RISCO

em epidemiologia

Dois dogmas da epidemiologia

1. Os casos de doença não têm distribuição aleatória no espaço e no tempo.
2. Existem *factores de risco* que influenciam o aparecimento e/ou desenvolvimento da doença
   (genéticos, comportamentais, ambientais ...)

Objectivos deste módulo

Definir *risco*

Avaliar factores de risco no contexto de “estudos transversais”

Lógica epidemiológica

1. Suspeita de que a exposição a um factor pode influenciar a ocorrência de uma doença
   - Observações da prática clínica
   - Observação da distribuição da incidência no espaço/tempo
   - Especulação teórica

2. Hipótese: determinado factor aumenta/diminui o **risco** de ter doença

3. Condução de estudo epidemiológico
   - Testa-se a hipótese sobre associação entre factor e doença
   - O estudo é feito por comparação entre indivíduos com e sem doença

4. Etapes do estudo:

   Existe ou não associação ?
   Medir (quantitativamente) a associação
   Testar se é (estatisticamente) significativa.

Risco

Risco é uma probabilidade e está, portanto, contido no intervalo [0, 1]

Mais concretamente...

Risco é a probabilidade de ocorrência de um acontecimento considerado indesejável, num intervalo de tempo e/ou num contexto específico

Em Epidemiologia,

"acontecimento indesejável" ≅ ser infectado e/ou desenvolver doença.

contexto ≅ exposição a um factor de risco.

Como medir o risco ?

Exemplo

60% dos participantes que comeram lasanha num jantar contraíram salmonelose

Significa isto que:

1 - A probabilidade (risco) de comer lasanha e contrair salmonelose é 0.6 ?

2 – Comer lasanha é um factor de risco para a salmonelose ?

1 – Sim

se uma pessoa inquirida aleatoriamente nos disser que comeu lasanha, a probabilidade de ter contraido salmonelose é 0.6

Risco = (número dos expostos que adoeceu) / (número exposto ao factor de risco) = 0.6

2 – Não necessariamente.

Suponhamos que 60% dos que NÃO comeram lasanha tambem contrairam salmonelose ... Comer lasanha já não parece tão arriscado !

Risco relativo (RR)

Conclusão:

É necessário comparar o risco dos que comeram lasanha com o risco dos que não comeram

Risco relativo (RR) = $\text{risco}_{\text{expostos}} / \text{risco}_{\text{não-expostos}}$

Quociente entre o risco dos que foram expostos ao (potencial) factor de risco e o risco dos que não foram expostos

Se RR > 1 o factor aumenta o risco
Se RR < 1 o factor tem efeito “protector” (diminui o risco)
Se RR = 1 o factor é indiferente

Organização dos cálculos

	Estado de doença		
Factor de risco	*Doente*	*não doente*	*Total*
Exposto	*a*	*b*	a+b
não exposto	*c*	*d*	c+d
Total	*a+c*	*b+d*	$\boxed{n}$

Total da amostra

Tabela de Contingência bidimensional, 2 x 2

Bidimensional porque tem 2 variáveis ditas categóricas
(Ser exposto ou não; adoecer ou não)

2 x 2 - Porque cada variavel tem duas classes ou categorias

Associação:
Saber que um indivíduo está exposto dá alguma informação sobre se está doente ? (ou vice-versa)

Cálculos

	Doente	*não Doente*	*Total*
Exposto	*a*	*b*	a+b
não Exposto	*c*	*d*	c+d
Total	*a+c*	*b+d*	*n*

Risco para os expostos = a/(a+b)
Risco para os não expostos = c/(c+d)

Risco relativo

$$RR = \frac{a/(a+b)}{c/(c+d)} = \frac{a(c+d)}{c(a+b)}$$

Exemplo

	Doente	*não doente*	*Total*
Comeu	60	40	100
não comeu	20	35	55
TOTAL	80	75	155

Risco para os expostos : 60/100 = 0.6
Risco para os não expostos : 20/55 = 0.36

RR = 0.6/0.364 = 1.65

Quem comeu tem uma probabilidade 1.65 vezes maior de adoecer que quem não comeu
Quem comeu tem um risco 1.65 maior de adoecer

Evitar alarmismo desnecessário

O RR pode ser muito alto mas ...
.... O risco absoluto nos expostos pode ser muito pequeno,

Explo:

Probabilidade de uma reacção alérgica na população: 10^{-8}
(1 em 100 milhões)

Probabilidade da mesma reação após toma de um medicamento: 10^{-7}
(1 em 10 milhões)

RR = $10^{-7} / 10^{-8} = 10$ *!*

Em geral o que obtemos são estimativas

	Doente	*não Doente*	*Total*
Exposto	*a*	*b*	a+b
não Exposto	*c*	*d*	c+d
Total	*a+c*	*b+d*	*n*

Risco para os expostos = a/(a+b)
Risco para os não expostos = c/(c+d)

Risco relativo

$$\boxed{RR = \frac{a/(a+b)}{c/(c+d)} = \frac{a(c+d)}{c(a+b)}} \quad [1]$$

Se *n* fôr uma amostra, [1] é uma estimativa do verdadeiro valor de RR na população

Quão maior que 1 tem a estimativa de ser para ser considerada “significativa “ ?

Já agora, *a/(a+b)* também é uma estimativa do verdadeiro risco dos expostos
c/(c+d) também é uma estimativa do verdadeiro risco dos não expostos

Intervalo de confiança para o risco

Construir um Intervalo de Confiança (IC) dentro do qual deve estar o verdadeiro risco com elevada probabilidade (em geral adopta-se 95%)

O risco ***(r)*** amostral estima o risco médio, o qual é uma proporção, logo o seu **erro padrão** é dado por

$$e(r) = \sqrt{r(1-r)/n}$$

Usando a aproximação Normal à binomial, um IC a 95% para o risco é dado por,

$$[r - 1.96\, e(r), \quad r + 1.96\, e(r)]$$

1.96 é o quantil da Normal que delimita dos dois lados da curva a área α = 5%

Este procedimento pode ser feito para $r_{expostos}$ e $r_{não\ expostos}$

Intervalo de confiança para uma proporção

Intervalo de Confiança a 95% para

- Proporção de incidência
- Proporção de prevalência
- Mortalidade por doença

$$\left[p - 1.96\sqrt{\frac{p(1-p)}{n}}\,,\quad p + 1.96\sqrt{\frac{p(1-p)}{n}}\right]$$

Pressupõe:

- Amostra grande (n>30)
- Proporção verdadeira não muito próxima de 0 ou de 1
  idealmente para situações em que $0,2 < p < 0,8$

Não funciona bem para risco de ocorrências raras: doenças raras
mortalidade

Alternativa para p~0 ou p~1 : intervalos de Agresti-Coull

Intervalo de confiança para o RR

O IC para o RR é um pouco mais complicado

O **erro padrão** do *Ln*(RR) é estimado por

$$e(\ln RR) = \sqrt{\frac{1}{a} - \frac{1}{a+b} + \frac{1}{c} - \frac{1}{c+d}}$$

Um IC a 95% para o *Ln*(RR) é dado por,

$$\ln RR \pm 1.96\, e(\ln RR)$$

donde, o IC para RR:

$$LI = e^{\ln RR - 1.96\, e(\ln RR)}$$ *Limite inferior*

$$LS = e^{\ln RR + 1.96\, e(\ln RR)}$$ *Limite superior*

Se o IC **não** incluir 1, há 95% probabilidade de o factor ser mesmo de risco (ou de protecção)

Associação estatisticamente significativa

IC do RR > 1 indica associação estatisticamente significativa entre factor de risco e doença, mas…

NÃO implica que o factor de risco seja a causa da doença

associação, correlação $\neq$ causalidade

Redução do risco absoluto (RRA) ou 'risk difference'

	doença	não-doença	risco
Expostos	a	b	a/(a+b)
Não-expostos	c	d	c/(c+d)

Redução do risco absoluto (RRA) = $|\text{risco}_{\text{expostos}} - \text{risco}_{\text{não-expostos}}|$

$$RRA = \left|\frac{a}{a+b} - \frac{c}{c+d}\right|$$

A diferença entre o risco dos expostos e o risco dos que não foram expostos.

Representa o acréscimo de probabilidade de adoecer quando um não-exposto passa a exposto

Número necessário tratar, NNT

Qual o número de expostos que tem de passar a não exposto para se evitar 1 caso de doença

ou,

Qual o número de ‘doentes’ que têm de ser ‘tratados’ para se evitar 1 caso de doença ?

A resposta é o “número necessário tratar” ou NNT

$$NNT = \frac{1}{RRA}$$

Exemplo numérico

	Com doença respiratória	Sem doença respiratória	Total
Fumadores	6568	14420	20988
Não fumadores	4397	13134	17531
Total	10965	27554	38519

$$RRA = \left|\frac{a}{a+b} - \frac{c}{c+d}\right| = |0{,}31 - 0{,}25| = 0{,}06$$

Por cada 100 fumadores que não tivesse sido fumador, observar-se-ia uma redução de 6 doentes

$$NNT = \frac{1}{RRA} = \frac{1}{0{,}06} = 16{,}666$$

por cada 16,6 indivíduos que deixassem de fumar evitar-se-ia 1 caso de doença respiratória.

Em vez de risco... o odds

Por vezes os epidemiologistas expressam a ocorrência de doenças em termos de *odds*

Risco = Nº de ocorrências “favoráveis” / Total de ocorrências

Explo: Nº que comeu lasanha e adoeceu / Nº total que comeu lasanha

Odds = Nº ocorrências “favoráveis” / Nº ocorrências “desfavoráveis”

Explo: Nº que comeu lasanha e adoeceu / Nº que comeu lasanha e não adoeceu

Exemplos

Nesta disciplina há 20 alunos inscritos

Se eu escolher um aluno aleatoriamente, qual a probabilidade de seres <u>tu</u> ?

Risco (probabilidade) = 1/20 = 0.05

Odds = 1/19 = 0.053

Em 100 alunos de Epidemiologia, 60 tiveram gripe em 2013-14

Qual o risco e o odds de ter gripe ?

Risco (probabilidade) = 60/100 = 0.6

Odds = 60/40 = 1.5

<u>Não</u> é uma probabilidade

“Odds”, “Odds Ratio” *(“excedências”, “razão de excedências”)*

	Doente	*não Doente*	*Total*
Exposto	*a*	*b*	a+b
não Exposto	*c*	*d*	c+d
Total	*a+c*	*b+d*	*n*

$$\textit{Odds} = \frac{\text{número de vezes que acontecimento ocorre}}{\text{número de vezes que não ocorre}}$$

Odds para os expostos = *a/b*
Odds para os não expostos = *c/d*

Estimador do *Odds ratio (OR)*

$$\hat{OR} = \frac{a/b}{c/d} = \frac{ad}{bc}$$

Se OR > 1 o factor aumenta mesmo o risco
Se OR < 1 o factor tem efeito “protector” (diminui o risco)

Intervalo de confiança para o OR

O **erro padrão** do OR é estimado por $$e(\ln OR) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

Um IC a 95% para o OR é dado por,

$$LI = e^{\ln OR - 1.96 e(\ln OR)}$$

$$LS = e^{\ln OR + 1.96 e(\ln OR)}$$

Se o IC **não** incluir 1, há 95% probabilidade de o factor ser mesmo de risco (ou de protecção)

RR ou OR ?

	Doente	*não Doente*	*Total*
Exposto	*a*	*b*	a+b
não Exposto	*c*	*d*	c+d
Total	*a+c*	*b+d*	*n*

Risco é uma probabilidade

Odds não é uma probabilidade

$$a/(a+b) \in [0,1]$$

$$a/b \in [0,+\infty[$$

Damos preferência ao risco, mas…

Nos estudos caso-controlo <u>não</u> se pode usar RR

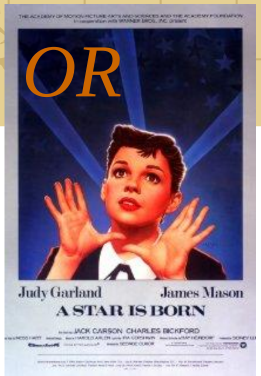
Contudo,

O OR é frequentemente uma boa aproximação ao RR

Se a doença fôr rara, *a* e *c* são muito pequenos

$$\frac{a/b}{c/d} \approx \frac{a/(a+b)}{c/(c+d)}$$

Nasceu uma estrela



O OR é utilizavel em todo o tipo de planeamentos epidemiológicos

O OR de doença é igual ao OR de exposição

Odds dos expostos contrairem doença relativa/ aos não expostos $$\hat{OR} = \frac{a/b}{c/d} = \frac{ad}{bc}$$

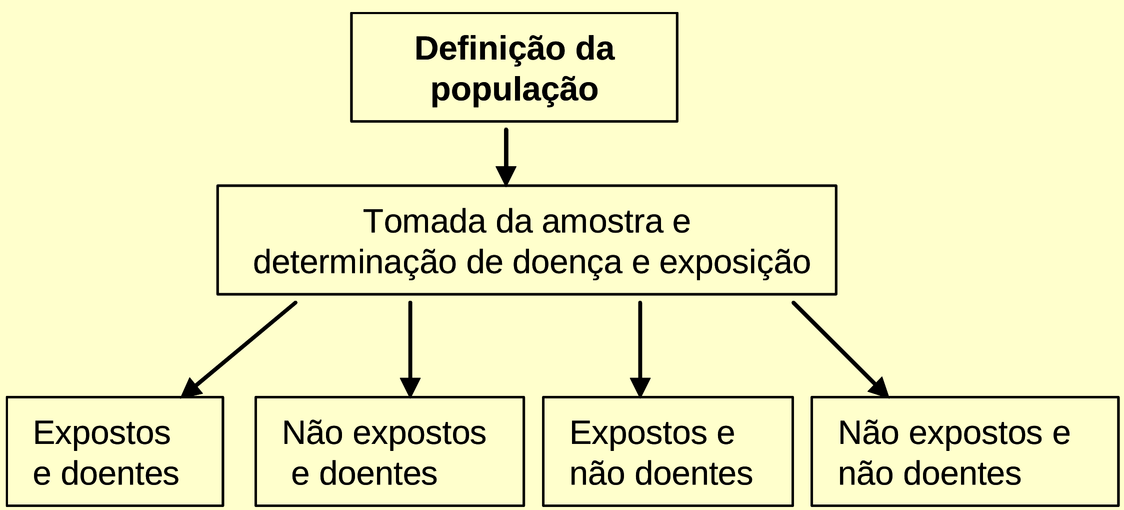
Odds dos doentes estarem expostos relativa/ aos não doentes $$\hat{OR} = \frac{a/c}{b/d} = \frac{ad}{bc}$$

O OR é obtido directamente em regressão logística

Numéricamente, OR > RR A diferença aumenta com a incidência da doença

Estudos transversais (cross-sectional)

- Uma *<u>única</u>* amostra tomada na população
- Durante espaço de tempo relativamente *<u>curto</u>*
- Mede-se *<u>simultaneamente</u>* – prevalência da doença e exposição ao factor risco



Impacto na comunidade

RR e OR medem associação entre o factor de risco e a doença

Nada dizem acerca do impacto sobre a **comunidade** de alterar o factor de risco

Exemplo

RR dos doentes com SIDA desenvolverem TB é RR=4
RR dos fumadores desenvolverem TB é RR = 2

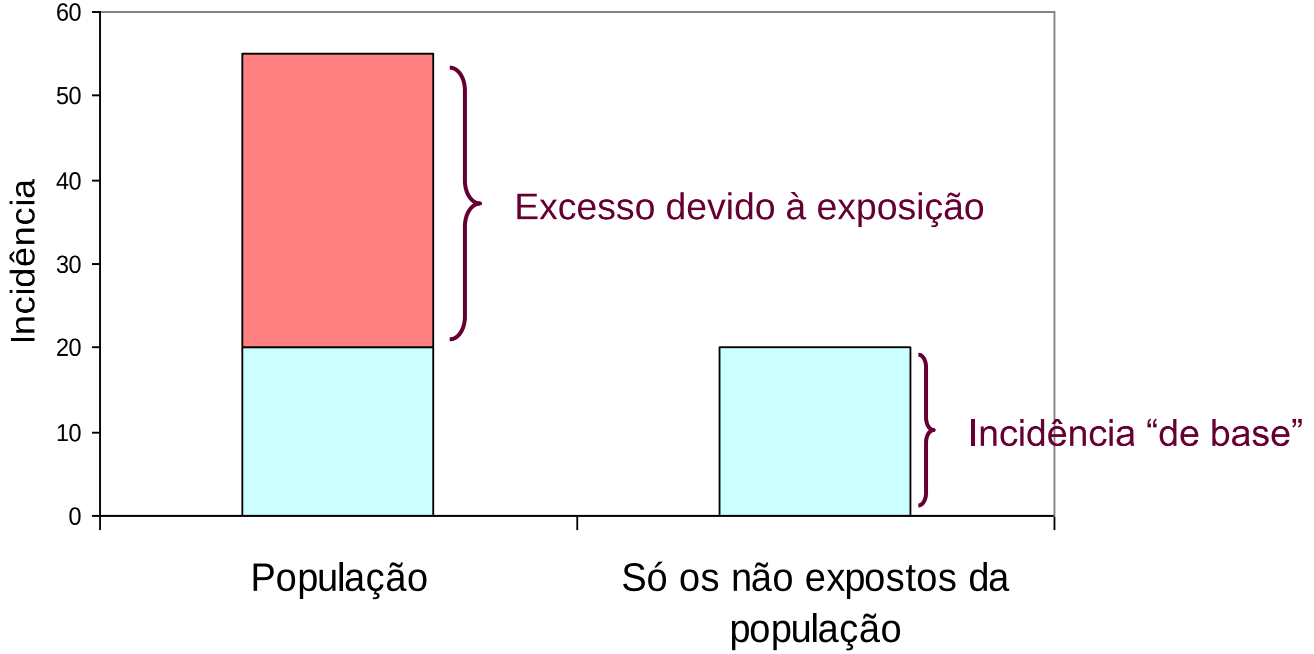
Retardar a evolução de VIH para SIDA tem maior impacto sobre a TB que promover campanhas anti-tabaco entre os jovens ?

Se a prevalência de fumadores for alta, provavelmente não !

Em saúde pública, há que ter em consideração simultaneamente:

1. O Risco Relativo
2. A prevalência do factor de risco na população

Risco atribuível ou ‘etiologic fraction’



- RA é a fracção de doença na população total atribuível ao factor de risco
- Mede a redução da doença na população caso o factor de risco fosse eliminado
- Ajuda a determinar medidas de saúde pública eficazes

O risco atribuível na população

Risco Atribuível na população

É uma combinação do RR com a prevalência do factor de risco
(há mais de uma forma de o fazer)

Seja T = proporção incidência (ou prevalência) total na população
T_{NE}= proporção incidência (ou prevalência) na população dos NÃO expostos ao factor de risco

$T–T_{NE}$ = excesso de casos de doença devidos ao factor de risco

$(T–T_{NE})/T$ = proporção de casos devidos ao factor de risco

Há mais de uma forma de construir este quociente !

Risco Atribuível - cálculo

	Doente	*não Doente*	*Total*
Exposto	*a*	*b*	a+b
não Exposto	*c*	*d*	c+d
Total	*a+c*	*b+d*	*n*

Assumindo que *n* é amostra aleatória da população:

(a+c)/n estima proporção total de doentes na população (risco total de doença)

c/(c+d) estima proporção de doentes entre os não expostos (risco entre os não expostos)

Risco Atribuível

$$\hat{RA} = \frac{\frac{a+c}{n} - \frac{c}{c+d}}{\frac{a+c}{n}} = \frac{r - r_{NE}}{r}$$

RA mede <u>proporção</u> de casos em excesso na <u>amostra</u>, logo,

1. Multiplicar por incidência da doença na população para ter <u>número</u> de casos em excesso na <u>população</u> (usar por explo stats oficiais)
2. Faz sentido calcular um IC para o verdadeiro RA

IC para RA(risco atribuível)

Um IC para RA pode ser obtido por:

$$\frac{(ad-bc)e^{\pm u}}{nc+(ad-bc)e^{\pm u}}$$

Para um IC de 95%:

$$u=\frac{1.96\,(a+c)(c+d)}{ad-bc}\sqrt{\frac{ad(n-c)+c^2b}{nc(a+c)(c+d)}}$$

Whittmore, AS. 1983. Estimating attributable risk from case-control studies. *Am J Epidemiol* **117:**76-85.

Ainda o Risco Atribuível – equação de Levin

Outra equação para o RA

$$RA = \frac{p_E(RR-1)}{1+p_E(RR-1)}$$

p_E = prevalência do factor de risco na população
RR = risco relativo

Levin ML. 1953. The occurrence of lung cancer in man. *Acta Intern Cancer* **9**:531

Leviton, A. 1973. Definitions of attributable risk. *Am J Epidemiol* **98**: 231.

Conduz aos mesmo resultados para RA, caso p_E e RR sejam obtidos a partir da tabela de contingência habitual

Mas,

Permite trabalhar com p_E e RR obtidos de fontes independentes, como estudos caso-controlo e estudos de coortes

Nota
O IC atrás apresentado NÃO é válido nestes casos !

Três medidas de associação

Variável 2

Variável 1

	Doente	*não Doente*	*Total*
Exposto	*a*	*b*	a+b
não Exposto	*c*	*d*	c+d
Total	*a+c*	*b+d*	*n*

RR e OR são uma das muitas formas de medir "Associação" entre as duas variáveis

Associação
O facto de se saber que um indivíduo pertence a uma categoria da variável 1 dá alguma informação sobre a categoria a que ele pertence na variável 2 ?

Se sim, existe associação. Se não, não existe

Outra medida de associação muito popular em tabelas é a estatística do qui-quadrado

$$\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Vantagem – permite avaliar se a associação é estatisticamente significativa

Desvantagens – é pouco "portavel"
- interpretação não imediata por não-estatísticos