

O Mito da Doença Rara

The Rare Disease Myth

Resumo

Demonstra-se que estudos caso-controle podem estimar sem viés os parâmetros risco relativo e fração atribuível, sem que a doença estudada seja rara. Argumenta-se que a pressuposição de raridade da doença para que as estimativas de odds ratio obtidas em estudos caso-controles se aproximem da medida de risco relativo é um mito ainda hoje difundido, que teve suas origens nos primeiros trabalhos, estabelecendo e consolidando o método caso-controle, em meados do século XX.

Palavras-chave: Estudo caso-controle. Estimabilidade. Risco relativo. Odds ratio. Fração atribuível.

Ricardo Cordeiro

Departamento de Medicina Preventiva e Social. Faculdade de Ciências Médicas. Universidade Estadual de Campinas – UNICAMP

Correspondência: Ricardo Cordeiro. Caixa Postal 6111, 13083-970 Campinas SP. E-mail: cordeiro@fcm.unicamp.br

Abstract

This paper shows that case-control studies can produce unbiased estimates of relative risk and attributable fraction without assuming that the studied disease is rare. It also discusses the still current myth that it is necessary to assume that a disease is rare in order to assure that odds ratio estimates are closer to relative risk measures in case control studies, showing that it comes from the beginning of the methodological development of case-control studies, in the mid 20th Century.

Key Words: Case control study. Estimability. Relative risk. Odds ratio. Attributable fraction.

O desenvolvimento dos estudos caso-controle constitui a maior contribuição metodológica da epidemiologia. Este tipo de estudo propicia uma maneira eficiente de obter estimativas populacionais de risco relativo (RR) e fração atribuível (FA). A razão desta eficiência advém da capacidade que o estudo caso-controle tem, a um custo relativamente baixo, de estimar a distribuição populacional de exposições a partir do seu grupo controle. Esta propriedade dispensa o consumo de tempo, bem como recursos financeiros e administrativos, na enumeração e seguimento de um grande conjunto de indivíduos, classificados de acordo com níveis de exposição.

A epidemiologia lida com relações causais entre exposições e doenças, que ocorrem em populações humanas. Os casos surgem em uma população, nem sempre facilmente identificável, chamada população fonte. A pedra de toque da eficiência do estudo caso-controle é a seleção dos controles. Para garantir a estimabilidade das medidas RR e FA, aparte flutuações randômicas, a distribuição de exposições entre controles deve ser a mesma da população fonte de casos.

Estimativa da Razão de Proporção de Incidência a Partir de Estudos Caso-Controle

Suponhamos uma população fechada, onde, para facilitar a argumentação, se deseja estudar a relação entre uma exposição dicotômica (presente ou ausente) e uma doença. Esta situação simples, e na prática infreqüente, pode ser estendida para exposições contínuas ou categóricas em vários níveis. Para tanto, a população é subdividida em duas coortes, uma exposta e outra não exposta ao fator estudado. Em um típico estudo de coorte, após um período de seguimento, obtém-se primeiramente a proporção de incidência¹ da doença nas coortes exposta e não exposta. Em seguida, obtém-se a razão entre essas duas incidências, medida de risco relativo chamada razão de proporção de incidência¹ (RPI). Esta expressa a

associação entre a exposição e a doença que, respeitados fortes pressupostos, é considerada uma medida do efeito da exposição sobre a população¹. As informações necessárias para medir a RPI em um estudo como este são: totais de casos expostos (A), casos não expostos (C), tamanho da coorte exposta no início do seguimento (N_1) e tamanho da coorte não exposta no início do seguimento (N_0).

$$RPI = \frac{\frac{A}{N_1}}{\frac{C}{N_0}} = \frac{\frac{A}{C}}{\frac{N_1}{N_0}} \quad (1)$$

Como foi visto em (1), a RPI pode também ser entendida, em um estudo de coorte com estas características, como a razão entre as medidas A/C e N_1/N_0 , que representam o odds de exposição respectivamente entre casos e a população fonte.

Uma abordagem alternativa seria a execução de um estudo caso-controle aninhado na população fonte acima ($N_1 \cup N_0$), onde os controles seriam amostrados da população fonte no início do estudo, num momento anterior à ocorrência dos casos. Este modo de amostrar o grupo controle é conhecido na literatura como “case-base sampling”² ou “case-cohort design”³, e abre a possibilidade de um mesmo indivíduo ser amostrado como controle e posteriormente vir a ser caso, fazendo assim parte de ambos os grupos, casos e controles, simultaneamente. Neste estudo a razão do odds de exposição entre casos e controles (OR) seria estimada como:

$$OR = \frac{\frac{A}{C}}{\frac{b}{d}} = \frac{Ad}{bC} \quad (2)$$

onde b e d representam controles expostos e não expostos amostrados da população fonte de casos, respectivamente. Frequentemente, os casos estudados (a, c) são uma amostra aleatória do total de casos incidentes na população fonte (A, C).

Nessa situação, sendo a seleção de controles independente da exposição estudada, condição essencial em estudos caso-contro-

le⁴, a fração amostral de controles na coorte exposta, aparte flutuações amostrais, é igual à fração amostral na coorte não exposta¹, aqui chamada r:

$$\frac{b}{N_1} = r \rightarrow N_1 = \frac{b}{r} \quad (3)$$

$$\frac{d}{N_0} = r \rightarrow N_0 = \frac{d}{r}$$

Substituindo (3) e (2) em (1), e desconsiderando flutuações amostrais, temos:

$$RPI = \frac{\frac{A}{N_1}}{\frac{C}{N_0}} = \frac{\frac{A}{b/r}}{\frac{C}{d/r}} = \frac{Ad}{bC} = OR \quad (4)$$

Esta igualdade encontra-se demonstrada em Greenland e Thomas⁵.

Assim sendo, em um estudo caso-controle, onde

- a população fonte é perfeitamente identificada (como, por exemplo, em um estudo de base populacional),
- os casos são todos os que incidiram na população fonte durante um determinado período de seguimento, ou uma amostra aleatória deles,
- os controles são alocados aleatoriamente, a partir da população fonte, no início do seguimento, antes de os casos que serão estudados incidirem,

a RPI coincide assintoticamente com o OR, sem que seja necessária nenhuma presunção sobre a raridade da doença estudada^{1,2,5-8}.

Estimativa de Razão de Taxa de Incidência a Partir de Estudos Caso-Controle

Suponhamos agora uma situação mais comum, onde se deseja investigar a relação entre a exposição e a doença acima referidas a partir de uma população dinâmica, com indivíduos entrando e saindo sem restrições. A razão das taxas de incidência da

doença entre expostos e não expostos (RTI), também uma medida de risco relativo que, respeitados fortes pressupostos representa o efeito da exposição sobre a população estudada¹, pode ser diretamente obtida em um típico estudo de coorte como:

$$RTI = \frac{\frac{A}{T_1}}{\frac{C}{T_0}} = \frac{\frac{A}{C}}{\frac{T_1}{T_0}} \quad (5)$$

onde T_1 e T_0 representam as somatórias pessoa.tempo geradas nas coortes exposta e não exposta, respectivamente.

Uma abordagem alternativa seria a execução de um estudo caso-controle contemplando a alocação dos controles aleatoriamente na população fonte no momento em que os casos vão ocorrendo e sendo identificados. Este modo de amostrar o grupo controle é conhecido na literatura como “risk-set sampling”⁹, “sampling from the study base”¹⁰ ou “density sampling”¹¹. Também aqui, este modo de amostrar os controles abre a possibilidade de um mesmo indivíduo participar do estudo como caso e controle, simultaneamente. Entretanto, diferentemente da alocação de controles anteriormente comentada (“case-base sampling”²), onde cada indivíduo da população fonte (população fechada) tem a mesma chance de ser alocado como controle, neste caso a probabilidade de ser amostrado é proporcional ao tempo de seguimento do indivíduo na população fonte (população dinâmica).

Analogamente à situação anterior, a medida OR da associação entre exposição e doença seria estimada conforme a equação (2). Sob estas condições, temos⁸

$$\frac{b}{T_1} = r' \rightarrow T_1 = \frac{b}{r'} \quad (6)$$

$$\frac{d}{T_0} = r' \rightarrow T_0 = \frac{d}{r'}$$

Isto é, sendo a alocação dos controles independente da exposição⁴, a razão entre controles expostos e somatória pessoa.tempo na

população fonte exposta é, aparte flutuações randômicas, igual à razão entre controles não expostos e a população fonte não exposta¹, razão esta aqui denominada r' .

Substituindo (2) e (6) em (5), e desconsiderando flutuações amostrais, temos:

$$RTI = \frac{\frac{A}{T_1}}{\frac{C}{T_0}} = \frac{\frac{A}{b/r'}}{\frac{C}{d/r'}} = \frac{Ad}{bC} = OR \quad (7)$$

Esta igualdade encontra-se demonstrada em Greenland e Thomas⁵.

Assim sendo, em um estudo caso-controle onde:

- a população fonte é perfeitamente identificada (como, por exemplo, em um estudo de base populacional),
- os casos são todos os que incidiram na população fonte durante um determinado período de seguimento, ou uma amostra aleatória deles,
- os controles são alocados aleatoriamente, a partir da população fonte, no momento em que os casos vão incidindo e sendo identificados,

a RTI coincide assintoticamente com o OR, sem que seja necessária nenhuma pressuposição sobre a raridade da doença estudada^{1,2,5-8}.

Estimativa de Fração Atribuível a Partir de Estudos Caso-Controle

A medida fração atribuível (FA) informa a fração de todos os casos de um agravo ocorridos na população estudada atribuível a um conjunto de exposições. Trata-se de um conceito epidemiológico que relaciona o risco relativo de um agravo com a prevalência das exposições que se supõe estarem causalmente relacionadas a ele. É usualmente interpretada como a porcentagem de casos ocorridos em uma população que seriam evitados se fosse eliminada a exposição a um fator causal da doença¹². Daí decorre sua utilidade em Saúde Pública, parti-

cularmente quando se faz necessário escolher entre estratégias alternativas de prevenção¹². O conceito foi formulado pela primeira vez por Levin em 1953¹³, com o nome de risco atribuível (attributable risk).

A FA pode ser estimada da seguinte maneira^{1,4,14}:

$$FA = \frac{P(RR - 1)}{P(RR - 1) + 1} \quad (8)$$

onde P é a prevalência da exposição de interesse na população fonte e RR é a medida de risco relativo utilizada. Como visto acima, em estudos caso-controle onde a população fonte é perfeitamente identificável, sendo os controles alocados aleatoriamente a partir dela, obtém-se estimativas de RR. Nesta situação, a série de controles estima também a prevalência de exposição na população fonte de casos⁴, uma vez que aquela é uma amostra aleatória desta. Assim, reúnem-se as informações necessárias para estimar a FA de acordo com a equação (8).

O Mito da Doença Rara

Os primeiros estudos caso-controle começaram a ser realizados na década de 1920 do século passado, nos Estados Unidos⁴. Cole refere que até o final dos anos 50 eles eram “raramente executados, pobremente entendidos e pouco conceituados”¹⁵.

A década de 50 foi decisiva para o desenvolvimento metodológico destes desenhos¹⁶. Em 1951, Cornfield publica o primeiro estudo sobre este método¹⁷, demonstrando que as freqüências de exposição entre casos e controles, medidas de fácil obtenção, podem estimar um parâmetro de grande interesse epidemiológico: “razão da freqüência de doença entre indivíduos expostos relativa àquela entre indivíduos não expostos”¹⁷. Este parâmetro ganhou vários nomes e diferentes interpretações nos anos subseqüentes, sendo hoje amplamente conhecido como risco relativo. Neste texto, Cornfield argumenta que a razão do odds de exposição entre casos e controles é um bom estimador do risco relativo, contanto que a freqüência da doença

estudada seja rara. Em 1954, Cochran¹⁸ fornece as bases da análise estratificada e mostra como é possível combinar duas ou mais tabelas 2x2 em um único indicador sumariante de associação. Em 1955, Woolf¹⁹ propõe as bases da análise de *odds-ratio* utilizando a transformação logarítmica. Em 1956, Cornfield deduz um método de obtenção de intervalos de confiança para OR²⁰, procedimento rotineiramente executado nos dias de hoje por pacotes estatísticos. Em 1959, Mantel e Haenszel²¹ sistematizam alguns dos pressupostos dos estudos caso-controle e indicam dois métodos para a estimativa de odds ratio em estudos estratificados, largamente utilizados até o presente.

Estas análises, que conferiram credibilidade aos estudos caso-controle, tinham como referência um desenho com seleção de casos prevalentes. Daí o termo “razão de *freqüência* de doença” originalmente cunhado, ao invés de “razão de *incidência* de doença”, que seria mais adequado à idéia de risco relativo. Sob este desenho, onde os controles são alocados após a ocorrência de todos os casos (entre “sobreviventes” da doença), e sem considerações a respeito da população fonte, a raridade da doença estudada é pressuposto necessário para a estimação do RR⁸. Este conceito se difundiu amplamente na literatura epidemiológica internacional, ecoando na literatura nacional até os dias de hoje. Entretanto, dando seqüência ao refinamento metodológico dos estudos caso-controle, além de considerações importantes a respeito da identificação e do controle de vieses, novas abordagens amostrais foram incorporadas na década de 70, particularmente a partir do trabalho “Estimability and estimation in case-referent studies”, de Miettinen⁷. Assim, desenhos amostrais do tipo “case-base sampling”²² ou “risk-set sampling”⁹, aboliram a necessidade da pressuposição da raridade da doença para estimar risco relativo em estudos caso-controle, como visto acima.

Conforme afirma Pearce⁸, o significado das estimativas de odds ratio obtidas em estudos caso-controle difere se os controles são amostrados a partir das pessoas em ris-

co no início do seguimento para a identificação de casos (case-base sampling²), da experiência pessoa. tempo da população fonte de casos (risk-set sampling⁹) ou da população não doente ao final do seguimento. Estudos caso-controle utilizando estes três métodos de seleção de controles estimam RPI, RTI e OR, respectivamente. Nenhuma dessas estimativas depende de qualquer pressuposição sobre a raridade da doença estudada. A pressuposição de raridade é necessária apenas para que a estimativa de OR seja uma aproximação do risco relativo em

estudos com o desenho tradicional de Cornfield, Mantel e Haenszel, isto é, com amostragem de controles entre não doentes após já identificados todos os casos⁸.

A utilização do termo odds ratio para referir resultados de um estudo caso-controle com qualquer um dos desenhos acima citados é estatisticamente correta porque se trata mesmo de uma razão de odds de doença entre indivíduos expostos e não expostos. Entretanto, mais informativo é referir o que, de fato, do ponto de vista epidemiológico, está sendo estimado.

Referências

1. Rothman KJ, Greenland S. *Modern epidemiology*. 2nd edition. Philadelphia, PA: Lippincott-Raven; 1998.
2. Miettinen OS. Designs options in epidemiologic research: un up-date. *Scandinavian Journal of Work, Environment and Health* 1982; 9(Suppl 1): 7-14.
3. Prentice RL. A case-cohort design for epidemiological cohort studies and disease prevention trials. *Biometrika* 1986; 73: 1-11.
4. Schlesselman JJ. *Case-control studies - Design, conduct, analysis*. New York: Oxford University Press; 1982.
5. Greenland S, Thomas DC. On the need for the rare disease assumption in case-control studies. *Am J Epidemiol* 1982; 116: 547-53.
6. Rodrigues L, Kirkwood BR. Case-control designs in the study of common diseases: updates on the demise of the rare disease assumption and the choice of sampling scheme for controls. *Int J Epidemiol* 1990; 19: 205-13.
7. Miettinen OS. Estimability and estimation in case-referent studies. *Am J Epidemiol* 1976; 103: 226-35.
8. Pearce N. What does the odds ratio estimate in a case-control study? *Int J Epidemiol* 1993; 22(6):1189-92.
9. Robins JM, Gail MH, Lubin JH. More on biased selection of controls for case-control analysis of cohort studies. *Biometrics* 1986; 42:1293-9.
10. Miettinen OS. *Theoretical epidemiology - Principles of occurrence research in medicine*. New York: John Wiley; 1985.
11. Kleinbaum DG, Kupper LI, Morgenstern H. *Epidemiologic research - principles and quantitative methods*. Belmont, CA: Lifetime Learning Publications; 1982.
12. Coughlin SS, Jacques B, Weed DL. Attributable risk estimation in case-control studies. *Epidemiol Rev* 1994; 16(1): 51-64.
13. Levin ML. The occurrence of lung cancer in man. *Acta Unio International Contra Cancrum* 1953; 9: 531-41.
14. Breslow NE, Day NE. *Statistical methods in cancer research. Volume 1 - The analysis of case-control studies*. Lyon: IARC; 1980. (Scientific Publications; 32).
15. Cole P. Introduction. In: Breslow NE, Day NE. *Statistical methods in cancer research*. Geneva: IARC; 1980: 14-40.
16. Rego MAV. Aspectos históricos dos estudos caso-controle. *Cad Saúde Pública* 2001; 17(4): 1017-24.
17. Cornfield J. A method of estimating comparative rates from clinical data. Applications to cancer of the lung, breast and cervix. *J Nat Cancer Inst* 1951; 11: 1269-75.
18. Cochran WG. Some methods for strengthening the common X² tests. *Biometrics* 1954; 10: 417-51.
19. Woolf B. On estimating the relation between blood group and disease. *Ann Hum Genet* 1955; 19: 251-3.
20. Cornfield J. A statistical problem arising from retrospective studies. In: Neyman J. *Proceedings of The Third Berkeley Symposium*. Berkeley: University of California; 1956.
21. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Nat Cancer Inst* 1959; 22: 719-48.

recebido em: 25/04/2005

versão final reapresentada em: 09/06/2005

aprovado em: 10/06/2005