nas características basais (externas) terão a mesma magnitude e o mesmo efeito, qualquer que seja o método de alocação; somente a direção das diferenças será afetada. É somente quando os números de grupos randomizados para cada intervenção são grandes, que a randomização provavelmente produzirá distribuições similares das características basais entre os grupos de intervenção. Assim, a análise de ensaios randomizados por *cluster* deve envolver métodos que levem em consideração a formação dos grupos (Omar e Thompson, 2000; Turner et al., 2001; Spiegelhalter, 2001), que são essenciais para se estimar adequadamente a quantidade de variabilidade introduzida pela randomização (dada uma hipótese sobre o tamanho dos efeitos do tratamento).

## ESTUDOS NÃO EXPERIMENTAIS

As limitações impostas pela ética e por custos restringem a maior parte da pesquisa epidemiológica a estudos não experimentais. Embora seja antiético para um pesquisador expor uma pessoa a uma causa potencial de doença, simplesmente para aprender sobre a etiologia da doença, frequentemente as pessoas se expõem, voluntária ou involuntariamente, a muitos fatores potencialmente nocivos. Considere o exemplo dos cigarros (MacMahon, 1979):

[As pessoas] escolhem uma faixa ampla de dosagens de uma variedade de substâncias potencialmente tóxicas. Considere o hábito de fumar cigarro, ao qual centenas de milhões de pessoas têm se exposto, em níveis variando de quase zero (para aqueles expostos exclusivamente por meio do fumo alheio) a três ou quatro cigarros por hora de vigília do dependente, e os consequentes dois milhões ou mais de mortes por câncer de pulmão no último século, somente neste país.

Além do tabaco, as pessoas nas nações industrializadas se expõem, entre outras coisas, a uma variedade de regimes de exercício, desde o sedentarismo até a exaustão, as dietas vegetarianas, até as derivadas quase totalmente de fontes animais, e a intervenções médicas para diversas situações. Cada uma dessas exposições pode ter consequências pretendidas e não pretendidas, que podem ser investigadas pela epidemiologia observacional.

Idealmente, desejaríamos que a força das evidências da pesquisa não experimental fosse tão grande como aquela passível de obtenção por um experimento bem delineado. Em um experimento, entretanto, o pesquisador tem o poder de alocar exposições de forma a ampliar a validade do estudo, ao passo que, na pesquisa não experimental, ele não pode controlar as circunstâncias da exposição. Se os expostos têm um risco maior ou menor para a doença do que os não expostos, uma comparação simples entre expostos e não expostos será confundida por essa diferença e,

assim, não refletirá de forma válida o efeito isolado da exposição. A comparação será confundida pelas diferenças externas em risco entre os grupos de exposição (i.e., diferenças que não são atribuíveis ao contraste da exposição em estudo).

A falta de randomização põe em questão a prática padrão de se analisar dados não experimentais com métodos estatísticos desenvolvidos para estudos randomizados. Sem randomização, a variação sistemática é um composto de todas as fontes não controladas de variação – inclusive qualquer efeito de tratamento – mas inclui, também, fatores de confundimento e outras fontes de erro sistemático. Como resultado, em estudos sem randomização, a variação sistemática estimada por métodos estatísticos padrão não é prontamente atribuível a efeitos de tratamento, nem pode ser confiavelmente comparada à variação que se espera ocorrer pelo acaso. A separação de efeitos de tratamento da mistura de variação sistemática não controlada, em estudos não randomizados (ou em estudos randomizados com falta de concordância), requer hipóteses adicionais sobre as fontes de erro sistemático. Em estudos não experimentais, essas hipóteses, geralmente, não são mais do que especulações, embora possam ser incorporadas à análise como distribuições anteriores em análise bayesiana, ou como ajustes de parâmetros em um viés de análise (Cap. 18 e 19). Nesse sentido, a inferência causal na ausência de randomização é altamente especulativa. A validade de tal inferência depende do quão bem as especulações sobre os efeitos de erros sistemáticos correspondam a seu efeito verdadeiro.

Visto que o pesquisador não pode designar a exposição em estudos não experimentais, deve basear-se fortemente na fonte primária de arbítrio que remanesce: a seleção de indivíduos. Se o paradigma da observação científica é o experimento, então o paradigma da pesquisa epidemiológica não experimental é o “experimento natural”, no qual a natureza imita o tipo de experimento que o pesquisador poderia ter conduzido, salvo por restrições éticas e de custos. De longe, o mais renomado é o elegante estudo do cólera em Londres, realizado por John Snow. Em Londres, em meados do século XIX, havia várias companhias que serviam água de beber encanada para os residentes, que frequentemente competiam, servindo clientela semelhante dentro dos distritos da cidade. Snow tirou vantagem desse experimento natural ao comparar as taxas de mortalidade por cólera em residentes assinantes de duas das maiores companhias de fornecimento de água: a Companhia Southwark e Vauxhall, que encanava água impura do rio Tâmisa, contaminada por esgotos, e a Companhia Lambeth, que em 1852 mudou seu ponto de coleta de Hungerford Market para o lado oposto do Tâmisa, em Thames Ditton, obtendo, assim, suprimento de água livre dos esgotos da cidade. Como Snow (1855) descreveu:

[...] a mistura do suprimento de água da Companhia Southwark e Vauxhall com o da Companhia Lambeth, numa parte extensa de Londres, permitiu que os sujeitos fossem separados de tal modo a gerar a prova mais incontestável, de um lado ou do outro. Nos subdistritos... supridos por ambas as companhias, a mistura dos suprimentos é do tipo mais íntimo. Os encanamentos de cada companhia passam por todas as ruas, e entram em quase todas as pracinhas e becos. Um poucas casas são supridas por uma companhia, e umas poucas pela outra, de acordo com a decisão do proprietário ou ocupante no tempo em que as Companhias de Água estavam em competição ativa. Em muitos casos, uma só casa tem um suprimento diferente daquelas em cada lado. Cada companhia fornece tanto para ricos como para pobres, tanto para casas grandes como para pequenas; não há diferença quer na condição, quer na ocupação das pessoas recebendo a água das diferentes companhias... é óbvio que nenhum experimento poderia ter sido concebido que testasse de maneira mais rigorosa o efeito do suprimento de água sobre a progressão do cólera do que esse.

O experimento, também, foi na mais larga escala. Não menos do que trezentas mil pessoas de ambos os sexos, de cada idade e ocupação, e de cada nível e estado social, desde nobres até os muito pobres, foram divididos em dois grupos sem sua opção, e, em muitos casos, sem seu conhecimento; um grupo sendo suprido por água contendo eflúvios da rede de esgoto de Londres, e entre esses o que quer que possa ter vindo dos pacientes com cólera, o outro grupo tendo água bastante livre de impurezas.

Para contabilizar esse experimento, tudo que era preciso era verificar o suprimento de água de cada casa individual onde um ataque fatal de cólera tivesse ocorrido[...]

Há dois tipos primários de estudos não experimentais em epidemiologia. O primeiro, o *estudo de coorte* (também chamado de estudo de seguimento ou estudo de incidência), é um estudo análogo direto do experimento. Grupos de exposição diferentes são comparados, mas (como no estudo de Snow) o pesquisador seleciona apenas os sujeitos a observar, e apenas classifica esses sujeitos pelo *status* de exposição, em vez de alocá-los a grupos de exposição. O segundo, o *estudo de caso-controle de incidência*, ou, simplesmente, o *estudo de caso-controle*, usa um passo extra de amostragem da população-fonte para os casos: enquanto um estudo de coorte incluiria todas as pessoas na população dando origem aos casos do estudo, um estudo de caso-controle seleciona somente uma amostra daquelas pessoas, e escolhe quem incluir, em parte com base em seu *status* de doença. Esse passo extra de amostragem pode tornar um estudo de caso-controle muito mais eficiente do que um estudo de coorte da mesma população, mas o primeiro introduz numerosas sutilezas e caminhos para viés, que estão ausentes nos estudos de coorte típicos.

Discussões mais detalhadas, tanto dos estudos de coorte como dos estudos de caso-controle e suas variantes, com exemplos específicos, são apresentadas nos Capítulos 7 e 8. Fornecemos aqui uma visão geral breve dos delineamentos.

## Estudos de coorte

No estudo de coorte paradigmático, o pesquisador define dois ou mais grupos de pessoas que estão livres de doença e que diferem de acordo com a extensão de sua exposição a uma causa potencial de doença. Esses grupos são referidos como as coortes do estudo. Quando dois grupos são estudados, pensa-se em um, usualmente, como a coorte exposta, ou índice – aqueles indivíduos que experimentaram a suposta exposição causal ou condição – e pensa-se no outro, então, como a coorte não exposta, ou de referência. Pode haver mais do que apenas duas coortes, mas cada coorte representaria um grupo com um nível, ou tipo, diferente de exposição. Por exemplo, um estudo de coorte ocupacional de trabalhadores químicos poderia compreender coortes de funcionários que trabalhassem em departamentos diferentes da fábrica, com cada coorte sendo exposta a um conjunto diferente de produtos químicos. O pesquisador mede os tempos de incidência e calcula as taxas de doença em cada uma das coortes do estudo e compara essas medidas de ocorrência.

No experimento natural de Snow, as coortes do estudo eram residentes de Londres que consumiam água da Companhia Lambeth ou da Companhia Southwark e Vauxhall, e que moravam em distritos onde havia tubulações de ambas as companhias de água. Snow foi capaz de estimar a frequência dos óbitos por cólera, usando domicílios como o denominador, separadamente para as pessoas em cada uma das coortes (Snow, 1855).

De acordo com uma resposta que foi feita ao Parlamento, a Companhia Southwark e Vauxhall supria 40.046 casas de 1<sup>o</sup> de janeiro a 31 de dezembro de 1853, e a Companhia Lambeth fornecia para 26.107 domicílios durante o mesmo período; conseqüentemente, como 286 ataques fatais de cólera aconteceram, nas primeiras quatro semanas da epidemia, em casas supridas pela primeira companhia, e somente 14 naquelas com abastecimento pela última, as proporções de ataques fatais para cada 10.000 residências foi a seguinte: Southwark e Vauxhall, 71, Lambeth, 5. Portanto, o cólera foi quatorze vezes mais fatal nesse período, entre as pessoas recebendo a água impura da Companhia Southwark e Vauxhall, do que entre aquelas tendo a água mais pura do Tâmesa, de Thames Ditton.

Muitos estudos de coorte começam com uma única coorte, que é heterogênea com relação à história de exposição. Comparações de experiência de doença são feitas dentro da coorte, através de subgrupos definidos por uma ou mais exposições. Os exemplos incluem estudos de coortes definidos a partir de listas de filiação a unidades administrativas ou sociais, tais como coortes de médicos ou de enfermeiras, ou coortes definidas por meio de registros de empregos, tais como coortes de operários de fábricas.

## Estudos de caso-controle

Os estudos de caso-controle são melhor compreendidos e conduzidos pela definição de uma população-fonte no início, o que representa uma população de estudo hipotética na qual um estudo de coorte poderia ter sido realizado. Se um estudo de coorte fosse feito, as tarefas primárias seriam identificar os expostos e a experiência dos não expostos do denominador, mensurados como unidades pessoa-tempo de experiência, ou como o número de pessoas em cada coorte do estudo, e então identificar o número de casos que ocorrem em cada categoria pessoa-tempo, ou em cada coorte do estudo. Em um estudo de caso-controle, esses mesmos casos são identificados, e seu *status* de exposição é determinado exatamente como em um estudo de coorte, mas os denominadores, dos quais as taxas poderiam ser calculadas, não são mensurados. Em vez disso, um grupo controle de sujeitos do estudo é tirado como uma amostra da população-fonte total, que deu origem aos casos.

O propósito desse grupo controle é determinar o tamanho relativo dos denominadores de expostos e não expostos na população-fonte. Exatamente como podemos tentar medir riscos ou taxas em uma coorte, os denominadores que a série de controle representa em um estudo de caso-controle podem refletir o número de pessoas nos subgrupos de expostos e de não expostos da população-fonte ou a quantidade de pessoa-tempo nesses mesmos subgrupos (Cap. 8). A partir do tamanho relativo desses denominadores, o tamanho relativo das taxas de incidência, ou as proporções de incidência, podem então ser estimados. Assim, os estudos de caso-controle geram estimativas diretas das medidas de efeito relativo. Visto que o grupo controle é usado para estimar a distribuição da exposição na população-fonte, o requisito principal da seleção de controles é que devem entrar na amostra, independentemente de seu *status* de exposição.

## Estudos prospectivos versus estudos retrospectivos

Os estudos ainda podem ser classificados como prospectivos ou retrospectivos, embora várias definições tenham sido usadas para esses termos. Os primeiros autores definiam estudos prospectivos e retrospectivos para denotar estudos de coorte e de caso-controle, respectivamente. Usar os termos *prospectivo* e *retrospectivo* dessa maneira não traz informações adicionais e deixa de destacar outros aspectos importantes de um estudo, para os quais a descrição como prospectivo ou retrospectivo poderia ser reveladora; portanto, um uso diferente se desenvolveu.

Um aspecto central do delineamento, que pode ser destacado pela distinção entre prospectivo e retrospectivo, é a ordem no tempo do registro de informações sobre exposição e a ocorrência de doença. Em alguns estudos, particularmente naqueles em que a exposição é medida perguntando-se às pessoas sobre sua história de exposição, é possível que a ocorrência de doença possa influenciar a recordação da exposição e introduzir viés nos resultados do estudo, por essa influência. Um estudo baseado em tal recordação merece o rótulo de retrospectivo, pelo menos no que diz respeito ao registro da informação de exposição, e talvez para o estudo como um todo. Avaliar a exposição por lembrança depois que a doença ocorreu é uma característica de muitos estudos de caso-controle, o que pode explicar por que, frequentemente, esses estudos são rotulados como retrospectivos. Um estudo com mensuração retrospectiva, nesse sentido, gera a preocupação de que a ocorrência da doença, ou de seu diagnóstico, tenha afetado a exposição da avaliação.

Entretanto, nem todos os estudos de caso-controle envolvem memória. Por exemplo, os estudos de caso-controle que avaliam exposição a drogas têm mensuração prospectiva, se a informação sobre exposições e outros fatores de risco é tirada de prontuários médicos, ou de registros de exposição que precedem ao desenvolvimento da doença. Esses estudos de caso-controle podem ser descritos mais apropriadamente como prospectivos, pelo menos com relação à mensuração da exposição.

Nem todas as variáveis de um estudo precisam ser mensuradas simultaneamente. Alguns estudos podem combinar mensuração prospectiva de algumas variáveis com mensuração retrospectiva de outras. Tais estudos podem ser vistos como sendo uma mistura de mensurações prospectivas e retrospectivas. Uma regra razoável pode ser descrever um estudo como prospectivo, se a mensuração

da exposição não puder ser influenciada pela doença, e retrospectivo, em caso contrário. Essa regra pode levar um estudo com uma mistura de variáveis mensuradas prospectiva e retrospectivamente a ser descrito de maneira diferente para análises distintas, o que é apropriado.

O acesso a dados pode afetar a validade de um estudo, tanto quanto o registro dos dados. A averiguação histórica tem implicações quanto a viés por seleção e por falta de dados, na medida em que registros ou dados podem estar faltando de modo sistemático. Por exemplo, preservar a informação sobre exposição que tenha sido registrada no passado (i.e., prospectivamente) pode depender da ocorrência da doença, como pode ser o caso se registros ocupacionais foram destruídos, exceto para os trabalhadores que apresentaram reclamações de incapacidade. Assim, informações registradas prospectivamente podem ter um componente retrospectivo para sua inclusão em um estudo, se esta depender da ocorrência de doença. Ao determinar se a informação em um estudo é obtida prospectiva ou retrospectivamente, a possibilidade de que a doença influencie o registro dos dados, ou sua via de ingresso no estudo, deve ser considerada.

Os termos *prospectivo* e *retrospectivo* também têm sido usados para se referir ao momento do pessoa-tempo acumulado com relação à condução do estudo. Nesse uso, quando pessoa-tempo se acumula antes da condução do estudo, diz-se que se trata de um estudo retrospectivo, mesmo que o *status* de exposição seja registrado antes da ocorrência da doença. Quando pessoa-tempo se acumula depois que o *status* começa, é dito que se trata de um estudo prospectivo; nessa situação, normalmente o *status* de exposição é registrado antes da ocorrência da doença, embora haja exceções. Por exemplo, o tipo de ocupação pode ser registrado para uma coorte ocupacional na abertura de um estudo, e quando os trabalhadores ingressam na coorte, mas um sanitarista industrial pode alocar categorias de exposição às categorias ocupacionais somente depois da conclusão do estudo e, portanto, depois que todos os casos de doença tenham ocorrido. Existe, então, o potencial para que a doença influencie a alocação realizada pelo sanitarista industrial.

Similarmente, outras nuances podem complicar a classificação de estudos como retrospectivos ou prospectivos, com relação à condução do estudo. Por exemplo, estudos de coorte podem ser conduzidos medindo-se os eventos de doença depois do início do estudo, pela definição das coortes como de algum tempo no passado e mensurando-se a ocorrência da doença no tempo antes do começo do estudo, ou uma combinação dos dois. De modo semelhante, os estudos de caso-controle podem basear-se em eventos de doença que ocorrem depois do início do estudo ou em eventos que tenham ocorrido antes que o estudo começasse, ou uma combinação. Assim, quer estudos de coorte, quer de caso-controle, podem averiguar eventos prospectiva ou retrospectivamente, do ponto de vista do tempo de começo do estudo. De acordo com esse uso, *prospectivo* e *retrospectivo* descrevem o tempo dos eventos sob estudo em relação ao tempo em que a pesquisa começa ou termina: *prospectivo* refere-se a eventos concorrentes ao estudo, e *retrospectivo* refere-se ao uso de eventos históricos.

Essas considerações demonstram que a classificação de estudos como prospectivos ou retrospectivos não é óbvia, e que esses termos não transmitem prontamente uma mensagem clara sobre o estudo. O aspecto mais importante do estudo que esses termos poderiam ajudar a esclarecer seria se a doença poderia influenciar as informações sobre exposição, e esse é o uso que recomendamos. Então, *prospectivo* e *retrospectivo* serão termos que poderiam descrever alguns estudos de coorte e de caso-controle. Sob as definições alternativas, estudos classificados como retrospectivos poderiam, na verdade, utilizar métodos que excluem a possibilidade de que a informação sobre exposição pudesse ter sido influenciada pela doença, e estudos classificados como prospectivos poderiam, realmente, usar métodos que não evitassem tal possibilidade. Visto que o termo *retrospectivo* frequentemente implica um delineamento inerentemente menos confiável, e que o termo *prospectivo* geralmente tem um significado oposto, a designação de classificação, sob as definições alternativas, nem sempre transmite com acurácia as forças ou as fraquezas do delineamento. O Capítulo 9 discute mais as vantagens e os defeitos de dados concorrentes e dados históricos e da mensuração prospectiva e retrospectiva.

## Estudos transversais

Um estudo que inclui como sujeitos todas as pessoas na população, ao tempo da averiguação, ou uma amostra representativa de todas essas pessoas, selecionadas sem levar em consideração o estado de exposição ou de doença, geralmente é referido como um *estudo transversal*. Um estudo transversal conduzido para estimar prevalência é denominado *estudo de prevalência*. Em geral, a exposição é averiguada simultaneamente com a doença, e subpopulações com exposições diferentes são comparadas em relação a sua prevalência de doença. Tais estudos não precisam ter objetivos etiológicos. Por exemplo, a prestação de serviços de saúde frequentemente requer conhecimento apenas de quantos itens serão necessários (tais como o número de leitos hospitalares), sem referência às causas da doença. No entanto, dados transversais são usados, frequentemente, para inferências etiológicas, para o que uma compreensão perfeita de suas limitações é essencial.

Um problema é que tais estudos frequentemente apresentam dificuldade em determinar o tempo de ordem dos eventos (Flanders et al., 1992). Outro problema, chamado de amostragem com viés de duração (Simon, 1980a), é que os casos identificados em um estudo transversal representarão em excesso os casos com longa duração e serão subrepresentativos daqueles com enfermidade de curta duração. Para entender, considere duas situações extremas envolvendo uma doença com uma duração altamente variável. Uma pessoa contraindo essa doença aos 20 anos e vivendo até os 70 anos pode ser incluída em qualquer estudo transversal durante 50 anos de doença. Uma pessoa contraindo a doença com a idade de 40 e falecendo um dia depois quase não tem chance de inclusão. Assim, se a exposição não altera o risco da doença, mas faz com que ela seja leve e prolongada quando contraída (de modo que a exposição tenha associação positiva com a duração), a prevalência de exposição será elevada entre os casos. Consequentemente, uma associação positiva exposição-doença será observada em um estudo transversal, mesmo que a exposição não tenha efeito sobre o risco de doença, e que seja benéfica se a doença ocorrer. Se a exposição não altera o risco de doença, mas faz com que a doença seja rapidamente fatal se contraída (de modo que a exposição seja negativamente associada à duração), então a prevalência de exposição será muito baixa entre os casos. Como resultado, a associação exposição-doença observada em um estudo transversal será negativa, embora a exposição não tenha efeito sobre o risco de doença e fosse prejudicial se a doença ocorresse. Há métodos analíticos para lidar com a relação potencial entre exposição e duração (p. ex., Simon, 1980a). Esses métodos requerem as datas de diagnóstico dos casos do estudo ou informações sobre a distribuição das durações da doença em estudo sobre níveis diferentes de exposição; tal informação pode estar disponível nas bases de dados médicos.

Os estudos transversais podem envolver a amostragem diferencial de sujeitos com relação ao *status* de doença, para aumentar o número de casos na amostra. Tais estudos às vezes são chamados de *estudos caso-controle de prevalência* porque seu delineamento é muito semelhante àquele dos estudos caso-controle de incidência, contudo, a série de casos compreende casos prevalentes, em vez de casos incidentes (Morgenstern e Thomas, 1993).

## Estudos de mortalidade proporcional

Um estudo de mortalidade proporcional inclui apenas sujeitos mortos. As proporções de sujeitos expostos falecidos designadas às causas-índice de morte são comparadas com as proporções de sujeitos não expostos falecidos designadas às causas-índice. A *razão de mortalidade proporcional* (proportional mortality ratio-PMR) resultante é a medida tradicional do efeito da exposição sobre as causas-índice de morte. Superficialmente, a comparação de proporções de sujeitos morrendo de uma causa específica, entre um grupo de expostos e um de não expostos, assemelha-se a um estudo de coorte mensurando incidência. A semelhança é enganosa, contudo, porque um estudo de mortalidade proporcional não envolve a identificação e o seguimento de coortes. Todos os sujeitos estão mortos no tempo de entrada no estudo.



O pressuposto de um estudo de mortalidade proporcional é que se a exposição causa (ou previne) uma doença fatal específica, deveria haver, proporcionalmente, mais (ou menos) mortes por aquela doença entre pessoas falecidas que tinham sido expostas, do que entre pessoas mortas que não tinham sido expostas. Esse raciocínio contém duas falhas importantes. A primeira delas diz respeito ao fato de que uma comparação de PMR não pode distinguir se a exposição aumenta a ocorrência de causas-índice de morte, previne a ocorrência de outras causas de morte ou produz alguma mistura desses efeitos (McDowall, 1983). Por exemplo, um estudo de mortalidade proporcional poderia encontrar um excesso proporcional de mortes por câncer entre grandes usuários de aspirina, em comparação com não usuários de aspirina, mas essa descoberta poderia ser atribuível a um efeito preventivo da aspirina sobre os óbitos cardiovasculares, que compreendem a grande maioria das mortes não causadas por câncer. Assim, um pressuposto implícito de um estudo de etiologia de mortalidade proporcional é que a taxa de mortalidade geral para outras categorias, que não a causa-índice, não esteja relacionada à exposição. A segunda maior falha em comparações de mortalidade é que elas não podem determinar a extensão em que a exposição origina as causas-índice de morte ou piora o prognóstico das doenças correspondentes às causas-índice. Por exemplo, uma associação do uso de aspirina com mortes por acidente vascular cerebral, entre todas as mortes, poderia ser consequência de um efeito da aspirina sobre a incidência de acidentes vasculares cerebrais, de um efeito da aspirina sobre a gravidade desses acidentes ou alguma combinação desses efeitos.

As ambiguidades na interpretação de uma PMR não são necessariamente uma falha fatal, porque tal medida frequentemente fornece entendimentos sobre relações causais que valem a pena ser perseguidos. Em muitas situações, só pode haver uma ou poucas causas estreitas de morte que sejam de interesse, e pode-se julgar implausível que uma exposição afete substancialmente o prognóstico, ou a ocorrência, de quaisquer óbitos não índice. Entretanto, muitas das dificuldades na interpretação de estudos de mortalidade proporcional podem ser mitigadas considerando-se tais estudos como uma variante do estudo de caso-controle. Proceder assim requer a conceituação de uma população combinada de indivíduos expostos e não expostos, nos quais os casos ocorreram. Os casos são aquelas mortes, tanto de expostos como de não expostos, na categoria-índice ou em outras categorias; os controles são os outros óbitos (Miettinen e Wang, 1981).

O princípio da seleção de séries de controle é escolher indivíduos que representem a população-fonte da qual os casos se originaram, para aprender a distribuição da exposição naquela população. Em vez de fazer a amostragem dos controles diretamente da população-fonte, podemos amostrar óbitos que nela ocorrem, desde que a distribuição da exposição entre as mortes amostradas seja a mesma da população-fonte; isto é, a exposição não deve estar relacionada a causas de óbito entre os controles (McLaughlin et al., 1985). Se mantivermos em mente os objetivos da seleção de controles, torna-se claro que não estamos presos a selecionar como controles todas as outras mortes que não os casos-índice. Em vez disso, podemos selecionar como controles um conjunto limitado de causas de morte de referência, com base na ausência presumida de associação com a exposição. Desse modo, outras causas de óbito, para as quais uma relação com a exposição é conhecida, suspeita ou meramente plausível, podem ser excluídas.

O princípio por trás da seleção das causas de controle de morte para inclusão no estudo é idêntico ao princípio de selecionar uma série de controle para qualquer estudo de caso-controle: as séries de controle devem ser selecionadas independentemente da exposição, com o objetivo de estimar a proporção da experiência da população-fonte que é exposta, como nos estudos de caso-controle de densidade (Cap. 8). Mortes por causas que não fazem parte da série controle podem ser excluídas do estudo ou podem ser estudadas como grupos de casos alternativos.

Tratar um estudo de mortalidade proporcional como um estudo de caso-controle pode, assim, aumentar sua validade. Isso também fornece a base para estimar as medidas epidemiológicas usuais de efeito, que podem ser derivadas de tais estudos (Wang e Miettinen, 1982). Por essas razões, os estudos de mortalidade proporcional são descritos e conduzidos de modo crescente como estudos de caso-controle. O mesmo tipo de delineamento e de análise tem reaparecido no contexto de verificar



eventos adversos relatados espontaneamente, em conexão com uso farmacêutico. A U.S. Food and Drug Administration mantém uma base de dados de relatos espontâneos, o Adverse Event Reporting System (AERS) (Rodriguez et al., 2001), que tem sido uma fonte de dados para estudos delineados para triar associações entre drogas e efeitos adversos não identificados previamente, usando técnicas de Bayes empíricas (Du Mouchel, 1999). Evans e colaboradores (2001) propuseram que esses dados devessem ser analisados do mesmo modo que os dados de mortalidade têm sido analisados em estudos de mortalidade proporcional, utilizando uma medida que eles chamaram de razão de relato proporcional, ou PRR, que seria análoga à PMR nos estudos de mortalidade proporcional. Essa abordagem, contudo, está sujeita aos mesmos problemas que acompanharam a PMR. Da mesma forma que com a PMR, esses problemas podem ser reduzidos pela aplicação dos princípios dos estudos de caso-controle à tarefa de vigilância dos dados de relato espontâneo (Rothman et al., 2004).

## Estudos ecológicos

Todos os tipos de estudos descritos até aqui compartilham a característica de que as observações feitas são pertinentes a indivíduos. É possível, e algumas vezes necessário, conduzir pesquisa na qual a unidade de observação é um grupo de pessoas, em vez de um indivíduo. Tais pesquisas são chamadas de *estudos ecológicos* ou *estudos agregados*. Os grupos podem ser classes de uma escola, fábricas, cidades, condados ou nações. O único requisito é que as informações sobre as populações estudadas estejam disponíveis para a mensuração das distribuições de exposição e de doença em cada grupo. Taxas de incidência ou de mortalidade são comumente usadas para quantificar a ocorrência de doenças em grupos. A exposição também é mensurada por um índice geral; por exemplo, o consumo de álcool no condado pode ser estimado a partir de dados do imposto sobre álcool, informações sobre *status* socioeconômico estão disponíveis nos tratados do censo do decênio, e dados ambientais (temperatura, qualidade do ar, etc.) podem ser obtidos local ou regionalmente. Esses dados ambientais são exemplos de exposições que são mensuradas por necessidade em um grupo, porque os dados em nível individual geralmente são indisponíveis e impraticáveis de coletar.

Quando a exposição varia entre indivíduos dentro dos grupos ecológicos, o grau de associação entre exposição e doença não precisa refletir associações em nível individual (Firebaugh, 1978; Morgenstern, 1982; Richardson et al., 1987; Piantadosi et al., 1988; Greenland e Robins, 1994; Greenland, 2001a, 2001b; Cap. 25). Além disso, o uso de medidas substitutas para exposição (p. ex., dados fiscais sobre o álcool, em vez de dados de consumo) e para doença (mortalidade em vez de incidência) distorce ainda mais as associações (Brenner et al., 1992b). Finalmente, os estudos ecológicos sofrem, geralmente, de indisponibilidade dos dados necessários para controle adequado do confundimento na análise (Greenland e Robins, 1994). Mesmo que o objetivo da pesquisa seja estimar efeitos de exposições em nível de grupo sobre resultados em nível de grupo, problemas de inadequação de dados, bem como agrupamento inapropriado, podem trazer sérios vieses às estimativas de estudos ecológicos (Greenland, 2001a, 2002b, 2004a). Todos esses problemas, combinados, podem produzir resultados de validade questionável, em qualquer nível. Apesar disso, os estudos ecológicos podem ser úteis para detectar associações de distribuições de exposição com ocorrência de doença, porque tais associações podem assinalar a presença de efeitos que mereçam maior investigação. Uma discussão detalhada de estudos ecológicos é apresentada no Capítulo 25.

## Geração de hipótese *versus* rastreamento de hipótese

Estudos nos quais a validade é menos segura têm sido referidos às vezes como estudos “geradores de hipótese”, para distingui-los dos “estudos analíticos”, nos quais a validade pode ser melhor. Os estudos ecológicos têm sido considerados, frequentemente, como estudos geradores de hipótese, por preocupação com vários vieses. A distinção, entretanto, entre estudos geradores de hipótese e estudos analíticos não é conceitualmente precisa. É o pesquisador, não o estudo, quem gera hipóteses, e quaisquer tipos de dados podem ser usados para testar hipóteses. Por exemplo, comparações inter-

nacionais indicam que as mulheres japonesas têm uma taxa de câncer de mama muito mais baixa do que as mulheres nos Estados Unidos. Esses dados são ecológicos e sujeitos às preocupações usuais sobre as muitas diferenças culturais que existem entre os dois grupos. Entretanto, o achado corrobora numerosas hipóteses, inclusive as teorias de que menarca precoce, dietas ricas em gorduras e tamanho grande da mama (todos mais frequentes em mulheres americanas do que em japonesas) podem ser fatores determinantes importantes do risco de câncer de mama (p. ex., ver Trichopoulos e Lipman, 1992). A diferença internacional nas taxas de câncer de mama não é geradora de hipótese nem de estudo analítico, pois as hipóteses surgiram independentemente desse achado. Assim, a diferença entre estudos geradores de hipótese e estudos analíticos seria melhor substituída por uma distinção mais acurada.

Uma proposta que enxergamos favoravelmente é se referir a estudos preliminares, de validade ou de precisão limitada, como *estudos de rastreamento de hipótese*. Em analogia com o rastreamento de indivíduos para doença, tais estudos representam um teste relativamente fácil e barato para a presença de uma associação entre exposição e doença. Se tal associação é detectada, ela está sujeita a testes mais rigorosos e dispendiosos, usando-se um delineamento mais válido, que pode ser chamado de estudo confirmador. Embora a analogia com o rastreamento não deva ser levada ao extremo, ela descreve melhor a progressão dos estudos do que a distinção estudo gerador de hipótese/estudo analítico.

# Estudos de coorte

Kenneth J. Rothman e Sander Greenland

## Definição de coortes e grupos de exposição 123

- Classificação de pessoa-tempo 125
- Exposições crônicas 126
- Tempo sem exposição em sujeitos expostos 127
- Categorizando a exposição 128
- Intensidade média e alternativas 130
- Pessoa-tempo imortal 130

Eventos pós-exposição 131

**Tempo dos eventos de desfecho 132**

**Custos 132**

**Localização de sujeitos 134**

**Coortes de exposição especial e da população geral 134**

A meta de um estudo de coorte é mensurar e, geralmente, comparar a incidência de doença em uma ou mais coortes estudadas. Conforme discutido no Capítulo 3, a palavra *coorte* designa um grupo de pessoas que compartilham uma experiência, ou condição, em comum. Por exemplo, uma coorte de nascimento compartilha o mesmo ano ou período de nascimento, uma coorte de fumantes tem a experiência de fumar em comum e uma coorte de vegetarianos compartilha uma certa dieta. Frequentemente, se há duas coortes no estudo, uma delas é descrita como a coorte exposta, constituída por indivíduos que experimentaram um evento ou com uma condição causal presumível, e a outra é descrita como a coorte não exposta, ou de referência. Se houver mais de duas coortes, cada uma pode ser caracterizada por um nível ou tipo diferente de exposição.

O presente capítulo foca elementos básicos para o delineamento e a condução de estudos de coorte. Maiores considerações para o delineamento de estudos de coorte são apresentadas nos Capítulos 9 a 11, ao passo que métodos de análise aplicáveis a estudos de coorte estão nos Capítulos 14 a 21. Muitos aspectos essenciais de avaliação da exposição, que não são discutidos aqui, podem ser encontrados em Armstrong e colaboradores (1992).

## DEFINIÇÃO DE COORTES E GRUPOS DE EXPOSIÇÃO

Em princípio, um estudo de coorte poderia ser usado para estimar riscos médios, taxas ou tempos de ocorrência. Exceto em certas situações, nas quais riscos médios e tempos de ocorrência não podem ser mensurados diretamente a partir da experiência de uma coorte. A observação de riscos médios ou de tempos de eventos específicos requer que a coorte inteira permaneça exposta ao risco e sob observação, durante todo o período de seguimento. A perda do acompanhamento de sujeitos durante o período de estudo impede mensurações diretas dessas médias, porque a evolução dos sujeitos perdidos é desconhecida. Sujeitos que falecem por riscos competitivos (outros desfechos que não aqueles de interesse), da mesma forma, impedem o pesquisador de estimar riscos condicionais diretamente (risco de um desfecho específico condicionado à não ocorrência de desfecho). Assim, a

única situação na qual é possível mensurar riscos médios e tempos de ocorrência diretamente em um estudo de coorte é aquela em que há pouca ou nenhuma perda de seguimento e risco competitivo escasso. Embora alguns ensaios clínicos forneçam essas condições, muitos estudos epidemiológicos não o fazem. Quando ocorrem perdas e riscos competitivos, pode-se ainda estimar a taxa de incidência diretamente, ao passo que o risco e o tempo médios de ocorrência devem ser estimados pelo uso de métodos de sobrevida (tábua de vida) (ver Caps. 3 e 16).

Diferentemente dos riscos médios, que são mensurados com indivíduos como a unidade no denominador, as taxas de incidência têm pessoa-tempo como a unidade de medida. O acúmulo de tempo, e não de indivíduos, no denominador de taxas permite flexibilidade na análise de estudos de coorte. Enquanto os estudos que estimam risco diretamente estão ligados conceitualmente à identificação de coortes específicas de indivíduos, os estudos com medidas de taxas de incidência podem, dentro de certas suposições, definir os grupos de comparação em termos de unidades pessoa-tempo, que não correspondem a coortes específicas de indivíduos. Um dado indivíduo pode contribuir pessoa-tempo a um, a dois ou a mais grupos de exposição em um dado estudo, porque cada unidade de pessoa-tempo contribuída ao seguimento por um dado indivíduo possui sua própria classificação com relação à exposição. Assim, um indivíduo, cuja experiência de exposição muda com o tempo, pode, dependendo dos detalhes da hipótese do estudo, contribuir tempo de seguimento a várias taxas diferentes, específicas por exposição. Em tal estudo, a definição de cada grupo de exposição corresponde à definição de elegibilidade pessoa-tempo para cada categoria de exposição.

Como resultado desse foco em pessoa-tempo, nem sempre faz sentido referir-se aos membros de um grupo de exposição, dentro de um estudo de coorte, como se o mesmo conjunto de indivíduos estivesse exposto em todos os pontos do tempo. Os termos *população aberta*, ou *população dinâmica*, descrevem uma população na qual a experiência de pessoa-tempo pode crescer a partir de uma lista cambiante de indivíduos (ver Cap. 3). (Algumas vezes usa-se o termo *coorte aberta*, ou *coorte dinâmica*, mas esse uso é conflitante com outro, no qual uma coorte é uma lista fixa de indivíduos.) Por exemplo, as taxas de incidência de câncer relatadas pelo Connecticut Cancer Registry (registro de Câncer de Connecticut) são derivadas da experiência de uma população aberta. Como a população de residentes de Connecticut está sempre mudando, os indivíduos que contribuem para essas taxas não são um grupo específico de pessoas acompanhadas ao longo do tempo.

Quando os grupos de exposição em um estudo de coorte são definidos no início do seguimento, sem movimento algum de indivíduos entre os grupos de exposição durante o seguimento, os grupos são chamados, algumas vezes, de *coortes fixas*. Os grupos definidos por alocação do tratamento em ensaios clínicos são exemplos de coortes fixas. Se o seguimento de coortes fixas sofrer com perdas, ou por riscos competitivos, as taxas de incidência ainda poderão ser mensuradas diretamente e utilizadas para estimar riscos médios e tempos de incidência. Se não ocorrerem perdas de uma coorte fixa, esta satisfaz a definição de uma *população fechada* (ver Cap. 3) e é chamada, frequentemente, de *coorte fechada*. Em tais coortes, os riscos não condicionais (que incluem o efeito de riscos competitivos) e os tempos médios de sobrevida podem ser estimados diretamente.

No estudo de coorte mais simples, a exposição seria uma condição permanente e facilmente identificável, fazendo do trabalho de alocar sujeitos a coortes expostas e a não expostas uma tarefa fácil. Infelizmente, as exposições de interesse para os epidemiologistas raramente são constantes e, frequentemente, são difíceis de mensurar. Considere como um exemplo os problemas de identificar para estudo uma coorte de usuários de um medicamento de prescrição específica. Identificar os usuários requer um método para localizar aqueles que recebem prescrições ou compram medicamento. Sem um sistema de registro de prescrições, isso se torna uma tarefa difícil. Mesmo com um sistema de registros, a identificação daqueles que recebem a prescrição, ou mesmo daqueles que adquirem o medicamento, não é equivalente à identificação daqueles que realmente o usam. Além disso, aqueles que são usuários desse medicamento hoje podem não sê-lo amanhã, e vice-versa. A definição do uso deve estar ligada ao tempo, porque a exposição pode se modificar. Finalmente, o efeito que está sendo estudado pode envolver um período de indução considerável. Nesse caso, o *status* da exposição em um dado tempo

estará relacionado a um possível aumento ou diminuição do risco da doença, somente algum tempo mais tarde. Assim, alguém que começasse a usar o medicamento hoje poderia experimentar um efeito relacionado a ele em 10 anos, mas poderia não experimentar qualquer efeito relacionado nos primeiros 5 anos depois da exposição.

É tentador pensar sobre a identificação de estudos de coorte como simplesmente um processo de identificar e classificar indivíduos quanto a seu *status* de exposição. O processo pode ser complicado, contudo, pela necessidade de classificar a experiência de um só indivíduo em diferentes categorias de exposição, e em tempos distintos. Se a exposição pode variar com o tempo, no mínimo o pesquisador precisa levar em conta o tempo experimentado por cada sujeito do estudo em cada categoria de exposição, na definição das coortes do estudo. A sequência ou o tempo da exposição também podem ser importantes. Se houver muitas sequências de exposição possíveis, cada indivíduo poderia ter uma sequência única de categorias de exposição, e assim poderia se definir uma coorte de exposição única, contendo apenas aquele indivíduo.

Um pressuposto simplificador comum na análise epidemiológica é que o único aspecto de uma exposição que determina o risco atual seja algum sumário numérico simples da história de exposição. Os sumários típicos incluem nível atual de exposição, exposição média e exposição cumulativa, isto é, a soma de cada categoria de exposição multiplicada pelo tempo passado naquele nível. Frequentemente, a exposição é “retardada” no sumário, o que significa que somente é contada a exposição em, ou até, algum tempo especificado antes do atual momento. Embora se tenha uma flexibilidade enorme na definição de sumários de exposição, métodos baseados na presunção de que apenas um só sumário é relevante, podem conter vieses graves, sob certas condições (Robins, 1987). Por enquanto, presumiremos que um sumário isolado é uma medida adequada de exposição. Com esse pressuposto, os estudos de coorte podem ser analisados definindo-se as coortes com base em pessoa-tempo em vez de em pessoas, de modo que uma pessoa pode ser um membro de coortes de exposição diferentes, em tempos diversos. No entanto, alertamos o leitor a ter em mente o pressuposto de sumário único, ao interpretar tais análises. O tempo que um indivíduo contribui para o denominador de uma ou mais taxas de incidência em um estudo de coorte é denominado, algumas vezes, *tempo em risco*, no sentido de estar em risco para o desenvolvimento da doença. Algumas pessoas e, conseqüentemente, todas as suas pessoas-tempo, não estão em risco para uma dada doença, porque são imunes ou porque não têm o órgão-alvo da doença em estudo. Por exemplo, as mulheres que sofreram uma histerectomia, e, por definição, todos os homens, não estão em risco para câncer uterino, porque não têm útero.

## Classificação de pessoa-tempo

O guia principal para a classificação de pessoas, ou de pessoa-tempo, é a hipótese do estudo, que deve ser definida tão detalhadamente quanto for possível. Se o estudo aborda a questão da extensão em que a ingestão de cenouras reduzirá o risco subsequente de câncer do pulmão, a hipótese em estudo é melhor declarada em termos de que quantidade de cenouras, consumida em qual período de tempo, prevenirá o câncer de pulmão. Além disso, a hipótese do estudo deve especificar um tempo de indução entre o consumo de uma dada quantidade de cenouras e o efeito subsequente: o efeito do consumo de cenouras poderia começar imediatamente, começar gradualmente ou só começar depois de um intervalo e se estender além do tempo em que um indivíduo parasse de comer cenouras (Rothman, 1981).

Em estudos com exposições crônicas (i.e., exposições que persistem durante um período prolongado de tempo), é fácil confundir o tempo durante o qual a exposição ocorre com o tempo em risco para efeitos da exposição. Por exemplo, em estudos ocupacionais, o tempo de emprego é confundido algumas vezes com o tempo em risco para efeitos da exposição. O tempo de emprego é o tempo durante o qual a exposição se acumula. Em contraste, o tempo em risco para efeitos da exposição deve vir, logicamente, depois do acúmulo de uma quantidade específica de exposição, porque somente depois daquele tempo a doença pode ser causada, ou prevenida, por aquela quantidade de exposição. As durações desses dois períodos de tempo não têm relação constante uma com a outra.

O tempo em risco de efeitos pode muito bem se estender para além do fim do emprego. É somente o tempo em risco de efeitos que deve ser contabilizado no denominador das taxas de incidência, para aquela quantidade de exposição.

A distinção entre tempo de acréscimo de exposição e o tempo em risco de efeitos da exposição é mais fácil de entender considerando-se um exemplo no qual a exposição é muito curta. Em estudos dos efeitos retardados da exposição à radiação emitida pela bomba atômica, a exposição foi quase instantânea, mas o período de risco durante o qual a exposição tem produzido efeito é longo, talvez uma vida inteira, embora o risco para certas doenças não aumente imediatamente após a exposição. O risco de câncer após a exposição à radiação só aumentou depois de um período de indução de no mínimo vários anos, dependendo do tipo de câncer. As taxas de incidência de câncer entre aqueles expostos a doses altas de radiação da bomba podem ser calculadas separadamente para tempos diferentes em seguida à exposição, de modo que se podem detectar elevações específicas ao período de indução abordado na hipótese do estudo. Sem estratificação por tempo desde a exposição, a taxa de incidência mensurada entre aqueles expostos à bomba seria uma taxa média refletindo períodos de efeito da exposição e períodos sem tal efeito, porque seria incluída no denominador alguma experiência da coorte exposta que corresponde ao tempo em que não houve risco aumentado da radiação.

Como o pesquisador deve estudar hipóteses que não especificam tempos de indução? Para essas hipóteses, os períodos de tempo apropriados, nos quais estratificar as taxas de incidência, não são claros. Não há maneira de estimar efeitos da exposição, contudo, sem fazer alguma suposição, implícita ou explicitamente, sobre o tempo de indução. A decisão sobre que tempo incluir para um dado indivíduo no denominador da taxa corresponde à suposição sobre tempo de indução. Se em um estudo dos efeitos retardados em sobreviventes das bombas atômicas no Japão, o denominador da taxa incluísse o tempo experimentado pelos sujeitos do estudo começando no dia após a exposição, a taxa forneceria uma estimativa de efeito diluída, a menos que o período de indução (inclusive o período “latente”) fosse pelo menos de um dia. Poderia ser mais apropriado permitir um período de indução mínimo de alguns meses ou anos depois da explosão da bomba.

E se o pesquisador não tiver qualquer base para hipótese de um período específico de indução? É possível aprender sobre o período, estimando-se os efeitos de acordo com categorias de tempo desde a exposição. Por exemplo, a taxa de incidência de leucemia entre sobreviventes da bomba atômica, relativa àqueles que estavam longe durante a explosão, pode ser examinada de acordo com anos desde a explosão. Em um estudo sem viés, esperaríamos que as estimativas de efeito se elevassem acima do valor de nulidade, quando o período mínimo de indução houvesse passado. Esse procedimento funciona melhor quando a própria exposição ocorre em um ponto, ou em um intervalo estreito, do tempo, mas pode ser usado mesmo que a exposição seja crônica, desde que haja um modelo para descrever a quantidade de tempo que deve passar antes que um dado acúmulo de exposição começasse a ter um efeito. Abordagens mais sofisticadas para analisar o tempo de indução são discutidas no Capítulo 16.

## Exposições crônicas

A definição de exposição crônica com base nos efeitos previstos é mais complicada do que quando a exposição ocorre apenas em um ponto do tempo. Podemos conceituar um período durante o qual a exposição se acumula até o suficiente para desencadear um passo no processo causal. Esse acúmulo de experiência com a exposição pode ser uma função complexa da intensidade da exposição e do tempo. O período de indução só começa depois que a exposição alcançou esse ponto hipotético de gatilho, e tal ponto provavelmente variará entre os indivíduos. Os epidemiologistas ocupacionais, frequentemente, mensuram o tempo de indução para exposição ocupacional a partir do tempo da primeira exposição, mas esse procedimento envolve a suposição extrema de que o primeiro contato com a exposição pode ser suficiente para produzir a doença. Qualquer que seja a suposição adotada, ela deve ser parte explícita da definição da coorte e do período de seguimento.

Consideremos os passos a tomar para identificar coortes de estudo, quando a exposição é crônica. Primeiramente, o pesquisador deve determinar quantos grupos de exposição serão estudados e estabelecer as definições para cada uma das categorias de exposição. A definição de categoria de exposição poderia ser baseada na intensidade máxima de exposição experimentada, na intensidade média durante um período de tempo ou em alguma quantidade de exposição cumulativa. Uma mensuração familiar do fumo de cigarros é a medida maços-anos, que é o produto do número de maços fumados por dia pelo número de anos de fumo. Essa medida indexa o número cumulativo de cigarros fumados, com um maço por dia igual ao produto de 20 cigarros por maço e 365 dias, ou 7.300 cigarros. Índices cumulativos de exposição e medidas ponderadas por tempo da intensidade média da exposição são métodos populares para mensurar a exposição em estudos ocupacionais. Essas definições de exposição devem ser ligadas ao período de tempo de um efeito da exposição, de acordo com a hipótese do estudo, levando em consideração, explicitamente, o período de indução.

Ao empregar medidas de exposição cumulativas ou médias, deve-se reconhecer a natureza composta das medidas e, se possível, analisar separadamente os componentes. Por exemplo, maços-anos é um composto de duração e de intensidade do fumo: 20 maços-anos podem representar meio maço por dia durante 40 anos, um maço por dia durante 20 anos ou 2 maços por dia durante 10 anos, assim como outras inúmeras combinações. Se os efeitos biológicos dessas combinações diferem em grau importante, o uso de maços-anos ocultaria essas diferenças e, talvez, até mesmo apresentaria uma impressão enganosa de padrões de doses-resposta (Lubin e Caporaso, 2006). Análises suplementares do tabagismo como duas variáveis de exposição, duração (anos de fumo) e intensidade (maços fumados por dia), forneceriam uma salvaguarda contra inadequações da análise de maços-anos. Outras variáveis de exposição que não são levadas em consideração pela duração e pela intensidade, tais como idade no início da exposição, idade na cessação da exposição e tempo de exposição relativo à doença (período de indução ou de retardo), também podem merecer separação nas análises.

Observemos um exemplo simplificado. Suponha a hipótese de estudo de que o fumo aumenta o risco de câncer de pulmão com um tempo de indução mínimo de 5 anos. Para uma dada categoria de fumo, o tempo experimentado por um sujeito não é a pessoa-tempo “exposta”, até que o indivíduo tenha alcançado aquele nível, e, então, 5 anos adicionais tenham passado. Somente então a experiência de câncer do pulmão daquele indivíduo relacionou-se ao fumo, de acordo com a hipótese do estudo. A definição da coorte de estudo com 20 maços-anos de fumo será a experiência de pessoa-tempo de indivíduos expostos, começando 5 anos depois que eles tenham fumado 20 maços-anos. Note que se o estudo de coorte mede taxas de incidência, o que significa que aloca a pessoa-tempo dos sujeitos individuais do estudo, os grupos de exposição são definidos pela alocação de pessoas-tempo, e não por listas de sujeitos individuais. A análise dessas taxas dependerá da suposição de que somente a exposição “atual”, definida como tendo fumado 20 maços-anos há 5 anos, é relevante, e que outros aspectos de história de exposição, tal como a quantidade fumada depois de 5 anos atrás, são irrelevantes.

## **Tempo sem exposição em sujeitos expostos**

O que acontece com o tempo experimentado por sujeitos expostos que não preenche a definição de tempo em risco de efeitos da exposição, de acordo com a hipótese do estudo? Especificamente, o que acontece com o tempo depois que os sujeitos expostos tornam-se expostos, e antes que a indução mínima tenha passado? Duas escolhas são razoáveis para lidar com essa experiência. Uma possibilidade é considerar qualquer tempo que não esteja relacionado à exposição como o tempo não exposto, e destiná-lo à coorte do estudo que representa não exposição. Objeções possíveis a essa abordagem seriam que a hipótese em estudo pode ser baseada em meros palpites sobre o limiar para efeitos da exposição e sobre o período de indução, e que o tempo durante o acúmulo de exposição, ou períodos de indução, podem, de fato, estar no risco de efeitos da exposição. Tratar a última experiência como não em risco de efeitos da exposição pode levar, então, a uma subestimação do efeito da exposição (ver Cap. 9 para uma discussão sobre erro de classificação da exposição). De modo alternativo, pode-se



simplesmente omitir do estudo a experiência de sujeitos expostos, que não estão em risco de efeitos da exposição, conforme a hipótese do estudo. Para essa alternativa ser prática, deve haver um número razoavelmente grande de casos observados entre sujeitos com nenhuma exposição.

Por exemplo, suponha a hipótese de um tempo de indução mínimo de 10 anos. Para indivíduos acompanhados desde o início da exposição, essa hipótese implica que nenhum efeito da exposição pode ocorrer dentro dos primeiros 10 anos de seguimento. Somente após os primeiros 10 anos de seguimento, um indivíduo pode experimentar doença causada pela exposição. Portanto, sob a hipótese, somente pessoas-tempo ocorrendo depois de 10 anos de exposição deveriam contribuir para o denominador da taxa, entre os expostos. Se a hipótese estivesse correta, deveríamos alocar os primeiros 10 anos de seguimento ao denominador da taxa de não expostos. Suponha, entretanto, que a hipótese estivesse seguimento, e que a exposição pudesse produzir casos em menos de 10 anos. Então, se os casos e pessoas-tempo dos primeiros 10 anos de seguimento fossem acrescentados aos casos e pessoas-tempo não expostos, a taxa resultante teria um viés em direção à taxa dos expostos, reduzindo assim as diferenças aparentes entre as taxas de expostos e não expostos. Se o cálculo da taxa dos não expostos fosse limitado a casos e pessoas-tempo verdadeiramente não expostos, esse problema seria evitado.

O preço de se evitar tal problema, entretanto, seria a precisão reduzida na estimativa da taxa entre os não expostos. Em alguns estudos, o número de casos verdadeiramente não expostos é pequeno demais para produzir uma comparação estável, e, assim, a experiência inicial de pessoas expostas é demasiado valiosa para ser descartada. Em geral, o melhor procedimento em uma dada situação dependeria da diminuição de precisão produzida pela exclusão da experiência inicial de pessoas expostas, e da quantidade de viés introduzido pelo tratamento da experiência precoce de pessoas expostas, como se ela fosse equivalente àquela de pessoas que nunca foram expostas. Uma alternativa que tenta abordar ambos os problemas é tratar o tempo de indução como se fosse uma variável contínua em vez de um tempo fixo, e modelar os efeitos da exposição como dependentes dos tempos de exposição (Thomas, 1983, 1988). Essa abordagem é discutivelmente mais realista, na medida em que o tempo de indução varia entre os indivíduos.

Assuntos semelhantes surgem se o *status* da exposição puder mudar de expostos para não expostos. Se a exposição cessa, mas é pensado que os efeitos dela continuam, não faria sentido colocar a experiência de um indivíduo anteriormente exposto na categoria de não exposto. Contudo, se é pensado que os efeitos da exposição ocorrem aproximadamente ao mesmo tempo que ela, o que significa que o período de indução é próximo de zero, então, alterações no *status* de exposição deveriam levar a mudanças correspondentes em como a experiência cumulativa é classificada com respeito à exposição. Por exemplo, se indivíduos tomando uma droga anti-inflamatória estão em risco aumentado de sangramento gastrointestinal somente durante o período em que estão usando o medicamento, então, apenas o tempo durante a exposição é equivalente ao risco de sangramento gastrointestinal em consequência da droga. Quando um indivíduo para de usar a droga, os eventos de sangramento e pessoas-tempo experimentados por ele devem ser reclassificados de exposto para não exposto. Aqui, o tempo de indução é zero e a definição de exposição não envolve história de exposição.

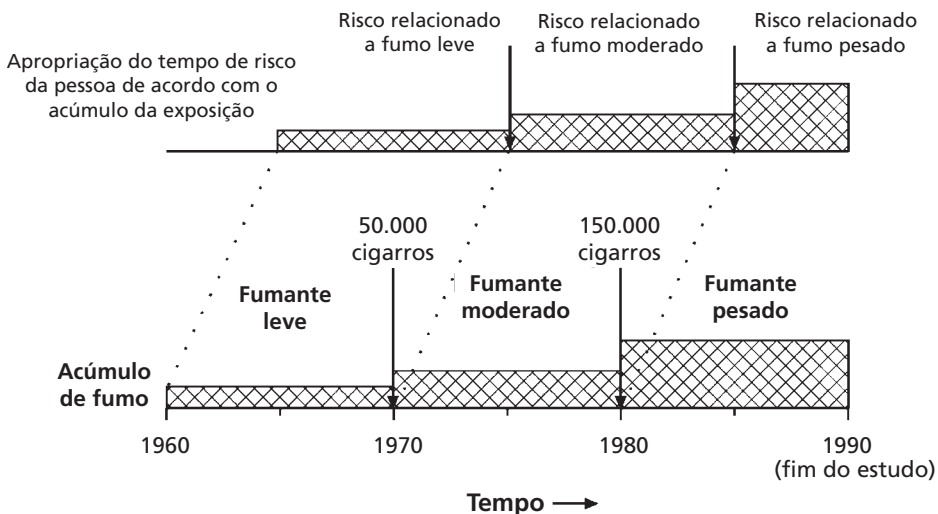
## Categorizando a exposição

Outro problema a considerar é que a hipótese do estudo pode não fornecer orientação razoável sobre onde traçar o limite entre expostos e não expostos. Se a exposição é contínua, não é necessário traçar limite algum. Em vez disso, pode-se usar as informações quantitativas de cada indivíduo completamente, ou pelo uso de algum tipo de nivelamento, tal como médias móveis (ver Cap.17), ou pela colocação da variável exposição dentro de um modelo de regressão como um termo contínuo (ver Caps. 20 e 21). Naturalmente, a última abordagem depende da validade do modelo utilizado para estimação. Cuidado especial deve ser tomado com modelos de exposições e confundidores mensurados repetidamente, os quais são chamados, algumas vezes, de modelos de dados longitudinais (ver Cap. 21).

A abordagem mais simples para calcular taxas diretamente requererá uma população de tamanho razoável dentro das categorias de exposição, se é desejado um resultado estatisticamente estável. Para se obter taxas de incidência, então, precisamos agrupar as experiências dos indivíduos em categorias relativamente grandes, para as quais podemos calcular taxas de incidência. Em princípio, seria possível formar várias coortes que correspondessem a várias categorias de exposição. Para uma medida cumulativa de exposição, contudo, a categorização pode introduzir dificuldades adicionais para a definição de coorte. Um indivíduo que, ao longo do caminho, passasse de uma categoria de exposição para outra mais alta teria, mais tarde, um risco de doença que, teoricamente, poderia se enquadrar na definição de mais de uma categoria de exposição.

Por exemplo, suponha que definimos fumante moderado como tendo fumado 50.000 cigarros (equivalente a cerca de 7 maços-anos) e que definimos fumante pesado como tendo fumado 150.000 cigarros (cerca de 21 maços-anos). Suponha que um homem fumou seu 50.000<sup>o</sup> cigarro em 1970 e seu 150.000<sup>o</sup> em 1980. Após o desconto do mínimo de 5 anos para o período de indução, classificariamos seu tempo como fumante moderado começando em 1975. Em 1980, ele tornou-se um fumante pesado, mas o período de indução de 5 anos para o fumo pesado ainda não passou. Assim, de 1980 a 1985 sua experiência ainda é classificada como fumo moderado, mas de 1985 em diante passou a ser de fumo pesado (Fig. 7.1). Geralmente, o tempo é alocado apenas à categoria mais alta de exposição que se aplica. Esse exemplo ilustra a complexidade da definição de coorte com uma hipótese que leva em conta tanto a quantidade de exposição cumulativa quanto o tempo mínimo de indução. Outros esquemas de apropriação poderiam ser imaginados com base em outras hipóteses sobre a ação da exposição, inclusive em hipóteses que permitissem que o tempo de indução variasse com a história de exposição.

Um esquema de alocação inválido incluiria no denominador da taxa de incidência dos expostos a experiência de não exposição de um indivíduo que, eventualmente, torna-se exposto. Por exemplo, suponha que, em um estudo ocupacional, a exposição é categorizada de acordo com a duração de emprego em um trabalho específico, e com a categoria de exposição mais alta tendo, pelo menos, 20 anos de emprego. Suponha que um trabalhador ficou empregado naquela atividade por 30 anos. É um erro alocar os 30 anos de experiência daquele trabalhador à categoria de exposição de



**FIGURA 7.1** • Linha de tempo mostrando como um fumante se movimenta para categorias mais altas de exposição cumulativa ao fumo e como o tempo em risco que corresponde a essas categorias é apropriado para levar em conta um período de indução mínimo de 5 anos.

20 ou mais anos de emprego. O trabalhador só atingiu aquela categoria de exposição depois de 20 anos no emprego, e somente os últimos 10 anos de sua experiência são relevantes para a categoria de exposição mais alta. Observe que se o trabalhador tivesse falecido após 10 anos de emprego, o óbito não poderia ter sido alocado à categoria de 20 anos de emprego, porque o trabalhador teria tido somente 10 anos de emprego.

Uma regra útil para lembrar é que o evento e o pessoa-tempo que está sendo acumulado no momento do evento devem ser alocados à mesma categoria de exposição. Assim, uma vez que o pessoa-tempo passada em cada categoria de exposição tenha sido determinado para cada sujeito do estudo, a classificação dos eventos (casos) de doença segue as mesmas regras. A categoria de exposição à qual um evento é alocado é a mesma categoria em que o pessoa-tempo para aquele indivíduo estava se acumulando no instante em que o evento ocorreu. A mesma regra – que a classificação do evento acompanha a classificação do pessoa-tempo – também se aplica em relação a outras variáveis do estudo, que podem ser usadas para estratificar os dados (ver Cap. 15). Por exemplo, o pessoa-tempo será alocado em uma categoria de idade diferente à medida que o indivíduo envelhece. A categoria de idade à qual um evento é designado deve ser a mesma categoria etária na qual o pessoa-tempo individual estava se acumulando por ocasião do evento.

## Intensidade média e alternativas

Pode-se também definir a exposição atual de acordo com a média (aritmética ou geométrica) de intensidade, ou de categoria de exposição, até o tempo atual, em vez de por uma medida cumulativa. No cenário ocupacional, a concentração média de um agente no ar ambiente seria um exemplo de intensidade de exposição, embora também devesse ser levado em conta qualquer equipamento protetor que afetasse a exposição de um indivíduo ao agente. Intensidade de exposição é um conceito que se aplica a um ponto no tempo e, tipicamente, variará com o decorrer do tempo. Os estudos que mensuram a intensidade da exposição podem usar uma média de intensidade ponderada para o tempo, o que requereria mensurações múltiplas da exposição durante o tempo. A quantidade de tempo na qual um indivíduo é exposto a cada intensidade forneceria sua ponderação no cálculo da média.

Uma alternativa à intensidade média é classificar a exposição de acordo com a intensidade máxima, com a intensidade mediana, com a intensidade mínima ou com alguma outra função da história de exposição. O tempo de acompanhamento que um indivíduo passa em uma dada intensidade de exposição poderia começar a se acumular assim que aquela categoria de exposição fosse atingida. O tempo de indução também deve ser levado em consideração. O ideal é que, a hipótese do estudo especifique um tempo de indução mínimo para efeitos da exposição, o que, por sua vez, implicará um intervalo de tempo apropriado a ser usado na classificação da experiência individual.

Os esquemas de alocação da exposição, cumulativos e por média, padecem de um problema especial, pois podem tornar impossível desenredar os efeitos da exposição dos efeitos de confundidores que variem com o tempo (Robins 1986, 1987). Métodos que tratem exposições e confundidores em um período como distinto da exposição, e confundidores em outros períodos, são necessários para evitar esse problema (Robins et al., 1992; ver Cap. 21).

## Pessoa-tempo imortal

Ocasionalmente, uma definição de coorte precisará que todos que se enquadrem na definição tenham sobrevivido por um período especificado. Tipicamente, esse período de imortalidade surge porque um dos critérios de entrada na coorte depende da sobrevivência. Por exemplo, uma coorte ocupacional pode ser definida como todos os operários que tenham sido empregados em uma fábrica por pelo menos 5 anos. Há certos problemas com tal critério de entrada, entre os quais ele garantirá que o estudo perderá efeitos entre os trabalhadores temporários, que podem ser designados para tarefas mais altamente expostas do que os empregados regulares, de longa permanência, e poderá incluir pessoas mais suscetíveis aos efeitos da exposição, que poderão se demitir mais cedo por causa de tais efeitos.

Vamos supor, entretanto, que somente os operários de longa permanência são de interesse para o estudo e que todas as exposições relevantes (inclusive aquelas durante os 5 anos iniciais de emprego) são levadas em consideração na análise.

O critério de 5 anos de ingresso garantirá que todos os trabalhadores na coorte do estudo sobreviveram a seus primeiros 5 anos de emprego, porque aqueles que morressem nunca satisfariam o critério de admissão e, assim, seriam excluídos. Segue-se que a análise de mortalidade de tais operários deveria excluir os primeiros 5 anos de emprego para cada trabalhador. Esse período de tempo é referido como *pessoa-tempo imortal*. É claro que os operários da fábrica não eram imortais durante esse tempo, porque eles poderiam ter morrido. O subgrupo de trabalhadores que satisfazem a definição da coorte, porém, é identificado, depois do fato, como aqueles que sobreviveram a esse período.

A abordagem correta para lidar com pessoa-tempo imortal em um estudo é excluí-la de qualquer denominador, mesmo que a análise não tenha mortalidade como foco. Essa abordagem é correta, porque a inclusão de pessoa-tempo imortal enviesará para baixo as taxas de morbidade estimadas e, conseqüentemente, trará viés às estimativas de efeitos obtidas das comparações internas. Como um exemplo, suponha que um estudo de mortalidade ocupacional inclua somente operários que trabalharam por 5 anos em uma fábrica, que 1.000 trabalhadores expostos e 1.000 não expostos preenchem esse critério de entrada, e que depois que o critério for satisfeito observamos 200 mortes entre 5.000 pessoas-ano expostas, e 90 mortes entre 6.000 pessoas-ano não expostas. A razão e a diferença de proporções corretas, comparando os expostos e não expostos, são então  $(200/5.000)/(90/6.000) = 2,7$  e  $200/5.000 - 90/6.000 = 25/1.000 \text{ ano}^{-1}$ . Se, entretanto, incluíssemos de forma incorreta as pessoas-ano imortais de 5.000 expostos e 5.000 não expostos nos denominadores, obteríamos uma razão com viés de  $(200/10.000)/(90/11.000) = 2,4$  e uma diferença com viés de  $200/10.000 - 90/11.000 = 12/1.000 \text{ ano}^{-1}$ . Para evitar esse viés, se um estudo tem um critério para uma quantidade mínima de tempo antes que um sujeito seja elegível para estar em um estudo, o tempo durante o qual o critério de elegibilidade é preenchido deveria ser excluído do cálculo das taxas de incidência. De um modo mais geral, o tempo de seguimento alocado a uma categoria de exposição específica deve excluir o tempo durante o qual a definição da categoria de exposição está sendo preenchida.

## Eventos pós-exposição

A alocação de tempo de seguimento a categorias específicas não deve depender de eventos que ocorram depois que o seguimento em questão tenha advindo. Por exemplo, considere um estudo no qual um grupo de fumantes é aconselhado a deixar de fumar, com o objetivo de estimar o efeito de abandonar o fumo *versus* continuar fumando sobre as taxas de mortalidade. Para um sujeito que fuma por algum tempo depois que o conselho é dado, e deixa de fumar mais tarde, o seguimento de como alguém deixou o fumo só deve começar no momento em que o sujeito para, e não quando recebe o conselho, porque é o efeito de deixar de fumar que está sendo estudado, não o efeito do aconselhamento (se o efeito do conselho estivesse em estudo, o seguimento começaria com o aconselhamento). Mas, como deve ser tratado um sujeito que deixa de fumar por um tempo e mais tarde volta a fumar?

Quando essa questão surgiu em um estudo real desse problema, os pesquisadores excluíram do estudo qualquer um que voltasse a fumar. A decisão foi equivocada, porque se o sujeito tivesse morrido antes de voltar a fumar, o óbito teria sido contabilizado no estudo, e o sujeito não teria sido excluído. O tempo de seguimento do sujeito foi excluído se ele voltou a fumar, algo que só ocorreu *depois* que o sujeito tinha acrescido tempo na coorte dos que deixaram de fumar. Uma análise apropriada deveria incluir a experiência daqueles que voltaram a fumar até o tempo em que retomaram o fumo. Se a propensão a voltar a fumar não fosse associada a risco, sua experiência subsequente à volta poderia ser excluída, sem introduzir viés. A taxa de incidência entre as pessoas-ano, enquanto tendo deixado, poderia então ser comparada com a taxa entre aqueles que continuaram fumando durante o mesmo período.

Como outro exemplo, suponha que os pesquisadores quisessem examinar o efeito de ser um ex-fumante por pelo menos 5 anos em relação a um fumante contínuo. Então, qualquer ex-fumante por pelo menos 5 anos que voltasse a fumar nesse período seria excluído. A experiência de pessoa-tempo para cada sujeito durante os primeiros 5 anos após deixar de fumar também deveria ser excluída, porque seria pessoa-tempo imortal.

## TEMPO DOS EVENTOS DE DESFECHO

Como pode ser aparente por discussão prévia, o tempo em que um evento de desfecho ocorre pode ser um determinante capital da quantidade de pessoa-tempo contribuída por um sujeito a cada categoria de exposição. É importante, portanto, definir e determinar o tempo do evento da maneira menos ambígua e mais precisa possível. Para alguns eventos, tal como morte, nenhuma das duas tarefas apresenta qualquer dificuldade. Para outros desfechos, tal como a soroconversão ao vírus da imunodeficiência humana (HIV), o tempo do evento pode ser definido de um modo razoavelmente preciso (o aparecimento de anticorpos HIV na corrente sanguínea), mas a mensuração do tempo é difícil. Para outros, tais como esclerose múltipla e aterosclerose, a própria definição do tempo de início pode ser ambígua, mesmo quando a presença da doença pode ser determinada de forma inequívoca. Da mesma forma, o tempo de perda do seguimento e outros eventos censurados podem ser difíceis de definir e de determinar. Determinar se um evento ocorreu até um dado momento é um caso especial em que se deve estabelecer quando um evento aconteceu, porque tal conhecimento requer saber que o tempo do acontecimento foi anterior ao dado tempo.

A abordagem dos problemas supramencionados depende fortemente de detalhes dos dados disponíveis e do estado atual de conhecimento do desfecho em estudo. Portanto, somente fazemos aqui algumas observações gerais sobre tópicos de tempo de desfechos. Em todas as situações, recomendamos que se comece com pelo menos um protocolo escrito, para classificar os sujeitos com base nas informações disponíveis. Por exemplo, o tempo de soroconversão pode ser mensurado como o ponto intermediário entre o último teste negativo e o primeiro positivo. Para eventos definidos inequivocamente, qualquer desvio de tempos reais da determinação do protocolo pode ser visto como um erro de mensuração (discutido no Cap. 9). Doenças temporalmente ambíguas, tais como cânceres ou condições vasculares, são consideradas, frequentemente, como ocorrendo na ocasião do diagnóstico, mas é aconselhável o uso de um período de intervalo mínimo, quando quer que um período longo de latência (sem diagnóstico, ou prodrômico) seja inevitável. Pode ser possível, algumas vezes, entrevistar sujeitos sobre o início mais precoce dos sintomas, mas tais recordações e sintomas podem estar sujeitos a erros consideráveis e à variabilidade entre as pessoas.

Alguns eventos com determinação ambígua do tempo são administrados com definições padrão, mesmo sendo um tanto arbitrárias. Por exemplo, em 1993, o início da síndrome de imunodeficiência adquirida (aids) foi redefinido como a ocorrência de qualquer moléstia ou evento clínico definidor de aids (p. ex., contagem de CD4  $<200/\mu\text{L}$ ). Como um segundo exemplo, a perda de tempo de seguimento é tomada, convencionalmente, como o ponto médio entre a última tentativa bem sucedida e a primeira tentativa mal sucedida de fazer contato. Qualquer dificuldade em determinar um tempo arbitrariamente definido de um evento é tratada, então, como um problema de mensuração, que pode ser abordado pelos métodos descritos no Capítulo 19. Deve-se reconhecer, entretanto, que a arbitrariedade da definição para o tempo de um evento representa outra fonte de erro de mensuração, com consequência de potenciais vieses, que serão discutidas no Capítulo 9.

## CUSTOS

Os estudos de coorte geralmente são grandes empreendimentos. A maioria das doenças afeta apenas uma proporção pequena de uma população, mesmo que seja acompanhada por muitos anos. Obter estimativas estáveis de incidência requer um número substancial de casos de doença e, portanto, o

pessoa-tempo que dá origem aos casos também deve ser de bom tamanho. Pessoa-tempo suficiente pode ser acumulada acompanhando-se coortes por um longo período de tempo. Algumas coortes com exposições especiais (p. ex., japoneses vítimas das bombas atômicas [Beebe, 1979]), ou com histórias médicas e pessoais detalhadas (p. ex., o estudo de coortes de Framingham, Massachusetts [Kannel e Abbott, 1984]), realmente têm sido acompanhadas por décadas. Entretanto, se for pretendido que o estudo forneça resultados mais apropriados, o requisito pessoa-tempo pode ser atingido pelo aumento do tamanho das coortes. Naturalmente, estudos longos de grandes populações são dispendiosos. É comum que os estudos de coortes custem milhões de dólares, e gastos acima de US\$100 milhões têm ocorrido. A maior parte do custo origina-se da necessidade de se estabelecer um sistema contínuo para monitorar a ocorrência de doença em uma população grande.

Os custos dos estudos de coortes frequentemente limitam a viabilidade. Quanto mais baixa a incidência da doença, pior a viabilidade de um estudo de coortes, a menos que recursos públicos, dedicados a registros de saúde, possam ser empregados produtivamente (ver Cap. 23). A viabilidade ainda é mais prejudicada por um período longo de indução entre a causa hipotética e seu efeito. Um período longo de indução contribui para uma baixa incidência geral, por causa do tempo de seguimento adicional necessário para que sejam obtidos casos relacionados à exposição. Para detectar algum efeito, o estudo deve abranger um período tão longo quanto o período mínimo de indução e, na prática, deve abranger um período consideravelmente mais longo do que o período mínimo de indução. Estudos de coortes são pouco adequados para estudar o efeito de exposições que, hipoteticamente, causam doenças raras, com longos períodos de indução. Tais estudos de coortes são caros em relação à quantidade de informações alcançadas, o que significa que não são eficientes.

As despesas com estudos de coortes podem ser reduzidas de várias maneiras. Uma delas é usar um sistema existente para monitorar a ocorrência de doença (ver Cap. 23). Por exemplo, um registro regional de câncer pode ser utilizado para verificar a ocorrência da doença entre membros de uma coorte. Se a despesa da verificação de casos já está sendo garantida pelo registro, o estudo será consideravelmente mais barato.

Outra maneira de reduzir os custos é se basear em coortes históricas. Em vez de identificar membros da coorte concomitantemente com o início do estudo, e de planejar que o período de seguimento ocorra durante o mesmo, o pesquisador pode optar por identificar membros de coorte com base em registros de exposição prévia. O período de seguimento até a ocorrência da doença pode ser total ou parcialmente no passado. Para verificar casos que ocorreram no passado, os pesquisadores podem se basear em registros para averiguar a doença em membros da coorte. Se o período de seguimento começa antes do período durante o qual o estudo é conduzido, mas se estende para dentro do período da pesquisa, então vigilância ativa ou um novo sistema de monitoramento para verificar novos casos de doença, podem ser desenvolvidos.

Na medida em que a seleção de sujeitos ocorre depois do período de seguimento sob observação (algumas vezes chamada de seleção retrospectiva de coorte; ver Cap. 6), o estudo geralmente custa menos do que uma pesquisa equivalente, na qual a seleção de sujeitos acontece antes do período de seguimento (algumas vezes chamada de prospectiva). Uma desvantagem dos estudos de coortes retrospectivos é sua dependência de registros, que podem padecer de informações faltantes ou mal registradas. Quando tais carências estão relacionadas às variáveis em estudo, este pode sofrer vieses de seleção similares àqueles que podem ocorrer em estudos de caso-controle (ver Cap. 8). Por exemplo, se os registros são eliminados sistematicamente com a morte de um membro da coorte, então todo o pessoa-tempo retrospectivo será imortal e, portanto, deveria ser excluído.

Uma terceira maneira de reduzir custos é substituir uma das coortes, especificamente a coorte não exposta, por informações da população geral. Em vez de coletar novas informações de uma grande população não exposta, os dados existentes de uma população geral são usados para comparação. Esse procedimento tem várias desvantagens. Em primeiro lugar, só é razoável se houver alguma segurança de que apenas uma pequena proporção da população geral está exposta ao agente em estudo, como é



o caso, frequentemente, com exposições ocupacionais. Na medida em que parte da população geral está exposta, haverá erro por erro de classificação, que introduzirá um viés na comparação, o qual, ordinariamente, é na direção de subestimar o efeito (ver Cap. 9). Um outro problema é que as informações obtidas para a coorte exposta podem diferir em qualidade dos dados existentes para a população geral. Se dados de mortalidade são utilizados, o atestado de óbito é, frequentemente, a única fonte de informações para a população geral. Se informações médicas adicionais fossem usadas para classificar as mortes em uma coorte exposta, os dados obtidos assim não seriam comparáveis com aqueles da população geral. Essa falta de comparabilidade pode reduzir ou aumentar o viés nas comparações resultantes (Greenland e Robins, 1985a). Finalmente, é provável que a coorte exposta seja diferente da população em muitas maneiras que não são mensuradas, levando, assim, ao confundimento incontrolável nas comparações. O clássico “efeito do trabalhador sadio” é um exemplo desse problema, no qual o confundimento surge porque os operários têm que satisfazer um critério mínimo de saúde (devem ser capazes de trabalhar), o que não acontece com a população geral.

Uma quarta maneira de reduzir as despesas de um estudo de coorte é conduzir um estudo de caso-controle dentro da coorte, em vez de incluir a população inteira da coorte no estudo (Cap. 8). Tais estudos de caso-controle “aninhados” podem ser conduzidos, frequentemente, com uma fração do custo de um estudo de coorte, e ainda produzir os mesmos achados, com quase a mesma precisão.

## LOCALIZAÇÃO DE SUJEITOS

Os estudos de coorte que duram muitos anos apresentam problemas logísticos que podem afetar adversamente a validade. Seja o estudo retrospectivo ou prospectivo, frequentemente é difícil localizar pessoas, ou seus registros, muitos anos depois que tenham sido inscritos em estudos de coortes. Em estudos prospectivos, pode ser possível manter contato periódico com os sujeitos do estudo, e dessa forma manter informações atualizadas sobre sua localização. Tal procura acrescenta custos aos estudos de coorte prospectivos, mas a mobilidade crescente da sociedade demanda maiores esforços para localizar os sujeitos. Um número substancial de sujeitos perdidos no seguimento pode levantar sérias dúvidas sobre a validade do estudo. Seguintos que localizam menos de cerca de 60% dos sujeitos geralmente são vistos com ceticismo, mas até mesmo o seguimento de 70 a 80%, ou mais, pode ser baixo demais para dar garantia suficiente contra viés, se houver razão para se acreditar que a perda do seguimento pode estar associada à exposição ou à doença.

## COORTES DE EXPOSIÇÃO ESPECIAL E DA POPULAÇÃO GERAL

Um aspecto atraente dos estudos de coorte é a sua capacidade de prover estudos de uma variedade de efeitos possíveis sobre a saúde, a partir de uma exposição isolada. Um seguimento de mortalidade pode ser conseguido tão facilmente para todas as causas de óbito, como para qualquer causa específica. A vigilância de saúde para uma doença como ponto final pode, algumas vezes, ser expandida para incluir muitos ou todos os pontos finais, sem muito trabalho adicional. Um estudo de coorte pode fornecer um quadro abrangente do efeito sobre a saúde de uma dada exposição. Tentativas de se obter tal informação abrangente sobre exposições motivam a identificação de coortes de “exposição especial”, que são grupos identificáveis com exposição a agentes de interesse. Exemplos de tais coortes de exposição especial incluem coortes ocupacionais sujeitas a exposições no local de trabalho, estudos de pescadores ou fazendeiros expostos cronicamente à radiação solar, vítimas de bomba atômica e a população em torno de Chernobyl expostas à radiação ionizante, a população ao redor de Seveso, na Itália, exposta à contaminação ambiental com dioxina, adventistas do sétimo dia “expostos” a dietas vegetarianas e populações expostas a estresse por calamidades, tais como terremotos e ataques terroristas. Tais exposições não são comuns e requerem a identificação de coortes expostas para fornecer informações suficientes para o estudo.



Algumas vezes, exposições comuns são estudadas por meio de estudos de coortes que pesquisam um segmento de população identificado sem relação com o *status* de exposição. Tais coortes de “população geral” têm sido usadas para estudar os efeitos do fumo, do uso de contraceptivos orais, da dieta e da hipertensão. É mais eficiente limitar um estudo de coorte da população geral a exposições do que uma proporção substancial de pessoas que tenha experimentado a exposição; caso contrário, a coorte de não expostos será ineficientemente grande em relação à coorte de expostos. Uma população pesquisada pode ser classificada de acordo com tabagismo, consumo de bebidas alcoólicas, dieta, uso de drogas, história médica e muitos outros fatores de potencial interesse. Uma desvantagem é que a informação da exposição deve ser obtida, geralmente, por entrevistas com cada sujeito, o oposto da obtenção de informações a partir de registros, como é feito frequentemente em coortes de exposição especial.

# Estudos de caso-controle

Kenneth J. Rothman, Sander Greenland e Timothy L. Lash

## Elementos comuns dos estudos de caso-controle 138

- Pseudofrequências e razão de chances 138
- Definição da população-fonte 140
- Seleção de casos 141
- Seleção de controles 141
- Falácias comuns na seleção de controles 142
- Fontes para séries de controles 143
- Outras considerações para seleção de sujeitos 147

## Variantes do delineamento de caso-controle 150

- Estudos de caso-controle aninhados 150

- Estudos de caso-coorte 150
- Estudos de caso-controle de densidade 152
- Estudos caso-controle cumulativos (epidêmicos) 152
- Estudos de casos, caso-hipotético e caso-cruzado 153
- Amostragem em dois estágios 154
- Estudos de mortalidade proporcional 155
- Estudos de caso-controle com casos prevalentes 155

O uso e a compreensão dos estudos de caso-controle são parte dos desenvolvimentos metodológicos mais importantes da epidemiologia moderna. Conceitualmente, há ligações claras entre os experimentos randomizados e os estudos de coorte não randomizados, e entre estes e os estudos de caso-controle. Os estudos de caso-controle, no entanto, diferem suficientemente do paradigma científico da experimentação que uma abordagem cuja condição foi descuidada, a interpretação é um convite à concepção errônea. Neste capítulo, revemos os delineamentos de caso-controle e contrastamos suas vantagens e desvantagens com os delineamentos de coorte. Consideramos também variações do delineamento básico do estudo de caso-controle.

O saber convencional sobre estudos de caso-controle é que eles não geram estimativas de efeitos tão válidas quanto as medidas obtidas em estudos de coorte. Esse pensamento pode refletir incompreensões comuns na conceituação de estudos de caso-controle, que serão esclarecidas mais adiante; pode também refletir preocupações com informações de exposição e de seleção viesadas, nos estudos de caso-controle. Por exemplo, se as informações sobre exposição vêm de entrevistas, os casos, geralmente, terão relatado tais informações depois de conhecer seu diagnóstico. O diagnóstico pode afetar o relato de várias maneiras, por exemplo, ao melhorar a memória, assim aumentando a sensibilidade entre casos, ou ao provocar mais falsas memórias da exposição, reduzindo, assim, a especificidade entre os casos. Além disso, a própria doença pode turvar a memória e reduzir, assim, a sensibilidade. Esses fenômenos são exemplos de *viés de recordação*. Entretanto, a doença não pode afetar informações de exposição coletadas antes de sua ocorrência. Dessa forma, informações sobre exposição extraídas de registros criados antes que a doença ocorresse não estão sujeitas ao viés de recordação, não importa se o delineamento do estudo é de coorte ou de caso-controle.

Inversamente, os estudos de coorte não estão imunes a problemas que pensava-se ser particulares aos estudos de caso-controle. Por exemplo, enquanto um estudo de coorte pode reunir informa-

ções sobre exposição para uma população-fonte inteira no início do estudo, ele ainda requer localização posterior de sujeitos para se verificar variações de exposição e de desfechos. Se o sucesso dessa localização estiver vinculado à exposição e ao desfecho, o viés de seleção resultante se comportará de forma análoga à que é levantada, frequentemente, como uma preocupação nos estudos de caso-controle (Greenland, 1977; Cap. 12). Similarmente, os estudos de coorte às vezes usam memórias para reconstruir ou imputar história de exposição (avaliação retrospectiva) e são vulneráveis ao viés de recordação, se essa reconstrução for feita depois da ocorrência da doença. Assim, embora mais oportunidades para vieses de recordação e seleção possam surgir em estudos típicos de caso-controle do que em estudos de coorte prospectivos típicos, cada estudo deve ser considerado em detalhe para se avaliar sua vulnerabilidade ao viés, independentemente do delineamento.

O saber convencional defende também que os estudos de coorte são úteis para se avaliar a amplitude de efeitos relacionados a uma só exposição, ao passo que os estudos de caso-controle fornecem informações somente sobre a doença que aflige os casos. Esse pensamento conflita com a ideia de que os estudos de caso-controle podem ser vistos simplesmente como estudos de coorte mais eficientes. Da mesma forma que se pode optar por mensurar mais de um desfecho de doença em um estudo de coorte, é possível conduzir uma série de estudos de caso-controle aninhados na mesma população, usando vários desfechos de doença como a série de casos. O estudo de caso-coorte (ver adiante) é particularmente adequado para essa tarefa, permitindo que um grupo controle seja comparado com várias séries de casos. Independentemente de o delineamento de caso-coorte ser a forma de estudo de caso-controle utilizada ou não, os estudos de caso-controle não precisam ser caracterizados como limitados em relação ao número de desfechos de doença que podem ser estudados.

Para doenças que são bastante raras, os estudos de coorte tornam-se impraticáveis, e os estudos de caso-controle passam a ser a única alternativa útil. Entretanto, se a exposição é rara, os estudos de caso-controle usuais são ineficientes, e é preciso usar métodos que recrutem seletivamente sujeitos expostos adicionais, tais como os estudos de coorte especiais ou os delineamentos em dois estágios. Se tanto a exposição quanto o desfecho são raros, os delineamentos em dois estágios podem ser a única opção informativa, pois empregam sobreamostragem tanto de sujeitos expostos quanto de enfermos.

À medida que a compreensão dos princípios dos estudos de caso-controle tem progredido, a reputação desses estudos tem melhorado. No passado, era comum se ouvir os estudos de caso-controle sendo designados depreciativamente como estudos “retrospectivos”, um termo que se aplicaria apenas a alguns estudos de caso-controle e que se aplica também a alguns estudos de coorte (ver Cap. 6). Embora os estudos de caso-controle realmente apresentem mais oportunidades para viés e inferências errôneas do que os estudos de coorte, tais oportunidades aparecem como resultado da facilidade relativa com que um estudo de caso-controle pode ser montado. Visto que a condução de um estudo de caso-controle não precisa ser extremamente dispendiosa nem consumir muito tempo, muitos trabalhos têm sido conduzidos por pesquisadores ingênuos que não compreendem, nem implantam, os princípios básicos do delineamento de caso-controle válido. Ocasionalmente, tais pesquisas descuidadas podem produzir resultados valiosos, porém, frequentemente, errados, porque princípios básicos foram violados. A má reputação sofrida no passado pelos estudos de caso-controle origina-se mais de má conduta e de hiperinterpretação de resultados do que de qualquer fraqueza inerente à abordagem.

Idealmente, um estudo de caso-controle pode ser conceituado como uma versão mais eficiente de um estudo de coorte correspondente. Sob essa conceituação, os casos no estudo de caso-controle são os mesmos que seriam incluídos normalmente no estudo de coorte. Em vez de incluir toda a experiência da população-fonte que deu origem aos casos (o estudo base), como seria a prática usual em um delineamento de coorte, os controles são selecionados da população-fonte. Wacholder (1996) descreve esse paradigma do estudo de caso-controle como um estudo de coorte com dados faltantes, aleatoriamente e por delineamento. A amostragem de controles, a partir da população que deu origem aos casos, supre o ganho em eficiência de um delineamento de caso-controle sobre um delineamento de coorte. Os controles fornecem uma estimativa da prevalência da exposição e covariáveis na população-fonte. Quando os controles são selecionados entre membros da população que estão

em risco da doença no começo do período de seguimento do estudo, a razão de chances (*odds ratio*) do caso-controle estima a razão de riscos que seria obtida de um delineamento de coorte. Quando os controles são selecionados entre membros da população que não eram casos ao tempo em que cada caso ocorreu, ou de outra maneira proporcional ao pessoa-tempo acumulado pela coorte, a razão de chances do caso-controle estima a razão de frequência que seria obtida de um delineamento de coorte. Finalmente, quando os controles são selecionados entre membros da população que não eram casos ao fim do período de seguimento do estudo, a razão de chances do caso-controle estima a razão de chances de incidência que seria obtida de um delineamento de coorte. Com cada estratégia de seleção de controles, o cálculo da razão de chances é o mesmo, mas a medida de efeito estimada pela razão de chances difere. Delineamentos que implementam cada um desses paradigmas de seleção de controles serão discutidos depois dos tópicos que são comuns a todos os delineamentos.

## ELEMENTOS COMUNS DOS ESTUDOS DE CASO-CONTROLE

Em um estudo de coorte, o numerador e o denominador de cada frequência de doença (proporção, taxa ou chances de incidência) são medidos, o que requer enumeração da população inteira e vigilância – ou uso de um registro existente – para identificar casos além do período de acompanhamento. Um estudo de caso-controle válido observa a população de maneira mais eficaz utilizando uma série de controles, em vez de um cálculo completo dos denominadores das frequências de doença. Os casos em um estudo de caso-controle devem ser das mesmas pessoas que seriam consideradas casos em um estudo de coorte da mesma população.

### Pseudofrequências e razão de chances

O objetivo primário para a seleção de controles é que a distribuição da exposição entre os controles seja a mesma da população-fonte de casos. O motivo lógico para esse objetivo é que, se ele for alcançado, podemos usar a série de controles no lugar das informações do denominador em medidas de frequência de doenças, para determinar a razão da frequência de doença em pessoas expostas, relativa àquela entre pessoas não expostas. O objetivo primário será alcançado se pudermos amostrar controles da população-fonte, de modo que a razão do número de controles expostos ( $B_1$ ) para a experiência de expostos da população-fonte seja a mesma que a razão do número de controles não expostos ( $B_0$ ) para a experiência de não expostos da população-fonte, exceto erro amostral. Para a maioria dos propósitos, esse objetivo só precisa ser seguido dentro de estratos de fatores que serão utilizados para estratificação na análise, tais como fatores usados para restrição ou pareamento (Caps. 11, 15, 16 e 21).

Usando pessoa-tempo para ilustrar, o objetivo requer que  $B_1$  tenha a mesma razão da quantidade de pessoas-tempo expostas ( $T_1$ ) que a quantidade de pessoas-tempo não expostas ( $T_0$ ) de  $B_0$ , exceto erro amostral:

$$\frac{B_1}{T_1} = \frac{B_0}{T_0}$$

Aqui  $B_1/T_1$  e  $B_0/T_0$  são as taxas de amostragem dos controle, isto é, o número de controles selecionados por unidade de pessoa-tempo. Suponha que os casos expostos  $A_1$  e os casos não expostos  $A_0$  ocorram durante o período do estudo. Então, as taxas de expostos e de não expostos serão

$$I_1 = \frac{A_1}{T_1} \quad \text{e} \quad I_0 = \frac{A_0}{T_0}$$

Podemos usar as frequências de controles expostos e não expostos como substitutos para os denominadores reais das taxas, a fim de obtermos razões de caso-controle específicas para exposição, ou *pseudotaxas*:

$$\text{Pseudotaxa}_1 = \frac{A_1}{B_1}$$

e

$$\text{Pseudotaxa}_0 = \frac{A_0}{B_0}$$

Essas pseudotaxas não têm interpretação epidemiológica por si próprias. Suponha, entretanto, que as taxas de amostragem de controles  $B_1/T_1$  e  $B_0/T_0$  são iguais ao mesmo valor  $r$ , como seria esperado se os controles fossem selecionados independentemente da exposição. Se essa taxa de amostragem comum  $r$  é conhecido, as taxas reais de incidência podem ser calculadas por simples álgebra, porque, exceto erro amostral,  $B_1/r$  deve ser igual à quantidade de pessoas-tempo expostas na população-fonte, e  $B_0/r$  deve ser igual à quantidade de pessoas-tempo não expostas na população-fonte:  $B_1/r = B_1/(B_1/T_1) = T_1$  e  $B_0/r = B_0/(B_0/T_0) = T_0$ . Para obter as taxas de incidência, precisamos multiplicar cada pseudotaxa pela taxa de amostragem comum,  $r$ .

Se a taxa de amostragem comum não é conhecida, o que frequentemente é o caso, ainda podemos comparar os tamanhos das pseudotaxas por divisão. Especificamente, se dividirmos a pseudotaxa para expostos pela pseudotaxa para não expostos, obteremos

$$\frac{\text{Pseudotaxa}_1}{\text{Pseudotaxa}_0} = \frac{A_1/B_1}{A_0/B_0} = \frac{A_1/[(B_1/T_1)T_1]}{A_0/[(B_0/T_0)T_0]} = \frac{A_1/(r \cdot T_1)}{A_0/(r \cdot T_0)} = \frac{A_1/T_1}{A_0/T_0}$$

Em outras palavras, a razão das pseudotaxas para os expostos e para os não expostos é uma estimativa das taxas de incidência na população-fonte, desde que a taxa de amostragem dos controles seja independente da exposição. Assim, usando-se o delineamento de caso-controle, pode-se estimar a razão de taxa de incidência em uma população, sem obter informações sobre cada sujeito na população. Derivações semelhantes, na seção seguinte sobre variantes de delineamentos de caso-controle, mostram que se pode estimar a razão de riscos pela amostragem de controles entre aqueles em risco de doença no começo do período de acompanhamento (delineamento de caso-coorte), e que se pode estimar a razão de chances de incidência pela amostragem de controles entre os que não foram casos ao fim do período de seguimento (delineamento de caso-controle cumulativo). Com esses delineamentos, as pseudofrequências correspondem às proporções de incidência e às chances de incidência, respectivamente, multiplicadas pelas taxas de amostragem comuns.

Há uma penalidade estatística pelo uso de uma amostra dos denominadores, em vez de mensurar a experiência de pessoa-tempo de toda a população-fonte: a precisão das estimativas da razão de taxa de incidência de um estudo de caso-controle é menor do que a precisão de um estudo de coorte da população total que deu origem aos casos (a população-fonte). No entanto, a perda de precisão que se origina da amostragem de controles será bem pequena se o número de controles selecionados por caso for grande (geralmente quatro, ou mais). Além disso, a perda é balanceada pela economia do custo de não ter que obter informações sobre todos os membros da população-fonte. A economia de custos pode permitir ao epidemiologista aumentar a população-fonte e, assim, obter mais casos, resultando em uma estimativa geral melhor da razão de taxa de incidência, do que seria possível usando-se os mesmos gastos para conduzir um estudo de coorte.

A razão das duas pseudotaxas em um estudo de caso-controle é escrita, geralmente, como  $A_1B_0/A_0B_1$  e é chamada, às vezes, de *razão dos produtos cruzados*. A razão dos produtos cruzados em um estudo de caso-controle pode ser vista como a razão de casos para controles entre os sujeitos expostos ( $A_1/B_1$ ), dividida pela razão de casos para controles entre os sujeitos não expostos ( $A_0/B_0$ ). Essa razão também pode ser vista como a chance de ser exposto entre os casos ( $A_1/A_0$ ) dividida pela chance de ser exposto entre os controles ( $B_1/B_0$ ), em cujo caso ela é denominada de *razão de chances de exposição*. Embora qualquer das duas interpretações gere o mesmo resultado, enxergar a razão de chances como a razão das razões de caso-controle mostra mais diretamente como o grupo controle substitui as informações do denominador em um estudo de coorte e como a razão das pseudofrequências dá o mesmo resultado que a razão das taxas de incidência, proporção de incidências, ou razão de chances de incidência na população-fonte, se a amostragem for independente da exposição.

Um ponto que queremos enfatizar é que, *em nenhum momento* da discussão precedente tivemos que presumir que a doença sob estudo fosse “rara”. Em geral, a presunção de doença rara *não* é necessária em estudos de caso-controle. Da mesma forma que para estudos de coorte, contudo, não se esperaria que a razão de chances de incidência, nem a razão de taxas, fosse uma boa aproximação da razão de riscos, ou que fosse dobrável ao longo dos estratos de um fator de risco (mesmo que o fator não fosse um confundidor), a menos que a proporção de incidências fosse menor do que cerca de 0,1 para cada combinação da exposição e do fator (Cap. 4).

## Definição da população-fonte

Se os casos compõem uma amostra representativa de todos os casos em uma população precisamente definida e identificada, e os controles são amostrados diretamente dessa população-fonte, diz-se que se trata de um estudo *baseado em população* ou um estudo de base *primária*. Para um estudo de caso-controle baseado em população, a amostragem aleatória de controles pode ser factível se existir, ou puder ser compilado, um registro da população. Quando a amostragem aleatória da população-fonte de casos é viável, é geralmente a opção mais desejável.

A amostragem aleatória de pessoas não significa, necessariamente, que cada pessoa deva ter uma probabilidade igual de ser selecionada para ser um controle. Como explicado, se o objetivo é estimar a razão de taxas de incidência, então empregariamos amostragem longitudinal (densidade), na qual a probabilidade de uma pessoa ser selecionada para controle é proporcional ao tempo em risco dessa pessoa. Por exemplo, em um estudo de caso-controle aninhado dentro de uma coorte ocupacional, os trabalhadores de uma lista de empregados teriam sido acompanhados por extensões de tempo variáveis, e um esquema de amostragem aleatória refletiria esse tempo variável para estimar a razão de taxas de incidência.

Quando não é possível identificar a população-fonte explicitamente, a amostragem aleatória simples não é factível, e outros métodos de seleção de controles devem ser usados. Tais estudos são chamados algumas vezes de estudos de bases *secundárias*, porque a população-fonte é identificada secundariamente à definição de um mecanismo de encontro dos casos. Uma população-fonte secundária, ou “base secundária”, é, portanto, uma população-fonte que é definida a partir de (secundária a) uma dada série de casos.

Considere um estudo de caso-controle no qual os casos são pacientes tratados para psoríase grave, na Clínica Mayo. Esses pacientes que vêm para a Clínica Mayo são de todos os cantos do mundo. Qual é a população-fonte específica que dá origem a esses casos? Para responder, teríamos que saber exatamente quem iria à Clínica Mayo se tivesse psoríase grave. Não podemos enumerar essa população-fonte, porque muitas pessoas nela não sabem se iriam à Clínica Mayo para tratamento de psoríase grave, a menos que realmente desenvolvessem essa condição. Essa fonte secundária poderia ser definida como uma população espalhada ao redor do mundo, constituída por pessoas que iriam para a Clínica Mayo se desenvolvessem psoríase grave. Essa é a fonte secundária da qual a série de controle para o estudo idealmente seria extraída. O desafio para o pesquisador é aplicar critérios de elegibilidade aos casos e controles, de modo que haja uma boa correspondência entre os controles e essa população-fonte. Por exemplo, os casos de psoríase grave e seus controles poderiam ser restritos àqueles condados dentro de uma certa distância da Clínica Mayo, de modo que pelo menos uma correspondência geográfica entre os controles e a população-fonte secundária pudesse ser assegurada. Essa restrição, contudo, poderia deixar poucos casos para o estudo.

Infelizmente, o conceito de uma base secundária frequentemente está conectado de modo tênue às realidades subjacentes e pode ser altamente ambíguo. Para a psoríase, por exemplo, para uma pessoa ir à Clínica Mayo são necessários muitos fatores que variam com o tempo, tais como se a pessoa é encorajada pelo seu médico regular e se pode arcar com as despesas de deslocamento. Não está claro, então, como, ou mesmo se, seria possível definir a base secundária, quanto mais uma amostra da mesma, e, assim, não está claro se seria possível garantir que os controles fossem membros da base por ocasião da amostragem. Preferimos, portanto, conceituar e conduzir estudos de caso-controle com

uma população-fonte bem definida, e então identificar e recrutar casos e controles representativos da doença e da experiência de exposição daquela população. Quando, em vez disso, toma-se uma série de casos como ponto de partida, incumbe-se ao pesquisador demonstrar que uma população-fonte pode ser definida operacionalmente, para que o estudo seja reformulado e avaliado com relação a essa fonte. Considerações similares aplicam-se quando se toma uma série de controles como ponto de partida, como é feito algumas vezes (Greenland, 1985a).

## Seleção de casos

Idealmente, a seleção de casos constará de uma amostragem direta dos casos dentro de uma população-fonte. Portanto, exceto na amostragem aleatória, presume-se que todas as pessoas na população-fonte que desenvolvam a doença de interesse sejam incluídas como casos no estudo de caso-controle. Nem sempre é necessário, entretanto, incluir todos os casos da população-fonte. Os casos, como os controles, podem ser amostrados de forma aleatória para inclusão no estudo de caso-controle, desde que essa amostragem seja independente da exposição em estudo, dentro dos estratos de fatores que serão utilizados para estratificação na análise. Para tanto, suponha que consideremos apenas uma fração,  $f$ , de todos os casos. Se essa fração for constante durante a exposição, e os casos expostos  $A_1$  e os não expostos  $A_0$  ocorrerem na população-fonte, exceto erro amostral, a razão de chances do estudo será

$$\frac{A_1/B_1}{A_0/B_0} = \frac{fA_1/(r \cdot T_1)}{fA_0/(r \cdot T_0)} = \frac{A_1/T_1}{A_0/T_0}$$

como antes. Naturalmente, se menos do que todos os casos são amostrados ( $f < 1$ ), a precisão do estudo será mais baixa proporcionalmente a  $f$ .

Os casos identificados em uma só clínica, ou tratados por um só médico, são séries de casos possíveis para estudos de caso-controle. A população-fonte correspondente para os casos tratados em uma clínica inclui todas as pessoas que seriam atendidas naquela clínica e seriam registradas com o diagnóstico de interesse se elas tivessem a doença em questão. É importante especificar “se elas tivessem a doença em questão”, porque clínicas servem a populações diferentes para doenças diferentes, dependendo dos padrões de encaminhamento e da reputação da clínica em áreas específicas de especialidade. Como observado, sem uma população-fonte identificada com precisão, pode ser difícil, ou impossível, selecionar controles de modo a não haver viés.

## Seleção de controles

A definição da população-fonte determina a população da qual os controles são amostrados. Idealmente, a seleção envolverá amostragem direta dos controles a partir da população-fonte. Com base em princípios explicados, com relação ao papel da série de controles, podem ser formuladas muitas regras gerais para a seleção dos controles. Duas regras básicas são que:

1. Os controles devem ser selecionados da mesma população – a população-fonte – que dá origem aos casos. Se essa regra não pode ser seguida, é preciso que haja evidências sólidas de que a população fornecendo os controles tem uma distribuição de exposição idêntica àquela da população que é fonte dos casos, o que é uma exigência muito restritiva e raramente demonstrável.
2. Dentro dos estratos de fatores que serão usados para estratificação da análise, os controles devem ser selecionados independentemente de seu *status* de exposição, de modo que a taxa de amostragem para os controles ( $r$  na discussão prévia) não varie com a exposição.

Se essas regras e as regras correspondentes para os casos forem satisfeitas, então a razão de pseudofrequências, exceto erro amostral, será igual à razão da medida de doença correspondente na população-fonte. Se a taxa de amostragem é conhecida, então as medidas reais de frequência de



doença também podem ser calculadas. (Se as taxas de amostragem diferem para casos expostos e para não expostos, ou controles, mas são conhecidas, as medidas de frequência de doença e suas razões ainda podem ser calculadas, usando-se fórmulas de correção especiais; ver Caps. 15 e 19.) Para uma discussão mais detalhada dos princípios para seleção de controles em estudos de caso-controle, ver Wacholder e colaboradores (1992a, 1992b, 1992c).

Quando se deseja que os controles representem pessoa-tempo, a amostragem de pessoa-tempo deve ser constante ao longo das categorias de exposição. Esse requisito implica que a *probabilidade* de qualquer pessoa ser amostrada como um controle deva ser proporcional à quantidade de pessoa-tempo que o indivíduo passa em risco de doença na população-fonte. Por exemplo, se na população-fonte uma pessoa contribui duas vezes mais pessoa-tempo, durante o período de estudo, do que outra, a primeira pessoa deverá ter uma probabilidade duas vezes maior do que a segunda de ser selecionada como um controle. Essa diferença na probabilidade de seleção é induzida automaticamente por controles de amostragem, em uma taxa constante por unidade de tempo durante o período em que os casos são amostrados (amostragem *longitudinal* ou de *densidade*), e não pela amostragem de todos os controles em um ponto no tempo (tal como o começo ou o fim do período de seguimento). Com a amostragem longitudinal de controles, um membro da população, cuja presença é duas vezes mais longa do que a de outro, terá uma chance duas vezes maior de ser selecionado.

Se o objetivo do estudo é estimar um risco ou razão de taxas, seria possível para uma pessoa ser selecionada como controle e ainda assim permanecer elegível para ser um caso, de sorte que tal pessoa poderia aparecer no estudo tanto como um controle quanto como um caso. Essa possibilidade pode parecer paradoxal, ou errônea, contudo, está correta. Isso corresponde ao fato de que, em um estudo de coorte, um caso contribui tanto para o numerador quanto para o denominador da incidência estimada.

Suponha que o período de seguimento abranja 3 anos, e uma pessoa livre de doença no ano 1 seja selecionada como um controle potencial neste mesmo ano. Essa pessoa deve, em princípio, permanecer elegível para se tornar um caso. Suponha que esse controle então desenvolva a doença no ano 2, e agora se torne um caso no estudo. Como deveria tal pessoa ser tratada na análise? Visto que realmente desenvolveu a doença durante o período do estudo, muitos pesquisadores a contariam como um caso, mas não como um controle. Se o objetivo é possibilitar que a razão de chances do caso-controle forneça a estimativa da razão de chances da incidência, então essa decisão seria apropriada. Lembre-se, porém, que se um estudo de seguimento estava sendo conduzido, cada pessoa que desenvolvesse a doença contribuiria não apenas para o numerador do risco, ou taxa, de doença, mas também para as pessoas, ou pessoas-tempo, contabilizadas no denominador. Queremos que o grupo controle forneça estimativas do tamanho relativo dos denominadores das proporções de incidência, ou das taxas (coeficientes) de incidência, para os grupos comparados. Esses denominadores incluem todas as pessoas que mais tarde se tornem casos. Portanto, cada caso em um estudo de caso-controle deve ser elegível para ser um controle antes do tempo de início da doença, cada controle deve ser elegível para se tornar um caso desde o tempo de sua seleção como um controle, e uma pessoa selecionada como um controle, que mais tarde desenvolva a doença e seja selecionada como um caso, deve ser incluída no estudo, tanto como um controle quanto como um caso (Sheehee, 1962; Miettinen, 1976a, Greenland e Thomas, 1982; Lubin e Gail, 1984; Robins et al., 1986a). Caso seja pretendido que os controles representem pessoa-tempo e sejam selecionados longitudinalmente, argumentos similares mostram que uma pessoa selecionada como um controle deve permanecer elegível para ser selecionada como um controle novamente e, assim, poderia ser incluída na análise repetidamente como um controle (Lubin e Gail, 1984; Robins et al., 1986a).

## Falácias comuns na seleção de controles

Em estudos de coorte, a população do estudo é restrita a pessoas em risco para a doença. Alguns autores têm visto os estudos de caso-controle como se fossem estudos de coorte feitos às avessas,

chegando mesmo a descrevê-los como estudos *trohoc*\* (Feinstein, 1973). Sob esse ponto de vista, adiantou-se o argumento de que os estudos de caso-controle deveriam ser restritos àqueles em risco de exposição (i.e., àqueles com oportunidade de exposição). Excluir mulheres estéreis de um estudo de caso-controle sobre um efeito adverso de contraceptivos orais, e parear para duração de emprego em um estudo ocupacional, são exemplos de tentativas de controlar oportunidade de exposição. Se o fator usado para restrição (p. ex., esterilidade) não estiver relacionado com a doença, não será um confundidor, por conseguinte não produzirá benefício à validade da estimativa de efeito. Além disso, se a restrição reduzir o tamanho do estudo, a precisão da estimativa de efeito será reduzida (Poole, 1986).

Outro princípio usado algumas vezes em estudos de coorte é que tais estudos devem ser “limpos” no início do seguimento, incluindo apenas pessoas que nunca tiveram a doença. A má aplicação desse princípio ao delineamento de caso-controle sugere que o grupo controle deva ser “limpo”, incluindo apenas pessoas saudáveis, por exemplo. O surgimento de doença após o começo do período de seguimento não é motivo para se excluir sujeitos da análise de uma coorte, e tal exclusão pode levar a viés; semelhantemente, controles com doença que apareceu após a exposição não devem ser removidos da série de controle. No entanto, em estudos da relação entre fumo de cigarros e câncer colorretal, certos autores recomendaram que o grupo controle devesse excluir pessoas com pólipos do cólon, porque estes estão associados ao tabagismo e são precursores do câncer colorretal (Terry e Neugut, 1998). Tal exclusão, na verdade, reduz a prevalência da exposição nos controles para abaixo daquela na população-fonte de casos e, portanto, envies as estimativas de efeito para cima (Poole, 1999).

## Fontes para séries de controles

Os seguintes métodos para amostragem de controles se aplicam quando a população-fonte não pode ser enumerada explicitamente, de modo que a seleção aleatória não é possível. Todos esses métodos só devem ser implementados sujeitos às reservas sobre bases secundárias descritas.

### Controles da vizinhança

Se a população-fonte não pode ser enumerada, pode ser possível selecionar controles ao longo da amostragem de domicílios. Esse método não é direto. Em geral, uma lista geográfica de residências não está disponível, de modo que um esquema deve ser planejado para amostrar residências sem enumerá-las. Por conveniência, os pesquisadores podem amostrar controles que sejam individualmente pareados a casos da mesma vizinhança. Isto é, depois que um caso é identificado, um ou mais controles que residam na mesma vizinhança são recrutados para o estudo. Se a vizinhança está relacionada à exposição, o pareamento deve ser levado em consideração na análise (ver Cap. 16).

Controles de vizinhança são usados, frequentemente, quando os casos são recrutados de uma fonte conveniente, tal como de uma clínica ou hospital. Tal prática pode introduzir viés, entretanto, pois os vizinhos selecionados como controles podem não ser da mesma população-fonte que os casos. Por exemplo, se os casos são de um dado hospital, os controles da vizinhança podem incluir pessoas que não teriam sido tratadas no mesmo hospital, caso tivessem desenvolvido a doença. Se o tratamento no hospital, a partir do qual os casos foram identificados, estivesse relacionado com a exposição em estudo, então o uso de controles da vizinhança introduziria um viés. Como um exemplo extremo, suponha que o hospital em questão fosse da Administração de Veteranos dos Estados Unidos. Os pacientes nesses hospitais tendem a diferir de seus vizinhos em muitas maneiras. Uma delas, obviamente, é em relação ao histórico de serviço militar. A maioria dos pacientes em hospitais da Administração dos Veteranos terá servido nas forças armadas dos Estados Unidos, ao passo que apenas uma minoria de seus vizinhos terá feito isso. Essa diferença em história de vida pode levar a diferenças em histórias de exposições (p. ex., exposições associadas a combate ou a manuseio de

\* *Trohoc* = cohort (coorte em inglês) escrito às avessas.

armas). Para qualquer estudo em particular, a conveniência de se usar controles da vizinhança deve ser avaliada com relação às variáveis do estudo que sejam enfoque da pesquisa.

### ***Discagem aleatória de dígitos***

A amostragem de domicílios com base na seleção aleatória de números de telefone é feita com a intenção de simular a amostragem aleatória da população-fonte. A *discagem aleatória de dígitos*, como esse método tem sido denominado (Waksberg, 1978), oferece a vantagem de abordar todos os domicílios em uma área designada, mesmo aqueles com números de telefone não listados, por meio de uma simples ligação telefônica. Contudo, o método requer atenção considerável a detalhes e não garante seleção sem viés. Em primeiro lugar, a elegibilidade dos casos deve incluir residência em uma casa que tenha um telefone, de modo que casos e controles possam vir da mesma população-fonte. Em segundo, mesmo que o pesquisador possa implantar um método de amostragem, de modo que cada telefone tenha a mesma probabilidade de ser chamado, não haverá, necessariamente, a mesma probabilidade de contato com cada sujeito elegível para controle, porque os domicílios variam quanto ao número de pessoas que residem neles, à quantidade de tempo em que alguém está em casa e ao número de telefones operantes. Em terceiro, fazer contato com um domicílio pode requerer muitas chamadas em vários momentos do dia e em vários dias da semana, demandando trabalho considerável; muitos telefonemas podem ser necessários para se obter um sujeito-controle que preencha as características de elegibilidade específicas (Wacholder et al., 1992b). Em quarto, alguns domicílios utilizam secretárias eletrônicas, caixa de mensagens ou identificação de chamadas e podem não responder ou não retornar chamadas não solicitadas. Em quinto, a substituição de telefones fixos por celulares, em alguns domicílios, abala a suposição de que membros da população possam ser selecionados aleatoriamente por discagem aleatória de dígitos. Finalmente, pode ser impossível distinguir acuradamente números de telefones comerciais de residenciais, uma distinção necessária para se calcular a proporção de não respondentes.

Os controles por discagem aleatória de números geralmente são pareados aos casos por código de área (nos Estados Unidos, os três primeiros dígitos do número de telefone) e prefixo (os três dígitos seguintes ao código de área). No passado, o código de área e o prefixo estavam relacionados com o local da residência e com o tipo de telefone (serviço fixo ou móvel). Assim, se a localização geográfica, ou a participação em planos de telefonia móvel, fosse provavelmente relacionada à exposição, então o pareamento deveria ser levado em conta na análise. Mais recentemente, as companhias telefônicas dos Estados Unidos designaram códigos de área superpostos e permitiram que os assinantes mantivessem seu número de telefone ao se mudar dentro da região, de modo que a correspondência entre números de assinatura de telefone e localização geográfica diminuiu.

### ***Controles de bases hospitalares ou clínicas***

Como observado, a população-fonte em estudos de caso-controle de bases hospitalares ou em clínicas frequentemente não é identificável, porque ela representa um grupo de pessoas que seria tratado em uma determinada clínica ou um hospital se elas desenvolvessem a doença em questão. Em tais situações, uma amostra aleatória da população geral não corresponderá, necessariamente, a uma amostra aleatória da população-fonte. Se os hospitais ou as clínicas que provêm os casos para o estudo tratam apenas uma proporção pequena de casos na área geográfica, então os padrões de encaminhamento para o hospital ou para a clínica são importantes de se levar em consideração na seleção dos controles. Para esses estudos, uma série de controle compreende pacientes dos mesmos hospitais ou clínicas, pois estes podem fornecer uma estimativa de efeito com menos viés do que os controles da população geral (tais como aqueles obtidos das vizinhanças dos casos, ou por discagem aleatória de números). A população-fonte não corresponde à população da área geográfica, mas apenas às pessoas que buscariam tratamento no hospital ou na clínica se desenvolvessem a doença em estudo. Embora a população da área geográfica possa ser difícil, ou impossível, de

quantificar, ou mesmo de definir muito claramente, parece razoável esperar que outros pacientes de hospital ou clínica representem a população-fonte melhor do que controles da população geral. O maior problema com qualquer amostragem não aleatória de controles é a possibilidade de que eles não sejam selecionados independentemente da exposição na população-fonte. Os pacientes que são hospitalizados com outras doenças, por exemplo, podem não ser representativos da distribuição da exposição na população-fonte, porque a exposição está associada à hospitalização, a outras doenças ou a ambos os casos. Por exemplo, suponha que o estudo vise a avaliar a relação entre tabagismo e leucemia usando casos hospitalizados. Se os controles são pessoas que estão hospitalizadas com outras condições, muitas delas terão sido internadas por problemas associados ao fumo. Uma variedade de outros cânceres, assim como doenças cardiovasculares e respiratórias, está relacionada com o fumo. Assim, uma série de controle de pessoas hospitalizadas por doenças que não a leucemia incluiria uma proporção mais alta de fumantes do que a população-fonte dos casos de leucemia.

Limitar os diagnósticos dos controles a condições para as quais não haja indicação prévia de uma associação com a exposição melhora a série de controles. Por exemplo, em um estudo de fumo e de casos de leucemia hospitalizados, seria excluído qualquer um que fosse internado com uma doença sabidamente relacionada com tabagismo. Tal política de exclusão pode afastar a maioria dos controles potenciais, porque a doença cardiovascular, por si própria, representaria uma grande proporção dos pacientes hospitalizados. No entanto, mesmo algumas categorias diagnósticas comuns seriam suficientes para se encontrar sujeitos controles suficientes, de modo que as exclusões não prejudicariam o estudo por limitar o tamanho da série de controle. Na verdade, limitando-se o escopo dos critérios de elegibilidade, é razoável excluir categorias de controles em potencial, mesmo sob suspeita de que uma dada categoria pudesse estar relacionada à exposição. Se errada, o custo da exclusão é que a série de controle torna-se mais homogênea com relação ao diagnóstico e talvez um pouco menor. Porém, se estiver correta, então a exclusão é importante para ultimar a validade do estudo.

Contudo, um pesquisador raramente pode ter certeza de que uma exposição não esteja relacionada a uma doença ou à hospitalização por um diagnóstico específico. Consequentemente, seria imprudente usar apenas uma categoria diagnóstica isolada como uma fonte de controles. O uso de uma variedade de diagnósticos tem a vantagem de diluir, potencialmente, os efeitos do viés de incluir um grupo diagnóstico específico que seja relacionado à exposição e permite o exame do efeito de se excluir certos diagnósticos.

A intenção de excluir uma categoria diagnóstica da lista de critérios de elegibilidade para identificação de controles é simplesmente melhorar a representatividade da série de controles com relação à população-fonte. Tal critério de exclusão não implica que deveria haver exclusões baseadas em história de doença (Lubin e Hartge, 1984). Por exemplo, em um estudo de caso-controle de fumo e de pacientes de leucemia hospitalizados, poderiam ser usados controles hospitalizados, mas seriam excluídos alguns que estivessem internados por causa de doença cardiovascular. Esse critério de exclusão para controles não implica que os casos de leucemia que tivessem doença cardiovascular seriam excluídos; somente se a doença cardiovascular fosse a causa da hospitalização, o caso deveria ser excluído. Para os controles, o critério de exclusão deve aplicar-se apenas à causa da hospitalização usada para identificar o sujeito do estudo. Uma pessoa que estivesse internada por causa de uma lesão traumática e que, assim, fosse elegível para ser um controle, não seria excluída se tivesse sido hospitalizada previamente por doença cardiovascular. A população-fonte inclui pessoas que tiveram doença cardiovascular, e elas devem ser incluídas na série de controle. Excluir tais pessoas levaria a uma subrepresentação do tabagismo relativa à população-fonte e produziria um viés para cima nas estimativas de efeito.

Se a exposição afeta diretamente a hospitalização (p. ex., se a decisão de hospitalizar baseia-se em parte na história de exposição), o viés resultante não pode ser remediado sem o conhecimento das taxas de hospitalização, mesmo se a exposição não estiver relacionada com a doença em estudo ou às doenças dos controles. Essa situação, de fato, foi um dos primeiros problemas de estudos de bases hospitalares a receber análise detalhada (Berkson, 1946) e é chamada, frequentemente, de viés

berksoniano; que será discutido mais sob os tópicos de viés de seleção (Cap. 9) e de viés de colisor (Cap. 12).

### **Outras doenças**

Em muitos cenários, especialmente em populações com registros de doença estabelecidos ou em bases de dados de reclamações de seguros, pode ser mais conveniente escolher controles a partir de pessoas que são diagnosticadas com outras doenças. As considerações necessárias para a seleção válida de controles a partir de outros diagnósticos são paralelas àquelas discutidas para controles hospitalares. É essencial excluir quaisquer diagnósticos que se saiba, ou se suspeite, estarem relacionados à exposição e, melhor ainda, incluir somente diagnósticos para os quais há alguma evidência indicando que não são relacionados à exposição. Esses critérios de exclusão e de inclusão só se aplicam ao diagnóstico que trouxe a pessoa para o registro ou para a base de dados, dos quais os controles foram selecionados. A história de uma doença relacionada à exposição não deve ser uma base para exclusão. Se, contudo, a exposição afeta diretamente a chance de entrada no registro ou na base de dados, o estudo estará sujeito ao viés berksoniano mencionado para estudos hospitalares.

### **Controles amigos**

Escolher amigos de casos como controles, da mesma forma que usar controles da vizinhança, é um delineamento que, inerentemente, usa o pareamento individual e precisa ser avaliado quanto às vantagens e desvantagens de tal pareamento (discutido no Cap. 11).

Afora as complicações do pareamento individual, há preocupações maiores originando-se do uso de controles amigos. Primeiramente, ser nominado como um amigo por um caso pode estar relacionado ao *status* de exposição do controle em potencial (Flanders e Austin, 1986). Por exemplo, os casos poderiam declarar preferencialmente como amigos seus conhecidos com os quais se engajam em atividades específicas, que poderiam se relacionar com a exposição. Atividade física, consumo de bebidas alcoólicas e exposição ao sol são exemplos de tais exposições. As pessoas que são mais solitárias podem ter menor probabilidade de ser nominadas como amigos, de modo que seus padrões de exposição serão subrepresentados em uma série de controle de amigos. Exposições mais comuns a pessoas extrovertidas podem se tornar hiper-representadas entre controles amigos. Esse tipo de viés foi suspeitado em um estudo de diabetes melito insulino-dependente, no qual os pais dos sujeitos dos casos identificaram os controles. Os casos tinham menos amigos do que os controles, tinham mais problemas de aprendizado e maior probabilidade de não gostar da escola. O uso de controles amigos poderia explicar esses achados (Siemiatycki, 1989).

Um segundo problema é que, diferentemente de outros métodos de seleção de controles, escolher amigos como controles cede muito da tomada de decisão sobre a escolha de sujeitos-controle aos casos ou seus representantes (p. ex., pais). O pesquisador que usa controles amigos geralmente pedirá uma lista de amigos e escolherá aleatoriamente a partir da lista, mas, para a criação da lista, ele dependerá totalmente dos casos ou de seus representantes. Essa dependência adiciona uma fonte potencial de viés ao uso de controles amigos, que não existe para outras fontes de controles.

Um terceiro problema é que o uso de controles amigos pode introduzir um viés que deriva da natureza superposta de grupos de amizade (Austin et al., 1989; Robins e Pike, 1990). O problema surge porque casos diferentes podem designar grupos de amigos que não são mutuamente exclusivos. Como resultado, as pessoas com muitos amigos serão hiper-representadas na série de controle, e quaisquer exposições associadas a tais pessoas também se tornam hiper-representadas (ver Cap. 11).

Em princípio, parear categorias deve formar uma separação mutuamente exclusiva e coletivamente exaustiva em relação a todos os fatores, tais como vizinhança e idade. Por exemplo, se o pareamento é por idade, viés devido a grupos de pareamento em superposição pode surgir de *viés de calibragem*, um termo que se refere a escolher controles que tenham um valor para o fator de pareamento dentro de uma faixa especificada do valor do caso. Assim, se o caso tem 69 anos de idade, seria possível parear controles que estivessem dentro de 2 anos da idade de 69. O viés de superposição pode ser evitado se são usadas

categorias de idade não superpostas para o pareamento. Assim, se o caso tem 69 anos de idade, poderiam ser escolhidos controles dentro da faixa etária de 65 a 69 anos. Na prática, entretanto, o viés devido à superposição de idade e categorias de vizinhança provavelmente é pouco importante (Robins e Pike, 1990).

### **Controles mortos**

Um controle morto não pode ser um membro da população-fonte de casos, porque a morte impede a ocorrência de uma nova doença. Suponha, contudo, que os casos estão mortos. A necessidade de comparabilidade é um argumento a favor do uso de controles mortos? Embora certos tipos de comparabilidade sejam importantes, escolher controles mortos representará mal a distribuição da exposição na população-fonte se a exposição causa ou previne óbitos em uma proporção substancial de pessoas ou se ela está associada a um fator não controlado que o faz. Se forem necessárias entrevistas e alguns casos estiverem mortos, será necessário usar respondentes substitutos para os casos mortos. Para aumentar a comparabilidade e evitar o problema de selecionar controles mortos, os respondentes substitutos também podem ser usados para controles vivos pareados aos casos mortos (Wacholder et al., 1992b). Todavia, a vantagem de informações comparáveis para casos e controles frequentemente é exagerada, como será abordado mais adiante. A principal justificativa para o uso de controles mortos é a conveniência, tal como em estudos baseados inteiramente em mortes (ver a discussão de estudos de mortalidade proporcional adiante e no Cap. 6).

## **Outras considerações para seleção de sujeitos**

### **Representatividade**

Alguns livros enfatizam a necessidade de representatividade na seleção de casos e controles. O conselho tem sido que os casos devem ser representativos de todas as pessoas com a doença, e que os controles devem ser representativos de toda a população não enferma. Tal conselho pode ser enganoso. Um estudo de caso-controle pode ser restrito a qualquer tipo de caso que possa ser de interesse: casos do sexo feminino, casos em idosos, casos gravemente enfermos, casos que morreram logo depois do início da doença, casos leves, casos de Filadélfia, casos entre operários de fábrica, e assim por diante. Em nenhum desses exemplos os casos seriam representativos de todas as pessoas com a doença, mas estudos de caso-controle perfeitamente válidos seriam possíveis para cada um (Cole, 1979). A definição de um caso pode ser bastante específica, contanto que ela tenha uma lógica sólida. A preocupação principal é a delineação clara da população que dá origem aos casos.

Geralmente, os controles devem representar a população-fonte para os casos (dentro de categorias das variáveis de extratificação), em vez de a população inteira de não enfermos. Esta população pode diferir vastamente da população-fonte para os casos em idade, em raça, em sexo (p. ex., se os casos vêm de um hospital da Administração de Veteranos), em *status* socioeconômico, em ocupação, e assim por diante – inclusive na exposição de interesse. Uma das razões para enfatizar as similaridades, em vez das diferenças, entre estudos de coorte e estudos de caso-controle é que numerosos princípios se aplicam a ambos os tipos de estudo, mas são mais evidentes no contexto de estudos de coorte. Em particular, muitos princípios relativos à seleção de sujeitos aplicam-se de forma idêntica a ambos os tipos de estudo. Por exemplo, é amplamente apreciado que os estudos de coorte possam ser baseados em coortes especiais, e não na população geral. Portanto, os estudos de caso-controle podem ser conduzidos por amostragem de casos e controles a partir dessas coortes especiais. Os controles resultantes devem representar a distribuição da exposição dentro daquelas coortes, e não da população geral, refletindo a regra mais geral de que os controles devem representar a população-fonte dos casos no estudo, e não a população geral.

### **Comparabilidade da acurácia de informação**

Alguns autores têm recomendado que informações obtidas sobre casos e controles devam ser de acurácia comparável ou igual para garantir que não haja um diferencial (distribuição igual) de erros de mensuração (Miettinen, 1985a; Wacholder et al., 1992a; MacMahon e Trichopoulos, 1996). A lógica



para esse princípio é a noção de que erros de mensuração não diferenciais enviesam a associação observada em direção à nulidade e assim não gerarão uma associação falsa, e que o viés em estudos com erro não diferencial é mais previsível do que em estudos com erro diferencial.

O princípio da comparabilidade de informações (acurácia igual) frequentemente é usado para guiar a seleção de controles e a coleta de dados. Por exemplo, ele é a base para o uso de respondentes substitutos em vez de entrevistas diretas com os controles vivos, quando se quer que a informação sobre os casos seja obtida de respondentes substitutos. Na maioria dos cenários, contudo, os argumentos para o princípio são logicamente inadequados. Um problema, discutido longamente no Capítulo 9, é que a não diferenciação do erro de mensuração de exposição está longe de ser suficiente para garantir que o viés será na direção da nulidade. Tais garantias requerem que os erros de exposição também sejam *independentes* de erros em outras variáveis, inclusive doença e confundidores (Chavance et al., 1992; Kristensen, 1992), uma condição que nem sempre é plausível (Lash e Fink, 2003b). Por exemplo, parece provável que as pessoas que escondem uso pesado de álcool também tendam a subrelatar outros comportamentos socialmente desaprovados, tais como fumo pesado, uso de drogas ilícitas, e assim por diante.

Outro problema é que os esforços para garantir acurácia igual de informações sobre exposição tenderão, também, a produzir acurácia de informações igual em outras variáveis. A direção do viés geral, produzido pelos erros não diferenciais resultantes nos confundidores e modificadores de efeito, pode ser mais larga do que o viés causado por erro diferencial de acurácia desigual das informações sobre exposição de casos e controles (Greenland, 1980; Brenner, 1993; Marshall e Hastrup, 1996; Marshall et al., 1999; Fewell et al., 2007). Além disso, a menos que a exposição seja binária, mesmo um erro não diferencial independente na mensuração de exposição não há garantia que será produzido viés na direção da nulidade (Dosemeci et al., 1990). Finalmente, mesmo quando o viés produzido por forçar acurácia de mensuração igual é na direção da nulidade, não há garantia de que tal viés seja menor do que aquele que teria resultado do uso de uma mensuração com erro diferencial (Greenland e Robins, 1985a; Drews e Greenland, 1990; Wacholder et al., 1992a). Por exemplo, em um estudo que usasse respondentes substitutos para casos, o uso de respondentes substitutos para os controles poderia levar a um viés maior do que o uso de entrevistas diretas com controles, mesmo que estas últimas resultassem em maior acurácia das mensurações dos controles.

O princípio da comparabilidade de informações (acurácia igual) só é aplicável, portanto, sob condições muito limitadas. Em particular, ele só pareceria útil quando confundidores e modificadores de efeito são mensurados com erro desprezível, e quando o erro de mensuração é reduzido pelo uso de fontes de informações igualmente acuradas. Caso contrário, o viés por forçar casos e controles a ter acurácia de mensuração igual pode ser tão imprevisível quanto o efeito de não fazê-lo e pode arriscar o erro diferencial (acurácia desigual).

### **Número de grupos controle**

Surgem situações nas quais o pesquisador pode encarar uma escolha entre dois ou mais grupos controle possíveis. Com frequência, haverá vantagens para um grupo que estarão faltando no outro, e vice-versa. Considere, por exemplo, um estudo de caso-controle baseado em uma série de casos hospitalizados. Visto que estão internados, os controles não seriam representativos da população-fonte, na medida em que a exposição esteja relacionada à hospitalização para as condições dos controles. Controles da vizinhança não sofreriam esse problema, mas poderiam ser não representativos de pessoas que iriam ao hospital se tivessem a doença do estudo. Então, que grupo controle é melhor? Em tais situações, alguns têm argumentado que mais de um grupo controle deva ser usado, em uma tentativa de abordar os vieses de cada grupo (Ibrahim e Spitzer, 1979). Por exemplo, Gutensohn e colaboradores (1975), em um estudo de caso-controle da doença de Hodgkin, usaram um grupo controle de cônjuges para controlar as influências ambientais durante a vida adulta, mas também usaram um grupo controle de irmãos para controlar o ambiente da infância e o sexo. Ambos os grupos de contro-



les são uma tentativa de representar a mesma população-fonte dos casos, mas têm vulnerabilidades diferentes a vieses de seleção e sobre confundidores potenciais diferentes.

O uso de múltiplos grupos de controle pode envolver trabalho considerável, por isso é mais a exceção do que a regra em pesquisas de caso-controle. Frequentemente, uma fonte de controles disponível é superior a todas as alternativas práticas. Em tais cenários, não se deve desperdiçar esforços coletando controles de fontes provavelmente viesadas. A interpretação dos resultados também será mais complicada, a menos que os grupos de controles diferentes gerem resultados similares. Se os dois grupos produzirem resultados diferentes, deveria ser encarado o problema de explicar as diferenças e tentar inferir qual foi a alternativa mais válida. Logicamente, então, o valor de se usar mais de um grupo controle é bastante limitado. Os grupos controle podem e devem ser comparados, mas a falta de diferença entre os grupos mostra apenas que ambos incorporam um viés semelhante. Uma diferença mostra apenas que pelo menos um é viesado, mas não indica qual o melhor ou qual o pior. Somente informações externas poderiam ajudar a avaliar a provável extensão do viés nas estimativas dos diferentes grupos-controle, e aquelas mesmas informações externas poderiam ter favorecido a seleção de apenas um dos grupos controle na fase de planejamento do estudo.

### **Momento de classificação e diagnóstico**

O Capítulo 7 discutiu extensamente alguns princípios básicos para classificação de pessoas de casos e de unidades pessoa-tempo de acordo com o *status* de exposição. Os mesmos princípios se aplicam aos casos e controles nos estudos de caso-controle. Se for pretendido que os controles representem pessoas-tempo (em vez de pessoas) na população-fonte, deve-se aplicar princípios para classificação de pessoa-tempo à classificação dos controles. Em especial, os princípios de classificação de pessoa-tempo levam à regra de que os controles devam ser classificados por seu *status* de exposição, no momento de sua seleção. Exposições acrescidas depois daquele tempo devem ser ignoradas. A regra que informações (tal como a história de exposição) sejam obtidas de maneira a permitir que se ignorem as exposições acrescidas depois do tempo de seleção. De modo semelhante, os casos devem ser classificados no momento do diagnóstico de início da doença, dando-se o desconto para qualquer período de latência embutido, ou para hipóteses de período de indução. Determinar o momento do diagnóstico, ou do início da doença, pode envolver todos os problemas e ambiguidades discutidas no capítulo anterior para os estudos de coorte e precisa ser resolvido pelo protocolo da pesquisa antes que as classificações sejam feitas.

Por exemplo, considere um estudo sobre uso de álcool e câncer de laringe, que também examinou o fumo como um confundidor e possível modificador de efeito, usou questionários aplicados por entrevistador para coleta de dados e utilizou controles da vizinhança. Para examinar o efeito de álcool e de fumo, ao presumir um período de latência de 1 ano (um tempo de indução de no mínimo 1 ano), o questionário teria de possibilitar a determinação dos hábitos de beber e de fumar até um ano antes do diagnóstico (para os casos), ou da seleção (para os controles).

O momento de seleção não precisa se referir à identificação do controle pelo pesquisador, mas, em vez disso, pode referir-se a um evento análogo ao momento de ocorrência para o caso. Por exemplo, o momento de seleção para controles que são casos de outras doenças pode ser tomado como o momento de diagnóstico para aquela doença; o momento de seleção de controles hospitalares pode ser tomado como o momento da internação. Para outros tipos de controles, pode não haver tal evento natural análogo ao momento de diagnóstico do caso, e o momento real de seleção terá que ser usado. Na maioria dos estudos, o momento de seleção precederá o momento em que os dados são coletados. Por exemplo, em estudos baseados em entrevistas, os controles podem ser identificados e, então, pode ocorrer um retardo de semanas ou de meses antes que as entrevistas sejam realizadas. Para evitar complicar as perguntas da entrevista, essa distinção frequentemente é ignorada, e os controles são indagados sobre hábitos em períodos anteriores à entrevista.

## VARIANTES DO DELINEAMENTO DE CASO-CONTROLE

### Estudos de caso-controle aninhados

Os epidemiologistas algumas vezes referem-se a estudos de caso-controle específicos como estudos caso-controle *aninhados*, quando a população na qual o estudo é conduzido é uma coorte completamente enumerada, o que permite que a amostragem aleatória formal de casos e controles seja efetuada. O termo normalmente é utilizado em referência a um estudo de caso-controle conduzido em um estudo de coorte, no qual informações adicionais (talvez a partir de exames dispendiosos) são obtidas na maioria, ou na totalidade, dos casos, mas, por economia, é obtida somente de uma fração dos membros remanescentes da coorte (os controles). Contudo, muitos estudos de caso-controle baseados em populações podem ser considerados aninhados em uma população-fonte enumerada. Por exemplo, quando há um registro de doença baseado na população e uma enumeração censitária da população servida pelo registro, pode ser possível usar os dados do censo para a amostragem aleatória dos controles.

### Estudos de caso-coorte

O *estudo de caso-coorte* é um estudo de caso-controle no qual a população-fonte é uma coorte, e (dentro de estratos de amostragem ou do pareamento) cada pessoa nessa coorte tem uma chance igual de ser incluída no estudo como um controle, independentemente de quanto tempo tenha contribuído para a experiência de pessoa-tempo da coorte ou de a pessoa ter desenvolvido a doença do estudo. Esse delineamento é uma maneira lógica de conduzir um estudo de caso-controle quando a medida de efeito de interesse é a razão de proporções de incidência, em vez da razão de taxas, como é comum em estudos perinatais. O risco médio (ou proporção de incidência) de adoecer durante um período especificado pode ser escrito

$$R_1 = \frac{A_1}{N_1}$$

para a subcoorte de expostos, e

$$R_0 = \frac{A_0}{N_0}$$

para a subcoorte de não expostos, onde  $R_1$  e  $R_0$  são as proporções de incidência entre os expostos e os não expostos, respectivamente, e  $N_1$  e  $N_0$  são os tamanhos iniciais das subcoortes de expostos e de não expostos. (Essa discussão aplica-se igualmente bem a variáveis de exposição com diversos níveis, mas, por simplicidade, consideraremos apenas uma exposição dicotômica.) Os controles devem ser selecionados de tal forma que a distribuição da exposição entre eles seja uma estimativa, sem viés, da distribuição na população-fonte. Em um estudo de caso-coorte, a distribuição que desejamos estimar está entre  $N_1 + N_0$  membros da coorte, não entre sua experiência de pessoa-tempo (Thomas, 1972; Kupper et al., 1975; Miettinen, 1982a).

O objetivo é selecionar controles a partir da coorte-fonte, de tal modo que a razão do número de controles expostos ( $B_1$ ) para o número de membros da coorte expostos ( $N_1$ ) seja a mesma que a razão do número de controles não expostos ( $B_0$ ) para o número de membros da coorte não expostos ( $N_0$ ), exceto o erro amostral:

$$\frac{B_1}{N_1} = \frac{B_0}{N_0}$$

Aqui,  $B_1/N_1$  e  $B_0/N_0$  são as frações de amostragem dos controles (o número de controles selecionados por membro da coorte). Exceto em caso de erro aleatório, essas frações de amostragem serão iguais, se os controles tiverem sido selecionados independentemente da exposição.

Podemos usar as frequências dos controles expostos e não expostos como substitutos para os denominadores reais das proporções de incidência, para se obter “pseudoriscos”:

$$\text{Pseudorisco}_1 = \frac{A_1}{B_1}$$

e

$$\text{Pseudorisco}_0 = \frac{A_0}{B_0}$$

Esses pseudoriscos não têm interpretação epidemiológica por si próprios. Suponha, entretanto, que as frações de amostragem dos controles sejam iguais à mesma fração,  $f$ . Então, exceto o erro amostral,  $B_1/f$  deveria ser igual a  $N_1$ , o tamanho da subcoorte dos expostos; e  $B_0/f$  deveria ser igual a  $N_0$ , o tamanho da subcoorte dos não expostos:  $B_1/f = B_1/(B_1/N_1) = N_1$  e  $B_0/f = B_0/(B_0/N_0) = N_0$ . Assim, para se obter as proporções de incidência, precisamos apenas multiplicar cada pseudorisco pela fração de amostragem comum,  $f$ . Se essa fração não for conhecida, ainda podemos comparar os tamanhos dos pseudoriscos por divisão:

$$\frac{\text{Pseudorisco}_1}{\text{Pseudorisco}_0} = \frac{A_1/B_1}{A_0/B_0} = \frac{A_1/[(B_1/N_1)N_1]}{A_0/[(B_0/N_0)N_0]} = \frac{A_1/fN_1}{A_0/fN_0} = \frac{A_1/N_1}{A_0/N_0}$$

Em outras palavras, a razão de pseudoriscos é uma estimativa da razão de proporções de incidência (razão de riscos) na coorte-fonte, se a amostragem dos controles for independente da exposição. Assim, usando-se um delineamento de caso-coorte, pode-se estimar a razão de riscos em uma coorte sem se obter informações sobre cada membro da coorte.

Até aqui, presumimos implicitamente que não há perda de seguimento, nem riscos competitivos, na coorte subjacente. Se houver, ainda é possível se estimar o risco ou as razões de taxa a partir de um estudo de caso-coorte, desde que tenhamos dados sobre o tempo passado em risco pelos sujeitos amostrados ou usemos certas modificações de amostragem (Flanders et al., 1990). Esses procedimentos requerem as suposições usuais para estimação de razão de taxas em estudos de coorte, ou seja, que as perdas de seguimento e os riscos competitivos não estejam associados com exposição ou com risco de doença.

Uma vantagem do delineamento de caso-coorte é que, a partir de uma só coorte, ele facilita a realização de um conjunto de estudos de caso-controle, todos os quais utilizam o mesmo grupo controle. Como uma amostra da coorte no arrolamento, o grupo controle pode ser comparado com qualquer número de grupos de casos. Se controles pareados são selecionados de pessoas em risco na ocasião em que surge um caso (como na amostragem por conjunto em risco que é descrita adiante), a série de controles deve ser adaptada a um grupo específico de casos. Se desfechos comuns precisam ser estudados, e se quer usar um só grupo controle para cada desfecho, um outro esquema de amostragem deve ser usado. A abordagem caso-coorte é uma boa escolha em tal situação.

Os delineamentos de caso-coorte têm outras vantagens, assim como desvantagens, em relação aos delineamentos alternativos de caso-controle (Wacholder, 1991). Uma desvantagem é que, por causa da superposição da condição de membro nos grupos de caso e controle (controles que são selecionados também podem desenvolver a doença e entrar no estudo como casos), é preciso selecionar mais controles em um estudo de caso-coorte do que em um estudo de caso-controle comum com o mesmo número de casos, se for desejável conseguir a mesma precisão estatística. Os controles extras são necessários porque a precisão estatística de um estudo é fortemente determinada pelos números de casos e de não casos distintos. Assim, se 20% dos membros da coorte fonte se tornarão casos, e se todos os casos serão incluídos no estudo, será necessário selecionar 1,25 vezes mais controles do que casos em um estudo de caso-coorte para garantir que haverá tantos controles que nunca se tornarão casos no estudo. Em média, apenas 80% dos controles em tal situação permanecerão não casos; os

outros 20% se tornarão casos. É claro que, se a doença é incomum, o número de controles extras necessários para um estudo de caso-coorte será pequeno.

## Estudos de caso-controle de densidade

Antes, descrevemos como as razões de chance do caso-controle estimarão as razões de taxas, se a série de controles for selecionada de modo que a razão dos denominadores de pessoa-tempo,  $T_1/T_0$ , for estimada validamente pela razão dos controles expostos para os não expostos,  $B_1/B_0$ . Isto é, para estimar razões de taxas, os controles deveriam ser selecionados de modo que a distribuição de exposição entre eles fosse, exceto um erro aleatório, a mesma que é entre a pessoa-tempo na população-fonte, ou dentro de estratos da população-fonte. Tal seleção de controles é chamada de *amostragem de densidade*, porque fornece dados para estimação de relações entre taxas de incidência, que têm sido chamadas de *densidades de incidência*.

Se a exposição de um sujeito pode variar com o tempo, então a história de exposição de um caso é avaliada até o tempo em que a doença ocorreu. A história de exposição de um controle é avaliada até um tempo índice análogo, tomado geralmente como o momento de amostragem; a exposição depois do momento de seleção deve ser ignorada. Essa regra ajuda a garantir que o número de controles expostos e de não expostos esteja em proporção à quantidade de pessoa-tempo de expostos e de não expostos na população-fonte.

O tempo durante o qual um sujeito é elegível para ser um controle deveria ser o tempo no qual também fosse elegível para se tornar um caso se a doença ocorresse. Assim, uma pessoa na qual a doença já tenha se desenvolvido, ou que tenha morrido, não é mais elegível para ser selecionada como um controle. Essa regra corresponde ao tratamento de sujeitos em estudos de coorte. Cada caso que é contabilizado no numerador de um estudo de coorte contribui para o denominador da taxa, até o tempo em que a pessoa se torna um caso, quando cessa a contribuição para o denominador. Uma maneira de implementar essa regra é escolher controles do grupo de pessoas na população-fonte que está em risco de se tornar um caso, no momento em que o caso é diagnosticado. Esse grupo algumas vezes é referido como o conjunto em risco para o caso, e esse tipo de amostragem de controles às vezes é chamado de amostragem por conjunto em risco. Os controles amostrados dessa maneira são pareados ao caso com relação ao momento de amostragem; assim, se o tempo está relacionado à exposição, os dados resultantes devem ser analisados como dados pareados (Greenland e Thomas, 1982). Também é possível conduzir amostragem de densidade não pareada usando-se métodos de amostragem de probabilidade se é conhecido o intervalo de tempo em risco para cada membro da população. Seleciona-se, então, um controle pela amostragem de membros com probabilidade proporcional ao tempo em risco e depois se amostra aleatoriamente um momento para medir a exposição dentro do intervalo em risco.

Como mencionado antes, uma pessoa selecionada como um controle que permaneça na população do estudo em risco depois da seleção, deverá continuar elegível para ser selecionada novamente como um controle. Assim, embora isso seja improvável em estudos típicos, a mesma pessoa pode aparecer no grupo controle duas ou mais vezes. Observe, entretanto, que incluir a mesma pessoa em ocasiões diferentes não leva necessariamente a que a informação de exposição (ou a de confundidor) seja repetida, porque essa informação pode mudar com o tempo. Por exemplo, em um estudo de caso-controle de uma epidemia aguda de doença intestinal, pode-se perguntar sobre alimentos ingeridos no dia ou em dias anteriores. Se um alimento contaminado foi uma causa da doença para alguns casos, então o *status* de exposição de um caso, ou controle, escolhido para o estudo há 5 dias, poderia ser bem diferente do que teria sido com 2 dias de estudo.

## Estudos caso-controle cumulativos (epidêmicos)

Em alguns cenários de pesquisa, os estudos de caso-controle podem abordar um risco que termina antes que comece a seleção de sujeitos. Por exemplo, um estudo de caso-controle de uma epidemia de doença diarreica após uma reunião social pode começar depois que todos os casos potenciais tenham

ocorrido (porque o tempo máximo de indução se esgotou). Em tal situação, um pesquisador poderia selecionar controles daquela parte da população que permaneceu depois da eliminação dos casos acumulados; isto é, selecionam-se os controles dentre os não casos (aqueles que continuaram não casos ao fim do seguimento da epidemia).

Suponha que a população-fonte é uma coorte, e que uma fração  $f$ , tanto de não casos expostos quanto de não expostos, seja selecionada para ser os controles. Então, a razão de pseudofrequências será

$$\frac{A_1/B_1}{A_0/B_0} = \frac{A_1/f(N_1 - A_1)}{A_0/f(N_0 - A_0)} = \frac{A_1/(N_1 - A_1)}{A_0/(N_0 - A_0)}$$

que é a razão de chances de incidência para a coorte. Essa razão fornecerá uma aproximação razoável da razão de taxa, desde que as proporções adoecendo em cada grupo de exposição, durante o período de risco, sejam baixas, isto é, menos de cerca de 20%, e que a prevalência de exposição permaneça relativamente constante durante o período de estudo (ver Cap. 4). Se o pesquisador preferir estimar a razão de riscos, em vez da razão de taxa de incidência, a razão de chances do estudo ainda pode ser usada (Cornfield, 1951), mas a acurácia dessa aproximação só é cerca de metade tão boa quanto aquela da aproximação da razão de chances da razão de taxas (Greenland, 1987a). O uso dessa aproximação no delineamento cumulativo é a base primária para o ensinamento errôneo de que é necessária uma suposição de doença rara para se estimar os efeitos dos estudos de caso-controle.

Antes da década de 1970, a conceituação padrão dos estudos de caso-controle envolvia o delineamento cumulativo, no qual os controles eram selecionados de não casos ao fim de um período de seguimento. Conforme discutido por numerosos autores (Sheehe, 1962; Miettinen, 1976a; Greenland e Thomas, 1982), os delineamentos de densidade e os de caso-coorte têm várias vantagens fora do cenário epidêmico agudo, inclusive, potencialmente, tem muito menos sensibilidade a viés por perda de seguimento relacionado à exposição.

## Estudos de casos, caso-hipotético e caso-cruzado

Há numerosas situações nas quais os casos são os únicos sujeitos usados para estimar ou testar hipóteses sobre efeitos. Por exemplo, algumas vezes é possível empregar considerações teóricas para construir uma distribuição prévia de exposição na população-fonte, e usar essa distribuição em lugar de uma série controle observada. Tais situações surgem naturalmente em estudos genéticos, nos quais as leis básicas de herança podem ser combinadas com certas suposições para derivar uma distribuição populacional ou parental específica de genótipos (Self et al., 1991). Também é possível estudar certos aspectos de efeitos conjuntos (interações) de fatores genéticos e ambientais, sem usar sujeitos controles (Khoury e Flanders, 1996); ver Capítulo 28 para detalhes.

Quando a exposição sob estudo é definida pela proximidade de uma fonte ambiental (p. ex., um fio de alta tensão), é possível construir um controle *especular* (hipotético) para cada caso, pela condução de um “experimento hipotético”. Ou o caso, ou a fonte de exposição, é hipoteticamente movido para outro local que seria igualmente provável se não houvesse efeito de exposição; a categoria de exposição do caso sob essa configuração hipotética é tratada, então, como a exposição do “controle” (pareada) para o caso (Zaffanella et al., 1998). Quando o controle especular surge pelo exame da experiência de exposição do caso fora do tempo, no qual a exposição poderia ser relacionada com a ocorrência de doença, o resultado é chamado de estudo de caso-cruzado (*case-crossover*).

O estudo cruzado (*crossover*) clássico é um tipo de experimento no qual dois (ou mais) tratamentos são comparados, como em qualquer estudo experimental. Em um estudo *crossover*, entretanto, cada sujeito recebe ambos os tratamentos, um seguindo o outro. Preferivelmente, a ordem na qual os dois tratamentos são aplicados é escolhida aleatoriamente para cada sujeito. Tempo suficiente deve

ser alocado entre as duas aplicações, de modo que o efeito de cada tratamento possa ser mensurado e possa regredir antes que o outro tratamento seja administrado. Um efeito persistente da primeira intervenção é chamado de *efeito carryover* (transportado de uma etapa para outra). Um estudo de caso-cruzado só é válido para estudar tratamentos para os quais os efeitos ocorrem dentro de um período de indução curto e não persistem, isto é, efeitos *carryover* devem estar ausentes, para que o efeito da segunda intervenção não se misture com o efeito da primeira.

O estudo *case-crossover* é um estudo análogo do caso-controle para o estudo *crossover* (Mackay, 1991). Para cada caso, um ou mais períodos de tempo pré-doença ou pós-doença são selecionados como períodos “controle” pareados para o caso. O *status* de exposição do caso no tempo do início da doença é comparado com a distribuição do *status* de exposição para aquela mesma pessoa nos períodos de controle. Tal comparação depende da suposição de que nem a exposição, nem confundidores, estão mudando com o tempo, de maneira sistemática.

Somente um grupo limitado de tópicos de pesquisa é tratável pelo delineamento de caso-cruzado. A exposição deve variar com o tempo para os indivíduos, em vez de ficar constante. Cor dos olhos, ou grupo sanguíneo, por exemplo, não poderiam ser estudados com um delineamento de caso-cruzado, porque ambos são constantes. Se a exposição não varia com a pessoa, então não há base para comparação de tempos de período de risco expostos e não expostos na pessoa. Como no estudo *crossover*, a exposição também deve ter um tempo de indução curto e um período transitório; caso contrário, exposições no passado distante poderiam ser a causa de um início de doença recente (um efeito *carryover*).

MacLure (1991) usou o delineamento de caso-cruzado para estudar o efeito da atividade sexual sobre o infarto do miocárdio incidente. Esse tópico é bem adequado para um delineamento de *case-crossover*, porque a exposição é intermitente, e, presume-se que tem um período de indução curto para o efeito hipotético. É presumível que qualquer aumento no risco de infarto do miocárdio consequente da atividade sexual seja classificado como um tempo curto depois desta. Um infarto do miocárdio é um desfecho que é bem adequado a esse tipo de estudo, porque se pensa que ele seja desencadeado por eventos em tempo próximo. Outras causas possíveis de um infarto do miocárdio, que poderiam ser estudadas por um estudo de caso-cruzado, seriam consumo de cafeína, consumo de álcool, exposição ao monóxido de carbono, exposições a drogas e exercício físico exaustivo (Mittleman et al., 1993), todos os quais ocorrem intermitentemente.

Cada caso e seu controle em um estudo de caso-cruzado são pareados automaticamente em todas as características (p. ex., sexo e data do nascimento) que não mudam nos indivíduos. A análise pareada dos dados de caso-cruzado faz o controle para todos os confundidores fixos, quer eles sejam mensurados ou não. Sujeito a suposições especiais, o controle para confundidores mensurados variáveis com o tempo pode ser possível usando-se métodos de modelagem para dados pareados (ver Cap. 21). Também é possível ajustar estimativas de caso-cruzado para o viés devido à tendência no tempo em exposição, por meio do uso de dados longitudinais de um grupo controle não doente (controles caso-tempo) (Suisa, 1995). Entretanto, esses ajustes de tendências dependem de suposições adicionais de não confundimento e podem introduzir viés se tais suposições não forem satisfeitas (Greenland, 1996b).

Há muitas variantes possíveis do delineamento *case-crossover*, dependendo de como os períodos de tempo de controle são selecionados. Essas variantes oferecem permutas entre potencial para viés, ineficiência e dificuldade de análise; ver Lumley e Levy (2000), Vines e Farrington (2001), Navidi e Weinhandl (2002) e Janes e colaboradores (2004, 2005), para discussão adicional.

## Amostragem em dois estágios

Outra variante do estudo de caso-controle utiliza amostragem em dois estágios, ou duas fases (Walker, 1982a; White, 1982b). Nesse tipo de estudo, a série de controle compreende um número relativamente grande de pessoas (possivelmente todas da população-fonte), das quais se obtêm informações de exposição ou, talvez, uma quantidade limitada de informações sobre outras variáveis relevantes. Então, somente para uma subamostra dos controles, são obtidas informações mais detalhadas sobre exposição ou sobre outras variáveis do estudo que podem precisar ser controladas na análise. Infor-



mações mais detalhadas também podem ser limitadas a uma subamostra dos casos. Essa abordagem em dois estágios é útil quando é relativamente barato obter as informações sobre exposição (p. ex., entrevistas por telefone), mas as informações sobre covariáveis são mais dispendiosas para se conseguir (p. ex., por exames de laboratório). Ela também é útil quando as informações sobre exposição já foram coletadas para toda a população (p. ex., histórias sobre o emprego para uma coorte ocupacional), mas são necessários dados de covariáveis (p. ex., genótipo). Essa situação surge em estudos de coorte, quando se precisa de mais informações além das coletadas na linha de base. Conforme será discutido no Capítulo 15, esse tipo de estudo requer métodos analíticos especiais, para se tirar vantagem total das informações coletadas em ambos os estágios.

## Estudos de mortalidade proporcional

Os estudos de mortalidade proporcional foram discutidos no Capítulo 6, onde foi assinalado que a validade de tais estudos pode ser melhorada, se forem delineados e analisados como estudos de caso-controle. Os casos são os óbitos ocorrendo na população-fonte. Os controles não são selecionados diretamente a partir da população-fonte, que é composta de pessoas vivas, mas são tirados de outras mortes na população-fonte. Essa série de controle é aceitável se a distribuição da exposição dentro desse grupo for similar àquela da população-fonte. Consequentemente, a série controle deve ser restrita a causas de morte que não sejam relacionadas com a exposição. Ver Capítulo 6 para uma discussão mais detalhada.

## Estudos de caso-controle com casos prevalentes

Os estudos de caso-controle às vezes baseiam-se em casos prevalentes, em vez de em casos incidentes. Quando é impraticável incluir apenas casos incidentes, ainda pode ser possível selecionar casos existentes da doença em um ponto no tempo. Se a razão de chances de prevalência na população for igual à razão de taxa de incidência, então a razão de chances de um estudo de caso-controle baseado em casos prevalentes pode estimar, sem viés, a razão de taxas. Como observado no Capítulo 4, entretanto, as condições necessárias para que a razão de chances de prevalência seja igual à razão de taxas são muito fortes, e não existe uma relação geral simples para razões específicas por idade. Se a exposição estiver associada à duração da doença ou à migração para fora da comunidade de prevalência, então um estudo de caso-controle baseado em casos prevalentes não pode, por si próprio, distinguir efeitos da exposição sobre a incidência da doença da associação da exposição com duração da doença, ou com a migração, a menos que se conheçam as forças das últimas associações. Se o tamanho da população exposta, ou da não exposta, muda com o tempo ou se há migração para dentro da comunidade de prevalência, a razão de chances de prevalência pode ser ainda mais afastada da razão de taxas. Consequentemente, é sempre preferível selecionar casos incidentes em vez de prevalentes ao se estudar a etiologia de uma doença.

Conforme discutido no Capítulo 3, os casos prevalentes normalmente são utilizados nos estudos de malformações congênitas. Em tais estudos, os casos verificados no nascimento são prevalentes, porque sobreviveram com a malformação desde o tempo de sua ocorrência até o nascimento. Seria mais útil, etiologicamente, verificar todos os casos incidentes, inclusive os abortos afetados, que não sobreviveram até o nascimento. Contudo, muitos desses não sobrevivem até que a verificação seja factível, e, assim, é praticamente inevitável que os estudos de caso-controle de malformações congênitas sejam baseados em casos prevalentes. Nesse exemplo, a população-fonte compreende todos os conceitos e os abortos, espontâneos ou provocados, representam emigração antes da data da verificação. Embora uma exposição não afete a duração de uma malformação, pode muito bem influir nos riscos de abortos espontâneos e provocados.

Outras situações nas quais os casos prevalentes são comumente usados são estudos de condições crônicas com tempos de início mal definidos e efeitos limitados sobre mortalidade, tais como obesidade, doença de Parkinson e esclerose múltipla, e estudos da utilização de serviços de saúde.



# Validade em estudos epidemiológicos

Kenneth J. Rothman, Sander Greenland e Timothy L. Lash

## Validade da estimação 156

### Confundimento 157

Confundimento como mistura de efeitos 157

Confundidores e substitutos de confundidores 158

Propriedades de um confundidor 159

Confundidores como fatores de risco externos à associação 159

Julgando o papel causal de um potencial confundidor 160

Critérios para um fator de confundimento 160

### Viés de seleção 163

Viés de autosseleção 163

Viés berksoniano 164

Distinguindo viés de seleção de confundimento 165

### Viés de informação 167

Erro de mensuração, erro de classificação e viés 167

Erro de classificação diferencial 168

Erro de classificação não diferencial 168

Erro de classificação de confundidores 175

As complexidades do erro de classificação simultânea 176

### Capacidade de generalização 177

## VALIDADE DA ESTIMAÇÃO

Uma estimativa epidemiológica é o produto final do delineamento, da condução do estudo e da análise de dados. Chamaremos todo o processo levando à estimativa (delineamento, condução do estudo e análise) de processo de estimação. A meta geral de um estudo epidemiológico pode então ser vista, geralmente, como acurácia na estimação. Mais especificamente, como descrito em capítulos anteriores, o objetivo de um estudo epidemiológico é obter uma estimativa válida e precisa da frequência de uma doença, ou do efeito de uma exposição sobre a ocorrência de uma doença na população-fonte do estudo. Inerente a esse objetivo é a visão de que a pesquisa epidemiológica é um exercício em mensuração. Frequentemente, um objetivo adicional é obter uma estimativa que seja generalizável a populações-alvo relevantes; esse objetivo envolve selecionar uma população-fonte para estudo que seja um alvo ou da qual se possa argumentar que experimenta efeitos semelhantes aos alvos.

Acurácia em estimação implica que o valor do parâmetro, que é o objeto da mensuração, seja estimado com pouco erro. Os erros na estimação são classificados tradicionalmente como *aleatórios* ou como *sistemáticos*. Embora erros aleatórios na amostragem e na mensuração de sujeitos possam levar a erros sistemáticos nas estimativas finais, princípios importantes do delineamento de estudos emergem da consideração separada de fontes de erros aleatórios e sistemáticos. Os erros sistemáticos em estimativas são referidos comumente como *vieses*; o oposto de viés é validade, de modo que uma estimativa que tenha pequeno erro sistemático pode ser descrita como *válida*. Analogamente, o oposto do erro aleatório é precisão, e uma estimativa com pequeno erro aleatório pode ser descrita como *precisa*. Validade e precisão são componentes da acurácia.

A validade de um estudo geralmente é separada em dois componentes: a validade das inferências realizadas, no que se refere aos membros da população-fonte (*validade interna*) e a validade das inferências pertinentes às pessoas fora daquela população (*validade externa ou capacidade de generalização*). A validade interna implica validade da inferência para a população-fonte de sujeitos do estudo. Em estudos de causalidade, ela corresponde à mensuração acurada de efeitos, não considerando a variação aleatória. Sob tal esquema, a validade interna é considerada um pré-requisito para a validade externa.

A maioria das violações da validade interna pode ser classificada em três categorias gerais: confundimento, viés de seleção e viés de informação, onde a última é o viés oriundo do erro de medida das variáveis do estudo. O confundimento foi descrito em termos gerais no Capítulo 4, ao passo que o viés de seleção específico e os problemas de mensuração o foram nos Capítulos 7 e 8. O presente capítulo descreve as formas gerais desses problemas em estudos epidemiológicos. O Capítulo 10 descreve como mensurar e limitar o erro aleatório, o Capítulo 11 aborda opções em delineamento que podem melhorar a acurácia geral e o Capítulo 12 mostra como vieses podem ser descritos e identificados pelo uso de diagramas causais. Depois de uma introdução à estatística nos Capítulos 13 e 14, os Capítulos 15 e 16 fornecem métodos básicos de ajuste para confundidores mensurados, enquanto o Capítulo 19 introduz métodos de ajuste para confundidores não mensurados, viés de seleção e erro de classificação.

A dicotomia da validade em componentes internos e externos poderia sugerir que a generalização é simplesmente uma questão de estender inferências sobre uma população-fonte para uma população-alvo. A seção final deste capítulo oferece uma visão diferente da capacidade de generalização, na qual a essência da generalização científica é a formulação de teorias abstratas (geralmente causais) que correlacionam as variáveis de um estudo. As teorias são abstratas no sentido de que não estão amarradas a populações específicas; em vez disso, elas se aplicam a um conjunto de circunstâncias mais geral do que as populações específicas em estudo. A validade interna em um estudo ainda é um pré-requisito para que o mesmo contribua de forma útil para esse processo de abstração, mas o processo de generalização é, afora isso, separado das preocupações com validade interna e com a mecânica de delineamento.

## CONFUNDIMENTO

O conceito de confundimento foi introduzido no Capítulo 4. Embora o confundimento ocorra na pesquisa experimental, ele é um tópico consideravelmente mais importante em estudos observacionais. Portanto, reveremos aqui os conceitos de confundimento e de confundidores, e então discutiremos outros assuntos sobre definição e identificação de confundidores. Como no Capítulo 4, nesta seção presumiremos que o objetivo é estimar o efeito que a exposição teve sobre os expostos na população-fonte. Esse é o efeito real (ou imaginado) da exposição. Indicaremos apenas brevemente como a discussão deveria ser modificada, ao se estimar efeitos contrafactuais (ou potenciais) de exposição, tais como o efeito que a exposição poderia ter sobre os casos não expostos. O Capítulo 12 examina o confundimento no contexto de diagramas causais, o qual não faz essas distinções explícitas.

### Confundimento como mistura de efeitos

No nível mais simples, o confundimento pode ser considerado uma confusão de efeitos. Especificamente, o efeito aparente da exposição de interesse é distorcido porque o efeito de fatores externos à associação é erradamente tomado como – ou misturado – com o efeito real da exposição (que pode ser nulo). A distorção introduzida por um fator de confundimento pode ser grande e pode levar à superestimação, ou subestimação de um efeito, dependendo da direção das associações que o fator de confundimento tenha com a exposição e com a doença. O confundimento pode modificar até mesmo a direção aparente de um efeito.

Uma definição mais precisa de confundimento começa por considerar a maneira pela qual os efeitos são estimados. Suponha que desejamos estimar o grau em que a exposição alterou a frequência de uma doença em uma coorte exposta. Para fazer isso, devemos estimar qual teria sido a frequência da doença nessa coorte caso a exposição tivesse estado ausente e comparar essa estimativa com a frequência observada sob exposição. Visto que a coorte foi exposta, essa ausência de exposição é *contrafactual* (contrária aos fatos), e assim, a frequência desejada de comparação com os não expostos não é observável. Então, em substituição, observamos a frequência de doença em uma coorte de não expostos. Raramente, porém, podemos tomar essa frequência em não expostos como representando bem qual teria sido a frequência na coorte de expostos se a exposição tivesse estado ausente, porque a coorte de não expostos pode diferir daquela de expostos em muitos fatores que afetam a frequência da doença, além da exposição. Para expressar esse problema, dizemos que o uso dos não expostos como um referencial para os expostos está *confundido*, porque a frequência da doença nos expostos difere daquela nos não expostos como resultado de uma mistura de dois ou mais efeitos, um dos quais é o efeito da exposição.

## Confundidores e substitutos de confundidores

Os fatores externos à associação que são responsáveis pela diferença na frequência da doença entre os expostos e os não expostos são chamados de *confundidores*. Além disso, fatores associados a esses fatores causais externos à associação, que podem servir como substitutos, são comumente chamados de confundidores. O exemplo mais extremo de tais substitutos é a idade cronológica. O aumento da idade está fortemente associado ao *envelhecimento* – acúmulo de mutações celulares e lesão tecidual que leva à doença – mas a idade crescente, por si só, não causa a maioria de tais alterações patogênicas (Kirkland, 1992), porque é apenas uma medida de quanto tempo passou desde o nascimento.

Não importa se um confundidor é uma causa da doença em estudo ou é meramente um substituto para tal causa, uma característica primária é que, se ele for perfeitamente mensurado, será preditivo da frequência da doença dentro da coorte de não expostos (de referência). Caso contrário, o confundidor não pode explicar por que a coorte de não expostos deixa de representar apropriadamente a frequência de doença que a coorte de expostos experimentaria na ausência de exposição. Por exemplo, suponha que todos os expostos foram homens e todos os não expostos foram mulheres. Se os homens não expostos tiverem a mesma incidência que as mulheres não expostas, o fato de que todos os não expostos foram mulheres em vez de homens não poderia ser responsável por qualquer confundimento que estivesse presente.

Em uma visão simples, o confundimento só ocorre se efeitos externos à associação se misturarem com o efeito sob estudo. Observe, entretanto, que o confundimento pode acontecer mesmo se o fator em estudo não tiver efeito. Assim, “mistura de efeitos” não deve ser tomada para implicar que a exposição em estudo tem um efeito. A mistura de efeitos surge de uma associação entre a exposição e os fatores externos à associação, não importando se a exposição tem um efeito.

Como um outro exemplo, considere um estudo para determinar se consumidores de bebida alcoólica experimentam uma incidência maior de câncer oral do que os abstêmios. O fumo é um fator externo à associação que está relacionado com a doença entre os não expostos (fumar tem um efeito sobre a incidência de câncer oral entre os abstêmios). O fumo também está associado ao consumo de álcool, porque há muitas pessoas que são abstinentes em geral, retraindo o consumo de álcool, o tabagismo e, talvez, outros hábitos. Consequentemente, os consumidores de bebidas alcoólicas incluem entre si uma proporção de fumantes maior do que a que seria encontrada entre os abstêmios. Visto que o fumo aumenta a incidência de câncer oral, os consumidores de álcool terão uma incidência maior do que os não consumidores, bastante à parte de qualquer influência do consumo de álcool por si próprio, simplesmente como uma consequência da maior quantidade de fumantes entre os consumidores de bebidas alcoólicas. Assim, um efeito aparente do consumo de álcool é distorcido pelo efeito do fumo; o efeito do tabagismo se mistura com o efeito do álcool, na comparação de consumidores de álcool com abstêmios. O grau de viés ou de distorção depende da magnitude do efeito do fumo, da força da associação

entre álcool e tabagismo e da prevalência de fumantes entre os abstêmios que não têm câncer oral. Uma ausência de efeito do fumo sobre a incidência de câncer oral, ou a ausência de uma associação entre tabagismo e consumo de álcool, levaria a nenhum confundimento. O fumo deve estar associado tanto ao câncer oral quanto ao consumo de álcool, para ser um fator de confundimento.

### Propriedades de um confundidor

Em geral, uma variável deve estar associada tanto à exposição quanto à doença em estudo, para ser um confundidor. Contudo, tais associações não definem um confundidor, pois uma variável pode possuir essas associações e, mesmo assim, não o ser. Há várias maneiras pelas quais isso pode acontecer. A mais comum ocorre quando a exposição sob estudo tem um efeito. Nessa situação, qualquer fator correlacionado com aquela exposição também tenderá a estar associado à doença, como uma consequência de sua associação com a exposição. Por exemplo, suponha que o consumo frequente de cerveja esteja associado ao consumo de *pizza* e que o consumo frequente de cerveja fosse um fator de risco para câncer do reto. Seria o consumo de *pizza* um fator confundidor? À primeira vista, poderia parecer que a resposta é sim, porque o consumo de *pizza* estaria associado tanto à ingestão de cerveja quanto ao câncer retal. Mas, se o consumo de *pizza* estivesse associado ao câncer do reto apenas por causa de sua associação ao consumo de cerveja, isso não seria confundimento; de fato, a associação de consumo de *pizza* com câncer colorretal, então, seria consequência inteiramente do confundimento pelo consumo de cerveja. Um fator de confundimento deve estar associado à ocorrência de doença à *parte* de sua associação com a exposição. Em particular, como explicado anteriormente, a variável com potencial de confundimento deve estar associada à doença entre os indivíduos não expostos (de referência). Se o consumo de *pizza* estivesse associado ao câncer do reto entre não consumidores de cerveja, então poderia confundir. Caso contrário, se estivesse associado ao câncer retal somente por conta de sua associação com a ingestão de cerveja, não poderia ser um confundidor.

Análogo a essa restrição na associação entre um potencial confundidor e a doença, o potencial confundidor deve estar associado com a exposição entre a população-fonte para os casos. Para essa associação com exposição, é como se os efeitos do potencial confundidor se tornassem misturados com os efeitos da exposição. Em relação a isso, deve ser notado que um fator de risco que é independente da exposição na população-fonte pode se tornar (e é provável que se torne) associado à exposição entre os casos; assim, não se pode tomar a associação entre casos como uma estimativa válida da associação na população-fonte.

### Confundidores como fatores de risco externos à associação

Também é importante esclarecer o que queremos dizer com o termo *externo à associação* na frase “fator de risco externo à associação”. Esse termo significa que a associação do fator com a doença surge de uma via causal que não a que se está estudando. Especificamente, considere o diagrama causal

Fumo → pressão arterial elevada → doença do coração

onde as setas representam causação. A pressão arterial elevada é um fator de confundimento? Ela certamente é um fator de risco para a doença e é também correlacionada com a exposição, porque pode resultar do fumo. É até mesmo um fator de risco para a doença entre indivíduos não expostos, porque a pressão arterial elevada pode resultar de outras causas que não do fumo. Não obstante, ela não pode ser considerada um fator confundidor, porque o efeito do fumo é mediado pelo efeito da pressão arterial. Qualquer fator que represente um passo na cadeia causal entre exposição e doença não deve ser tratado como um fator confundidor externo à associação, mas, em vez disso, requer tratamento especial como um fator *mediador* (Greenland e Neutra, 1980; Robins, 1989; ver Cap. 12).

Finalmente, uma variável pode satisfazer todas as condições precedentes, mas pode não fazê-lo após controle para alguma outra variável de confundimento, e, assim, pode não mais ser um confun-

didor dentro dos estratos do segundo confundidor. Por exemplo, pode acontecer que (a) o primeiro confundidor não esteja mais associado à doença dentro dos estratos do segundo ou (b) o primeiro não esteja mais associado à exposição dentro dos estratos do segundo confundidor. Em qualquer dos casos, o primeiro confundidor é apenas um substituto para o segundo. De modo mais geral, o *status* de uma variável como um confundidor pode depender de que outras variáveis são controladas quando a avaliação é feita; em outras palavras, ser um confundidor é condicional sobre o que mais é controlado.

## Julgando o papel causal de um potencial confundidor

Considere o caso simples, porém comum, de uma variável de exposição binária, com interesse focado no efeito da exposição em uma população exposta em particular, em relação ao que teria acontecido se tal população não tivesse sido exposta. Suponha que uma população não exposta seja selecionada como o grupo de comparação (de referência). Um potencial confundidor é, então, um fator que está associado à doença entre os não expostos, e que não é afetado pela exposição, nem pela doença. Podemos verificar o último requisito, se soubermos que o fator precede a exposição e a doença. Associação com a doença entre os não expostos é um critério mais difícil de decidir. Afora os potenciais confundidores simples, e agora óbvios, tais como idade, sexo e hábito de fumar, os dados epidemiológicos disponíveis frequentemente são ambíguos quanto à previsibilidade, mesmo quando estabelecem a ordem no tempo. Decidir, simplesmente, se a previsibilidade se mantém com base em um teste estatístico é insensível demais para detectar todos os confundidores importantes e, em consequência, pode produzir estimativas altamente confundidas, como exemplos reais demonstram (Greenland e Neutra, 1980).

Uma resposta à ambiguidade e à insensibilidade dos métodos epidemiológicos para detectar confundidores é evocar outras evidências relativas ao efeito do potencial confundidor sobre a doença, inclusive dados não epidemiológicos (p. ex., clínicos ou sociais), e, talvez, teorias mecanicistas sobre os possíveis efeitos dos potenciais confundidores. Incertezas sobre a evidência ou o mecanismo podem justificar o manuseio de um fator potencial confundidor como confundidor e como não confundidor em análises diferentes. Por exemplo, ao se avaliar o efeito do café sobre doença do coração, não está claro como tratar os níveis séricos de colesterol. Os níveis elevados constituem um fator de risco e podem estar associados ao uso de café, mas o colesterol sérico pode mediar a ação do uso de café sobre o risco de cardiopatia. Isto é, o colesterol elevado pode ser um fator mediador na sequência etiológica em estudo. Se a ordem no tempo do uso de café e elevação do colesterol não puder ser determinada, podem-se conduzir duas análises, uma na qual o colesterol sérico está controlado (o que seria apropriado se o café não afetasse o colesterol do soro) e outra na qual o colesterol sérico não está controlado, ou é tratado como um mediador (o que seria mais apropriado se o café afetasse o colesterol sérico e não fosse associado a determinantes não controlados dos níveis de colesterol). A interpretação dos resultados dependeria de qual das teorias sobre o colesterol sérico estivesse correta. Grafos causais provêm um meio útil para ilustrar essas relações multivariáveis e, como será explicado no Capítulo 12, permitem a identificação de confundidores para controle da estrutura do grafo.

## Critérios para um fator de confundimento

Podemos fazer um resumo até aqui com a observação de que, para uma variável ser um confundidor, ela deve ter três características necessárias (mas não suficientes nem definidoras), as quais discutiremos em detalhes. Então, assinalaremos algumas limitações dessas características na definição e na classificação do confundimento.

1. *Um fator de confundimento deve ser um fator de risco externo à associação para a doença.*

Conforme mencionado antes, um fator de potencial confundimento não precisa ser uma causa real da doença, e se não for, deve ser um substituto para tal causa, que não seja a exposição. Essa condição implica que a associação entre o potencial confundidor e a doença deva ocorrer dentro de níveis da exposição do estudo. Em particular, um fator potencial de confundimento deve ser um fator

de risco dentro da categoria de referência da exposição em estudo. Os dados podem servir como um guia para a relação entre o potencial confundidor e a doença, mas é a relação real entre esses, e não a relação aparente observada nos dados, que determina se o confundimento pode ocorrer. Em estudos grandes, que estão sujeitos a menos erros amostrais, esperamos que os dados reflitam mais de perto a relação subjacente, porém, em estudos pequenos, os dados representam um guia menos confiável, e devem-se considerar outras evidências externas (“conhecimento prévio”) concernentes à relação do fator com a doença.

O exemplo seguinte ilustra o papel que o conhecimento anterior pode desempenhar na avaliação do confundimento. Suponha que em um estudo de coorte sobre fibras de vidro veiculadas pelo ar e câncer de pulmão, os dados mostrem mais fumo e mais cânceres entre os altamente expostos, mas nenhuma relação entre fumo e câncer de pulmão dentro das categorias de exposição. A última ausência de uma relação não significa que um efeito do fumo não tenha sido confundido (misturado) com o efeito estimado das fibras de vidro: pode ser que parte, ou o total, do excesso de câncer nos altamente expostos fosse produzido somente pelo tabagismo, e que a falta de uma associação fumo-câncer no estudo de coorte fosse produzida por um confundidor daquela associação não mensurado nessa coorte ou por erro aleatório.

Como um exemplo inverso, suponha que conduzimos um estudo de exposição à luz solar e melanoma. Nossa melhor informação atual indica que, após o controle para idade e a área geográfica de residência, não há relação entre número de inscrição no Seguro Social e ocorrência de melanoma. Assim, não consideraríamos o número do Seguro Social como um confundidor, independentemente de sua associação com melanoma na coorte de exposição de referência, porque pensamos que ele não é um fator de risco para melanoma nessa coorte, dada a idade e a área geográfica (i.e., achamos que os números de inscrição no Seguro Social não afetam as taxas de melanoma e não são fatores de risco para melanoma, exceto idade e área). Mesmo se o controle para número do Serviço Social mudasse a estimativa do efeito, a estimativa resultante do efeito seria menos acurada do que a que ignora o número do Seguro Social, dada nossa informação prévia sobre a falta de confundimento real por tal numeração.

No entanto, visto que a informação externa geralmente é limitada, os pesquisadores, frequentemente, confiam em seus dados para inferir a relação de potenciais confundidores da doença. Essa confiança pode ser racionalizada se houver um bom motivo para suspeitar que a informação externa não seja muito relevante para seu próprio estudo. Por exemplo, uma causa de doença em uma população não será relacionada causalmente à doença em outra população que careça de causas componentes complementares (i.e., fatores de suscetibilidade; ver Cap. 2). Uma discordância entre os dados e a informação externa sobre um fator de risco suspeito, ou conhecido, pode, portanto, sinalizar uma inadequação nos detalhes da informação sobre fatores de interação, e não um erro nos dados. Tal explicação pode ser menos crível para variáveis tais como idade, sexo e hábito de fumar, cuja relação com a doença se pensa, frequentemente, que seja bem estável nas populações. De modo paralelo, uma informação externa sobre a ausência de um efeito para um possível fator de risco pode ser considerada inadequada, se tal informação for baseada em estudos que tiveram um viés considerável na direção da nulidade.

*2. Um fator de confundimento deve estar associado com a exposição sob estudo, na população-fonte (a população em risco da qual os casos são derivados).*

Para produzir confundimento, a associação entre um fator potencial confundidor e a exposição deve estar na população-fonte dos casos do estudo. Em um estudo de coorte, a população-fonte corresponde à coorte do estudo, e, assim, esse pré-requisito implica apenas que a associação entre um fator de confundimento e a exposição exista entre sujeitos que compõem a coorte. Assim, em estudos de coorte, a associação exposição-confundidor pode ser determinada exclusivamente a partir dos dados do estudo, e, nem mesmo teoricamente, depende de conhecimento prévio, se nenhum erro de mensuração estiver presente.



Quando a exposição em estudo houver sido alocada de forma aleatória, algumas vezes se pensará, erroneamente, que o confundimento não possa ocorrer, porque a randomização garante que a exposição será independente de (não associada a) outros fatores. Infelizmente, essa garantia de independência é apenas *sobre a média*, por meio de repetições do procedimento de randomização. Em quase qualquer randomização isolada (alocação), inclusive naquelas em estudos reais, haverá associações casuais da exposição com fatores de risco externos à associação. Em consequência, o confundimento pode ocorrer, e ocorre, em ensaios randomizados. Embora esse confundimento casual tenda a ser pequeno nos grandes ensaios randomizados, frequentemente será grande em ensaios pequenos e em subgrupos pequenos de grandes ensaios (Rothman, 1977). Além disso, a falta de adesão, ou de concordância (falha em seguir o protocolo de tratamento designado) de modo intenso, ou a desistência de participação, pode resultar em confundimento não aleatório considerável, mesmo em grandes experimentos randomizados (ver Cap. 12, especialmente Fig. 12.5).

Em um estudo de caso-controle, a associação de exposição com o potencial confundidor deve estar presente na população-fonte que deu origem aos casos. Se a série de controle for grande, e não houver viés de seleção nem erro de mensuração, os controles fornecerão uma estimativa razoável da associação entre a variável de potencial confundimento e a exposição na população-fonte, e pode ser verificada em relação aos dados do estudo. Em geral, todavia, os controles podem não estimar adequadamente o grau de associação entre o potencial confundidor e a exposição na população-fonte que produziu os casos do estudo. Se a informação estiver disponível na associação dessa população, pode ser usada para ajustar os achados da série de controle. Infelizmente, informações externas confiáveis sobre as associações entre fatores de risco na população-fonte raramente estão disponíveis. Assim, em estudos de caso-controle, preocupações sobre o grupo-controle terão que ser consideradas, ao se estimar a associação entre a exposição e o fator potencialmente confundidor, por meio da análise de viés (Cap. 19).

Considere um estudo de caso-controle aninhado de exposição ocupacional à fibras de vidro veiculadas pelo ar e a ocorrência de câncer de pulmão. Esse estudo fez uma amostragem aleatória de casos e controles, a partir de casos e pessoas em risco, em uma coorte ocupacional. Suponha que conheçamos a associação entre exposição e fumo no total da coorte, como se essa informação estivesse registrada para a coorte inteira. Poderíamos, então, usar a discrepância entre a associação verdadeira e a associação exposição-fumo observada nos controles, como uma medida da extensão na qual a amostragem aleatória deixou de produzir controles representativos. Não importa o tamanho dessa discrepância, se não houve associação entre fumo e exposição na coorte-fonte, o tabagismo não seria um confundidor verdadeiro (mesmo que aparentasse sê-lo nos dados de caso-controle), e a estimativa não ajustada seria a melhor estimativa disponível (Robins e Morgenstern, 1987). De modo mais geral, poderíamos utilizar qualquer informação da coorte inteira para fazer ajustes na estimativa de caso-controle, de modo análogo aos estudos em dois estágios (Caps. 8 e 15).

### 3. *Um fator de confundimento não deve ser afetado pela exposição ou pela doença. Em particular, ele não pode ser um passo intermediário no caminho causal entre a exposição e a doença.*

Esse critério é satisfeito automaticamente se o fator preceder a exposição e a doença. Caso contrário, o critério requer informações fora dos dados. O pesquisador deve considerar evidências ou teorias relativas a se a exposição ou a doença poderiam afetar o fator. Se o fator é um passo intermediário entre a exposição e a doença, ele não deve ser tratado simplesmente como um fator de confundimento; em vez disso, é necessária uma análise mais detalhada, que leve em conta sua natureza mediadora (Robins, 1989; Robins e Greenland, 1992; Robins et al., 2000).

Embora as três características dos confundidores sejam tomadas, algumas vezes, para definir um confundidor, é um erro fazê-lo, tanto por motivos conceituais quanto técnicos. Confundimento é a confusão, ou mistura, de efeitos externos à associação com o efeito de interesse. As duas primeiras características são, simplesmente, consequências lógicas da definição básica, propriedades que um fator precisa satisfazer a fim de confundir. A terceira propriedade exclui situações nas quais os efeitos não podem ser esclarecidos de modo direto (exceto em casos especiais). Tecnicamente, é possível



que um fator possua as três características e, ainda assim, não tenha seus efeitos misturados com a exposição, no sentido de que um fator pode produzir excesso, ou déficit, não espúrio da doença entre os expostos, apesar de sua associação com a exposição e de seu efeito na doença. Esse resultado pode ocorrer, por exemplo, quando o fator é apenas um dentre vários potenciais confundidores, e o excesso de incidência produzido por ele entre os expostos é balanceado perfeitamente pelo excesso de incidência provocado por um outro fator nos não expostos.

A discussão acima omite numerosas sutilezas que surgem na determinação qualitativa de quais variáveis são suficientes controlar, a fim de eliminar o confundimento. Esses aspectos qualitativos serão discutidos usando-se diagramas causais, no Capítulo 12. Entretanto, é importante lembrar que o grau de confundimento é mais digno de preocupação do que sua mera presença ou ausência. Em um estudo, uma razão de taxas de 5 pode se tornar 4,6 depois do controle para idade, ao passo que, em outro estudo, uma razão de taxas de 5 pode mudar para 1,2 após esse mesmo controle. Embora idade seja um confundidor em ambos os estudos, no primeiro, o montante do confundimento é relativamente desprezível, ao passo que no último, o confundimento é responsável por quase toda a associação bruta. Métodos para avaliação quantitativa de confundimento serão descritos nos Capítulos 15 e 19.

## VIÉS DE SELEÇÃO

Os vieses de seleção são distorções que resultam de procedimentos usados para selecionar sujeitos e de fatores que influenciam a participação no estudo. O elemento comum entre tais vieses é que a relação entre exposição e doença é diferente para aqueles que participam e para aqueles que, teoricamente, seriam elegíveis para o estudo, inclusive para aqueles que não participam. Visto que estimativas de efeito estão condicionadas à participação no estudo, as associações observadas em um estudo representam uma mistura das forças que determinam a participação com as forças que determinam a ocorrência da doença.

O Capítulo 12 examina o viés de seleção no contexto de diagramas causais. Esses diagramas mostram que às vezes (mas nem sempre) é possível desembaraçar os efeitos da participação dos determinantes da doença, usando-se métodos padrão para o controle de confundimento. Empregar tal controle analítico requer, entre outras coisas, que os determinantes da participação sejam mensurados acuradamente e não sejam afetados pela exposição, nem pela doença. Contudo, se aqueles determinantes forem afetados pelos fatores do estudo, o controle de tais determinantes não corrigirá o viés e poderá mesmo torná-lo pior.

Algumas formas genéricas de viés de seleção em estudos de caso-controle foram descritas no Capítulo 8. Elas incluem o uso de grupos de controle incorretos (p. ex., controles compostos de pacientes com doenças que são afetadas pela exposição em estudo). Consideramos aqui alguns tipos adicionais.

### Viés de autosseleção

Uma fonte de viés comum é *autosseleção*. Quando o Centers for Disease Control (Centro para Controle de Doenças) investigou a incidência de leucemia entre soldados que tinham estado presentes no Teste Atômico Smoky, em Nevada (Caldwell et al., 1980), 76% da tropa identificada como membros daquela coorte tinham desfechos conhecidos. Desses, 82% foram localizados pelos pesquisadores, mas os outros 18% fizeram contato por sua própria iniciativa, em resposta à publicidade sobre a pesquisa. Esse autoencaminhamento de sujeitos é considerado, normalmente, uma ameaça à validade, porque as razões para o autoencaminhamento podem estar associadas ao desfecho em estudo (Criqui et al., 1979).

No estudo do Teste Atômico Smoky, houve 4 casos de leucemia entre os  $0,18 \times 0,76 = 14\%$  dos membros da coorte que se autoencaminharam, e 4 entre os  $0,82 \times 0,76 = 62\%$  de membros da coorte localizados pelos pesquisadores, para um total de 8 casos entre os 76% da coorte com desfecho conhecidos. Esses dados indicam que o viés por autosseleção foi um problema pequeno, porém

real, no estudo Smoky. Se os 24% da coorte com desfechos desconhecidos tivessem uma incidência de leucemia como aquela dos sujeitos localizados pelos pesquisadores, esperaríamos que somente  $4(24/62) = 1,5$ , ou cerca de 1 ou 2 casos ocorressem entre esses 24%, para um total de apenas 9 ou 10 casos em toda a coorte. Se, em vez disso, presumíssemos que os 24% com desfechos desconhecidos tivessem uma incidência de leucemia igual àquela dos sujeitos com desfechos conhecidos, calcularíamos que  $8(24/76) = 2,5$ , ou cerca de 2 ou 3 casos ocorressem entre esses 24%, para um total de 10 ou 11 casos no total da coorte. Poderia ser, entretanto, que todos os casos entre os 38% ( $= 24\% + 14\%$ ) da coorte que não foram localizados estivessem entre os autorrelatados, não deixando caso algum entre aqueles com desfecho desconhecido. O número total de casos na coorte inteira seria, então, de apenas 8.

A autosseleção também pode acontecer antes que os sujeitos sejam identificados para o estudo. Por exemplo, é rotineiro verificar que a mortalidade de trabalhadores em atividade é menor do que aquela da população como um todo (Fox e Collier, 1976; McMichael, 1976). Esse “efeito trabalhador-sadio” presumivelmente é derivado de um processo de rastreamento, talvez largamente por autosseleção, que permite pessoas relativamente saudas a se tornarem, ou permanecerem, trabalhadores, ao passo que aqueles que permanecem desempregados, aposentados, incapacitados ou, por qualquer outro motivo, fora da população de trabalhadores em atividade, sejam, como um grupo, menos saudáveis (McMichael, 1976; Wang e Miettinen, 1982). Embora o efeito trabalhador-sadio tenha sido classificado, tradicionalmente, como um viés de seleção, pode-se ver que não reflete um viés criado pelo condicionamento à participação no estudo, mas sim pelo efeito de outro fator que influencia tanto o *status* de trabalhador, como algum grau de saúde. Como tal, o efeito trabalhador-sadio é um exemplo de confundimento, e não um viés de seleção (Hernan et al, 2004), conforme será explicado mais adiante.

## Viés berksoniano

Um tipo de viés de seleção, que foi descrito primeiramente por Berkson (1946) (embora não no contexto de um estudo de caso-controle) e que veio a ser conhecido como *viés de Berkson*, ou *viés berksoniano*, ocorre quando tanto a exposição quanto a doença afetam a seleção e, especificamente, porque elas afetam a seleção. É paradoxal o porquê de isso poder gerar um viés para baixo, quando tanto a exposição quanto a doença aumentam a chance de seleção; esse viés para baixo pode induzir uma associação negativa no estudo se a associação na população-fonte for positiva, mas não tão grande quanto o viés.

Um exemplo marcante de viés berksoniano surgiu na controvérsia inicial sobre o papel de estrogênios exógenos em causar câncer do endométrio. Vários estudos de caso-controle tinham relatado uma associação forte, com um aumento no risco de cerca de 10 vezes para as mulheres tomando estrogênios regularmente por vários anos (Smith et al., 1975; Ziel e Finkle, 1975; Mack et al., 1976; Antunes et al., 1979). A maioria dos pesquisadores interpretou esse aumento de risco como uma relação causal, mas outros sugeriram que os estrogênios estavam meramente causando o diagnóstico dos cânceres, e não sua ocorrência (Horwitz e Feinstein, 1978). O argumento baseava-se no fato de que os estrogênios induzem sangramento uterino. Portanto, sua administração presumivelmente levaria as mulheres a buscarem assistência médica, assim proporcionando que uma variedade de condições ginecológicas fosse detectada. O viés resultante foi referido como um viés de detecção.

O remédio que Horwitz e Feinstein propuseram para o viés de detecção foi usar uma série de controle de mulheres com doenças ginecológicas benignas. Esses pesquisadores pensaram que as condições ginecológicas benignas também estariam sujeitas ao viés de detecção, e que, portanto, usar uma série de controle composta por mulheres com problemas benignos seria preferível a usar uma série de controle de mulheres com outras doenças malignas, com doenças não ginecológicas, ou sem doença, como estudos anteriores haviam feito. A falha nesse raciocínio foi a presunção incorreta de que os estrogênios causariam o diagnóstico de uma proporção substancial de cânceres do endométrio, que, do contrário, permaneceriam sem diagnóstico. Mesmo se a administração de estrogênio adiantasse a data do diagnóstico de câncer do endométrio, tal avanço no tempo do diagnóstico não

precisaria, por si próprio, levar a qualquer viés substancial (Greenland, 1991a). Possivelmente, uma pequena proporção de casos preexistentes de câncer do endométrio que, do contrário, não teriam sido diagnosticados viesse à atenção, mas é razoável supor que o câncer de endométrio que não é *in situ* (Horwitz e Feinstein excluíram os casos *in situ*) geralmente progride para causar sintomas que levem ao diagnóstico (Hutchison e Rothman, 1978). Embora um estágio inicial permanente, não progressivo, de câncer do endométrio seja uma possibilidade, os estudos que excluíram tais casos *in situ* da série de casos ainda encontraram uma associação forte entre administração de estrogênios e risco de câncer do endométrio (p. ex., Antunes et al., 1979).

O grupo-controle alternativo proposto compreendia mulheres com condições ginecológicas benignas que, se presumia, não causassem sintomas levando ao diagnóstico. Tal grupo forneceria uma superestimativa da proporção da população-fonte de casos expostos a estrogênios, porque sua administração realmente causaria o diagnóstico de uma proporção substancial das condições benignas. O uso de uma série de controle composta de condições ginecológicas benignas produziria, assim, um viés que subestimaria fortemente o efeito dos estrogênios exógenos sobre o risco de câncer do endométrio. Uma outra solução que Horwitz e Feinstein propuseram foi examinar a associação entre mulheres que tivessem se apresentado com sangramento vaginal, ou tivessem sido submetidas a tratamento para tal sangramento. Posto que tanto a exposição (estrogênios exógenos) quanto a doença (câncer do endométrio) aumentam fortemente o risco de sangramento, a restrição a mulheres com sangramento, ou em tratamento para sangramento, resulta em um viés berksoniano tão grave que poderia, facilmente, diminuir em 5 vezes o risco relativo observado (Greenland e Neutra, 1981).

Uma grande lição a ser aprendida dessa controvérsia é a importância de se considerar vieses de seleção quantitativamente, em vez de qualitativamente. Sem apreciação da magnitude potencial dos vieses de seleção, a escolha de um grupo controle pode resultar em um viés tão grande que uma associação forte seja obscurecida; alternativamente, uma associação desprezível poderia, com a mesma facilidade, ser exagerada. Os métodos para consideração quantitativa de vieses são discutidos no Capítulo 19. Uma outra lição é que se corre o risco de induzir, ou piorar, o viés de seleção, quando quer que se usem critérios de seleção (p. ex., exigir a presença ou a ausência de certas condições) que sejam influenciados pela exposição em estudo. Se tais critérios também estiverem relacionados com a doença em estudo, é provável que sobrevenha um grave viés berksoniano.

## Distinguindo viés de seleção de confundimento

Viés de seleção e confundimento são dois conceitos que, dependendo da terminologia, frequentemente se superpõem. Por exemplo, em estudos de coorte, os vieses resultantes de seleção diferencial no início do seguimento frequentemente são chamados de viés de seleção, porém, em nossa terminologia, eles são exemplos de confundimento. Considere um estudo de coorte comparando a mortalidade por doenças cardiovasculares entre estivadores e trabalhadores de escritórios. Se indivíduos com boa capacidade física se autosselecionassem para o trabalho de estivador, esperaríamos que os estivadores tivessem uma mortalidade cardiovascular mais baixa do que a dos funcionários de escritórios, mesmo que o trabalho de um estivador não tenha efeito sobre a mortalidade cardiovascular. Em consequência, a estimativa bruta de tal estudo não poderia ser considerada uma estimativa válida do efeito da labuta como estivador, em relação ao trabalho de escritório, sobre a mortalidade cardiovascular.

Suponha, entretanto, que a condição física de um indivíduo que se torna um estivador pudesse ser mensurada e comparada com a forma física dos trabalhadores de escritório. Se tal mensuração fosse feita acuradamente em todos os sujeitos, a diferença em forma física poderia ser controlada na análise. Assim, o efeito da seleção seria removido pelo controle dos confundidores responsáveis pelo viés. Embora o viés resulte da seleção de pessoas para as coortes, ele é, de fato, uma forma de confundimento.

Visto que mensurações de aptidão física ao ingressar em uma ocupação geralmente não estão disponíveis, os esforços do pesquisador em tal situação seriam focados na escolha de um grupo de

referência que experimentasse as mesmas forças de seleção que a ocupação-alvo. Por exemplo, Paffenbarger e Hale (1975) conduziram um estudo no qual compararam a mortalidade cardiovascular entre grupos de estivadores que se engajavam em níveis diferentes de atividade física no emprego. Paffenbarger e Hale presumiram que os fatores de seleção para entrar na ocupação fossem semelhantes para os subgrupos engajados em tarefas demandando atividade alta ou baixa, porque as distribuições de trabalho eram feitas após o ingresso na profissão. Esse delineamento reduziria, ou eliminaria, a associação entre boa forma física e se tornar um estivador. Ao comparar grupos com intensidades diversas de exposição dentro de uma ocupação (comparação interna), os epidemiologistas ocupacionais reduzem a diferença nas forças de seleção que acompanham as comparações por meio de categorias ocupacionais, e diminuem, assim, o risco de confundimento.

Infelizmente, nem todo viés de seleção em estudos de coorte pode ser tratado como se faz com o confundimento. Por exemplo, se a exposição afeta a perda do seguimento, e a perda do acompanhamento afeta o risco, ocorre viés de seleção, porque a análise está condicionada a uma consequência comum (permanecer em seguimento está relacionado tanto com a exposição quanto com o desfecho). Esse viés pode surgir em um estudo de mortalidade ocupacional, se a exposição causou o abandono precoce da ocupação pelas pessoas (p. ex., ser transferido de um trabalho ativo para um trabalho burocrático, ou ser aposentado), e isso, por sua vez, levou tanto à perda do seguimento quanto a um risco aumentado de morte. Aqui, não há variável basal (confundidor) criando diferenças de risco entre grupos expostos e não expostos; em vez disso, a própria exposição está gerando o viés. Tal tendenciosidade seria irremediável sem mais informações sobre os efeitos da seleção, e mesmo com tais informações, o viés não poderia ser removido pelo simples controle de covariáveis. Essa possibilidade reforça a necessidade de seguimento rigoroso em estudos de coorte, requerendo, geralmente, um sistema para vigilância de desfechos na coorte. Se não existe tal sistema em operação (p. ex., um sistema de reclamações de seguros), o estudo terá que implantar seu próprio sistema, o que pode ser dispendioso.

Em estudos de caso-controle, as preocupações sobre a escolha de um grupo controle têm como foco fatores que possam afetar a seleção e o recrutamento para o estudo. Embora fatores confundidores também devam ser considerados, eles podem ser controlados na análise, se forem mensurados. Se fatores de seleção que influenciam a escolha de casos e controles não são afetados pela exposição (p. ex., sexo), qualquer viés de seleção que eles produzam também pode ser dominado pelo controle desses fatores na análise. A chave, então, para evitar confundimento e viés de seleção devido a covariáveis pré-exposição, é identificar antecipadamente e mensurar tantos confundidores e fatores de seleção quanto seja praticável. Fazer isso requer bom conhecimento do assunto em pauta.

Em estudos de caso-controle, entretanto, os sujeitos frequentemente são selecionados depois que a exposição e o desfecho ocorrem e, por isso, há um potencial elevado para viés, devido a efeitos combinados de exposição de doença e efeitos sobre a seleção, como ocorreu nos estudos sobre estrogênio e câncer do endométrio, que restringiram sujeitos a pacientes com sangramento (ou a pacientes recebendo procedimentos médicos específicos para tratar o sangramento). Como será mostrado pelo uso de grafos causais (Cap. 12), o viés por tais efeitos de seleção conjuntos normalmente não pode ser resolvido com o controle de covariáveis básicas. Esse viés também pode surgir em estudos de coorte, e até mesmo em ensaios randomizados nos quais os sujeitos são perdidos no seguimento. Por exemplo, em um estudo de mortalidade ocupacional, a exposição poderia causar o afastamento precoce da pessoa de sua ocupação, o que, por sua vez, poderia produzir tanto falta de localização da pessoa (e sua consequente exclusão do estudo), quanto risco aumentado de morte. Essas forças resultariam em uma redução da chance de seleção entre os expostos, com uma diminuição mais alta entre os casos.

Nesse exemplo, não há covariável (confundidor) basal criando diferenças de risco entre grupos de expostos e de não expostos; em vez disso, a própria exposição está ajudando a gerar o viés. Tal viés seria irremediável sem informações adicionais sobre os efeitos da seleção, e mesmo com tais informações, ele não poderia ser removido pelo simples controle de covariáveis. Essa possibilidade reforça a necessidade de uma avaliação minuciosa do desfecho na população-fonte, em estudos de

caso-controle; se nenhum sistema de avaliação estiver em operação (p. ex., um registro de tumores para um estudo sobre câncer), o estudo deverá implantar seu próprio sistema.

Posto que muitos tipos de vies de seleção não podem ser controlados na análise, a prevenção desse tipo de vies pela seleção apropriada dos controles pode ser crítica. A estratégia usual para essa prevenção envolve tentar selecionar um grupo controle que esteja sujeito às mesmas forças seletivas que o grupo de casos, na esperança de que os vieses introduzidos pela seleção de controles cancelem aqueles introduzidos pela seleção dos casos nas estimativas finais. O cumprimento dessa meta, mesmo que aproximadamente, raramente pode ser garantido; no entanto, frequentemente essa é a única estratégia disponível para abordar preocupações sobre vies de seleção. Essa estratégia e outros aspectos da seleção de controles foram discutidos no Capítulo 8.

Em resumo, a seleção diferencial que ocorre antes da exposição e da doença leva a confundimento e pode ser controlada, assim, por ajustes para os fatores responsáveis pelas diferenças de seleção (ver, p. ex., os métodos de ajuste descritos no Cap. 15). Contudo, o vies de seleção como descrito, normalmente, em epidemiologia (assim como na literatura de delineamento experimental) surge da seleção afetada pela exposição em estudo e pode ficar além de qualquer ajuste prático. Entre esses vieses de seleção, podemos distinguir ainda o vies berksoniano, no qual tanto a exposição quanto a doença afetam a seleção.

Alguns autores (p. ex., Hernan et al., 2004) tentam usar grafos, a fim de fornecer uma base formal para separar o vies de seleção do confundimento, por meio da equiparação do vies de seleção com um fenômeno chamado de *vies de colisor*, uma generalização do vies berksoniano (Greenland, 2003a; Cap. 12). Nossa terminologia está mais de acordo com designações tradicionais, nas quais o vies de seleção pré-exposição é tratado como uma forma de confundimento. Essas distinções são mais discutidas no Capítulo 12.

## VIÉS DE INFORMAÇÃO

### Erro de mensuração, erro de classificação e vies

Uma vez que os sujeitos a serem comparados tenham sido identificados, deve-se obter as informações sobre eles para a análise. Vies ao estimar um efeito pode ser causado por erros na mensuração das informações necessárias. Tal vies frequentemente é chamado de *vies de informação*. A direção e a magnitude dependem fortemente de a distribuição de erros para uma variável (p. ex., exposição ou doença) estar ligada ao valor real da variável, aos valores reais de outras variáveis ou a erros na mensuração de outras variáveis.

Para variáveis discretas (variáveis com somente um número contável de valores possíveis, tais como indicadores para sexo), o erro de mensuração geralmente é chamado de *erro de classificação*. O erro de classificação que depende dos valores reais de outras variáveis é denominado *erro de classificação diferencial*. O erro de classificação que não depende dos valores reais de outras variáveis é chamado de *erro de classificação não diferencial*. O erro de classificação que depende dos erros em mensuração ou classificação de outras variáveis é designado como *erro dependente*; caso contrário, ele é chamado de *erro independente* ou *erro não dependente*. *Erro correlato* algumas vezes é usado como um sinônimo para erro dependente, mas, tecnicamente, se refere aos erros dependentes que tenham um coeficiente de correlação diferente de zero.

Muito da discussão seguinte lidará com o erro de classificação de variáveis binárias. Nessa situação especial, a *sensibilidade* de um método de mensuração de exposição é a probabilidade de que alguém que seja realmente exposto seja classificado como exposto pelo método. A *probabilidade falso-negativa* do método é a probabilidade de que alguém realmente exposto seja classificado como não exposto; ela é igual a 1 menos a sensibilidade. A *especificidade* do método é a probabilidade de que alguém que, na realidade é não exposto, seja classificado como não exposto. A *probabilidade falso-positiva* é a probabilidade de alguém que realmente seja não exposto ser classificado como exposto; ela é igual a 1 menos a especificidade. O *valor preditivo positivo* é a probabilidade de que

alguém que é classificado como exposto seja realmente exposto. Finalmente, o *valor preditivo negativo* é a probabilidade de que alguém que é classificado como não exposto seja verdadeiramente não exposto. Todos esses termos também podem ser aplicados para descrições dos métodos para classificação de doenças ou para classificar um potencial confundidor ou modificador.

## Erro de classificação diferencial

Suponha que um estudo de coorte é realizado para comparar taxas de incidência de enfisema entre fumantes e não fumantes. O enfisema é uma doença que pode passar não diagnosticada sem atenção médica especial. Se os fumantes, por causa de preocupação sobre efeitos do fumo relativos à saúde ou como uma consequência de outros efeitos do tabagismo sobre a saúde (p. ex., bronquite), buscarem atenção médica em um grau maior do que os não fumantes, o enfisema poderia ser diagnosticado mais frequentemente entre fumantes do que entre não fumantes, simplesmente em consequência de maior atenção médica. O fumo realmente causa enfisema, porém, a menos que fossem tomadas medidas para garantir um seguimento comparável, esse efeito seria superestimado: uma parte da incidência excessiva de enfisema não seria um efeito biológico do fumo, mas um efeito do fumo sobre a *deteção* do enfisema. Esse é um exemplo de erro de classificação diferencial, porque o subdiagnóstico de enfisema (falha em detectar casos verdadeiros), que é um erro de classificação, ocorre mais frequentemente em não fumantes do que em fumantes.

Em estudos de caso-controle de malformações congênitas, as informações às vezes são obtidas a partir de entrevistas com as mães. As mães-casos recentemente deram à luz um bebê malformado, ao passo que a grande maioria das mães-controles pariu recentemente um bebê aparentemente sadio. Outra variedade de erro de classificação diferencial, referida como *viés de recordação*, pode resultar se as mães dos bebês malformados lembrarem, ou relatarem, exposições reais, diferentemente das mães de bebês saudáveis (sensibilidade aumentada à recordação da exposição entre as mães-casos) ou, mais frequentemente, lembrarem ou relatarem o que, em verdade, não ocorreu (especificidade reduzida da recordação da exposição entre as mães-casos). Supõe-se que o nascimento de uma criança com malformação sirva como um estímulo para a mãe recordar e relatar todos os eventos que poderiam ter desempenhado algum papel no desfecho infeliz. Presumivelmente, tais mulheres lembrarão e relatarão exposições tais como doença infecciosa, trauma e medicações mais frequentemente do que as mães de bebês saudáveis, que não tiveram um estímulo comparável. Uma associação não relacionada com qualquer efeito biológico resultará desse viés de recordação.

O viés de recordação é uma possibilidade em qualquer estudo de caso-controle que dependa da memória dos sujeitos, porque os casos e os controles são, por definição, pessoas que diferem na sua experiência com doença ao tempo de sua lembrança, e essa diferença pode afetar a recordação e o relato. Klemetti e Saxen (1967) encontraram que a duração do tempo decorrido entre a exposição e a recordação foi um indicador importante da acurácia da recordação; estudos nos quais o tempo médio desde a exposição foi diferente para os casos e para os controles poderiam, assim, sofrer um erro de classificação diferencial.

O viés causado por erro de classificação diferencial pode exagerar ou subestimar um efeito. Em cada um dos exemplos supracitados, o erro de classificação normalmente exagera os efeitos sob estudo, mas exemplos inversos também podem ser encontrados.

## Erro de classificação não diferencial

O erro de classificação não diferencial da exposição ocorre quando a proporção de sujeitos mal classificados na exposição não depende do *status* do sujeito com relação a outras variáveis na análise, inclusive sobre a doença. O erro de classificação não diferencial da doença acontece quando a proporção de sujeitos mal classificados na doença não depende do *status* do sujeito no que diz respeito a outras variáveis em análise, incluindo a exposição.



O viés introduzido pelo erro de classificação não diferencial, independentemente de uma exposição binária ou doença, é previsível na direção, ou seja, tendendo para o valor de nulidade (Newell, 1962; Keys e Kihlberg, 1963; Gullen et al, 1968; Copeland et al., 1977). Por causa dos efeitos relativamente imprevisíveis do erro de classificação diferencial, alguns pesquisadores passam por procedimentos elaborados para garantir que o erro de classificação seja não diferencial, tais como cegar as avaliações de exposição com relação ao *status* do desfecho, na crença de que isso garantirá um viés tendendo para a nulidade. Infelizmente, mesmo em situações em que o cegamento é possível, ou em estudos de coorte em que os desfechos de doença ainda não ocorreram, o agrupamento de dados contínuos ou categóricos de exposição em menos categorias pode mudar o erro de classificação de não diferencial para diferencial (Flegal et al., 1991; Wacholder et al., 1991). Mesmo quando se consegue o erro de classificação não diferencial, ele pode vir à custa de aumento do viés total (Greenland e Robins, 1985a; Drews e Greenland, 1990).

Finalmente, como será discutido, o caráter não diferencial sozinho não garante viés na direção da nulidade. Ao contrário de conceitos populares equivocados, o erro de classificação não diferencial de exposição ou de doença pode produzir, algumas vezes, o afastamento do viés do nulo, se a variável da exposição ou de doença tiver mais de dois níveis (Walker e Blettner, 1985; Dosemeci et al., 1990), ou se os erros de classificação dependerem de erros feitos em outras variáveis (Chavance et al., 1992; Kristensen, 1992).

### **Erro de classificação não diferencial de exposição**

Como um exemplo de erro de classificação não diferencial, considere um estudo de coorte comparando a incidência de câncer de laringe entre consumidores de bebidas alcoólicas com a incidência em abstêmios. Presuma que os consumidores realmente tenham uma taxa de incidência de 0,00050 ano<sup>-1</sup>, ao passo que os abstêmios tenham uma taxa de incidência de 0,00010 ano<sup>-1</sup>, apenas um quinto da primeira. Presuma, também, que dois terços da população do estudo consistem em consumidores, mas somente 50% o admitem. O resultado é uma população na qual um terço dos sujeitos são identificados (corretamente) como consumidores, e têm uma incidência de doença de 0,00050 ano<sup>-1</sup>, mas os dois terços restantes da população são compostos por números iguais de consumidores e abstêmios, todos os quais são classificados como abstêmios, e entre os quais a incidência média seria de 0,00030 ano<sup>-1</sup> em vez de 0,00010 ano<sup>-1</sup>, (Tab. 9.1). A diferença de taxa foi reduzida por erro de

**TABELA 9.1**

**Efeito do erro de classificação não diferencial do consumo de álcool na estimação da diferença de taxa de incidência e a razão de taxas de incidência para câncer de laringe (dados hipotéticos)**

	Taxa de incidência ( $\times 10^5$ a)	Diferença de taxa ( $\times 10^5$ a)	Razão de taxa
Sem erro de classificação			
1.000.000 de consumidores	50	40	5.0
500.000 abstêmios	10		
Metade dos consumidores classificados como abstêmios			
500.000 consumidores	50	20	1.7
1.000.000 de "abstêmios" (50% na verdade são consumidores)	30		
Metade dos consumidores classificados como abstêmios e um terço dos abstêmios classificados como consumidores			
666.667 "consumidores" (25% na verdade são consumidores)	40	6	1.2
833.333 "abstêmios" (60% na verdade são bebedores)	34		



classificação de  $0,00040 \text{ ano}^{-1}$  para  $0,00020 \text{ ano}^{-1}$ , ao passo que a razão de taxas foi reduzida de 5 para 1,7. Esse viés na direção da nulidade resultou do erro de classificação não diferencial de alguns consumidores de álcool como abstêmios.

O erro de classificação pode ocorrer simultaneamente em ambas as direções; por exemplo, os abstêmios também poderiam ter sido classificados incorretamente como consumidores. Suponha que, além de a metade dos consumidores ser classificada equivocadamente como abstêmios, um terço dos abstêmios também fosse mal classificado como consumidores. As taxas de incidência resultantes seriam de  $0,00040 \text{ ano}^{-1}$  para aqueles classificados como consumidores, e  $0,00034 \text{ ano}^{-1}$  para os classificados como abstêmios. Assim, o erro de classificação adicional obscurece quase completamente a diferença entre os grupos.

Esse exemplo mostra como o viés produzido por erro de classificação não diferencial de uma exposição dicotômica será na direção da nulidade (de nenhuma relação), se o erro de classificação for independente de outros erros. Se o erro de classificação for bastante grave, o viés pode eliminar completamente uma associação e, até mesmo, reverter a direção da associação (embora a reversão só ocorra se o método de classificação for pior do que classificar as pessoas aleatoriamente como “expostas” ou “não expostas”).

Considere como um exemplo a Tabela 9.2. O painel do topo da tabela mostra os dados esperados de um estudo de caso-controle hipotético, com a exposição mensurada como uma dicotomia. A razão de chances é 3,0. Agora suponha que a exposição é mensurada por um instrumento (p. ex., um questionário) que resulte em uma medida de exposição que tenha 100% de especificidade, mas somente 80% de sensibilidade. Em outras palavras, todos os sujeitos realmente não expostos estão classificados corretamente como não expostos, mas há uma chance de 80% de que um sujeito exposto seja classificado corretamente como tal, e, portanto, uma chance de 20% de que um sujeito exposto seja classificado incorretamente como não exposto. Presumimos que o erro de classificação é não diferencial, o que significa, para esse exemplo, que a sensibilidade e a especificidade do método de mensuração da exposição são as mesmas para casos e controles. Também presumimos que não há erro na mensuração da doença, do que se segue, automaticamente, que os erros de exposição são independentes dos erros de doença. Os dados resultantes são exibidos no segundo painel da tabela. Com a diminuição da sensibilidade na mensuração da exposição, a razão de chances está viesada, em que seu valor esperado aproximado cai de 3,0 para 2,6.

No terceiro painel, presume-se que a especificidade da medida de exposição é 80%, de modo que há uma chance de 20% de que alguém que, em verdade seja não exposto, seja classificado incorretamente como exposto. Os dados resultantes produzem uma razão de chances de 1,9, em vez de 3,0. Em termos absolutos, mais da metade do efeito foi eliminada pela classificação errônea no terceiro painel: o excesso da razão de chances é de  $3,0 - 1 = 2,0$ , ao passo que ela é de  $1,9 - 1 = 0,9$  com base em 80% de sensibilidade e em 80% de especificidade, no terceiro painel.

O quarto painel da Tabela 9.2 ilustra que, quando a soma de sensibilidade e especificidade é igual a 1, a estimativa esperada resultante será nula, independentemente da magnitude do efeito. Se a soma da sensibilidade e especificidade é menor que 1, então a estimativa esperada resultante estará na direção oposta ao efeito real. O último painel da tabela mostra o resultado quando tanto a sensibilidade quanto a especificidade são iguais a zero. Essa situação equivale a rotular todos os sujeitos expostos como não expostos, e vice-versa. Ela leva a uma razão de chances esperada que é o inverso do valor correto. Tal erro drástico de classificação ocorreria se a codificação das categorias de exposição fosse invertida durante a programação por computador.

Conforme visto nesses exemplos, a direção do viés produzido pelo erro de classificação não diferencial, independente de uma exposição dicotômica, tende para a nulidade, e se o erro de classificação for extremo, o erro de classificação pode ir além da nulidade e reverter a direção. Com uma exposição que é mensurada pela sua divisão em mais de duas categorias, entretanto, o exagero de uma associação pode ocorrer como resultado de um erro de classificação não diferencial independente (Walker e Blettner, 1985; Dosemeci et al., 1990). Esse fenômeno é ilustrado na Tabela 9.3.

**TABELA 9.2****Erro de classificação não diferencial com duas categorias de exposição**

	Expostos	Não expostos
Dados corretos		
Casos-	240	200
controle	240	600
	RC = 3,0	
Sensibilidade = 0,8		
Especificidade = 1,0		
Casos-	192	248
controle	192	648
	RC = 2,6	
Sensibilidade = 0,8		
Especificidade = 0,8		
Casos-	232	208
controle	312	528
	RC = 1,9	
Sensibilidade = 0,4		
Especificidade = 0,6		
Casos-	176	264
controle	336	504
	RC = 1,0	
Sensibilidade = 0,0		
Especificidade = 0,0		
Casos-	200	240
controle	600	240
	RC = 0,33	

RC, razão de chances.

**TABELA 9.3****Erro de classificação não diferencial com três categorias de exposição**

	Não expostos	Baixa exposição	Alta exposição
Dados corretos			
Casos-	100	200	600
controle	100	100	100
		RC = 2	RC = 6
40% de alta exposição → 4 baixa exposição			
Casos-	100	440	360
controle	100	140	60
		RC = 3,1	RC = 6

RC, razão de chances.

Os dados esperados, corretamente classificados na Tabela 9.3, mostram uma razão de chances de 2, para baixa exposição, e 6, para alta exposição, em relação a nenhuma exposição. Agora suponha que há uma chance de 40% de que uma pessoa com alta exposição seja classificada incorretamente na categoria de baixa exposição. Se esse for o único erro de classificação e for não diferencial, os dados esperados seriam aqueles vistos no painel de baixo da Tabela 9.3. Observe que só muda a estimativa para alterações de baixa exposição; ele contém agora uma mistura de pessoas com baixa exposição, e pessoas com alta exposição, mas que foram designadas incorretamente para baixa exposição. Visto que as pessoas com alta exposição portam com elas o risco maior de doença associada à exposição, a estimativa de efeito resultante para baixa exposição está viesada para cima. Se alguns indivíduos com baixa exposição houvessem sido classificados incorretamente como tendo tido exposição alta, então a estimativa do efeito da exposição, para a categoria de alta exposição, seria viesada para baixo.

Esse exemplo ilustra que, quando a exposição tem mais de duas categorias, o viés do erro de classificação não diferencial da exposição para uma dada comparação pode ser afastado da nulidade. Quando a exposição é politômica (i.e., tem mais de duas categorias), e há um erro de classificação não diferencial entre duas das categorias e nenhuma outra, as estimativas de efeito para aquelas duas categorias serão viesadas em direção uma da outra (Walker e Blettner, 1985; Birkett, 1992). Por exemplo, o viés na estimativa de efeito para a categoria de baixa exposição na Tabela 9.3 é na direção da estimativa da categoria de alta exposição, afastando-se da nulidade. Também é possível, para erro de classificação não diferencial independente, enviesar estimativas de tendência para longe da nulidade ou inverter uma tendência (Dosemeci et al., 1990). Tais exemplos são incomuns, contudo, porque a inversão de tendência não pode ocorrer se a medida média da exposição aumentar com a exposição verdadeira (Weinberg et al., 1994d).

É importante notar que a presente discussão é pertinente a resultados *esperados* com um tipo particular de *método* de mensuração. Em um dado estudo, flutuações aleatórias nos erros produzidos por um método podem levar a estimativas que estão mais longe da nulidade do que estariam se nenhum erro estivesse presente, mesmo que o método satisfaça todas as condições que garantem o viés em direção da nulidade (Thomas, 1995; Weinberg et al., 1995; Jurek et al., 2005). O viés se refere apenas à direção *esperada*; se não sabemos quais foram os erros no estudo, na melhor hipótese só podemos dizer que a razão de chances observada está, provavelmente, mais próxima da nulidade do que ela estaria, se os erros estivessem ausentes. À medida que o estudo aumenta, diminui a probabilidade de que um resultado em particular se desvie substancialmente de sua expectativa.

### **Erro de classificação não diferencial de doença**

Os efeitos do erro de classificação não diferencial de doença são semelhantes àqueles do erro de classificação não diferencial de exposição. Na maioria das situações, o erro de classificação não diferencial de um desfecho binário de doença produzirá viés na direção da nulidade, desde que a má classificação seja independente de outros erros. Há, entretanto, alguns casos especiais nos quais o erro de classificação não produz viés na razão de riscos. Além disso, o viés na diferença de riscos é uma função simples da sensibilidade e da especificidade.

Considere um estudo de coorte no qual 40 casos ocorrem realmente entre 100 sujeitos expostos, e 20 casos ocorrem realmente entre 200 sujeitos não expostos. Então, a razão de riscos real é  $(40/100)/(20/200) = 4$ , e a diferença de riscos real é  $40/100 - 20/200 = 0,30$ . Suponha que a especificidade de detecção da doença é perfeita (não há falso-positivos), mas que a sensibilidade é de apenas 70% em ambos os grupos de exposição (i.e., a sensibilidade de detecção da doença é não diferencial e não depende de erros na classificação da exposição). Os números esperados detectados seriam, então,  $0,70(40) = 28$  casos expostos e  $0,70(20) = 14$  casos não expostos, o que gera uma estimativa de razão de riscos esperada de  $(28/100)/(14/200) = 4$ , e uma estimativa de diferença de riscos esperada de  $28/100 - 14/200 = 0,21$ . Assim, o erro de classificação da doença não produziu viés na razão de riscos, mas a estimativa de diferença de riscos esperada é de apenas  $0,21/0,30 = 70\%$  da diferença de riscos real.

Esse exemplo ilustra como o erro de classificação não diferencial independente, com especificidade perfeita, não enviesará a estimativa da razão de riscos, mas enviesará para baixo a magnitude absoluta da estimativa da diferença de riscos, por um fator igual à probabilidade falso-negativa (Rodgers e MacMahon, 1995). Com esse tipo de erro de classificação, a razão de chances e a razão de taxas permanecerão viesadas na direção da nulidade, embora o viés seja pequeno quando o risco de doença for baixo ( $<10\%$ ) em ambos os grupos de exposição. Essa aproximação é uma consequência da relação da razão de chances e da razão de taxas com a razão de riscos, quando o risco da doença é baixo em todos os grupos de exposição (ver Cap. 4).

Considere em seguida a mesma coorte de estudo, mas agora com sensibilidade perfeita de detecção da doença (sem falso-negativos) e especificidade imperfeita de  $80\%$ . O número esperado de casos aparentes será, então,  $40 + (1 - 0,80)(100 - 40) = 52$ , entre os expostos, e  $20 + (1 - 0,80)(200 - 20) = 56$ , entre os não expostos. Sob essa formulação, os numeradores dão uma estimativa da razão de riscos esperada de  $(52/100)/(56/200) = 1,9$  e uma estimativa da diferença de riscos esperada de  $52/10 - 56/200 = 0,24$ . Ambas as medidas estão viesadas em direção à nulidade, com a estimativa da diferença de riscos esperada igual a  $0,24/0,30 = 80\%$  do valor real. Esse exemplo ilustra como o erro de classificação não diferencial independente com sensibilidade perfeita enviesará ambas as medidas, com a magnitude absoluta da estimativa de diferença de riscos viesada para baixo por um fator igual à probabilidade falso-positiva (Rodgers e MacMahon, 1995).

Com sensibilidade e especificidade imperfeitas, o viés na magnitude absoluta da diferença de riscos produzido pelo erro de classificação não diferencial de doença, que é independente de outros erros, será igual à soma das probabilidades falso-negativas e falso-positivas (Rodgers e MacMahon, 1995). Os vieses nas medidas de efeito relativas não têm uma fórmula simples, nesse caso.

Queremos enfatizar que, quando tanto a exposição quanto a doença são mal classificadas de forma não diferencial, mas os erros de classificação são dependentes, é possível se obter um viés substancial na direção oposta à nulidade (Chavance et al., 1992; Kristensen, 1992), e as relações simples de viés que acabamos de dar não mais se aplicam. Erros dependentes podem surgir facilmente em muitas situações, tais como em estudos nos quais o *status* de exposição e doença são ambos determinados a partir de entrevistas.

### **Disseminação da má interpretação de efeitos do erro da classificação não diferencial**

O viés por erro da classificação não diferencial independente de uma exposição dicotômica é sempre na direção da nulidade, de modo que se esperaria ver uma estimativa maior se a classificação errônea estivesse ausente. Como resultado, muitos pesquisadores se satisfazem ao conseguir o erro de classificação não diferencial em lugar da classificação acurada. Essa postura pode ocorrer, em parte, porque alguns pesquisadores consideram mais aceitável relatar equivocadamente uma associação como ausente, quando ela de fato existe, do que fazer um relato errado de uma associação como presente, quando na verdade ela não existe, e consideram o erro de classificação não diferencial como favorecendo o primeiro tipo de relato errado sobre o outro. Outros pesquisadores escrevem como se resultados positivos afetados por erro de classificação não diferencial fornecessem evidências mais fortes para uma associação do que o indicado por estatísticas não corrigidas. Todavia, há várias falhas em tais interpretações.

Em primeiro lugar, muitos pesquisadores esquecem que mais do que o caráter não diferencial é necessário para assegurar o viés em direção à nulidade. Precisa-se também de independência e de algumas outras restrições, tais como a variável ser binária. Em segundo, poucos pesquisadores parecem conscientes de que a categorização de variáveis contínuas (p. ex., usar quintis em vez de quantidades reais de comida ou de nutrientes) pode mudar o erro não diferencial para diferencial (Flegal et al., 1991; Wacholder et al., 1991), ou que a falta de controle de fatores relacionados com a mensuração pode fazer a mesma coisa, mesmo que tais fatores não sejam confundidores.

Ainda que o erro de classificação satisfaça todas as condições para produzir um viés em direção à nulidade na estimativa pontual, ele não produz, necessariamente, um viés correspondente para cima no valor  $P$  para a hipótese nula (Bross, 1954; Greenland e Gustafson, 2006). Em consequência, estabelecer que o viés (se presente) foi na direção da nulidade não aumentaria a evidência de que uma associação não nula estivesse presente. Além do mais, viés na direção da nulidade (tal qual o viés afastando-se da nulidade) ainda é uma distorção, a qual variará entre dos estudos. Em particular, ele pode produzir distorções sérias em revisões da literatura e em metanálises, mascarar diferenças reais entre estudos, exagerar diferenças ou criar diferenças falsas. Essas consequências podem ocorrer porque diferenças em características secundárias do estudo, tais como prevalência de exposição, afetarão o grau no qual o erro de classificação produz viés em estimativas de diferentes estratos ou estudos, mesmo que a sensibilidade e a especificidade da classificação não variem entre os estratos, ou estudos (Greenland, 1980). Situações típicas são pioradas pelo fato de que sensibilidade e especificidade, bem como prevalência de exposição, variarão entre os estudos (Begg, 1987).

Frequentemente, essas diferenças em desempenho de mensuração surgem de desigualdades aparentemente inócuas na maneira pela qual as variáveis são avaliadas ou categorizadas, com o pior desempenho surgindo a partir de categorizações de exposição excessivamente simplificadas ou não processadas. Por exemplo, suponha que tomar aspirina transitoriamente reduza o risco de infarto do miocárdio. A palavra “transitoriamente” implica um período de indução breve, com nenhum efeito preventivo fora daquele período. Para um dado ponto no tempo, ou unidade pessoa-tempo, na história de um sujeito, a classificação ideal daquele tempo como exposto ou não exposto à aspirina seria baseada no fato de a aspirina ter sido usada antes daquele tempo, mas dentro do período de indução para seu efeito. Por esse padrão, um infarto do miocárdio subsequente ao uso de aspirina dentro do período de indução seria classificado apropriadamente como um caso exposto à aspirina. Contudo, se nenhuma aspirina fosse usada dentro do período de indução, o caso seria adequadamente classificado como não exposto, mesmo que a aspirina tivesse sido tomada em um tempo anterior, ou posterior.

Essas classificações ideais refletem o fato de que o uso fora do período de indução é irrelevante do ponto de vista causal. Muitos estudos, entretanto, focam qualquer uso (em qualquer tempo durante a vida de um indivíduo), ou algum uso com duração de vários anos. Tais índices cumulativos durante um período de tempo longo aumentam a exposição possivelmente relevante com a exposição irrelevante, e podem introduzir, assim, um viés (geralmente na direção da nulidade) que é paralelo ao viés devido ao erro de classificação não diferencial.

Um viés semelhante pode vir de uma definição ampla de desfecho. Em particular, garantias injustificáveis da falta de qualquer efeito podem emergir facilmente de estudos nos quais uma faixa larga de desfechos não relacionados etiologicamente são agrupados. Em estudos de coorte em que há categorias de doença com poucos sujeitos, os pesquisadores são tentados, ocasionalmente, a aumentar o número de sujeitos em cada análise, dessa forma ganhando precisão. Esse agrupamento de categorias pode obscurecer efeitos sobre categorias de doença mais estreitamente definidas. Por exemplo, Smithells e Shepard (1978) investigaram o potencial teratogênico da droga *Bendectin*, um medicamento usado para náuseas da gravidez. Visto que somente 35 bebês em seu estudo de coorte nasceram com uma malformação, sua análise foi focada em um só desfecho: malformação. Contudo, nenhum agente teratogênico causa todas as malformações; se tal análise deixa de encontrar um efeito, a falha pode ser simplesmente o resultado do agrupamento de muitas malformações não relacionadas com *Bendectin* com aquelas que o são. De fato, apesar da alegação dos autores de que “seu estudo fornece evidência substancial de que o *Bendectin* não é teratogênico no homem”, seus dados indicaram uma relação forte (embora imprecisa) entre *Bendectin* e malformações cardíacas.

Erro de classificação que tem produzido, discutivelmente, viés na direção da nulidade é uma preocupação maior na interpretação de estudos que parecem indicar a ausência de um efeito. Consequentemente, em estudos que indicam pouco ou nenhum efeito, é crucial para os pesquisadores tentar estabelecer a direção do viés, para determinar se um efeito real poderia ter sido obscurecido. Ocasionalmente, os críticos de um estudo argumentarão que dados de exposição pobres, ou má classificação

de doença, invalidam os resultados. Esse argumento, entretanto, é incorreto se os resultados indicarem uma associação diferente de zero e se for possível ter certeza de que erros de classificação produziram viés em direção à nulidade, porque o viés estará na direção de subestimar a associação. Nessa situação, a tarefa maior será estabelecer que os erros de classificação foram realmente do tipo que produziria viés em direção à nulidade.

Inversamente, o erro de classificação que produziu, discutivelmente, viés afastando-se da nulidade é uma preocupação maior na interpretação de resultados que parecem indicar um efeito. O quadro nessa direção é turvado pelo fato de que forças que levam ao erro diferencial e a viés para longe da nulidade (p. ex., viés de recordação) são contrabalançadas, em extensão desconhecida (possivelmente por inteiro), por forças que levam o viés na direção da nulidade (p. ex., simples deterioração da memória com o passar do tempo). Mesmo apenas com variáveis binárias, uma análise quantitativa detalhada da recordação diferencial pode ser necessária para se obter qualquer ideia da direção do viés (Drews e Greenland, 1990), e mesmo com dados internos de validação, a direção do viés raramente pode ser clara. Discutiremos métodos analíticos para avaliar esses problemas no Capítulo 19.

A importância de se apreciar a provável direção do viés foi ilustrada pela interpretação de um estudo sobre espermicidas e defeitos de nascença (Jick et al., 1981a, 1981b). Esse estudo relatou uma prevalência aumentada de vários tipos de distúrbios congênitos entre mulheres que foram identificadas como tendo recebido uma prescrição para espermicidas, durante um intervalo especificado antes do nascimento. As informações sobre exposição constituíram somente uma correlação grosseira com o uso real de espermicidas durante um período de tempo teoricamente relevante, mas o erro de classificação que resultou tinha probabilidade de ser não diferencial e independente de erros na avaliação do desfecho, porque a informação sobre a prescrição foi registrada na memória de um computador antes que o desfecho fosse conhecido. Uma das críticas levantadas sobre o estudo foi de que imprecisões nas informações sobre exposição lançavam dúvidas sobre a validade dos achados (Felarca et al., 1981; Oakley, 1982). Essas críticas, entretanto, não abordaram a direção do viés resultante e, assim, são inapropriadas, se a estrutura do erro de classificação indica que o viés é para baixo, pois então aquele viés não poderia explicar a associação observada (Jick et al., 1981b).

Como um exemplo, é incorreto descartar um estudo que relata uma associação, simplesmente porque há erro de classificação não diferencial independente de uma exposição binária, porque, sem a má classificação, a associação observada provavelmente seria até mesmo maior. Assim, as implicações do erro de classificação não diferencial independente dependem fortemente da percepção do estudo como sendo “positivo” ou “negativo”. Ênfase sobre avaliação quantitativa, em vez de uma descrição qualitativa dos resultados do estudo, diminui a probabilidade de má interpretação, por isso exploraremos os métodos para avaliação quantitativa de viés no Capítulo 19.

## Erro de classificação de confundidores

Se uma variável de confundimento for mal classificada, a capacidade de controle do confundimento é dificultada (Greenland, 1980; Kupper, 1984; Brenner, 1993; Marshall e Hastrup, 1996; Marshall et al., 1999; Fewell et al., 2007). O erro de classificação não diferencial independente de uma variável de confundimento dicotômica reduzirá o grau no qual o confundidor pode ser controlado e, assim, causará um viés na direção do confundimento pela variável. O resultado esperado ficará entre a associação não ajustada e a associação ajustada corretamente (i.e., a que teria sido obtida se o confundidor não tivesse sido mal classificado). Esse problema pode ser visto como de confundimento de resíduo (i.e., confundimento deixado depois do controle das variáveis confundidoras disponíveis). O grau de confundimento de resíduo deixado nos estratos do confundidor mal classificado geralmente diferirá ao longo de tais estratos, o que distorcerá o grau aparente de heterogeneidade (modificação de efeito) entre os estratos (Greenland, 1980). O erro de classificação não diferencial independente, ou do confundidor, ou da exposição, pode, portanto, dar origem ao aparecimento de modificação de



efeito-medida (interação estatística) quando de fato não há nenhuma, ou mascarar o aparecimento de tal modificação quando de fato estiver presente.

Se o erro de classificação for diferencial ou dependente, a associação ajustada resultante pode nem mesmo cair entre a associação não processada e a correta ajustada. Então, o problema se torna não só de confundimento de resíduo, mas de distorção adicional produzida por seleção diferencial de sujeitos em estratos de análise diferentes. Infelizmente, os erros dependentes entre variáveis de exposição são comuns, especialmente em estudos baseados em questionários. Por exemplo, em estudos epidemiológicos de nutrientes e doença, as ingestões de nutrientes são calculadas a partir da ingestão de alimentos, e qualquer erro de avaliação da ingestão de comida será traduzido em erros dependentes entre os nutrientes encontrados nos mesmos alimentos. Similarmente, em estudos epidemiológicos de ocupação e doença, as exposições químicas geralmente são calculadas pelas histórias ocupacionais, e falhas na avaliação dessas histórias resultarão em erros dependentes entre exposições encontradas nos mesmos empregos.

Se o confundimento é forte e a relação exposição-doença é fraca, ou zero, o erro de classificação do confundidor pode produzir resultados extremamente enganosos, mesmo que este seja independente e não diferencial. Por exemplo, dada uma relação causal entre fumo e câncer da bexiga, uma associação entre fumo e beber café tornaria o tabagismo um confundidor da relação entre beber café e câncer da bexiga. Posto que o controle do confundimento pelo tabagismo depende de informações acuradas sobre o hábito de fumar, e como o erro de classificação das informações relevantes sobre o fumo é não pode ser evitado, não importa como o fumo seja mensurado, algum confundimento de resíduo pelo tabagismo não pode ser evitado (Morrison et al., 1982). O problema do confundimento de resíduo será ainda pior se a única informação disponível sobre o tabagismo for uma dicotomia simples do tipo “já fumou” *versus* “nunca fumou”, porque a falta de especificação detalhada do fumar impede o controle adequado do confundimento. O confundimento de resíduo é especialmente complicador, porque, para muitos pesquisadores e leitores, parecerá que o confundimento pelo fumo foi completamente controlado.

## As complexidades do erro de classificação simultânea

Continuando com o exemplo precedente, considere o erro de classificação do uso de café, assim como do fumo. Por um lado, se o erro de classificação do café for não diferencial com relação ao fumo, e independente de erros quanto a este último, o efeito provável seria diminuir ainda mais a associação observada fumo-café e, assim, reduzir adicionalmente a eficácia do ajuste para o fumo. O resultado seria um confundimento de resíduo para cima ainda maior do que quando somente o fumo foi mal classificado. Por outro lado, se as mensurações foram feitas por meio de questionários, os erros do café e do fumo poderiam estar positivamente associados e não independentes, contrabalançando potencialmente o fenômeno antes mencionado a um grau desconhecido. Também, se os erros quanto ao café forem não diferenciais com relação ao câncer de bexiga, e independentes de erros diagnósticos, produziriam, mais provavelmente, um viés para baixo na associação observada.

No entanto, se as mensurações foram a partir de um questionário aplicado após o diagnóstico, a não diferenciação dos erros, tanto do fumo quanto do café com relação ao câncer de bexiga, se tornaria questionável. Se os controles tenderam a subrelatar esses hábitos mais do que os casos, a diferenciação resultante provavelmente agiria em uma direção para cima das associações, tanto do café como do fumo, com o câncer, cancelando parcialmente o viés para baixo pelo erro de classificação do café, e o viés para cima pelo confundimento de resíduo do fumo; porém, se os casos tendessem a subrelatar esses hábitos mais que os controles, a diferenciação provavelmente agravaria o viés para baixo pelo erro de classificação do café, e o viés para cima pelo confundimento de resíduo do fumo.

O resultado desses efeitos seria quase impossível de prever, dada a falta usual de informações precisas sobre as taxas de erro de classificação. Enfatizamos que essa imprevisibilidade está acima daquela do erro aleatório presumido pelos métodos estatísticos convencionais; portanto, esse erro não se reflete nos intervalos de confiança convencionais, porque estes são dirigidos apenas para a variação aleatória na seleção de sujeitos e na exposição real, e presumem que erros na mensuração do consumo de café e tabagismo estejam ausentes.



## CAPACIDADE DE GENERALIZAÇÃO

Os físicos trabalham sob a suposição de que as leis da natureza são as mesmas em toda parte, e, portanto, que o que eles aprendem sobre a natureza tem aplicabilidade universal. Na pesquisa biomédica, às vezes parece que a nossa suposição é o oposto, isto é, que os achados de nossa pesquisa somente se aplicam a populações intimamente semelhantes àquelas que estudamos. Essa visão origina-se da experiência de que os efeitos biológicos podem diferir, e diferem, entre populações e subgrupos diferentes. Assim, o pesquisador cauteloso inclina-se a refrear a generalização de resultados além das circunstâncias que descrevem o cenário do estudo.

Em consequência, muitos estudos epidemiológicos são delineados para amostrar sujeitos de uma população-alvo de interesse particular, de modo que a população do estudo seja representativa da população-alvo, no sentido de ser uma amostra probabilística daquela população. Inferência para esse alvo também poderia ser obtida pela superamostragem de alguns subgrupos, e então padronizando, ou voltando a ponderar, os dados do estudo, para combinar com a distribuição da população-alvo. Os delineamentos em dois estágios (Caps. 8 e 15) são exemplos simples de tal estratégia.

Levada ao extremo, entretanto, a busca da representatividade pode impedir que sejam identificados validamente as relações causais. Se a generalização dos resultados do estudo é limitada, literalmente, às características daqueles estudados, então as inferências causais não podem ser generalizadas além daqueles sujeitos que foram estudados e durante o período de tempo no qual o estudo ocorreu. Contudo, mesmo os físicos reconhecem que aquilo que consideramos como sendo leis físicas universais poderia variar com o tempo, ou sob condições fronteiriças e, portanto, podem não ser verdadeiramente universais. O processo de generalização na ciência envolve fazer suposições sobre o domínio ao qual os resultados do estudo se aplicam.

O peso da ênfase sobre representatividade da amostra na pesquisa epidemiológica provavelmente é derivado de experiências iniciais com inquéritos, para os quais o objetivo de inferência era apenas a descrição da população pesquisada. Os cientistas sociais frequentemente realizam e se baseiam em inquéritos com amostras probabilísticas, porque as decisões sobre o que é relevante para generalização são mais difíceis nas ciências sociais. Além disso, as questões de interesse para os cientistas sociais podem ser concernentes apenas para uma população em particular (p. ex., eleitores em um município em um ponto do tempo), e as populações são consideravelmente mais diversificadas em fenômenos sociológicos do que biológicos.

Nos laboratórios de ciências biológicas, contudo, é rotina para os pesquisadores conduzir experimentos usando animais com características selecionadas para aumentar a validade do trabalho experimental, e não para representar uma população-alvo. Por exemplo, cientistas de laboratório conduzindo experimentos com *hamsters* preferirão, frequentemente, estudar *hamsters* geneticamente idênticos do que uma população representativa dos *hamsters* do mundo, a fim de minimizar preocupações sobre variação genética que afetasse os resultados. Essas restrições podem gerar inquietação sobre a capacidade de generalização, mas isso só se torna importante depois que os resultados do estudo tiverem sido aceitos como válidos para o grupo restrito que foi estudado.

De modo semelhante, os delineamentos epidemiológicos normalmente são mais consistentes se a seleção de sujeitos é guiada pela necessidade de se fazer uma comparação válida, o que pode demandar restrição intensa de sujeitos admissíveis a uma faixa estreita de características, e não por uma tentativa de tornar os sujeitos representativos, no entendimento de uma amostra de pesquisa, das populações-alvo potenciais. A seleção de grupos de estudo que são representativos de populações maiores no entendimento estatístico tornará mais difícil, frequentemente, fazer inferências válidas internamente, por dificultar o controle do confundimento por fatores que variam em tais populações, tornar mais difícil garantir níveis de cooperação uniformemente altos e assegurar mensurações uniformemente acuradas.

Para minimizar as ameaças de validade que discutimos, seria mais desejável selecionar grupos de estudo por homogeneidade com relação a confundidores importantes, por comportamento altamente cooperativo e pela disponibilidade de informações precisas, e não por uma tentativa de

serem representativos de uma população natural. Exemplos clássicos incluem o Estudo de Médicos Britânicos sobre fumo e saúde e o Estudo da Saúde das Enfermeiras, nenhum dos quais foi remotamente representativo da população geral, no que tange a fatores sociodemográficos. Presumiu-se que sua falta de representatividade não estivesse relacionada à maioria dos efeitos estudados. Se houve dúvidas sobre essa presunção, elas só se tornariam importantes se ficasse claro que as associações observadas eram estimativas de efeito válidas dentro dos próprios estudos.

Uma vez que a natureza e, pelo menos, a ordem de magnitude de um efeito são estabelecidas por estudos delineados para maximizar a validade, a generalização a outros grupos não estudados torna-se mais simples. Essa generalização é, em medida ampla, uma questão quanto a se os fatores que distinguem esses outros grupos dos grupos estudados modificam, de alguma forma, o efeito em questão. Ao responder essa questão, os dados epidemiológicos ajudarão e poderão ser essenciais, mas outras fontes de informação, tais como fisiopatologia básica, podem desempenhar um papel ainda maior. Por exemplo, embora a maioria dos dados decisivos conectando o fumo ao câncer de pulmão fosse derivada de observações em homens, ninguém duvidou de que os fortes efeitos observados seriam transportáveis, pelo menos aproximadamente, para mulheres, pois os pulmões de homens e mulheres parecem ser semelhantes, se não idênticos, em detalhes fisiológicos. Entretanto, dadas as enormes diferenças sexuais em perda de ferro, seria insensato generalizar livremente para os homens a respeito da suplementação com ferro observada em mulheres em pré-menopausa.

Tais exemplos contrastantes sugerem que, talvez até mais do que com a inferência (interna) sobre populações restritas, a generalização válida deve lançar mão do conhecimento de diversos ramos da ciência. Conforme temos enfatizado, a representatividade frequentemente é um empecilho à execução de um estudo internamente válido, e considerações de ciências afins demonstram que ela nem sempre é necessária para a generalização válida. Assim, prevenimos que a perseguição cega da representatividade geralmente leva ao desperdício de recursos preciosos dos estudos.

# Precisão e estatística em estudos epidemiológicos

Kenneth J. Rothman, Sander Greenland e Timothy L. Lash

**Erro aleatório e precisão estatística 179**

**Tamanho do estudo 180**

**Eficiência do estudo 181**

**Precisão e estratificação 181**

**Abordagens para o erro aleatório 182**

Testes de significância e testes de hipótese 182

Estimação estatística 188

**Conclusão 201**

Como descrito no Capítulo 9, dois tipos de erro, sistemático e aleatório, prejudicam a acurácia. O Capítulo 9 focalizou a compreensão das fontes de erro sistemático. Neste capítulo, discutimos métodos para mensurar, limitar e levar em conta o erro aleatório em um estudo epidemiológico e como interpretá-los apropriadamente. No Capítulo 11, abordamos opções no delineamento que possam reduzir a quantidade de erro aleatório (i.e., melhorar a precisão) de um estudo, dentro de certas restrições de custo e de viabilidade.

## ERRO ALEATÓRIO E PRECISÃO ESTATÍSTICA

O que é erro aleatório? Frequentemente é igualado à variação casual ou aleatória, a qual raramente é bem definida. Muitas pessoas acreditam que o acaso desempenha um papel fundamental em todos os fenômenos físicos e, por implicação, nos biológicos. Para alguns, a crença no acaso é tão dominante que coloca as ocorrências aleatórias em um papel importante, como causas componentes de tudo que experimentamos. Outros acreditam que a causalidade pode ser vista como determinística, significando que uma elaboração completa dos fatores relevantes, em um dado conjunto de circunstâncias, levará inabalavelmente, com análise suficiente, a uma predição perfeita dos efeitos resultantes dessas causas. Sob o último ponto de vista, toda experiência é predestinada a se desenrolar de uma maneira teoricamente previsível, que se desenvolve inexoravelmente a partir do padrão prévio de ações. Mesmo com essa visão determinística extrema, contudo, deve-se encarar o fato de que ninguém pode adquirir conhecimento suficiente para prever efeitos perfeitamente, salvo para padrões triviais de causa-efeito. A resultante previsibilidade incompleta de determinados desfechos torna sua variabilidade indistinguível das ocorrências aleatórias.

Uma descrição unificante da previsibilidade incompleta pode, assim, ser forjada, igualando a variação aleatória à ignorância sobre as causas dos desfechos de nossos estudos, uma ignorância que é inevitável, quer a chance física esteja entre as causas ou não. Por exemplo, prever o desfecho de uma moeda lançada para o alto representa um problema físico cuja solução é viável por meio da aplicação de leis físicas. Se as fontes de variação que não podemos explicar são na verdade fenômenos casuais, faz pouca diferença: tratamos tal variação como aleatória até que possamos explicá-la e, dessa forma, reduzi-la, relacionando-a a fatores conhecidos.

Em um estudo epidemiológico, a variação aleatória tem muitos componentes, mas o principal contribuidor é o processo de selecionar os sujeitos do estudo específico; que é geralmente designado como *amostragem*; a variação aleatória presente é conhecida como *variação de amostragem*, ou *erro amostral*. Os estudos de caso-controle frequentemente envolvem um processo de amostragem física, ao passo que os estudos de coorte muitas vezes não o envolvem. No entanto, é uma prática padrão tratar todos os estudos epidemiológicos, inclusive os estudos de coorte, como tendo erro amostral. Dessa maneira, os sujeitos em um estudo, fisicamente amostrados ou não, são vistos como uma amostra figurativa de pessoas possíveis de terem sido incluídas no estudo ou de diferentes experiências possíveis que os sujeitos do estudo poderiam ter tido. Mesmo que todos os indivíduos em uma população fossem incluídos em um estudo, os sujeitos do estudo são vistos como uma amostra da experiência biológica potencial de uma população conceitual ainda mais ampla. Com essa visão, o dito estatístico, de que não há erro de amostragem se uma população inteira (ao contrário de sua amostra) é estudada, não se aplica aos estudos epidemiológicos, mesmo que toda uma população seja incluída no estudo. Conceitualmente, os sujeitos reais são sempre considerados uma amostra de uma experiência de interesse mais ampla – embora eles raramente satisfaçam, verdadeiramente, a definição de uma amostra aleatória que sirva de base para as ferramentas estatísticas comumente usadas para mensurar a variação aleatória (Greenland, 1990, 2005b).

A amostragem é somente uma fonte de erro aleatório que contribui para as imprecisões imprevisíveis dos estudos epidemiológicos. Outra fonte é a variação inexplicável em medidas de ocorrência, tais como as taxas de incidência ou as proporções de prevalência observadas. Por exemplo, quando o estado de exposição não é distribuído aleatoriamente, o confundimento (Cap. 4) pode levar a desvios de associações estimadas de efeitos-alvo que excedem em muito aquilo que modelos estatísticos padrão presumem como prováveis. A mensuração errônea de variáveis-chave do estudo também contribui para a imprecisão geral, tanto de modo aleatório quanto sistemático. Em consequência de todas essas fontes extras de variação, e por causa da fraca sustentação teórica para a conceituação de sujeitos de um estudo como uma amostra de uma experiência mais ampla, as ferramentas estatísticas usuais que utilizamos para mensurar variação aleatória, em seu melhor desempenho, fornecem estimativas mínimas da incerteza real que deveríamos ter sobre o objeto da estimação (Greenland, 1990, 2005b). Um modo elementar para melhorar a quantificação de nossa incerteza é através da *análise de viés*, que discutiremos no Capítulo 19.

Uma medida comum de variação aleatória em um processo de mensuração ou de estimação é a *variância* do processo, que é discutida no Capítulo 13. A *precisão estatística* de (ou *informação estatística* em) uma mensuração ou de um processo é tomada frequentemente como sendo o inverso da variância das mensurações ou das estimativas que o processo produz. Nesse sentido, a precisão é o oposto do erro aleatório. A precisão da estimação pode ser melhorada (o que quer dizer que a variância pode ser reduzida) pelo aumento do tamanho do estudo. A precisão também pode ser aprimorada pela modificação do delineamento do estudo, a fim de diminuir a variância dado um número total fixo de sujeitos; esse processo é chamado de *melhoramento da eficiência estatística* do estudo. Ele será introduzido aqui e discutido mais detalhadamente no Capítulo 13. Talvez o exemplo epidemiológico mais comum de tal melhoria do delineamento seja o uso de um estudo de caso-controle em vez de um estudo de coorte, porque, para um tamanho de estudo fixo, a variância de uma estimativa de efeito depende muito da proporção de sujeitos no estudo que são casos.

## TAMANHO DO ESTUDO

Um modo comum de reduzir o erro aleatório ou melhorar a precisão de uma estimativa epidemiológica é aumentar o tamanho do estudo. Restrições práticas sobre recursos limitam inevitavelmente o tamanho do estudo, de modo que é preciso planejar em conformidade. Um método utilizado para se planejar o tamanho de um estudo é calcular tal tamanho com base nas fórmulas estatísticas convencionais de “tamanho de amostra” (p. ex., ver Schlesselman, 1974; Rothman e Boice, 1982; Gre-

enland, 1985b, 1988a). Essas fórmulas relacionam o tamanho de um estudo a seu delineamento, à população, ao poder ou à precisão desejados.

As fórmulas de tamanho de estudo, sendo puramente matemáticas, não levam em conta qualquer coisa que não esteja incluída como variável na fórmula. No máximo, servem apenas para prover diretrizes grosseiras e, em algumas situações, podem ser enganosas em uma perspectiva mais ampla. Por exemplo, as fórmulas convencionais não pesam o valor da informação obtida de um estudo contra seu uso de recursos. Ainda no planejamento do tamanho do estudo, um problema focal é determinar como equilibrar o valor da maior precisão dos resultados de uma pesquisa contra o seu custo maior. Assim, a solução do problema envolve uma análise de custo-benefício de gastar mais esforços ou fundos para ganhar mais precisão. Maior precisão tem um valor para os beneficiários da pesquisa, mas tal valor é indeterminado, porque é sempre incerto com relação a quantos beneficiários haverá. Além do mais, os benefícios potenciais envolvem complicações intrínsecas de muitos fatores sociais, políticos e biológicos, que quase nunca são quantificadas. Consequentemente, apenas palpites quanto ao tamanho custo-eficiente para um estudo epidemiológico são possíveis. Embora a determinação do tamanho do estudo possa ser ajudada por fórmulas convencionais, a escolha final também deve incorporar restrições práticas não quantificadas e implicações dos vários tamanhos de estudo.

## EFICIÊNCIA DO ESTUDO

Outra maneira de reduzir o erro aleatório, ou de aumentar a precisão em uma estimativa epidemiológica é modificar o delineamento do estudo. Um aspecto que pode ser, frequentemente, manipulado pelo delineamento é a alocação dos sujeitos em grupos de estudo. Quando o fator em estudo não tem efeito e nenhum ajuste é necessário, a distribuição igual em grupos de exposição é o delineamento de coorte mais eficiente (Walter, 1977). Por exemplo, na ausência de um efeito ou de confundimento, um estudo de coorte de 2.000 pessoas será mais eficiente se forem selecionadas 1.000 pessoas expostas e 1.000 não expostas, para investigação. Da mesma forma, em um estudo de caso-controle, na ausência de um efeito ou de confundimento, será mais eficiente estatisticamente ter um número igual de casos e controles. Na presença de um efeito, a alocação ótima para eficiência estatística difere da alocação igual por uma quantidade que é função do parâmetro sendo estimado (Walter, 1977). Entretanto, quando são necessários ajustes para confundimento (como quase sempre é o caso), esses resultados não mais se aplicam estritamente. Além disso, tais resultados presumem que nenhum efeito esteja presente – o que, é claro, não se sabe se é verdadeiro em qualquer aplicação real (caso contrário não haveria necessidade para o estudo). Assim, tais resultados não devem ser considerados mais do que diretrizes rudimentares para o delineamento.

## PRECISÃO E ESTRATIFICAÇÃO

Em muitas análises epidemiológicas, os dados brutos são divididos em estratos para examinar efeitos em subcategorias de uma outra variável ou para controle do confundimento. A eficiência de um estudo pode ser bastante afetada pela estratificação dos dados. Um estudo que tenha uma razão de alocação geral que seja favorável à precisão (que será uma razão de 1,0 se não há efeito nem confundimento) pode, contudo, ter razões de alocação dentro de estratos que variam intensamente de valores baixos para altos. Não é incomum se ver alguns estratos com as razões de alocação extremas de 0 e infinito (p. ex., nenhum caso em alguns estratos, e nenhum controle em outros). Quanto menores os números dentro dos estratos, mais extrema é provavelmente a variação da razão de alocação entre os estratos. Os valores extremos resultam de zero sujeito, ou unidade pessoa-tempo, para um grupo, em um estrato. Números pequenos dentro dos estratos resultam em muito poucos sujeitos em relação ao número de estratos criado. Esse problema de *dados esparsos* pode se desenvolver mesmo com estudos grandes, porque o número de estratos necessários na análise aumenta geometricamente com o número de variáveis usadas para estratificação. Na verdade, os dados esparsos representam uma limitação importante

da análise estratificada, embora o mesmo problema também afete negativamente os modelos de regressão. Métodos para se lidar com os dados esparsos estão descritos nos Capítulos 15 e 21.

Quando comparações dentro de estratos forem essenciais, e se espera muita variação da razão de alocação entre os estratos, então o *pareamento* das variáveis de estratificação será uma maneira de se manter uma razão de alocação eficiente dentro dos estratos, e de se reduzir os problemas de dados esparsos sem aumentar o tamanho do estudo. Quando o pareamento de todas as variáveis de estratificação não é factível, aumentar o número geral de sujeitos pelo menos reduzirá a escassez de dados e melhorará a precisão, mesmo que apenas um grupo (p. ex., os controles) possa ser expandido.

## ABORDAGENS PARA O ERRO ALEATÓRIO

A estatística e seu papel na análise de dados têm sofrido uma transformação gradual, mas profunda, em tempos recentes. Há uma distinção essencial entre um objetivo de estudo qualitativo (responder “sim” ou “não” a uma pergunta) e um quantitativo (mensurar algo). Tal distinção reflete uma preferência crescente pelo último objetivo e pelos métodos estatísticos compatíveis com ele. Até os anos 1970, a maioria das aplicações da estatística na epidemiologia tinha como foco a decisão se “acaso” ou “erro aleatório” podiam ser isoladamente responsáveis por uma associação observada. Os métodos usados para essa decisão eram os clássicos *testes de significância*, predominantes nas aplicações britânicas, e os *testes de hipótese* de Neyman-Pearson, predominantes em aplicações norte-americanas (Goodman, 1993; Gigerenzer, 2004). Por causa de suas similaridades, o termo *teste de significância* frequentemente é aplicado a ambos os métodos de coleta.

Essas aplicações de testes, que foram sujeitas a algum criticismo inicial (Boring, 1919; Berkson, 1938, 1942; Hogben, 1957), foram criticadas por epidemiologistas e estatísticos ao longo dos anos 1970 e 1980. Os críticos enfatizam que a maioria das aplicações epidemiológicas, se não todas, precisa mais do que uma decisão de o acaso, isoladamente, poder ter produzido uma associação. Mais importante é a estimação da magnitude da associação, inclusive uma avaliação da precisão do método de estimação. A ferramenta de estimação utilizada pela maioria dos autores é o intervalo de confiança, que fornece uma faixa de valores para a associação, sob a hipótese de que apenas a variação aleatória criou discrepâncias entre o valor verdadeiro da associação em estudo e o valor observado nos dados (Altman, et al., 2000; ver Caps. de 13 a 16). Outros autores, embora a favor do movimento pela estimação de intervalo, destacam que os intervalos de confiança padecem das mesmas imperfeições associadas aos testes de significância e defendem outras abordagens à estimação de intervalos (Goodman e Royall, 1988; Berger e Berry, 1988; Royall, 1997; Greenland, 2006a; Cap. 18).

## Testes de significância e testes de hipótese

Há quase 70 anos, Berkson (1942) escreveu:

Difícilmente é um exagero dizer que a estatística, como ensinada atualmente na escola dominante, consiste quase totalmente em testes de significância, embora nem sempre apresentada assim, alguns comparativamente simples e diretos, outros elaborados e de difícil compreensão.

O uso universal de valores *P* e de referências a “estatisticamente significativo”, encontrados na literatura médica atual, demonstra o papel dominante que os testes estatísticos de hipótese ainda desempenham na análise de dados, em alguns ramos das ciências biomédicas. Muitos pesquisadores ainda acreditam que seria infrutífero submeter à publicação qualquer artigo que carecesse de testes estatísticos de significância. Essa crença não é inteiramente sem fundamento, porque muitos editores e críticos de revistas científicas ainda confiam em testes de significância como indicadores de análise estatística sofisticada e significativa, assim como o meio primário de avaliação da variabilidade da amostragem em um estudo. A *significância estatística* baseia-se, em geral, no valor *P* (descrito abaixo): os resultados são considerados “significativos” ou “não significativos” conforme o valor *P*

seja menor a maior do que um valor de corte arbitrário, geralmente 0,05, que é denominado o *nível alfa* do teste.

A preocupação com os testes de significância deriva dos interesses de pesquisa dos estatísticos que foram pioneiros no desenvolvimento da teoria estatística, no início do século XX. Seus problemas de pesquisa eram primariamente industriais e agrícolas e envolviam, tipicamente, experimentos randomizados ou inquéritos de amostra aleatória, que formavam a base para escolha entre dois ou mais cursos de ação alternativos. Tais estudos eram delineamentos para produzir resultados que possibilitassem uma tomada de decisão, e os métodos estatísticos empregados tinham a intenção de facilitar tal decisão. Os conceitos que cresceram dessa herança são aplicados hoje em pesquisa clínica e epidemiológica e refletem esse cenário de tomada de decisão.

Os testes de significância estatística de associações focam geralmente a *hipótese nula*, a qual normalmente é formulada como uma hipótese de nenhuma associação entre duas variáveis em uma *superpopulação*, a população a partir da qual os grupos de estudo observados eram propositadamente amostrados de modo aleatório. Por exemplo, pode-se testar a hipótese de que a diferença de riscos na superpopulação é 0 ou, de forma equivalente, que a razão de riscos é 1. Note que essa hipótese é sobre a superpopulação e *não* sobre os grupos de estudo observados. Os testes podem focar, alternativamente, qualquer outra hipótese específica, por exemplo, que a diferença de riscos é 0,1, ou que a razão de riscos é 2. Para hipóteses não nulas, testes sobre uma medida (p. ex., uma diferença de riscos) geralmente não são equivalentes a testes sobre uma outra medida (p. ex., uma razão de riscos), de modo que se deve escolher uma medida de interesse para realizar um teste de não nulidade.

Uma má interpretação comum dos testes de significância é alegar que não há diferença entre dois grupos observados, porque o teste de nulidade não é estatisticamente significativo, visto que  $P$  é maior do que o ponto de corte para declaração da significância estatística (de novo, usualmente 0,05). Essa interpretação confunde um assunto descritivo (se os dois grupos observados diferem) com uma inferência sobre a superpopulação. O teste de significância refere-se apenas à superpopulação, não aos grupos observados. Dizer que a diferença não é estatisticamente significativa expressa apenas que não se pode rejeitar a hipótese nula de que os grupos da superpopulação são os mesmos; isso *não* implica que os dois grupos observados sejam o mesmo.

É preciso apenas olhar para os dois grupos observados para ver se são diferentes. O teste de significância é concernente, em vez disso, ao fato de a diferença observada dever levar à conclusão de que há uma diferença entre os grupos correspondentes na superpopulação. Além do mais, mesmo que a diferença observada não seja estatisticamente significativa, os grupos da superpopulação podem ser diferentes (i.e., o resultado não implica que a hipótese nula esteja correta). Em vez disso, a diferença não significativa observada expressa apenas que não se pode excluir a hipótese nula, caso o modelo estatístico usado para construir o teste seja aceito.

Entretanto, é uma má interpretação alegar que existe uma associação na superpopulação, porque a diferença observada é estatisticamente significativa. Em primeiro lugar, o teste pode ser significativo somente porque o modelo usado para o cálculo está errado (p. ex., pode haver muitas fontes de viés não controlado). Em segundo, o teste pode ser significativo somente por obra do acaso; por exemplo, mesmo sob condições perfeitas, o uso de um nível alfa de 0,05 confere uma diferença estatisticamente significativa 5% das vezes se a hipótese nula estiver correta.

Como enfatizamos, o ponto de corte alfa é uma convenção arbitrária e questionável; pode-se passar sem ele simplesmente relatando o valor  $P$  real a partir do teste, o que discutiremos agora, em detalhe. Então, exploraremos e criticaremos ainda mais a teoria que levou ao uso disseminado de pontos de corte arbitrários dos testes na pesquisa.

## Valores $P$

Há dois tipos principais de valores  $P$ : monocaudal e bicaudal. Além disso, há dois tipos de valores  $P$  monocaudal: superior e inferior. Definições e interpretações acuradas dessa estatística são sutis e, portanto, raramente fornecidas em textos epidemiológicos. Em consequência, a má interpretação de



valores  $P$  é comum em epidemiologia, bem como em outros campos. Assim, dedicaremos grande parte deste capítulo e do Capítulo 13 à discussão dessa estatística.

Um valor  $P$  monocaudal *superior* é a probabilidade de que uma quantidade correspondente calculada a partir dos dados, conhecida como a *estatística do teste* (tais como o teste  $t$  ou um teste de qui-quadrado), será maior ou igual a seu valor observado, presumindo-se que (a) a hipótese do teste esteja correta e (b) não haja fonte de viés na coleta de dados nem nos processos de análise. Similarmente, um valor  $P$  monocaudal *inferior* é a probabilidade de que a estatística do teste correspondente será menor do que, ou igual a, seu valor observado, novamente presumindo-se que (a) a hipótese do teste esteja correta e (b) não haja fonte de viés na coleta de dados nem nos processos de análise (algumas vezes descreve-se dizendo que o modelo estatístico subjacente está correto). O valor  $P$  bicaudal geralmente é definido como duas vezes o menor dos valores  $P$  superior e inferior, embora definições mais complicadas tenham sido usadas. Sendo uma probabilidade, um valor  $P$  monocaudal deve cair entre 0 e 1; o valor  $P$  bicaudal como recém-definido, entretanto, pode exceder a 1. Os comentários seguintes se aplicam a todos os tipos de valores  $P$ . Alguns autores referem-se a valores  $P$  como “níveis de significância” (Cox e Hinkley, 1974), mas é melhor evitar o último termo, porque tem sido usado por outros autores como se referindo a níveis alfa.

Nos testes de significância clássicos, supõe-se que valores  $P$  pequenos indicam que pelo menos uma das suposições usadas para derivá-lo é incorreta, isto é, qualquer ou ambas as hipóteses de teste (suposição 1) ou o modelo estatístico (suposição 2) está incorreto. Frequentemente, o modelo estatístico é tomado como uma certeza, de modo que um valor  $P$  pequeno é tomado como indicativo de um baixo grau de compatibilidade entre a hipótese de teste e os dados observados. Essa incompatibilidade origina-se do fato de que um valor  $P$  pequeno representa uma probabilidade baixa de se obter uma estatística do teste como extrema, ou mais extrema do que a estatística observada, se a hipótese de teste for verdadeira e nenhum viés estiver presente. Supõe-se, portanto, que valores  $P$  pequenos indicam que a hipótese em teste não é uma explicação aceitável para a associação observada nos dados. Essa interpretação comum tem sido extensamente criticada, porque não leva em conta explicações alternativas e sua aceitabilidade (ou falta disso); para exemplos, ver Berkson (1942) e críticas epidemiológicas mais tardias por Goodman e Royall (1988), Greenland (1990), Goodman (1993) e Gigerenzer (2004). Uma interpretação menos hipotética e mais cautelosa, então, é que um valor  $P$  pequeno indica que há um problema com a hipótese em teste, ou com o estudo, ou com ambos (Fisher, 1943).

Uma das interpretações mais comuns e ingênuas de valores  $P$  é que eles representam probabilidades de testes de hipóteses. Em muitas situações, pode-se calcular uma probabilidade bayesiana ou credibilidade (ver Cap. 18), para o teste de hipóteses, mas ela quase sempre estará longe do valor  $P$  bicaudal (Berger e Delampady, 1987; Berger e Sellke, 1987). Um valor  $P$  monocaudal pode ser usado para colocar um limite inferior na probabilidade bayesiana de certas hipóteses compostas (Casella e Berger, 1987) e sob certas condições se aproximarão da probabilidade bayesiana de que a associação verdadeira é o oposto da direção observada (Greenland e Gustafson, 2006). No entanto, um valor  $P$  para um teste de hipótese simples (p. ex., em que exposição e doença não estão associadas) não é uma probabilidade daquela hipótese: geralmente, aquele valor  $P$  é muito menor do que tal probabilidade bayesiana e, assim, pode facilmente levar alguém ao equívoco de rejeitar a hipótese em teste, inapropriadamente (Berger e Sellke, 1987; Goodman e Royall, 1988).

Uma das mais comuns interpretações errôneas sobre o valor  $P$  é a de que seja a probabilidade dos dados observados sob a hipótese em teste. Essa probabilidade é conhecida como a verossimilhança da hipótese em teste; ver Goodman e Royall (1988), Royall (1997), Edwards (1992) e a discussão seguinte. A verossimilhança de uma hipótese geralmente é muito menor do que o valor  $P$  para esta, porque o valor  $P$  inclui não só a hipótese em teste, mas também as probabilidades para todas as outras configurações de dados possíveis, nas quais a estatística do teste foi mais extrema do que o observado.

Uma má interpretação comum e sutil de um valor  $P$  para testar uma hipótese nula é que ela representa a probabilidade de que os dados mostrassem uma associação tão forte quanto a observada ou mais forte se a hipótese nula fosse correta. Essa má interpretação pode ser encontrada em muitos