



EVIEDVET - Promoção do Ensino e Formação  
em Medicina Veterinária Baseada na Evidência



# Aprender a identificar a relevância de um artigo científico

Nuno Henrique Franco, PhD



# Os artigos científicos e a evidência



- Em ciência, avaliamos a veracidade de uma reivindicação com base na evidência que a sustenta
- Artigos científicos publicados em revistas com revisão por pares frequentemente vistos como evidência suficiente de apoio a uma hipótese.

## Hipótese

“Vacinas causam autismo”



## Evidência?

**RETRACTADO**  
Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and  
pervasive developmental disorder in childhood

A J Wakefield, S M Hersh, A Anthony, J Smith, D M Casson, M Malik, M Berelowitz, A P Dhillon, M A Thomson,  
P Harvey, M Henderson, S E Davies, J A Walker-Smith

Wakefield et al (1998) *The Lancet* 351, pg. 637-641

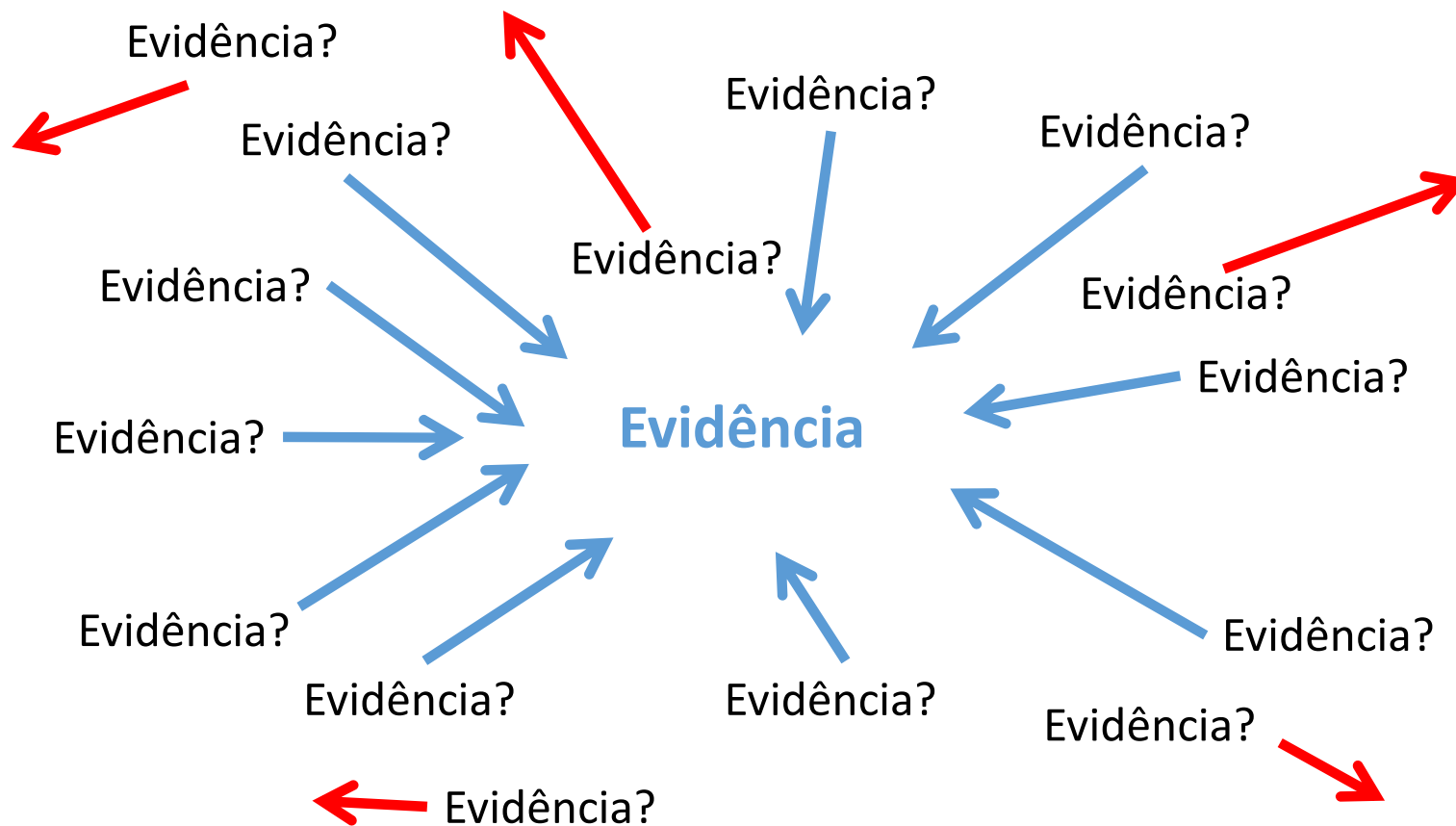
- Equipa composta por 13 clínicos reputados
- Revista medico-científica muito reputada
- Dados apresentados para sustentar esta hipótese

# Os artigos científicos e a evidência

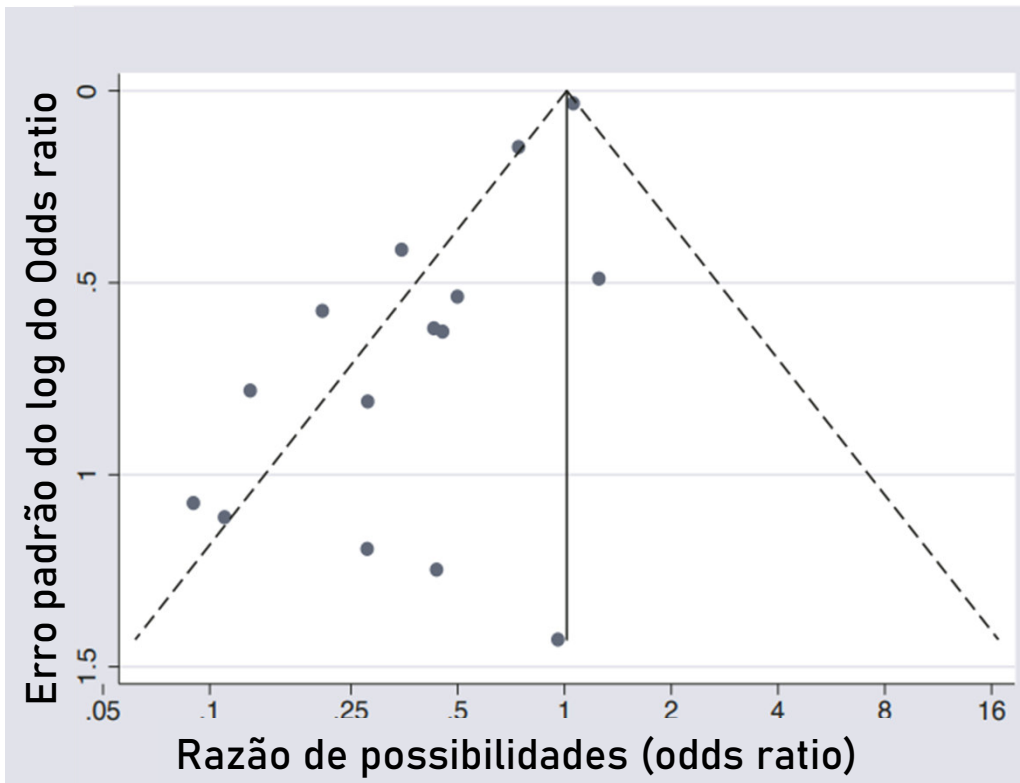


- Razões apresentadas pela revista para retractar o artigo:
  - Não houve aprovação ética do estudo (envolvendo métodos invasivos), contrariamente ao reportado
  - Os pacientes não haviam sido individualmente referenciados a Wakefield, contrariamente ao reportado
- Razões pela qual **nunca deveria ter sido publicado**
  - Amostra de 12 crianças – menos que o número de autores (!)
  - Amostra não-aleatória – crianças cujos pais associavam vacina MMR a autismo
  - Conflito de interesses – Wakefield recebeu 55.000£ de soc. de advogados dos pais

# Os artigos científicos e a evidência



# Há evidência em estudos pequenos?



## Funnel plot

- Fornece informação para avaliar risco de viés
- Relaciona efeito reportado com tamanho do estudo

## Exemplo

Efeito de terapia com magnésio no seguimento de enfarte do miocárdio na redução da mortalidade.

- Primeiros estudos pareciam confirmar hipótese
- Estudos maiores, mais precisos, contrariam hipótese
- “Vazio” à direita do funil sugere viés de publicação, ou efeito de outros vieses

# Estudo publicado = Evidência?

Confiamos que estudos publicados constituam evidência devido por confiarmos:



Método científico



Revisão por pares



Mecanismos de autocorreção



“O que é que poderia correr mal?”



# Estudo publicado = Evidência?

## Método científico

- Erros delineamento experimental
- Potência estatística insuficiente
- Erros na análise estatística
- Exacerbação das conclusões
- Divulgação selectiva
- *P-hacking*
- Fraude

## Revisão por pares

- Elevada subjectividade
- Falta de formação estatística
- Tempo reduzido para revisão
- Favorecimento de resultados “surpreendentes”
- Viés contra “resultados negativos”
- Ausência de instruções para revisores
- Artigos sem detalhes suficientes
- Dados não facultados
- Revisão fraudulenta

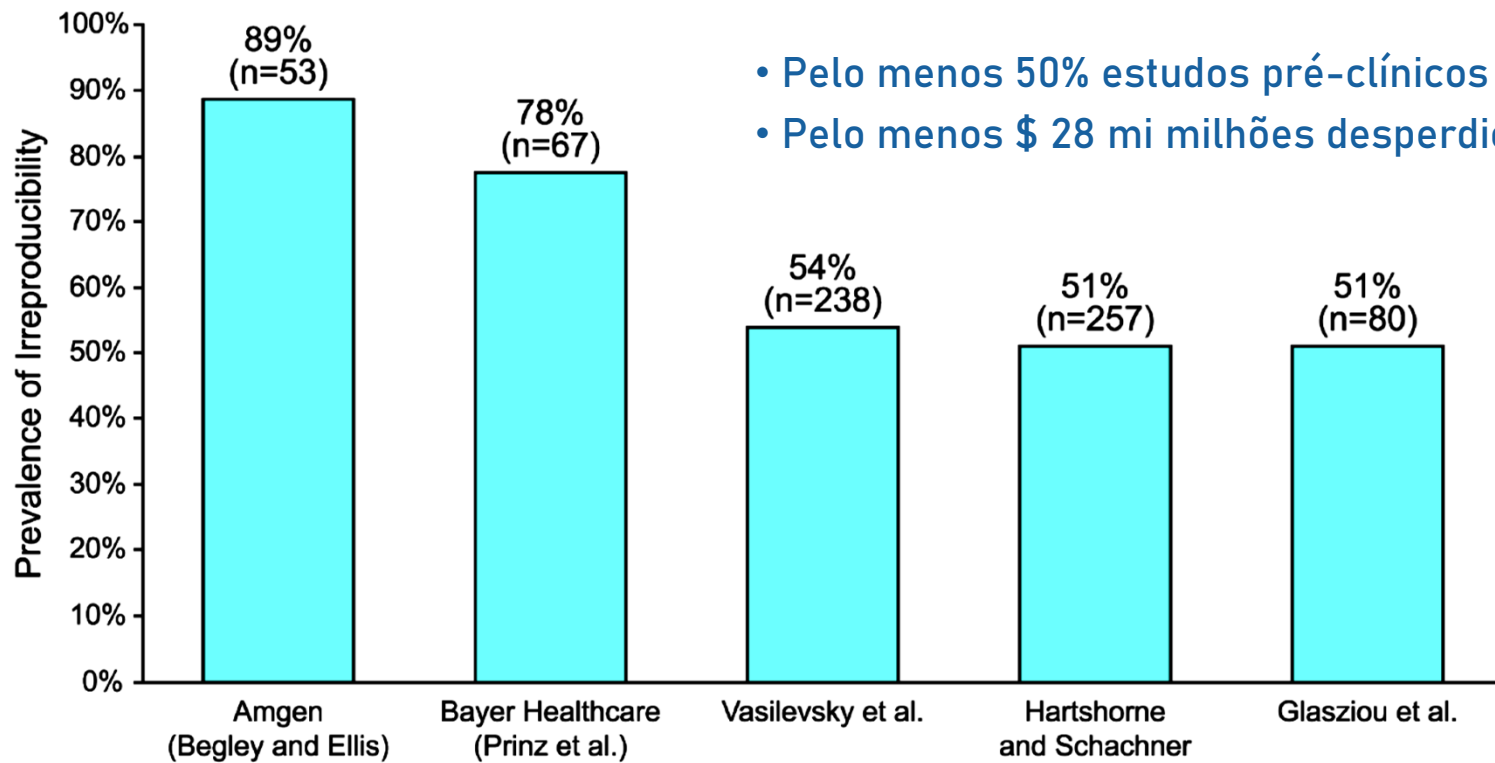
## Mecanismos de autocorreção

- **Sim, mas demora demasiado!**
- Poucos estudos confirmatórios
- Insucesso em reproduzir resultados não reportado
- Falta de incentivos para publicar “resultados negativos” ou estudos confirmatórios
- Ciência baseada em informação disponível, erros acumulam.





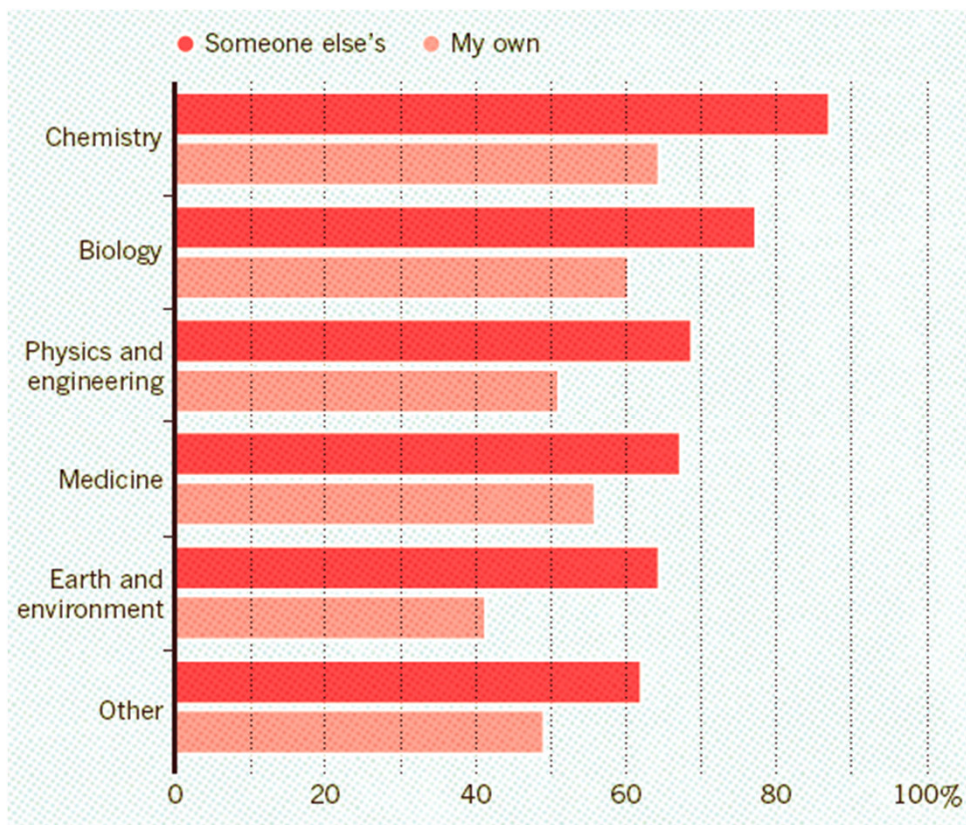
# Quão reprodutíveis são os estudos pré-clínicos?



- Pelo menos 50% estudos pré-clínicos irreprodutíveis
- Pelo menos \$ 28 mi milhões desperdiçados nos EUA por ano

Freedman et al. (2015) *PLoS Biol* 13: e1002165

# *Já falharam ao tentar reproduzir resultados de estudos científicos?*



- Sondagem da *Nature* a 1576 cientistas
- 90% consideram haver uma crise de reprodutibilidade!
- Não é uma “crise”. É um problema crónico.

Baker (2016) *Nature* 533: 452-454

# Quais as causas da irreprodutibilidade?



## WORLD VIEW A personal take on events



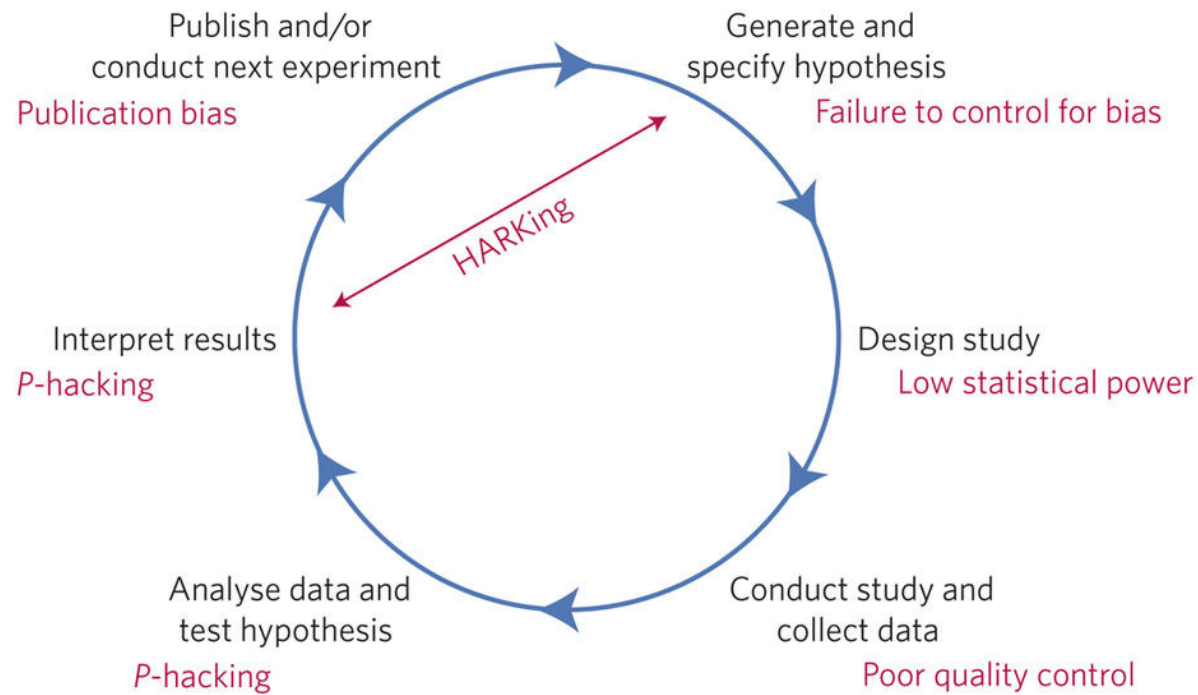
### Rein in the four horsemen of irreproducibility

Dorothy Bishop describes how threats to reproducibility, recognized but unaddressed for decades, might finally be brought under control.

- Viés de publicação
- Potência estatística insuficiente
- P-hacking
- HARKing – *Hypothesis after the Results Are Known*
- Fraca qualidade metodológica
- Ausência de cegamento e ocultação de alocação
- Tratamentos não randomizados
- Problemas de bem-estar animal
- Unidade experimental mal definida
- Outros vieses
- Fraude e má-conduta



# Possible causes of poor reproducibility



Munafò et al. (2017) *Nature Hum Behav* 1: 0021

# Viés de publicação

Potência estatística 80%  
 $\alpha=0.05$

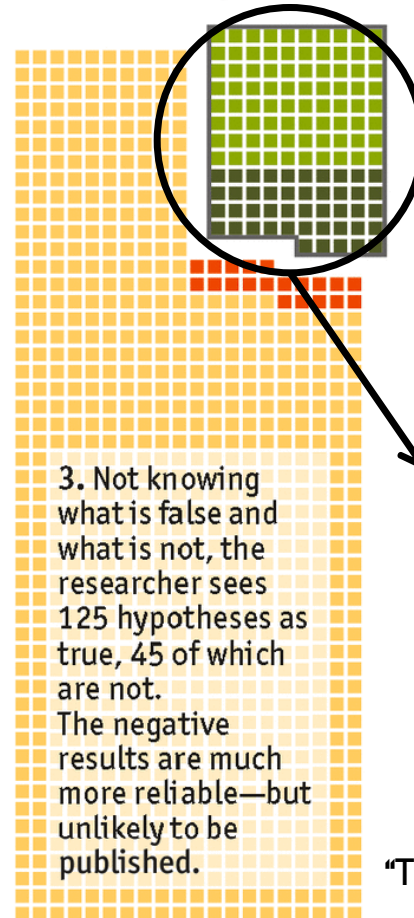
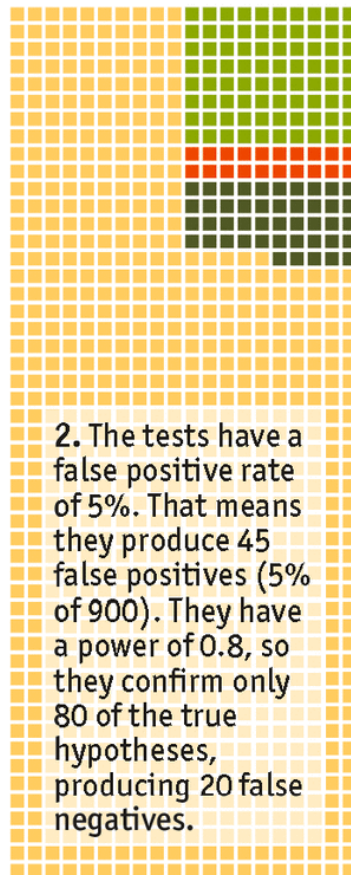
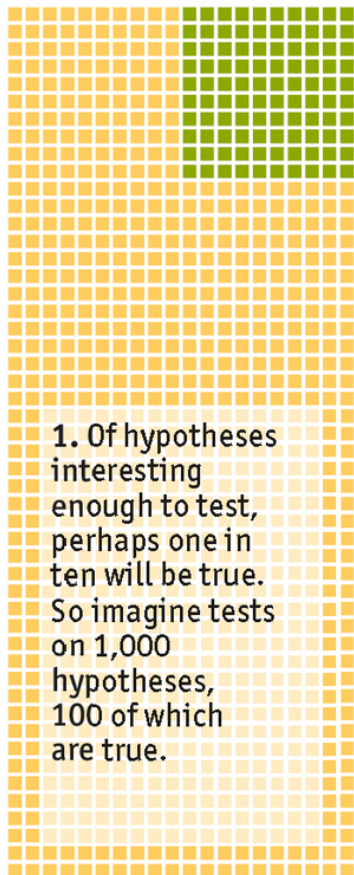


False

True

False negatives

False positives



36% falsos-positivos apenas por razões aleatórias (assumindo proporção 10% de efeitos verdadeiros)

Distorção da evidência por viés de publicação

“Trouble at the lab” *The Economist*

“E aqui pomos os resultados negativos...”



# Potência estatística insuficiente

- Potência estatística - probabilidade (condicional) de um teste detectar um efeito, se o mesmo for verdadeiro. **Estudos com baixa potência têm maior probabilidade maior de "falsos negativos"**.
- Potência estatística baixa quando o tamanho da amostra é insuficiente para detectar um efeito de determinado dimensão. Quanto menor o efeito (verdadeiro) ou maior a variabilidade na amostra, mais animais são necessários para detectar o efeito.
- Falha em detectar um efeito verdadeiro
  - Desperdiça animais, tempo e dinheiro
  - Reduz o valor preditivo dos estudos com animais
  - Pode levar a vias de investigação fechadas de forma inadequada.



# Potência estatística insuficiente

Se um estudo não tem potência estatística suficiente, e o cientista “encontra” um efeito, é provável que esteja inflacionado – “A maldição do vencedor”

Se numa área de investigação a potência estatística é reduzida, proporção global de falsos positivos aumenta, devido ao viés de publicação.

## Exemplo:

- Teste a 1000 drogas (“falso”= não há efeito), Potência 80%,  $\alpha = 0.05$
- 10% probabilidade de um efeito verdadeiro



→ “Positivos” reais reportados : 50

→ Falsos-positivos reportados: 45

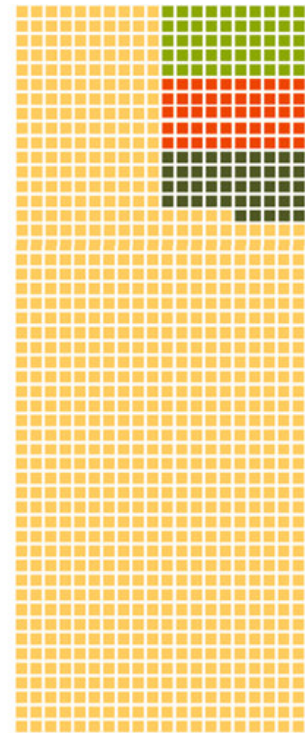
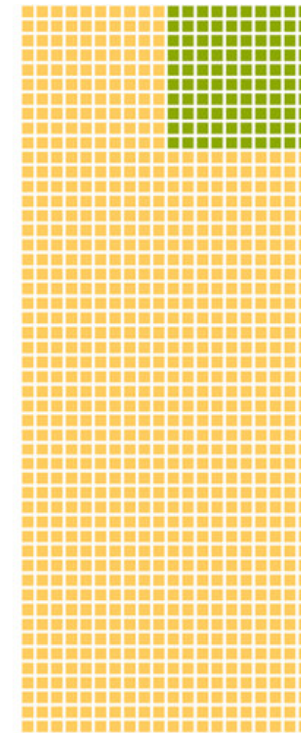
- 47% dos “resultados positivos” na literatura serão falsos

False

True

False negatives

False positives





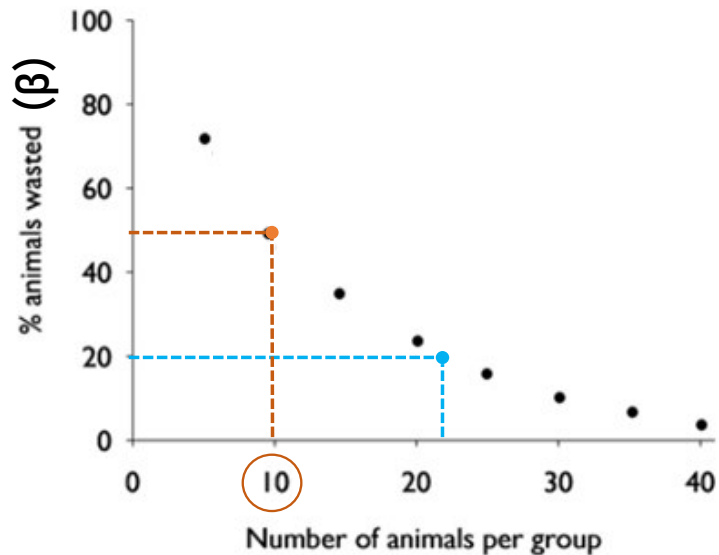
# Potência estatística insuficiente

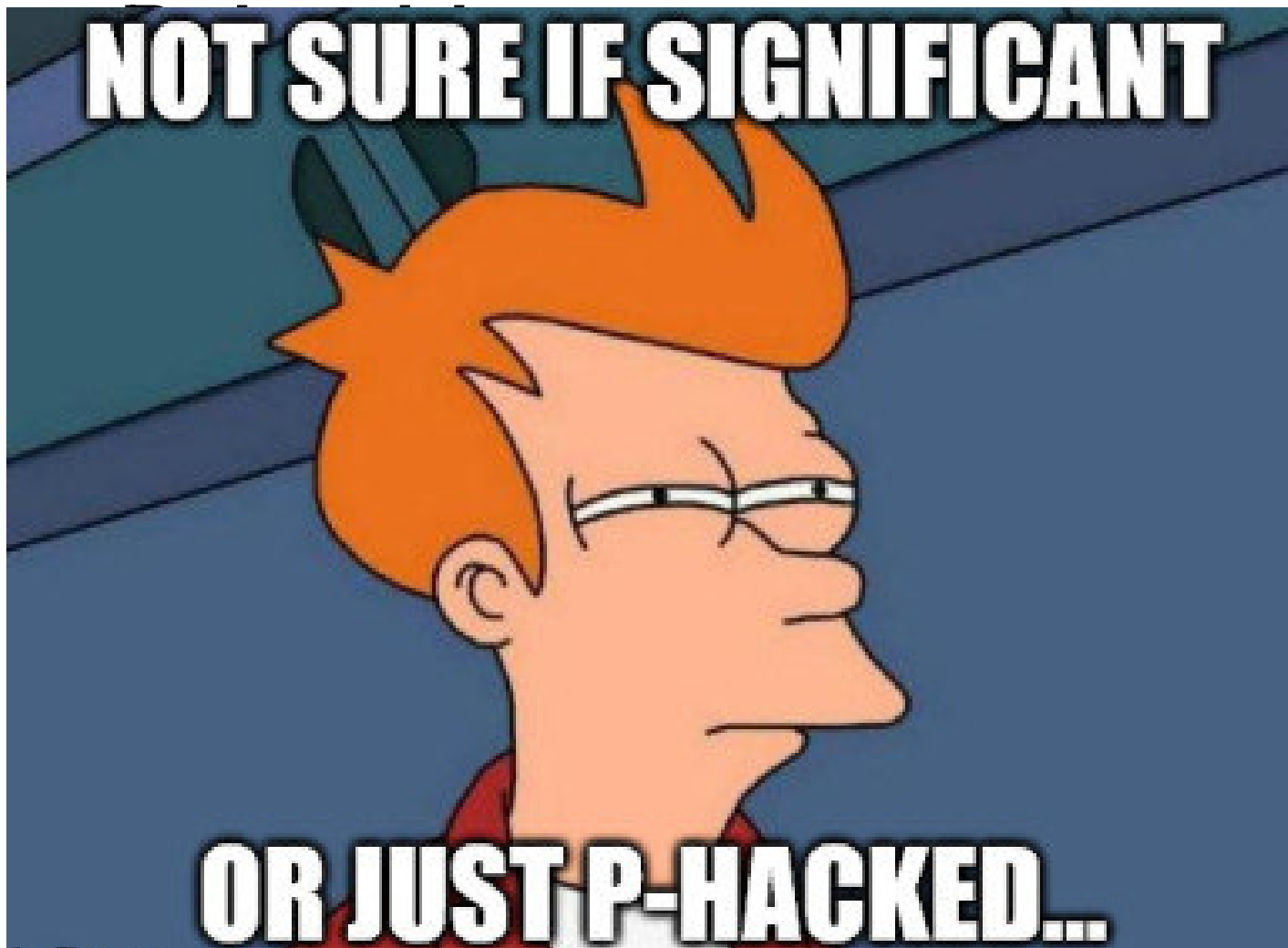
Potência estatística mediana nas neurociências: **entre 8%-31%**  
(K. Button et al 2013 *Nat Rev Neur*)

## Exemplo

Comparação entre 2 grupos. Detectar 30% aumento sobrevivência. Desvio-padrão 40%.  
Potência 80%.  $\alpha=0.05$

Tamanho de grupo recomendado: **23 animais**





- Repetir estudo para aumentar amostra
- Analisar apenas parte dos dados
- Incluir mais variáveis
- Ajustar dados (ex. dividir por peso)
- Transformação dos dados (i.e. logarítmica)
- Remoção outliers
- Definir outro grupo como controlo
- Comparar outra variável
- Mudar o teste estatístico



“

If you torture the data long enough, it will confess to anything.

”

RONALD COASE

# HARKing- *Hypothesis after results are known*



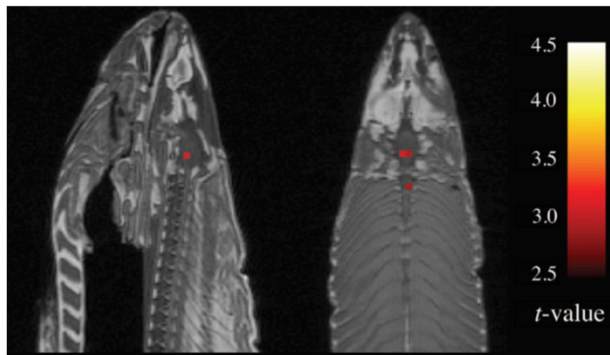
HARKing é como mudar os postes a meio do jogo. É batota.



- Só porque os resultados não confirmam a hipótese não quer dizer que não sejam relevantes
- Podem ser apresentadas teorias alternativas, para serem testadas posteriormente
- É inaceitável reportar teorias feitas *a posteriori* como sendo a hipótese original
- Também difícil de detectar numa publicação

# Demasiadas comparações

- Animais colocados em frente a duas imagens diferentes
- Ressonância magnética funcional (fMRI) a 1000 regiões no cérebro
- $\alpha$  definido para 0.001, duas regiões “significantes” detectadas
- Efeito real ou coincidência?



O “animal” era um cadáver de salmão

Levanta também a importância de considerar a plausibilidade da hipótese.

Bennett et al. (2010) Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction



# Demasiadas comparações

Um controle, muitos grupos de tratamento:

- Grupo saudável
- Grupo com patologia
- Patologia + tratamento A, patologia + tratamento B, etc.

Muitas vezes combinado com múltiplos outcomes e/ou múltiplos time-points.

Quais comparações entre grupos são relevantes depende da questão científica:

- Comparar todos os grupos com patologia vs. saudáveis?
- Comparar grupos tratados versus não-tratados?
- Comparar diferentes tratamentos?

Estas escolhas comparação **implicam diferentes questões científicas**. Comparações múltiplas não são intrinsecamente erradas, mas é sempre necessário corrigir (por exemplo, correcção de Bonferroni, Sidak)



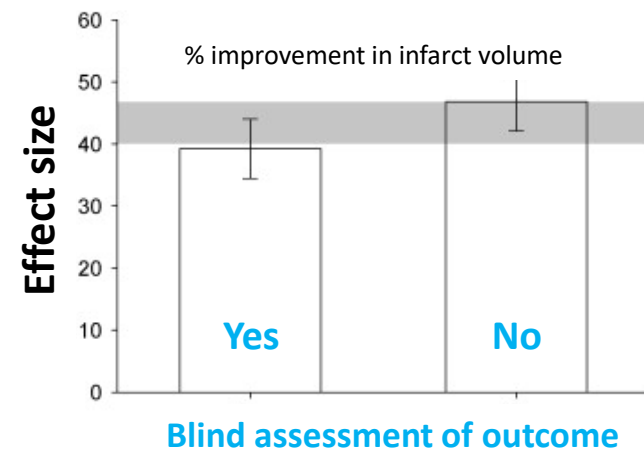
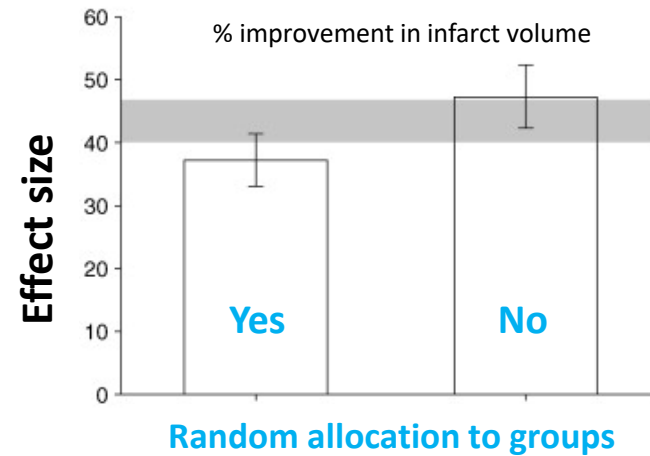
# Viés selecção/performance/detecção

Ausência de randomização/ocultação de alocação /cegamento

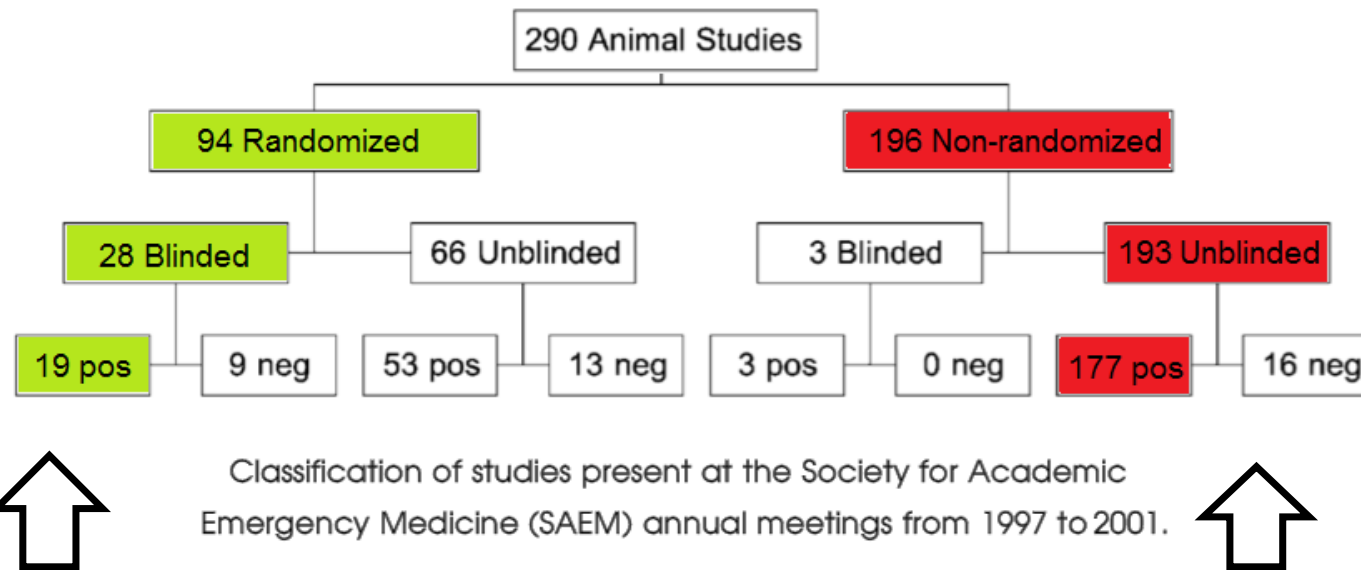


Couzin-Frankel (2013) *Science* 342

van der Worp, et al 2007 *Brain*



# Viés selecção/performance/detecção



**68%** dos estudos randomizados e com cegamento reportavam “resultados positivos”

**90%** dos estudos não-randomizados e cegamento reportavam “resultados positivos”



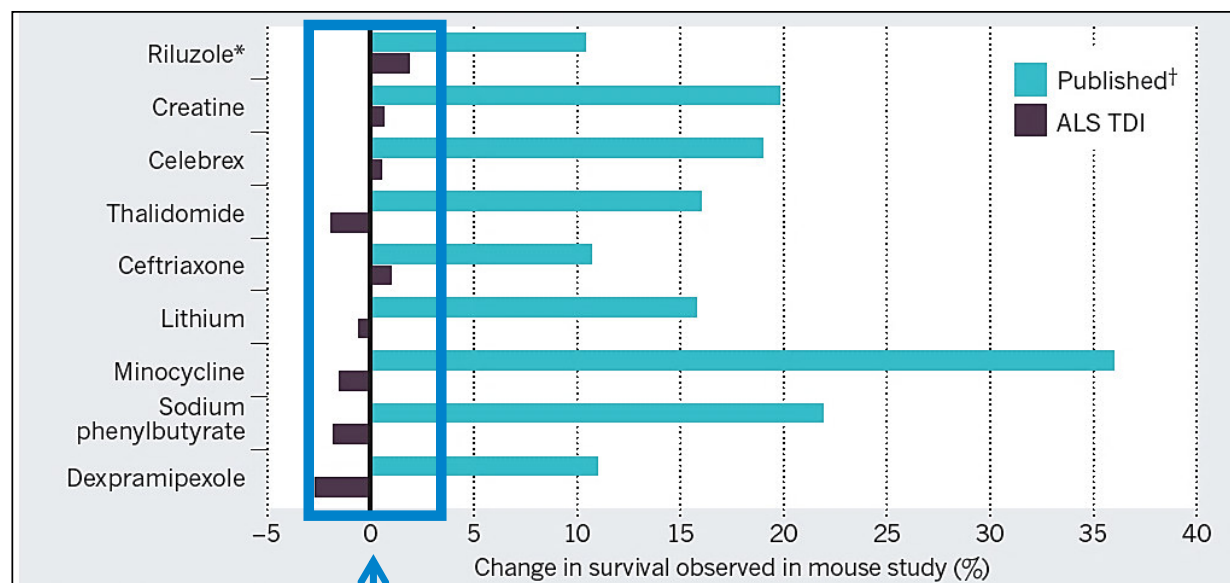
# Viés selecção/performance/detecção



# Irreprodutibilidade em ciência e falhas translacionais



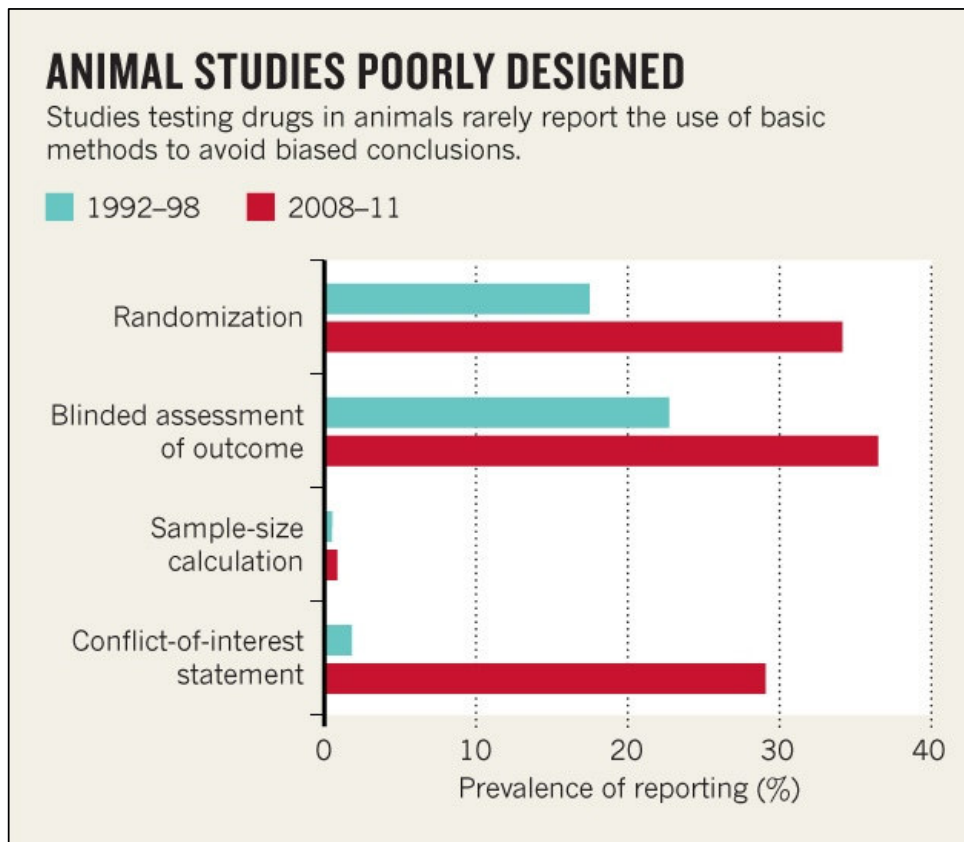
O Amyotrophic lateral sclerosis *Therapy Development Institute* re-testou mais de 50 drogas que tinham tido resultados promissores em modelos animais **mas devidamente!**



Perrin, S. 2014  
*Nature*

**Nenhum efeito encontrado em estudos com potência estatística suficiente, devidamente randomizados e com análise cega dos resultados!**

# Qualidade metodológica de estudos pré-clínicos



*MR Macleod et al (2015) PLoS Biol.  
10.1371/journal.pbio.1002273  
[cited in Cressey, D (2015) Nature News ]*

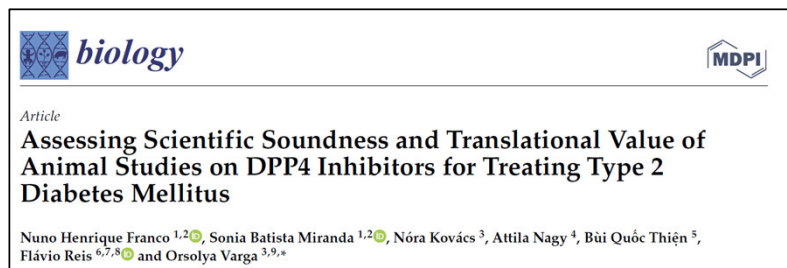
# Qualidade metodológica de estudos pré-clínicos



Fernandes, Franco et al 2019

N=569 estudos publicados em modelos de ALS

- 13.2% reportaram randomização de tratamentos
- 25.7% reportaram cegamento dos avaliadores.
- Apenas 10 (1.7%) justificaram tamanho amostral, quatro deles de forma pouco clara



Franco et al 2021

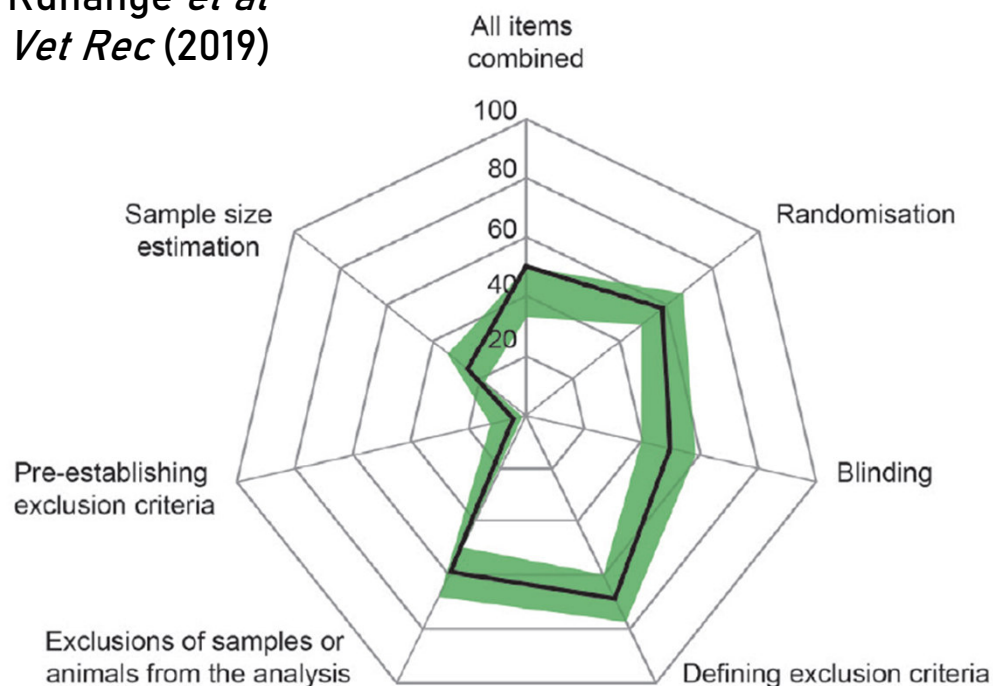
N=124 estudos em modelos roedores de Diabetes Tipo-2

- 51% reportaram randomização de tratamentos
- Apenas em três artigos (2.4%) tamanho da amostra justificado
- Tamanho mediano dos grupos: n=8 ratos, n=10 murganhos

# Qualidade metodológica de ensaios clínicos em medicina veterinária



Rufiange *et al*  
*Vet Rec* (2019)



Revisão sistemática N=120, 10 revistas (2016–2017)

Randomização  
Cálculo tamanho amostral  
Cegamento  
Justificação para exclusão

Requisitos mínimos

Distinção entre *fully reported* e *partially reported*

## Resultados

Mediana de 50% dos itens *fully reported*

Mediana de 18.4% dos itens *partially reported*

Apenas 1 artigo com dados disponibilizados

Apenas 1 artigo com todos parâmetros *fully reported*

# Qualidade metodológica de ensaios clínicos em medicina veterinária



VETERINARY  
ANAESTHESIA  
AND ANALGESIA

RESEARCH PAPER | ARTICLES IN PRESS

Randomization, blinding, data handling and sample size estimation in papers published in *Veterinary Anaesthesia and Analgesia* in 2009 and 2019

Brittany A. Munro • Paige Bergen • Daniel S.J. Pang  

Published: September 28, 2021 • DOI: <https://doi.org/10.1016/j.vaa.2021.09.004>

N=115 artigos

- *Full reporting* de randomização aumentou de 13.6% para 85.7%
- Estimaco de tamanho amostral aumentou de 0% para 20%
- Cegamento e justificaco de excluso de animais no variou significativamente
- Nenhum artigo em 2009 e apenas 2 em 2019 disponibilizavam os dados
- Nenhum artigo reportou os itens minimamente exigveis.

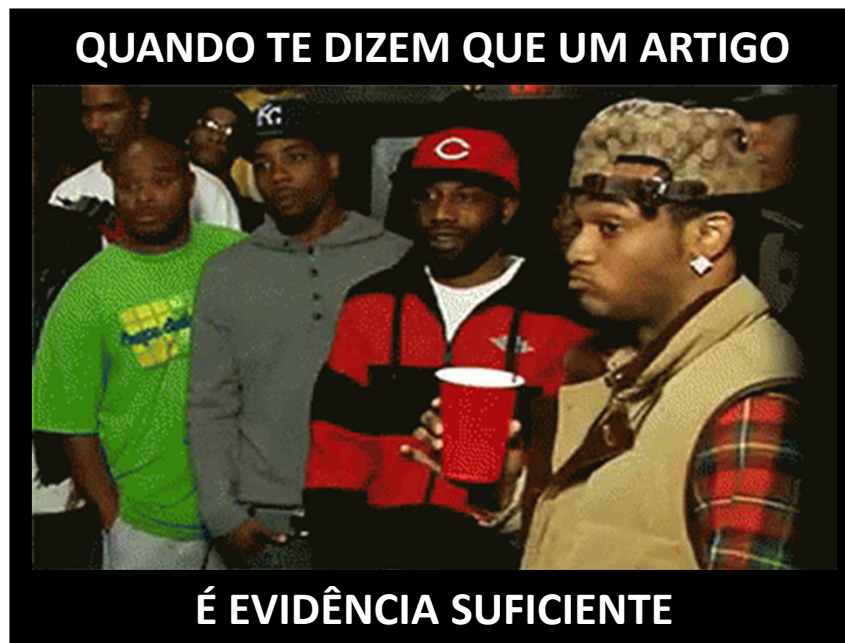
Qualidade metodolgica melhorou, mas mantm-se baixa, com elevado risco de vis

# A reter: o que avaliar num artigo

- Hipótese é clara e bem definida?
- Amostra utilizada
  - Tamanho da amostra reportado? E justificado?
  - Tamanho da amostra suficiente?
  - Há atrição não reportada?
  - Amostragem devidamente descrita e vieses prevenidos?
- Metodologia
  - Randomização dos tratamentos aos grupos é reportada?
  - Ocultação de alocação reportada?
  - Cegamento dos observadores reportado?
  - Unidade experimental devidamente definida?
- Artigo tem uma declaração de conflitos de interesses? aprovação ética?
- Análise
  - Testes estatísticos adequados ao delineamento experimental e dados?
  - Correção para múltiplas comparações?



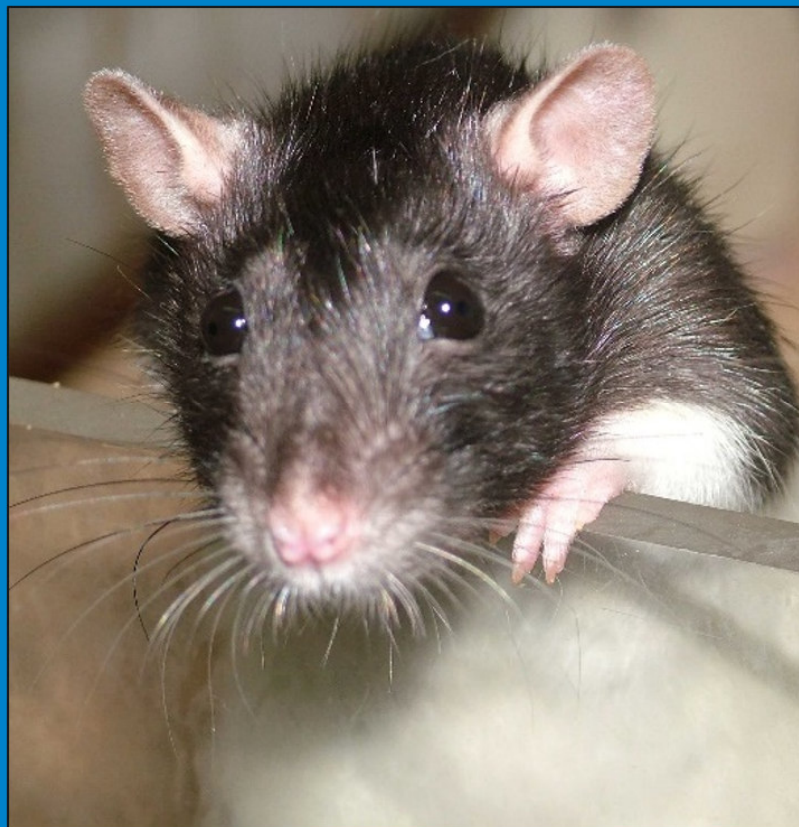
# A reter



- Por regra, não nos devemos basear num único artigo (ou apenas alguns) como evidência suficiente de um fenómeno.
- Independentemente do prestígio dos autores, da revista onde é publicado, das conclusões a que chega, ou da atenção dos *media*.
- Alguns artigos, contudo, são claramente mais fidedignos que outros.



Obrigado pela vossa atenção!



INSTITUTO  
DE INVESTIGAÇÃO  
E INOVAÇÃO  
EM SAÚDE  
UNIVERSIDADE  
DO PORTO



@nuno\_h\_franco



nfranco@ibmc.up.pt