

METODOLOGIAS DE INVESTIGAÇÃO EM CIÊNCIAS DA SAÚDE

Manuel Carmo Gomes

**Pós-graduação em Gestão e Administração em Saúde
Faculdade de Economia, Universidade do Algarve**

2012

A investigação e os seus métodos

A investigação é o processo sistemático e rigoroso que tem por objectivo descrever fenómenos, desenvolver explicações e testar teorias. Em última instância, a investigação visa contribuir para alargar o conhecimento de uma determinada área científica e, neste curso, centramo-nos nos métodos de investigação em ciências da saúde.

A crescente exigência de avaliações de custo-efectividade e de qualidade dos serviços prestados em ciências da saúde, tornam a investigação prioritária nesta área. Ao procurar evidência em favor de qualquer prática clínica ou terapêutica, apenas os estudos que seguem de perto os padrões da metodologia de investigação científica em todas as fases, são realmente de valor. São particularmente importantes o planeamento do estudo, a forma como os dados são obtidos, a análise estatística dos mesmos e a sua interpretação. Este curto curso tem por objectivo apresentar os principais métodos usados para planear estudos e testar hipóteses em ciências da saúde, ajudando o estudante a escolher a metodologia que melhor se adequa a conseguir obter respostas para as suas perguntas.

Não há metodologias absolutamente melhores que outras, há apenas metodologias mais adequadas para uma determinada pergunta. Se a pergunta é descritiva, por exemplo, qual o estado de saúde da população X ? então é necessário efectuar uma observação transversal da população em causa e desde logo é necessário decidir se se observam todos os indivíduos ou se se toma uma amostra da população. Os resultados, uma vez obtidos, poderão ainda permitir responder a perguntas secundárias como, por exemplo, existem diferenças entre os sexos ou entre grupos etários quanto ao estado de saúde ?

Se a pergunta diz respeito à eventual associação entre o estado de saúde e uma determinada medicação ou um determinado tratamento, a metodologia mais adequada é diferente. Por exemplo, os indivíduos que fizeram quimioterapia após diagnóstico de cancro e que mudam para uma dieta predominantemente vegetariana, têm menor probabilidade de vir a ter uma recidiva ao longo da vida do que os que não mudam os seus hábitos ? é uma pergunta que requer um estudo longitudinal, no qual os dois grupos (os que mudam e os que não mudam a dieta) são acompanhados ao longo do tempo e é registada a frequência de recidivas em cada um.

Se as perguntas equacionam a tentativa de estabelecer uma relação causa-efeito, muitas vezes o planeamento experimental ou, como se diz em ciências da saúde, o 'ensaio clínico' pode ser o método mais adequado. Por exemplo, a fisioterapia tem efeitos benéficos para os doentes com osteoartrite do joelho ? é uma pergunta que pode ser investigada organizando dois grupos de doentes de forma aleatória, um faz fisioterapia e o outro serve de grupo de controlo, sendo depois os resultados do

primeiro grupo comparados com os dos controlos. Nem sempre este tipo de método experimental é exequível, realista ou apropriado, e as relações causa-efeito por vezes têm de ser aproximadas a partir de abordagens retrospectivas (o que aconteceu no passado), não aleatorizadas (os membros de cada grupo não são amostras aleatórias) ou com indivíduos que participam no estudo durante períodos muito diferentes (são recrutados em instantes diferentes, ou desistem de participar no estudo, ou ficam incomunicáveis...). A metodologia a usar depende das perguntas, dos meios logísticos, do tempo e, quase sempre, do financiamento disponível.

Os métodos de investigação não devem ser vistos isoladamente. A complexidade das ciências da saúde obriga muitas vezes a combinação de metodologias para investigar facetas diferentes do mesmo problema e isso contribui para aumentar a validade externa das conclusões alcançadas, por outras palavras, a sua representatividade e a forma como podem ser extrapoladas para outras comunidades ou ambientes. A multidisciplinaridade das ciências da saúde conduziu ao desenvolvimento de uma panóplia muito vasta de métodos. Esta diversidade é enriquecedora, pelos motivos apontados, mas torna a sua apreensão mais difícil para o estudante que se inicia na investigação. Neste curso procura-se por isso apresentar de forma sistematizada os principais grupos de métodos de investigação e as situações em que, tipicamente, são mais adequados. Todos os métodos têm vantagens e inconvenientes, os quais são sempre apresentados. A comparação de vantagens e inconvenientes, não só é uma forma de comparar métodos entre si, mas também serve de guia para efectuar a avaliação da validade dos próprios resultados.

Sinopse das matérias cobertas pelo curso

Na investigação em ciências da saúde, tem-se assistido a grandes avanços resultantes da coalescência dos estudos epidemiológicos com os estudos clínicos. A epidemiologia estuda a distribuição dos casos de doença na população e a sua associação a factores de risco. Investiga perguntas como, por exemplo, o número de casos de cancro do pâncreas tem vindo a aumentar ? aumentaram mais nos homens ou nas mulheres ? existem precauções que se possam tomar para diminuir a probabilidade de desenvolver cancro ao longo da sua vida ? Trata-se de uma ciência de números, na qual se contabilizam casos de doença e a forma como surgem ao longo do tempo e se distribuem na população por sexo, idade, área geográfica, etc. Muitos métodos de investigação foram desenvolvidos nesta ciência, mas são agora também aplicados na clínica para avaliar a efectividade de tratamentos e de fármacos novos, em particular sempre que é necessário ... contar casos.

Este curso inicia-se com uma introdução à terminologia que, sendo há muito familiar aos epidemiologistas, está agora generalizada a todas as áreas das ciências da saúde em que existe necessidade de contar casos de doença e a forma como estes se distribuem

ao longo do tempo, no espaço geográfico, entre diferentes grupos de indivíduos, etc.. Após a terminologia, são introduzidos conceitos que permitem comparar a probabilidade de contrair doença entre indivíduos que diferem quanto a um critério pré-definido como, por exemplo, se fazem ou não uma determinada terapia ou medicação, o que comem, onde habitam, etc. Na segunda parte, apresentam-se de forma sistemática as principais metodologias de investigação em ciências da saúde. Apresentam-se os estudos transversais (ou 'cross-section'), os estudos caso-controle, os estudos de coortes e, finalmente, os ensaios clínicos. A apresentação destes métodos é aproveitada para gradualmente introduzir conceitos elementares importados de áreas diversificadas do conhecimento, como a epidemiologia, a teoria da amostragem, os testes de hipóteses e até a teoria filosófica do conhecimento científico.

Metodologias de Investigação em Ciências da Saúde

<http://webpages.fc.ul.pt/~mcgomes/aulas/UA1g>

Manuel Carmo Gomes
mcgomes@fc.ul.pt



Horário Set 2012

6ª f 18:30-20:00h; 20:30-22:00h
Sáb 9:30-11:00h; 11:30-13:00h

14-15 Set
21-22 Set



Investigação científica

Processo sistemático e rigoroso de descrição, medição, experimentação e de teste de hipóteses que visa, em última instância, o avanço do conhecimento.

Metodologia científica - conjunto de técnicas e procedimentos para efectuar investigação científica.

Avanço do conhecimento em...
CIÊNCIAS DA SAÚDE



Não há “o melhor” método

O(s) método(s) a adoptar depende(m) da pergunta

No Algarve a prevalência de diabetes é superior ao resto do país ?
(Depende do sexo ? Varia com os grupos etários ? ... etc)

A ingestão sistemática de uma dieta rica em vegetais, no seguimento de quimioterapia, diminui a probabilidade de recidiva do cancro da mama ?

Dados 3 broncodilatadores, qual o mais eficaz para controlar um ataque de asma ?

O medicamento Tafamidis impede o desenvolvimento da paramiloidose (doença dos pezinhos ?)



Epidemiologia e clínica



Estudos epidemiológicos

grupos, populações

Estudos clínicos

indivíduo,
médico-doente

Métodos de investigação em Ciências da Saúde

Medicina Baseada na Evidência (MBE)

Tópicos de MI em CS



Módulo	Metodologia de Investigação em CS
0	Apresentação
2	Epidemiologia descritiva
3	Estudos transversais
4	Estudos caso-controlo
5	Estudos de coortes
17	Ensaio clínico
6	Testes de diagnóstico

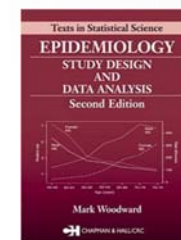
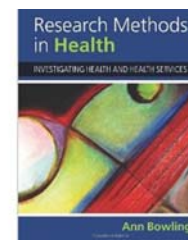
The screenshot shows a website for 'Metodologias de Investigação em Ciências da Saúde'. It includes a header with 'MCG Manuel Carmo Gomes' and a URL: <http://webpages.fc.ul.pt/~mcgomes/aulas/UAIG/>. The main content area features a sidebar with navigation links (HOME, OBJETIVOS, AÇÕES, INVESTIGAÇÃO, TESTES, SÍMBOLOS, ÍNDICE) and a central section titled 'Metodologias de Investigação em Ciências da Saúde'. Below this, there is an 'AVISOS' section with a 'Calendário de aulas em 2011' listing dates: 21 e 22 OUT-2011 and 28 e 29 OUT-2011. A 'Sinopse da unidade curricular' section follows, listing three topics: 1. Apresentação (Organization of the unit: investigation and methodologies), 2. Introdução à epidemiologia descritiva (Definitions and elementary concepts: incidence, prevalence, mortality, rates and standardizations), and 3. Risco e estudos 'cross-sectional' (Risk and 'cross-sectional' studies). Each topic is accompanied by a small icon and a date of update (9 Oct 11).

Leituras recomendadas



Bowling, A. 2009. 3rd Ed. *Research Methods in Health. Investigating Health and Health Services*. McGraw-Hill. Open Univ Press.

Woodward, M. 2004, 2nd Ed. *Epidemiology, Study Design and Data Analysis*. Chapman & Hall





Epidemiologia descritiva

Definições, taxas, confundimento



Epidemiologia

Estudo da *distribuição* (temporal e espacial) das doenças e dos seus *determinantes*.

Distribuição

Casos de doença por grupos etários, área geográfica, sexo, raça, hábitos de higiene...
- *Epidemiologia Descritiva*

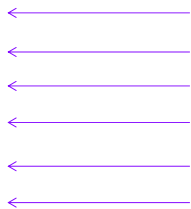
Determinantes (ou factores de risco)

Factores que promovem o aparecimento da doença.

Explo - o tétano é causado pela bactéria *Clostridium tetani*.
Quais os factores que promovem a formação de feridas e contacto c/ a bactéria ?
Quais os factores que promovem a não vacinação ?

- *Epidemiologia Analítica (medidas de associação)*

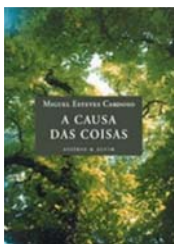
Associações doença – factores de risco



Objetivos

- Definições e terminologia básica da epidemiologia
- Planear experiências para medir associação entre doença e seus determinantes
- Medir a associação e decidir se é significativa

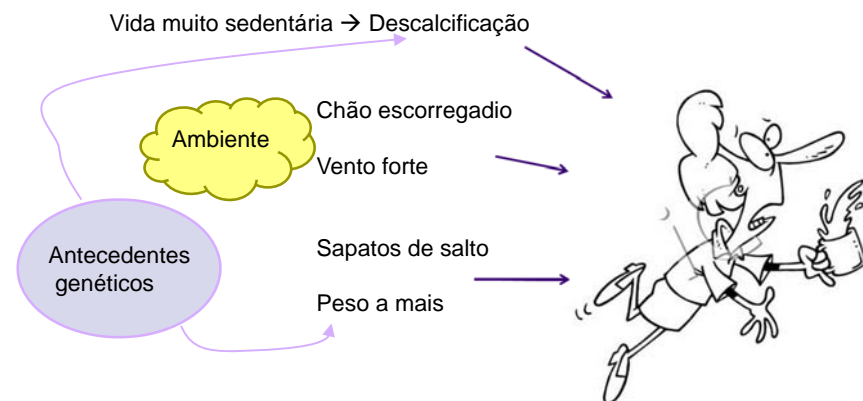




Para a maioria das doenças:

- O conceito de 1 causa => 1 efeito é ingénuo
- Há múltiplas causas - Necessárias
 - Suficientes
- As causas ambientais e genéticas são omnipresentes
- As causas interagem
- A ordem pela qual as causas surgem é em geral relevante
- Os tempos que decorrem entre cada causa e a doença variam

A causa das coisas é sempre multifactorial...



Casos de doença



Definição clínica

Baseada na sintomatologia observada pelo profissional de saúde sem recurso ao laboratório

Definição laboratorial

- Baseada em diagnóstico lab (citologia, radiologia, cirurgia, etc...)
- Baseada no isolamento do agente etiológico da doença ou dos seus produtos (ácidos nucleicos, antígenos, toxinas) a partir de produtos fisiológicos (sangue, urina, etc) do possível doente.

(mais demorada, mais cara)



Yersinia pestis

Caso possível – preenche algum(s) sinal clínico

Caso provável – preenche todo ou quase todo o quadro clínico

Caso confirmado – foi isolado o agente etiológico

Contagem de casos I



Duas formas de contabilização:

Incidência

Número de **novos** casos de doença por unidade tempo
É a incidência acumulada ao longo dum intervalo de tempo

Prevalência

Número de indivíduos infectados ou doentes num dado instante de tempo

Usada em especial para infecções de longa duração (tuberculose, hepatite B, VIH ...)

Mortalidade I



Mortalidade

Número absoluto de mortes por doença numa unidade de tempo (em geral 1 ano) ...ou, o mesmo, mas por habitantes

Expo: número de mortes por tuberculose em 2011 por 100 mil habitantes

Mortalidade específica

Número de mortes por doença num grupo específico

Expo: número de mortes por pneumonia em > 65 anos

número de mortes por pneumonia em > 65 anos por 100 mil habitantes

o numerador e denominador têm de se referir ao mesmo grupo específico (> 65 anos) !

Mortalidade II



Mortalidade por doença, ou Letalidade por doença, ou "Case fatality rate" (CFR)

Número absoluto de mortes por doença, por unidade de tempo, dividido pelo número de casos de doença onde as mortes ocorrem

Expo:

$$\text{CFR da TB} = \frac{\text{Número de mortes por TB em 2010}}{\text{Número de doentes com TB em 2010 (= prevalência)}}$$

No fundo, é uma percentagem, mas referente a uma unidade de tempo

Mortalidade absoluta e mortalidade por doença (CFR) podem ser muito diferentes

Expo: a mortalidade por meningite na Europa é baixa, a sua CFR é elevada

Exemplo: Gripe aviária



O relato da CFR com conhecimento imperfeito do denominador pode ser muito perturbador.

Em 20 Janeiro de 2012:

582 casos de gripe aviária (H5N1) em humanos
343 mortes

CFR aparente $59\% = (343/582) \times 100$

O denominador está correcto ?

Não se sabe quantos infectados com H5N1 existem sem terem dado entrada hospitalar !

Outros indicadores: incidência elevada em adolescentes e adultos pneumonias directas pelo vírus



"Taxas"



Incidência e prevalência são em geral apresentadas na forma de proporções que os epidemiologistas designam por "taxas" (rates)

$$\text{"Taxa" de Incidência} = \frac{\text{Número de novos doentes durante o ano}}{\text{População a meio do ano}} \times 100 \text{ mil}$$

$$\text{"Taxa" de Prevalência} = \frac{\text{Número de doentes a meio do ano}}{\text{População a meio do ano}} \times 100 \text{ mil}$$

"Taxa" : (variável aleatória / constante)

idealmente:

Denominador: só a população em risco !

Exercício



Em 2001, de acordo com o censo realizado nesse ano, havia em Portugal 10356117 habitantes. No mesmo ano, morreram dentro do território 105582 portugueses, dos quais 242 por tuberculose. Estima-se que o número de pessoas que estavam em tratamento de tuberculose durante o ano rondou os 4 mil.

- Qual foi a taxa de mortalidade em Portugal (por 100 mil habitantes) ?
- Qual foi a taxa de mortalidade por tuberculose (por 100 mil) ?
- Qual foi a prevalência de tuberculose (por 100 mil) ?
- Qual foi a taxa de letalidade da tuberculose ("case fatality rate")?
- Sabendo que a incidência anual da tuberculose foi de 43,42 casos por 100 mil habitantes, estime quantos casos novos de tuberculose houve em 2001.
- Os dados fornecidos nas alíneas acima são todos verídicos. Quais os que lhe inspiram menor confiança?

Populações e amostras



População alvo

- população acerca da qual pretendemos tirar conclusões

População de estudo

- população específica a partir da qual os dados são recolhidos

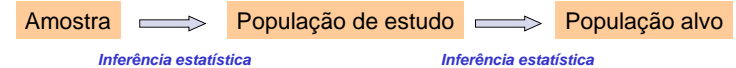
A população de estudo é *representativa* da população alvo?

Tem a mesma distribuição das características determinantes da doença ?
exemplo: mesma distribuição de idades ?
mesmas classes sócio-económicas ?

Amostra

- Conjunto de indivíduos tomados aleatoriamente a partir da população de estudo

É *representativa* da população de estudo ?



Exercício



Um investigador pretende conhecer a percentagem de aves selvagens que são seropositivas para o vírus H1N1 da gripe na região de Lisboa e Vale do Tejo (RLVT). Para isso delimita uma área no estuário do Tejo e, durante 5 dias, dirige-se lá e apanha 10 aves aleatoriamente por dia, retirando-lhes uma amostra de sangue para análise. Qual das seguintes afirmações é verdadeira,

- A população de estudo é a população de aves da RLVT
- A população alvo é o conjunto das 50 aves às quais foi retirado sangue
- A inferência da população de estudo para a população alvo, pressupõe que a área delimitada no estuário teve certos critérios em consideração (que critérios, já agora ?)

Confundimento no cálculo de taxas



Onde é que a taxa de mortalidade anual é mais elevada ?

Centro urbano industrializado



Retiro turístico 'year-round'



Variáveis de confundimento

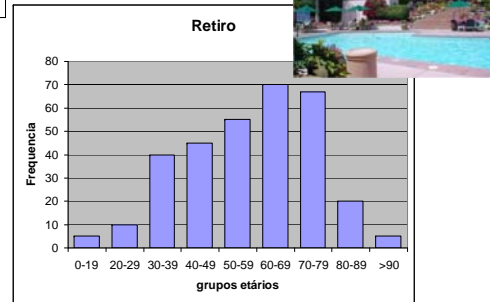
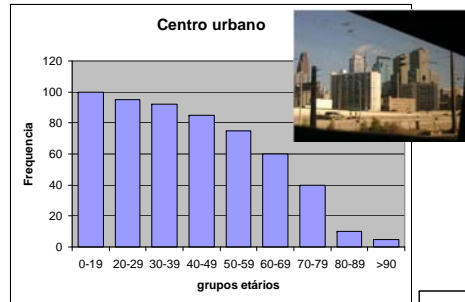
A idade, neste exemplo, é uma variável de **confundimento**

Altera a relação entre outras duas variáveis:

- Taxa de mortalidade
- Tipo de ambiente

de forma “escondida”

A observação da associação entre as duas variáveis em estudo *dentro dos estratos* da variável de confundimento, pode ser diferente da associação existente quando se ignoram esses estratos



Confundimento

Confundimento significa que ...

Existe uma variável (variável 2, idades), diferente daquela que está sob investigação (variável 1, meio ambiente), que se distribui de forma diferente dentro dos “estratos” da variável 1.

A proporção (jovens:velhos) é diferente no centro urbano e no retiro turístico.

Mesmo que a variável 1 não esteja associada à doença, a variável 2 vai causar diferenças entre expostos e não-expostos à variável 1

A variável de confundimento, além de estar associada à doença, tem comportamento diferente entre os estratos da variável sob investigação

Um exemplo detalhado

Melhorias no saneamento, assistência médica, esgotos ...

1970 → 2000

Contudo, ...

1970			2000		
População	Mortes	Mortes por 100 mil	População	Mortes	Mortes por 100 mil
900000	862	96	900000	1130	126

Idades (anos)	1970			2000			
	População	Mortes	Taxa de mortalidade	População	Mortes	Taxa de mortalidade	
<i>todas</i>	900000	862	96	<i>todas</i>	900000	1130	126
20-39	500000	60	12	20-39	300000	30	10
40-59	300000	396	132	40-59	400000	400	100
60+	100000	406	406	60+	200000	700	350

Padronização directa



Se a estrutura etária em 2000 fosse a mesma que em 1970, qual teria maior mortalidade total ?

1. Construir uma "população padrão" (PP) onde a variável de confundimento (idade) seja conhecida
2. Calcular as taxas de mortalidade **esperadas** na PP, aplicando as taxas de mortalidade de 1970 e 2000 **dentro dos estratos** de idade
3. Comparar mortalidade total esperada entre 1970 e 2000

Exemplos de PP:

A soma (idade a idade) das populações sob comparação

População mundial padrão fornecida pelas Nações Unidas

Padronização directa: aplicação



PP= soma das duas populações

Idades	População padrão	Taxa de 1970 por 100 mil	Mortes esperadas pela taxa de 1970	Taxa de 2000 por 100 mil	Mortes esperadas pela taxa de 2000
todas	1800000				
20-39	800000	12	96	10	80
40-59	700000	132	924	100	700
60+	300000	406	1218	350	1050
Total			2238		1830
Taxas na população-padrão ajustadas pela idade					
1970	$2238/1800000 = 124,3$				
2000	$1830/1800000 = 101,7$				

$$96 = (12/100\ 000) \times 800\ 000$$



RISCO

em epidemiologia e ciências da saúde

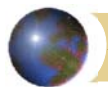


Factor de risco

Um factor (ou “exposição”) que se suspeita estar associado a uma doença, no sentido em que a sua presença aumenta a probabilidade de doença

Pode ser:

Comportamental
Genético
Ambiental
Social
Etc.



Lógica epidemiológica-clínica

1. Suspeita de que a exposição a um factor pode influenciar a ocorrência de uma doença, devido a...
 - Observações da prática clínica
 - Observação da distribuição da incidência no espaço/tempo
 - Especulação teórica
2. Formulação de uma hipótese
3. Condução de estudos analíticos
Testa-se hipótese sobre a associação entre factor e doença
4. Progressão do estudo:
*Existe ou não associação ?
Medir 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 Ciências da Saúde,

“acontecimento indesejável” \cong ser infectado e/ou desenvolver doença.
contexto \cong exposição a um factor de risco.



Como medir o risco ?



Exemplo

60% dos participantes que comeram marisco num jantar contraíram salmonelose

Significa isto que:

- 1 - A probabilidade de comer marisco no jantar e contrair salmonelose é 0.6 ?
- 2 - Comer marisco é um factor de risco para a salmonelose ?

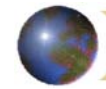
1 - Sim

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

$$\text{Risco} = (\text{número dos expostos que adoeceu}) / (\text{número exposto ao factor de risco}) = 0.6$$

2 - Não necessariamente.

Suponhamos que 60% dos que NÃO comeram marisco também contraíram salmonelose ... Comer marisco já não parece tão arriscado !



Risco relativo (RR)

Conclusão:

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

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

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

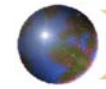
Factor de risco	Estado de doença		Total
	Doente	não doente	
Exposto	a	b	a+b
não exposto	c	d	c+d
Total	a+c	b+d	n

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 variável tem duas classes ou categorias



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

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



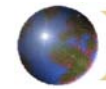
Exemplo

	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



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 = $c/(c+d)$

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

Se n for 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

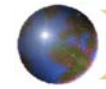
Construir um **Intervalo de Confiança (IC)** dentro do qual deve estar o verdadeiro risco com elevada probabilidade (em geral adopta-se 95%)

Um IC a 95% para o risco (r) é dado por,

$$r \pm 1.96 \sqrt{\frac{r(1-r)}{n}}$$

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

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



Intervalo de confiança para o RR

O IC para o RR é um pouco mais complicado

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

o IC para RR:

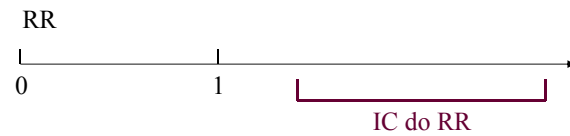
$$LI = e^{\ln \hat{RR} - 1.96 \hat{e}(\ln \hat{RR})} \quad \text{Limite inferior}$$

$$LS = e^{\ln \hat{RR} + 1.96 \hat{e}(\ln \hat{RR})} \quad \text{Limite superior}$$

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



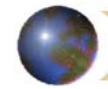
Associação estatisticamente significativa



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



Exercício

Numa aldeia sem água canalizada, surgem em poucas semanas vários casos de cólera que despertam a atenção das autoridades. Um investigador enviado ao local, suspeita que a origem da doença está na água de um poço e escolhe *aleatoriamente* 40 pessoas, às quais pergunta se usaram ou não o poço recentemente e se adoeceram ou não. Os resultados resumem-se na seguinte tabela,

Factor de risco	Estado de doença		Total
	Doente	não doente	
Usou o poço	8	10	18
não usou o poço	2	20	22
Total	10	30	40

- Qual é o risco de adoecer nesta aldeia ? Construa um IC a 95% para este risco.
- Qual é o risco de usar o poço e adoecer ?
- Qual é o RR e o OR de quem usou o poço adoecer ?
- Construa IC's a 95% para o RR e o OR
- Existe associação entre beber do poço e adoecer ? A água do poço é a causa da doença ?



Outra medida de associação: odds

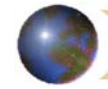
Por vezes os investigadores expressam a ocorrência de doenças em termos de *odds*

Risco = N^o de ocorrências "favoráveis" / Total de ocorrências

Expo: N^o que comeu marisco e adoeceu / N^o total que comeu marisco

Odds = N^o ocorrências "favoráveis" / N^o ocorrências "desfavoráveis"

Expo: N^o que comeu marisco e adoeceu / N^o que comeu marisco e não adoeceu



Exemplos

Nesta disciplina há 20 alunos inscritos

Se eu escolher um aluno aleatoriamente, qual a probabilidade de seres tu ?

Risco (probabilidade) = $1/20 = 0.05$

Odds = $1/19 = 0.053$

Em 100 alunos da UAIG, 60 tiveram gripe em 2010

Qual o risco e o odds de ter gripe ?

Risco (probabilidade) = $60/100 = 0.6$

Não é uma probabilidade

Odds = $60/40 = 1.5$

“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

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

$$\begin{aligned} \text{Odds para os expostos} &= a/b \\ \text{Odds para os não expostos} &= c/d \end{aligned}$$

Odds ratio (OR)

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

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

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

Um IC a 95% para o OR é dado por,

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

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

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

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

Odds não é uma probabilidade

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

Damos preferência ao risco, mas...

Nos estudos caso-controlo não se pode usar RR

Contudo,

O OR é frequentemente uma boa aproximação ao RR

Se a doença for rara, a e c são muito pequenos

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

Estudos transversais (cross-sectional)

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





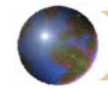
Exercício

Um estudo clássico de Doll and Peto (1976) sobre os médicos britânicos masculinos, determinou que as taxas de mortalidade por cancro do pulmão e por doença coronária eram, respectivamente, 150 por 100 mil e 1082 por 100 mil. Os autores estimaram a proporção de fumadores e de não fumadores entre os mortos, para as duas causas de morte. A tabela abaixo resume os resultados,

	Taxa de mortalidade (por 100 mil)	
	cancro pulmão	doença coronária
fumadores	140	669
não fumadores	10	413
	150	1082

de Doll and Peto. 1976. *Br Med J* 2:1525-1536

- Tome a taxa de mortalidade como uma medida de risco e responda,
- Qual é o RR dos fumadores, relativamente aos não fumadores, para o cancro do pulmão ? e para a doença coronária ?
 - Qual é o risco atribuível para cada uma das doenças ?
 - Compare o RR e o risco atribuível devido ao tabaco entre as duas doenças e suponha que um dia, por pura magia, conseguia que os médicos deixassem de fumar. Espera com isso evitar mais mortes por cancro do pulmão ou por doença coronária ? porquê ?



Confundimento

Variável de confundimento

Uma variável externa, não tida em consideração no estudo e que se distribui de forma diferente entre expostos e não-expostos. Provoca diferente ocorrência de doença entre expostos e não-expostos, (independentemente do efeito do factor de risco sob estudo).

Pode promover associação onde esta não existe, ou pode mascarar a associação onde esta existe.



Confundimento: exemplo

Factor de risco	Estado de doença		Risco
	Doente	não doente	
Exposto	81	29	0,736
não exposto	28	182	0,133
RR			5,52

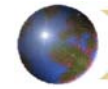
Quando a variável C é ignorada, existe forte associação doença-factor (RR=5.52)

Factor de risco	Tem C			Não tem C		
	Doente	não doente	Risco	Doente	não doente	Risco
Exposto	1	9	0,100	80	20	0,800
não exposto	20	180	0,100	8	2	0,800
RR	1,000			1,000		

Mas veja-se o que acontece quando C é tida em atenção !

Os indivíduos com C são mais comuns entre os não expostos

Os indivíduos sem C são mais comuns entre os expostos



Associação doença-variável C

Factor de risco	Tem C			Não tem C		
	Doente	não doente	Risco	Doente	não doente	Risco
Exposto	1	9	0,100	80	20	0,800
não exposto	20	180	0,100	8	2	0,800
RR	1,000			1,000		

A associação era entre C e a doença

Como obter uma estimativa global de RR na tabela completa ?

- Método de Mantel-Haenszel -

Estudos caso-control

Planeamento e análise

Avaliação da associação entre factores de risco & doença

Módulo	Metodologia de Investigação em CS
0	Apresentação
2	Epidemiologia descritiva
3	Estudos transversais
4	Estudos caso-control
5	Estudos de coortes
17	Ensaio clínico
6	Testes de diagnóstico

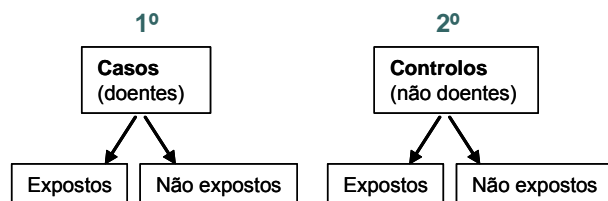
Estudos ...

- Transversais
- Caso-control
- de Coortes
- Interventivos

Estudos caso-control

1. Toma-se uma amostra de indivíduos c/ a doença (os **casos**)
2. Toma-se outra amostra de indivíduos do "mesmo tipo" mas sem a doença (**controles**)

Investiga-se casos e controles quanto à exposição ao factor de risco



Se houver associação doença-factor risco, espera-se maior proporção de expostos entre os casos que entre os controles

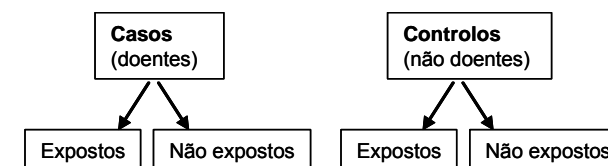
Transversais vs. Caso-control



Transversais (cross-sectional)

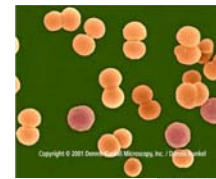
Amostra única (n) tomada de uma vez, em pouco tempo
Mede-se simultaneamente exposição e prevalência da doença

Caso-control



Escolha dos casos

- Casos seleccionados após definição prévia de caso
Existe tendência para alargar a definição a meio do processo, a fim de incluir mais casos
- Usar todos os encontrados (doenças raras) ou uma amostra de casos
- Se for amostra, deve ser representativa
- Podem ser localizados em hospitais, clínicas, ficheiros, etc.
Os casos podem provir de fontes múltiplas (e.g. vários hospitais)



Meningococo
(*Neisseria meningitidis*)

DOENÇA MENINGOCÓCICA - DEFINIÇÃO DE CASO*

Descrição Clínica

Quadro clínico compatível com doença meningocócica, por exemplo meningite e/ou meningococémia que pode degenerar rapidamente em púrpura fulminante, choque e morte. São possíveis outras manifestações.

Critérios laboratoriais para o diagnóstico

- Isolamento de *Neisseria meningitidis* de um local normalmente estéril (por exemplo, sangue ou líquido céfalo-raquidiano (LCR), ou, menos habitualmente, do líquido articular, pleural ou pericárdico).
- Detecção de ácido nucleico da *N. Meningitidis* a partir de um local normalmente estéril.
- Detecção do Antígeno da *N. meningitidis* a partir de um local normalmente estéril.
- Demonstração por microscópio de diplococos gram-negativos a partir de um local normalmente estéril.

Para um caso provável:

- Apenas um título elevado de anticorpos meningocócicos no soro convalescente.

Classificação do caso

Possível: Não aplicável.

Provável: Um quadro clínico compatível com doença meningocócica invasiva sem qualquer confirmação laboratorial, ou com identificação de *N. meningitidis* a partir de local não estéril, ou com níveis elevados de anticorpos meningocócicos em soro convalescente.

Confirmado: Um caso clinicamente compatível confirmado laboratorialmente.

É de salientar que os portadores assintomáticos não deverão ser notificados.

* Definição de caso publicada no Jornal Oficial das Comunidades Europeias a 03-04-2002.

Escolha dos controlos

- Devem provir da mesma população em risco de doença que os casos e ser representativos dessa população-alvo (**representatividade**)
- Devem ter características idênticas às dos casos, i.e., mesma distribuição de idades, sexo, origem socio-económica, etc. (**comparabilidade**)
- A sua selecção envolve muitas vezes uma fase não-aleatória
Em geral, os casos não são amostra aleatória da população e os controlos espelham isso.

Na prática, a **comparabilidade** entre controlos e casos é difícil de conseguir. A formação do grupo de controlo é em geral a parte difícil deste tipo de estudos.
A **comparabilidade** é mais importante que a representatividade

Explo – fontes de casos e controlos

Casos	Controlos
Todos os casos encontrados na população	Não-casos, numa amostra aleatória da população
Todos os casos num hospital	Amostra de pacientes no mesmo hospital que não têm a doença
Casos encontrados oportunisticamente em população difícil de definir	Esposos, vizinhos, familiares, amigos dos casos

Em qualquer das situações, os controlos são seleccionados de forma a ter:

- Mesma distribuição de idades que os casos
- Mesmo sex-ratio
- Mesma origem social
- Mesmos hábitos
- ...

Avaliação da exposição ao factor de risco de risco

Exposição ao factor de risco é avaliada retrospectivamente

- Em geral é estimada, a menos que existam registos
- Assume-se que exposição teve lugar antes e/ou durante o desenvolvimento da doença
- Frequentemente sujeita a erros de memória e/ou enviesamento do entrevistador

Situação simples: 1 só factor de risco

	doentes	Não-doentes	
Factor de risco	CASOS	CONTROLOS	Total
Exposto	a	b	a+b
não exposto	c	d	c+d
Total	a+c	b+d	n

Calcular RR como de costume ?
NÃO !

n já não é uma amostra aleatória da população

A proporção de doentes $(a+c)/n$ já NÃO representa essa mesma proporção na população

Por que razão não se pode calcular RR

Sejam A, B, C, D os valores absolutos na população
Seja f_1 e f_2 as proporções de doentes e controlos amostradas

Factor de risco	(a) POPULAÇÃO			(b) Valores esperados na amostra		
	Doentes	Não doentes	Total	Casos	Controlos	Total
Exposto	A	B	A+B	$A f_1$	$B f_2$	$A f_1 + B f_2$
não exposto	C	D	C+D	$C f_1$	$D f_2$	$C f_1 + D f_2$
Total	A+C	B+D	N	$(A+C) f_1$	$(B+D) f_2$	n

RR verdadeiro:

$$RR = \frac{A(C+D)}{C(A+B)}$$

RR no conjunto das duas amostras:

$$\frac{A f_1 (C f_1 + D f_2)}{C f_1 (A f_1 + B f_2)} \neq \frac{A(C+D)}{C(A+B)}$$

Pode-se calcular OR ?

Sejam A, B, C, D os valores absolutos na população
Seja f_1 e f_2 as proporções de doentes e controlos amostradas

Factor de risco	(a) POPULAÇÃO			(b) Valores esperados na amostra		
	Doentes	Não doentes	Total	Casos	Controlos	Total
Exposto	A	B	A+B	$A f_1$	$B f_2$	$A f_1 + B f_2$
não exposto	C	D	C+D	$C f_1$	$D f_2$	$C f_1 + D f_2$
Total	A+C	B+D	N	$(A+C) f_1$	$(B+D) f_2$	n

OR verdadeiro:

$$OR = \frac{AD}{BC}$$

OR no conjunto das duas amostras:

$$OR = \frac{A f_1 D f_2}{B f_2 C f_1} = \frac{AD}{BC} \quad \text{Pode !}$$

Múltiplos factores de risco

Expo: Surto de 37 casos de gastroenterite num self-service
 Seleccionaram-se aleatoriamente 58 comensais para controlos
 Não havia certeza sobre o dia em que se deu a infecção (15 ou 16 Ago ?)

	Gastroenterite		Sem gastroenterite	
	Comeu	Não comeu	Comeu	Não comeu
dia 15 Agosto	6	31	9	48
dia 16 Agosto	18	19	14	43
salada	12	24	5	52
sandwiches	16	21	14	44
ovos	4	33	4	54

Soma não dá 37

Soma não dá 58

Houve respostas "não me lembro"



A tabela original é decomposta para a análise de cada factor

	Gastroenterite		Sem gastroenterite	
	Comeu	Não comeu	Comeu	Não comeu
dia 15 Agosto	6	31	9	48
dia 16 Agosto	18	19	14	43
salada	12	24	5	52
sandwiches	16	21	14	44
ovos	4	33	4	54

Factor de risco	Gastroenterite	não doente	Total
comeu em 15 Ago	6	9	15
não comeu	31	48	79
Total	37	57	94

$$31 = 37 - 6$$

$$48 = 57 - 9$$

$$OR = (6/9) / (31/48) = 1,03$$

Intervalo de confiança para o OR constrói-se como habitualmente

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

Um IC a 95% para o OR é dado por,

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

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

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

Estudos caso-controlo: vantagens

- ❑ Bons para doenças raras – permitem reunir numa amostra os poucos casos encontrados
- ❑ Muito mais rápidos e baratos que os estudos de coortes (estes acompanham indivíduos ao longo dos anos). São por isso bons para doenças de longa latência
- ❑ Podem-se estudar vários factores de risco em simultâneo. Basta interrogar casos e controlos sobre mais de um factor

... e desvantagens

- ❑ Várias fontes de enviesamento:
 - Os controlos não são totalmente equivalentes aos casos
 - A 'memória' de exposição dos casos é em geral muito melhor que a dos controlos. Às vezes é "boa" demais.
- ❑ O odds ratio é apenas uma aproximação ao RR
- ❑ Ser um caso pode ter mais a ver com ser "sobrevivente" que ser um doente
(em doenças que matam, podemos não conseguir incluir no estudo os doentes mais expostos ao factor de risco porque já morreram)
- ❑ É difícil estabelecer uma sequência temporal 1º exposição, 2º doença

Alguns estudos históricos do tipo caso-controlo

1950's

- Tabaco e cancro do pulmão

1970's

- Dietil stilbestrol (DSE) e adenocarcinoma vaginal
- Estrogéneos post-menopausa e cancro do endométrio

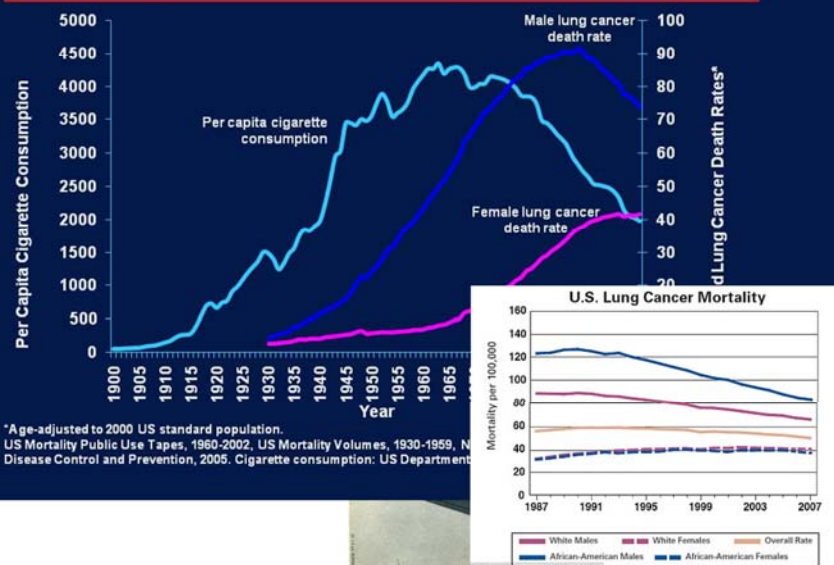
1980's

- Aspirina e síndrome de Reyes
- Certo tipo de tampões menstruais e Síndrome do Choque Tóxico (TSS)
- Práticas sexuais que provocam rompimento das mucosas e o VIH/SIDA

1990's

- Eficácia vacinal
- Dieta e cancro
- *Helicobacter pylori* e cancro do estômago

Tobacco Use in the U.S., 1900-2002



Houve também enganar históricos (1982)

Casos – 20 homossexuais com sarcoma de Kaposi
Controlos – 40 homossexuais sem sarcoma
Factor de risco – uso/não uso de “poppers” (estimulantes sexuais)

OR= 10 !!

Durante anos discutiu-se a possibilidade dos estimulantes serem mutagénicos !

Condunimento !

A toma de estimulantes estava associada a práticas sexuais de risco

Escolha de controlos (revisão)

Preferível:

Controlos escolhidos na mesma comunidade cultural e sócio-económica de onde os casos são oriundos

Frequente: emparelhamento

Familiares, amigos ou vizinhos escolhidos a dedo por terem a mesma idade, sexo, hábitos etc.

Não confundir: emparelhamento com comparabilidade

Comparabilidade: tomada de amostra aleatória com idêntica distribuição de idades, sexos, etc.

Emparelhamento: escolha individualizada, não aleatória, de pares para cada caso (em geral 2 ou 3 emparelhamentos por caso) (O emparelhamento garante sempre comparabilidade).
Com emparelhamento, a análise estatística é um pouco diferente

Quantos controlos ?

Número de casos – em geral limitados pelos que há disponíveis
Número de controlos – em geral existe possibilidade de escolha

Recordar $\hat{\epsilon}(\ln OR) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$ $LI = e^{\ln OR - 1.96\hat{\epsilon}(\ln OR)}$
 $LS = e^{\ln OR + 1.96\hat{\epsilon}(\ln OR)}$

Número de controlos demasiado pequeno
Gera estimativas de OR com um IC demasiado grande

Número de controlos demasiado grande
Significa gasto de recursos sem contrapartidas em termos de precisão

Noções fundamentais sobre... ... o método científico !

Suponhamos que se suspeita que o factor de risco X está associado à doença.

Formalmente o procedimento estatístico evolui da seguinte forma:

- (1) Coloca-se a hipótese de que X NÃO está associado à doença
É a chamada hipótese nula, ou H_0
- (2) Concebe-se um teste para rejeitar H_0
por explo: construir um IC 95% em torno do OR e comparar com 1
- (3) O teste pode rejeitar ou não a H_0
- (4) - Se rejeitar, dizemos “os dados fornecem evidência de que H_0 é errada”
Temporariamente aceitamos H_1

- Se não rejeitar dizemos “não existe evidência suficiente para rejeitar H_0 ”

Dois erros possíveis no teste

Qualquer estudo científico deve evitar dois erros:

- (1) Tipo I - Concluir que a exposição está associada à doença quando de facto não está

A probabilidade de cometer este erro chama-se “nível de significância”, α

o nosso teste controla α , ao construir o IC a 95%, estabelece $\alpha=0.05$

- (2) Tipo II - Concluir que exposição não está associada à doença, quando de facto está

A probabilidade de cometer este erro denomina-se β ; e $1-\beta$ chama-se a “potência” do estudo

Se, na realidade, a exposição tem um RR diferente de 1, a potência é a probabilidade de que o OR seja significativamente diferente de 1.
O aumento da potência consegue-se com o aumento da amostra (casos e/ou controlos)



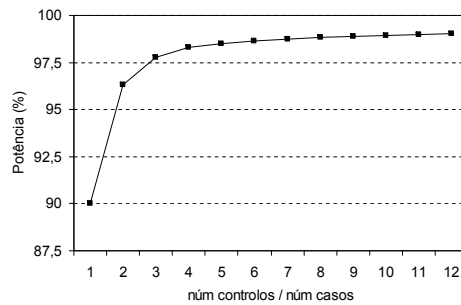
		Teste de H_0	
		Não rejeita H_0	Rejeita H_0
Mundo real	H_0 é verdadeira	$1-\alpha$	α
	H_0 é falsa	β	$1-\beta$ (potência)

Se o teste rejeita, sabemos que a probabilidade de errar é α

Se o teste não rejeita, em geral não conhecemos β



Aumento da potência com o número de controlos

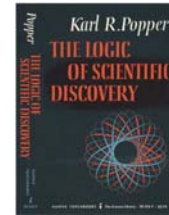


Potência
 Probabilidade de que a hipótese nula (H_0 : não existe associação factor-doença) seja rejeitada quando é falsa

Em geral **2 a 3 controlos por caso** é o ideal
 Raramente vale a pena ter mais de 4 controlos por caso



Raciocínio “Popperiano”



(c) 1934

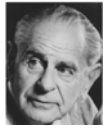
Racionalismo crítico

Para ser genuinamente científica, uma hipótese (ou teoria) tem de ser “falsificável”

i.e. tem de ser possível conceber um teste (observacional ou experimental) que ponha a hipótese à prova.

Um teste não exclui a possibilidade de uma teoria não rejeitada vir a ser rejeitada mais tarde por um teste mais apurado.

Os nossos testes não podem “demonstrar” as teorias, apenas as podem rejeitar.



Sir Karl Popper, 1980



Sir Karl Popper, 1990



Potência

O aumento do número de controlos (ou, em geral, de uma amostra) aumenta a potência de um teste feito com base na amostra.

Procedimento geral em testes

Hipótese nula (H_0)

Explo: Não existe associação entre factor de risco e doença

Erro Tipo I: H_0 é verdadeira, mas o teste rejeita-a.

probabilidade deste erro = α . Em geral adoptamos $\alpha = 0.05$

Erro Tipo II: H_0 é falsa, mas o teste não a rejeita.

probabilidade deste erro = $(1 - \text{potência do teste})$
 probabilidade de NÃO cometer este erro = **potência do teste**

Estudos de coortes

Planeamento e análise

Avaliação da associação entre factores de risco & doença

Estudos ...

- Transversais
- Caso-controlo
- de Coortes
- Interventivos

Módulo	Metodologia de Investigação em CS
0	Apresentação
2	Epidemiologia descritiva
3	Estudos transversais
4	Estudos caso-controlo
5	Estudos de coortes
17	Ensaio clínicos
6	Testes de diagnóstico

A coorte

Coorte de nascimento = Classe anual

Conjunto de todos os indivíduos nascidos no mesmo intervalo de tempo ao longo de toda a sua vida.
Em geral o intervalo de tempo considerado é 1 ano

Idades	1996	1997	1998	1999	2000	2001
0	1000					
1		800				
2			600			
3				450		
4					300	
5						150

Coortes e população



Coorte de 1996

Coorte de 1995

	1996	1997	1998	1999	2000	2001
Coorte de 1996	1000	1100	980	1020	975	1151
Coorte de 1995	750	800	735	811	820	740
	620	625	600	583	605	570
	435	455	425	450	470	440
	315	280	326	276	300	305
	128	155	162	130	155	150

Coorte (lato sensu)

Coorte

Conjunto de indivíduos **definidos segundo determinado critério**, ao longo de um período de tempo.

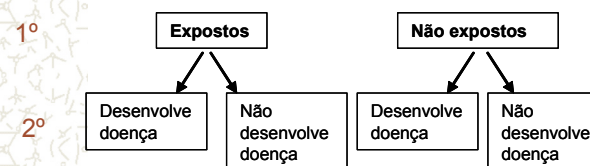
(Em geral o período para o qual o critério é válido)

Expos

- Os alunos que dão entrada na FE-UAlg e usam a cantina
- Os alunos que dão entrada na FE-UAlg e não usam a cantina
- Os alunos que entraram no MGUS em 2011 ao longo do curso
- Os homens fumadores com 35 a 60 anos de idade de uma vila algarvia
- Os trabalhadores da secção de pinturas de uma fábrica de automóveis

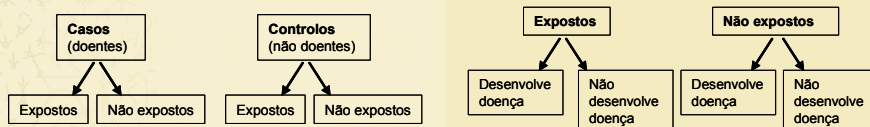
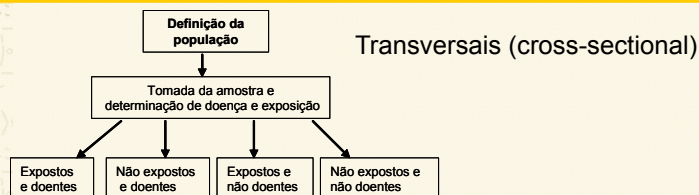
Estudos de coortes

1. Seleccionam-se 2 grupos de indivíduos não doentes. Um grupo está exposto ao factor de risco. Outro não está.
2. Acompanham-se os dois grupos ao longo do tempo e, no fim, a incidência da doença é comparada entre os dois grupos



Se houver associação factor-doença, espera-se maior proporção de doentes no grupo dos expostos que nos não-expostos

Transversais, Caso-controlo, Coortes



Caso-controlo (retrospectivos)

Coortes (longitudinais)

Exemplo

513 toxicodependentes foram testados para HIV e seguidos durante 2 anos
Registando-se se desenvolveram TB-doença

Estado inicial	Desenvolveu TB ?		total
	sim	não	
HIV +	8	207	215
HIV -	0	298	298



No início, os indivíduos devem ser testados para TB-doença e excluídos caso tenham sintomas.
(Fundamental para se poder sequer falar em causalidade)



Neste explo, a exposição (ou não) ao factor não é óbvia, tem de ser investigada

Variantes

... normalmente visam poupar tempo ou dinheiro

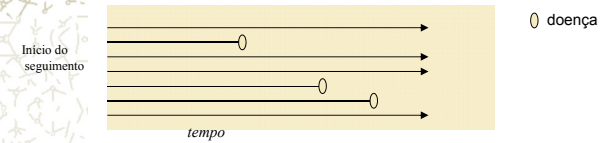
- Estudar **mais de 1 factor** de risco em simultâneo (mesmas pessoas agrupadas de forma diferente)
- Usar como **grupo de controlo a população geral** de onde provém o grupo exposto. Assim, basta seguir o grupo exposto.

Inconv: as estatísticas oficiais não fornecem dados tão detalhados sobre a população geral como se terá sobre os expostos.

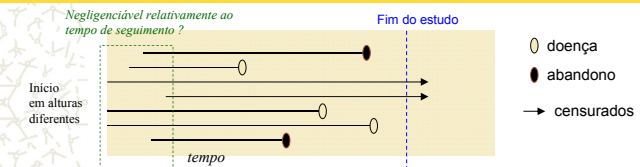
- **Formar os grupos com base no seu passado** de exposição ou não (Estudos *retrospectivos*)

Inconv: Falhas de memória. Há indivíduos expostos no passado que já morreram devido à exposição.

Coorte fixa vs coorte variável

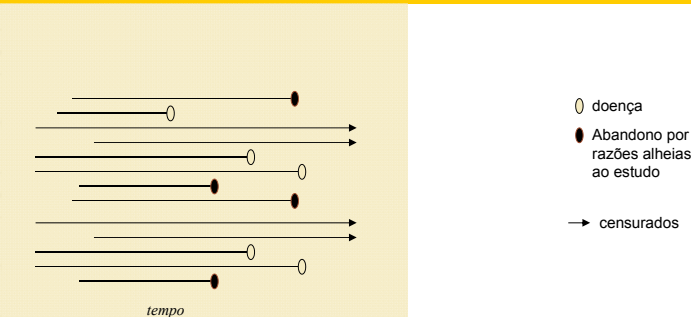


Coorte fixa – Todos começam ao mesmo tempo e são seguidos durante o mesmo tempo (ou até que ocorra doença)



Coorte variável – Nem todos começam ao mesmo tempo. Nem todos chegam ao fim, apesar de não terem adoecido ("abandonos") e alguns são "censurados"

Amostras grandes: riscos em competição



Coortes de amostras grandes Explo: ensaios clínicos de medicamentos

Os "Riscos em competição" tornam-se muito relevantes
A probabilidade de aparecimento de abandonos por razões alheias à doença aumenta.

Cálculos para coorte fixa

	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

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

RR pode ser calculado porque se define n à partida, quando ainda ninguém é doente

Intervalos de confiança calculam-se como habitualmente

Pessoas-tempo

Coorte variável - Os indivíduos entram e saem do estudo em instantes diferentes
 - Os riscos em competição provocam abandonos
Como contar os indivíduos em estudo ?

Ideia : usar o somatório de tempos de estadia no estudo (em vez do número de indivíduos)

1. Para cada indivíduo: medir tempo de estadia até adoecer ou abandonar o estudo
2. Calcular Z = soma de "pessoas-tempo" para cada grupo (expostos e não-expostos)

Exemplo

8 seropositivos para a hepatite B, foram seguidos em alturas diferentes, registrando-se se adoeciam com cirrose.

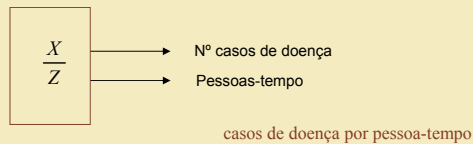
Registro do seguimento:

Indivíduo	Início	Fim	Seguimento	
			Tempo de seguimento	Doença ?
1	05-Out-56	01-Dez-93	446	sim
2	10-Out-69	31-Dez-97	339	sim
3	10-Jun-79	31-Dez-97	223	não
4	30-Ago-84	28-Set-94	121	não
5	08-Mai-62	08-Jul-91	350	sim
6	01-Nov-66	10-Mai-79	150	sim
7	21-Mar-54	30-Jun-91	447	não
8	08-Jun-61	29-Jul-65	50	sim
total:			2126	sim= 5
			meses	não=3

Z = 2126 pessoas-mês

Taxa de incidência, X/Z

Taxa de Incidência (TI):



Exposto	Adoeceu ?		Total (pessoas-tempo)
	sim	não	
sim	a		Z dos expostos
não	c		Z dos não-expostos

Totais marginais têm natureza diferente. Não são a soma das linhas da tabela.

TI dos expostos: a/Z exp
 TI dos não expostos: c/Z não-exp

Características da taxa de incidência

Não é uma proporção

X não é limitado superiormente

X/Z aumenta quando a unidade de Z alarga:

$$\frac{X}{\text{pessoas} \times 365 \text{ dias}} \quad \frac{X}{\text{pessoas} \times 12 \text{ meses}} \quad \frac{X}{\text{pessoas} \times 1 \text{ ano}}$$

Unidades do denominador

São arbitrárias: - pessoa-ano
 - pessoa-mês
 - 1000 pessoas-ano

Taxa instantânea

Pode ser interpretada como uma velocidade

60 Km/hora dá a velocidade instantânea, embora expressa para um período não-instantâneo (1 h)

Unidades da TI

Tempo⁻¹

IC para a taxa instantânea, com X grande

$$TI = \frac{X}{Z}$$

Se os casos de doença ocorrem aleatoriamente ao longo do tempo, devem ser bem descritos por um processo de Poisson. Se X for grande ($X > 20$), um intervalo de confiança a 95% é dado por,

$$\left(\frac{X}{Z}\right) \pm 1.96 \sqrt{\frac{X}{Z^2}}$$

IC para a taxa instantânea, com X pequeno

$$TI = \frac{X}{Z}$$

Se os casos de doença ocorrem aleatoriamente ao longo do tempo, devem ser bem descritos por um processo de Poisson. Se X for pequeno ($X < 20$), um intervalo de confiança devia ser construído por métodos exactos. Uma boa aproximação, contudo, é dada pela fórmula de Byar,

$$\frac{(X + 0.5) \left(1 - \frac{1}{9(X + 0.5)} \pm \frac{1.96}{3} \sqrt{\frac{1}{X + 0.5}} \right)^3}{Z}$$

(1.96 na fórmula é para um IC de 95%)

Da TI ao risco

Para converter uma TI em risco:

$$r = 1 - e^{-TI \times t}$$

Se $TI = 0$, obtém-se risco = 0
Se $TI = +\infty$, obtém-se risco = 1

Assume-se que

$$\frac{dN}{dt} = -TI N$$

(EDO de 1ª ordem)
Ver apontamentos teóricos

RR com taxas instantâneas

$$RR = \frac{r_1}{r_2} = \frac{1 - e^{-TI_1 t_1}}{1 - e^{-TI_2 t_2}}$$

Risco Relativo de doença nos expostos ao fim do tempo t

RR :
Quantas vezes número de casos ocorridos em uma unidade de tempo é superior entre expostos relativamente a não expostos".

Se o expoente ($TI t$) for pequeno, verifica-se que

$$1 - e^{-TI t} \approx TI t$$

$TI t$	$1 - e^{-TI t}$
0.01	0.01
0.05	0.05
0.10	0.10
0.15	0.14
0.20	0.18

Logo,

$$RR = \frac{r_1}{r_2} \approx \frac{TI_1}{TI_2}$$

NOTA – Se t for grande, esta simplificação não deve ser usada

IC para o RR com taxas instantâneas

$$RR = \frac{r_1}{r_2} \approx \frac{X_1/Z_1}{X_2/Z_2}$$

X_1/Z_1 é a TI nos expostos

X_2/Z_2 é a TI nos não expostos

Intervalo de confiança aproxim a 95%

$$RR_{inf,sup} = e^{\ln RR \pm 1.96 \sqrt{1/X_1 + 1/X_2}}$$

Retomando este exemplo...

8 seropositivos para a hepatite B, foram seguidos em alturas diferentes, registrando-se se adoeciam com cirrose.

Seguimento					
Registro do seguimento:			Tempo de seguimento	Doença ?	
Indivíduo	Início	Fim			
1	05-Out-56	01-Dez-93	446	sim	
2	10-Out-69	31-Dez-97	339	sim	
3	10-Jun-79	31-Dez-97	223	não	
4	30-Ago-84	28-Set-94	121	não	
5	08-Mai-62	08-Jul-91	350	sim	
6	01-Nov-66	10-Mai-79	150	sim	
7	21-Mar-54	30-Jun-91	447	não	
8	08-Jun-61	29-Jul-65	50	sim	
			total:	2126	sim= 5
			meses		não=3

Z = 2126 pessoas-mês

Exemplo (conti.)

Outro grupo de 14 indivíduos seronegativos, foi também seguido durante um total de 4725 pessoas-mês, tendo 3 adoecido e 11 não adoecido.

Construindo a tabela de contingência,

	Cirrose	Não cirrose	total	IC a 95% para as TI		
				TI	L inf	L sup
HBsAg +	5		2126	0,00235	0,00029	0,00441
HBs Ag -	3		4725	0,00063	-0,00008	0,00135
	8		6851	0,00117	0,00036	0,00198

$$RR = 0.00235/0.00063 = 3.7$$

X (casos doença)

Z (pessoas tempo)

$$\left(\frac{X}{Z}\right) \pm 1.96 \hat{e} \left(\frac{X}{Z}\right)$$

Exemplo (conti.)

	Cirrose	não cirrose	total
HBsAg +	5		2126
HBs Ag -	3		4725
	8		6851

Os IC's para o RR calculam-se por: $RR_{inf,sup} = e^{\ln RR \pm 1.96 \sqrt{1/X_1 + 1/X_2}}$

RR	Ln RR	RaizQ(1/X ₁ + 1/X ₂)	LI Ln(RR)	LS (RR)	LI RR	LS RR
3.70	1.309	0.730	-0.122	2.741	0.88521	15.500

Estudos de coortes: vantagens

- ❑ Acompanham a sequência de acontecimentos – permitem o estabelecimento de causalidade (nos estudos transversais e caso-controlo, já há doentes à partida)
- ❑ Possível estudar várias doenças ao mesmo tempo. Basta registar a sua incidência ao longo do tempo.

... e desvantagens

- ❑ Estudos longos e caros. Especialmente sério em doenças de longa latência (TB, zona, VIH, etc.)
- ❑ Não apropriados para doenças raras. Estas requerem o seguimento de um número demasiado grande de indivíduos.
- ❑ Pode haver mudanças de comportamento dos indivíduos ao longo do estudo que alteram a sua relação c/ o factor de risco (mudança de dieta, de hábitos de higiene, de práticas sexuais, etc.)

Análise de Sobrevivência (“Survival Analysis”)

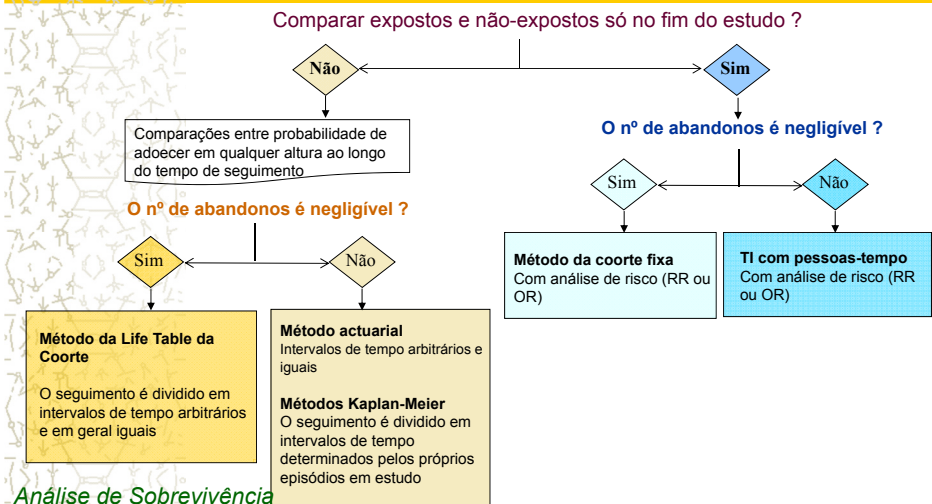
Calcula-se RR dentro de intervalos de tempo definidos ao longo do período total de seguimento

Estes intervalos podem ser todos iguais – análise da “life table” da coorte

ou

Os intervalos variam e são definidos pelos próprios acontecimentos de doença – métodos Kaplan-Meier

Métodos em estudos de coortes



Ensaio Clínico

Introdução ao planeamento e análise

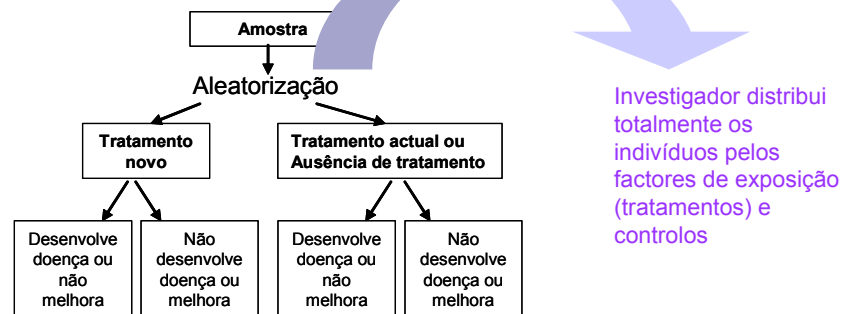


Estudos Interventivos

- Experiência feita com um grupo de pacientes com características totalmente controladas pelo investigador.
- Estudos Comunitários
 - tratamentos aplicados a grupos
- Ensaio Clínico (*Clinical trials*)
 - tratamentos aplicados individualmente
 - Conotado com testes de drogas pela indústria farmacêutica

Atribuição de tratamentos aos indivíduos é totalmente planeada, em geral seguindo critérios de aleatoriedade

Ensaio Clínico



1. Toma-se uma amostra de indivíduos e distribui-se pelos "tratamentos" e controlos
2. Investiga-se expostos e controlos quanto à doença

Exemplos

- Teste da eficácia da droga A , comparativamente a B, C, etc
- Teste da eficácia da vacina X, comparativamente a não vacinados
- Teste de novo método para diagnosticar clinicamente uma doença, comparativamente ao método tradicional
- Teste de tratamento por radioterapia, comparativamente a cirurgia
- Etc.



Exemplo 2 (Planeamento)

Toma de suplementos vitamínicos melhora o raciocínio em crianças de idade escolar ?

Grupo de Placebo			Grupo tratado		
Individuo	Resultado dos testes		Individuo	Resultado dos testes	
	Antes	Depois		Antes	Depois
1	89	83	1	70	87
2	82	97	2	91	91
3	107	107	3	106	104
4	95	101	4	92	87
5	110	100	5	103	114
...
etc..	etc..	etc..	etc..	etc..	etc..

Crianças distribuídas aleatoriamente entre os dois grupos

Número idêntico de crianças em cada grupo

Variáveis de confundimento (idade, origem sócio-económica) bem distribuídas

Existem **controles**: placebos

Crombie et al. 1990. *Lancet* 335:744-7

Exemplo 2 (Análise)

Grupo de Placebo			Grupo tratado		
Individuo	Resultado dos testes		Individuo	Resultado dos testes	
	Antes	Depois		Antes	Depois
1	89	83	1	70	87
2	82	97	2	91	91
3	107	107	3	106	104
4	95	101	4	92	87
5	110	100	5	103	114
...
etc..	etc..	etc..	etc..	etc..	etc..
\bar{X}_m média das diferenças			\bar{X}_m média das diferenças		

Comparação de duas médias (*teste t-student*)

H_0 : as duas médias são iguais

H_1 : as duas médias são diferentes (2 "caudas") ou

H_1 : a média dos tratados é maior (1 "cauda")

Ensaios clínicos: vantagens

- ❑ Apropriados para investigar causalidade. A "causa" antecede o "efeito"
- ❑ Controlo de variáveis de confundimento: podemos decidir quais as características dos indivíduos que recebem cada tipo de tratamento.
- ❑ Pode-se garantir máxima potência (controle de n)
- ❑ É possível estudar o efeito de combinações de tratamentos

... e desvantagens

- ❑ Estudos demorados (custos, abandonos...) Em geral menos que os de cortes, mais que os caso-controlo.
- ❑ Envolvem muitas vezes problemas éticos. Podem torna-los impraticáveis em Epidemiologia de DTs.

Considerações éticas

Requisitos éticos dos ensaios clínicos

Declaração de Helsínquia, 1964 e actualizações
(*World Medical Association*)

Regime jurídico de ensaios clínicos com medicamentos de uso humano
Lei 46/2004 de 19 Agosto

- necessidade de consentimento informado
- necessidade de Protocolo → a enviar para comité ético sectorial e/ou nacional
- necessidade de ponderar benefícios e riscos
- prioridade na protecção da saúde dos voluntários

Fases dos ensaios clínicos



Produtos farmacêuticos para utilização em humanos estão sujeitos a fases de ensaio.

Fase I Testa-se segurança básica (toxicidade, efeitos secundários...) e não eficácia. Em geral com algumas dezenas de voluntários saudáveis.

Se passar...

Fase II Determinação da melhor formulação, coadjuvantes, etc. Estudos dose-resposta. Em geral com 1 a 2 centenas de doentes.

Se passar...

Fase III Ensaio clínico propriamente dito. Destina-se a avaliar a eficácia. Envolve número muito elevado de indivíduos. Produto fica pronto para licenciamento.

Se passar...

Fase IV Produto já no mercado. Estuda-se Efeito a longo prazo do produto (eficácia e efeitos secundários)

O Protocolo

- Requisito de todos os ensaios clínicos
- Justifica o ensaio
- Documento a apresentar a comités éticos, agências de financiamento...
- Guia de referência para todos os participantes

Justificação	Estado da arte do tratamento. O que traz o tratamento de novo ?
Objectivos	O que é que o ensaio pretende esclarecer? (eficácia? dosagem? efeitos secundários ?)
Plano de estudo	Descrição do planeamento experimental (aleatorização, paralelo/cruzado, cego)
Seleção indivíduos	Tamanho das amostras e critérios de inclusão e exclusão de indivíduos.
Drogas e dosagens	Lista dos produtos (e/ou tratamentos) a serem tomados pelos indivíduos. Dosagens, administração, etc.
Medições	Que variáveis-resposta serão medidas durante o ensaio?
Procedimentos	A parte longa do protocolo. Tarefas a fazer e quem as faz, especificadas por fase do EC.
Análise	Métodos estatísticos que se planeia usar para análise dos dados
Confidencialidade	Declaração sobre confidencialidade

Boas práticas de planeamento

- Controlos
- Aleatorização
- Cegueira (= ocultação)
- Tipo de estudo (paralelo/cruzado)

Planeamento: controlos

Estudos devem ser comparativos,

entre tratamento e ausência de tratamento (usar placebo)

entre novo tratamento e tratamento tradicional

entre novos tratamentos alternativos



Planeamento: cegos



Ensaio diz-se,

Cego / com ocultação (single-blind)

Se cada indivíduo a tratar não sabe que tratamento recebeu

Duplamente cego / dupla ocultação (double-blind)

Se o médico, investigador ou quem está a medir os resultados do tratamento também não sabe quem recebeu tratamento e/ou que tratamento.

Triplamente cego / tripla ocultação (triple-blind)

Quem analisa os resultados, por exemplo um epidemiologista, também é ignorante sobre o tratamento tomado por cada indivíduo.

Planeamento: aleatorização

A distribuição dos indivíduos pelos tratamentos deve ser ALEATÓRIA

i.e., impossível saber ao certo para que tratamento vai o próximo indivíduo

Aleatorização global

- aleatorização completa, aleatorização enviezada, alternância
utilizar tabelas de números aleatórios ou geradores computacionais

Aleatorização estratificada

Aleatorização global

Completa

Atribuem-se os tratamentos através de um mecanismo de geração de números aleatórios (moeda ao ar, gerador de máquina de calcular...)

Possíveis inconvenientes (amostras pequenas):

Número desigual por tratamento, distribuição desigual das variáveis de confundimento

Alternação

Ordena-se aleatoriamente os indivíduos e atribui-se os tratamentos alternadamente: A, B, C, A, B, C...

Possíveis inconvenientes (amostras pequenas):

Distribuição desigual das variáveis de confundimento

Aleatorização estratificada

Há variáveis de confundimento ? (VCs)
Sim !

Estas variáveis devem estar equitativamente distribuídas pelos tratamentos para evitar o confundimento

- 1º Subdividir a amostra por *estratos* formados com base nas VC's
Estrato = um grupo onde as VCs são homogêneas
Expo: homens com >65 anos que vivem em meio rural
- 2º Aleatorizar os tratamentos *dentro* de cada estrato

Exemplo

Amostra de n= 1000
VCs : sexo, idade (criança/adulto)

	Crianças	Adultos	total
homens	360	240	600
mulheres	300	100	400
	660	340	1000

Aleatorização
intra-estrato

Tratamento A				Tratamento B			
	Crianças	Adultos		Crianças	Adultos		
homens	180	120	300	180	120	300	
mulheres	150	50	200	150	50	200	
	330	170	500	330	170	500	

Estudos paralelos

Subdividem-se os indivíduos aleatoriamente por tratamentos.
Um indivíduo só recebe o tratamento atribuído ao seu grupo

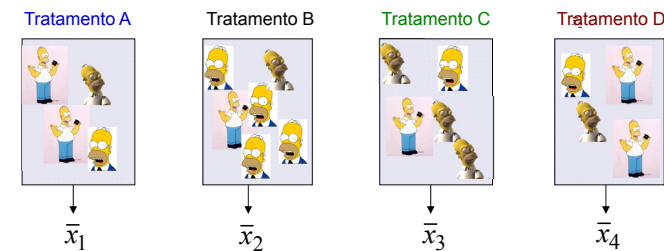


Número indivíduos por grupo pode ser diferente, ...
... embora seja desejável que seja pouco diferente

Inconveniente: variabilidade inter-indivíduos na reacção ao mesmo tratamento
provoca variabilidade intra-grupo

Estudos paralelos, análise...

- Análise de variância (ANOVA)



H_0 : Os efeitos médios dos tratamentos são todos iguais

H_1 : Há pelo menos um tratamento que origina um efeito médio diferente

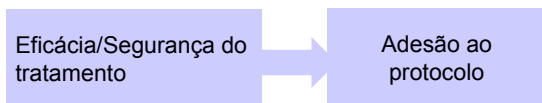
Não-adesão ao protocolo: o problema

Abandonos as falhas de adesão em ECs *raramente são aleatórios*



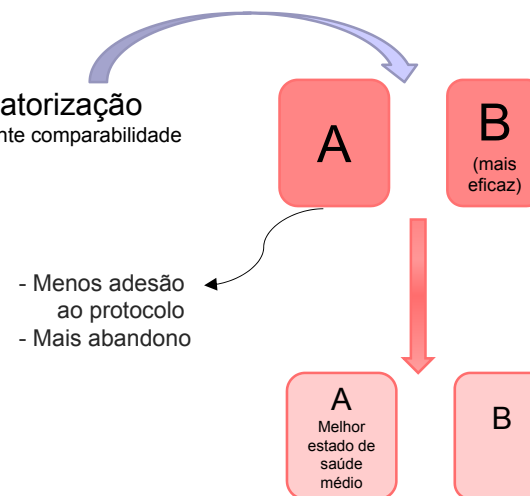
- Não aderem porque:
- tratamento não faz efeito
 - já se sentem curados
 - maus efeitos secundários
 - ...

Há associação entre:



A não-adesão modifica a aleatorização inicial

Aleatorização garante comparabilidade



Duas análises em ECs: APP e ITT

Analysis Per-Protocol, APP

Analisados os resultados apenas dos indivíduos que aderiram ao protocolo.

EXPERIÊNCIA LABORATORIAL CONTROLADA – MEDE EFICÁCIA

Intention-To-Treat, ITT

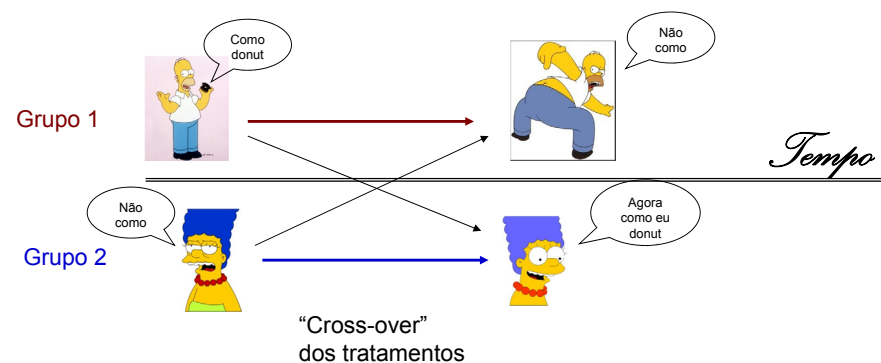
FDA, EMEA

Analisados os resultados de todos os indivíduos com base na *intenção inicial* de tratamento

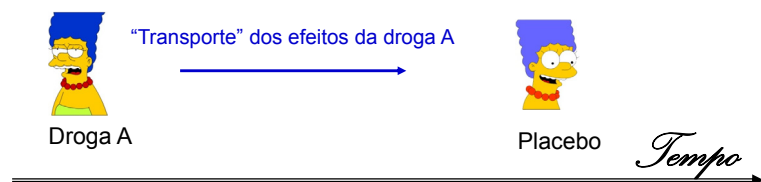
EXPERIÊNCIA 'PRAGMÁTICA' – MEDE EFECTIVIDADE NO MUNDO REAL

Estudos cruzados (*cross-over studies*)

Subdividem-se os indivíduos aleatoriamente por grupos. Cada indivíduo recebe todos os tratamentos espaçados no tempo. A ordem dos tratamentos difere entre grupos



Efeito de “transporte”, “lavagem”, cegueira



Período de “lavagem” necessário para os efeitos da droga A desaparecerem

O estudo pode ser cego.

Paciente pode receber (A+Placebo), (B+Placebo), (Placebo+Placebo)...

Estudos cruzados: vantagens

- ❑ Evitam a variabilidade *inter-indivíduos* (benéfico, assumindo que é maior que a variabilidade *intra-indivíduos*)
 - ❑ Permite uma amostra maior. Cada indivíduo permite uma comparação entre tratamentos. São feitas comparações *intra-indivíduo*.
- (Nos ensaios paralelos cada comparação requer tantos indivíduos quantos os tratamentos)

Estudos cruzados: desvantagens

- ❑ Só utilizáveis para doenças de longa duração para as quais tratamento provoca efeitos temporários (e nunca a cura)

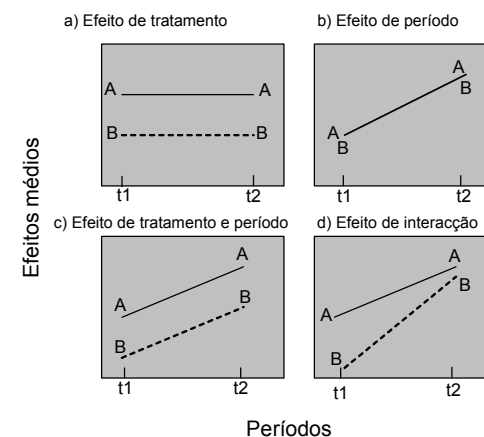
Ex: asma, herpes labial, bronquite, enxaquecas, micose recorrente.

- ❑ Indivíduos permanecem no estudo muito tempo, originando muitos abandonos, custos altos, evolução da própria doença.

- ❑ O efeito de “transporte” pode não desaparecer

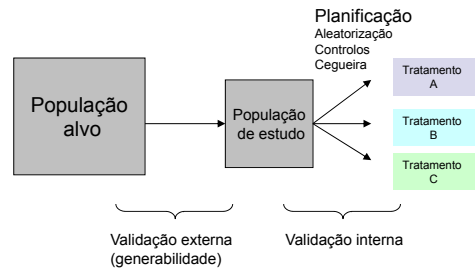
Ensaio cruzado: análise gráfica

- a) Tratamento A provoca efeito mais alto que B, independentemente da ordem
- b) Efeito médio de A e B é o mesmo, mas o efeito de ambos é mais alto qdo dado em 2º lugar
- c) Simultaneamente a) e b)
- d) Simultaneamente a) e b), mas o efeito do tempo em B é mais forte



As linhas unem o *resultado médio do mesmo tratamento* nos 2 tempos

Validade e generalização



Validação interna

O estudo foi bem planeado ?

Validação externa

A amostra é representativa ?